LING229 Islands Seminar
McCloskey & Wagers, UCSC Linguistics Twenty-ten

Introduction, 6 January 2010

(1) Core questions:
   i. Generally, how do we account for the fact that certain strings are unacceptable?
   ii. In particular, how do we account for island phenomena?

Where islands refer to grammatical domains that block A’ dependency formation
   iii. How do we deal with gradience in acceptability measurements?

   How do we assign blame between different components of our theory?

(2) Three modes of inquiry

What’s at stake, I

(3) Unbounded movement
   i. What did John buy?
   ii. What do you think [that John bought __]?
   iii. What do you think [that John believes [he should buy __]]?

   etc.

(4) Some islands
   i. *What do you wonder [whether John bought __]?
   ii. *What did you make [the claim that John bought __]?
   iii. *What do you think [the speech about __] interrupted the TV show?
   iv. *What do you worry [if John buys __]?

(5) Some (non-exclusive) solutions to the problem of ‘overgeneration’ and islands
   i. General combinatory rules augmented with filters or conditions on rule applicability. E.g., i.a.,

   Subjacency: A cyclic rule cannot move a phrase from position Y to position X
   \[... X \ldots [\alpha \ldots [\beta \ldots Y \ldots ] \ldots ] \ldots X \ldots , \] where
   Where $\alpha, \beta$ are cyclic nodes

   Ch. ‘73
ii. Interface conditions are formulated to exclude island-generating derivations. Islands lead to contradictions in relating dominance and linear order. Fox & Pesetsky (2004), Uriagereka (1999)

iii. An island-generating derivation is available in principle, but not in practice. Syntactic competence is masked by properties of language processing.

(6) Some sentences are difficult or impossible to understand
   
   (i) The sentence the linguist the article criticized analyzed contained an island.
   (ii) The linguist published in the prestigious journal celebrated.

   Comprehension difficulty linked to perceptual mechanisms (Miller & Chomsky, 1963, Frazier & Fodor 1978, ...). Both for analytical and empirical reasons. Building a grammar to exclude (i) is difficult, (ii) is a ‘morphological accident’. Medial V drop leads to acceptability improvement:

   (iii) The sentence the linguist the article criticized contained an island.

(7) Is the impossibility of extraction from island-domains a reflection of the processing difficulties they pose?

(8) i. What are the sources of difficulty?
    ii. What is the time-scale at which difficulty operates?

(9) Multiple routes from perceptual difficulty to constraints
   i. ‘Adaptive grammar’/’grammaticization’ accounts. Certain kinds of movements would lead to real-time computational difficulties. The content of the grammar prohibits those movements, and is in that way adaptive to the real-time grammar user. Fodor (1978), Berwick & Weinberg (1984), Hawkins (1999):

   E.g. Fodor (1981) (though not referring to island constraints)
   ... the grammar is complicated by restrictions on otherwise quite generally applicable rules, and the complication is directed specifically towards cases where the rules threaten to create ambiguous constructions.

“On the basis of these arguments, we suggest a grammar of filler-gap dependencies (FGDs) that is far simpler than that generally countenanced in the syntactic literatures, more ‘minimal’ even than Chomsky’s (1995) ‘Minimalist’ Program. In our view, a competence grammar licenses a superset of the possible filler-gap sentences, leaving it to processing factors to explain why only a proper subset of these grammatically licensed sentences ... are fully acceptable.”

[Hofmeister & Sag, ms, p. 3]

iii. **Functionalist accounts that are less clearly committed to a time-scale**
Syntactic restrictions are replaced in favor of requirements of discourse cohesiveness, well-formed conceptual structures, etc. Notions of cohesiveness/well-formedness for other levels of representation are, in turn, sometimes justified with reference to notions like spreading activation. Kuno (1973), Erteschik-Shir (1982), Deane (1991), etc.

(10) *What do you wonder [whether/if John bought __]?*

e.g. Kluender & Kutas (1993), et seq.
The lowered acceptability of island-violation containing sentences reflects the compounding of number of factors that independently contribute difficulty.

i. **whether/if clauses are more costly than clauses headed by that, even if there is no movement dependency**

Who knows whether John bought a car? >> (more difficult, less acceptable)
Who knows that John bought a car?

ii. **Longer dependencies are more costly than shorter ones.**

Who knows that John bought a car? >>
Who does John know that Susan met?

(11) Research strategy

i. Characterize the claims in (10)a-b (and similar claims).

ii. Specify mechanisms.
Data from Sprouse, Wagers & Phillips, *in prep.* Technique is magnitude estimation.

(13) How do we parse filler-gap dependencies?  
Some suggested processes. Which of these processes is used, and in which order?

i. Decide a constituent is a filler.  
ii. Encode/store the filler (and its context) appropriately.  
iii. Decode gap positions.  
iv. Consult/retrieve the/filler encoding.  
v. Evaluate appropriateness of the filler with respect to local conditions  
vi. Evaluate fit of the filler with respect to global conditions.  
vii. Update encodings of related constituents.

(14) Clearing gap positions (Fodor, 1978)

i. The dog that Mary would like to buy at the store  
ii. The dog that Mary would like to walk to the store  
iii. The dog to whom Mary mentioned that Susan read a story

(15) Storing a filler

i. Are there special memory devices associated with movement dependencies?

Wanner & Maratsos (1978):  
Fillers occupy a distinguished representational state (the HOLD cell) until the dependency is completed.

Fillers are ‘retrieved’ from memory once a gap is posited, but they are not in a special state. They are simply a recent encoding in memory like any other.
ii. Are unsatisfied movement dependencies ‘costly’?

“Filler-gap dependencies are difficult structures to process ... Identifying the gap is not easy. It is an empty element with no surface manifestation and its presence must be inferred from its immediate environment. At the same time, the filler must be held in working memory, and all other material on the path from filler to gap must be processed simultaneously, and the gap must be correctly identified and filled.” [Hawkins, 1999, p. 246-7]

“Now in a sentence involving extraction, we have hypothesized that the extraction site and the extracted NP must command attention. Of course, they would elicit attention most readily in the absence of competing elements (distractors).” [Deane, 1991, p. 36]

“It was the fearless passengers [1] who the able sailor [2] advised [3] about the lifeboats [4] although the heavy storm was quickly abating [5]”

![Asymptotic Accuracy for 1st NP as a function of Probe Position](image)

**Fig. 4.** Asymptotic accuracy for the (synonym) probe recognition task with cleft constructions (filled squares), simple main clause constructions (open squares), and sentential complement constructions (open circles).
(17) How could parsing be limited by finite resources to yield severe unacceptability?
(18) Multiple processes compete for a single 'pool' of resources

(E.g., Capacity Constrained Comprehension, Just & Carpenter, 1992)

(19) Familiar orderings of acceptability
    Chomsky (1973)
    i. Who did you see pictures of?
    ii. Who did you see the pictures of?
    iii. Who did you see John’s picture of?

    Kluender (1992) [C&M83]
    iv. This is the paper that we really need to find someone who understands
    v.  This is the paper that we really need to find a linguist who understands
    vi. This is the paper that we really need to find the linguist who understands
    vii. This is the paper that we really need to find his advisor, who understands
    viii. This is the paper that we really need to find John, who understands

(20) Establishing the intervener in a discourse model varies in difficulty/resources consumed. The easier the first process, the more resources 'left over' to complete the dependency.
Individual variability in capacity

More generally
i. Linking fillers/gaps requires the participation of working memory and the coordination of attentional processes.

ii. In a broad sense, we can think of working memory as the set of processes and representations that manages access to recently encoded information. [Without making any particular claim about the architecture of WM; nothing said about special buffers, separability from LTM, etc.]

iii. Assume these processes have a non-zero failure rate.
iv. As more processes are required to parse a sentence, that failure rate is compounded.
v. The success rate in parsing translates onto an acceptability scale.

Islands do suggest, however, a distinction between the *computability of a representation* at some level and the *interpretability of a string*.

i. That is, the sense that extraction from islands is unacceptable is at least partially independent of the ease with which the extraction could be interpreted.

ii. A hard to quantify, but nonetheless salient, intuition that the difficulty presented by islands is not the same as that presented by, say, center embedding.

iii. Parasitic gap phenomena are thus of interest here. As is resumption.
What’s a stake, II

(24) Acceptability judgments are gradient. What does this mean?

(25) *E.g.* *Barriers* (Chomsky, 1986)

\[ \beta \text{ is } n\text{-subjacent } \alpha \text{ iff there are fewer than } n+1 \text{ barriers for } \beta \text{ that exclude } \alpha \]

(26)

i. What did Simon spread the rumor that they started? 2-subjacent
ii. What did Harold wonder whether they had ruined? 2-subjacent
iii. What did the captain give the command to start? 1-subjacent
iv. Who did Adele wonder whether to invite? 1-subjacent

In contrast to an explanation couched in terms of processing, a grammar-based characterization of the amply documented, pervasive graded acceptability of sentences with FGDs [filler-gap dependencies] would therefore require some highly specialized linguistic machinery. This may include FDG-specific constraints like Subjacency, as well as a system for calculating fine-grained acceptability differences. Notably, such a calculator would in all likelihood serve no function other than to tally the result of aggregating grammaticality violations. Additionally, such a system must also provide for the possibility that the interaction of grammatical constraints can lead to interactions in sentence processing. [Hofmeister & Sag, p. 35]

The notion “acceptable” is not to be confused with “grammatical.” Acceptability is a concept that belongs to the study of performance, whereas grammaticality belongs to the study of competence. ... That is, the generative rules of the language assign an interpretation to them in exactly the way in which they assign an interpretation to somewhat more acceptable sentences (1). Like acceptability, grammaticality is, no doubt, a matter of degree (cf. Chomsky, 1955, 1957, 1961), but the scales of grammaticality and acceptability do not coincide. Grammaticality is only one of many factors that interact to determine acceptability. [Chomsky, 1965]

In conclusion, we would say that it much more likely to be useful to adopt a model in which:

• The grammar parses what it can of an input stream, spewing out the unparsable material as an “error” (or, perhaps, simply ignoring it).
• The *listener* may use a wide variety of higher-order cognitive processes to attempt to determine the source of the error. These would include at least: knowledge about the speaker (where is this person from? how do people talk there? is the speaker intoxicated?); knowledge about the likely communicative intent of the speaker (s/he must have meant ‘kicked Peter’ when s/he said ‘picked Keter’; s/he was going to finish that sentence in some way like this, but got distracted by the fact that her/his hair was on fire. . . ); how does this string differ from an error-free one which is like it (e.g., if one just discards the error as a total mistake, is the resulting parsed string coherent, given the source of the string and his/her likely communicative intent), or must one interpret the error as an attempt to say something other than what was said; etc.
• Making these judgments may entail sending several alternative versions of the string heard – constructed by the listener by manipulating the string in light of the information considered above – through the grammar for parsing until a *minimally divergent* but *maximally coherent* (given the above considerations) string is generated.
• The grammar’s outputs consist solely of the parsed string and any residual unparsed material (identified by the fact that it is not in the parse) – there is no gradedness in those representations. [Hale & Reiss, 2008]
Miller (1988) discusses a related problem in cognitive psychology in general:

The issue of whether human information-processing is discrete or continuous is enormously complex theoretically, for the three reasons outlined in Part 1. First, discreteness and continuity are not a dichotomy, but rather a dimension on which there are intermediate possibilities. Second, stages can be discrete or continuous in at least three different senses: information representation, transformation, and transmission. Third, some of the stages used to perform a task may be discrete and others continuous, even in a given sense. Thus, there is a graded multidimensional space of information-processing models differing with respect to discreteness and continuity. Future comparisons of models, both theoretical and empirical, must acknowledge the great complexity of this space to avoid overgeneralization and equivocation.

Consider an acceptability rating task, with sample data as below:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Unacceptable &lt; - - - - &gt; Acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj/NP</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>(Island)</td>
<td>124 71 34 18 16 6 15</td>
</tr>
<tr>
<td>Obj/NP</td>
<td>66 43 62 28 37 15 33</td>
</tr>
<tr>
<td>Subj</td>
<td>19 20 25 20 34 44 122</td>
</tr>
<tr>
<td>Obj</td>
<td>10 13 21 22 31 44 143</td>
</tr>
</tbody>
</table>

Subj/NP = subextraction from NP in subject position
Obj/NP = subextraction from NP in object position
Subj = gap in subject position
Obj = gap in object position

What is the significance of the fact that raters use all categories?

There are not only multiple cognitive components that determine the acceptability rating assigned to a sentence, but multiple possible mapping principles.

For example, we could assume truly binary ‘grammaticalness’, and variable confidence in the system’s estimation of grammaticalness.
(36) (This actually turns out to be a reasonable model ...)

(37) *Lesson:* The graded-ness of acceptability judgments provides relatively weak constraints on our models.

(38) It will be likely necessary to draw together other kinds of measurements/tasks.
   i. where are the ‘anchors’
   ii. where is the covariation
   iii. how does information used in judgment become available over time

(39) **Priming**
   Luka & Barsalou (2005): Obtains for moderately complex sentences like:
   *Who said my brother was kept tabs on by the FBI?*

(40) Does priming occur for all sentence types?
   i. Satiation obtained in ‘Yes’-‘No’ tasks:
      Sprouse (2009): Response equalization when stimuli are overwhelmingly bad.
ii. Suggestive evidence from using a 7-pt scale that islands don't prime. More research is needed.

\[(41)\] Time course measures

<table>
<thead>
<tr>
<th>Violation Type</th>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
<th>Example 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subcategorization violation</td>
<td>T1. It was the essay that the writer knew the editor had glosted.</td>
<td>T2. It was the essay that the writer knew the editor had amazed.</td>
<td>T3. It was the essay that the writer knew the editor had admired her.</td>
<td>T4. It was the essay that the writer scolded the editor who admired.</td>
</tr>
<tr>
<td>Thematic role violation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filled-gap violation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Island violation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Graph showing ungrammatical filler-gap constructions]
Figure 2. A sample trial sequence illustrating the response-signal speed-accuracy trade-off variant of the grammaticality judgment task used in Experiment 2. Following a 500-ms fixation point, the words in a string are presented sequentially at a rate of 250 ms/word. A tone sounds at one of seven times (0.014, 0.1, 0.243, 0.5, 0.8, 2, and 3 s) after the onset of the final word in the string. Participants were trained to respond to the interruption tone within a 100–300-ms window.