Computational Neurolinguistics

Sita

University of Toronto

April 2019

What is Computational Neurolinguistics[1]

Arbib, Caplan - 1979[1]

1. Faculty models: represent entire task oriented process
2. Process models: fractionate psycholinguistic tasks and ascribe the components to particular brain regions
3. Representational models: specific linguistic representations to build a psycholinguistic analysis

Arbib et al. - 2000s [...]

1. Damage to Broca's area and agrammatism
   - Depends on gramatical structure of patient's language.
2. Construction grammar over innate Universal Grammar

Embick, Poeppel - 2015 [4]

Ways in which Computational/representational (CR) theories connect with neurobiological (NB) foundations

1. Correlational (NB correlates with CR)
2. Integrated (NB guides choice of CR)
3. Explanatory (NB explains CR)

---

[1] Caveat emptor...

Outline

1. Brennan, Hale - “Hierarchical structure guides rapid linguistic predictions during naturalistic listening” [2]
2. Lopopolo et al. - “Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain” [7]
3. Hale et al. - “Finding syntax in human encephalography with beam search” [5]

Encephalon = Brain

Brennan, Hale - *Hierarchical structure guides rapid linguistic predictions during naturalistic listening*[2]

Using electroencephalography (EEG) to test whether linguistic predictions reflect hierarchy, not just linear sequences

Tools of the trade

1. EEG: Measures voltage fluctuation across a field of electrodes
   - N400 ERP (Event response profile)
2. fMRI: Based on NMR. Nuclei in perturbed by a weak oscillating magnetic field produce an electromagnetic signal characteristic of the nucleus.
3. $\textit{surprisal} = -\log_2(p(w|C))$, $\textit{perplexity} = 2^{\text{mean}(\text{surprisal}(w_i, i \in [n]))}$

State at time of [2]

1. Prior experiments using reading times show mixed evidence
2. fMRI results give some evidence for hierarchical structures
3. N400 ERP correlates with expectation based on sequential models

Idea: Correlate surprisal estimates (PoS/syntactic surprisal) from different models across EEG signal field.

Models (Unlexical)

1. Trigram model, $C = w_{i-1}, w_{i-2}$
2. SRN [Frank et al], $C = w_{i-j} : j < i$
3. CFG [Hale 2014, Stanford Parser Probabalistic CFG], $C =$ all grammatical structures consistent with word seen so far

+$$ for listening "Alice in Wonderland"

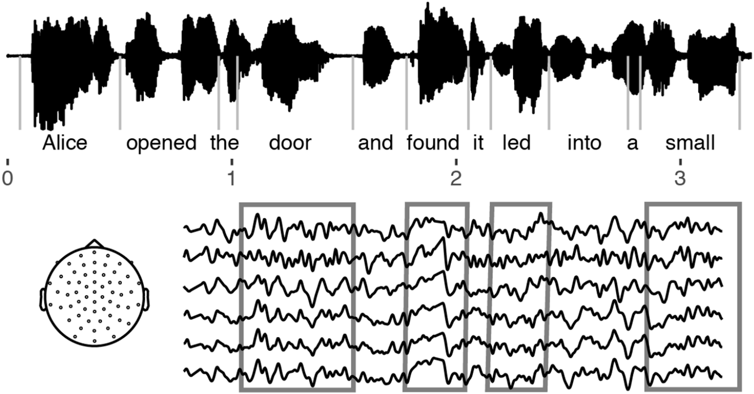

Figure 1: from Brennan and Hale[2]

Statistically

1. CFG has highest mean surprisal, and least amount of information in context

Questions

1. Which signals correlate with surprisal?
2. Which signals are better reflect hierarchical structures?

Analysis

- ▷ Selecting electrodes
  1. $y = \beta X + \epsilon$
  2. Multiple regression with control variables + surprisal
  3. Null model: Rows of design matrix are randomly permuted
- ▷ Measuring model quality on adding hierarchical information[2]
  1. Use Hierarchical Bayesian regression modelling
  2. Widely Applied Information Criterion: Generalization of Akaike information criterion which depends on # parameters and maximum value of likelihood function for the model.

[2] Statistical esoterica

Model comparison

Does CFG improve the fit beyond SRN + Ngram?

▷ Yes: Indicated by falling WAIC

– Over some regions not all, but including regions identified without CFG

– Surprisingly: Need word class (function vs. content) to see this over the baseline model (regression model using SRN + Ngram only).

Similarly...

▷ SRN, Ngram both improve over CFG.

– Word class does not seem to matter?

▷ SRN and Ngram give very similar models so don't improve over each other

Lopopolo et al. - "Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain" [7]

Consider information existing at syntax, lexico-semantic, phonology levels and then map where each generates brain activity.

State at time of [7]

1. Different parts of the brain may encode for different statistical information: entropy, surprisal.
2. There may exist sensory modality independent brain processing. E.g. for auditory and visual stimuli.

Questions

1. Use one measure on different levels of encoding on one text to identify level specific sensitivity.
2. Supramodal processor in brain: Does there exists a region of the brain sensitive to the stochastic measure independent of the level generating it?

Setup

1. Subjects listen to three chosen texts + texts reversed in the MRI scanner.
2. Three levels of information:
   - Syntactic: PoS
   - Phonological: phonemes sequence
   - Lexical: lexical forms
3. Trigram model

Idea

1. fMRI: the blood-oxygen-level dependent (BOLD) contrast
2. Account for haemodynamic response lag
3. Correlate variance in each voxel against the three signals
   - Control: reversed version of the story
4. Low correlation between input streams themselves

Results

1. Identification of different areas of brain that respond to different levels of information
2. No area significantly responsive to all three
3. Phonological stream analysis is less robust

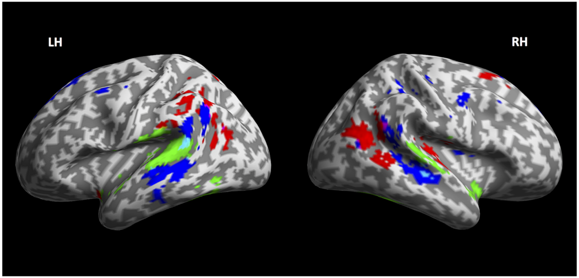

Figure 2: Lexical (green), syntactic (blue) and phonological (red) streams (from Lopopolo et al.[7])

Hale et al. - *Finding syntax in human encephalography with beam search* [5]

Want: cognitively-plausible incremental processing that matches physiological signals

◦ Combine RNN grammars with an incremental parsing algorithm quantify language comprehension difficulty and correlate against EEG

– RNN Grammars are probabilistic models that generate trees. Complexity measures based on “parser state”

RNN Grammars [Dyer et al.[3, 6]]

1. (N, Σ, Θ)
2. Parser: Words $x_i \rightarrow$ parse tree y
   - Stack + input buffer + transition set
3. Generator: Stochastically generates parse trees and terminal symbols
   - Replace reading terminal symbols by GEN operation.
4. RNNG defines a joint distribution on words + trees:

$$p(x, y) = \prod_{1}^{|a(x,y)|} p(a_t | a_{<t})$$

where $a(x, y)$ is parse tree in terms of generation transitions.

Parametrize p(x,y) by r_{a_i}, u_i, b_{a_i} where
$u_i = \tanh(W[\text{algorithm state}] + c)$

$$p(x,y) = \prod_{1}^{|a(x,y)|} \frac{\exp r_{a_t}^T u_t + b_{a_t}}{\sum_{a':\text{valid transition}} \exp r_{a'}^T u_t + b_{a'}}$$

Magic: algorithm state = Stack + buffer + history grows unboundedly but be encoded using an RNN/LSTM

– Train to maximize the likelihood of corpus of parse trees

Beam seach

Mix BFS + DFS: carry along a list of possible parse trees ("beam"). A new input could be incompatible with one but not all.

◦ Naive beam search fails: spend a lot more time exploring structures (e.g. starts new constituents) than words.

– Stern et al[8] Word Synchronous Beam Search: keep searching through structural actions until "enough" high-scoring parser states finally take a lexical action

Analysis

1. Same data as [2]
2. RNNGs were trained to match the output trees provided by the Stanford parser
3. Incremental complexity metrics:
   distance: count of actions required to synchronize k analyses at the next word[3]
   surprisal: log-ratio of summed forward probabilities for analyses in the word beam
4. Regression analysis over all electrodes, various models, varying beam sizes

[3] How many actions till k lexical actions can be executed yield next words?

Results

1. Distance and suprisal measures are reliably associated with electro-physiological signals
   LSTM surprisal is not
2. Distance measure with RNNG$_{-COMP}$ “discovers” P600 ERP

Michael A. Arbib and David Caplan.
Neurolinguistics must be computational.
Behavioral and Brain Sciences, 2(3):449–460, 1979.

Jonathan R. Brennan and John T. Hale.
Hierarchical structure guides rapid linguistic predictions during naturalistic listening.
PLOS ONE, 14(1):1–17, 01 2019.

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith.
Recurrent neural network grammars.
CoRR, abs/1602.07776, 2016.

David Embick and David Poeppel.
Towards a computational(ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics.
Language, Cognition and Neuroscience, 30(4):357–366, 4 2015.

John Hale, Chris Dyer, Adhiguna Kuncoro, and Jonathan Brennan.
Finding syntax in human encephalography with beam search.
In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 2726–2735, 2018.

Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A. Smith.
What do recurrent neural network grammars learn about syntax?
CoRR, abs/1611.05774, 2016.

Alessandro Lopopolo, Stefan L. Frank, Antal van den Bosch, and Roel M. Willems.
Using stochastic language models (slm) to map lexical, syntactic, and phonological information processing in the brain.
PLOS ONE, 12(5):1–18, 05 2017.

Mitchell Stern, Daniel Fried, and Dan Klein.
Effective inference for generative neural parsing.
CoRR, abs/1707.08976, 2017.