SPAWNing priming predictions with an uncertainty-based reanalysis mechanism reveals distinct representations of reduced and full relative clauses in English.

Grusha Prasad (gprasad@colgate.edu)
Colgate University

Tal Linzen (linzen@nyu.edu)
New York University

Introduction: How can theoretical syntax facilitate the study of the representations that people construct when processing sentences? To explore this question, recent work [1] proposed an ACT-R based serial parser (SPAWN) that can be used to evaluate the extent to which syntactic theories align with behavior by generating priming [2] predictions from these theories. As a case study, this work generated priming predictions from two competing theories of reduced relative clauses, Whiz-Deletion [3] and Participial-Phase [4], and found that predictions from neither theory could capture the pattern of behavior observed in a comprehension-to-production priming experiment (Fig2). The authors hypothesized that the Whiz-Deletion theory might better capture empirical patterns of behavior with a reanalysis mechanism that allowed for backtracking of multiple parsing decisions. In this work we implement such a mechanism and test whether predictions from either theory under the new model can capture human behavior. Additionally, we study the extent to which the strength of prior knowledge influences the fit to behavior.

Model description: In SPAWN, every word is associated with one or more CCG-type syntactic "tags" which determines the parse states it can combine with. For example, the verb *examined* is associated with an active or passive tag (Tab1). For every word, the parser retrieves a tag with the highest activation and tries to combine it with the current parse state. If the combination is successful, it moves to the next word. If not, it retrieves another tag and repeats the process. If a combination is not possible with any of the tags, the model has to reanalyze some previous word. The original version of SPAWN adopted a brute force approach to reanalysis by backtracking one word at a time and discarding impossible combinations. While this approach is guaranteed to give the correct parse, it is cognitively implausible. In our new version, the model uses an uncertainty-based sampling approach to decide which word to reanalyze: words with more tags have a higher probability of being sampled. This allows the model to backtrack more than one parsing decision at a time, but is not guaranteed to give the correct parse by itself: for a word with many tags, if the incorrect tag has a high prior probability, then the model can get stuck in a loop of always selecting the incorrect tag. To avoid this, we made two additional modifications: First, when a word is being reanalyzed, we added inhibition to tags that were previously retrieved; this inhibition decays over time, just like activation (Eqn1). Second, after some *m* iterations, the parser ignores its prior knowledge, thus making it less likely for the model to keep sampling the same incorrect tag; we set m to 2000.

Generating predictions: As in [1], we created 1280 instances of Whiz-Deletion and Participial-Phrase versions of SPAWN by sampling different model parameters (see [5]). We modulated the strength of prior knowledge by training models on 0, 100 and 500 sentences, and used the same method as in [1] to generate predicted priming effects under the two theories.

Results: As the prior knowledge weakened, the Participial-Phase models predicted the graded pattern observed in humans both qualitatively and quantitatively, whereas the Whiz-Deletion models exhibited a qualitatively different graded pattern from humans (see Fig3 for interpretation of the patterns). These results suggest that, in contrast with predictions from [1], <u>the Participial-Phrase theory, which assumes that reduced and full RCs have different representations, better captures the representations that people construct.</u> More generally, this work highlights how jointly modeling the grammar and parser can help adjudicate between competing theories and improve our computational models of sentence processing.

References:

[1] Prasad, Grusha and Linzen, Tal (2022). Studying relative clause representations: a novel parsing model and priming paradigm. Oral Presentation at the 36th Annual Human Sentence Processing Conference.
[2] Branigan, Holly P & Martin J Pickering (2017). An experimental approach to linguistic representation. Behavioral Brain Sciences
[3] Chomsky, Noam (1957). Syntactic Structures. The Hague: Mouton.
[4] Harwood, William (2018). Reduced Relatives and Extended Phases: A Phase-Based Analysis of the Inflectional Restrictions on English Reduced Relative Clauses. Studia Linguistica.
[5] Vasishth, Shravan and Felix Engelmann (2021). *Sentence comprehension as a cognitive process: A computational approach*. Cambridge University Press

Parse state for "the cat"	Tag for "examined"	CCG notation	Combination
DP	Active	(TP\DP)/DP	TP/DP (looking for DP)
DP	Passive	VoiceP/PP	Not possible (trigger reanalysis now)
DP/VoiceP	Active	(TP\DP)/DP	(TP/VoiceP)/DP (looking for DP, then PP; this is not possible and so trigger reanalysis later)
DP/VoiceP	Passive	VoiceP/PP	DP/PP (looking for PP)
DP/CP	Active	(TP\DP)/DP	(TP/VoiceP)/DP (looking for DP, then PP; this is not possible and so trigger reanalysis later)
DP/CP	Passive	VoiceP/PP	Not possible (trigger reanalysis now)

Table 1: Illustration of how the two tags of "examined" can combine with different parse states.

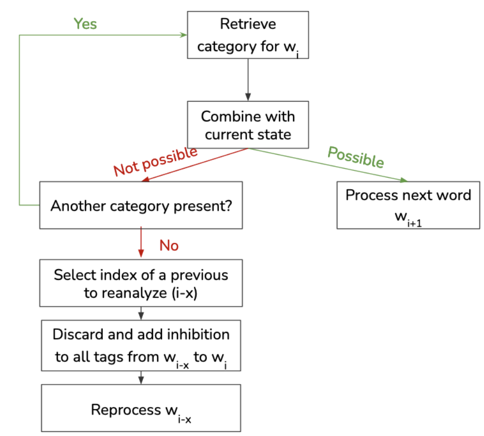

Figure 1: Schematic of reanalysis mechanism

Base level activation of tag j for word i and decay d

$B_{ij} = log \sum_{k=1}^{K} T_{ij_k}^{-d}$ where T_{ij_k} is the time between the *k*-th encounter of the tag and the current word.

Lexical activation for tag j for word i

L_{ij} = P(tag$_j$ | word$_i$) * max activation

Inhibition for tag j for word i in sentence s and decay d

$I_{ijs} = log \sum_{k=1}^{K} T_{ijs_k}^{-d}$ where T_{ijs_k} is the time between the *k*-th time the tag was discarded when processing s and the current time.

Overall activation of a tag j for word i in sentence s

Act(tag$_j$ | w$_{i,s}$) = $B_{ij} + L_{ij} - I_{ijs}$ + Normal(0.35,1)

Equation 1: SPAWN formulae for activation and inhibition

RRC prime: The cat examined by the doctor was skittish.
ProgRRC prime: The cat being examined by the doctor was skittish.
FRC prime: The cat who was examined by the doctor was skittish.
AMV prime: The cat examined the doctor and was skittish.

RRC target: The defendant examined ___

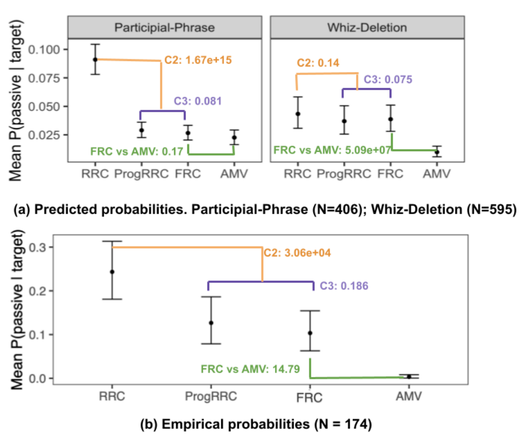

Figure 2: Predicted and empirical results taken from [1].

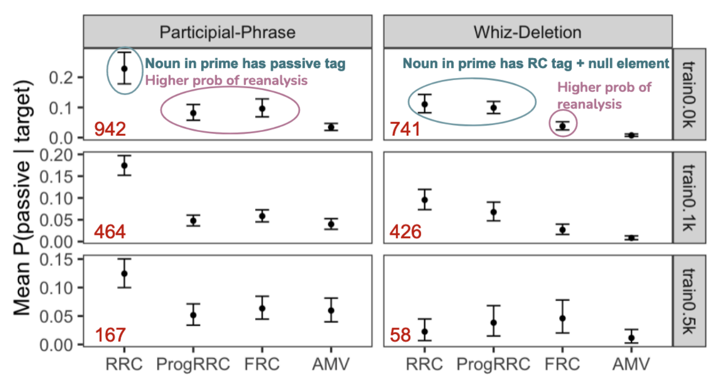

Figure 3: Posterior P(passive | target) in different prime conditions for models trained with 0, 100 or 500 sentences. Only models that generated at least one passive production were included (red numbers). Pink and teal text explain why the qualitative patterns are observed.