Evaluating Parsing Strategies in Sentence Processing

Sarah VanWagenen[1,2], Jonathan Brennan[3], Edward P. Stabler[1]

1-University of California at Los Angeles, 2-Technion, Israel Institute of Technology, 3-New York University

Introduction

- *Terminology*
  - *Parser:* Formal model(s)
  - *Parsing Complexity:* Defined measures of parser effort
  - *Recognition:* Human recognition of syntactic structure
  - *Recognition Complexity:* Measures of human effort
  - *Prediction:* Anticipating upcoming categories/words
  - *Incrementality:* Word by word integration into existing structure

Background: Significant psycholinguistic evidence for top-down parsing, but few models of sentence processing incorporate standard complexity measures and properties of such a parser. (e.g.,Rayner 1983, Taraban and McClelland 1988, Tanenhaus et al 1995, Sturt and Lombardo 2005)

Research Question: Can we find neurophysiological evidence bearing on what parser provides the ***best model of human syntactic recognition*** by comparing **recognition complexity** to **parsing complexity**?

Goal: Ultimate-- Find the (Grammar, Parser) pair that maximizes data coverage.

Proximal-- Proof of Concept. Using an existing dataset, a theoretically motivated grammar and a psychologically motivated parsing strategy, are standard measures of parsing complexity at all correlated with neural activity in reasonable regions of the brain during sentence processing?

Minimalist Grammars

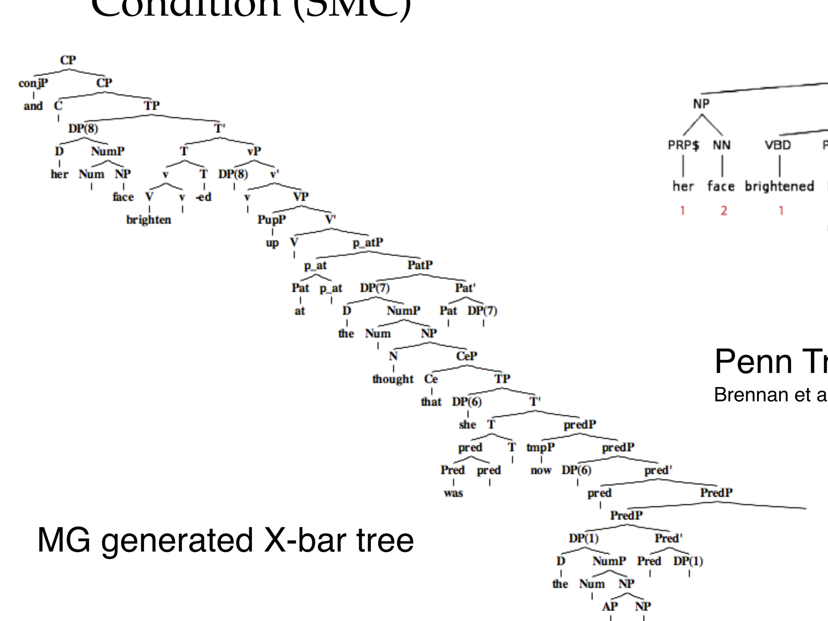
Most sentence comprehension models use Context-Free Grammars like the one used to parse the Penn Treebank

Emerging consensus that human languages must be at least Mildly Context Sensitive (MCS)

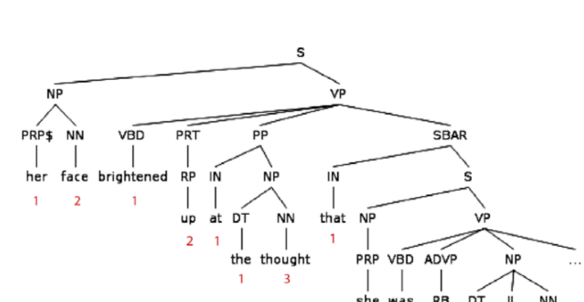
Minimalist Grammars (MGs) provide linguistic sophistication at the right level of complexity

Lexically defined, feature driven Merge and Move operations (Stabler 1991)

- An MG is a 4-tuple (V, Cat, Lex, $\mathscr{F}$)
  - V is a finite set of vocabulary items
  - Cat is a finite set of features of the following types:
    - x is a category feature (selectee)
    - =x is a selector feature
    - -x is a movement feature (licensee)
    - +x is a movement triggering feature (licensor)
  - $Lex \subset (V^* \times Cat^*)$ is a finite set of lexical items
  - $\mathscr{F}$ = {merge,move}: the set of generating functions
    - merge is a binary operation which takes two expressions and puts them together
    - move is a unary operation which takes a single expression and rearranges it parts
- Can implement any standard theoretical proposal
- Features of our MG:
  - binary branching
  - head movement
  - wh-movement
  - remnant movement
  - pied-piping
  - vp shells
  - left and right adjunction
  - Shortest Move Condition (SMC)
- Analyses used:
  - ditransitives (Larson 1988)
  - dative alternation (Baker 1997)
  - relative clauses (Kayne 1994)
  - passives (Baker, et al 1989)
  - case checking (Haegemann 1991)
  - PRO (Stabler 2003)
  - genitives, raising, ECM, control, quantification, wh-movement, expletives (Sportiche, Koopman, Stabler 2006)



MG generated X-bar tree

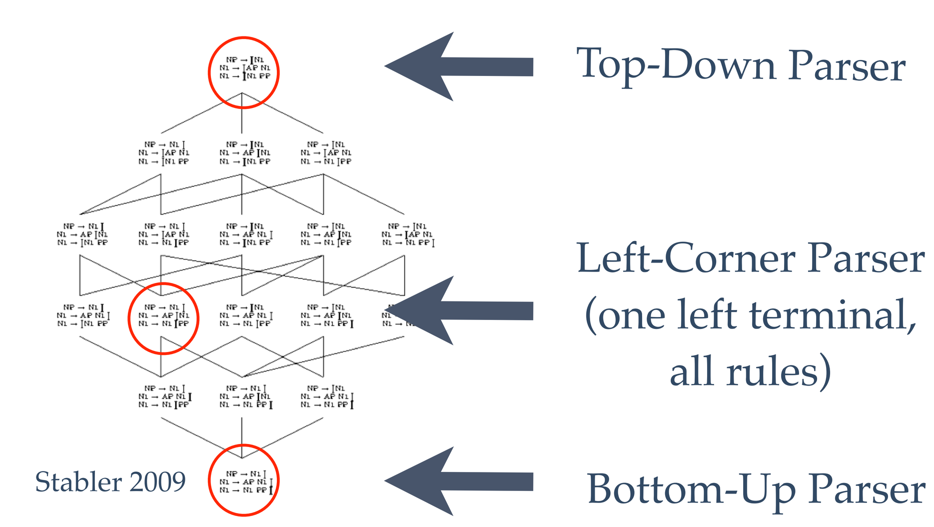


Penn Treebank tree
Brennan et al 2010

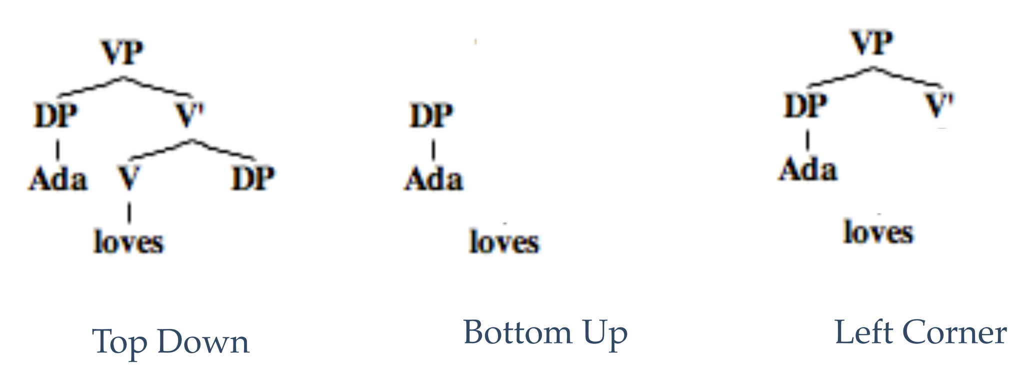
Logic and Control

- Logic: how the parser enumerates nodes and arcs of a tree
- Control: how the parser decides which rule to use when faced with non-determinism
- Information-theoretic measures of ambiguity account for a significant degree of recognition complexity
- Maintain distinction and focus on "logic" of the parser

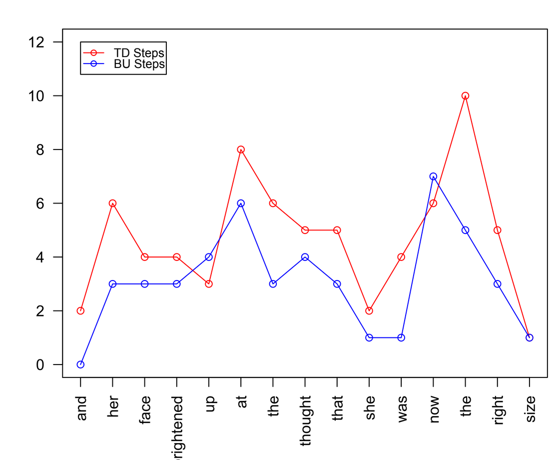
Parsing Strategies



- Broad agreement that parsing is *predictive* and *incremental*
- Only Top-Down parser is *fully* incremental and predictive



Parsing Complexity



Number of computational steps performed between the scan of each non-empty leaf node.

Top-Down and Bottom-Up

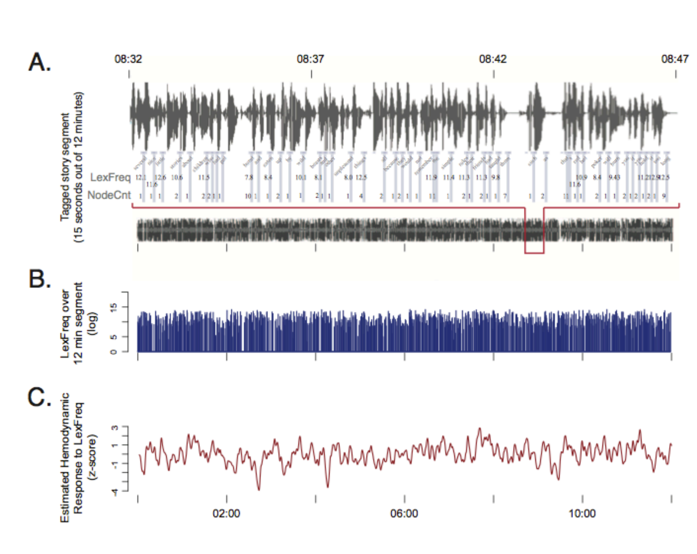
Recognition Complexity

- 30 minute audio text of *Alice in Wonderland* collected by Brennan and colleagues (Brennan 2010)
- BOLD signal measured (fMRI) while subjects listen to story
- Predictors generated from parsing complexity measures

A. Word boundaries annotated with values of predictors

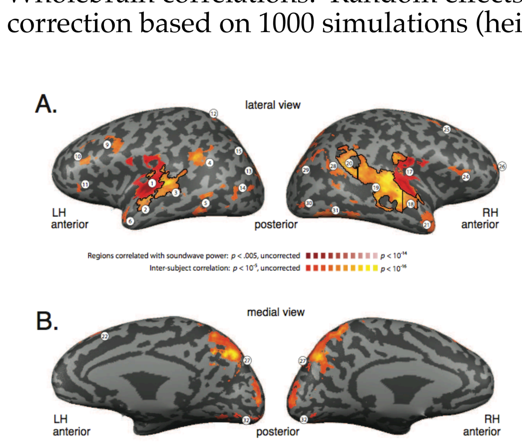
B. Value of predictors assigned to each time point :
Computational Steps
Stack Depth
Lexical Frequency (as measure of lexical access)
Word occurrence (binary)
End of Sentence (binary)

C. Predictors convolved with function describing estimated hemodynamic response (Boynton 1996)



Correlations

Wholebrain correlations: Random effects analysis, corrected for serial correlations, cluster correction based on 1000 simulations (height t(9) = 3.1, cluster size = 21 3x3x4 voxels)



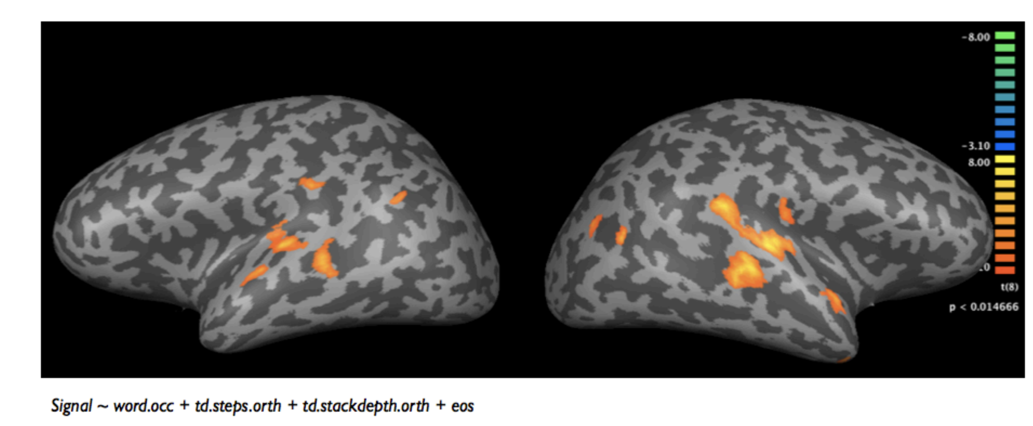
Inter-Subject Correlations (ISC): time course of activity from one subject used to predict activity of other subjects

High correlations between predicted and actual activity assumed to be function of external stimulus

Sound power correlations in audio cortex (red), ISC (orange)

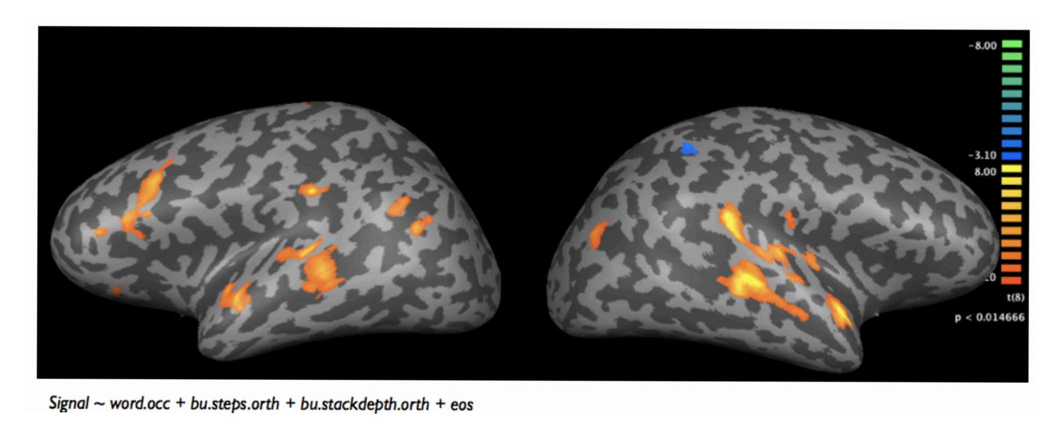
Multiple Regression Analysis

Regions of Activity Highly Correlated with Top-Down Steps



Signal ~ word.occ + td.steps.orth + td.stackdepth.orth + eos

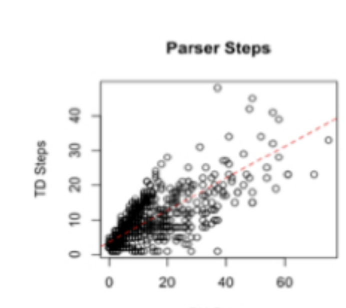
Regions of Activity Highly Correlated with Bottom-Up Steps



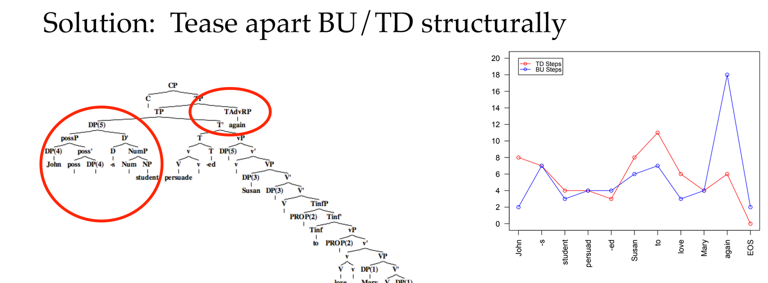
Signal ~ word.occ + bu.steps.orth + bu.stackdepth.orth + eos

Issues

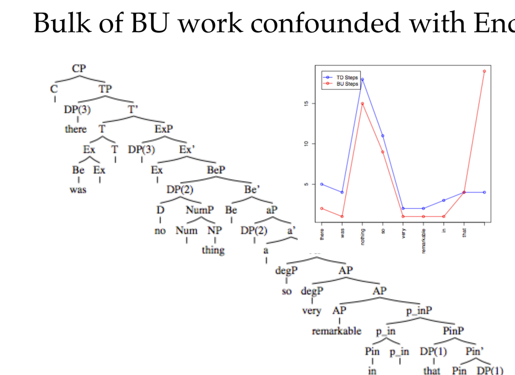
TD and BU Steps are highly correlated with each other, making model comparison difficult



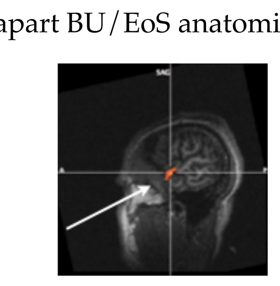
Solution: Tease apart BU/TD structurally



Bulk of BU work confounded with End of Sentence



Solution: Tease apart BU/EoS anatomically



Conclusion

- Proof of Concept: Standard parsing complexity measure over reasonable grammars DOES correlate with neural activity in reasonable brain regions.
- Fruitful, then, to refine grammars and continue to test (Grammar, Parser) pairs against this rich source of data.