Extraction Without Traces

Ivan A. Sag (Stanford University)
Janet Dean Fodor (Graduate Center, CUNY)*

1. Introduction

Phonetically unexpressed syntactic constituents have been a familiar ingredient of syntactic analyses of extraction for many years. But traceless analyses of long distance filler-gap dependencies have also been formulated, by Gazdar et al. (1984), Steedman (1987, 1988), and Kaplan and Zaenen (1989). In this paper we present a new traceless analysis cast within the framework of Head-driven Phrase Structure Grammar (Pollard and Sag 1987, 1994), which permits a simple and elegant, lexically-based account of unbounded dependency constructions. It provides a basis for dealing with a variety of constraints on extraction, with idiosyncratic lexical extraction 'signatures', and with strong crossover phenomena, and it is compatible with contraction phenomena at 'gap' sites, and with the current psycholinguistic results about how extraction constructions are processed. None of this requires the postulation of phonologically empty constituents of any kind.

Earlier HPSG analyses did assume the existence of WH-trace, which was imported into GPSG, and thence into HPSG, on the basis of a long tradition of 'independent evidence' for empty elements in extraction gaps. But somewhat suprisingly, when one takes stock and examines carefully what the empirical support for WH-trace amounts to, it turns out that there is none. WH-trace has been motivated primarily by

*We thank Peter Sells, Geoff Pullum, and Josef Taglicht for comments on an earlier version of this paper. Our collaboration was supported in part by NSF Grant SBR-9309588 to Stanford University.

theory-internal considerations which, however telling they may be, do not travel well from one theoretical framework to another. For instance, a theory which employs binding principles to constrain extraction gaps will want gaps to contain empty pronominals/anaphors, but this won't be a relevant consideration in a theory that imposes comparable constraints by other means. Also, a theory which posits one kind of empty constituent could find it more parsimonious, in terms of the generalizations expressed by UG principles, to posit **more** than one type of empty constituent. But again, such an argument may have no bite in the context of a different set of theoretical assumptions. This is why special interest has always attached to the theory-independent arguments in favor of WH-trace that have been developed over the years. These could settle the issue about empty constituents, and about many other matters besides, inasmuch as the treatment of extraction is interlinked with that of many other related linguistic phenomena. Evidence for WH-trace, if decisive, would direct theory development towards those frameworks into which empty constituents do fit naturally. The fact that traceless analyses of the extraction facts can be constructed would be a minor curiosity but not of central interest.

The major lines of evidence that have been cited for WH-trace, other than the extraction facts themselves, are: (i) failure of *wanna* contraction over a gap; (ii) failure of auxiliary contraction preceding a gap; (iii) impossibility of floated quantifiers preceding a gap; and (iv) psycholinguistic results. We will address and update the discussion of each of these phenomena. In each case, it turns out that the force of the arguments has declined the longer they have been considered. Our summary will show that the preponderance of evidence at present is either against, or at least neutral with respect to, the existence of WH-trace.[1]

2. Evidence for WH-Trace

2.1. Wanna Contraction

Want and *to* contract to *wanna* in (1)a and (1)b where PRO intervenes (in theories which assume PRO), but not in (1)c, where WH-trace (__) intervenes.

(1) a. Does Kim_i want PRO_i to (wanna) go to the movies?

b. Who_j does Kim_i want PRO_i to (wanna) go to the movies with $\_\__j$?

[1] Our discussion is limited to WH-trace; we do not discuss other empty constituents except as they happen to be relevant to WH-constructions. Note, though, that evidence of types (i)–(iv) has generally been seen as **strongest** for WH-trace, and weaker or non-existent for other empty elements such as PRO or NP-trace.

c. Who$_j$ does Kim$_i$ want $\_\_\,_j$ to (*wanna) go to the movies?

It has been widely accepted that the explanation for (1)c is that WH-trace is visible to the contraction rule which creates *wanna*. The invisibility of PRO (and NP-trace) for purposes of this rule has been attributed to lack of Case: only Case-marked NPs appear in PF – alternatively, NPs not requiring Case are free to appear in other PF-positions where they would not interfere with contraction. The generality of such an account is compromised, however, by the fact that nominative Case-marked WH-trace does not block contraction of an auxiliary onto a preceding verb (Bresnan, 1971; Schachter, 1984).

(2) a. Who$_i$ does Kim think $\_\_\,_i$ is (think's) beneath contempt?

b. Who$_i$ does Kim think $\_\_\,_i$ will (think'll) be late?

c. What$_i$ does Kim imagine $\_\_\,_i$ has (imagine's) been happening?

The contrast between (1)c and the examples in (2) is a serious problem. Chomsky (1986) proposed a rule ordering solution. Empty constituents are deleted at PF; *think's*-contraction applies after they are deleted, but *wanna*-contraction applies before. As Chomsky notes, this would fit well with current theorizing only if the ordering could be shown to follow from general principles, but so far it has not.

It is often supposed that Bresnan's (1971) account of auxiliary contraction can explain this difference between nominative and accusative WH-trace. Bresnan proposed that, unlike the reduced *to* in *wanna*-contraction, the reduced auxiliary in *think's*-contraction is a proclitic; in (2)a, for instance, the *'s* is associated not to *think* on its left but to *beneath* on its right. Then, of course, the WH-trace does **not** intervene between the auxiliary and its host in (2). However, as noted by Lakoff (1972), the reduced auxiliary exhibits phonetic sensitivity to the verb on its left, not to the word on its right. In (2)a the *'s* is unvoiced after */k/*, while in (2)c it is voiced after */n/*; in both examples it is followed by voiced */b/*. Zwicky (1970) noted also that there is sensitivity to the syntactic properties of the left context. For auxiliary contraction following an overt subject there are contrasts such as *You'd enjoy it* but **Kim and you'd enjoy it.* Attempts to save the procliticization analysis (e.g. Klavans 1985) run into other problems; see Fodor (1993) for details. The fact is that Bresnan's proposal was made prior to the introduction of empty constituents into syntactic theory; it works well in a theory without traces, but cannot be adapted successfully to fit into a theory with traces.

Think's-contraction is extremely general; it applies to all matrix verbs which permit a tensed complement clause without *that*, and to all

auxiliaries (with some individual variation with respect to the extent of the reduction). But *wanna*-contraction is extremely limited; it applies to only a handful of matrix verbs that take *to*-complements. Consider *gonna*, *hafta*,**intenna* (*intend to*), **lufta* (*love to*), **meanna* (*meant to*). The arbitrariness of this set suggests that there is no rule, but just a collection of lexical facts. A small number of subject-control verbs alternate with a colloquial form (e.g., *wanna*) which has exactly the same syntactic and semantic properties except that the complement it selects is a base form without *to*. The alternating form no doubt had its historical source in contraction and merger of *to* with the verb, but no synchronic derivation is motivated.

Note that though a verb such as *wanna* is irregular in certain ways, every one of its oddities has some precedent. The lack of *to* is evidenced by modals, *let*, and *make* (*I let him go*, *I made him cry*). The lack of inflection (**He wannas leave*) and the limitation to contexts where inflection is not required (*We wanna leave*,**He wanna leave*) are shared by the exceptional verb *beware* (*Beware of the dog!*,**John bewares the dog.***John beware the dog.*; see Fodor, 1972) and by the 'serial verb' *go* (*Every day I go fetch the paper*,**I have often gone fetch(ed) the paper*). Thus, though verbs such as *wanna* and *hafta* (with its third-singular present form *hasta*) are clearly exceptional, there is no reason to suppose that they are impossible. And the assumption that the English lexicon contains these verbs solves a number of problems that attend the postulation of a **rule** of *wanna*-contraction (e.g. the anomalous regressive assimilation required to derive *hafta* and *hasta* – cf. the progressive pattern of other verb combinations, e.g. *eats [s]*, *craves [z]*). In addition, this lexical analysis predicts the configurational restrictions on the distribution of *wanna* discussed by Postal and Pullum (1982) and references therein.

Now if there is a rule of *think's*-contraction and no rule of *wanna*-contraction, then there remains no evidence for the blocking of any rule by any intervening empty constituent. The reason (1)c is ungrammatical is not that a trace is present, not that a rule is blocked, but just that *wanna* is a subject control verb (as in (1)a and (1)b), and in any case a lexical item cannot contain within it the gap associated with a WH-filler.[2] The apparently contradictory observations that accusative WH-trace does block contraction while nominative WH-trace does not, are now very simply resolved, on the assumption that inter-

[2]It is often claimed that there are 'dialects' where examples like (1)c are grammatical. In such a variety, *wanna* contraction presumably **is** a phonological rule, one that applies only in the configuration: [*want* [*to* VP]], a configuration shared by all of (1)a-c in an analysis (like the one we present below) which countenances neither PRO nor WH-trace.

vening empty constituents are irrelevant to phonological contraction processes. Whether they are irrelevant because they are not visible at PF, or because they do not exist at any linguistic level, cannot be determined on these grounds, but must be decided in some other fashion.

2.2. Auxiliary Contraction

WH-trace has also been held to play a crucial role in the explanation of the observation by King (1970) that the contraction of auxiliaries is prohibited immediately preceding an extraction gap or the site of ellipsis.

(3) a. The butcher is laughing and the baker is (*baker's) too. [VP-Ellipsis]

b. How tall do you think she is (*'s) __ ? [WH-Extraction]

Bresnan's (1971) analysis (see above) assumed that the auxiliary cliticizes onto the following word **prior** to ellipsis or extraction, and thereby prevents the ellipsis or extraction from occurring. More recent treatments within theories that employ empty constituents typically assume that ellipsis/extraction precedes cliticization, and that cliticization onto a phonologically unrealized constituent is impossible. However Selkirk (1984) (also Sells 1983; see also unpublished work by Sharon Inkelas and Draga Zec) shows that these facts follow from known principles of metrical phonology, which account for many other facts besides. As noted by Barss (1993), Selkirk's theory subsumes the think's-contraction phenomenon discussed above, as well as King's observations.[3] We follow Barss's exposition here.

The essence of Selkirk's proposal is that the English phonological rule of Monosyllabic Destressing is subject to a constraint that the grid column to which it applies may not be followed by a silent demibeat. The Monosyllabic Destressing rule applies to function words such as auxiliaries. Once stress is reduced, the auxiliary becomes eligible for vowel reduction, which in turn is the precondition for auxiliary contraction. However, silent demibeats are obligatorily added to the right edge of a metrical grid, as shown in (4)b.

[3] Barss (1993: nt. 24) suggests a syntactic treatment of *think's*-contraction which feeds the prosodic component, and which rests crucially on movement traces. However, one of its premises seems to be incorrect. Barss claims that *Who do you think's available?* is more acceptable than *Who do you think's altruistic?* because *available* is a stage-level predicate while *altruistic* is an individual-level predicate. Consideration of further data suggests, to the contrary, that *think's contraction* is more acceptable when it does not create a sequence of adjacent stresses.

(4) a.

				x	
	x			x	
x	x	x		x	x
x	x	x	x	x	x
the	ba	ker	is	laugh	ing

(the baker's laughing)

b.

	x			
x	x	x	x	
x	x	x	x	x
the	ba	ker	is	

(*the baker's)

Thus auxiliary contraction is impossible at the right edge of a metrical grid, e.g. at the end of a clause. Likewise, auxiliary contraction is blocked in the context of an appositive, as in (5), since appositives are preceded and followed by silent demibeats that inhibit the contraction.

(5) *The baker's, I think, laughing outside.

For present purposes the most important aspect of Selkirk's treatment of these contraction phenomena is that it makes no reference to phonetically empty syntactic entities. Contraction is blocked not by the **presence** of an empty constituent, but the **absence** of any overt constituent following the auxiliary. The descriptive scope of Selkirk's analysis is arguably superior to any treatment based on empty elements (and is certainly no less explanatory), but it is completely neutral with respect to the existence of such elements. Thus these contraction facts do not offer any support for the existence of WH-trace.[4]

2.3. Floated Quantifiers

Floated quantifiers may not appear directly before an extraction site:

(6) a. They (all) were (all) completely satisfied.

b. How satisfied do you think they all were __ ?

c.*How satisfied do you think they were all __ ?

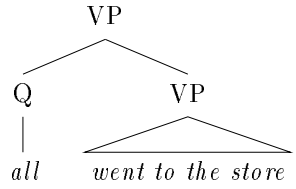
These contrasts have also been explained (Baker 1971, Sag 1980) in terms of WH-trace. The precise nature of the trace-based account of

[4]Selkirk's is not the only prosodically based approach imaginable. For example, an alternative, stress-based account of these same auxiliary contraction contrasts has been suggested by Inkelas and Zec (1993). Certain unresolved issues remain, e.g. the exact nature of the prosodic boundary in examples like *I'll prepare lunch, and you can | dinner* (our thanks to Josef Taglicht for examples). Such cases may warrant appeal to independent principles requiring some degree of stress on remnants in gapping and pseudogapping constructions. However, as far as we are aware, none of the likely approaches to these problems need or would benefit from the postulation of traces.

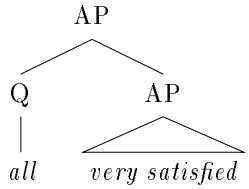
these facts has been a matter of some debate (see Ernst 1983 and the references cited there). But in each analysis, the presence of a WH-trace interacted with some other rule or constraint to ensure that examples like (6)c are never derived. Baker discussed proposals for enriching the grammar to allow traces to be visible in some contexts and invisible in others; Sag's proposal included a surface filter in addition to the traces he assumed.

Interestingly, there is a straightforward account of the data in question that involves no such complexities. Brodie (1983) and Dowty and Brodie (1984) have proposed that floated quantifiers be treated as base-generated adjoined modifiers, as shown in (7).

(7) a.



b.



This analysis thus likens floated quantifiers syntactically to VP-adjoined adverbs.[5] As Dowty and Brodie argue at length, this syntactic analysis provides an appropriate basis for an account of the full range of semantic constraints on the floated quantifier construction. In addition, it provides a solution to certain longstanding problems facing movement-based alternatives, e.g. the fact that *None of the classes have all finished the exam* would have to be derived from a D-structure that should also give rise to **All (of) none of the classes have finished the exam.*

And crucially, though this point is not made by Dowty and Brodie, the treatment of floated quantifiers as adjoined modifiers interacts perfectly with a traceless theory of extraction. If extraction and ellipsis involve no trace (e.g. if (6) involves no AP trace), then there is no node for these elements to adjoin to. The ungrammaticality of examples like (6) above is then a simple and direct consequence of the Brodie/Dowty proposal, just in case the grammar of extraction posits no traces for floated quantifiers or other adjoined modifiers to modify. This is a much simpler and more elegant account than any of the trace-based alternatives in the literature.

[5] There are similar contrasts involving VP modifiers like *(n)ever* (*How satisfied do you think they ever are* __ *?*; **How satisfied do you think they are ever* __ *?*). These pose the same problem as floated quantifiers and are amenable to the same solution.

2.4. Psycholinguistic Results

The experimental data on sentence processing that have been offered as evidence for movement traces indicate that at a gap position, the meaning of the filler (antecedent) phrase is mentally activated. This has been argued for NP-trace by Bever and McElree (1988) and MacDonald (1989), and for WH-trace by Nicol and Swinney (1989), Hickok (1993), Nicol (1993) and others. This work had its origin in previous studies which had shown that overt pronouns and anaphors activate their antecedents during sentence processing. The idea was that if an extraction gap harbors a silent species of pronoun, it should show just the same sort of behavior as an overt pronoun. Antecedent activation does indeed occur in both cases. However, this fact does not suffice to establish that an empty constituent is present in the extraction examples. The experimental findings are **compatible** with the presence of an empty constituent, but no more than that. They are equally compatible with any other linguistic analysis which supports a processing model in which a fronted WH-phrase is semantically integrated into the sentence meaning when the processor associates it with the gap position. This could of course be so even if there is no empty category at that position. In either case, the processor would have to be using local information about the verb and any other arguments or adjuncts associated with it, in order to determine that this **is** the gap position, and to integrate the WH-phrase into the sentential semantics. Thus on almost any model, with or without traces, the relevant processing could be expected to occur at or around the gap position.

What would it take to show, on the basis of sentence processing, either that there is, or that there is not, a syntactic entity in an extraction gap? Fodor (1993) noted that no psycholinguistic finding could even in principle qualify as evidence for empty constituents unless it were established that the data pertained to the **syntactic** processing or representation of the sentence. But for all we know at present, the experimental techniques that have been used to study gap processing to date might be providing information about **semantic** processing only. If so, they would be completely uninformative about traces, because they wouldn't tell us whether the sentence meaning is computed from a trace or from a traceless gap in the syntax. This objection is particularly telling in the case of NP-trace, since its only sturdy experimental support derives from a post-sentential recall task, which almost certainly taps semantic representations in memory rather than syntactic structure or on-line syntactic processing. By contrast, WH-constructions show antecedent reactivation even when the experimental paradigm (cross-modal priming) is one that more plausibly does tap on-

line syntactic processing. But the logical point still stands. Even if the syntactic sensitivity of the cross-modal experiments were definitively established, this would not – **could** not – demonstrate the existence of a syntactic **entity** in the gap. This is because any syntactic activity detected at a gap site would be equally explicable as due to the syntactic processes by which the parser detects the gap, i.e., deduces that a constituent is 'missing' at that point.

Pickering and Barry (1991) have argued recently for the contrary position, that sentence processing facts demonstrate that there is **no** empty element in an extraction gap. The empirical basis for their claim resides in certain indications that WH-phrase interpretation occurs not when the gap site is processed but when the governing verb (or other thematic role assigner) is processed. Often these coincide, as in example (8) from Nicol and Swinney 1989; the role-assigning verb is *accused*, and the gap immediately follows it (assuming canonical word order; see Pickering and Barry for discussion). But the verb and the gap are separated when a prepositional phrase argument of a transitive verb is fronted, or the second object of a double object dative construction. In (9), the role-assigning verb is *gave*, and the gap occurs 34 words later.

(8) The policeman saw the boy that the crowd at the party accused GAP of the crime.

(9) That's the prize which we gave [every student capable of answering every single tricky question on the details of the new and extremely complicated theory about the causes of political instability in small nations with a history of military rulers] GAP.

(10) We gave [every student capable of answering every single tricky question on the details of the new and extremely complicated theory about the causes of political instability in small nations with a history of military rulers] [a prize].

Pickering and Barry note that (9) does not, at least intuitively, exhibit the usual awkwardness that occurs when a long constituent is followed by another constituent, as in (10). No experimental data are available, but informal impressions suggest that the word *which* in (9) is assigned its semantic role when *gave* is received. There is no sign that the processor is attaching any constituent, even a phonologically null constituent, into the tree after the long noun phrase. Pickering and Barry also observe that a sentence like (11), in which two gap sites occur at the end, creating a doubly center-embedded structure, does not exhibit the extreme difficulty normally associated with doubly center-embedded constructions, as in (12). Again, this suggests that no constituents, empty or otherwise, are being attached after *tea* in (11).

(11) John found the saucer [on which Mary put the cup [into which I poured the tea GAP] GAP].

(12) John found the saucer [which Mary put the cup [which I poured the tea into GAP] on GAP].

Pickering and Barry wanted to conclude from these facts that WH-trace is not just unnecessary in extraction gaps but is provably absent.

There are all sorts of details here that might be questioned. But it is more useful to put minor objections aside and focus on the general line of argument, which is novel and interesting. It is not, however, convincing. Gorrell (1993) and Gibson and Hickok (1933) have pointed out, in defense of empty constituents, that the Pickering and Barry facts are equally compatible with a sentence processing device which posits WH-trace whenever it has sufficient information to do so (or at least, to make a reasonable guess), and builds the trace into the correct position in the tree though in advance, sometimes, of encountering words that precede it structurally. Thus a WH-trace could be postulated as *gave* is processed in (9), but the trace could be located structurally to the right of the complex indirect object NP whose details the processor must then continue to fill in. What we have here is a dissociation of the **time** at which the trace is postulated and the **position** in the string at which it is placed. But such dissociation is very common in sentence processing. It has been a familiar assumption for years that predictable tree nodes can be constructed in advance – for instance, that on encountering a nominative NP the processor could construct an associated VP node, though not yet knowing what the VP will consist of or whether other words will intervene first. Pickering (1993) acknowledges this point and retreats to a parsimony argument: the data give 'very good grounds to abandon empty categories, since there is then no **processing** reason to postulate their existence' (p. 175 [Pickering's emphasis]).

In short, it appears that a theory with empty elements has the flexibility to absorb data showing antecedent reactivation at the gap site **or** at the verb, and that only a theory without traces, such as Pickering and Barry's, or the HPSG analysis we present below, actually engages with the psycholinguistic facts. Pickering (p.193) says that his model 'makes some clear empirical predictions. In particular, it should not be possible to find experimental effects of any kind at the assumed gap location.' What little experimental data there are do not support this strong claim, and indeed it is not clear that Pickering's model is or should be committed to it. Nicol (1993) tested sentences such as (13), where four words separate the gap from the verb.

(13) [To which pilot] did [the soldier who was recovering from an ill-

ness in San Francisco] send [some new popular cassettes] GAP on Wednesday morning?

Nicol observed reactivation of the antecedent noun *pilot* both following the verb and at the gap. Though she notes that additional controls are needed to eliminate alternative possibilities, she suggests that a thematic role is tentatively assigned to the WH-phrase when the verb is processed, and that this assignment is checked at the gap site, giving rise to antecedent activation at both points. This seems very plausible. For the processor to check the gap position is essential. At the verb, it can only guess the role of the WH-phrase, and there are many examples in which an initial guess later proves to be incorrect. For instance in (14)a the processor typically assumes that *which book* is the object of *read*, and only later discovers that it must be the object of *from*. In (14)b there is a tendency to guess that *which boys* is the direct object of *introduce* and then be surprised on finding that *introduce* is followed by an overt direct object.

(14) a. Which book did you read to the children from?

b. Which boys did you introduce Sandy to?

Thus, the argument structure of the verb can suggest a **possible** role for the WH-phrase, but that role can be maintained only if there is no other phrase that needs it, and no obligatory role elsewhere in the sentence that lacks lexical realization. This is a fact about sentences, and any efficient processing system must respect it, regardless of whether the grammar that the processor operates with acknowledges WH-trace or not.

Thus the final judgement, at least until or unless other forms of argument can be devised, is that the processing data are neutral with respect to the linguistically important issue of whether a syntactic constituent is or is not present in an extraction gap. There is agreement now on both sides of the debate that antecedent reactivation is not decisive evidence in favor of empty constituents, and also that 'it is not possible to **disprove** the existence of empty categories' on the basis of Pickering and Barry's examples (Pickering, p. 175). The two accounts – no empty categories versus empty categories anticipated by the parser – 'are empirically indistinguishable with respect to the type of data under consideration' (Gibson and Hickok, p. 160). This doesn't mean that there are no facts of interest about filler–gap constructions that psycholinguistic experiments can reveal. For instance, the fact that NP-movement gaps and WH-movement gaps behave differently in the cross-modal priming experiments is potentially of some significance. It suggests that the two kinds of dependency have a different

linguistic status, as is the case in HPSG (see below). Experimantally distinguishing different kinds of linguistic dependency (e.g. syntactic, lexical, or semantic) is valuable and is something that psycholinguistic methods may allow us to do. However, that is not the same as distinguishing syntactic phenomena with empty constituents from syntactic phenomena without empty constituents, which, at the present level of experimental sophistication, we have no idea how to do.

3. An HPSG Extraction Analysis

Following Pollard and Sag (1994: chap. 9), we assume a lexical rule that removes an element from a word's COMPS list, placing that complement instead onto the value (a list) of the feature SLASH.[6] Intuitively, SLASH encodes information about the element(s) that are missing from a phrase (i.e. about the 'gaps' that the 'slashed' phrase contains). This complement extraction lexical rule, shown in (15), thus creates lexical items whose valence is reduced by one complement, but whose SLASH value contains precisely the information associated with the missing complement.

(15) Complement Extraction Lexical Rule:

$$\begin{bmatrix} \text{COMPS} & \langle \ldots, \boxed{1}, \ldots \rangle \\ \text{SLASH} & \langle\,\rangle \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \text{COMPS} & \langle \ldots, \ , \ldots \rangle \\ \text{SLASH} & \langle \boxed{1} \rangle \end{bmatrix}$$

Within a lexical rule, all properties of the input (e.g. semantic role assignment) that are not explicitly modifed remain unchanged in the corresponding output. Thus, the effect of this rule will be to map lexical entries like $hates_1$ into counterparts like $hates_2$:

(16) hates_1 $\qquad\qquad\qquad\qquad$ hates_2

$$\begin{bmatrix} \text{HEAD} & \text{verb[fin]} \\ \text{SUBJ} & \langle \boxed{1}\text{NP[nom]}_{3s} \rangle \\ \text{COMPS} & \langle \boxed{2}\text{NP[acc]} \rangle \\ \text{ARG-S} & \langle \boxed{1}, \boxed{2} \rangle \\ \text{SLASH} & \langle\,\rangle \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \text{HEAD} & \text{verb[fin]} \\ \text{SUBJ} & \langle \boxed{1}\text{NP[nom]}_{3s} \rangle \\ \text{COMPS} & \langle\,\rangle \\ \text{ARG-S} & \langle \boxed{1}, \boxed{2} \rangle \\ \text{SLASH} & \langle \boxed{2}\text{NP[acc]} \rangle \end{bmatrix}$$

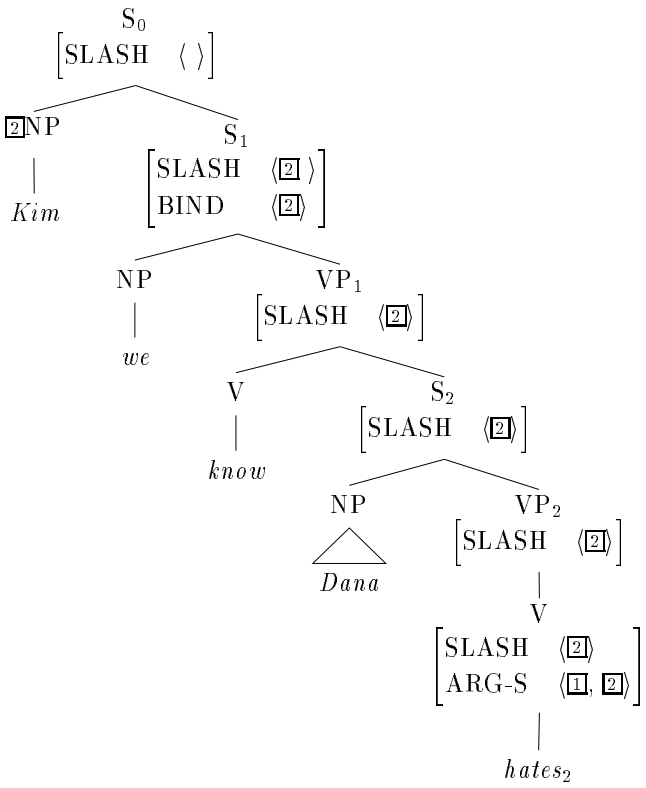
By default, a lexical entry has the empty list as its value for a nonlocal feature such as SLASH. Lexical entries that result from the lexical rule in (15), however, contain non-empty SLASH specifications

[6] For expository convenience, we will treat SLASH values as synsem objects, ignoring the important distinction between LOCAL and SYNSEM values.

that will percolate up the tree until an appropriate binding environment is found in accordance with principles we explicate in a moment. Thus only *hates*$_2$, not *hates*$_1$, may terminate a filler-gap dependency, as illustrated in (17). Through a cascade of identities, the filler at the top of the dependency is linked to the relevant position of the argument structure of *hates*$_2$. Extraction is thus treated entirely in terms of constraint satisfaction, rather than transformational derivation.

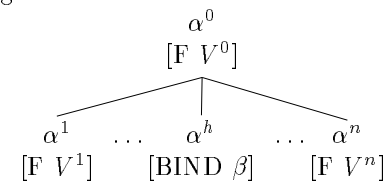
What guarantees that the SLASH value lexically specified on *hates*$_2$ percolates to higher elements in a tree like (17)? Universal grammar contains a principle – the NONLOCAL Feature Principle – which governs the 'percolation' of all NONLOCAL features. Here we restrict our attention to the NONLOCAL feature SLASH, which functions much as it did in the analysis first proposed by Gazdar (1981). The value of SLASH is a list, rather than a category (as it was in Gazdar's analysis), and NLFP defines a phrase's SLASH value in terms of an 'append' operation that collects the SLASH values of the phrase's daughters in left-to-right (first-to-last) order.

(17)



(18) NONLOCAL Feature Principle (NLFP)

For each NONLOCAL feature F in a local structure α^0 whose head-daughter is α^h:



$V^0 = append(V^1, \ldots, V^n) - \beta$

(18) says that the SLASH value of a phrase is the list that results from first appending the daughters' SLASH values – in left-to-right order – and then removing from that list whatever element (possibly none) the head-daughter binds off (specified as the value (possibly empty) of the feature BIND).[7]

To see how this works, let us reconsider the tree in (17), bottom up. (For convenience, empty list values of SLASH and BIND are suppressed; thus, for example, the V node over $hates_2$ has an implicit empty list value for BIND.) Consider VP_2. Since it has only one daughter, the append of all the daughters' SLASH values is identical to that daughter's SLASH value, i.e. is just the list $\langle \boxed{2} \rangle$. And since that daughter is also the head-daughter whose BIND value is the empty list, VP_2's SLASH value is $\langle \boxed{2} \rangle$ minus nothing at all, i.e. just $\langle \boxed{2} \rangle$. The NP sister of VP_2 is specified as [SLASH $\langle\ \rangle$], hence if we consider S_2, we see that the append of its two daughter's SLASH values is again $\langle \boxed{2} \rangle$; and since its head daughter (VP_2) has an empty BIND value, it follows from the NLFP that S_2's SLASH value is also the singleton list $\langle \boxed{2} \rangle$. By similar reasoning, NLFP guarantees that $hate_2$'s SLASH value is passed all the way up to S_1 in (17).

How does SLASH-binding happen in (17)? This is a consequence of the schema in (19) that licenses the highest local tree in (17):

[7] For the sake of uniformity, we treat BIND values as possibly empty lists and – as list difference. In practice, however, the BIND values (like the list values of the feature SUBJ) contain at most one member.

(19) Slash-Binding Schema:

$$X \rightarrow \boxed{2}\text{Filler-Dtr}, \quad \text{Head-Dtr} \begin{bmatrix} \text{SLASH} & \langle \boxed{2} \rangle \\ \text{BIND} & \langle \boxed{2} \rangle \end{bmatrix}$$

The $\boxed{2}$s in (19) represent feature structure identities between the filler-daughter and the head-daughter's SLASH and BIND values. These identities, taken together with the NLFP, cause the SLASH-binding to be just as indicated in (17). The SLASH value of S_0 is the empty list: $\langle \boxed{2} \rangle$ (the append of the filler-daughter's empty SLASH value and the head-daughter's nonempty SLASH value) minus $\langle \boxed{2} \rangle$ (the BIND value of the head-daughter).[8] SLASH percolation thus stops at S_0, a complete sentential constituent.

In the remaining paragraphs, we comment briefly on several features of the extraction analysis just sketched. First, since our analysis (like its GPSG and HPSG predecessors) encodes information about extraction dependencies via the percolation of SLASH specifications, such information is systematically registered on nodes that dominate the extraction site. This in turn predicts that information about the extraction should be available for selection by lexical items that appear at intermediate levels within an extraction dependency. The wealth of evidence that such selection does indeed exist throughout the world's languages (see, e.g. Hukari and Levine 1994) is striking confirmation of this approach.

Second, because our analysis, unlike earlier SLASH-based accounts, treats the value of the SLASH feature as a list (rather than simply a category or a set of categories), the nested dependency condition (Fodor 1978; Pesetsky 1982) can be simply stated within the grammar. That is, the contrast between (20)a,b can be made to follow from the fact that binding like that specified in the lexical entry of an *easy*-type adjective can only access the **first** member of the (infinitival complement's) SLASH value.

(20) a. [Gorbachev's innocence]$_j$, [my friend in Minsk]$_i$ was easy to talk to $\_\__i$ about $\_\__j$.

b.*[My friend in Minsk]$_i$, [Gorbachev's innocence]$_j$ was easy to talk to $\_\__i$ about $\_\__j$.

[8] Specialized instances of (19), which we cannot discuss here, allow an account of various SLASH-binding constraints particular to relative or interrogative constructions. Also, a second clause must be added to the formulation of NLFP in (18) in order to take account of parasitic gaps.

The NLFP ensures that this initial SLASH element (lexically coindexed with the subject of *easy*) will always be associated with the leftmost gap within the infinitival complement (e.g. the missing object of *to* in the phrase *talk to* __ *about* __), forcing the higher binder (the topicalized NP) to bind the rightmost gap.

Third, the lexically based theory of extraction just sketched provides a natural account of lexical exceptions – both positive and negative. For example, the contrast between (21)a,b (first noted by Kayne (1984)) is treated entirely within the lexicon.

(21) a. This student, I can assure you __ to be trustworthy.

b.*I can assure you [this student] to be trustworthy.

The lexical entry for *assure* selects a subject, one object, and a VP complement. *Assure* is also specified as [SLASH ⟨NP⟩], where the the NP on this SLASH list is coindexed with the VP complement's unexpressed subject. This lexical entry can thus only occur at the bottom of a filler–gap dependency. This analysis seems superior to the one considered by Fodor (1992), where *assure* takes an otherwise nonexistent subcategorization frame: two objects (the second marked as phonetically null, i.e. a WH-trace) and a VP complement. And there are apparently varieties of English where the object of certain verbs, e.g. *make* and *let* are unextractable:

(22) a. You let Sandy go home early.

b. %Who did you let go home early?

c. Who did you let Sandy go to visit?

In such varieties, *let* and/or *make* are simply exceptions to the complement extraction lexical rule. Note that this does not mean that extraction is blocked from the VP complement, as in (22)c.

And finally, our traceless extraction analysis is compatible with an account of strong crossover stated in terms of Principle C. The HPSG Binding Theory spelled out in Pollard and Sag (1994: sec. 6.8.3) is stated in terms of a nonconfigurational notion of obliqueness-command, (o-command), which is argued at length to be superior to approaches based on the familiar notion of c-command. O-command relations are defined recursively, and hence extend to the argument structures of verbs within embedded constituents. Thus the highest subject (*we*) in a topicalized structure like (17) is the least oblique element of the highest argument structure (ARG-S list of the highest verb *know*) and *we* o-commands all the elements in the argument structure lists of the

subordinate verbs. Since the filler in (17) is identified with the object in the lowest verb's argument structure, the filler is also o-commanded by the higher subject *we*. Principle C, which bars a nonpronominal from being coindexed with an o-commanding expression, thus rules out the standard cases of strong crossover:

(23) a. *Clinton$_i$, he$_i$ thinks we should support.

b. *Who$_i$ does she$_i$ think we should support.

Because the complement extraction lexical rule removes elements from a verb's COMPS list, but leaves its ARG-S list intact, the argument structure relations necessary for the binding theory account of strong crossover remain intact – without appeal to WH-traces.

4. Summary and Conclusions

Trace theory has gone largely unquestioned within Government Binding theory, and was so well-entrenched that it became a feature also of other theoretical frameworks. However, it is appropriate for any theory to take stock from time to time to see whether its principles and theoretical postulates are continuing to do useful work as the system has evolved. Except for Government Binding theory, it appears that the theory-internal motivation for phonologically empty syntactic entities is minimal. In Combinatory Categorial Grammar, extraction is accomplished through judicious use of combinators, rendering traces quite unnecessary. Similarly, traces have little to contribute to an LFG extraction analysis based on the notion of functional uncertainty.

In HPSG it has turned out on closer scrutiny that the real work of constructing and constraining long distance dependencies is done by the feature SLASH which marks the path from filler to gap; the existence of a syntactic element at the bottom of the path in the gap position is not critical. Because of this, it is very important to note that syntactic theories are not under any **external** pressure to postulate empty constituents in WH-gaps. The arguments for empty constituents that have been based on *wanna* contraction, auxiliary contraction, and floated quantifiers were ingenious, and were the right kinds of arguments to bring to bear on the issue, but as further work has been directed toward them they have simply evaporated. And though psycholinguistic experimentation may one day, as we must hope, be able to reveal the mental representation of extraction constructions more clearly, the current verdict must be that psycholinguistics has not yet succeeded in finding a technique for addressing the theoretically crucial difference between **empty** and **absent** constituents. It is important to pursue all these lines of investigation, and we can't anticipate what interesting

arguments for or against empty constituents might yet emerge. But at present our conclusion is that analyses of extraction must be evaluated on their descriptive and explanatory merits in accounting for the extraction facts; other domains provide no evidence that will decide the matter.

References

Altmann, G. T. M. and Richard Shillcock, eds. 1993. *Cognitive Models of Speech Processing: The Second Sperlonga Meeting*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baker, C. Leroy. 1971. Stress level and auxiliary behavior in English. *LI* 2: 167–181.

Barss, Andrew 1993. Transparency and visibility: sentence processing and the grammar of anaphora. In Altmann and Shillcock, eds. Pp. 401–451.

Bever, Thomas G. and Brian McElree. 1988. Empty categories access their antecedents during comprehension. *LI* 19: 35–43.

Bresnan, Joan. 1971. Contraction and the transformational cycle in English. Unpublished ms. distributed by Indiana University Linguistics Club.

Brodie, Belinda. 1983. *English Adverb Placement in Generalized Phrase Structure Grammar*. M.A. Thesis, Ohio State University.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Dowty, David R. and Belinda Brodie. 1984. The semantics of "floated" quantifiers in a transformationless grammar. WCCFL 3. Stanford: Stanford Linguistics Association.

Ernst, Thomas. 1983. More on adverbs and stressed auxiliaries. *LI* 14: 542–549.

Fodor, Janet D. 1972. Beware. *LI* 3: 528–535.

Fodor, Janet D. 1978. Parsing strategies and constraints on transformations. *LI* 9: 427–473.

Fodor, Janet D. 1992. Islands, learnability, and the lexicon. *Island Constraints: Theory, Acquisition, and Processing*, ed. Helen Goodluck and Michael Rochemont. Boston: Kluwer Academic Publishers. Pp. 109–180.

Fodor, Janet D. 1993. Processing empty categories: a question of visibility. In Altmann and Shillcock, eds. Pp. 351–400.

Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *LI* 12: 155–184.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1984. Foot features and parasitic gaps. *Sentential Complementation*, ed. W. Geest and Y. Putseys. Dordrecht: Foris. Pp. 83–94.

Gibson, Edward and Hickok, G. 1993. Sentence processing with empty categories. *Language and Cognitive Processes* 8, 147-161.

Gorrell, Paul. 1993. Evaluating the direct association hypothesis: A reply to Pickering and Barry. *Language and Cognitive Processes* 8: 129–146.

Hickok, G. 1993. Parallel parsing: evidence from reactivation in garden-path sentences. *Journal of Psycholinguistic Research* 22: 239–250.

Hukari, Thomas and Robert D. Levine. 1994. Adjunct extraction. *WCCFL 12*.

Inkelas, Sharon and Draga Zec. 1993. Auxiliary reduction without empty categories: a prosodic account. *Working Papers of the Cornell Phonetics Laboratory* 8: 205–253.

Kaplan, Ronald and Annie Zaenen. 1987. Long-distance dependencies, constituent structure, and functional uncertainty. *Alternative Conceptions of Phrase Structure*, ed. Mark Baltin and Anthony Kroch. Chicago: U. of Chicago Press.

Kayne, Richard. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.

King, Harold V. 1970. On blocking the rules for contraction in English. *LI* 1: 134–136.

Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61: 95–120.

Lakoff, George. 1972. The arbitrary basis of transformational grammar. *Language* 48: 76–87.

MacDonald, Maryellen C. 1989. Priming effects from gaps to antecedents. *Language and Cognitive Processes* 4: 1–72.

Nicol, Janet L. 1993. Reconsidering reactivation. In Altmann and Shillcock, eds. Pp. 321–347.

Nicol, Janet and David A. Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18: 5–19.

Pesetsky, David. 1982. *Paths and Categories*. Doctoral dissertation, MIT.

Pickering, Martin. 1993. Direct association and sentence processing: a reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes* 8: 163–196.

Pickering, Martin and Guy Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6: 229–259.

Pollard, Carl and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. CSLI Lecture Notes Series No. 13. Stanford: Center for the Study of Language and Information. [Distributed by University of Chicago Press]

Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.

Postal, Paul M. and Geoffrey K. Pullum. 1982. The Contraction Debate. *LI* 13: 122–138.

Sag, Ivan A. 1980. A further note on floated quantifiers, adverbs, and extraction sites. *LI* 11: 255-257.

Schachter, Paul. 1984. Auxiliary reduction: An argument for GPSG. *LI* 15: 514–523.

Selkirk, Elizabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Sells, Peter. 1983. Juncture and the phonology of auxiliary reduction in English. U. Mass. *Occasional Papers in Linguistics. Vol. 8*, ed. Toni Borowsky and Dan Finer. Amherst: GLSA.

Steedman, Mark. 1987. Combinatory grammars and parasitic gaps. *NLLT* 5: 403–439.

Steedman, Mark. 1988. Combinators and grammars. *Categorial Grammars and Natural Language Structures*, ed. Richard Oehrle, Emmon Bach, and Deirdre Wheeler. Dordrecht: Reidel.

Zwicky, Arnold M. 1970. Auxiliary reduction in English. *LI* 1: 323–336.