Modeling ungrammaticality: a self-organizing model of islands

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

Acceptability judgments are gradient, while grammaticality is traditionally conceived of as categorical (sentences are either grammatical or ungrammatical but cannot be "partially" (un)grammatical). Degrees of acceptability are thus traditionally attributed to extra-grammatical factors (e.g. memory limitations). We present a new way to account for degrees of acceptability: a self-organized sentence processing model (SOSP; e.g. [1]). On this model, gradient acceptabilities can be generated by the grammar itself. Indeed, grammaticality and ungrammaticality are seen as two endpoints along a continuum, rather than as discrete notions [2,3].

SOSP offers an independently motivated way of accounting for degrees of acceptability: linguistic structures result from continuous interaction and competition amongst sentential elements terminating when bonds are formed. When syntactic and semantic combination requirements cannot be satisfied (viz. no optimal bond is available, as in ungrammatical sentences and difficult garden paths), the system *forces* the formation of (sub-optimal) structures, giving rise to various more or less (un)acceptable outcomes.

We focus on islands, encapsulated syntactic environments out of which nothing, or very little, can be extracted, arguably the most prototypical yet challenging case of unacceptability. Islands come in two flavors, strong and weak [4]. Strong islands ban any extraction, while weak islands have been argued to allow for certain extractions. In particular, linguists have noted that they seem to allow for the extraction of a D(iscourse)-linked wh-element (*which NP*), while disallowing the extraction of non D-linked element (e.g. *what*) (e.g. [5,6]). In this work, we focus on two island types: subject islands (1d), which are strong, and whether islands (2d), which are weak. We address three empirical facts, that, together, present a challenge to traditional grammatical and parsing theories. First, weak island acceptability is gradient. Using a 2x2 factorial design for island effects [7,8] – in which the island effect is isolated from two processing factors – (i) DEPENDENCY LENGTH (long vs. short) and (ii) embedded STRUCTURE TYPE (island vs. non-island) (cfr. (1) an (2)) – it has been shown that D-linked whether islands are more acceptable than non D-linked ones, and yet still not fully acceptable (Fig.1) [9]. Second, D-linking interacts with island types: while D-linking ameliorates the acceptability of weak islands, it does not help strong islands (Fig.1) [9]. Third, D-linked weak islands with an intransitive embedded verb (*Which car do you wonder whether John* ***slept****?*) are less acceptable than those with a transitive (*Which car do you wonder whether John* ***bought****?*), while no such contrast is detected in non D-linked weak islands ([10] shows this in wh-islands). We take this as evidence that weak islands, though ungrammatical, are interpreted. Therefore, the dependency between the extracted wh- and the gap inside the island is established. This claim is challenging for traditional models which capture island effects by barring the formation of dependencies inside islands.

We ran 20 runs of the model which generated the observed data pattern (see Figure 1, blue lines). The model succeeds by coercing D-linked whether islands into a non-island structure via coercion of "wonder" into an approximation of "think", which licenses the propagation of a filler, and of "whether", which syntactically blocks chain formation, into "that", which allows it, but at the cost of lowering the overall grammaticality. For non D-linked whether islands, the coercion does not happen because non D-linked wh- lack sufficient featural richness to cause the system to discover coercion, resulting in failure to propagate the wh- inside the island, and very low grammaticality. For subject islands, no suitable structural analogy is available, resulting in a systematic failure to propagate the wh- inside the subject island and very low grammaticality. All in all, we argue that SOSP offers a valuable new way of approaching the relationship between grammar and processing. It is closely related to generative linguistic theory, but it differs in non-trivial ways from traditional assumptions, notably continuity, and a central role for processing in grammatical explanation. We hope our results will spur new discussion on these topics.

(1) Factorial design **subject islands**

a. NON-ISLAND, SHORT
Who/Which leader __ thinks the speech interrupted the TV show?
b. NON-ISLAND, LONG
What/Which speech does the leader think __ interrupted the TV show?
C. ISLAND, SHORT
Who/Which leader __ thinks the speech by the president interrupted the TV show?
d. ISLAND, LONG
Who/Which politician does the leader think the speech by __ interrupted the TV show?

(2) Factorial design **whether islands**

a. NON-ISLAND, SHORT
Who/Which woman __ thinks that John bought a car?
b. NON-ISLAND, LONG
What/Which car do you think that John bought__?
C. ISLAND, SHORT
Who/Which woman __ wonders whether John bought a car?
d. ISLAND, LONG
What/Which car do you wonder whether John bought __?

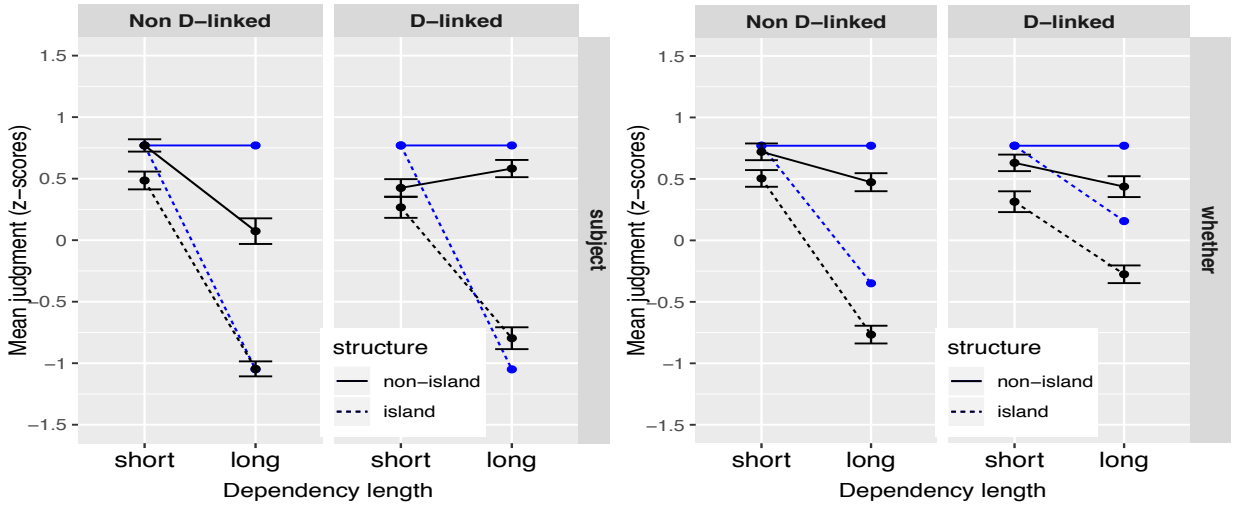

Fig. 1 Interaction plots between dependency length (long vs. short) and embedded structure type (island vs. non-island) - the term 'island' here does not refer to an island-violating structure, but to the presence of a structural domain that does not tolerate extraction (e.g. whether embedded clause). Only the long/island conditions violate island constraints. The island effect can be defined as a statistical interaction between the two factors (it is what remains after the linear sum of the two processing factors). All four interactions are significant, but the interaction for the whether island in the D-linked condition is reduced compared to the non D-linked condition. Empirical results are in black (data from Sprouse & Messick 2015) and results from the model's simulation are in blue (model variance within conditions was negligible).

References. [1] Smith & Tabor. 2018. Proceedings paper, International Conference on Cognitive Modeling. Madison, WI. [2] Kempen, G., & Vosse, T. (1989). Connection Science, 273-290. [3] Smolensky, P. (1986). Parallel distributed processing, volume I (pp. 194–281). MIT Press. [4] Szabolcsi. 2006. The Blackwell companion to syntax, 479-531; [5] Rizzi 1990. The MIT Press; [6] Cinque 1990. MIT, Cambridge; [7] Sprouse 2007. College Park, MD: University of Maryland dissertation; [8] Sprouse et al 2012. Language, 88, 83-123.; [9] Sprouse & Messick 2015. Poster presented at NELS 46; [10] Villata et al. 2015. Poster presented at Cuny.