Strong and weak islands: a theory of graded accessibility of linguistic domains

Whitney Tabor, Sandra Villata, & Jon Sprouse (University of Connecticut)
whitney.tabor@uconn.edu

Structural approaches to linguistic islands generally treat them as domains out of which extraction is prohibited. On this view, islandhood is an all-or-nothing thing. Yet psycholinguistic results show that naive subjects reliably assign intermediate judgments to certain island violations. For example, *whether* and Complex NP islands with lexically specified (i.e., "D(iscourse)-linked") Wh-elements (*which* N) (1)-(2) are rated worse than perfectly grammatical controls, but are not rated as low as their non-D-linked (*what*/*who*) counterparts [1]. Additionally, if we control for processing factors by defining "island effect size" as the decrement caused by extraction from an island beyond these factors [1], then the D-linked weak islands just mentioned have smaller effect sizes than (strong) *because*-Adjunct and Subject islands ((1)-(4), Fig. 1). These data provide evidence that intermediacy exists strictly within the island domain. Such data make it important that the theory of language address acceptability judgment gradience.

In seeking a principled approach to gradience, we note that researchers have invoked semantics to account for exceptions to broad syntactic generalizations about extraction constraints [3][4]. These approaches generally claim that semantics and syntax cooperate in producing binary judgment values so they do not address gradience. Inspired by [5][6][7], we propose a different relation between syntax and semantics: Self-Organized Sentence Processing (SOSP) [8][9]—see also[10][11]). In this framework (Fig. 2), each perceived word activates a syntactic treelet with vectors of semantic and syntactic features on its mother and daughter nodes. Bonds, which have continuous valued activations, form between the treelets in all possible ways and grow or decline in strength competitively. When the system encounters an ungrammatical sentence, it doesn't necessarily fail to form a parse—it can form a tree with some feature clash, giving rise to intermediate grammaticality. The role of semantics is not dual to that of syntax, but rather both kinds of features guide structure-formation; there is only one representational space.

In the case of weak and strong islands, this system predicts the judgment values found by [2] as shown in Fig. 1 (blue). The treelets are derived from a lexical dependency grammar for English. Long-distance dependencies are modeled via slash-propagation [12]. The model generates intermediate acceptability values for the extractions from D-linked *whether* islands because, although the grammar does not permit slash features to pass through *wonder*+CP or *whether*, these treelets are semantically similar to treelets like *think* and *that* which allow slash passing. This means *wonder* and *whether* can be coerced into linking filler and gap, resulting in an interpreted structure with mild feature clash. Extractions from Complex NPs behave similarly because, for example, *hear the report* is coerced to *hear*. On the other hand, extractions from *because*-Adjunct and Subject islands have no close analogies, so they fail to be parsed.

Although intermediate acceptabilities are not generated by the dual syntax-semantics accounts above, a variant of these gives gradient predictions: syntax fails to link filler and gap, but then semantics discovers a substitution that works (e.g, "hear the report" $\rightarrow$ "hear") and the sentence gets an intermediate rating because semantics succeeds while syntax fails. SOSP differs from this account in claiming that semantics guides parsing. To test this prediction, we ran a Maze Task: participants read a sentence like "Which necklace did the detective wonder whether Alice stole..." word by word, and then they had to choose between two words up to the end of the sentence ("the/in gems./April."). SOSP predicts that, on many, (but not all) trials (due to noise), participants will posit a gap after "stole" and thus choose the PP ("in April") instead of the NP ("the gems") continuation. This is predicted less with non-D-linked Wh because of lower stability of abstract noun phrases and not at all in strong islands because these are not coercible. Indeed, Maze data for *whether,* Complex NP, and Adjunct-islands showed intermediate acceptability values and a D-linking effect in the first two types, but no such intermediacy or D-linking in *because*-Adjunct islands (Fig. 3). In sum, SOSP suggests that exploring a general theory of gradience may help address theoretically challenging issues surrounding islands.

(1) **Whether Island Extractions** [Weak] (*/?/✓ = traditional judgments, islands in brackets)
* What$_i$ /✓ Which necklace$_i$ does the detective wonder [whether Alice stole t$_i$]?
(2) **Complex NP Island Extractions** [Weak]
* What$_i$/? Which pie$_i$ did you hear [the rumor that Jeff baked t$_i$]?
(3) *because*-**Adjunct Island Extractions** [Strong]
* What$_i$/* Which snack$_i$ did the brothers wink [because the sisters abhored t$_i$]?
(4) **Subject Island Extractions** [Strong]
* Who$_i$/* Which politician$_i$ does the retiree think [the speech by t$_i$] interrupted the show?

Fig. 1. Difference of difference scores from Likert scale acceptability judgments on D-linked islands and controls in [2]. Each bar indicates the residual disapproval of an island violating sentence after processing effects have been subtracted out. (Gray=human, Blue=model)

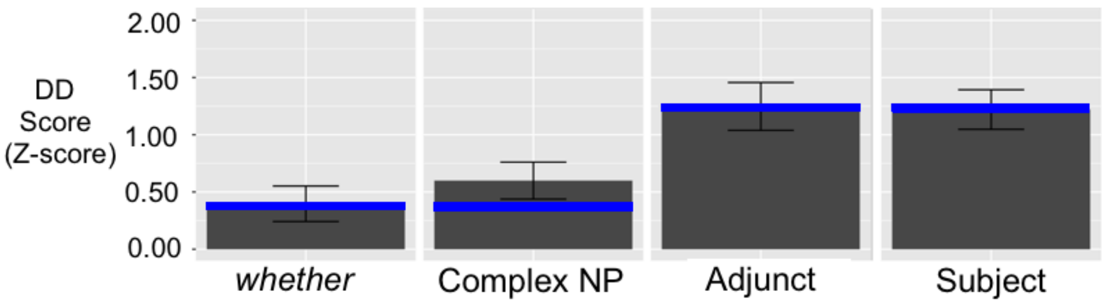
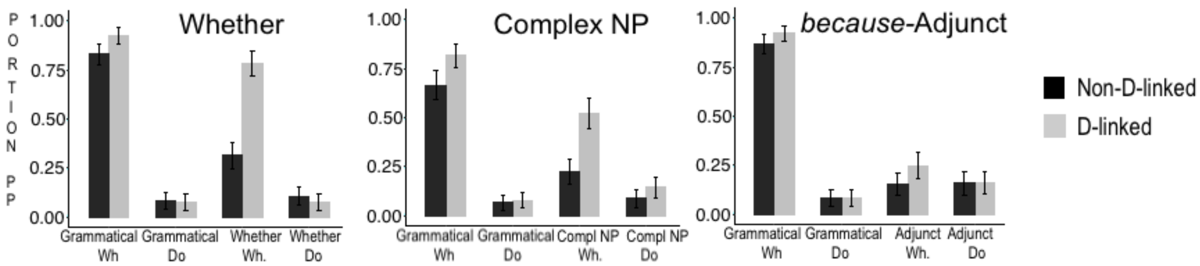


Fig. 2. Let $\mathbf{x}$, a vector, encode the state of the language processor. The model is a set of differential equations in $\mathbf{x}$ governed by a potential function called "Harmony" ($H(\mathbf{x})$). Each harmony peak corresponds to a stable configuration of bonds (fully grammatical structure, suboptimal structure that is a coherent tree, or failed parse [mix of partially completed trees]):

A structure's harmony = product of bond harmonies (each reflecting degree of clash)	$h_i = \prod_{l \in bonds} \left(1 - \frac{dist(\mathbf{f}_{l,daughter},\, \mathbf{f}_{l,mother})}{nfeat}\right)$
Radial basis functions define peaks at each structural locus.	$\phi_i(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x}-\mathbf{c}_i)^\top(\mathbf{x}-\mathbf{c}_i)}{\gamma}\right)$
The global harmony at any point is the height on the flank of the locally dominant peak.	$H(\mathbf{x}) = \max_{i \in 1 \ldots n} h_i \phi_i(\mathbf{x})$
Competitive bond formation and feature specification is noisy hill-climbing on the landscape.	$\frac{d\mathbf{x}}{dt} = \nabla_{\mathbf{x}} H(\mathbf{x}) + \epsilon$

Fig. 3. Portion of PP continuations in the Maze Task experiment. Wh = Wh-question, Do = Yes-No question. The third pair of bars in each panel shows the island extractions.



Refs. [1] Sprouse, J. et al. (2016). NLLT 34:307–44. [2] Sprouse, J. & Messick, T. (2015). How gradient are island effects?. NELS 46. [3] Szabolsci, A., & Swart, F. (1993). Weak Islands. Nat. Lang. Sem. 1, 235-284. [4] Davies, W. D., & Dubinsky, S. (2003). On extraction from NPs. NLLT 21 (1), 1-37. [5] Ross, H. (1987). Islands and Syntactic Prototypes. CLS 23: 309-320. [6] Kalyan, S. (2012). Similarity in linguistic categorization. Cog. Ling. 23(3). 539–554. [7] Brown, J. (2015). Blackholes and subextraction... CLS 51: 67-81 [8] Smith, G., & Tabor, W. (2018). 16th ICCM (p. 138-143). [9] Villata, S., Sprouse, J., & Tabor, W. (2019). 41st Cogsci, 1178-1184. [10] Levy, R. (2008). A noisy channel model... EMNLP 13, 234–243. [11] Cho, P. W., Goldrick, M. A., & Smolensky, P. (2017). . . . gradient symbolic computation. Linguistic Vanguard, 3, 76-96. [12] Gazdar, G. (1981). Phil. Trans. (Series B) of RS, 295, 267-83.