

## Processing Complexity in Subjacency Violations: The Complex Noun Phrase Constraint\*

Ivan A. Sag, Philip Hofmeister and Neal Snider  
Stanford University

We may make an intuitive judgment that some linguistic expression is odd or deviant. But we cannot in general know, pretheoretically, whether this deviance is a matter of syntax, semantics, pragmatics, belief, memory limitations, style, etc. [Chomsky (1977:4)]

### 1 Introduction

At least since the pioneering work of Chomsky and Miller (1963), it has been generally agreed within discussions of grammatical theory that processing has a role to play in explaining why certain linguistic examples are unacceptable. That is, certain examples that conform to the rules of the competence grammar, e.g. (1), are nevertheless judged to be unacceptable because they are difficult to process.

(1) [That [that [that Kim left] bothers Sandy] upsets them] is a problem.

Indeed, a processing (or ‘performance-based’) account of the deviance of such examples has much to recommend it. It streamlines the competence grammar, allowing it to contain simple recursive rules that ‘overgenerate’. At the same time, the appeal to processing difficulty as the explanation for the deviance of directly self-embedded structures has the virtue of explaining a gradient phenomenon in terms of a gradient mechanism. That is, a double self-embedding (as in (1)) is worse than a single self-embedding, as in (2):

(2) [That [that Kim left] bothers Sandy] upsets them.

The processing account makes sense of this observation, because processing-based explanations are inherently gradient: if a given on-line cognitive task is difficult, performing multiple such tasks is even harder – processing difficulty is cumulative. In addition, processing difficulty can be eased by the manipulation of interacting factors. Similarity and finiteness are just two of the factors that influence processing difficulty. Hence, modifying the embedded clause of (2) to make it non-finite, and hence less similar to its embedding context, as in (3), reduces processing difficulty, enhancing acceptability as a result:

(3) [That [for Kim to leave now] would bother Sandy] upsets them.

---

\*For useful discussion, we thank Ted Gibson, Perry Rosenstein, Tom Wasow and audiences at the 2007 LSA Meeting in Anaheim and at CLS. We also gratefully acknowledge research support from Stanford University.

Processing-based accounts, stated in terms of the continuous measure of processing difficulty, thus have the potential to provide more satisfying accounts of graded phenomena. The eventual goal of course is to actually ground the explanation in terms of real-time cognitive mechanisms and the competence grammar that they consult.

With the success of processing-based explanations introduced by Chomsky, Miller, and others, it is perhaps surprising that throughout the early history of generative grammar, it was assumed that island constraints should be explained in terms of constraints on the competence grammar, indeed *universal* constraints on the application of transformational rules. This assumption is evident in the original island proposal (the ‘A-over-A Condition’) of Chomsky (1962), in such well-worn studies as Ross 1967, Chomsky 1977, and Postal 1998, and in present-day transformationalist proposals. Yet, throughout the tradition of island studies, researchers have observed (often in footnotes) variations and uncertainty of judgment – notably examples that simply sound much better than they should, given the structure-based analyses that have been under discussion for the last half century.<sup>1</sup> Although it is now generally recognized that much of the critical island-related data is graded in nature, and that an empirically adequate competence grammar of subjacency effects must recognize degrees of violation,<sup>2</sup> there have been few attempts to explain island phenomena in terms of processing. (But see Kluender 1991, Kluender and Kutas 1993b, and Kluender 1998.)

This paper is part of an ongoing research program that seeks to recalibrate the data space that forms the basis of theorizing about island phenomena. As we will show, the literature has underappreciated the effect of various factors that systematically cause increased processing difficulty and hence degraded acceptability. Once such factors are carefully controlled for, we find that certain subjacency effects (those involving nominal complement clauses) are attenuated, or else disappear altogether.<sup>3</sup> The conclusion that some island effects can effectively be reduced to the aggregate processing difficulty posed by multiple sources does not mean that *all* syntactic island phenomena owe their status to principles of processing. Nevertheless, controlling for the factors known to influence acceptability will invariably lead to a clearer picture of what, if any, constraints on filler-gap dependencies should be included in a competence grammar. Here, we present and discuss the results of our reading time experiments and controlled acceptability studies on complex noun phrase violations. These findings support a processing-based account of subjacency effects, rather than a grammar-based one.

## 2 Processing Factors

The standard view is so firmly ingrained that some syntacticians may find it hard to imagine that island effects are to be explained by anything other than grammatical constraints. This is certainly true for familiar subjacency effects of the sort

---

<sup>1</sup>For some discussion, see Kluender 1991.

<sup>2</sup>For example, Chomsky’s (1986) analysis is based on the notion of ‘n-subjacency’, which is defined as follows:  $\beta$  is **n-subjacent** to  $\alpha$  iff there are fewer than n+1 barriers for  $\beta$  that exclude  $\alpha$ .

<sup>3</sup>In fact, the same is true for relative clause instances of the CNPC, though the extra filler-gap dependency introduces further processing difficulty.

illustrated in (4):

- (4) a. \*What did President Bush hear [the rumor [that the EU invaded \_ ]]?
- b. \*This was a puzzle that we met [the man [who solved \_ ]].
- c. \*What did Bill Clinton wonder [whether Hilary would like \_ ]?

Yet, as we hope to show, once the various interacting, nongrammatical factors are brought to light, the role of purely structural grammatical constraints is greatly diminished, or in the present case, rendered unnecessary.

A wide array of experiments and experimental methodologies in psycholinguistics has confirmed the difficulty of processing long-distance dependencies. Inside long-distance dependencies, reading times and response times to various tasks increase. (Wanner & Maratsos 1978, King & Just 1991, Kluender & Kutas 1993a, 1993b, Hawkins 1999). Island constructions, however, compound this baseline difficulty of filler-gap processing by imposing numerous other additional obstacles for efficient and easy processing.

For example, many island constructions require processing numerous referential entities and relatively long-term storage of syntactic predictions for incomplete propositions, while simultaneously retaining the filler-phrase in memory and searching for the correct gap site, which has no phonological manifestation. In this way, syntactic islands are akin to the center-embedding examples discussed by Chomsky and Miller (1963), which represent the paradigmatic case of performance difficulties.

According to storage-based theories of processing difficulty (e.g. Chen *et al.* 2005), increasing the cognitive resources applied to storage costs decreases the resources available for other sentence processing tasks. Hence, in examples like (5), at the most deeply embedded subject, syntactic predictions for three upcoming verbal heads as well as on object NP must be stored:

- (5) The reporter the senator Mary met attacked questioned the president.

This storage requirement presumably exceeds the capacity of the language processor. At clausal boundaries, therefore, expectations for new constituents get ‘added to the stack’, while previously expected constituents must be ‘pushed down on the queue’, so to speak. On this view, island phenomena can be traced back to the “necessity to maintain a filler in working memory while encountering an additional referential processing burden at the embedded clause boundary” (Kluender 1998).

There are a number of distinct factors that can contribute to processing difficulty in filler-gap constructions. We list some of these here, with only brief discussion:

**Distance.** In general, the longer the distance between filler and gap, the harder the retrieval and integration is at the relevant gap site. Thus, even contrasts like the following are reliably distinguished in experimental settings:

- (6) a. The reporter who \_ sent the photographer to the editor hoped for a good story.
- b. The reporter who the photographer sent \_ to the editor hoped for a good story.

The longer distance between the filler and gap in (6b) results in increased processing time, slower response times to various types of on-line tasks, etc. Researchers have proposed different ways of measuring the distance relevant for explaining such contrasts (Gibson 1998, 2000; Vasishth & Lewis 2006), but the basic effect is clear.

**Semantic Complexity.** Less salient, more complex NPs intervening between a filler and its gap increase processing difficulty. (Warren and Gibson 2002, 2005). This factor is also evident in center-embeddings that differ with respect to the complexity of the embedded subject NPs. (7a) is easier to process (and hence more acceptable ( $\geq$ )) than (7b):

- (7) a. The girl [someone [I knew] brought] left.  $\geq$   
b. The girl [the boy [the host knew] brought] left.

First and second person pronouns are the easiest NPs to process and among non-pronominal NPs, those with less content (e.g. *someone*, *anything*) are easiest.<sup>4</sup>

**Informativity.** Less informative fillers (e.g. *who* vs. *which man*) are harder to retrieve at the gap site (Hofmeister et al. 2007; Hofmeister 2007). This result holds also for NPs that are not always regarded as fillers, e.g. a clefted element or the nominal phrase heading a relative clause. Thus, the more informative indefinite NP in (8a) leads to faster retrieval at the gap site than its less informative counterpart in (8b):

- (8) a. It was [an influential communist-leaning dictator] that Sandy said she liked \_\_.  $\geq$   
b. It was [a dictator] that Sandy said she liked \_\_.

**Frequency.** Less frequent lexical items contribute to processing difficulty (Hale 2001; Jurafsky 2003). Hence (9a) is considerably easier to process than (9b):

- (9) a. Which letter did the judge decide to send back immediately?  $\geq$   
b. Which epistle did the magistrate opt to remand forthwith?

**Similarity.** As noted in the introduction, similarity-based interference can increase processing difficulty (Gordon et al. 2004). Thus, embedding a definite NP within a definite NP, as in (10b) creates a processing burden that (10a) lacks:

- (10) a. The doctor that some nurse with a limp consoled treated one of the patients.  $\geq$   
b. The doctor that the nurse with the limp consoled treated the patient.

**Finiteness.** Finiteness is another factor which influences processing (Ross 1967; Kluender 1992), as we have already noted. Eventive finite clauses are the hardest to process; infinitival clauses the easiest, with modal finites being intermediate:

- (11) a. What were you unsure how to file?  $\geq$

---

<sup>4</sup>Reducing the semantic complexity or “richness” of linguistic elements in general (not just NPs) intervening between filler and gap has been identified by numerous authors as a means for improving extraction (see Deane 1991 for a discussion).

- b. What were you unsure how you should file? ≥
- c. What were you unsure how they had filed?

Gibson (2000) speculates that the increased processing burden of finite verbs stems from their eventive nature. That is, the cost of introducing a reference to an event is similar to that of an NP that evokes reference to a discourse entity.

**Contextualization Difficulty:** All examples bring with them presuppositions, some of which are easier to contextualize than others. For example, as noted by Kroch (1989), the reduced acceptability of an example like (2a) is better explained simply by noting the difficulty of accommodating its presupposition in (2b):

- (12) a. How much money was John wondering whether to pay?
- b. There was a sum of money about which John was wondering whether to pay it.

This is a further source of processing difficulty that is rampant and uncontrolled in the generative literature. Note that a main-clause interrogative potentially introduces a difficulty of this kind, as one must imagine a context where the question's presuppositions are satisfied. This difficulty can generally be eliminated by constructing an analogous example where the interrogative clause is embedded, leaving the example as a whole free of complex presuppositions.

Further factors that introduce processing complexity may include collocational frequency, i.e. the frequency of the lexical material in the filler phrase appearing with the subcategorizing element:

- (13) a. Which **country** do you think it's likely that they **visited** last month? ≥
- b. Which **country** do you think it's likely that they **criticized** last month?

In addition to basic considerations of 'plausibility', the explanation of such contrasts may be grounded in the notions of 'surprisal' or 'expectation' (Hale 2001; Levy 2005).

Just as eliminating a clausal embedding in the center-embedding structures facilitates processing, a variety of factors in island constructions can be manipulated so as to maximize overall processing ease. Under such conditions, acceptability may rise to levels comparable to the acceptability of minimally different non-islands. For example, reading-time evidence suggests that the usage of more informative filler-phrases facilitates the processing and acceptability of long-distance dependencies into syntactic islands.

So the question we have set out to investigate is whether processing difficulty can explain island effects without competence grammar constraints. Controlling for some of the factors discussed in this section leads us to consider the following examples:

- (14) a. I wonder which Mid-East country they heard rumors that we had invaded \_\_ .
- b. This was the only crossword puzzle that we've ever found anyone who could solve \_\_ .

- c. It was that Ibsen play that they wondered whether we would like \_\_ .

We hope it is clear that these examples are more acceptable (and easier to process) than their counterparts that we used to begin this section, namely those in (4) above. Yet there is no reason to think that the structure of these examples is different. Once one appreciates this fact, a processing-based account of whatever difference in acceptability remains between (4a-c) and (14a-c) appears more plausible.

### 3 Complex Noun Phrases: Self-Paced Reading

Ross (1967) introduces the class complex noun phrase constraint (CNPC), exemplified previously in the examples above:<sup>5</sup>

(15) *The Complex NP Constraint*

No element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation.

In the same text, however, Ross himself notes certain exceptions to the complex noun phrase constraint.

- (16) a. You are making the claim that the company squandered a large amount of money.  
b. How much money are you making the claim that the company squandered?

Subsequently, other researchers pointed to further apparent counterexamples to this constraint (as well as the formulations that came to replace Ross's original version, such as Subjacency (Chomsky 1973) and Barriers (Chomsky 1986)). Chung & McCloskey (1982), for instance, observe that the nature of the extracted element affects the overall acceptability of the construction. For example, extraction of an indefinite like *a paper* is preferred over extraction of a *wh*-phrase:

- (17) a. This is a paper that you really need to find someone you can intimidate with.  $\geq$   
b. *Which paper* do you really need to find someone you can intimidate with?  $\geq$   
c. *How many papers* do you really need to find someone you can intimidate with?  $\geq$   
d. *What* do you really need to find someone you can intimidate with?

Of course, a strictly syntactic formulation of the CNPC would fail to predict any difference in these examples.

On top of these exceptions to the constraint in English, other languages exhibit an abundance of counterexamples to this supposedly universal constraint. It is well-known, for instance, that Swedish allows CNPC violations like (18) (Allwood 1976):

---

<sup>5</sup>Ross attributes the constraint to observations from Edward Klima.

- (18) Den där gamla skräphögen kanner ja killen som köpte?  
 That old piece junk know I the guy who bought.  
 ‘That old piece of junk, I know the guy who bought (it).’

In addition to the putative counterexamples, the acceptability of an extraction out of a complex noun phrase appears intuitively to vary with the type of the island-forming NP. Kluender (1992) orders the acceptability of CNPC violations in the following way from best to worst:

- (19) a. This is the paper that we really need to find *someone* who understands.  
 $\geq$   
 b. This is the paper that we really need to find *a linguist* who understands.  
 $\geq$   
 c. This is the paper that we really need to find *the linguist* who understands.  $\geq$   
 d. This is the paper that we really need to find *his advisor*, who understands.  $\geq$   
 e. This is the paper that we really need to find *John*, who understands.

As Kluender notes, "a referentially specific head noun in an embedded complex NP necessitates the mental identification of an extra referent in addition to the logical subject of the entire predication" (1992: 239). Essentially, this harks back to the factor of semantic complexity mentioned in the previous section. More complex constituents intervening between a filler and its gap consume processing resources that could otherwise be applied to processing the filler-gap dependency.

In this section, we examine how systematically manipulating these two non-syntactic factors – the content of the filler and the determiner type of the island-forming NP – impacts the processing of CNPC violations. To quantify processing difficulty, we employ the self-paced reading methodology, where participants read sentences at their own pace. The results of this comprehension experiment can then be compared to the findings of an acceptability study which we discuss subsequently.

### 3.1 Materials

The stimuli in this experiment consisted of 36 embedded CNPC violations, which varied with respect to properties of the filler-phrase and the island-forming NP type. For each item, each participant saw one of seven experimental conditions (2 x 3 + 1). Across conditions, the extracted *wh*-phrase was either a bare *wh*-item (= BARE condition), such as *who* or *what*, or a comparatively more informative *which-N*' phrase (= WHICH condition). Precisely half of the items presented animate *wh*-phrases, while the other half contained inanimate *wh*-phrases. The other factor considered in this experiment was the effect of NP type on subsequent sentence processing. Subjects read one of three kinds of island-forming NPs: a definite NP (DEF), an indefinite plural (PL), or an indefinite singular (INDEF). Additionally, a baseline for each item was included that lacked the island-forming NP. (20) shows a sample experimental item with all seven conditions:

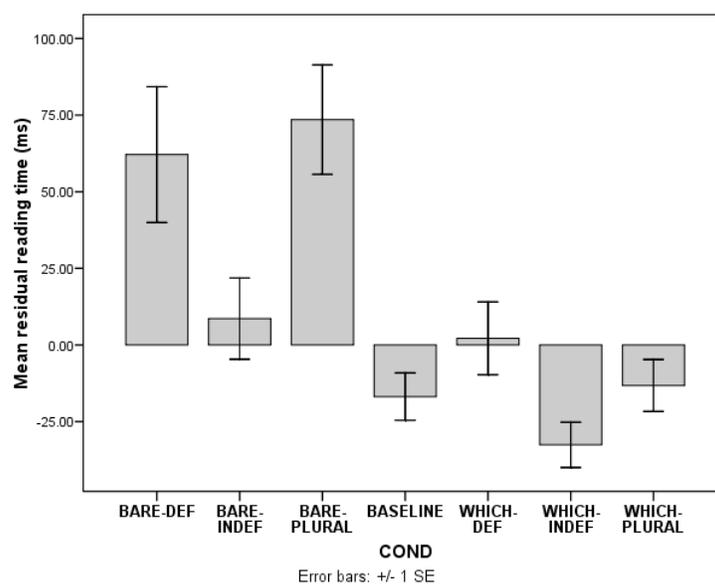
- (20) BARE-DEF: I saw **who** Emma doubted the report that we had captured in the nationwide FBI manhunt.  
BARE-PL: I saw **who** Emma doubted reports that we had captured in the nationwide FBI manhunt.  
BARE-INDEF: I saw **who** Emma doubted a report that we had captured in the nationwide FBI manhunt.  
WHICH-DEF: I saw **which convict** Emma doubted the report that we had captured in the nationwide FBI manhunt.  
WHICH-PL: I saw **which convict** Emma doubted reports that we had captured in the nationwide FBI manhunt.  
WHICH-INDEF: I saw **which convict** Emma doubted a report that we had captured in the nationwide FBI manhunt.  
BASELINE: I saw **which convict** Emma doubted that we had captured in the nationwide FBI manhunt.

All reading-time results were analyzed in a 3 x 2 repeated measures ANOVA design. Post-hoc comparisons were also made to the baseline as necessary. Due to poor question-answer accuracy in some conditions of some items (7 out of 252 cells, or 2.77% of the item analysis data set), the missing data for those cells were replaced with the linear trend for that point using the *Replace Missing Values* command in SPSS 15.0. This method of data imputation inserts the predicted values of empty data cells by regressing the existing series on an index variable scaled 1 to  $n$ . For this experiment, outliers were removed via data trimming – in this case, eliminating the five most extreme reading times from each condition – which affected 4.9% of the total data, after responses corresponding to incorrectly answered stimuli were removed.

We exclusively report on residual reading times, which take into account the effect of word length on reading. This effectively reduces variability due to individual differences in reading times. For each experimental item, a yes/no comprehension question checked that subjects followed the text and read for understanding. The results presented here consist only of reading times from items which were correctly answered.

### 3.2 Participants

Twenty-five Stanford University undergraduates, all native speakers of English, were paid \$15 to complete this experiment along with a short off-line survey that was unrelated to this experiment. Combined, the two experiments lasted between twenty and thirty-five minutes. For the entire experiment, global average reading time per word was 398.9 milliseconds ( $SD = 83.2$ ). Average question-answer accuracy for all items, including fillers, equaled 79.45%. There were no main effects of condition in the accuracy scores (BARE-DEF: 81.10%,  $SE = 3.49$ ; BARE-INDEF: 80.47%,  $SE = 3.52$ ; BARE-PL: 75.00%,  $SE = 3.84$ ; WHICH-DEF: 80.77%,  $SE = 3.47$ ; WHICH-INDEF: 75.20%,  $SE = 3.82$ ; WHICH-PL: 83.72%,  $SE = 3.26$ ; BASELINE: 76.74%,  $SE = 3.73$ ).



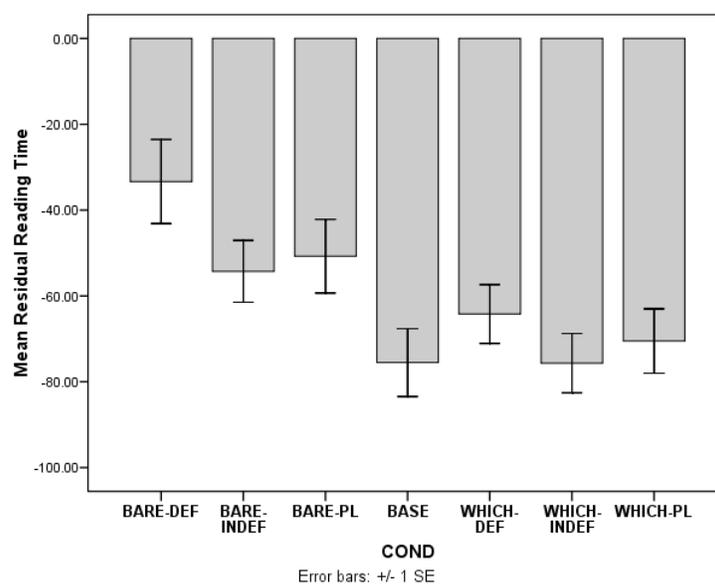
**Figure 1:** Residual reading times at complementizer

### 3.3 Results

At the word immediately after the *wh*-phrase, informativity of the *wh*-phrase does not significantly impact reading times (NP type is, of course, irrelevant as the NPs have not been reached yet). Yet, on the second word after the *wh*-phrase – the first verb encountered after the filler – there is a main effect for *wh*-phrase informativity: the BARE conditions (which are *less informative* than the WHICH conditions) lead to significantly faster reading times early on in the sentence processing ( $F(1,24) = 8.335, p < .01$ ;  $F(1,35) = 6.152, p < .05$ ).

Beginning at the complementizer *that*, in contrast, the WHICH conditions are read faster than the BARE conditions ( $F(1,24) = 13.776, p = .001$ ;  $F(1,35) = 18.953, p < .001$ ). In all three NP-type conditions, the WHICH version is processed faster than the corresponding BARE version, as shown in Figure 1.

This processing advantage for the WHICH conditions extends beyond the complementizer – it covers the subsequent pronominal subject, the embedded auxiliary and verb (see Figure 2), as well as the regions after the verb where the presence of the gap is confirmed. Looking at the embedded pronominal subject, for instance, the statistical analysis reveals a main effect for the informativity of the *wh*-phrase ( $F(1,24) = 5.028, p < .05$ ;  $F(1,35) = 11.972, p = .001$ ). Likewise, at the auxiliary, reading times are fastest in the WHICH conditions ( $F(1,24) = 16.283, p < .001$ ;  $F(1,35) = 8.245, p < .01$ ). The WHICH conditions also produce faster reading times at the subcategorizing verb ( $F(1,24) = 7.013, p < .05$ ;  $F(1,35) = 5.249, p < .05$ ) and at the subsequent word, where readers receive overt evidence of the missing constituent. Also of importance is the fact that, throughout the most embedded clause, the average reading time for the WHICH conditions does not significantly differ from that of the baseline. This stands in stark contrast to the BARE condition, which remains consistently slower than the baseline.



**Figure 2:** Residual reading times at auxiliary + verb

At only two word regions do the results indicate an effect of the type of island-forming noun phrase. First, at the complementizer, there is a significant effect of NP type. A closer inspection of the data reveals that the effect at this particular region largely stems from drastically slowed reading times in the BARE-DEF and BARE-PL conditions, as shown in Figure 1. The slower reading times for these two conditions are responsible for the main effect of NP type at the complementizer ( $F(1,19) = 9.415, p = .001$ ;  $F(1,34) = 11.619, p < .001$ ). This effect recurs at the first word after the verb, e.g. the word *in* in (20). In this case, the reduced reading times associated with the definite NPs are responsible for the effect, as the BARE-DEF and WHICH-DEF conditions produce the slowest reading times of any condition.

In sum, the reading results identify an initial, but brief, slowdown after processing the more informative *wh*-phrases, but beginning with the complementizer, this disadvantage reverses and becomes a highly significant processing advantage. This advantage lasts until several words after the subcategorizing verb that the *wh*-phrase must be associated with. Effects due to the type of the island-forming NP are much more localized (and relatively weaker), appearing only at the complementizer and the first word after the verb.

### 3.4 Discussion

Reading times in this study show a strong influence of the informational content of the extracted element. In cases where the informational content of the extracted *wh*-phrase was high, reading times were comparable to that of the baseline condition that contained no syntactic islands. Following a brief slowdown after the *which-N'* phrase (which is undoubtedly due to the cost of building a more complex representation), processing becomes markedly faster in the WHICH conditions. This advantage begins at the complementizer and extends beyond the subcategoriz-

ing verb.

It is worth considering why the advantage for the WHICH conditions begins at the complementizer. One possibility is that *which-N'* phrases reduce the chances of garden-pathing the filler-gap dependency. Along the filler-gap path in these complex noun phrase constructions, another verb appears whose complement is the complex noun phrase containing the gap site. In some of the stimuli, such as the example appearing in (21), the bare *wh*-item could easily be interpreted as a nominal argument of the intervening verb *announced*. By contrast, a lexically richer *which-N'* phrase like *which structure* is unlikely to be interpreted as a nominal argument of the verb, as propositions and facts get announced, but concrete objects do not.

- (21) a. He published which structure Brittany announced plans that we would build to replace the condemned building.  
b. He published what Brittany announced plans that we would build to replace the condemned building.

In the case of the bare *wh*-phrase, however, the filler can be associated with either a propositional- or an individual-denoting gap. This temporary ambiguity can lead to an early misparsing in the BARE conditions, where subjects initially assume the bare *wh*-item to be an argument of the first verb encountered (*announced* in (21a,b)). Upon receiving counterevidence to this initial analysis, a reanalysis must take place – one which consumes extra processing resources. The complementizer is the first word after the head noun, which is the first piece of evidence confirming that a complex NP, rather than the *wh*-phrase, is the argument of the first verb. As soon as the head noun is reached, therefore, the process of reanalysis begins and the ensuing processing difficulty spills over onto subsequent constituents, including the complementizer. This resource expenditure may in fact incur a cost over a relatively long period of time. Such a scenario would thus explain the finding that *wh*-phrases of varying levels of informativity result in processing differences beginning at the complementizer.<sup>6</sup>

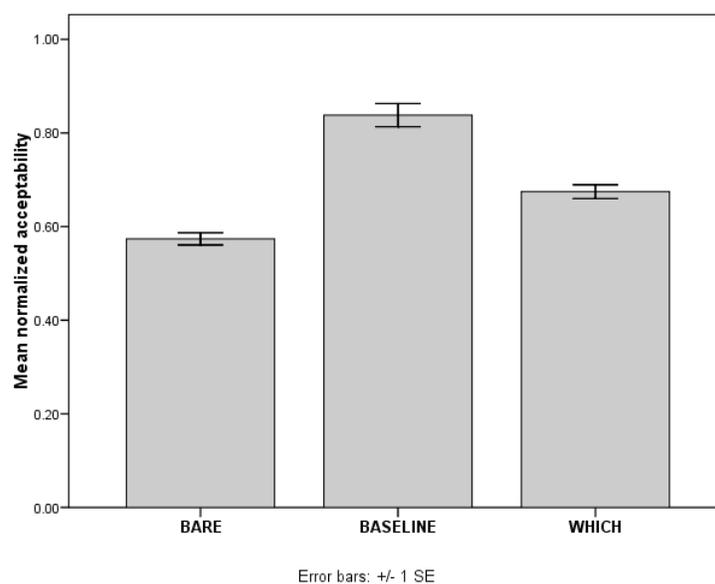
In contrast to the filler effects, which start early and remain significant throughout nearly the entire filler-gap dependency, manipulating the type of the island-forming NP generates only small and temporary effects during the processing of the filler-gap dependency. Because of the sporadic and temporary nature of these effects, further experimentation is required in order to understand the effect of NP type on the processing of the filler-gap dependencies. What can be said, however, is that NP type plays less of a role in processing than filler-informativity, at least in extractions out of complex NPs.

#### 4 Complex Noun Phrases: Acceptability

To complement the self-paced reading experiment, we also conducted a controlled acceptability study using the exact same stimuli, including the same fillers. Sixteen subjects (none of whom had participated in the previous reading time study) were

---

<sup>6</sup>Other factors, however, may also be contributing to the informativity effects observed in this experiment. Differences in retrieval from memory across conditions, for instance, may be responsible for the continued informativity effects seen at the verb and in the subsequent spillover regions (see Hofmeister 2007).

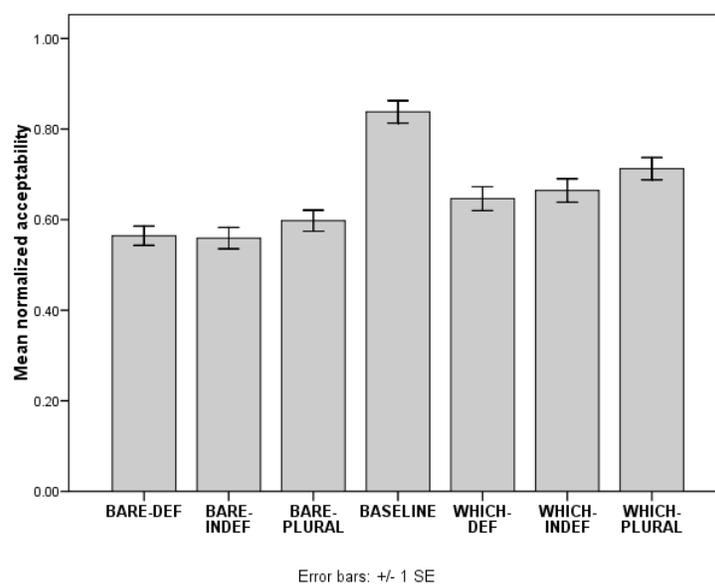


**Figure 3:** Mean normalized acceptability ratings of complex noun phrase constructions

asked to rate the sentences for naturalness on a scale of 1 to 7. They were specifically instructed not to rate the sentences according to prescriptive grammar rules. The subjects in this study were given course credit for their participation.

In concert with the results of the comprehension study, the acceptability ratings reflect a preference for constructions with extracted *which-N'* phrases over bare *wh*-words, as depicted in Figure 3. As expected, the baseline received significantly higher ratings than either of the syntactic island conditions. For the most part, these acceptability results parallel the findings of the self-paced reading study. One exception is that, in the acceptability study, the baseline and WHICH conditions are significantly different – the baseline received a higher mean rating of acceptability. In the reading time study, however, the baseline did not consistently produce significantly faster reading times inside the critical embedded clause. The second difference between the results of the two studies concerns the effect of NP type: in the acceptability study, plural NPs raise the acceptability of extraction out of the complex noun phrase to a marginal degree, as shown in Figure 4. In contrast, plural indefinites did not lead to faster processing in the reading time study. What differences of this kind show is that acceptability judgments can be sensitive to factors whose effect does not surface in more localized on-line processing measures. Reading times in particular may not reflect the effects of semantic, pragmatic, and/or discourse processing that extend beyond the first-pass reading of the sentence. Such effects usually manifest themselves not only as lower acceptability judgments, but also as higher response times in answering comprehension questions. Indeed, we find that question response times for the baseline case in datasets like (20) are about 150 milliseconds faster than those of the WHICH-conditions.<sup>7</sup>

<sup>7</sup>Thus, claiming that processing difficulty influences acceptability does not mean that that we



**Figure 4:** Mean normalized acceptability ratings of complex noun phrase constructions

What the acceptability study shows is that systematically manipulating these two non-syntactic factors produces results parallel to the outcome of the reading-time study. Making the reasonable assumption that this correlation is not accidental, the next logical step is to infer a direction of causality for these particular results: Does the reduced grammaticality of these subjacency violations create processing difficulty or does processing difficulty create the reduced acceptability of these sentences? The fact that processing factors of the sort observed here have been independently observed to cause uncontroversially grammatical examples to be judged less acceptable strongly suggests that processing difficulty plays a causal role here as well. This in turn has the welcome effect of eliminating grammatical stipulations (the Subjacency Condition), thus simplifying our account of linguistic knowledge while maximizing the explanatory consequences of independently verifiable extra-grammatical factors.<sup>8</sup>

## 5 Conclusion

The evidence presented here supports the view that Subjacency effects can be explained in processing terms. Mitigating the factors known on independent grounds to be sources of processing difficulty causes acceptability to rise. In fact, removing

---

expect acceptability ratings to be absolutely identical to online processing measures, like reading times. Some differences are definitely expected on the assumption that acceptability judgments constitute the end-result (or involve the input) of a number of cognitive processes, including sentence processing effort. It should also be noted that the fewer number of words contained in the baseline condition is another likely reason for the higher mean acceptability rating of the baseline sentences (e.g. the last one shown in (20)).

<sup>8</sup>For a critical discussion of attempts to explain filler informativity differences in grammar-internal terms (D-Linking, referentiality, etc.), see Hofmeister 2007.

the difficulty posed by just a single factor – the informativity of the filler phrase – is sufficient to reduce reading times to baseline levels and judgments of acceptability to near-baseline levels (cf. the factors ameliorating self-embedded structures).

It should be stressed once again that, despite the evidence that processing plays a large role in the acceptability of CNPC violations, we are not suggesting that all syntactic island phenomena can or should be explained via principles of processing. Although additional experimental research (Hofmeister *et al.* 2007, Hofmeister 2007) strongly suggests that processing difficulty explains the reduced acceptability of, for example, *wh*-islands and adjunct islands, there are other island effects that may simply be the product of grammatical constraint. For instance, the Conjunct Constraint (part of Ross's Coordinate Structure Constraint) is a likely candidate for a syntactic explanation, as intuitions regarding violations of this constraint appear to be quite stable. That is, controlling non-syntactic factors like the ones discussed above do little to ameliorate these violations.

Thus, although we propose to purge from competence grammar any constraint like the Subjacency Condition or the Complex NP Constraint, the point of our research program is not to explain away all syntactic island constraints. Rather, it is to exploit the knowledge of independently motivated processing constraints in order to better assess what needs to be included in the grammar. After controlling for factors whose contribution to processing difficulty has been established, the constraints of the grammar can be seen as much more homogenous and transparent. At the end of the day, the gradient data is explained but the grammar is simpler.

## References

- Allwood, Jens. 1976. The Complex NP-Constraint as a non-universal rule. In *University of Massachusetts Occasional Reports No. 2*. Amherst, MA: University of Massachusetts, Dept of Linguistics. Also in E. Engdahl and E. Ejerhed, eds. (1982), *Readings on Unbounded Dependencies in Scandinavian Languages*, Acta Universitatis Umensis and Almqvist & Wiksell International, Stockholm and Umeå.
- Chen, Evan, Edward Gibson, & Florian Wolf. 2005. Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52.144–169.
- Chomsky, Noam. 1962. The logical basis of linguistic theory. In *Preprints of Papers for the Ninth International Congress of Linguists, August 27–31, 1962*, ed. by Morris Halle, 509–674. Cambridge, MA: Morris Halle.
- . 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. by Stephen Anderson & Paul Kiparsky, 232–286. New York: Holt, Rinehart, and Winston.
- . 1977. On *Wh*-movement. In *Formal Syntax*, ed. by Adrian Akmajian, Peter Culicover, & Thomas Wasow, 71–132. New York: Academic Press.
- . 1986. *Barriers*. Cambridge, MA: MIT Press.
- , & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In *Handbook of Mathematical Psychology*, ed. by R. Duncan Luce, Robert R. Bush, & Eugene Galanter, volume 2, 269–321. New York: Wiley.
- Deane, Paul. 1991. Limits to attention: a cognitive theory of island phenomena. *Cognitive Linguistics* 2.1–63.

- Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68.1–76.
- . 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In *Image, Language, Brain*, ed. by W. O’Neil Y. Miyashita, A. Marantz, 95–126. Cambridge: MIT Press.
- Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.
- Hawkins, John. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75.244–285.
- Hofmeister, Philip. 2007. *Facilitating Memory Retrieval in Natural Language Comprehension*. Stanford University dissertation.
- , T. Jaeger, Ivan Sag, Inbal Arnon, & Neal Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in Search of its Evidential Base*, ed. by Sam Featherston & Wolfgang Sternefeld. Berlin: Mouton de Gruyter.
- Jurafsky, Dan. 2003. Probabilistic modeling in psycholinguistics: linguistic comprehension and production. In *Probabilistic Linguistics*, ed. by Rens Bod, Jennifer Hay, & Stefanie Jannedy. Boston: MIT Press.
- King, Jonathan, & Marcel Just. 1991. Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language* 30.580–602.
- Kluender, Robert. 1991. *Cognitive Constraints on Variables in Syntax*. University of California, San Diego dissertation.
- . 1992. Deriving island constraints from principles of predication. In *Island Constraints: Theory, Acquisition and Processing*, ed. by Helen Goodluck & Michael Rochemont, 241–279. Dordrecht: Kluwer.
- . 1998. On the distinction between strong and weak islands: A processing perspective. In *The Limits of Syntax*, ed. by Peter Culicover & Louise McNally, volume 29 of *Syntax and Semantics*, 241–279. San Diego: Academic Press.
- , & Marta Kutas. 1993a. Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5.196–214.
- , & Marta Kutas. 1993b. Subjacency as processing phenomenon. *Language and Cognitive Processes* 8.
- Kroch, Anthony. 1989. Amount quantification, referentiality, and long *wh*-movement. Manuscript, University of Pennsylvania; appears in Dimitriadis et al. (eds.), 1998, *Penn Working Papers in Linguistics*.
- Levy, Roger. 2005. *Probabilistic Models of Word Order and Syntactic Discontinuity*. Stanford University dissertation.
- Postal, Paul M. 1998. *Three Investigations of Extraction*. Cambridge, MA: MIT Press.
- Ross, John R., 1967. *Constraints on Variables in Syntax*. MIT dissertation. [Published in 1986 as *Infinite Syntax!* Norwood, N. J.: Ablex].
- Vasishth, S., & R.L. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language* 82.767–794.
- Wanner, Eric, & Michael Maratsos. 1978. An ATN approach to comprehension. In *Linguistic Theory and Psychological Reality*, ed. by Morris Halle, Joan Bresnan, & George A. Miller, 119–161. Cambridge, MA: MIT Press.
- Warren, Tessa, & Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85.79–112.
- , & ———. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20.751–767.