PROCESSING COMPLEXITY AND FILLER-GAP DEPENDENCIES ACROSS GRAMMARS

JOHN A. HAWKINS

University of Southern California

This article examines crosslinguistic variation in FILLER-GAP DEPENDENCIES (wh-questions and relative clauses) from a processing perspective, and integrates research findings from psycholinguistics, language typology and generative grammar. Numerous implicational universals and hierarchies are proposed that receive a natural explanation in terms of processing and complexity. Filler-gap domains are complex in proportion to their size and in proportion to the amount of simultaneous syntactic and semantic processing that is required in addition to gap identification. They are simplified by making the gap easier to identify and process, or by avoiding a gap structure altogether. When grammatical variation is viewed from this perspective many descriptive insights and implicational patterns can be motivated that have either been stipulated or that have gone unnoticed hitherto. This approach provides an alternative to the assumption of innate parameterized subjacency constraints in this area.*

INTRODUCTION. This article is concerned with the grammar and processing of FILLER-GAP DEPENDENCIES, principally wh-questions and relative clauses, across languages. Examples of the relevant structures in English are:

(1) a. Who do you think that Mary saw?
   b. The person you think that Mary saw is Fred.

In many analyses of these structures, a filler is matched with a co-indexed empty position, or gap, as shown in 1'.

(1') a. Whoi [do you think that Mary saw O]  
   b. the personi [whoi you think that Mary saw O,] . . .

There are severe restrictions on grammaticality in these constructions. Ross (1967) was the first to define constraints of generality, and in the thirty years since then we have seen the growth of a large research literature. More languages have been investigated, various formalisms have been developed for their analysis, and a number of crosslinguistic generalizations have been derived from language samples. There has also been a growing interest in the processing of wh-questions and relative clauses, following Fodor 1978.

When one takes a broad look at these structures across different language types, however, there are universals and patterns of variation that are still not being adequately accounted for. Quite a few languages are counterexamples to general constraints such as Ross’s COMPLEX NP CONSTRAINT (CNPC) and to subsequent reformulations in terms

* I would like to thank the editor of Language, Mark Aronoff, associate editor Johanna Nichols, and two anonymous referees for their helpful comments in connection with the preparation of this article. The material presented here has also benefited from discussions with many individuals, including Elaine Andersen, Joseph Aoun, Chris Barker, Bernard Comrie, Gisbert Fanselow, Gary Holland, Jim Hurford, Larry Hyman, Stig Johansson, Ron Kaplan, Andreas Kathol, Paul Kay, Ed Keenan, Simon Kirby, Robert Klunder, Ekkehard König, George Lakoff, Maryellen MacDonald, John Moore, Masha Polinsky, and Jean-Roger Vergnaud. Presentations at the following institutions provided valuable feedback: UC Berkeley, UC San Diego, USC, the Free University (Berlin), Humboldt University (Berlin), and the Universities of Potsdam, Copenhagen, Oslo, and Reading. None of the individuals I have named necessarily agrees with the theoretical position advanced here. Financial support for this research was provided by the German Science Foundation Innovationskolleg project on Formal Models of Cognitive Complexity at the University of Potsdam in 1996 (DFG grant INK 12/A1), and by a sabbatical leave from the University of Southern California in 1997, both of which are gratefully acknowledged.
of subjacency (Chomsky 1981) and barriers (Chomsky 1986). The ungrammaticality of a sentence like 2 reveals the operation of this constraint in English, but corresponding sentences in Swedish, Norwegian, Japanese and Akan are not ungrammatical (cf. §4.1).

(2) *Which book, [do you know [the professor [who wrote Oi]]]

If constraints like this are intended to be universal, they are too strong. They are also too weak. There are languages that make a distinction between the WH-question of 1a and the relative clause of 1b. In many dialects of German, including the standard language, the direct translation of 1a is grammatical, the direct translation of 1b is not (Kvam 1983).¹

(3) a. Weni [glaubst du, dass Maria Oi gesehen hat]
   who think you that Mary seen has
   ‘Who do you think that Mary saw?’

   the person who you think that Mary seen has
   ‘the person who you think that Mary saw’

The same distinction has been noted for Russian in Comrie 1973. Most current models assume a single parameter setting for a given language, and do not predict this kind of construction-specific difference. More generally, many crosslinguistic patterns in relative clause constructions were noticed by Keenan and Comrie (1977) and led to the formulation of their ACCESSIBILITY HIERARCHY, yet these patterns have been largely ignored in the formal syntax literature. Hierarchically arranged variants do not fit well with a parameterized Universal Grammar (UG) when they do not involve the kinds of bounding nodes (S and NP, etc.) that generative grammarians postulate. There are many other crosslinguistic regularities like this that are not currently being predicted.

There is also a large amount of stipulation, and an absence of motivation or explanation for constraints of generality. Why should complex NPs be less hospitable environments for gaps than other structures? Why are there so-called WH-islands? Why are resumptive pronouns found in place of gaps in some syntactic positions and not others? We are simply told that this is the way things are. It is also apparent that there are numerous factors that contribute to wellformedness in this area—syntactic, semantic, and lexical—yet no unifying principle has been proposed for why all of them should be relevant and for why they combine together in the ways they do to give us the variation we observe.

In this article I am going to pursue a different approach to these problems. Filler-gap dependencies are difficult structures to process. This intuition has been implicit in some of the grammatical literature and it has now been made explicit in models of language processing. I shall argue that this difficulty is directly reflected in the conventions of grammars. Grammars define sets of permissible structures for language use and processing. But, in contrast to Chomsky 1965, I believe that the conventions themselves have been profoundly shaped by processing, and that by adding processing as a central component to our grammatical theory we can build a richer theory: better predictions can be made for grammatical variation and for possible versus impossible grammars; and a lot of what is currently stipulated can be motivated and derived.

Gaps in complex NPs are ungrammatical in many languages because they are hard structures to process. In those languages that permit them, all simpler structures are

¹ The southern German dialects permit more extensive filler-gap dependencies than the northern dialects. The distinction between 3a and 3b is based, inter alia, on native speaker judgements by Germans who are considered by Kvam (1983) to be speakers of Standard German.
predicted to be grammatical. If a gap is avoided in a structural environment in favor of some easier-to-process alternative, such as a co-indexed resumptive pronoun, then all more complex environments should have grammaticalized a simplifying alternative as well, down to the point of complete ungrammaticality. And ungrammaticality in a complex environment implies ungrammaticality in all more complex counterparts. When this intuition is defined explicitly and tested, the variation in filler-gap dependency conventions across grammars turns out to be surprisingly principled and predictable.

1. The Grammar and Processing of Filler-Gap Dependencies. In standard transformational grammar (Chomsky 1965), and in much work derived from it, it has been assumed that the WH-phrase in questions, relative clauses and similar constructions is moved from a grammatically determined original position into the leftmost position of the relevant clause. In government-binding theory (Chomsky 1981), movement leaves a ‘trace’ which is coindexed with the WH-phrase, as shown in 1’a. In the relative clause there is further coindexation between the moved WH-element and the head of the relative, as shown in 1’b. Other models, such as generalized phrase structure grammar (Gazdar et al. 1985), assume no actual movement of the WH-phrase, but maintain the notion of a trace.

In psycholinguistics the terms filler and gap have been used for the moved element and its trace respectively (following Fodor 1978, 1989), and it is assumed that an association is set up between these items in on-line processing. Some processing theories assume instead that an association is made between the WH-phrase and the actual word that subcategorizes for it (Pickering et al. 1994), as shown in 4.

(4) Who$_{i}$ [do you think that Mary saw$_{i}$]

These ‘direct association’ theories are based on trace-free models of grammar such as categorial grammar (Moortgat 1988, Steedman 1987) and head-driven phrase-structure grammar (Pollard & Sag 1993). In what follows I shall incorporate the central insight of filler-subcategorizer theories, by assuming that the subcategorizer must be accessed for subcategorized gaps, but I shall still postulate an empty position, in accordance with filler-gap theories. In this way both subcategorized and nonsubcategorized gap structures can be described by a shared mechanism (the gap), even though the manner of its activation is different. A subcategorized gap is activated by a lexical cooccurrence frame, a nonsubcategorized gap by a phrase structure cooccurrence possibility sanctioned by general syntactic rules. One consequence of this is that the processing domain for filler-gap identification may not need to extend to the gap itself when there is a subcategorizer, but can generally be shorter and more efficient as a result of the direct association. It is not significant here whether movement or base generation of the WH-phrase is assumed. In both cases a filler in surface structure must be matched with a subcategorizer or its gap at some distance from the filler.

Numerous issues involving the processing of these dependencies have not yet been resolved in the psycholinguistic literature (cf. Goodluck & Rochemont 1992 for a useful summary). One such involves the precise time course at which island information becomes available in parsing (compare Pickering et al. 1994 and Traxler & Pickering 1996). But there appears to be a consensus on the following basic point. Filler-gap dependencies are difficult structures to process, and they are characterized by a heightened processing load and a constant effort to relate the filler to its appropriate gap. Identifying the gap is not easy. It is an empty element with no surface manifestation and its presence must be inferred from its immediate environment. At the same time
the filler must be held in working memory, all the other material on the path from filler to gap must be processed simultaneously, and the gap must be correctly identified and filled. This basic insight has been central to psycholinguistic theorizing ever since Wanner and Maratsos (1978) provided experimental evidence for a measurable processing load within a filler-gap domain. Numerous subsequent experiments have refined it, and there has also been neurolinguistic support using event-related brain potentials or ERPs (Kluender & Kutas 1993a, b, King & Kutas 1992, 1993).

One important finding is that there appears to be a first resort strategy in parsing. A gap is postulated as soon as it can be and is filled with the filler, thereby relieving the strain on working memory. Consider 5.

(5) Which student did you ask Mary about?
(5') Which studenti [did you ask (Oi) Mary about Oi]

There are two possible gap sites for which student in the on-line parse: immediately after ask and after about. A slow reading of 5 confirms that which student is first interpreted as the direct object of ask, i.e. it is assigned to the first subcategorizer encountered. This interpretation is then revised as soon as Mary is parsed, since Mary has to be understood as the direct object, and which student is returned to its unfilled filler status until the preposition about is encountered, to which it can be assigned as a complement. Clifton and Frazier define this parsing process as the active filler hypothesis.

(6) Active Filler Hypothesis: When a filler of category XP has been identified in a nonargument position, such as COMP, rank the option of assigning its corresponding gap to the sentence over the option of identifying a lexical phrase of category XP. (Clifton & Frazier 1989:292)

As an illustration of the kinds of processing effects that this predicts in on-line experiments, consider the following sentences from Stowe 1986.

(7) a. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

b. My brother wanted to know [who], Oi will bring us home to Mom at Christmas.

c. My brother wanted to know [who], Ruth will bring (*Oi) us home to Oi at Christmas.

Stowe found that subjects took significantly longer to process us in 7c than in 7a or 7b. This is explained as follows. The parser predicts a gap in the object position of bring in 7c and fills it with who. This filled gap is then immediately contradicted by the presence of the overt pronoun us, leading to a reanalysis and longer processing times. Exx. 7a and 7b involve no such reanalysis.

Summarizing, a filler must be held in working memory until its corresponding gap has been identified. As words in the path from filler to gap are processed there appears to be a constant effort to find the gap, relate the filler to it, and thereby unload the filler from working memory by resolving the uncertainty over gap location. Filler-gap dependencies are complex structures to process.

My primary interest in this context is in the grammatical conventions that distinguish permissible from impermissible filler-gap dependencies in different languages. Have grammars responded to this complexity? What kinds of structural properties could have this effect, and do we find such properties being conventionalized?

I will argue that crosslinguistic variation patterns reveal the impact of processing in three ways. First, processing is easier the more reduced, syntactically and semantically,
a filler gap domain is. In this way there is less additional processing simultaneously with gap identification, and a more constrained search space. Reduction motivates hierarchies of gap-containing environments, with grammars cutting off and with progressively fewer exemplifying languages down each hierarchy. Second, grammars can conventionalize ways of making gap identification and processing easier, for example by positioning the filler so that it will be encountered and processed prior to its gap site. Third, grammars can avoid a filler-gap structure altogether by conventionalizing structural alternatives, such as resumptive pronouns. These preferences are in partial opposition to the expressive motivation for having moved wh-words and head nouns that are semantically modified by relative clause sisters containing a grammatical position co-indexed with the head.

2. **Filler-gap domains.** Let us begin by defining a filler-gap domain, so that we can pick out that portion of a sentence that is relevant for the resolution of a filler-gap dependency.

(8) **Filler-Gap Domain** [FGD]: An FGD consists of the smallest set of terminal and nonterminal nodes dominated by the mother of a filler and on a connected path that must be accessed for gap identification and processing; for subcategorized gaps the path connects the filler to the gap’s subcategorizer and includes, or is extended to include, the gap’s dependent and disambiguating arguments (if any); for nonsubcategorized gaps the path connects the filler to the gap site; all constituency relations and cooccurrence requirements holding between these nodes belong to the description of the FGD.

The reason for defining the path of the FGD from filler to subcategorizer, when there is one, is that the subcategorizer activates a gap site in the position of its subcategorized phrase. It also assigns properties to the gap, such as case and θ-roles, and is responsible for the different ‘thematic entailments’ of Dowty 1991 and Primus 1999. Hence it is not necessary to activate the gap itself in order to process the filler-gap dependency: all the properties of the gap are dependent either on the filler or on the subcategorizer, both of which are in the FGD. This direct association of filler and subcategorizer generally results in more efficient domains. If FGDs were defined from filler to gap in a VOS language like Malagasy, which has head-initial relative clauses, then relativization on a subject (which is the only position relativizable in this language) would have the largest possible FGD with the gap being furthest from the filler, clearly a counterintuitive result. But if the filler is connected to the subcategorizer of the gap, the verb, then the path from the head of the relative to the verb is short and efficient. Similar considerations apply to the mirror-image situation in SOV languages with prenominal relatives, as in Japanese. Subject relatives would again have the largest possible FGDs if the domain were defined from gap to filler, whereas the clause-final verb is adjacent to the head. Most relativizations in English, an SVO language, will also have smaller FGDs when the FGD is defined from filler to subcategorizer.

In some structures and languages (a minority) the FGD will be lengthened by a direct association with the subcategorizer, for example if the filler precedes the gap which in turn precedes its subcategorizer. Such a situation arises in Dutch, for which Frazier 1987 and Frazier & Flores d’Arcais 1989 have presented evidence that subject-object gap ambiguities can be resolved in favor of subject assignments to the gap prior to a clause-final lexical verb. But the θ-roles assigned to NP sequences in Dutch and their precise thematic entailments are determined by the verb and admit of various combinatorial possibilities (see Primus 1999 for illustration from numerous languages), whether these NPs are overt or not, so that processing of the verb is still required for full gap interpretation, which supports the definition in 8.
Since the path for a subcategorized gap is defined from filler to subcategorizor in 8, and since most gap sites stand outside the FGD so-defined, I shall assume, for consistency, that any gap that does occur on the path is not included in the node count and that the FGD is limited to overt constituents only (i.e. those that are actually perceived and processed) and to the dominating structure they project to. For nonsubcategorized gaps FGDs are defined from filler to gap site, but the gaps themselves will again be removed from the node count, for similar reasons.

An FGD is also defined to include a gap’s dependent arguments and disambiguating arguments (if any). Dependent arguments are those upon which a gap depends, for syntactic or semantic property assignments. A direct object in a language like English requires the cooccurrence of a subject, and is asymmetrically dependent on it syntactically and semantically (Dowty 1991, Primus 1999, Hawkins 1994:50–56). A subject, by contrast, can occur with one-place predicates, and it is not dependent on a direct object in two-place predicates. The FGD for a direct object gap will therefore always include, or be extended to include, an overt subject. The motivation for this is similar to that for including the subcategorizor: a direct object gap requires access to the subject for full interpretation and processing. It is, for example, thematically dependent upon the subject and the verb for its precise interpretation (Dowty 1991, Primus 1999). In English the SVO word order guarantees that an overt subject will always be in the FGD for an object gap, whereas an overt object will not be in the FGD for a subject gap. Compare books, [that professors write O], with professors, [that O, write books] (FGDs are italicized). In a VSO language, however, the FGD for the object gap will be extended to include the overt subject, books, [that write professors O], whereas the FGD for the subject gap need not include the overt object, i.e. professors, [that write O, books]. The intuition behind this is that the FGD in professors that write is complete and processable, even when the subcategorizor is transitive and takes an object. The head of the relative, professors, is recognizable as the subject and agent of write and does not depend on the object for any syntactic or semantic property assignments. By contrast, although the head, books, is recognizable as the object and patient of write in books that write, the processing of the object is incomplete without the subject, and one is left wondering who performed this action and brought the books into existence (on account of their thematic dependence on the agent).

There is now extensive psycholinguistic support for the greater difficulty of relativizing on a direct object versus on a subject (see §3.2), and this is consistent with 8: the FGD for subject relatives is smaller and less internally complex than for object relatives. Even languages with identical surface word orders for both show the subject preference. Frauenfelder et al. 1980 and Holmes & O’Regan 1981 have compared subject relatives and stylistically inverted object relatives in French, e.g. la femme [qui connaît ma mère] (‘the woman who knows my mother’) and la femme [que connaît ma mère] (‘the woman that my mother knows’). These constructions use the same RelPro-V-NP order and differ only in the form of the relative pronoun (qui versus que) and in the grammatical relations assigned to the identical string (SVO versus OVS). They found that the subject relatives were still easier to process than the object relatives. According to 8, the direct object relatives will contain a dependent overt subject in their FGDs, whereas the subject relatives will not contain an overt object.

There are circumstances, however, in which a gap’s nondependent arguments must also be included in the FGD, namely when the filler and the subcategorizor alone are ambiguous between a subject gap and an object gap interpretation. Depending on case marking (and verb agreement) the sequence cats that saw rats in a VSO language could
be assigned a subject relative interpretation \([\text{cats}, \text{that saw O, rats}]\), or an object interpretation \([\text{cats}, \text{that saw rats O}]\). In such cases the FGDs for both interpretations will include the disambiguating overt NP within the relative clause, since this NP is required for gap identification and processing. In the mirror-image verb-final structure of Japanese, \([(\text{NP, O}) \text{ V} \text{NP}_1]\), the FGD will similarly need to include a disambiguating lower NP.

The definition in 8 states that all constituency relations and cooccurrence requirements holding between nodes in the FGD belong to the description of the domain. Constituency relations can be stated in terms of immediate domination (ImD), from which further structural relations such as domination, sisterhood and c-command can be derived.\(^3\) Cooccurrence requirements (RQ) will principally be those defined by verbs and prepositions and may be syntactic (e.g. transitivity versus intransitivity), morphosyntactic (e.g. case assignment), or semantic (e.g. \(\theta\)-role assignment).

Consider an illustrative FGD for the relative clause in 1b whose trace-enriched surface structure is shown in 9. Phrasal groupings are limited to so-called maximal projections, or major distinct phrasal types, without regard for any further phrase-internal branching, and employ traditional category and constituency labels. The FGD for 9 is given in \(9'\).\(^4\)

\[
(9') \quad \text{filler} = [\text{person}]_{\text{Ni}} \\
\text{mother of filler} = \text{NP}_1 \\
\text{subcategorizer} = [\text{saw}]_{\text{V}_2}
\]

\(^3\) If a node A ImD (immediately dominates) B, and B ImD C, then A dominates C. If A ImD both B and C, then B and C are sisters. If A ImD B and C, and C dominates D, then B (asymmetrically) c-commands D. And so on.

\(^4\) Where there are several instances of a category or phrase within the FGD, a numbered index is assigned, \(\text{NP}_1, \text{NP}_2, \text{etc.}\), reflecting the temporal order of the on-line parse.
The filler for relative clause constructions is assumed here to be the head noun \textit{person}, rather than the \textit{wh}-pronoun that is coindexed with it, in contrast to \textit{wh}-questions in which the \textit{wh}-word is the filler. Most types of relative clauses across languages do not employ a \textit{wh}-pronoun but instead prefer a subordination marker such as \textit{that}, a nonfinite verb form, or no surface marking at all (Lehmann 1984). The only possible surface filler in such cases is the head of the relative clause, and the surface evidence suggests a deletion rather than a movement analysis for the gap. English employs the \textit{wh}-strategy in a proper subset of its relative clauses, and for reasons of both internal and crosslinguistic consistency I shall regard the nominal head as the filler of the FGD in all cases. The pronoun \textit{who} will be referred to as a FILLER COPY within the relative clause which, by being fronted to a position adjacent to the head, leaves a gap in the position relativized on, just as the deletion strategy leaves a gap. The particular types of cooccurrence requirements imposed by \textit{think} and \textit{saw} are left deliberately vague in 9', i.e. their theta-role and case assignments, and so on. These details need not concern us at the moment.

Notice that only the nodes on the path from filler to subcategorizer are included in the FGD. If the relative clause of 1b had contained additional material to the right of \textit{saw}, e.g. the person who you think that you saw yesterday, then yesterday would not have been in the FGD. The determiner \textit{the} occurs to the left of the filler in 9 and is also not included, though different constituency assumptions could have grouped \textit{the} with \textit{person} under a separate NP immediately dominated by NPI, this whole NP then being the head and filler. I take no stand on this constituency issue here.

3. REDUCE FGD SIZE: A CLAUSE-INTERNAL HIERARCHY. If filler-gap constructions make more demands on working memory, and involve more simultaneous processing decisions and a constant effort to locate the gap, then we should expect performance data to show a preference for simpler rather than more complex instances of this difficult structural type. One way to simplify is to reduce the set of nodes in the FGD. A smaller FGD means that the filler remains active for a shorter period, and there are fewer items to process simultaneously with gap identification. This preference is defined in 10.

(10) \textbf{Minimize FGDs:} The human processor prefers FGDs to be as small as possible.

Minimization can be at variance with the demands of expressive power, since more complex relativizations like 1b above are sometimes needed for referent identification. The expressor and the processor are in competition in these cases. The preference in 10 may also be in competition with other processing preferences, if certain additional nodes contribute directly to the ease of gap identification (§6.4). All things being equal, however, FGDs should be as small as possible.

3.1. DOMAIN MINIMIZATION IN LINEAR ORDERING UNIVERSALS. In Hawkins 1990, 1994, and 1997 I have argued that linear ordering universals provide direct evidence for node reduction and minimality, and for a correlation between performance and grammars. Some orderings reduce the constituent recognition domains for their containing phrases, and it is systematically these orders that are preferred in performance when languages
have a choice, and in grammars when one order is fixed. For example, the Greenbergian cross-category universals define a preference for head-adjacent orderings such as $\text{vp}[V \, pp[P \, NP]]$ and $[[\text{NP} \, P] \, \text{pp} \, V]\text{vp}$, as opposed to $\text{vp}[V \, [\text{NP} \, P] \, pp]$ and $[\text{pp} \, [P \, NP] \, V]\text{vp}$. Assuming that $V$ projects to, or constructs, a VP, and $P$ a PP, then just two words, $V$ and $P$, suffice in the head-adjacent orders for the recognition of these two immediate constituents of the VP, whereas all the terminal categories dominated by VP have to be viewed in order to derive the same constituency information when the heads are nonadjacent. The adjacent orders result in smaller constituent recognition domains and have higher and more efficient ‘IC-to-word ratios’. In multiple branching structures such as NPs that contain an adjective, a head noun and a relative clause, e.g. intelligent students [that I teach at USC], we see a similar massive preference across languages for the relative adjacency of the adjective, the noun and whatever category constructs the relative clause (here the subordinator that), both in head-initial and head-final languages. The principle that defines these ordering preferences is called EARLY IMMEDIATE CONSTITUENTS (EIC), see Hawkins 1994:ch. 3.

When complexity is permitted, it appears to be hierarchically organized, with grammaticality cutting off in different languages at fixed and implicationally arranged points. For example, it has been well known since Miller & Chomsky 1963, Bever 1970, and Kimball 1973 that center embeddings are hard structures to process. What is less well known is that there are hierarchies of tolerance for center embeddings of different degrees of complexity in the grammatical conventions of languages. In the center-embedding environment between a preposition and a head noun within NP, i.e. $pp[P \, NP[\ldots \, N]]$, we find the following implicational pattern: if a language permits a prenominal relative clause, it permits the typically shorter and less internally complex prenominal possessive phrase; if it permits the possessive phrase, it almost always has a prenominal single-word adjective. As the complexity of the center embedding increases, in the form of longer distances between $P$ and $N$ (resulting in lower IC-to-word ratios for the recognition of PP), the number of attested languages declines. And if the grammar permits a more complex center embedding in this structure at all, it permits the less complex counterparts as well.

For filler-gap dependencies, which are also hard to process, this generalization suggests the following hypothesis:

\[11\] **Filler-Gap Complexity Hypothesis:** If an FGD of complexity $n$ on a complexity hierarchy $H$ is grammatical, then FGDs for all less complex variables on $H (n-1)$ will also be grammatical.

In order to test this we need to first motivate hierarchies of complexity for FGDs.

### 3.2. The Accessibility Hierarchy

I have argued (Hawkins 1994:37–42) that the ACCESSIBILITY HIERARCHY (AH) of Keenan and Comrie (1977, 1979) involves increas-

---

5 In the calculation procedure of Hawkins 1994:69–83 an efficient ordering such as $\text{vp}[went \, pr[to \, the \, movies]]$ would have an IC-to-word ratio for the VP of $\frac{2}{2} = 100\%$. Two ICs, $V$ and $PP$, are recognized on the basis of a parse string of just two words, went to. An inefficient ordering such as $\text{vp}[went \, [the \, movies \, to] \, pr]$ would have a lower ratio of $\frac{2}{4} = 50\%$, since recognition of the same two ICs would require access to all four words and the recognition domain would need to proceed from the verb went to the postposition to constructing the PP.


7 See Hawkins 1994:315–21 for data and discussion on this center-embedding hierarchy for prepositional phrases and on other center-embedding hierarchies.
ingly complex domains for relativization, measured in terms of the number of nodes and structural relations that need to be computed in order to match the relative clause head with the position relativized on.⁸

(12) **Accessibility Hierarchy** (AH): SU > DO > IO/OBL > GEN

Keenan and Comrie proposed that relativization becomes more difficult the lower one goes down the AH, and that certain languages cut off at each of the implicationally permitted points.⁹ Their precise predictions are formulated as a set of hierarchy constraints, defined in terms of relative clause ‘strategies’. Their data can be explained in terms of the relative complexity of the positions relativized on. This claim is not straightforward, given the differences between languages in the surface coding of these positions, i.e. in their case marking, linear ordering and syntactic configuration. Nonetheless there is a clear general correlation between the number of dominating and cooccurring nodes required for relativizable positions and their AH ranking. Subjects generally occupy the highest argument position in a tree and c-command other NP positions (generally asymmetrically). A direct object requires the cooccurrence of a (c-commanding) subject and is generally dominated by VP as well as S. FGDs for object gaps will contain overt dependent subjects, as discussed in §2, whereas FGDs for subject gaps need not contain overt objects. An indirect object is regarded in models of relational grammar (Blake 1990) as the third argument of a three-place predicate, requiring the cooccurrence of both subject and object, and in the present context I will assume that overt subjects and direct objects are required in FGDs with indirect object gaps.¹⁰

⁸ Keenan and Comrie (1977, 1979) formulated the AH as follows: SU > DO > IO > OBL > GEN > OCOMP. I have excluded OCOMP since the coding of this position is highly variable across languages (Stassen 1985), and since Keenan and Comrie do not have systematic data for it in their fifty-language sample. The version of the AH given in 12 is the one used in Comrie 1989, with IO and OBL positions collapsed. This collapsing is also supported by the complexity metric proposed here, which does not distinguish between them.

⁹ This intuition was captured more straightforwardly in Keenan & Comrie 1972, prior to the revisions of the 1977 paper. Sample languages cited in Keenan & Comrie to support the implicational structuring of the cutoff points for relativization crosslinguistically (i.e. aggregating over all strategies) are:

- SU only: Malagasy, Maori
- SU & DO only: Luganda, Indonesian
- SU & DO & IO only: Basque
- SU & DO & IO & OBL only: North Frisian, Catalan
- SU & DO & IO & OBL & GEN: English, Hausa, Kera

¹⁰ The asymmetric dependency of an indirect object on a subject is clear. Indirect objects require a cooccurring subject, are typically asymmetrically c-commanded by it and thematically dependent on it (Dowty 1991, Primus 1999). The relationship between indirect and direct objects is more complex. Indirect objects typically require a cooccurring direct object, whereas direct objects may or may not require an indirect object. But depending on the coding of indirect objects as case-marked NPs or as PP, there are different c-command (and ordering) relations with a direct object (Primus 1997), and indirect objects have higher thematic roles than direct objects on the thematic hierarchy of Primus 1999. They have lower cases on Primus’s 1993, 1999 case hierarchies, however. From the processing perspective of the present article the obligatory cooccurrence property is the most important one since it affects the number of overt NPs that are assumed to be in an FGD, and hence it impacts the latter’s size as defined in 8 above. The intuition we wish to capture is that a three-place predicate in an FGD with a direct object gap that lacks an overt indirect object is more complete and processable than an FGD with an indirect object gap that lacks an overt direct object. Compare essays [that professors assigned O₁ (to students)], in which gap identification and processing are successful without to students, with essays [that professors assigned (essays) to O₂], in which access to essays is required. Obviously there will be lexical variation among three-place predicates in terms of how obligatory the overt direct object is (students that professors wrote (…) to is OK), but the point is that the FGD with the direct object gap is always complete and processable without the overt indirect object, and this supports the asymmetry assumed here.
Indirect objects are not readily distinguishable from obliques in terms of the complexity metric proposed here, however, and so the two positions are collapsed. Oblique NPs are often embedded within a PP, which gives them extra depth, but they differ from indirect objects in not generally requiring a direct object, though they do generally require a subject. Finally, a genitive NP is dominated by a possessive phrase within a higher dominating NP in a variety of structural positions, the aggregated complexity of which puts the genitive at the bottom of the hierarchy.

We can illustrate this by calculating the minimal FGD sizes for relativizations on each of the AH positions. Minimal FGDs contain the smallest set of nodes that are grammatically required for relativization on each position, including any overt dependent arguments but not disambiguating arguments (whose presence reflects lexical idiosyncrasies of the subcategorizer). Since all the nodes selected are obligatory for each relativizable position, differences between FGD sizes should remain constant when optional phrases are added. Assume that direct (NP_o) and indirect (NP_i) objects are VP-dominated, with the latter requiring the cooccurrence of the former, both requiring a subject (NP_s). Assume that oblique NPs (NP_obl) are VP- and PP-dominated and require a subject. Assume that subjects are VP-external, and that genitives (NP_g) can occur embedded within each NP position. These configurational assumptions are illustrated in 13, which collapses all the relativizable positions into a single tree representation and which does not assume any additional wh-movement within the relative clause (i.e. no relative pronouns, simply a complementizer or other subordination indicator which has been omitted here since it is constant across all the positions). The linear orderings of 13 assume an SVO clause structure and a noun-initial NP, but exactly the same relative ranking of positions emerges from other orderings when FGDs are calculated in accordance with these assumptions. Terminal nodes are omitted for convenience.

(13)  
```
NP
  N
  S
  NP_i
    N_i
    PossP
      Poss
      NP_g
    PossP
      Poss
      NP_g
    NP_o
    PossP
      Poss
      NP_g
    N_o
    PossP
      Poss
      NP_g
    N_o
    PossP
      Poss
      NP_g
    N_obl
    Poss
    NP_g
```  

For discussion and illustration of this point, see Hawkins 1994:38–39.
The minimal FGD for each relativizable position will include (according to 8) the filler N, the subcategorizer of N’s gap (V, P or Poss), any overt arguments on which the gap is dependent, and all nodes dominated by the mother of N (NP) that are required for grammaticality and that are on the path from N to the rightmost constituent of the FGD (excluding the phrase dominating the gap itself). For relativizations on a genitive I have assumed that the containing SU, DO, and so on, exhibits the same asymmetrical cooccurrence and dependency requirements that I have discussed for gaps, and that the subcategorizing verb is required for a containing subject, since otherwise that subject is unprocessable, cf. §6.4 (GEN-SU stands for relativization on a genitive within a subject NP, GEN-DO for relativization on a genitive within a direct object, etc).

(14) Minimal FGDs for relativizations on:

SU = 5 \{N, NP, V, VP, S\}
DO = 7 \{N, NP, N_o, NP_o, S, V, VP\} (requires SU)
IO = 9 \{N, NP, N_o, NP_o, S, V, VP, N_o, NP_o\} (requires SU & DO)
OBL = 9 \{N, NP, N_o, NP_o, S, V, VP, P, PP\} (requires SU)
GEN-SU = 9 \{N, NP, N_o, NP_o, S, Poss, PossP, V, VP\}
GEN-DO = 11 \{N, NP, N_o, NP_o, S, V, VP, N_o, Pos, PossP\} (requires SU)
GEN-IO = 13 \{N, NP, N_o, NP_o, S, V, VP, N_o, N_o, NP_o, Poss, PossP\} (requires SU & DO)
GEN-OBL = 13 \{N, NP, N_o, NP_o, S, V, VP, P, PP, N_obl, NP_obl, Poss, PossP\} (requires SU)

As 14 indicates, the number of nodes increases down the AH, and as the nodes increase there are more structural relations and cooccurrence requirements to compute and more morphosyntactic and semantic operations that apply, such as case assignment, \(\theta\)-role assignment and thematic dependency computations. There will also be more terminal nodes to process in larger domains, with more words to recognize and process phonologically and morphologically. Node quantity is therefore just one index of the relative complexity of an FGD, but it is a fundamental one with numerous correlating properties that involve additional processing operations.

Performance support for the AH comes from a repetition experiment conducted by Keenan and S. Hawkins (1987) on English speakers, both children and adults. Their prediction was that repetition accuracy would correlate with positions on the AH. Their data, shown in 15, bear this out.

(15) Repetition accuracies for relativization

<table>
<thead>
<tr>
<th></th>
<th>SU</th>
<th>DO</th>
<th>IO</th>
<th>OBL</th>
<th>GEN-SU</th>
<th>GEN-DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>64%</td>
<td>62.5%</td>
<td>57%</td>
<td>52%</td>
<td>31%</td>
<td>36%</td>
</tr>
<tr>
<td>Children</td>
<td>63%</td>
<td>51%</td>
<td>50%</td>
<td>35%</td>
<td>21%</td>
<td>18%</td>
</tr>
</tbody>
</table>

The relative ranking SU>DO has been corroborated by a number of studies in the psycholinguistic literature, following the initial experiments of Wanner and Maratsos (1978) supporting the greater processing load of object relatives compared with subject relatives in English. Holmes and O’Regan (1981), for example, found more comprehension errors and more regressive eye movements when relativizing on direct objects (in French). Ford (1983) pointed to longer lexical decision times in English object relatives. King and Just (1991) found lower comprehension accuracy and longer reading times in self-paced reading experiments. Pickering and Shillcock (1992) also found significant
reaction time differences between the two positions in self-paced reading experiments, both within and across clause boundaries. Neurolinguistic support for the difference between subject relatives and object relatives using ERPs has been presented in King & Kutas 1992, 1993. Finally, textual frequencies for relativizations on all the positions of the AH have been quantified in Keenan 1975 and are consistent with it. Keenan's aggregates are: SU = 46%, DO = 24%, IO/OBL = 15%, GEN = 5%.

3.3. \([-\text{Case}]\) GAPS. Two relative clause strategies are considered different in Keenan & Comrie (1977, 1979) if they differ either in terms of the ordering of the relative clause and head, or in terms of case-coding. A strategy is \([+\text{Case}]\) if it contains a nominal element in the relative clause that unequivocally indicates which NP position is being relativized on. This can be done by means of a case-coding relative pronoun, as in Latin or Russian, by pronoun retention (as in the Hebrew example 18 below), or by the presence of a preposition governing the position relativized on (e.g. the man that I gave the book to or to whom I gave the book in English). Relative clauses that contain no such nominal element are \([-\text{Case}]\) (e.g. the man that I know). \([-\text{Case}]\) relatives typically contain a subordinating complementizer or relativizer (English that), a participial verb form (Tamil), or else no special subordination marking apart from adjacency of the relative clause to its head (Japanese). A \([-\text{Case}]\) strategy can also contain a relative pronoun, but it must be case-invariant (such as die in Dutch or who in dialects of English).

These different strategies involve subtly different syntactic operations, with very different consequences for the ease or difficulty of processing. \([+\text{Case}]\) strategies with resumptive pronouns employ a pronoun in place of the gap, so these relative clauses are not filler-gap constructions, and I will return to them in §3.4. \([+\text{Case}]\) strategies with relative pronouns involve both a (displaced) pronoun and a gap and will be considered in §3.5. The \([-\text{Case}]\) languages consistently employ a gap strategy only, so all of these languages are relevant for any predicted differences between grammaticalized gap positions.

The filler-gap complexity hypothesis in 11 predicts that a grammatical gap in any low position on the clause-internal hierarchy of 12 implies the grammaticality of a gap in all higher positions. This follows because structural complexity increases down 12, so if a gap is possible in a low position, it should be possible in all higher, easier-to-process positions as well. For relative clauses this prediction is formulated as 16.

\[(16)\] Relative Clause Gap Hierarchy: If a relative clause gap is grammatical in position P on a complexity hierarchy H, then gaps will be grammatical in all higher positions on H.

This prediction is supported exceptionlessly in Keenan and Comrie's data. All their languages with \([-\text{Case}]\) strategies (I counted 40) limit the gap strategy to all and only the implicationally permitted relativization possibilities in 17.\[^{12}\]

\[(17)\] Languages with \([-\text{Case}]\) relative clause gaps in Keenan and Comrie's data

\[\text{SU only:} \quad \text{Aoban, Arabic (Classical), German, Gilbertese, Iban, Javanese, Kera, Malagasy, Maori, Minang-Kabau, Tagalog, Toba Batak}\]

\[^{12}\] Prediction 16 remains intact when we add the P-subcategorized gaps of Keenan and Comrie's \([+\text{Case}]\) strategies to the V-subcategorized gaps of the \([-\text{Case}]\) languages. Languages like English, N. Frisian, and Swedish now exemplify gap strategies on SU & DO & IO/OBL positions, i.e. this extends their gap strategies down the AH compared with the V-subcategorized gaps, but it does not change the set of relativizable gap cooccurrences given in 17.
Keenan and Comrie’s sample is a convenience one, of course, which does not control for genetic and areal bias. We cannot draw reliable conclusions about relative frequencies from it, therefore. But it does contain enough languages to illustrate all the predicted cut-off points, and since any single language is a potential counterexample it is interesting that no language shows an unpredicted combination of gap positions.

For Keenan and Comrie it is significant that all the [−Case] strategies apply to subjects, making them primary strategies in their terminology, to which the primary relativization constraint applies: if a primary strategy can apply to a low position on the AH, then it can apply to all higher positions. In this way they predict the distribution of 17 and the original intuition of Keenan & Comrie 1972 (if you can relativize low [using any strategy], you can relativize high) is preserved for this subset of the data. For all nonprimary strategies, a continuous segment prediction requires only that any distinct strategy (where distinctness is defined by both [−Case] and ordering) should apply to a continuous segment of the AH.

But why are [−Case] relatives always primary strategies (i.e., why do they always apply to subjects)? This is neither required nor predicted by Keenan and Comrie’s constraints, since the continuous segment prediction allows for any strategy, including those with [−Case], to apply to DO & IO/OBL or to IO/OBL & GEN or to all three or to any one alone, none of which occur. Can we eliminate this overpermissiveness and derive all and only the options of 17? The filler-gap complexity hypothesis of 11 does this, on the assumption that the AH is a complexity ranking of relativizable positions and that the gap strategy is a complex structure for processing. If a gap is permitted in a more difficult environment, it will also appear in all simpler ones, subjects being the simplest. Hence the gap strategy must go all the way to the top of the hierarchy, if it exists at all.

3.4. A REVERSE IMPLICATIONAL STRUCTURING FOR COPY PRONOUNS. Many languages employ a coreferential pronoun in the position relativized on, thereby indicating the role of the head explicitly. Hebrew is an example.

(18) ha-isha, [she-Yoav natan la, et ha-sefer]
the-woman that-Yoav gave to-her DO the-book
‘the woman that Yoav gave the book to’

The presence of this resumptive or copy pronoun can be argued to make the processing of head and relative clause easier, since an empty category no longer needs to be inferred from its environment but is given formal expression, and this is the basic insight that Keenan and Comrie propose.

Within the present framework, note that pronoun-retaining relatives replace a filler-gap dependency with a filler-pronoun dependency that proceeds from the filler to the resumptive pronoun. These pronouns avoid a gap structure, and so I shall first formulate a hypothesis for gap avoidance in general. The hypothesis assumes that gaps will be
avoided in proportion to the processing difficulty of the environments in which they would have occurred if grammatical.\(^ {13}\)

(19) **Gap Avoidance Hypothesis**: If a (potential) FGD of complexity \(n\) on a hierarchy \(H\) is avoided by employing a semantically equivalent alternative structure, then gaps in all more complex environments on \(H\) (\(n + 1\)) will be avoided as well.

When a gap is avoided, we expect the semantically equivalent alternative to be one that makes the processing of a complex environment easier in some way. A pronoun copy makes recognition of the head’s role in the relative clause as clear as it can possibly be, and this leads to the prediction in 20 for what is, in effect, a reverse implicational structuring for copy pronouns.

(20) **Relative Clause Copy Pronoun Hierarchy**: If a copy pronoun in a relative clause is grammatical in position \(P\) on a complexity hierarchy \(H\), then copy pronouns will be grammatical in all lower positions that can be relativized at all.

Twenty-seven of Keenan and Comrie’s languages have a pronoun-retaining strategy, and these languages support prediction 20 without exception.

(21) Pronoun-retaining languages in Keenan and Comrie’s data

- GEN only: Japanese, Javanese, Korean, Malay, Roviana, Turkish, Yoruba
- IO/OBL & GEN only: Fulani, Greek, Hausa, Minang-Kabau, Shona, Toba Batak, Welsh, Zurich German
- DO & IO/OBL & GEN only: Aoban, Arabic, Chinese (Pekingese), Czech, Genoese, Gilbertese, Hebrew, Kera, Persian, Slovenian, Tongan\(^ {14}\)
- SU & DO & IO/OBL & GEN: Urhobo

Prediction 20 allows pronoun retention not to go down to the bottom of the hierarchy, in the event that relativization is completely ungrammatical on the lowest positions. For this clause-internal hierarchy all 27 languages do go all the way.

The pattern of 21 was noticed by Keenan and Comrie, but it is not predicted by their hierarchy constraints. Pronoun retention per se does not count as a separate strategy, so no predictions are made for it. But even if it were a separate strategy, the distribution of copy pronouns would not follow from the hierarchy constraints. The continuous segment prediction allows all strategies to occur in adjacent positions anywhere on the hierarchy, which means that \([ + \text{Case}]\) relatives could in principle apply to higher positions than \([ - \text{Case}]\), and their constraint 3 (strategies that apply at any one point may in principle cease to apply at any lower point) misses the generalization that relativization in 21 can cut off at any higher point.

We clearly have two implicational patterns going in opposite directions in these relative clause data: gaps from low to high, copy pronouns from high to low. The gaps cut off in more complex environments, while the pronouns cut off in simpler environments. This can be seen vividly when we examine the 24 languages in the

---

\(^ {13}\) In addition to resumptive pronoun strategies such as 18 in Hebrew there are additional gap-avoiding structures on which 19 could be tested, including head-internal relatives as in Tibetan (Lehmann 1984, Keenan 1985, Basilico 1996). These alternatives generally keep the filler ‘in situ’, in the same way that an unmoved wh-word remains ‘in situ’ in Japanese (Kuno 1973). They also involve processing difficulties of their own, however, in the form of structural ambiguity over which NP is actually the semantic head (see e.g. Lehmann 1984:111).

\(^ {14}\) For Tongan, an ergative language, I assume the hierarchy ABS > ERG > IO/OBL > GEN, cf. Primus 1993, and hence the second position down on the hierarchy is ERG and not DO.
sample that have both. The two strategies are largely complementary to each other, with overlapping options only at the points of transition, as shown in Table 1 (in which the languages are arranged in terms of declining copy pronouns). The distribution of gaps to pronouns decreases down \( \text{AH} \) (100% to 65% to 25% to 4%) while that of pronouns to gaps increases (0% to 35% to 75% to 96%).

<table>
<thead>
<tr>
<th>Language</th>
<th>SU</th>
<th>DO</th>
<th>IOOBL</th>
<th>GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aoran</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Arabic</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Gilbertese</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Kera</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Chinese (Pek)</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Genoese</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Hebrew</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Persian</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Tongan(^b)</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Fulani</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Greek</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Welsh</td>
<td>gap</td>
<td>pro</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Zür. German</td>
<td>gap</td>
<td>*</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Toba Batak</td>
<td>gap</td>
<td>*</td>
<td>pro</td>
<td>pro</td>
</tr>
<tr>
<td>Hausa</td>
<td>gap</td>
<td>gap</td>
<td>gap/pro</td>
<td>pro</td>
</tr>
<tr>
<td>Shona</td>
<td>gap</td>
<td>gap</td>
<td>gap/pro</td>
<td>pro</td>
</tr>
<tr>
<td>Minang-Kabau</td>
<td>gap</td>
<td>*</td>
<td>*/pro</td>
<td>pro</td>
</tr>
<tr>
<td>Korean</td>
<td>gap</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
</tr>
<tr>
<td>Roviana</td>
<td>gap</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
</tr>
<tr>
<td>Turkish</td>
<td>gap</td>
<td>gap</td>
<td>gap</td>
<td>pro</td>
</tr>
<tr>
<td>Yoruba</td>
<td>gap</td>
<td>gap</td>
<td>0</td>
<td>pro</td>
</tr>
<tr>
<td>Malay</td>
<td>gap</td>
<td>gap</td>
<td>RPgap</td>
<td>pro</td>
</tr>
<tr>
<td>Javanese</td>
<td>gap</td>
<td>*</td>
<td>*</td>
<td>pro</td>
</tr>
<tr>
<td>Japanese</td>
<td>gap</td>
<td>gap</td>
<td>gap/pro</td>
<td>pro</td>
</tr>
</tbody>
</table>

Gaps = 24 [100%] 17 [65%] 6 [25%] 1 [4%]  
Pros = 0 [0%] 9 [35%] 18 [75%] 24 [96%]

**Key:**
- gap = \([-\text{Case}]\) strategy as defined by Keenan & Comrie 1977, 1979
- pro = copy pronoun retention (as a subinstance of \([\text{+Case}]\)
- * = obligatory passivization to a higher position prior to relativization
- 0 = position does not exist as such
- RPgap = see Table 2

\(^a\) The \([-\text{Case}]\) gap classifications and copy pronoun classifications are all from Keenan & Comrie 1977, 1979. \([-\text{Case}]\) gap languages may employ a general subordination marker, no subordination marking, a fronted case-invariant relative pronoun, or a participial verb form.

\(^b\) For Tongan the top two positions of AH are Absolutive and Ergative respectively, not SU and DO (Primus 1995).

The implicational structuring of the two strategies (cf. 17 and 21) and their relative distribution in Table 1 support the proposed complexity ranking of AH positions and the proposed difference between gaps and pronouns in relation to this complexity. What has not yet been explained is the fact that gaps are preferred at all in the higher positions, despite the fact that they are hard-to-process empty categories. Why does pronoun retention not regularly extend all the way up?

A possible answer is that gaps have the advantage of economy of expression (Haiman 1983) and require less form processing. There are numerous structures in all languages that attest to the advantages of brevity in environments that permit recoverability of the relevant information (control structures, coordinate deletions, and so on), and this
appears to be what motivates the gaps in higher hierarchy positions. We have a competing motivation, therefore, between (reduced) form processing and explicit dependency marking, with structural complexity increasingly requiring the latter (see Hawkins 1998 for further examples and discussion of this competition).

3.5. [+Case] RELATIVE PRONOUNS. Most of Keenan and Comrie’s [+Case] strategies that do not employ a copy pronoun use a relative pronoun instead, as in Russian.

(22) a. Ivan videl devušku [kotoraja, O ljubit Petra]
Ivan saw girl who (SU) loves Peter
b. Ivan videl devušku [kotoruju, Petr ljubit O]
Ivan saw girl whom (DO) Peter loves

The relative pronoun indicates the role of the head in the relative clause unambiguously, but unlike the copy pronoun it has been moved to left-peripheral position, leaving a gap. Ex. 22 is a filler-gap structure, therefore, but with a co-indexed filler copy (i.e. relative pronoun) adjacent to the head. In addition to its head disambiguation function, the relative pronoun serves to construct the relative clause (Hawkins 1994) and to formally mark the dependency to the head. In some languages, like Arabic, these latter functions are performed independently of case-coding and the relative pronoun is case-ambiguous (see Keenan & Comrie 1979:333), as it is for die in Dutch and for who in many English dialects.

The case-distinctive relative pronoun strategies are interesting in the present context because they combine features of both the gap and the copy pronoun, and this places them in an intermediate position in terms of processing complexity. Like a copy pronoun, a relative pronoun can signal the function of the matrix filler in the subordinate clause. But by displacing it, a gap still needs to be recognized and filled. Using a simple feature system, if [−Case] relatives are [+gap, −pro], and [+Case] relatives with copy pronouns are [−gap, +pro], then [+Case] relative pronoun strategies will be [+gap, +pro].15

Consider the languages in Keenan and Comrie’s data that combine a [−Case] gap with a [+Case] relative pronoun + gap (abbreviated here as RPgap). I counted 12, using Keenan and Comrie’s [−Case] and [+Case] classifications and taking the relative pronoun classification for [+Case] languages from Maxwell 1979, in which (case-distinctive) relative pronouns are recognized as a separate strategy. The [+Case] rela-

15 If pied piping applies to move, for example, a subcategorizing preposition along with a relative pronoun (the person with whom John talked), then the gap within the relative clause will be co-indexed with the containing PP. The result is two separate coindexation domains, one between the head and the relative pronoun, the other between the moved PP and its gap (the person, [[with whom], John talked O]). Non-pied-piped counterparts involve a single index assigned to the head, the relative pronoun and the gap (the person, [who, John talked with O]). The pied piping structures are highly preferred crosslinguistically. The relative pronoun remains adjacent to its subcategorizer and close to its head (Hawkins 1998), so even though there are two coindexation domains, both are short and the normal adjacency of subcategorizer to subcategorizee is preserved. The pied piping structure also avoids garden paths resulting from competing gap sites (§6.3) and it conforms to the same Fillers First preference as moved relative pronouns alone (§6.1). The preposition stranding structure has the advantage of a single co-indexation, but this is outweighed in most languages by the competing advantages of pied piping. It may not be an accident that preposition stranding is productive in English, a language in which prepositions are highly dependent on verbs for their interpretation and processing (Hawkins 1998, 1999). By keeping the preposition in situ, this interdependency can be processed within a smaller domain. This predicts that the ratio of stranding to pied piping in English should be proportional to the degree of dependency between V and P. The overall processing advantages and disadvantages of each of these structures, and their consequences for performance and grammars, need to be further pursued.
Table 2. Languages combining [-Case] gaps with [+Case] relative pronouns + gaps.  

<table>
<thead>
<tr>
<th>Language</th>
<th>SU</th>
<th>DO</th>
<th>IO/OBL</th>
<th>GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>gap/RPgap</td>
<td>RPgap</td>
<td>RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>Finnish</td>
<td>gap/RPgap</td>
<td>gap/RPgap</td>
<td>RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>English</td>
<td>gap</td>
<td>gap</td>
<td>RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>Dutch</td>
<td>gap</td>
<td>gap</td>
<td>RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>Italian</td>
<td>gap</td>
<td>gap</td>
<td>RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>Swedish</td>
<td>gap</td>
<td>gap</td>
<td>RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>Malay</td>
<td>gap</td>
<td>gap</td>
<td>RPgap</td>
<td>pro</td>
</tr>
<tr>
<td>Spanish</td>
<td>gap</td>
<td>gap</td>
<td>gap/RPgap</td>
<td>RPgap</td>
</tr>
<tr>
<td>Catalan</td>
<td>gap</td>
<td>gap</td>
<td>gap/RPgap</td>
<td>–</td>
</tr>
<tr>
<td>Roviana</td>
<td>gap</td>
<td>gap</td>
<td>gap/RPgap</td>
<td>pro</td>
</tr>
<tr>
<td>Greek</td>
<td>gap</td>
<td>*</td>
<td>*/RPgap</td>
<td>–</td>
</tr>
<tr>
<td>Maori</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

| Gaps | 12 [86%] | 10 [83%] | 3 [20%] | 0 [0%] |
| RPgaps | 2 [14%] | 2 [17%] | 12 [80%] | 8 [100%] |

Key: gap = [-Case] strategy as defined by Keenan & Comrie 1977, 1979, cf. Table 1  
pro = copy pronoun retention, cf. Table 1  
RPgap = relative pronoun + gap (as a substance of Keenan & Comrie’s [+Case])  
* = obligatory passivization  
– = relativization impossible using any strategy  

* The [-Case] gap classifications are all from Keenan & Comrie 1977, 1979, as are the copy pronoun classifications. Within the set of [+Case] languages defined in Keenan & Comrie the relative pronoun classification is from Maxwell 1979 and refers to fronted case-distinctive or preposition-governed relative pronouns that leave a (V-subcategorized, P-subcategorized or adjunct) gap.

tive pronouns prefer the lower AH positions, just as the copy pronouns do, as shown in Table 2. The RPgap strategies increase down the AH (14% to 17% to 80% to 100%), while the gaps decline (86% to 83% to 20% to 0%), and the two strategies overlap only at the point or points of transition. (The data in Table 2 are arranged according to declining relative pronoun strategies.)

Since relative pronouns also involve a gap, they should be subject to the relative clause gap prediction of 16: any gap on a low position implies the grammaticality of gaps in all higher positions. Table 2 reveals that this is the case. The RPgaps generally occur lower than [-Case] gaps but provide a continuous gap strategy from low to high in each case.  

Table 2 also reveals that it is possible to extend the copy pronoun prediction of 20 to (case-distinctive) relative pronouns: if a pronoun is grammatical in position P on a hierarchy, then pronouns will be grammatical in all lower positions that are grammatical. Either a relative pronoun or a copy pronoun occurs all the way down AH for as long as relativization is possible at all. Hence, RPgaps appear to obey the implicational predictions for both gaps and (copy) pronouns, and since these predictions go in opposite directions, it is no accident that we find RPgaps intervening between gaps and copy pronouns in languages like Malay and Roviana that combine all three.

4. REDUCE FGD SIZE: A CLAUSE EMBEDDING HIERARCHY. The grammatical variation in filler-gap or filler-pronoun dependencies considered so far has been definable in terms of the relative complexity of different positions within a clause. There is also

16 For Maori I need to assume either that this prediction is satisfied prior to obligatory passivization of the DO position, or that obligatory passivization removes DO from the hierarchy on this occasion, so that the only higher position to IO/OBL is SU.
principled variation across languages in the distribution of filler-gap and filler-pronoun dependencies that cross clause boundaries.

4.1. A CLAUSE EMBEDDING HIERARCHY FOR FGDS. Ross’s (1967) original constraints blocked movements out of complex NPs such as 2 above and 23.

(23) a. *Who₁ [did you discuss [the fact [that the professor taught O₁]]]
    b. *The student₁ [that you discussed [the fact [that the professor taught O₁]]]

His complex NP constraint was defined as follows:

(24) Complex NP Constraint (CNPC): No element contained in an S dominated by an NP with a lexical head noun may be moved out of that NP by a transformation.

Subsequent reformulations in terms of subjacency (Chomsky 1981, 1986) prevent movements out of all environments that involve crossing more than one bounding node, e.g. NP or IP (= S), with the selection of such nodes being subject to parametric variation across languages (Rizzi 1982). The precise predictions of subjacency depend on a number of theory-internal assumptions, some of which are quite controversial like the successive cyclic movement of WH (Bach 1977). What is significant for our purposes is that general constraints of grammar have been proposed that block filler-gap dependencies in certain environments with large FGDs, while smaller FGDs may be unconstrained and allowed to run free.

Some languages permit violations of the CNPC, however. The Scandinavian languages provide convincing examples, like 25 from Allwood 1982.

(25) ett benₐ [somₐ jag ser [en hund [som gnager på Oₐ]]] (Swedish)
    a bone which I see a dog which is-gnawing on

Productive counterexamples are also found in Japanese as discussed by Kuno (1973: 239–40), including 26.

(26) [[[Oₐ osiete-ita] seito ga] rakudaisita] sensei₁
    teaching-was student flunked teacher
    ‘the teacher who the students that (he) was teaching flunked’

And CNPC violations in Akan are exemplified in Saah & Goodluck 1995. There has not yet been a quantification of the languages that permit such violations, though Maxwell (1979) gives an initial classification and cites Korean and Tamil as further counterexamples. Impressionistically CNPC-violating languages are a minority, and their productivity in performance is often restricted and attests to their complexity.¹⁷

On the other hand, many languages impose grammatical constraints on FGDs that are stronger and more restrictive than those of English. The ungrammatical relative clause gap in the German finite subordinate clause of 3b contrasts with its grammatical English counterpart 1b. Gaps in finite subordinate clauses are also ungrammatical in Russian, as shown in 27a from Comrie 1973:297. The corresponding 27b, with the gap in an infinitival phrase, is grammatical.

¹⁷ Kvam (1983:125) mentions that the heads of complex NPs in CNPC-violating extractions in Norwegian are typically indefinite (see §5.2 for discussion of the significance of this point). A detailed discussion of performance factors influencing acceptability in Swedish CNPC violations can be found in Allwood 1982, Andersson 1982, and Engdahl 1982. Gaps in complex NPs in Akan are less acceptable than in finite complements of the verb (Saah & Goodluck 1995). For Japanese, Kuno (1973:239–40; 244–60) discusses restrictions in terms of the appropriateness or otherwise of a theme interpretation for the filler of the relative clause whose gap is within the complex NP.
(27) a. *Vot [ogurcyi s[kotorye, ja obeščal s[čto prinesu Oi]]
   'Here are the cucumbers which I promised that I'd bring.'
 b. Vot [ogurcyi s[kotorye, ja obeščal vp[prinesti Oi]]
   'Here are the cucumbers which I promised to bring.'

Corresponding relative clause gaps in infinitival phrases are also possible in German (Kvam 1983), and of course in English too (as in the translation of 27b). French behaves like English and permits relative clause gaps in both finite and infinitival complements.18

(28) a. la femmei s[quei j’espère s[que vous allez épouser Oi]]
   the woman who I hope that you are-going to-marry
 b. la femmem s[quei j’espère vp[épouser Oi]]
   the woman who I hope to marry

For Wh-questions the constraints of German and Russian are closer to those of English and French, with gaps being generally acceptable in both finite and infinitival clauses, though there is some dialect variation in both languages. A finite clause gap in German is illustrated in 3a, and infinitival gaps are straightforwardly possible (Kvam 1983). For Russian, Comrie (1973:302) gives example 29a and points out that the corresponding 29b (which does not carry the echo question interpretation that it does in English) is preferred.

(29) a. (?Kogo s[ty skazal s[čto on ljubit]]
   'Who do you say that he loves?'
 b. s[ty skazal s[čto on ljubit kogo]

It appears that infinitival phrases are most hospitable to gaps, that finite subordinate clauses are more resistant, while complex NP environments are most resistant of all. Hence we have a hierarchy for gaps in clause embeddings supported by the languages cited so far.

(30) Clause embedding hierarchy

S represents the clause that is immediately adjacent to the filler. Relative clause constructions cut off in Russian and German when the gap is in a finite S’, while in English and French they cut off when the gap is in a complex NP (31). For Wh-questions, Russian, German, English and French all cut off at the complex NP position (32).

(31) Permitted FGDs for relative clause heads
 a. S[. . . VP' . . .] Swedish, Japanese, English, French, Russian, German
 c. S[. . . NP[Det N S'] . . .] Swedish, Japanese

(32) Permitted FGDs for Wh-question fillers19

18 French also permits Wh-question gaps in finite and infinitival subordinate clauses.

(i) Qui, s’est-ce que vous espérez s[que j’épouserais Oi] who is-it that you hope that I will-marry
(ii) Qui, s’est-ce que vous espérez vp[épouser Oi] who is-it that you hope to-marry

19 Japanese has filler-gap structures only for relative clause heads as fillers, but keeps its Wh-question words in situ and so has no filler-gap construction at all in this case, see §§5.1, 6.1
The ranked positions of 30 appear to correlate with increasing FGD size. Consider the relevant structures in English:

\[(33)\]
\[a. \text{Who}_i \text{s[do you try } \text{VP'}\text{[to see } \text{O}_i]\text{]?} \]
\[b. \text{The person}_i \text{s[that you tried } \text{VP'}\text{[to see } \text{O}_i]\text{] is Harry.} \]

\[(34)\]
\[a. \text{Who}_i \text{s[do you hope } \text{S'}\text{[that you would see } \text{O}_i]\text{] is Harry.} \]
\[b. \text{The person}_i \text{s[that you hoped } \text{S'}\text{[that you would see } \text{O}_i]\text{] is Harry.} \]

\[(35)\]
\[a. *\text{Who}_i \text{s[do you know } \text{NP}\text{[the professor } \text{S'}\text{[that taught } \text{O}_i]\text{]?} \]
\[b. *\text{The student}_i \text{s[who, you know } \text{NP}\text{[the professor } \text{S'}\text{[that taught } \text{O}_i]\text{] is Hany.} \]
\[c. *\text{What}_i \text{s[do you regret } \text{NP}\text{[the fact } \text{s[that he stole } \text{O}_i]\text{]?} \]
\[d. *\text{The money}_i \text{s[that you regret } \text{NP}\text{[the fact } \text{s[that he stole } \text{O}_i]\text{] is missing.} \]

In 33 the FGDs proceed from the filler to the gap’s subcategorizer see, and include a higher verb try with its subject you, and an embedded infinitival phrase (labeled here VP’). The FGDs of 34 are more complex. The subcategorizer of the gap now occurs within a finite subordinate clause (labeled S’), and this S’ contains an overt subject (you), finiteness marking (would), and the complementizer that. The nonfinite VP’ in 33 contains no overt subject and no finiteness marking, though it does contain what can be analyzed as an infinitival complementizer (to). The FGDs for 34 therefore contain more nodes. The complex NP environments of 35 have the largest FGDs of all. They include an additional dominating NP over S’ and an additional lexical head noun that c-commands the gap, compared with 34.\(^{20}\)

These differences in FGD sizes motivate 30. As with the AH (12) there are more nodes to be processed down the hierarchy, and these additional nodes involve phonological, morphological, syntactic, and semantic processing operations that apply simultaneously with filler-gap processing. The node increases have been illustrated here on the basis of minimal FGD size differences between positions. Ultimately, of course, it is the aggregate processing load associated with these positions in performance that leads to their ranking and to the cut-off points in grammatical conventions. Some experimental support for the greater processing complexity of gaps in finite subordinate clauses versus matrix clauses (i.e. \[s[...S'[...O_i...]]\] versus \[s[...O_i...]]\) is presented in Pickering & Shillcock 1992. The greater processing difficulty of gaps in complex NP environments versus finite subordinate clauses has been shown in Akan, a language that permits CNPC violations, in a controlled acceptability experiment by Saah and Goodluck (1995). Gaps in complex NPs are grammatical in this language, but they were still judged significantly less acceptable than gaps in simpler environments. Saah and Goodluck tested the same sentence types on speakers of English and found, predictably, that English CNPC violations were completely unacceptable, on account of their ungrammaticality. The processing difficulty posed by gaps in complex NPs can also be seen in Swedish and Norwegian. These structures are made as easy as possible in performance by minimizing all other complexity factors (see the references in n.17). I have not seen experimental results

\(^{20}\) In 35, a and b also involve more complexity than c and d: in addition to the co-indexation between who/the student, and O₁, there is further co-indexation between the professor and its gap within S’, which has not been indicated here and which will become relevant for a further prediction in §5.1.
of relevance to the infinitival/finite complement distinction, but impressionistically
infinitival gaps are much more frequent in English performance.

The filler-gap complexity hypothesis (11) does not require that all structural instances
of a less complex structural type on a hierarchy should be grammatical when a more
complex counterpart is grammatical. This is because there are often other interacting
complexity considerations. If a finite subordinate clause undergoes WH-fronting, the
result may be a WH-island, and an environment that normally permits a relative clause
may no longer do so (§5.1), even in a language that permits gaps within the lowest
structural position of hierarchy 30 (complex NPs). Hence the grammaticality of a gap
in a low hierarchy position implies the grammaticality of all higher hierarchy positions
in principle, or all things being equal, and it is sufficient for the prediction if just some
FGDs of the simpler type are grammatical.

Notice that English gaps in complex NP environments can sometimes be rescued by
pronoun retention: *I met the man, who, I had almost forgotten the fact that he, was our
prime minister* is not too bad. On the other hand, some complex environments are
completely ungrammatical, even with pronoun retention. Erteschik-Shir (1992:90) stars
36, with a copy pronoun in a relative clause complex NP in subject position.

(36) *This is the girl, [who, [the man who raped her,] had escaped from prison].

A containing relative clause will be argued to be a more complex environment than a
containing NP complement like *the fact that* . . . in §5.1, and a containing subject
introduces a center embedding and lacks a subcategorizer for the subject within the
filler-pronoun domain (see §6.4), whereas both of these disadvantages can be avoided
by a containing object. Copy pronouns are preferred to gaps in English in environments
of intermediate complexity, it seems, but even pronouns are ungrammatical when com-
plexity is greatest.

My approach in this article predicts that hierarchies of processing complexity will
be as relevant for filler-pronoun dependencies as they are for FGDs, and we saw support
for this in the AH data in §3.4. A copy pronoun is easier to process than a gap, but
there is still a processing load associated with filler-pronoun domains. The filler must
be held in working memory for the duration of the dependency, and there is more
simultaneous processing in the path from filler to pronoun than there is in corresponding
sentences without the filler. What is easier is identification of the position relativized
on. As a result, low and complex hierarchy positions, such as complex NPs, may not
be relativizable at all, especially when the demands of expressiveness are minimal.
When pronoun retention is grammatical, therefore, it obeys a high-to-low prediction.
But if pronoun retention is ungrammatical at a low hierarchy point, then no further
relativization will be possible at all down the hierarchy, since pronoun retention is the
most helpful and easy-to-process strategy there is.

Consider Hebrew. In this language copy pronouns within complex NPs are regularly
possible, as illustrated in 37 (from Keenan 1972a).

(37) Ani roa et ha-ish, [she-Sally ma'aamina NP[la-shmoa S[ she-Miriam
hikta oto,]]

I see do the-man that-Sally believes the-rumor that-Mary
hit him

'I see the man that Sally believes the rumor that Mary hit him.'

In Toba Batak, on the other hand, copy pronouns are not possible in complex NPs
corresponding to the Hebrew 37, even though they are possible in finite subordinate
clauses, and on IO/OBL and GEN within clauses (see Table 1). Keenan (1972b) gives the following example of a grammatical copy pronoun in a finite subordinate clause:

(38) boruboru₄ i₅[ ima na dirippu si Bissar ₛ[na manussi abit ibana₄]]
woman the namely that thought by Bissar that washes clothes she
‘the woman who Bissar thought that she washed the clothes’ (Toba Batak)

Pronouns do not go all the way down the clause embedding hierarchy (30) in Toba Batak, therefore, though they do go down the AH.

We now see why copy pronouns can lead to confusion when grammatical constraints are considered in isolation from the processing considerations that motivate their existence. Languages like Hebrew led Ross (1967) to exclude pronoun retention from his CNPC, limiting it to gap strategies only. But pronouns in complex NPs are ungrammatical in Toba Batak, suggesting that his CNPC (and subjacency) are still operative. All of this crosslinguistic variation is predicted by processing complexity: the fully ungrammatical structures are those at the lower end of complexity hierarchies, with ungrammaticality at a high point implying ungrammaticality all the way down, using any strategy; pronoun retention is grammatical from high to low positions for as long as relativization is possible at all; and gaps are grammatical from low to high all the way up to the highest position.

4.2. CROSSLINGUISTIC VARIATION IN GRAMMATICAL CONVENTIONS. It should be clear from the data considered that the conventions of grammars vary considerably with respect to the distribution of gaps, copy pronouns, and complete ungrammaticality. This variation is constrained and predictable, however, on the basis of processing complexity. In those languages whose grammars permit extensive filler-gap dependencies, we see performance evidence for degrees of relative complexity, and these relative degrees, I hypothesize, are constant across the human processor and are consequences of its basic architecture. What varies is the conventionalized response in different speech communities. Processing ease declines down the AH and as it does so the distribution of gaps to pronouns declines. Almost all languages have gaps on SU and pronouns on GEN, with the variation being clustered around DO and IO/OBL. For the clause embedding hierarchy, the distribution of gaps to copy pronouns appears to follow the same pattern.

The conventions of grammars have been set differently, therefore, in response to relative complexity. Even within a single language family, Germanic, we see the complete spectrum of conventionalized variation in the permitted FGDs for relative clause fillers (see 31). At the one end many (northern and central) dialects of German are quite constrained in their gap possibilities, at the other end Swedish and Norwegian are extremely liberal. English occupies an intermediate position, along with the southern German dialects. To anyone who knows these languages it is clear that the conventions are different, just as the conventions for other shared structures such as verb-particle constructions have been set differently following the breakup of Proto-Germanic. These conventionalized differences result in striking differences in grammaticality judgments among speakers of the Germanic languages. We are not dealing here with performance differences alone. The grammars are different, just as the grammars of English and Akan were found to be different in Saah and Goodluck’s grammaticality judgment

21 Constructions such as call NP[the boy] up exhibit two alternating patterns in English and Norwegian, V NP Part and V Part NP, just one each in Danish (V NP Part) and Swedish (V Part NP), and a different pair in German and Dutch (NP Part V and V NP Part). There are real differences here in conventionalized grammatical rules.
PROCESSING COMPLEXITY AND FILLER-GAP DEPENDENCIES ACROSS GRAMMARS 267

experiment (1995). The grammar of gaps varies, therefore, the grammar of pronoun retention varies, and the set of grammatical versus ungrammatical strings generated by different grammars also varies. This variation isUltimately motivated by processing, and is constrained and predicted by it.

There may also be certain processing conventions in different language communities that define possibly different degrees of acceptability for the structures generated by otherwise identical grammars, as argued in Kluender 1997. The speakers of some languages may regularly prefer small FGDs and more frequent applications of passive and other promotional rules converting lower (underlying) gap positions to the higher positions actually relativized on (cf. the variants illustrated in Keenan & Comrie 1977). Some positions of the AH permit both gaps and copy pronouns in certain languages (see Table 1), and different speech communities may exhibit different preferences for the one or the other. But the structures available to a language have been conventionalized in its grammar, and though grammars respond to the same hierarchies of relative complexity, the grammatical conventions themselves also vary, depending on, for example, the availability of alternative structures for expressive purposes. The responses to hierarchies such as the AH (12) and 30 have fixed position-specific conventions for gaps and copy pronouns. The conventions also appear to be specific to different filler types. WH-question fillers occur lower on hierarchy 30 than relative clause head fillers in Russian and German dialects. This is not predicted by 30 alone, but will be predicted in conjunction with an independent complexity difference between relative clause FGDs and WH-question FGDs (§5.1).

Notice that once the grammar conventionalizes a gap possibility in a certain construction type, for example a WH-question filler with a gap in a finite subordinate clause, then that possibility of necessity extends to gaps contained within several such clauses, because grammars cannot count (see Frazier 1985 for discussion). Hence structures such as *Who do you think that Fred believes that Bill suspects . . . that Mary saw? are grammatical in English. And though the increasing depth of a gap will be associated with increasing processing complexity in performance, its grammaticality remains intact. By contrast, if an environment that is not conventionally sanctioned occurs anywhere on the path from filler to gap, the sentence will be ungrammatical (e.g. *Who do you think that Fred believes that Bill heard the rumor that Mary saw O?).

In place of Chomsky's bounding nodes and subjacency condition (1981, 1986), I suggest that languages have conventionalized permitted versus nonpermitted gap and copy pronoun environments, defined in terms of different filler types and node combinations as in 30. The grammatical possibilities for each construction are constrained by the hierarchies and are explainable by processing complexity. Once a convention has been set within a grammar, the FGDs of all sentences generated are checked for compatibility or incompatibility with it. The successive cyclic movement of WH is one way to describe this checking processing throughout a domain, but this is not assumed in the present framework. FGDs can simply be searched for fillers and node configurations below the cutoff point for the relevant language and construction.

Subjacency is not a universal condition on grammatical conventions, therefore. There are counterexamples in several languages, in the form of CNPC violations, and at the same time there is a rich array of universal patterns in filler-gap and filler-pronoun dependencies that are not being accounted for. There are also subtle distinctions of the kind illustrated for German and Norwegian in Kvam 1983 that are not predicted. This lack of universality is evidence against the proposed innateness of subjacency (see Hoekstra & Kooij 1988 for the innateness argument), while the crosslinguistic patterns
and language-internal details point instead to the explanatory role of processing complexity.

Processing may also be able to motivate some hitherto mysterious restrictions on embedded subject extractions. Comrie (1989:162) points out that embedded objects are freely relativizable in the Hungarian 39a, whereas there is native speaker uncertainty over embedded subject relatives such as 39b, which goes against the ease of relativization that is correctly predicted for clause-internal relativizations by the AH.

(39) a. a pénz [amit mondtam, hogy a fiú elvett]
    the money which-ACC I-said that the boy took-away

b. ?a fiú [aki mondtam, hogy elvette a pénzt]
    the boy who I-said that took-away the money-ACC

German exhibits a similar asymmetry (Kvam 1983:84). The full set of acceptability distinctions elicited by Kvam points to an interaction of two factors that increase complexity: length of FGD; and (in)compatibility of the moved wh-word with the matrix verb. Moved wh-words are fully acceptable in German when the subcategorizer of the gap is intransitive and the wh-word is compatible with the matrix verb (was in German is ambiguously nominative or accusative).

(40) Was glaubst du, dass geschehen ist?
    what think you that happened has
    ‘What do you think has happened?’

In this way the FGD is short and does not contain a direct object preceding the subcategorizer (i.e. the subordinate verb in clause-final position), and the moved wh-word can actually be assigned to the matrix verb on-line as an accusative object (with the meaning ‘what do you think?’) prior to its eventual nominative interpretation as an embedded subject. If the subcategorizer is transitive and the wh-word is incompatible with the matrix, complete ungrammaticality results (wer is unambiguously nominative).

(41) *Wer glaubst du, dass das Spiel gewonnen hat?
    who think you that the game won has
    ‘Who do you think won the game?’

Intermediate acceptability results when one or the other of these conditions is not met (as in ?was glaubst du, daß den Jungen beeinflußt hat? ‘what do you think that has influenced the boy’, which contains a transitive subcategorizer). Independent support for the desirability of on-line matrix compatibility (which is in accordance with the active filler strategy of 6) comes from partial movement structures in German such as Was glaubst du, wie sie das gemacht hat? (literally ‘what think you, how she did that?’) which translates ‘how do you think (that) she did that?’ in English. These structures reduce a long and complex FGD into two shorter domains corresponding to ‘what do you think’ and ‘how did she do that’ using a dummy filler was which is processable on-line as the grammatical ‘what do you think’. There is probably no garden path here since ‘what do you think’ is part of the interpretation.

The that-trace effect of English (Chomsky & Lasnik 1977) also exhibits some performance variation. *What do you think that has happened? is ungrammatical, in contrast to the German 40. But Culicover (1993) points out that these structures are relatively well formed when additional material intervenes between that and the gap (John is the kind of person, who, I suspect that after a few drinks O, would be unable to walk straight). It is not clear why the extra weight after that should improve these sentences. What is clear is that there are additional factors that impact acceptability in embedded gap environments and that these factors interact and partially conflict with the AH.
ranking. When *that* is removed in extractions from finite subordinate clauses in English (*What do you think has happened?*), the AH prediction that subject extractions will be easier to process than objects has been confirmed in Pickering and Shillcock’s experiments (1992): they found subject gaps easier to process than object gaps both within and across clause boundaries.

5. **Reduce Additional Syntactic and Semantic Processing in FGDs.** It is not just the relative size of an FGD that makes it complex (see 10). Even in surface domains containing the same quantity of nodes for processing there may be more or less complexity as a function of the particular syntactic and semantic properties of the items in the domain.

5.1. **Reduce Additional Syntactic Processing.** FGDs of the same size exhibit subtle degrees of preference that suggest a more general minimization hypothesis for syntactic representations.

(42) **Reduce Additional Syntactic Processing in FGDs:** The human processor prefers to minimize the syntactic rules and processing operations that apply in FGDs.

One manifestation of this preference can be seen in the *wh*-island effects of Chomsky 1973. A subordinate clause that has undergone the syntactic process of *wh*-fronting and coindexation with a gap, e.g. in an indirect question, is more difficult to extract out of than a simple (declarative) *that*-clause in English.

(43) a. *What* did you hope [that they would bake *Oil*?
b. *What* did you wonder [how, they would bake *Oil*?]

Having one filler-gap dependency to resolve within the subordinate clause evidently makes it harder to resolve another across the clause boundary. The preference of 42, which is supported by all the evidence for node minimization, thus leads to implicational prediction 44.

(44) **Subordinate Gap/No Gap Hierarchy:** If a matrix filler can be matched with a gap in a subordinate clause of complexity *n* containing another gap, then it can be matched with a gap in a subordinate clause of complexity *n* containing no other gap.

When we convert the finite subordinate clauses of 43 to infinitival complements, the indirect question infinitival is no longer a *wh*-island, and both the declarative and the indirect question complement can contain a gap, as in 45.

(45) a. *What* did you hope [to bake *Oil*]
b. *What* did you wonder [how, to bake *Oil*

Other infinitival complements do constitute a *wh*-island (e.g. *what, did you ask [why, to bake *Oil*]), but infinitival complements permit more *wh*-island violations than finite complements. We have already seen that infinitival complements permit more gaps than finite subordinate clauses in Russian and German (§4.1), so the greater productivity of 45b compared to 43b in English should come as no surprise.

More generally, the difference between 45b and 43b is explained by the combinatorial effect of two hierarchies: the clause embedding hierarchy of $30 \ VP' > S' > NP-S'$, and 44 nongap-clause > gap clause. English has limited its extractions from gap clauses to infinitival complements (45b), and these extractions cut off at the finite complement point on hierarchy 30 (cf. 43b), whereas extractions from nongap clauses go down lower and include both finite (43a) and infinitival complement environments (45a).
Other languages have conventionalized the combinatorial options differently. Swedish and Norwegian are more liberal than English and permit extractions from structures corresponding to both 45b and 43b, see Engdahl 1982. German is more restrictive and permits them in neither, see Kvam 1983. In all these languages both hierarchies are respected and the grammatical conventions fall within a limited set at the intersection of each. No language can exist, for example, in which extractions are more productive from gap clauses than from nongap clauses, or more productive from finite clauses than from infinitival clauses.

Hierarchy 44 makes a further prediction in combination with hierarchy 30. Movements out of relative clauses should be worse than out of NP complement constructions, since the latter do not contain a second gap linked to the head of the complex NP. Recall the examples of 35, repeated in 35’ with all filler-gap dependencies indicated.

(35’)

a. *Who$_1$$_s$ [do you know$_{NP}$ [the professor$_{NP}$ [that$_{0}$ taught$_{O_1}$]]$_{O_1}$]?

b. *The$_4$ student$_{NP}$ [who$_1$ [you know$_{NP}$ [the professor$_{NP}$ [that$_{0}$ taught$_{O_1}$]]$_{O_1}$] is Harry].

c. *What$_5$$_s$ [do you regret$_{NP}$ [the fact$_{S’}$ [that$_{0}$ he$_{O}$ stole$_{O_1}$]]$_{O_1}$]?

d. *The money$_{NP}$ [that you regret$_{NP}$ [the fact$_{S’}$ [that$_{0}$ he$_{O}$ stole$_{O_1}$]]$_{O_1}$] is missing.

The c and d sentences are indeed more acceptable than a and b and even border on grammaticality. Certain lexicalized phrases such as make the claim that permit gaps quite grammatically, as Ross 1967 originally pointed out.

Consider now the differences between relative clauses and WH-questions in Russian and German in relation to the reduced syntactic processing preference of 42. Filler-gap dependencies in relative clauses are more complex than their WH-question counterparts. Both contain a filler that is peripheral to the clause that contains a gap, and both require identification of the gap. But the filler of the relative clause gap (i.e. the nominal head) is also a constituent of the matrix and it undergoes all the rules and processes that apply to matrix NP constituents: case-assignment, θ-role assignment, positioning, and so forth. When the filler is matched with the gap, these matrix properties must often be systematically undone within the relative clause, since the case, θ-role and position of the gap can be quite different. By contrast, the WH-question filler receives grammatical properties within one clause only, namely those that are appropriate for its gap site, and does not undergo all the additional processes that are characteristic of a relative clause head. The relative clause filler-gap structure involves more simultaneous syntactic operations and processing than the WH-question, therefore, and it is on this basis that we predict that it will be less or equally productive.

(46) **Relative Clause and WH-Question Filler:** In languages that have filler-gap structures for both relative clauses and WH-questions, if a gap is grammatical for a relative clause filler in an FGD of complexity n on a hierarchy H, then a gap will also be grammatical for a WH-question word filler in an FGD of complexity n.

Russian and German choose the less option: relativization cuts off at the S’ position of the clause embedding hierarchy, WH-question formation is still possible here. The definition in 46 allows for both to be equal, as in English, but the reverse of Russian

---

22 Swedish permits gaps in WH-clauses all the way down the clause embedding hierarchy 30, including within complex NPs containing indirect questions, as in (i), from Allwood 1982:20.

(i) Vem$_1$_$s$ [hade$_{NP}$ [en$_{NP}$ stilla undran$_{NP}$ [om$_{NP}$ Stina verkligen sett$_{O_1}$]]$_{O_1}$]$_{O_1}$

Who$_1$ had$_{NP}$ a quiet wonder$_{NP}$ whether$_{NP}$ Stina really$_{NP}$ had-seen$_{NP}$

"Who$_1$ did$_{NP}$ I$_{NP}$ wonder$_{NP}$ quietly$_{NP}$ whether$_{NP}$ Stina$_{NP}$ really$_{NP}$ had-seen$_{NP}$?"
and German, in which relative clauses are more productive than wh-questions, is predicted not to occur.

The processing complexity of relative clause filler-gap structures has also been supported by acquisition studies beginning with Sheldon 1974 that motivate the preferred parallel function of the nominal head and the gap (see MacWhinney & Pluh 1988, Clancy et al. 1986, and also Kirby 1999 for discussion of the relationship between parallel function and the AH). Parallel function is interesting because it is the additional and possibly different cases and 0-roles assigned to relative clause fillers and their gaps that, I claim, makes these structures more complex than wh-questions. When the properties of the filler and the gap are parallel, the relative clause filler is similar to a wh-question filler, whose properties are identical to those of its gap, and fewer additional properties are assigned within the FGD as a result. Some crosslinguistic effects of parallel function can be found in languages like Persian and Ancient Greek. In Persian the head of a relative can optionally acquire the case of its gap when this latter conflicts with the case of the head, while in Ancient Greek the case of a relative pronoun attracts to that of its head (see Comrie 1989:153–54 for data summary).

Note that the prediction in 46 applies only to languages that actually have filler-gap structures for both types of fillers. If there is no filler-gap structure for wh-questions and if they remain in situ in languages like Japanese (Kuno 1973), then relative clause filler-gaps will be more productive by default. There are many languages that pattern this way, with relative clause filler-gap dependencies and no wh-word gaps, but, it seems, no languages of the opposite type: languages with a wh-word filler-gap dependency cooccurring with systematic gap avoidance in relative clauses, e.g. through pronoun retention.

Why should gap avoidance be easier for wh-questions than for relative clauses? I believe this is connected to the parallel function of filler and gap in wh-questions. Relativization involves a noun-modification structure with co-indexed NP positions in two separate clauses, each of which contracts grammatical relations and dependencies within its clause. But a moved wh-word either remains immediately peripheral to the clause that contains its gap or it moves into a higher clause in which it contracts no such relations and dependencies. The grammatical properties of the wh-word are always those of the gap, and this makes it unnecessary to have two grammatical positions for the one set of grammatical functions. Wh-words can remain in situ, therefore, or alternatively they can move to clause-peripheral position adjacent to the clause over which they have scope (i.e. an iconic motivation, see Haiman 1983). The parallel function of wh-fillers and their gaps can therefore explain both the greater complexity of relative clause filler-gap dependencies (i.e. 46 and the Russian and German data) and the dispensability of wh-question gaps altogether.

5.2. REDUCE ADDITIONAL SEMANTIC PROCESSING. In addition to the preference for reduced syntactic processing in FGDs there is substantial evidence for a similar preference for less semantic processing as well.

(47) Reduce FGDs Semantically: The human processor prefers to minimize the amount of semantic information that needs to be processed in an FGD. Kluender (1992) ties together a number of semantic generalizations about the material that can stand in the path from filler to gap, and he proposes that the more semantic processing that is required within the domain, the more unacceptable and/or ungrammatical the filler-gap dependency becomes.

Consider the so-called bridging verbs that permit their clausal complements to contain
gaps. Across languages, the bridging verbs that are grammatical seem always to include semantically weak verbs like say rather than whisper, or nonfactives (think/believe) more productively than factives (realize/regret). For English Kluender gives the following examples adapted from Culicover & Wilkins 1984 (where > means ‘better than’):

(48) a. How angry did Mary say that John was? >
    b. How angry did Mary say softly that John was? =
    c. How angry did Mary whisper that John was?

Even in a language with extensive extractions, like Norwegian, long-distance FGDs exhibit the same preference for semantically weak bridging verbs (Kvam 1983). This can be expressed in 49.

(49) **Bridging Verb Hierarchy:** If a bridging verb or verb complex V of semantic specificity n is grammatical in an FGD, then all verbs or verb complexes V’ with less semantic specificity than V will also be grammatical.

Degrees of specificity could be defined in terms of semantic components or features. Whispers combines components of meaning associated with both say and softly, and is accordingly semantically richer and more specific. Specificity could also be defined in terms of entailment: whisper entails say, but not vice versa. Entailment would also define a specificity difference between factive and nonfactive verbs, since factives (John realizes that he is sick) entail corresponding sentences with nonfactives (John believes that he is sick), whereas the converse fails. Factives are more difficult to extract out of in English and in many other languages exhibiting so-called weak island effects (Kiparsky & Kiparsky 1970, Hukari & Levine 1995).

Another clear example of the preference in 47 can be seen in the declining grammaticality of the examples in 50 cited in Kluender 1992 and adapted from Chomsky 1973 and Erteschik-Shir & Lappin 1979.

(50) a. Who did you see a picture of? >
    b. Who did you see the picture of? >
    c. Who did you see John’s picture of?

An NP with an indefinite head readily permits a gap in an of-complement. The definite head is worse, and a possessive modifier makes the filler-gap structure worst of all. A definite NP adds a uniqueness claim to the existential claim of the indefinite, as well as pragmatic appropriateness conditions guaranteeing the satisfaction of uniqueness within some portion of the universe of discourse (Hawkins 1991). A possessive modifier contains an additional referring expression and defines a relation of possession or association between this referent and the head. The amount of semantic processing increases from 50a to 50c, and as it does so a filler-gap structure becomes increasingly impossible. Neurolinguistic support for this increase in processing using ERP measurements has been given in Kluender & Kutas 1993b and in Neville et al. 1991. Parallel to 49, therefore, is hierarchy 51.

(51) **Head Noun Phrase Hierarchy:** If an NP with head noun (phrase) N of semantic specificity n on the path from filler to gap is grammatical, then all head noun (phrases) N’ with less semantic specificity than N will also be grammatical.

Further evidence for 51 comes from complex NP gaps in Swedish and Norwegian. In performance they generally involve indefinite heads, as in example 25 above.

Kluender (1992) proposes a similar hierarchy for the complementizers of 52.
What did John doubt (a) that she would win? >
(b) if she would win? > (c) whether she would win

He justifies this hierarchy as follows: ‘The complementizer that merely signals that a proposition follows, while if indexes a possible state of affairs from among an infinite set of such possible states, and whether indexes one of only two possible (alternative) states of affairs, and can thus in a way be said to be more referentially specific in character’ (1992:240). Neurolinguistic support for the heightened processing load down 52 using ERP measurements is given in Kluender & Kutas 1993 a, b.

Notice how all of these semantic specificity hierarchies interact with syntactic hierarchies. If the conventions of a particular language allow gaps in finite subordinate clauses at all, and if there are restrictions in this environment involving semantic types of permitted bridging verbs, then the restricted verb types will be the semantically more specific ones. In this way, both the clause embedding hierarchy 30 and the bridging verb hierarchy 49 are respected. Similarly, any restrictions in the head noun status of complex NPs will respect 30 and hierarchy 51.

6. IDENTIFYING THE GAP. Domain minimization and reduced syntactic and semantic processing limit the amount of processing that needs to be undertaken in the path from filler to subcategorizer or gap, and this reduces the simultaneous processing and working memory load required for a difficult structural type. Some further crosslinguistic universals suggest that grammars have conventionalized ways of making it easier to actually identify and process the gap. Pronoun retention avoids the gap altogether, but in structures that retain the gap there are ways of making identification and processing easier.

6.1. Fillers First. Let us begin with an idea first proposed by Fodor (1983) on the basis of data from English: it is easier to process fillers before gaps than gaps before fillers. A gap is simply the absence of something. Its very existence must be inferred. But a filler has both form and content which can be used to initiate an FGD and activate the search for the gap, checking for the appropriateness of a match when each possible subcategorizer or gap site is encountered. If the gap comes first, it will be harder to recognize and there may be considerable backtracking when the filler is encountered. Fodor’s principle has even greater force in languages that have more deletion and scrambling than English.

(53) Fillers First: The human processor prefers to process fillers before their subcategorizers or gaps.

There is clear evidence for 53 in both WH-questions and relative clauses across languages. Consider the former. The languages that move WH-words to clause-peripheral position invariably move them to the left.

(54) Question-word Movement Universal: All languages that move a WH-question word to clause-peripheral position move it to the left, not to the right. There are two large language samples that support this: Greenberg 1966 and Ultan 1978. Greenberg gives data on WH-question word positioning in thirty languages, Ultan in seventy-nine. I have merged the two samples in 55, giving the correlations with basic verb position, and eliminating any cases of disagreement over basic verb position or WH-fronting. I have also eliminated languages Ultan describes as having questionable fronting. Eighty-seven languages remain, forty-nine of which have fronting (56%).
(55) **wh-fronting and verb position**

V-initial: 17/20 languages = 85%

SVO: 25/34 languages = 73.5%

SOV: 7/33 languages = 21%

Only one language, Khasi, is claimed (by Ultan) to postpose the wh-word to final position, but this classification appears to be erroneous. Other languages claimed to have wh-postposing have been raised as counterexamples, but Petronio and Lillo-Martin (1997) dismiss these claims. The preference for Fillers First appears to be exceptionless in wh-questions, therefore.

Languages with basic SOV generally avoid the wh-fronting option. They either keep the wh-word in situ or else move it to a focus position immediately preceding the verb (Kim 1988). We can make a prediction for the basic word order correlates of wh-fronting if we combine the preference for Fillers First (53) with the preference for domain minimization (10). If an SOV language like Japanese were to have wh-fronting, its basic verb position would result in a considerable distance between the filler and its subcategorizer, since the verb would be preceded by all other constituents of the sentence, including adjuncts and left-branching embedded clauses. I therefore propose

(56) **Question-word Fronting and Basic Word Order:** A fronted wh-question word will be preferred across languages in proportion to the distance between wh and the subcategorizing verb in its basic order: the smaller the FGD, the more the wh-word is fronted.

The distributional pattern of VSO > SVO > SOV in 55 bears this out: 56 also predicts that rigid SOV languages (like Japanese) will be more resistant to fronting than non-rigid SOV, i.e. VSO > SVO > SOVnr > SOVr. Impressionistically, I believe this is the case, though there are too few SOV languages with wh-fronting in the combined Greenberg-Ultan sample to permit general conclusions to be drawn.

Consider now the effects of Fillers First in relative clauses. This preference is in competition with basic word order in languages that are head final. Japanese places the relative clause before the head noun, i.e. [. . . O, . . . ] N, and this structure is preferred by Early Immediate Constituents (Hawkins 1990, 1994). When NP is embedded within PP, for example, a head noun can be adjacent to its governing postposition in Japanese, resulting in 100% IC-to-word ratios for the PP domain. If the relative clause follows the head noun and precedes P, i.e. [NP[N S'] P]pp, the recognition domain for PP will be longer and will have low IC-to-word ratios. Fillers First can therefore conflict with the preferences of EIC. For wh-questions there is generally no EIC-motivated contrary pull to position the wh-word filler to the right of the clause over which it takes its scope. But for relative clauses, the positioning of the head noun to the right is motivated by head-final orders in phrases that contain NP, so that there is a conflict here and this.

---

23 Ultan cites Nagaraja 1985 as his source, but the examples that Nagaraja gives of final-position wh (pp. 12–13) show no evidence of movement: they are in their unmoved basic order. A direct object ayu ‘what’ is going to be in final position in this SVO language if it remains in situ. At the same time Khasi clearly does show signs of possessing a wh-fronting rule (e.g. for baley ‘why’ questions), which means that Khasi behaves like many other languages in having both wh-fronting and wh-in situ.

24 If the gap is located in a left-branching subordinate clause of an SOV language, then a further processing problem will regularly arise: the path from a left-peripheral wh-filler to its gap will not include the matrix verb and will fail to make clear the structural relationship of the subordinate clause to its matrix, making all such embeddings similar to sentential subjects in English (see §6.4 for discussion of the valency completeness preference).
conflict results in variation. Lakhota (SOV) inverts the Japanese order to N₁ [. . . O₁ . . ] (Lehmann 1984:80–82), and 59% of SOV languages are like Lakhota, according to Dryer (1992:86), while 41% are like Japanese.

Proof of this competing motivation comes from a comparison with head-initial languages. These latter are almost exceptionlessly consistent in having head + relative orders (Dryer 1992:86). The explanation is that there is now no conflict: N₁ [. . . O₁ . . ] is preferred both by Fillers First and by EIC (the initial Nᵢ of NP is adjacent to a preposition within a containing PP, etc). But head-final languages exhibit no consistency whatsoever, and this variation alone suggests that no single strategy is optimal for these languages. A processing perspective can explain why.

Processing can also make sense of many further characteristics of prenominal relatives. Their syntactic form and content is much more restricted than that of their postnominal counterparts. Lehmann (1984:168–73) points out that prenominal relatives are more strongly nominalized, with more participial and nonfinite verb forms (as in Dravidian, Lehmann 1984:50–52), with more deletion of arguments, and more obligatory removal of modal, tense and aspect markers, and so on. All of this makes them typically shorter and reduces the amount of simultaneous processing. It also reduces temporary ambiguity, by explicitly signaling subordination status within the relative clause (cf. Hawkins 1994:323–28, Antinucci et al. 1979, Clancy et al. 1986).

The Japanese structure is bad for Filler First, therefore, but good for EIC. Lakhota is bad for EIC but good for Filler First. Languages like Lakhota will improve their IC-to-word ratios if they incorporate a rule of extraposition from NP that can position the center-embedded relative clause to the right of a postposition when Nᵢ [. . . O₁ . . ] is embedded in a PP (as in the German example dem Haus gegenüber, in dem er wohnt ‘the house opposite in which he lives’, i.e. opposite the house in which he lives). Adjacency of the heads N and P is achieved through such extrapositions at the same time that the initial filler preference is maintained. The result is still not fully optimal, however, on account of the separation and discontinuity of the head noun from its relative clause.

Fillers First is a major contributor to the grammatical structure of relative clauses in SOV languages, therefore, but its role must be seen in conjunction with the EIC preference for noun-final NPs in such languages. Languages with highly grammaticalized head-final word orders across all phrasal categories, i.e. rigid head-final languages like Japanese, will be those for which adherence to EIC is most advantageous, since there will be more head-final containing phrases whose processing is improved by noun-final NPs with prenominal relatives, and these languages will resist Fillers First to the greatest extent. As a result, postnominal relatives should be found predominantly in nonrigid SOV languages.

Fillers First receives further support from a number of other linear precedence asym-

---

25 Chinese is the only counterexample combining prenominal relative clauses with head-initial structures, but this language is not consistently head-initial (see Li 1990).

26 In Hawkins 1990 I proposed a less general principle in competition with EIC in head-final languages, namely garden path avoidance through Minimal Attachment (Frazier 1979, 1985). Fillers First claims that the positioning of a relative clause head before its subcategorizer or gap is to be preferred whether there is a garden path in the reverse ordering or not, for the reasons outlined in the main text and in greater detail in Hawkins 1998.

27 See Hawkins 1994:196–210 for clear performance data from German illustrating the decreasing textual frequency of Extraposition from NP structures, the larger the distance is between the head noun and an S’ relative clause modifier or complement.
metries across languages involving lexical categories and their dependent empty categories, for example control structures, and from other forms of ‘strong’ dependencies (cf. Hawkins 1998).

Notice finally that relative pronouns, i.e. filler copies of a nominal head, obey the same Fillers First asymmetry as WH-question words. They stand invariably to the left of their subcategorizers or gaps, never to the right, and they occur exclusively in post-nominal relatives (Lehmann 1984). In this way they can be both adjacent to the filler and obey Fillers First.

6.2. MORPHOLOGICAL, PHONOLOGICAL, AND SYNTACTIC SIGNALING OF AN FGD. A further way to facilitate gap identification is to assign special surface structure properties to FGDs. In this way the extent of the FGD is signaled explicitly, and the processor is guided to the gap. Many languages appear to have grammaticalized such explicit signals.

Irish exhibits an interesting alternation between complementizers that are within an FGD and those that are not. McCloskey (1979:150–51) discusses the contrast in 57.

(57) a. Shíl mé [goN mbeadh sé ann]  
thought I COMP would-be he there  
‘I thought that he would be there.’

b. an fear, [aL shíl mé [aL bdeadh Oi ann]]  
the man COMP thought I COMP would-be there  
‘the man that I thought that (he) would be there’

In the path from filler to gap in 57b a special complementizer is used, aL (where L indicates lenition), which contrasts with the normal complementizer with nasal mutation (goN) illustrated in 57a. The aL complementizer serves as a signal to the processor that the FGD is still ongoing and that the gap has not yet been encountered. As soon as it is encountered, any complementizers to the right of the FGD must revert to goN, as shown in 58.

(58) an fear, [aL shíl Oi [goN mbeadh sé ann]]  
the man COMP thought COMP would-be he there  
‘the man that (he) thought that he would be there’

Having alternating complementizers like this indicates which clause the gap is in.

Another type of morphological response can be seen in Chamorro, which employs special verb marking on all verbs in the path from filler to gap (Chung 1982, 1994). The explicit marking of FGDs through phonological means has been documented in Kikuyu. Clements et al. 1983 points out that FGDs are distinguished from corresponding domains without gaps by the suppression of tonal downstep. When the WH-word is left in situ, the downstep reappears.

A common syntactic strategy for marking FGDs is through verb fronting, as in 59, a German example from Grewendorf 1991:75.

(59) Was, [sagte Hans [glaubte Karl [fürchtete Maria [werde Peter Oi tun]]]]  
what said Hans believed Karl feared Maria would Peter do  
‘What did Hans say Karl believed Maria feared Peter would do?’

The fronting of the verb in each embedded S’ signals that that S’ is in the FGD. Other languages with similar verb fronting rules in their FGDs are Yiddish (Diesing 1990), French (Kayne & Pollock 1978) and Spanish (Torrego 1984).

6.3. AVOID COMPETING GAP SITES. The identification of a gap is made more difficult by the existence of competing gap sites in the FGD. Misidentification of the gap adds to the processing load of an FGD by causing garden path effects. When the correct
gap site is uniquely determined in the on-line parse no such misanalysis and subsequent correction need arise. How can competitor gap sites be eliminated or reduced? One way is for the grammar to limit the number of subcategorizers that permit their complements to be empty.

Notice that the on-line ambiguity of 5 in English arises only because English allows preposition stranding, i.e. a gap site after prepositions.

(5') Which studenti [did you ask (Oi) Mary about Oi]

In the alternative, pied-piping structure (Ross 1967), there is no on-line misanalysis.

(5") [About which studenti, [did you ask Mary Oi]

If a language pied-pipes its PPs obligatorily and disallows gaps subcategorized by prepositions, the processor will not even consider assigning a fronted NP to a preposition and only a verb-subcategorized gap site will be available. Such is the case in German:

(60) Welche Frau [hast du über Fritz gefragt] which woman have you about Fritz asked

‘Which woman have you asked about Fritz?’

If welche Frau ‘which woman’ originates in a PP, the whole PP must be fronted.

(61) [über welche Frau] hast du Fritz gefragt? about which woman have you Fritz asked

‘About which woman have you asked Fritz?’

The great majority of the world’s languages behave like German rather than English in this respect (see van Riemsdijk 1978 and Koopman 1984 for discussion of some exceptional languages in Germanic and Kru), and conform to 62.

(62) Avoid Competing Subcategorizers: The human processor prefers to avoid garden paths that result from competing subcategorizers within an FGD. Garden paths of this and other kinds are hard to process (Frazier 1979, 1985).

Notice how 62 dovetails with domain minimization (10): gaps in smaller FGDs on the AH (12) or clause embedding hierarchy (30) will contain fewer competing subcategorizers to be encountered and eliminated on the path from filler to gap. Domain minimization is independently supported, however, by the fact that smaller domains are preferred even when they don’t contain alternative subcategorizers that are syntactically or semantically compatible with the filler in question.

Why should it be the prepositional gaps rather than the verbal gaps that are more constrained crosslinguistically? This makes sense from the current perspective because prepositional gaps are typically deeper in the clause structure than verbal gaps and will have larger and more complex FGDs (recall the positioning of OBL on Keenan and Comrie’s AH). In addition, verbs have more subcategorized arguments and are more frequently occurring than prepositions (every clause has a verb, but not every clause has a PP). Hence the more complex and less frequently occurring gap option is eliminated in most grammars. We therefore have the hierarchy in 63.

(63) Gap Subcategorizer Hierarchy: If a gap can be subcategorized by P, it can be subcategorized by V.

6.4. VALENCY COMPLETELESS. English and certain other languages obey a SENTENTIAL SUBJECT CONSTRAINT (SSC) (Ross 1967), operating as shown in 64a.

(64) a. *Whoi s[did s[that Mary disliked Oi] v[ Surprise Sue]]]

b. Whoi s[did it v[ Surprise Sue s[that Mary disliked Oi]]]]

The path from the filler to the subcategorizer of the gap proceeds from whoi to disliked
in both examples. But the verb that subcategorizes for the containing sentential complement \(\text{[that Mary disliked } O]\), namely surprise, stands outside the FGD in 64a, since it occurs to the right of the gap. In the extraposition structure 64b surprise does stand in the path from filler to gap-subcategorizer and is in the FGD for who. I shall say that the processing domain for the filler-gap dependency in 64b is ‘valency complete’: the subcategorizers for all phrases containing the gap are present in the FGD and will have been processed prior to gap identification. I hypothesize that valency completeness facilitates processing by making it clear how all the phrases containing the gap are structurally connected within the FGD, and I attribute the SSC and other subject-object asymmetries such as 65 in English to this processing motivation.

(65) a. *What, s[what] [did NP[the title of } O]\, VP[amuse John]]
   b. What, s[what] [did John VP[read NP[the title of } O]]

Converting a sentential subject into an infinitival complement does not help on this occasion, plausibly because the complement’s subcategorizor, fascinate, is still not in the FGD.

(66) a. *What, s[what] [did VP,[to read } O]\, VP[fascinate Sue]]
   b. What, s[what] [did it VP[fascinate Sue VP,[to read } O]]

The preference in 67, therefore, seems to obtain.

(67) Valency Completeness: The human processor prefers FGDs to include the subcategorizers for all phrases within the domain that contain the gap.

Sixty-seven makes a prediction for languages that have conventionalized the SSC: they should be those, like English, whose basic word order positions the subcategorizer for a sentential complement outside the FGD for a gap within this complement. If the subcategorizer occurs within the FGD in the basic order, then there need be no subject-sensitive constraint (though there might, of course, be a general constraint on extraction out of all finite subordinate clauses). Consider Malagasy which does not obey the SSC (Keenan 1972b:175):

(68) ny vehivavy, s[what] [did noheverin-dRakoto VP,[to read } O]\, s[fa nividy vary O]]
   the woman that (was) thought by Rakoto that bought rice
   ‘the woman who was thought by Rakoto that (she) bought rice’

Malagasy allows only subject gaps, including (subject) gaps within sentential subjects. The finite subordinate clause S’ in 68 is a (passivized) sentential subject and it contains a gap in its subject position coreferential with vehivavy. The basic word order of Malagasy is VOS, which means that a sentential subject already stands in what is in effect the extraposed order of the English 64b, and the subcategorizing matrix verb noheverin is necessarily encountered before the subcategorizer of the gap (nividy). Hence noheverin is available to clarify the structural role of the containing S’ by the time the gap

---

\(^{28}\) That valency completeness is a preference rather than a universal requirement is suggested by the following example from German:

(i) Was, s[what] [did VP,[to kaufen] versucht]
   what have you to buy tried
   ‘What have you tried to buy?’

The FGD proceeds from the was to the subcategorizer for the gap, the verb kaufen in VP’. This domain excludes versucht, therefore, which is the subcategorizer for the VP’ complement. On the other hand, the ‘control domain’ holding between the matrix subject du and the infinitival kaufen is shorter than in the corresponding extraposed sentence (Was hast du versucht zu kaufen?\?), and this provides a competing motivation for (i).
PROCESSING COMPLEXITY AND FILLER-GAP DEPENDENCIES ACROSS GRAMMARS 279

subcategorizor within it is recognized. Japanese also permits gaps in sentential subjects, even though gaps precede fillers in relative clauses in this language. It is significant that the subcategorizor for the containing S’ can again be in the FGD in this reverse ordering, just as it can be in Malagasy and in the extraposed sentential structure of English. Example 69, taken from Kuno 1973:241, illustrates this.

(69) [[kare ga Oi kaita kotoIs, ga yoku sirarete-iruIs bun_i he su wrote that su well known-is article
‘the article which that he has written (it) is well-known’

The FGDs of English that avoid gaps within containing subjects result in larger domains. The FGD of 64b is longer than that of 64a, yet 64b is better, which goes against the normal domain minimization and reduction preference. But the extra material within 64b includes the subcategorizor for the containing clause, and this subcategorizor facilitates processing by clarifying the structural relations in the path from filler to gap, as we have seen. Extra length is therefore tolerated when it directly aids the processing of structural relations, in the same way that extra material in the form of pronoun retention is welcome when it clarifies the filler’s role in the dependent sentence (§3.4). Domain minimization and reduction apply in the normal case, unless they are in competition with independently motivated processing preferences such as 67. This means that constituents and operations that add to the processing load of filler-gap dependencies will always be minimized, whereas those that facilitate processing may not be.

7. CONCLUSIONS. I have argued that crosslinguistic universals and variation patterns in filler-gap and filler-pronoun dependencies are ultimately explained by processing complexity. Processing preferences lead to statistical and nonstatistical universals, to implicational hierarchies in the cutoff points for grammatical conventions, and to variation in structures for which independently motivated preferences are in competition. A processing approach can also unify the many different factors that contribute to wellformedness in this area, syntactic, semantic and lexical, and can explain why grammatical conventions make the distinctions they do. They are all relevant for defining the relative ease or difficulty of gap identification and processing.

The manner in which grammars have responded to processing through conventionalization turns out to be language- and also construction-specific, but is constrained by processing complexity. For any given language and for any given construction, lower positions on each complexity hierarchy will never be more productive in the grammar of that language than higher positions. Within the set of grammatical constraints fixed by different languages there may also be conventionalized processing preferences and different levels of acceptability for comparable structures generated by identical grammars (cf. §4.2).

The preferences that emerge from this crosslinguistic comparison are, I submit, the same preferences that we see in performance in languages that allow structural options and alternatives. Grammatical conventions are ‘frozen’ processing preferences. The preferred word orders in languages permitting choices, for example, are generally those that become grammaticalized in the languages and structures with fixed ordering conventions (Hawkins 1994, 1997). A historical change in one area of the grammar can result in new performance preferences within another, and these new preferences can lead to further changes in the grammatical system, as processing efficiency is incorporated into the grammar itself. A shift in verb position, for example, has consequences for the efficiency of wh-fronting and for the choice of postnominal versus prenominal relative clauses (section 6.1). In a verb-initial language, wh-fronting and postnominal
relatives will be preferred, in a rigid verb-final language they will not be. Seen from this explanatory perspective, grammars and grammatical evolution can be viewed as complex adaptive systems (Gell-Mann 1992), with ease of processing driving the adaptation, in response to prior changes (see Kirby 1994, 1999 for an illustration of this using computer simulations of grammars in evolution).

It is perhaps because the conventions of particular languages differ with respect to complexity that linguists have often been reluctant to accept processing explanations for grammar. If difficulty is supposed to explain the absence of structure X in language Y, why can X show up as grammatical in language Z? But difficulty is a matter of degree, not an absolute prohibition. And a structure that is difficult with respect to one processing preference (Fillers First) can have advantages with respect to another (cross-category parallelism and EIC), or can be motivated by the absence of expressive alternatives. The hierarchies and distributional preferences across languages reflect these degrees of processing difficulty, while the competing motivations are apparent in the variation patterns. You cannot see these preferences and patterns in just a handful of languages, however; larger samples and a diversity of language types are required.

The conventions of each language still need to be described, of course, and for this we need the best descriptive apparatus we can get. A processing approach provides a principled basis for choosing between alternative analyses. For example, any description needs to be sensitive to possible differences between relative clauses and WH-questions and should not necessarily set constraints for a whole language. There can be differences between environments whose complexity reduces to different processing causes, such as WH-islands versus sentential subjects. The former permit gaps in infinitival complements in English, the latter do not. Languages with more productive gaps than English will generally extend their use down several complexity hierarchies at the same time, as in Scandinavian. But possible discrepancies are explicitly predicted in our approach, both between filler types and between environment types, and grammars will need to capture the permitted structural combinations in the most general, but descriptively adequate, manner.

Finally a brief word about learnability. The nonuniversality of subjacency is evidence against its innateness (see §4.2). If, instead, the variation in filler-gap dependencies across languages is more adequately explained by complexity and efficiency, then innate grammatical constraints on gaps will not be available to the learner. The architecture of the human processor can be innate, of course, and so can certain other properties that figure in a filler-gap dependency. But I see the constraints themselves—on size, on additional syntactic operations, and so forth—as resulting from conventionalized grammatical rules that have responded to processing in the evolution of different languages. This conclusion will be unwelcome to learnability theorists, since subjacency constraints have always been a prime example of how innateness makes available grammatical restrictions that (it is claimed) cannot be learned from positive evidence alone, see for example Hoekstra & Kooij 1988. My approach suggests a different solution to the negative evidence problem (which is a genuine problem regardless of one’s theoretical position on innateness, cf. Bowerman 1988). Instead of regarding the constraints on filler-gap dependencies as innate, we can hypothesize that learners will comprehend their input and will postulate grammars in accordance with ease of processing. Smaller domains will precede larger ones in grammar acquisition, parallel function filler-gap structures will come before nonparallel function ones, and so on. In this way the simplest environment on each hierarchy will be postulated first on account of processing ease, and any extension beyond that to lower hierarchy positions will be forced on the learner
by positive evidence from the relevant language. Hierarchies of complexity like the AH have already been shown to predict the order of language acquisition (see Hawkins 1987). I propose now that the processing complexity that motivates these hierarchies may also explain the negative evidence problem by providing a principled limit on initial hypotheses and by requiring extensions beyond this to be justified by the data of experience. A similar sequence of events has been proposed in Berwick’s (1985) SUBSET PRINCIPLE, for which processing can be argued to provide the motivation (see Saah & Goodluck 1995 for a similar conclusion). It is also possible that learning is triggered by encountering grammatical instances of a low hierarchy position gap with a certain filler, whereupon the grammaticality of all simpler instances of that construction can be inferred. These questions require further research.

REFERENCES


Antinucci, Francesco; Alessandro Duranti; and L. Gebert. 1979. Relative clause structure, relative clause perception, and the change from SOV to SVO. Cognition 7.145–76.


1999. The relative ordering of prepositional phrases in English: Going beyond manner-place-time. Los Angeles: University of Southern California, MS.


1993. Bridging gaps with longer spans: Enhancing ERP studies of parsing. Poster presented at the Sixth Annual CUNY Sentence Processing Conference, University of Massachusetts, Amherst.


