

## rnnpyscholing - Particle Shift (#11024)

### Author(s)

Richard Futrell (MIT/UC Irvine) - rfutrell@uci.edu  
Ethan Wilcox (Harvard University) - wilcoxeg@g.harvard.edu  
Takashi Morita (MIT) - tmorita@alum.mit.edu  
Roger Levy (MIT) - rplevy@mit.edu

Created: 05/18/2018 09:43 AM (PT)

Public: 10/26/2018 11:17 AM (PT)

### 1) Have any data been collected for this study already?

It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

### 2) What's the main question being asked or hypothesis being tested in this study?

An LSTM model shows locality effects in particle shift, i.e. shifted particles are especially high-surprisal when there is a long NP before them.

### 3) Describe the key dependent variable(s) specifying how they will be measured.

LSTM surprisal at the phrasal verb particle, and also total sentence surprisal, as measured using the Gulordava et al. (2018) LSTM and the Jozefowicz et al. (2016) LSTM.

### 4) How many and which conditions will participants be assigned to?

2x2 conditions crossing:

- \* particle shifted or not,
- \* object NP is short or long.

e.g.,

- \* Bob threw out the trash (unshifted/short)
- \* Bob threw the trash out (shifted/short)
- \* Bob threw out the old trash that had been sitting in the kitchen (unshifted/long)
- \* Bob threw the old trash that had been sitting in the kitchen out (shifted/long)

### 5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will predict the surprisal at the particle and total sentence surprisal using linear regression. A locality effect is manifested as a positive interaction of conditions shifted and long, indicating that the long-shifted case is superadditively bad.

### 6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will exclude data from sentences that turn out to contain UNKs before or at the critical particle.

### 7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We have constructed 32 items. We believe this will be enough.

### 8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Surprisal values from the Gulordava et al. (2018) LSTM have been collected but not looked at or analyzed.