WH-QUESTIONS IN ENGLISH: A PARSING-ORIENTED THEORY OF LOCALITY

Gregory F. Coppola
Simon Fraser University

1. Introduction

In English, the wh-word (or interrogative word) in the canonical question form appears preposed, at the front of the sentence. In such a question, there is an empty syntactic position later in the sentence at which the wh-word is to be interpreted. Following Fodor (1978) we will refer to such questions as *filler-gap* sentences. The preposed wh-word is called the *filler* and the empty argument position where the filler is to be interpreted is called the *gap*. Several researchers, especially Ross (1967), Chomsky (1973) and Huang (1982) have noted constraints, called *locality constraints*, on where the gap can occur in sentences of, among other languages, English[1].

This paper will be concerned with deriving three of these constraints from general principles. These general principles are simple, stateable in terms of surface structure without recourse to facts about the derivation of a sentence, and can be motivated on the grounds of parsing feasibility. Further, these general principles predict other, attested syntactic phenomena, a fact which attests to their generality.

The first type of locality constraint under consideration is Ross' (1967) "Complex Noun Phrase Constraint" (CNPC), which bars a gap from occurring inside an S which is dominated by an NP, as in (1a). (In what follows, the gap will always be represented by the symbol t_i. t_i is coindexed with the wh-word which is to be interpreted in its place.) The second is Ross' "Sentential Subject Constraint" (SSC), which bars a gap from occurring inside a sentential subject, as in (1b). The third is Huang's (1982) "Adjunct Condition", which will henceforth often be referred to as the "Adjunct Island Constraint" (AIC), which bars a gap from occurring inside an adjunct[2], as in (1c).

(1) a. * What$_i$ did Mary like [$_{NP}$ the boy [$_S$ who ate t_i]]

[1] In their terms they gave constraints on 'movement' or 'extraction' of the wh-word, which they held to originate in its canonical position. However, the assumption that syntactic items 'move,' in some meaningful sense, during the production of a sentence is a theory-internal one. While the notion of movement may be correct, our account of wh-constraints does not depend on facts about the derivation and so, not needing to prejudge the issue, is stated in a filler-gap framework.

[2] Here, adjunct positions are equated prepositional phrases that are directly dominated by sentences.

Actes du congrès annuel de l'Association canadienne de linguistique 2008.
Proceedings of the 2008 annual conference of the Canadian Linguistic Association.

b. * What$_i$ did [$_S$ for Ernie to win t_i] seem unlikely

c. * What$_i$ did Bush order the attack [$_{PP}$ after [$_S$ Saddam mentioned t_i]]

We will have need of a convenient label for the three constraints under discussion and, for lack of a better name and for the purposes of this document alone, let us refer to these three constraints collectively as the *RH constraints* (i.e. the Ross-Huang constraints).

The major early approaches to reducing the RH constraints to more general principles, such as Chomsky 1981, 1986, Huang 1982 and Rizzi 1990, posited that ungrammaticality arose due to facts about (or at least stated in terms of) the derivation of a sentence. Later, Chomsky (2000, 2001), drawing on Uriagereka (1998), proposed that locality constraints are ultimately due to an aspect of the speech production architecture that allows an optimization in the production of phonological forms. In other words, *locality constraints arise due to the nature of the speech* production *architecture*.

Another tradition, which began with Fodor (1978), focused on the *parsing* of wh-questions. Hawkins (1994, 1999, 2004) applied the intuitions of this psycholinguistic tradition to give a parsing-oriented formal account of locality phenomena, deriving typological facts from general principles. Hawkins basic thesis is that, “[f]iller-gap dependencies are difficult to process,” (1999, p. 245).

But, since optimization of sentence structure for the benefit of the parser “can be at variance with the demands of expressive power,” (1999, p. 251), languages *conventionalize* a compromise, allowing the possibility of filler-gap sentences but preventing the gap from occurring in positions that excessively tax the parser.

This paper follows the line of reasoning of Fodor and Hawkins and reduces the RH constraints to general principles which can be motivated on the grounds that they prevent the burden on the parser from growing beyond the point of feasibility.

This paper omits analysis of an important kind of locality constraint, namely the “Wh-island Constraint” noticed by Chomsky (1973) and illustrated in 0:

(2) ?What$_i$ did Mary say [$_S$ who saw t_i]

Basically the Wh-island constraint dictates that, in a question with multiple wh-words, the preposed wh-word must be the one whose canonical position is earliest (or left-most).

However, informants are not unanimous about the extent of the deviance of such examples. More importantly, Wh-Island Constraint violations arise from the interaction of multiple wh-words and occur regardless of the syntactic structure involved. In contrast, the RH constraints under discussion here are due only to the syntactic structure of the question. Thus, the RH constraints are more likely to illuminate facts about sentence structure generally.

In section 2, a specific set of principles are given that derive a (more accurate) variant of the RH constraints. These general principles predict other, as yet unnoticed kinds of deviant question sentences, as we will see. Then, in Section 3 we will discuss how this theory fits in with a broader theory of sentence parsing and we will compare, on theoretical grounds, the notion of a parsing-oriented theory of locality constraints to the speech-production-oriented theory of Chomsky (2000, 2001).

2. The Analysis

Following Chomsky 1957, we will analyze sentences as strings of words contained in recursively-embedded constituents so that each sentence corresponds to a (graph-theoretic) tree where terminal nodes correspond to words and non-terminal nodes correspond to constituents (or *phrases*). All phrases posited will be constituents at the level of surface structure. And, as we are unconcerned as to how the surface structure was produced, will not refer to transformations. See (4) for an example tree.

We will make use of only four phrasal categories: Sentence (S), Noun Phrase (NP), Prepositional Phrase (PP) and Adjective Phrase (AP). That is, all internal nodes on the tree will bear a label of one of these types.

I am commonly asked to justify this choice of phrasal categories, which constitutes an abrupt departure from Chomsky's (1970) X-bar theory, which is currently popular in North America. A justification is given in Section 3.1, coming more naturally after the theory is developed. For now, I ask the reader to simply suspend disbelief.

Recall that the 'gap' in a filler-gap question is the 'empty syntactic position at which the wh-word must be interpreted.' As such, the gap does not actually exist as an overt entity in the input signal received by the hearer. The gap is, technically, an absence. In actual fact, we assume that the parser decides where the gap was after reading *past* it. For example, consider:

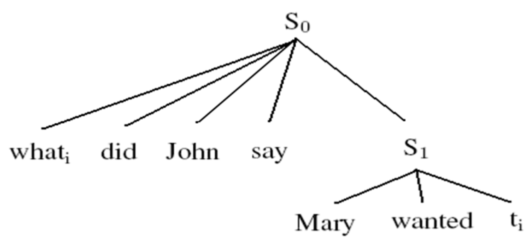
(3) a. Who did Mary say t_i saw Ernie?

b. Who did Mary say Ernie saw t_i yesterday?

c. Who did Mary say Ernie saw t_i?

In (3a), the parser would realize at *saw* that the subject is missing. In (3b) it would realize at *yesterday* that the object is missing. In (3c) it would realize at the end of the sentence that the object is missing.

While the gap is technically an absence, it turns out that for the purposes of stating our theory here, it is easiest to speak of and draw the gap as an overt object. So, on the tree, we will draw the gap as a terminal node. For example, in (4), the gap, drawn as t_i, is a child to S1, which is of type S.



(4)

In what follows, each non-terminal node has a *name* and a *type*. The type is the kind of phrasal category that that node is an instantiation of (either S, NP, PP or AP). The name is used to label and identify the node. For example, in 0, we see a node of type S whose name (and therefore label) is S_0. Though we will not develop formal conventions, the type should be clear from the label even if not explicitly stated. Further, in some cases, the node's name will also be its type (e.g. a node of type NP might be called NP), a slight abuse of notation from which no confusion should result.

One complication is required, however. We will require that each phrase contain more than one overt element, and the gap, t_i, is not an overt element. So, any phrase of the form XP = $[_{XP}$ X $t_i]$[3] must be *collapsed*—i.e. the non-terminal node XP is deleted and the children of XP become children to XP's mother. (To be precise, if YP = $[_{YP}$ A $[_{XP}$ X t_i] B] then, when XP is collapsed, we will have YP = $[_{YP}$ A X t_i B].)

So, consider, for example:

(5) $[_S$ Who$_i$ did John go home $[_{PP}$ with t_i]] → $[_S$ Who$_i$ did John go home with t_i]

The symbol → will always indicate that the collapse operation has occurred. Here, the PP is deleted and its children become children to the S. From now on, we will use square brackets to indicate constituent structure rather than drawing trees.

We are now ready to state the constraints that will derive (a more accurate variant of) the RH constraints, as well as predict the existence of other unacceptable wh-question forms:

(6) **Sentential Recursion Constraint (SRC)**:

a. All of the gap's mother's ancestors must be of type S

b. Every S that is an argument has no non-children to its right

From now on let us refer to (6a) as SRC1 and (6b) as SRC2. By non-children we mean ancestors and siblings. By M being "to the right of N" it is meant that M is a right-more sibling of N or else is the right-more sibling of an

[3] In English, it happens that the phrase $[_{XP}$ t_i X] never occurs.

ancestor of N. (Under these definitions, SRC2 is somewhat redundant but more clear.)

The mechanics of the SRC may be unclear to the reader at this point but the matter will be clarified as we apply the SRC to examples.

Let us start by considering a CNPC violation:

(7) *[$_S$ What$_i$ did John believe [$_{NP}$ the fact [$_{S1}$ that Bill wanted t_i]]]

Here, the gap's mother is S1. But, S1's mother is NP, which is of type NP. Thus, S1 has an ancestor which is not of type S. Thus, it is not the case that the gap's mother's ancestors are all of type S. Thus, (7) is an SRC1 violation. All CNPC violations are derived similarly.

Consider next the A(djunct)IC violation:

(8) *[$_S$ What$_i$ did George order [$_{NP}$ the attack] [$_{PP}$ after [$_{S1}$ Saddam took t_i]]]

Here, again, the gap's mother is S1, which has a PP ancestor. PP is not of type S, so (8) is barred by SRC1.

Note, however, that our account differs slightly in its predictions compared to the blunter Adjunct Condition of Huang because, under the SRC, the following sentences are allowed:

(9) a. [$_S$ Who$_i$ did George go home [$_{PP}$ with t_i]] → [$_S$ Who$_i$ did George go home with t_i]

b. [$_S$ What$_i$ did you do that [$_{PP}$ for t_i]] → [$_S$ What$_i$ did you do that for t_i]

Gaps are allowed in contexts like these because, in each case, the PP contains only one overt element and so is collapsed and thus the gap's mother is the matrix S. The matrix S has no ancestors and so it trivially satisfies the condition that all of its ancestors are of type S. (Note that, even if each PP were not collapsed, the examples of (9) would still not constitute a SRC1 violation since each of each PP's ancestors would still be of type S.)

My informants unanimously accept (9). Thus, we have an interesting empirical discrepancy between the predictions of the SRC one the one hand and those of Huang's (descriptive) Adjunct Condition and those theories that sought to explain it on the other. Both Huang 1982 and Chomsky 1986 attributed the prohibition on extraction of a wh-word from an adjunct to the fact that wh-words could not be extracted out of domains that were not 'governed[4]' in appropriate ways. But, examples such as (9) suggest that gaps can occur in adjuncts, so the approach of barring extraction from all adjuncts due to the nature of the adjunct position is too coarse.

[4] 'Government' was a notion in vogue for a period after Chomsky 1981. Argument positions aside from the subject were typically governed while adjuncts were not.

Also, note that a word which introduces a subordinate sentence that is not subcategorized by a verb—e.g., *since*, *because*, *if*—is treated as a preposition here, correctly predicting the deviance of sentences such as:

(10) a. * $[_S$ What$_i$ does Mark eat fish $[_{PP}$ if $[_{S1}$ Ernie does t_i]]]

b. * $[_S$ What$_i$ does Mark eat fish $[_{PP}$ $[_{S1}$ because Ernie does t_i]]]

Before leaving the topic of PPs, let us note that one *can* find examples of questions with structures identical to those in (9) that are deviant, such as:

(11) ?$[_S$ What$_i$ did you come home $[_{PP}$ after t_i]] → ? $[_S$ What$_i$ did you come home after t_i]

It seems, however, that this deviance is a semantic/pragmatic issue. (11) involves more presupposition than its acceptable counterpart *When did you come home?* (i.e., because (11) presupposes the answerer came home *after* something). Hawkins (1999, see pp. 271—273), following Kleunder (1992), argues that, roughly, for question sentences in general, questions with 'too many' entailments are unacceptable[5]. This is the same phenomenon illustrated in the distinction in acceptability in (12) and (13):

(12) a. Who did Mary see *a* picture of?

b. ? Who did Mary see *the* picture of?

(13) a. How angry did Mary *say* that John was?

b. ? How angry did Mary *whisper* than John was?[6]

In (12b), the word *the* adds a presupposition of uniqueness in addition the presupposition of existentiality. In (13b), *whisper* adds the presupposition that something was not only said but said softly. More precisely, the relevant principle may phrased thus: if sentence S is acceptable and the entailments of T are a subset of those in S, then T is also acceptable; if S is unacceptable and the entailments of T are a superset of the entailments of S, then T is also unacceptable.

---

[5] His actual thesis is that, "The human [parser] prefers to minimize the amount of semantic information that needs to be processed" in a filler gap sentence (p. 271). For our purposes, we need not dwell on the fact that 'semantic information that needs to be processed,' 'presupposition,' and 'entailment' denote concepts which actually differ in important ways and therefore may correspond to different theses about the nature of the deviance of 'the pictures of' questions. We simply require here that the deviance is due to extra-syntactic considerations.

[6] This example is from Hawkins (1999), which is in turn adapted from Cullicover and Wilkins (1984).

Some might argue that the deviances of (12b) and (13b) are a result of hidden syntactic structure. However, since there is clearly a correlation between acceptability and number of entailments, the use of hidden structure, especially if the presence of the hidden structure in turn correlates with the number of entailments, seems superfluous, unless one's theory is necessarily committed to it.

A final prediction made by the SRC1 is that a gap cannot be found inside an S dominated by an AP, resulting in what we might call an *Adjective Island Constraint*, a phenomenon which has apparently gone unnoticed so far:

(14) a. * [$_S$ What is Mary [$_{AP}$ upset [$_S$ that Ernie took t_i]]]

b. * [$_S$ How is Mary [$_{AP}$ upset [$_S$ that Ernie took the cake t_i]]]

Turning to the Sentential Subject Constraint, consider:

(15) [$_S$ What$_i$ is [$_{S1}$ that Ernie will eat t_i] likely]

This is simply an SRC2 violation as S1 is an argument of type S that has material to its right. SSC violations are derived this way in general.

However, the SRC2 also predicts other types of unacceptability (i.e. it is a more general principle than the Sentential Subject Constraint):

(16) a. ? [$_S$ Who$_i$ does [$_{S1}$ that Ernie will eat cake] bother t_i]

b. ? [$_S$ Who$_i$ did Mark say [$_{S1}$ Ernie was dating t_i] to Mary]

c. ? [$_S$ Who$_i$ did Mark say [$_{S1}$ Ernie was dating Candy] to t_i]

In each case, S1 has material to its right and so violates SRC2.

At this point, we should stop to verify that certain acceptable questions can still be formed without violating the SRC. The reader should be able to verify at this point that the examples in (17) are each predicted to be acceptable.

(17) a. [$_S$ What$_i$ did Mary say [$_S$ Dave wanted t_i]]

b. [$_S$ Who$_i$ did Mary say [$_S$ t_i wanted Dave]]

c. [$_S$ Who$_i$ did Mary say [$_S$ Dave gave a gift [$_{PP}$ to t_i]]] →
[$_S$ Who$_i$ did Mary say [$_S$ Dave gave a gift to t_i]]

Next consider sentences with infinitival constituents:

(18) a. What did Mary say she wanted to eat?

b. What did Mary persuade Ernie to eat?

Supposing we analyze infinitivals as being of type S, we get the correct predictions. The following would then be the parses of (18), showing no SRC violations:

(19) a. $[_S$ What$_i$ did Mary say $[_S$ she wanted $[_S$ to eat t_i]]]

b. $[_S$ What$_i$ did Mary persuade Ernie $[_S$ to eat t_i]]

(20) $[_S$ Who$_i$ did Mary say $[_S$ she wanted $[_S$ to go to the party $[_{PP}$ with t_i]]]] → $[_S$ Who$_i$ did Mary say $[_S$ she wanted $[_S$ to go to the party with t_i]]]

Finally, we consider the possibility of a gap occurring inside a 'pictures of' NP context—i.e. an NP that contains what we might analyze as a PP. Consider (21):

(21) a. $[_S$ What$_i$ does Mary like $[_{NP}$ pictures $[_{PP}$ of t_i]] → $[_S$ What$_i$ does Mary like $[_{NP}$ pictures of t_i]]

b. $[_S$ What$_i$ does Mary like $[_{NP}$ stories $[_{PP}$ about t_i]] → $[_S$ What$_i$ does Mary like $[_{NP}$ stories about t_i]]

After each PP is collapsed, the gap's mother becomes the NP, whose only ancestor is the matrix S, thus satisfying SRC1. This contrasts with the behavior of a complex noun phrase, in which the gap's mother is an S dominated by an NP, violating the SRC1.

The reader will note that it was only in these 'pictures of' contexts that the rule of collapsing phrases of the form XP = $[_{XP}$ X t_i] has had any practical effect in the course of this paper. If it were not for the collapsing rule, the theory would make incorrect predictions, predicting unacceptability in too many cases. Given its limited range of application, the collapsing rule is at this point stipulatory.

On this point, we note that sentences such as (22) seem unacceptable:

(22) a. * $[_S$ What$_i$ does Mary like $[_{NP}$ pictures $[_{PP}$ of t_i] $[_{PP}$ by Pete]]] → $[_S$ What$_i$ does Mary like $[_{NP}$ pictures of t_i $[_{PP}$ by Pete]]]

b. * $[_S$ What$_i$ does Mary like $[_{NP}$ pictures $[_{PP}$ by Pete] $[_{PP}$ of t_i]]] → $[_S$ What$_i$ does Mary like $[_{NP}$ pictures $[_{PP}$ by Pete] of t_i]]

The unacceptability of these sentences probably parallels the unacceptability of (12b) repeated here:

(23) ? Who did Mary see *the* picture of?

That is, perhaps the questions in (22) have too many entailments.

Whatever the reason, it may be that the only allowable 'pictures of' contexts are of the form [$_{NP}$ Det N [$_{PP}$ P t_i]] (or [$_{NP}$ Det N P t_i]). Perhaps a better analysis might be achieved by making use of this fact in stead of employing the collapsing rule.

In any case, assuming that spoken sentences are generated in some way corresponding to a system of Context Sensitive Rules of the sort discussed in Chomsky 1957, there is not necessarily anything to prevent redundant rules in the system. So while more general syntactic theories are certainly preferable to less general ones, it is not antecedently clear how far this process of generalization will go nor, conversely, how much of syntactic description will be 'stipulatory'.

3. Discussion

3.1 Motivating the Set of Phrasal Categories S, NP, AP, PP

A common point of discomfort for audiences when presented with this theory it is not given in the context of a well-established framework. Especially salient, given the times, is its departure from Chomsky's (1970) X-bar theory. So, let us take a moment to defend this departure.

Of course, it would be preferable if this theory were an extension of some theory that had already demonstrated its ability to account for an array of interesting syntactic phenomena. But, every theory must start somewhere. And, assuming that this theory is successful in the domain of its application (i.e. it derives the RH constraints and more), it deserves an amount of consideration on the grounds of this success alone.

Also, while it is true that more work has been done along principles-and-parameters/X-bar/minimalist grounds, modern minimalist theory, as far as I understand, does not have an unproblematic account for what we have been calling the RH constraints, as is discussed in Section 3.3, certainly a demerit considering that research along these lines has been going on for over thirty years.

In fact, there is a growing chorus of voices asserting that the principles-and-parameters thesis (advanced in Chomsky 1981) is a failure. Newmeyer (2005, pp. 76—102) gives a well-reasoned critique to this end. Thus, it may be time to start considering openly new, alternative frameworks for analysis than the minimalist analysis, which is, after all, simply "one of a number of alternatives being explored" (Chomsky 2001, p. 1).

Taking a less combative line, one may suppose that there *is* more structure to syntactic trees (or analysis generally) than represented here but that it is simply not relevant for deriving the RH constraints. (One may even suppose all of X-bar structure.) That is, the phrases S, NP, AP and PP would be similar to the *bounding nodes* of Chomsky 1986 or the *phases* of Chomsky 2001 in which only a subset of the phrases considered to exist generally are singled out in the analysis of locality.

Thus, one need not accept that S, NP, AP and PP are the *only* syntactic categories in order to accept this analysis. In particular, there seems to be a closer connection between the verb and its internal (or post-verbal) arguments than there is between the verb and its subject, which one may model using something like a V(erb) P(hrase). In such a case, this VP is simply not involved in the statement of the SRC and may be a different sort of structure than, e.g., NP altogether.

In any case, depending on the specifics of the reader's favorite theory, he or she may be able to map these categories to analogous categories or structures in that theory, ignoring particulars.

Further, the set of categories S, NP, AP and PP can be motivated as a natural class. As will be discussed in Section 3.2, this is a parsing-oriented theory and there is particular interest in which *right edges* of constituents the parser must remember to expect to find if it is to recreate the syntactic tree as it parses a sentence left-to-right.

Consider a simplified modern X-bar sentence structure:

(24) $[_{CP}\ [_{TP}\ [_{VP}\ [_{vP}\ \ ]]]]$

In a structure like this, the right edges of the CP, TP, VP and *v*P all *necessarily* coincide (at least for the English speaker). Finding the right edge of the CP entails finding the right edges of TP, VP and *v*P (except possibly in extremely rare sentence structures[7]). It would be a waste of working memory for the parser to record that it must find the right edge of each of these individually.

Instances of each of the phrasal categories chosen here (i.e. S, NP, AP and PP) can occur with their right edges free (i.e. not coinciding with the right edge of an instance of another category). The requirement that all phrasal categories in a theory can have instances occur with their right edges free specifies a *natural class of theories*, of which the theory developed here is a member. Thus, the class of phrasal categories S, NP, AP and PP is not arbitrary.

3.2 The SRC Allows Parsing Optimization and Bounds on Computational Complexity

It was claimed at the outset that the SRC, used to derive the RH constraints, can be motivated on parsing grounds, a matter which we demonstrate now.

We will assume that, in order to correctly reconstruct the argument structure of a sentence, the parser must (and so does) reconstruct (at least) some of the sentence structure. Without knowing the structure of the sentence the parser would presumably be unable to, e.g., know whether a given verb was, e.g., the matrix verb or the verb in a relative clause.

---

[7] One example of a sentence in which the right edge of a presumed VP arguably does not coincide with the right edge of its containing CP is *Jon wanted to come home, and come home he did*. In response, let us note that sentences like this are extremely rare and marked and it would be curious if a significant price in working memory were paid on the parsing of *every* sentence to accommodate outliers like these.

We assume that parsing happens online (or ‘word-by-word’) and in a left-to-right fashion (where words more to the ‘left’ are those encountered earlier). As stated earlier, we assume that words are the smallest relevant syntactic units (for our purposes) and that words must be grouped into recursively embedded constituents such that the sentence corresponds to a tree. Thus, in order to recreate the sentence structure, the parser must discover a syntactic tree and so correctly identify the left-most and the right-most word in each constituent.

For simplicity, let us suppose that, given some matrix sentence S_M with structure P, the parser is going to be able to, upon encountering some word W, correctly identify W as the left-most word in any constituents that W actually *is* the left-most word in in the correct structural analysis, P, of S_M.

Consider, as an example:

(25) $[_S$ $[_{NP}$ The duke] likes Mary]

So, here we are assuming that after reading the word *The*, the parser recognizes that *The* is the left-most word in an S constituent because it actually *is* the left-most word in an S constituent in the correct parse of the sentence. And thus, the parser must find the right-most word in the S.

Also, since determiners introduce noun phrases, we assume the parser also realizes that *The* is the first word in an NP constituent. (Note that *The* is the first word in more than one constituent.) Thus, the parser must find the right-most word in (or right *edge* of) the NP.

In order to find the right edge of each *open constituent*—i.e. of each constituent C such that, after parsing up to some word W, the left edge of C has been encountered but the right edge of C has not yet been encountered—the parser must store in memory enough information to know which open constituents must have right edges found for them and in which order.

For the sake of argument, let us suppose that the parser keeps a list of the open constituents, ordered so that constituents whose right edges must be found sooner are listed earlier. Call this list the *stack*. (Whether the stack should be interpreted literally is discussed in Section 3.3.)

So, while parsing (25), after reading *The*, the parser’s stack is <NP, S>, since it must find a right edge for NP and a right edge for S, in that order. (The convention for drawing the stack should be obvious, the list is enclosed in angled brackets and the constituents whose right edges must be found sooner are left-more in the list.)

Now, note that by SRC1, if a gap is found, it must either be in a context like (26a) or like (26b):

(26) a. $[_S$ ... $[_{Se}$ X t_i Y]]

b. $[_S$ ... $[_{Se}$ U $[_A$ X t_i Y] Z]]

Here, U, X, Y and Z are strings of terminals. Se is of type S and A has any type other than S. That Se in (26a) and A in (26b) are each dominated only by nodes of type S is guaranteed by SRC1.

Also, if one node of type S, S_{child}, is directly dominated by another of type S, then, in particular, S_{child} is not dominated by a PP and so, we will assume, S_{child} must be an argument to its containing S. Thus, SRC2 applies to dictate that there is nothing to the right of S_e (this happens in both (26a) and (26b)).

The fact that a gap can be only found in the contexts in (26) has important ramifications. Suppose the parser is parsing a filler-gap sentence left-to-right and is currently parsing some phrase of type S, call it S' (i.e. S' was the last phrase added to the stack). Now suppose that the parser encounters a word which is the left edge of a sentential argument to S'. Call this sentential argument S''. An example of this situation is illustrated in (27), where the parser recognizes that *Ernie* is at the left edge of an embedded S argument, called S'':

(27) [$_{s'}$ What did Mary say [$_{s''}$ Ernie wanted t_i to do]]

Now, by the SRC2, the parser can assume that no material is to the right of S''. That is, the right edge of S'' is the same as the right edge of S' (the right edge of an S does not *generally* coincide with the right edge of its mother, but the SRC2 guarantees this in the case of a filler-gap sentence).

Because of this, the parser can perform an optimization. The parser can *replace* S' on the stack with S'' instead of adding S'' to the stack *in addition to* S' because both are guaranteed to have the same right edge (which also is the end of the sentence). In the (27) example, the stack after reading *Ernie* can be <S''> instead of <S'', S'>. Thus, when recursing into sentential arguments to find a gap, the stack never grows.

Now, let us suppose that the parser is able to recognize the gap (empty argument position) after it has read the word following the gap, which is the word *to* in (27). When the parser finds a gap, this gap is, again, either in a context like (26a) or one like (26b). First, suppose the gap is in a context like (26a). Here, when the gap is found, the stack has only one thing on it, which would be Se in the terms of (26a) or S'' in the terms of (27). Thus, the stack has a size of 1 when the gap is found.

Second, suppose the gap is in a context like (26b). In the terms of (26b), after parsing up to the end of the string U, the parser is, for reasons discussed above, in a state where its stack has size 1. The parser will find the gap inside A, where it will have a stack size of 2.

Thus, by assuming the SRC, the parser can be guaranteed that the stack size when the gap is found will never be above 2. Note that this bound on stack size is only possible because of the stack replacement optimization that the SRC allows. That is, SRC2 allows the stack replacement algorithm. Then SRC1 in combination with the possibility of the stack replacement algorithm allows the bound on stack size when the stack is found.

What is the importance of the number 2 as the bound on the stack size? The most important thing is probably that this bound (of 2) is *constant*—the bound on the stack size when the gap is found is not a function of the sentence's length or (parse-tree) depth, as it would be barring some constraint like the SRC. Also of importance is that the number 2 is low but, depending on how one wants to count things, it might as well be, e.g., 1 or 3, so long as the bound is fixed and constant for all acceptable sentences.

3.3 A Parsing-Oriented vs. a Speech-Production-Oriented Account of Locality Constraints

Wanner and Maratsos (1978) demonstrated that parsing a filler-gap sentence places a significant strain on working memory. Stowe (1986) demonstrated that, while parsing left-to-right, the parser will sometimes, due to the structure of the sentence, posit a gap earlier than it should and then be forced to reanalyze the sentence as more of it is seen. In other words, parsing a filler-gap sentence is a significantly more difficult activity than parsing a declarative sentence.

As noted, the 'stack' discussed in the last section may only be a metaphor. However, there would seem to be a number of important aspects of computational complexity that grow as the stack size grows, including parse-tree depth, number of right edges to be found in order to close constituents and number of alternative possible parses to be adjudicated between.

Given the demonstrable strain involved in parsing any filler-gap sentence, it would seem plausible that something will need to prevent these aspects of complexity from growing without bound, or else the task of finding the gap may become infeasible.

Well, if we assume that we have shown that *some* kind of constraints on filler-gap sentences will need to exist for the benefit of the parser, it would seem to be a theoretically more parsimonious account if *all* constraints exist for the benefit of the parser. Indeed, the SRC is very simple and both parts of it are needed to derive the optimization and complexity bound. At the same time, these two principles derive the most famous and most-studied examples of structural filler-gap constraints, namely the CNPC, SSC and AIC. Thus, there seems to be no need of the hypothesis that locality constraints exist for reasons of speech-production.

Indeed, suppose we ask the question the other way around: is there evidence that, if the parser could handle filler-gap sentences of arbitrary complexity, speech-producers would be unable to produce locality constraint violations? I am not aware of anything more than conjecture that this is the case.

Arguments in favor of phase theory, such as those by Chomsky (2001), have the following form:

(29) If we assume that derivation happens 'in phase' *to allow a speech-production optimization*, then we predict grammatical behavior X and, in fact, we see behavior X, therefore derivation happens in phase *to allow a speech-production optimization*.

Even if the theories of this form were empirically adequate in what we might call a *grammatical sense*—meaning that they correctly predicted acceptability and deviance of the various sentence types—we could still account for the data equally well by dropping the 'to allow a speech-production optimization' part from (29) to gain a simpler theory with the same empirical coverage. *Independent evidence*, especially psycholinguistic evidence—i.e. evidence obtained by measuring language users in action—that a speech production optimization such as this is might be going on is completely lacking, a fact pointed out by Labelle (2007).

In contrast, as mentioned, there is independent psycholinguistic evidence that filler-gap sentences are hard to parse. Admittedly, it may be easier to experimentally demonstrate parsing difficulties than it is to find evidence for speech-production optimizations but the discrepancy in terms of independent psycholinguistic evidence remains nevertheless.

Further, it does not seem that the derivation by phase analysis is adequate even in only a grammatical sense—at least not if it is ultimately intended that the RH constraints be captured under such a framework. Phase-based derivation cannot account for the Sentential Subject Constraint or the Adjunct Island Constraint. These sorts of extraction would presumably still be barred by the nature of subject and adjunct positions as they were in previous theories, a fact which would make incorrect predictions in the cases of an example like (9).

The Complex Noun Phrase Constraint cannot be attributed to DP (the analog of what is called NP here) being a phase, out of which extraction cannot occur, or else extraction would not be possible out of 'pictures of' contexts like those in (21). To attribute the CNPC to an interaction between a DP and a contained CP (the analog of what is called S here) would just be a return to the theory employing Subjacency in Chomsky 1986, in which case the concept of a phonologically-helpful phase would be in jeopardy.

Thus, it seems that the notion of locality for the benefit of phonological optimization has little if any meaningful evidence in its favor, whereas the notion of locality constraints for the benefit does have independent motivation.

4. Conclusion

The Sentential Recursion Constraint comprises two simple principles which derive a more accurate version of what we have called the RH constraints—the Sentential Subject Constraint, the Complex Noun Phrase Constraint and the Adjunct Island Constraint. The generality of these two principles is demonstrated by the fact that the SRC predicts other behavior beyond that predicted by the RH constraints alone.

The SRC can be motivated on the grounds that it ensures the feasibility of the parsing of a filler-gap sentence. In contrast with Chomsky's (2001) conjecture that locality constraints are ultimately due to an optimization during speech production, which has no independent evidence, the need for constraints on filler-gap sentences for the benefit of the parser has independent, psycholinguistic justification.

References

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar*, (pp. 184–221). New York: Holt, Rinehart, and Winston.
Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle*, (pp. 232–286). Amsterdam: North-Holland.
Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
Chomsky, N. (2000). Minimalist Inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, (pp. 89–155). Cambridge: MIT Press.
Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language*, (pp. 1–52). Cambridge: MIT Press.
Cullicover, P. W., & Wilkins, W. K. (1984). *Locality in linguistic theory*. Orlando: Academic Press.
Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 4, 155–209.
Hawkins, J. A. (1994). *Cambridge studies in linguistics*, volume 73. Cambridge: Cambridge University Press.
Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, 75, 244–285.
Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Huang, C.-T. J. (1982). Logical relations in chinese and the theory of grammar. Doctoral Dissertation, MIT.
Kleunder, R. (1992). Deriving island constraints from principles of predication. In *Island constraints: Theory, acquisition and processing*, (pp. 223–258). Dordrecht: Kluwer.
Labelle, M. (2007). Biolinguistics, the minimalist program, and psycholinguistic reality. Snippets Online Journal. URL http://www.ledonline.it/snippets/allegati/snippets14002.pdf.
Newmeyer, F. (2005). *Possible and probable languages*. Oxford: Oxford University Press.
Rizzi, L. (1990). *Relativized minimality*. Cambridge: MIT Press.
Stowe, L. A. (1986). Parsing wh-questions: Evidence for on-line gap location. *Language and Cognitive Processes*, 34, 319–340.
Uriagereka, J. (1999). Multiple spell-out. In S. D. Epstein & N. Hornstein (Eds.), *Working minimalism*, (pp. 251–282). Cambridge: MIT Press.
Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle,
J. Bresnan, & G. A. Miller (Eds.), *Linguisitic theory and psychological reality*, (pp. 119–161). Cambridge: MIT Press.