EEG correlates of syntactic expectation reflect both word-to-word and hierarchical dependencies

Jonathan Brennan  
Max Cantor  
Rachael Eby  
John T. Hale  

University of Michigan  
University of Colorado Boulder  
University of Michigan  
Cornell University
What kind of syntactic information guides expectations during every-day sentence comprehension?

Adapted from Caplan, 1992 p. 272
Evidence from standard controlled experiments

...for hierarchical structures

- Classical structural ambiguities, garden path effects etc.
- Grammatical constraints guide online resolution of pronominal reference (e.g. Nicole & Swinney, 1989; Sturt & Lombardo 2002) and long-distance dependencies (e.g. Stowe 1986, Phillips 2006; reviewed by Lewis & Phillips 2015)
- Structure-based expectations predict reading times (e.g. Hale 2006)

...for surface heuristics

- Interpretation may reflect surface sequences that conflict with structure (Reviewed by Sanford & Sturt 2002; Ferreira & Patson 2007; Karimi & Ferreira 2016; cf. Kuperberg 2007)
- ERPs sensitive to linear expectations, but not hierarchical structure (Frank et al. 2015)
Information used during comprehension may be contingent on task structure and listener goals

linguistic representations built in the course of language processing are only good enough to tackle the task at hand

Karimi & Ferreira, 2015

Do every-day comprehension tasks like skimming the newspaper or passively listening to a story demand hierarchical representations?
Mixed evidence for hierarchy during naturalistic comprehension

- Eye-tracking data from newspaper text (Dundee corpus); test correlation with *word expectations* conditioned by different information

- **Con** Frank & Bod 2011: Expectations based on context free grammars don’t improve regression fits over surface-level models

- **Pro** Fossum & Levy 2012, Van Schijndel & Schuler: Better fits for grammar-based expectations with alternate model parameters (also Demberg et al., 2012)

- Our contribution: Passively record EEG during extended naturalistic listening to test for neural signals modulated by surface and/or hierarchical expectations
Current study

Participants listen to a story during EEG recording

Quantify expectations based on surface-level and hierarchical representations with Surprisal (Hale 2001)

Correlate surprisal estimates with EEG signals (cf. Frank et al., 2015)

Test contribution of hierarchical representations above-and-beyond surface-level models
Methods
Protocol

• 17 participants listened to ch. 1 of *Alice in Wonderland* audiobook

• 12.6 m; 2129 words

• Comprehension quiz at end

• EEG recorded with 61 active electrodes (M10 montage)
Data

Linked-mastoid reference

Epoch at word onset
-0.3 - 1 s

ICA to remove eye-movement artifacts

Visual inspection to reject trials with excessive noise

Subtract baseline
-0.1 - 0 s

Filter from 0.5 - 40 Hz
Three probabilistic language models

• Surface-based Markov Model: \textit{ngram}

• Surface-based Recurrent Neural Net: \textit{rnn}

• Hierarchical Context-Free Grammar: \textit{cfg}
Surface-level Markov Model: *ngram*

- $\Pr(w \mid w_{-1} \ldots w_{-(n-1)})$

- Defined over Part of Speech (POS) tags
  (Stanford parser; Klein & Manning 2003)

- Trigram model ($n = 3$) estimated with OpenGRM, Witten-Bell smoothing
Surface-level Recurrent Neural Network: *rnn*

- Recurrent hidden layer; prior network state and current input both condition output activation

- Pr(\(w \mid \text{full observed left-context}\) )
  POS tags

- Estimated with the rwthlm toolkit,
  Hidden layer with 500 sigmoid units,
  SGD+BPTT learning (Sundermeyer et al., 2014; cf. Frank 2013)

```
PRP → VBP → WRB → JJ → NNS → PRP → VBP → [?]```

12
Hierarchical Context-Free Grammar: $\textit{cfg}$

- Probabilistic context-free grammar (Stanford parser; Klein & Manning 2003)
  
- $\Pr(w \mid \text{structure consistent with left-context})$
  
- Excludes terminal rules to match POS-based surface-level models (cf. “POS Surprisal”; Roark et al., 2009)
  
- Estimated with the EarleyX implementation of Stolcke’s probabilistic Earley parser (Luong et al., 2013; Stolcke, 1995)
Training

• All models trained on entire text of *Alice in Wonderland* (punctuation and chapter headings removed)

• Reflects comprehender sensitivity to local and genre-specific statistics in the stimulus (e.g. Fine et al. 2013)
Word-by-word POS Surprisal: $-\log_2(\Pr(\text{POS} \mid \ldots))$

**ngram**

VBP → WRB → [?]

**rnn**

... → VBP → WRB → JJ → NNS → [?]

**cfg**
Correlations between target variables

freq

ngram

rnn

cfg
Statistical Analysis

\[ EEG_{e,t} = \beta \text{ sound power} + \beta \text{ trial order} + \beta \text{ freq}_n + \beta \text{ freq}_{n-1} + \beta \text{ freq}_{n+1} + \beta \perp \text{ngram}_n + \beta \perp \text{rnn}_n + \beta \perp \text{cfg}_n + \epsilon \]

• Test higher-level contributions **above-and-beyond** lower models using orthogonalization (⊥)

• Content and function words modeled separately (ask me about function words)

• All predictors centered

• Group-level correction across times and sensors with cluster-based permutation test (10,000 reps; Maris & Oostenveld, 2007)
Results
Content Words

*p < 0.05, 462 - 604 ms
Content Words

*p < 0.05, 496 - 666 ms
Content Words
Content Words
Estimated ERPs for different POS Surprisal values

- 7 bits
- 4 bits
- 1 bit
- 8 bits
- 2 bits
- 5 bits

\( p = 0.07 \)
\( * p < 0.05 \)

\( \downarrow \) ngram
Estimated ERPs for different POS Surprisal values

- Cz

- ngram

- cfg

- p = 0.07

- *p < 0.05

- n.s.
Summary

- Surface *ngram* surprisal modulates central negativity (content words, 450–600 ms, marginal 360–440 ms)

- Surface *rnn* surprisal shows additional positive shift in right posterior electrodes (500–660 ms)

- Hierarchical *cfg* surprisal modulates central negativity above-and-beyond lower order models (500–700 ms)
Take-aways

- EEG signals reflect part-of-speech expectations conditioned by surface-level and hierarchical structure by ~500 ms

- Evidence that phrase-structure conditions online expectations during naturalistic listening
Thank you

Research funded in part by a grant from the Office of the Dean for the College of LS&A at the University of Michigan

Data collected with assistance from Claudia Kassouf and Salim Huerta
Extra
I wonder how many miles I've fallen by this time.

The hot day made her feel very sleepy and stupid.

Surprisal, Bits

A sort of mixed flavor of cherry tart, custard, pineapple roast, turkey, toffee, and hot buttered toast.

To come out among the people that walk with their heads downward.
Function Words

*p < 0.05, 804 - 928 ms
<table>
<thead>
<tr>
<th></th>
<th>This Study</th>
<th>Frank et al. 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>17</td>
<td>split 12/12</td>
</tr>
<tr>
<td>stimulus</td>
<td>Narrative</td>
<td>Isolated sentences from novels</td>
</tr>
<tr>
<td>model domain</td>
<td>POS</td>
<td>Words &amp; POS</td>
</tr>
<tr>
<td>CFG parser</td>
<td>EarlyX; parent-child probabilities</td>
<td>Roark; ancestor + sibling probabilities</td>
</tr>
<tr>
<td>model training</td>
<td>text-specific</td>
<td>large corpus of british english text</td>
</tr>
<tr>
<td>baseline</td>
<td>subtracted from data</td>
<td>included in regressions</td>
</tr>
<tr>
<td>corpus accuracy</td>
<td>not applied</td>
<td>filter for significant results</td>
</tr>
<tr>
<td>statistics</td>
<td>All sensors &amp; time-points</td>
<td>6 components based on previous literature</td>
</tr>
<tr>
<td>word class</td>
<td>split content &amp; function words</td>
<td>analyze all words; post-hoc split</td>
</tr>
<tr>
<td>results</td>
<td>cfg &gt; rnn &gt; ngram for POS</td>
<td>ngram &gt; cfg = rnn for words but not POS</td>
</tr>
</tbody>
</table>
End