+Q/-Q and the Ambiguity Hypothesis of *wh*-Islands

Kyle Grove
Cornell University

1. Introduction

This paper reexamines the Ross (1967) claim that while interrogative (+Q) CPs are islands for *wh*-extraction, as shown in 1a and 1b, declarative (-Q) CPs are not, as shown in 2a and 2b. Alexopoulou and Keller (2007) report that English-speaking subjects find extractions in the Island condition to have degraded acceptability as compared to the Non-island condition, in non-embedded and embedded contexts, as seen in 2.

(1) a. ? Who did Mary wonder whether we will fire?
    b. ? Who did Jane think that Mary wonder whether we will fire?

(2) a. Who did Mary claim that we will fire?
    b. Who did Jane think that Mary claim that we will fire?

Bresnan (1970) noted two natural classes of sentential-embedding verbs: verbs like *claim*, which subcategorize for -Q complements (*that* or $\emptyset$); and verbs like *wonder*, which subcategorize for a wide array of +Q continuations (*whether*, *if*, and embedded questions), as shown in Example 3.

(3) a. * Mary wondered that the supervisor will fire the employee.
    b. Mary wondered whether the supervisor will fire the employee.
    c. Mary wondered if the supervisor will fire the employee.
    d. Mary wondered who the supervisor will fire.
    e. Mary wondered who will fire the employee.
    f. Mary wondered when the supervisor will fire the employee.

Hypothesis: Extraction from embedded interrogatives is difficult because +Q Ambiguity (Bresnan, 1970) compounds the filler-gap ambiguity.

We argue that the Ambiguity Hypothesis accounts for the finding of Alexopoulou and Keller (2007) that *wh*-extraction from +Q produces degraded Magntiude Estimates when compared to *wh*-extraction from -Q. The account is also congruent with several other findings in the literature, which we report in the Discussion section.

2. Test

The Ambiguity Hypothesis suggests that Islands exhibit greater incremental ambiguity than Nonislands due to the +Q Ambiguity operant in Islands. To formalize the respective roles of competence and performance in this prediction, we adopt two linking hypotheses. On the competence side, we implement Bresnan (1970) using the Derivational Minimalist Grammars formalism (Stabler, 1997). On the performance side, we adopt the Entropy Reduction (Hale, 2006) complexity metric. The ERH operationalizes the linguist's notion of ambiguity by modeling the intrinsic sentence processing work required to parse a word as the amount of information uncertainty reduced by the word. The comprehender's analytical state is modeled as a weighted intersection of the competence grammar and the unfolding sentence. This weighted intersection is itself a probabilistic grammar whose weights

*This material is based upon work supported by the National Science Foundation under Grant No. 0741666.

correspond to degrees of belief attached to competing parses of the unfolding sentence; information-processing work has occurred when weight accrues toward certain analyses.

=C S	**claim** =CC V-Q	=>V v	**Mary** D
=T C	**claim** V-Q	**that** =T CC	**who** D -wh
=Q C	**wonder** =P V+Q	**whether** =T CQ	**about** =D P
will =v-Q =D T	**wonder** =CQ V+Q	**if** =T CQ	**with** =D P
=v-Q =D T	**wonder** =Q V+Q	=T +wh Q	**why** P -wh
will =v+Q =D T	**wonder** =CC V+Q	**do** =T +wh Q	**how** P -wh
=v+Q =D T	=>V+Q v+Q	**must** =T +wh Q	**when** P -wh
V-Q =CC V-Q	=>V-Q v-Q	**fire** =D V	**where** P -wh
V-Q V-Q	=>V-Q v-Q	**fire** V	T << P

Figure 1: Sample Unweighted Minimalist Grammar

2.1. Minimalist Grammars

The Minimalist Grammars framework of Stabler (1997) formalizes transformational grammars written by Chomskyan linguists for use in psycholinguistic complexity metrics such as the Entropy Reduction Hypothesis. In Minimalist Grammars, each lexical item has a distinct set of syntactic (SYN) features, which uniquely determine a movement chain that the lexical item can participate in (Hale and Stabler, 2005). Thus, while Minimalist Grammars are not context-free, they possess a context-free backbone which can be weighted and estimated as a context free grammar can.

2.1.1. A Minimalist Grammar of Bresnan (1970)

We implement a Derivational Minimalist Grammar (Stabler, 1997) which distinguishes between +Q and -Q complementizers, deriving the paradigm in Example 3. This grammar constitutes an analysis of the problem space encountered by the human parser on the sentence in Alexopoulou and Keller (2007), so the amount of work required to incrementally parse this problem space can be calculated.

As seen in 1, the grammar implements Bresnan (1970)'s central observation that there exist two natural classes of embedded complement phrases (-Q embedded declaratives, +Q embedded interrogatives) and that clause-embedding verbs can pattern differently with respect to which complements they select. For example, verbs such as 'wonder' select only -Q continuations, whereas verbs such as 'think' select +Q continuations. The grammar also implements filler-gap wh-questions structures and verb transitivity ambiguity.

2.2. Entropy

Hale (2006)'s Entropy Reduction Hypothesis suggests that the difficulty of sentence processing is predicted by positive decreases in entropy. According to the ERH, difficult sentence processing tasks involve the rapid reduction of a large, entropic parse forest to a less ambiguous parse forest. The parse forest congruent with a prefix w includes all continuations of w; at each word the human sentence processor must reject predictions which are suddenly no longer congruent with the sentence. Sudden reductions in the entropy of the prediction space triggered by a word result in greater cognitive load at that word. Hale (2006) develops a methodology for computing entropies of probabilistic intersections of context free grammars and automata for use in modeling incremental sentence processing. Entropy is often conceptualized in terms of a single random variable. The entropy of a simple random variable is equal to $-\Sigma_i p_i \log_2 p_i$. A fair coin, for example, has an entropy of $-(.5\ log\ .5) + (.5\ log\ .5)$, i.e., 1.0 bits of entropy, since each of the two possible outcomes (heads, tails) has a 0.5 probability of occurring. For the purposes of modeling linguistic difficulty, it is helpful to conceptualize three separate aspects of difficulty (ambiguity) which the entropy models. First, entropy models the fact that, all else being equal, with equiprobable outcomes are more difficult than cases where probability is skewed towards certain outcomes. Second, entropy models the fact that decisions with many (equiprobable) outcomes are more difficult than decisions with fewer (equiprobable) outcomes. Finally, decisions with outcomes dependent on other decisions are more uncertain than simple random variables. Hale (2006) combines Grenander-style entropy computation with the discovery of Billot and Lang (1989) that intersections of CFGs and automata are themselves CFGs. The probabilistic extension of Billot and Lang (1989) operationalizes the

ABOUT = [lemma = wonder & pos = VVD|Z][lemma = about]
WHETHER = [lemma = wonder & pos = VVD|Z][lemma = whether]
IF = [lemma = wonder & pos = VVD|Z][lemma = if]
Q = [lemma = wonder & pos = VVD|Z][(lemma = wh.* & (pos=W.* | pos =IN | pos = RB)) | lemma = how]
THAT = [lemma = wonder & pos = VVD|Z][lemma = that & pos = IN]

Figure 2: Corpus Queries for complements of 'wonder','claim' on NYT

notion of linguistic entropy that a comprehender possesses in an incremental parse of the sentence. That the derivation tree languages of mildly context sensitive grammars are context-free (Hale and Stabler, 2005) allows the extension of this methods to Stabler (1997)'s Minimalist Grammars, as demonstrated in Hale (2006). We refer the reader to Hale (2006) for further detail about the computation of entropies of probabilistic grammars.

2.3. Weighted Minimalist Grammars

The Entropy Reduction Hypothesis requires a probabilistic formal grammar which intersects a theory in the form of a grammar with the problem space of an experiment. We implement this problem space via a weighted corpus which we parse up into a Minimalist mini-treebank.

We built a weighted training corpus which reflects statistical subcategorization facts for the verbs in the experiment. Using the CQP (Christ, 1993) queries shown in Fig. 2, we collected counts from *the New York Times* corpus (Sandhaus, 2008) for different complements subcategorized for by *wonder* and *claim*. We also obtained counts for embedded verb token transitivity. The training corpus utilizes a factorial design, such that each sentence varies across four parameters: Matrix Verb (*claim* vs. *wonder*); Complement Type (*if*, *whether*, embedded question, *that*, *about*); Embedded Verb Transitivity (transitive, intransitive); and Adjunct/Argument extraction. All but the last parameter was estimated as described above; the Adjunct/Argument parameter was set at 0.5 for the experiment described here. Each sentence is weighted by the products of counts from each parameter, so that the weighted corpus represents an intersection of the weighted grammar and problem space of *wh*-island parsing.

3. Results

We used *mcfgcky* (Grove, 2010) to train on the above minicorpus and test on the Alexopoulou and Keller (2007) sentences. We tested whether a Minimalist Grammar which encoded the greater diversity of +Q continuations would result in greater processing effort of wh-slands using the Entropy Reduction (Hale, 2006) metric. We obtained greater total entropy reductions for the Island condition than for the Control condition, as seen in Fig. 4.

Almost all of the difference is due to the +Q ambiguity. As seen in Fig. 5, Island and Control conditions exhibit almost the same Filler-Gap ambiguity, but the Island condition exhibits the additional +Q ambiguity. This is verified through a post-hoc analysis of incremental parser state output immediately following the matrix verb, so that the probabilistic parse forest corresponding to "Who did Mary wonder/claim *" was obtained. From this parse forest, we extracted the MG category which corresponds to a matrix verb phrase with a nominal gap ($v_{wonder} - q$ for Island, $v_{wonder} - q$ for Control). This category was found to have different entropy and branch weightings depending upon the experimental condition. Island condition matrix verb phrases ($v_{wonder} - q$) have 2.556 bits of total entropy, whereas Control condition matrix verb phrases ($v_{claim} - q$) have 0.011 bits of total entropy. However, this difference in entropy is not just due to the encoding of complementizers in the grammar, nor the simple statistical distribution of sentences, but is rather due to multiple dimensions of the +Q ambiguity. The Island condition spreads probability across three different branches, as seen in Fig. 6, and the one-step rewriting entropy of these branches is 1.290 bits. The particular branch which corresponds to embedded *wh*-questions is encoded with Q -q. This category itself has 2.328 bits of entropy, contributing 1.222 bits of entropy to its parent. Our interpretation is of this fact is that the main result is driven by such hierarchical uncertainties stemming from syntactic complexity.

+Q wonder	sentence frame	-Q claim
0.00720	”who does Jane wonder/claim about”	0.00047
0.01240	”who does Jane wonder/claim if Mary will punish”	3.1965e-5
0.00150	”who does Jane wonder/claim if Mary will punish with”	3.8877e-6
0.01240	”who does Jane wonder/claim if Mary will punish Mary with”	3.1966e-5
0.01334	”who does Jane wonder/claim who will punish”	8.47745e-5
0.01334	”who does Jane wonder/claim who will punish with”	8.47745e-5
0.00314	”who does Jane wonder/claim when Jane will punish”	1.99785e-5
0.00038	”who does Jane wonder/claim when Jane will punish with”	2.4298e-6
0.00314	”who does Jane wonder/claim when Jane will punish Jane with”	1.99785e-5
0.00314	”who does Jane wonder/claim where Jane will punish”	1.99785e-5
0.00038	”who does Jane wonder/claim where Jane will punish with”	2.4298e-6
0.00314	”who does Jane wonder/claim where Jane will punish Jane with”	1.99785e-5
0.00314	”who does Jane wonder/claim how Jane will punish”	1.99785e-5
0.00038	”who does Jane wonder/claim how Jane will punish with”	2.4298e-6
0.00629	”who does Jane wonder/claim how Jane will punish Jane with”	1.99785e-5
0.00314	”who does Jane wonder/claim why Jane will punish”	1.99785e-5
0.00038	”who does Jane wonder/claim why Jane will punish with”	2.4298e-6
0.00314	”who does Jane wonder/claim why Jane will punish Jane with”	1.99785e-5
0.00019	”who does Jane wonder/claim that Mary will punish”	0.01710
2.3326e-5	”who does Jane wonder/claim that Mary will punish with”	0.00208
0.00019	”who does Jane wonder/claim that Mary will punish Mary with”	0.01710
0.00677	”who does Jane wonder/claim whether Mary will punish”	0.0
0.00081	”who does Jane wonder/claim whether Mary will punish with”	0.0
0.00671	”who does Jane wonder/claim whether Mary will punish Mary with”	0.0

Figure 3: Sample Minicorpus for Embedded Verb ’punish’

	Entropy Reduction Summed Across the Sentence
+Q	$\sum H \downarrow = 3.379$ bits
-Q	$\sum H \downarrow = 1.899$ bits
ΔQ	1.480 bits

Figure 4: Total Entropy Reductions with Argument/Adjunct Parameter at 0.50/0.50

4. Discussion

We report in the following section some background results which we argue are congruent with the account developed in this work.

4.1. wh-Islandhood is Combinatorial in Nature

Kluender and Kutas (1993) found that embedded interrogatives impose more cognitive load than embedded declaratives even when no extraction from the embedded sentential takes place. They crossed a Question-type factor with two conditions, a Yes/No Question control Condition and a *wh*-Question condition, with an embedding factor with three conditions: a ‘claim that’ condition, a ‘wonder whether’ condition, and an embedded question (‘wonder who’) condition. Subjects participated in an neurolinguistic EEG task and a speeded acceptability judgment task. The Yes/No Question Condition controls for extraction out of the embedded sentential while preserving the interrogative mood. As seen in Fig. 7, subjects in the neurolinguistic EEG task exhibited greater deflection in ERP in the ‘wonder whether’ and the ‘wonder who’ conditions than the ‘claim that’ condition, even when extraction out of the embedded sentential did not occur. Kluender and Kutas (1993) propose that these results indicate separate, discrete correlates for *wh*-processing and island-structure building.

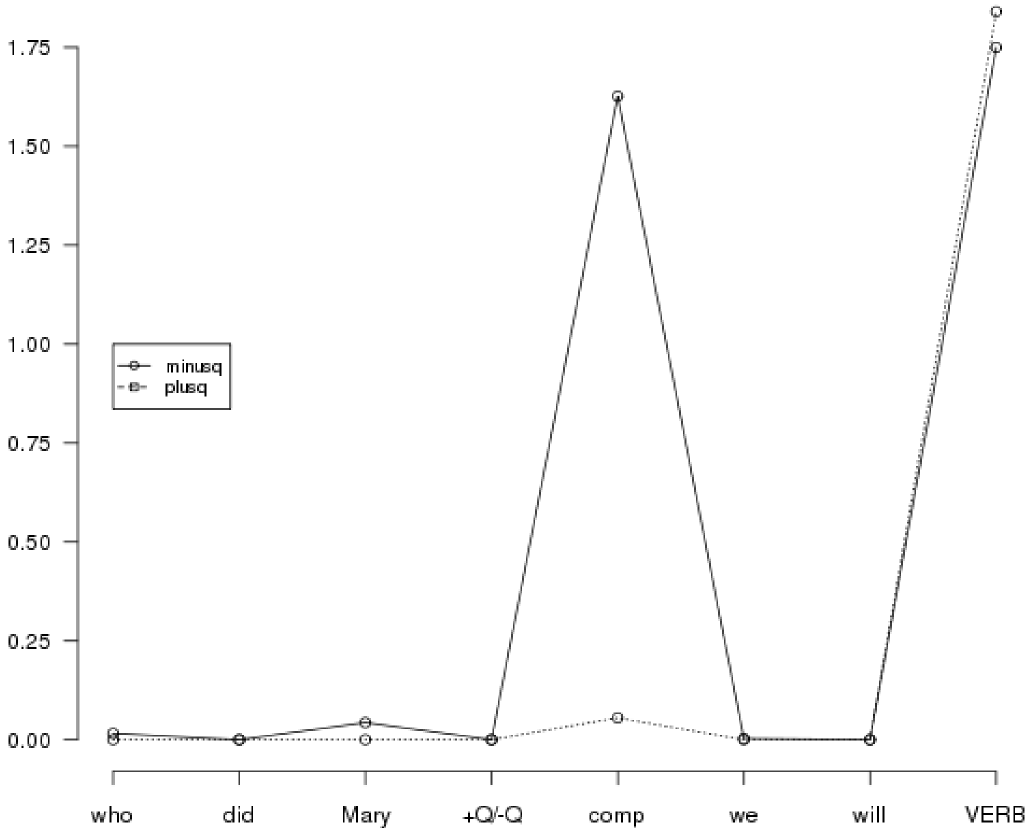

Figure 5: Incremental Entropy Reductions for Island and Control Conditions

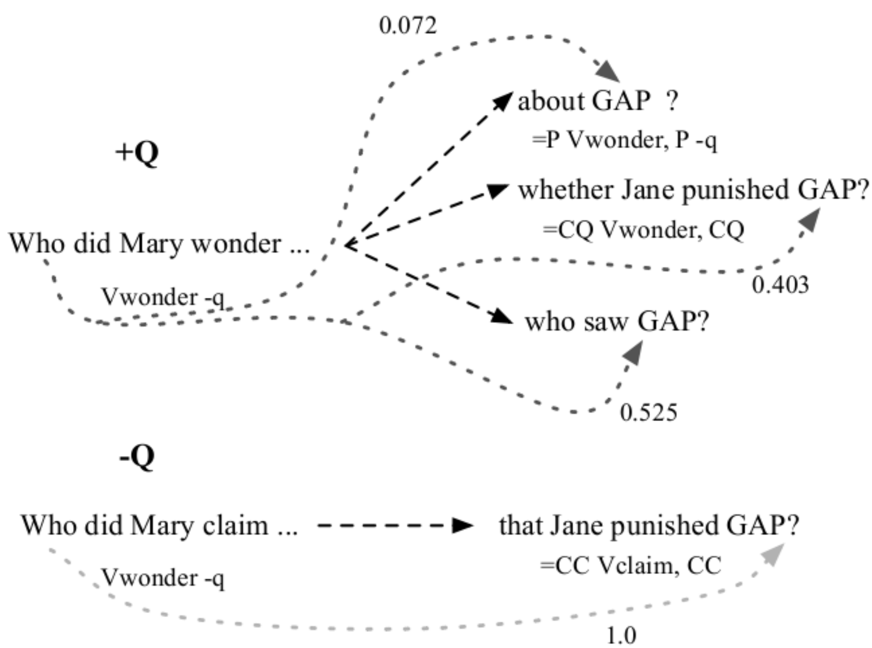

Figure 6: Branch Probabilities for Island and Control Conditions After Verb

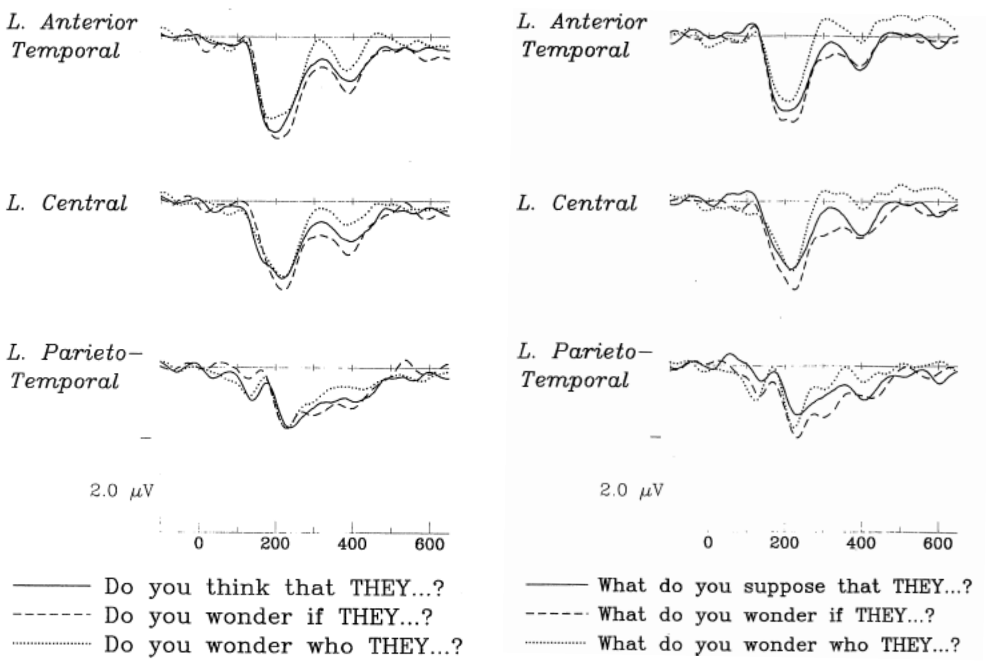

Figure 7: (Kluender and Kutas, 1993)

This effect straightforwardly follows from the +Q Ambiguity (Bresnan, 1970), which our model captializes on.

4.2. wh-Islandhood is Ameliorated by Extragrammatical Factors

Hofmeister and Sag (2010) found that potential islands with ‘Complex’ extraposed NP arguments (such as in 4a) are processed more quickly than Simple cases (such as in 4b).

(4) a. Which employee did Albert learn whether they dismissed after the annual performance review? (p.30)
    b. Who did Albert learn that they dismissed after the annual performance review? (p.30)

That this clearly extragrammatical level of treatment remedies the *wh*-island effect suggests that the proper explanation of the effect itself is extragrammatical and therefore a performance artifact. Hofmeister and Sag (2010) suggest that the effect may be due to working memory, but do not fully outline how.

We argue that the *wh*-island effect is due to the intrinsic work required of the parser by the structure-assigning task at hand. The asymmetry in the amount of intrinsic processing work required by the grammar originates from a structural ambiguity which is present in the Island condition but not the Control condition. Because our account predicts that the *wh*-island effect is the sum of the +Q ambiguity and the Filler-Gap ambiguity, it predicts that savings in the ambiguity budget for either ambiguity should yield easier processing. Hofmeister and Sag (2010)’s effect would then fall out as a case where +Q ambiguity is operant but Filler-Gap processing is ameliorated.

4.3. wh-Islandhood is not a Working Memory Phenomenon

However, Sprouse et al. (to appear) found no significant correlation between working memory capacity and the ability to process *wh*-islands. Neither serial recall nor n-back capacity predicted perceived acceptability on a variety of island processing tasks, including *wh*-islands. Given the tendency of acceptability judgments and on-line processing measures to correlate, this result would belie a working memory basis for the *wh*-island effect. Thus Sprouse et al. (to appear) suggest, contra Hofmeister and Sag

(2010), that island processing is a grammatical, not a processing phenomenon. The results of Hofmeister and Sag (2010) and Sprouse et al. (to appear) would appear at first glance to be irreconcilable.

As our account offers ambiguity rather than working memory as an explanation for the *wh*-island effect, it is not falsified by the results of Sprouse et al. (to appear) (to appear). We point out that the hierarchical processing of nested structures such as *wh*-islands bears little resemblance to the n-back and serial order working memory tasks employed in Sprouse et al. (to appear), which are predominantly tests for short term capacity for memorization of discrete items.

4.4. Conclusion

An additive-factors approach (Sternberg, 1969) to empirical experiments could tease apart the complexity of *wh*-island processing. The modeling methodology of this paper integrates with such an empirical program in a way which provides cohesion to the myriad number of factors involved in islandhood. It also provides for a possibly more informative set of results than appealing to the modularity of the grammar or the processor as the sole province of islandhood. One possible avenue of exploration would be to expand the repertoire of +Q and -Q verbs commonly used in experiments. Island experiments are almost entirely limited to the most common matrix verbs, such as ‘wonder’ (+Q), ‘claim’ (-Q), and ‘say’ (ambiguous); Hofmeister and Sag (2010) is a notable exception in that it uses twelve verbs with ambiguous +Q/-Q status, but it lacks a control condition with pure -Q verbs. By compiling experimental results where conditions utilize verbs across the +Q/-Q continuum, verb ambiguity provides us with a new independent variable we can modulate in experiment.

References

Theodora Alexopoulou and Frank Keller. Locality, Cyclicity, and Resumption: At the Interface between the Grammar and Human Sentence Processor. *Language*, 83(1):110–160, 2007.

Sylvie Billot and Bernard Lang. The structure of shared forests in ambiguous parsing. In *Proceedings of the 27th annual meeting on Association for Computational Linguistics*, pages 143–151. Association for Computational Linguistics Morristown, NJ, USA, 1989.

Joan W. Bresnan. On complementizers: Toward a syntactic theory of complement types. *Foundations of Language*, 6(3):297–321, 1970. ISSN 0015-900X.

Oliver Christ. A Modular and Flexible Architecture for an Integrated Corpus Query System. *COMPLEX’94, Budapest, 1994*, 1993.

Kyle Grove. mcfgcky. Unpublished manuscript, Cornell University, November 2010.

John T. Hale. Uncertainity About The Rest of The Sentence. *Cognitive Science*, 2006.

John T. Hale and Edward P. Stabler. Strict deterministic aspects of minimalist grammars. *Logical aspects of computational linguistics*, pages 162–176, 2005.

Philip Hofmeister and Ivan A. Sag. Cognitive constraints and Island effects. *Language*, 86(2):366–415, 2010. ISSN 1535-0665.

Robert Kluender and Marta Kutas. Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8 (4):573–633, 1993.

John R. Ross. Constraints on variables in syntax. *Unpublished doctoral dissertation, Massachusetts Institute of Technology*, 1967.

Evan Sandhaus. The New York Times annotated corpus. *Linguistic Data Consortium, Philadelphia*, 2008.

John Sprouse, Matt Wagers, and Colin Phillips. A test of the relation between working memory capacity and syntactic island effects. *Language*, to appear.

Edward P. Stabler. Derivational minimalism. In *Logical Aspects of Computational Linguistics.*, pages 68–95. Springer, 1997.

Saul Sternberg. The discovery of processing stages: Extensions of donders’ method. *Acta psychologica*, 30:276–315, 1969.

Proceedings of the 29th West Coast Conference on Formal Linguistics

edited by Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman

Cascadilla Proceedings Project Somerville, MA 2012

Copyright information

ISBN 978-1-57473-451-5 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Grove, Kyle. 2012. +Q/-Q and the Ambiguity Hypothesis of *wh*-Islands. In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, ed. Jaehoon Choi et al., 346-352. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2720.