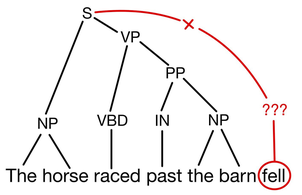
Particle Filtering with Neural Language Models: Modelling the Effects of Memory on Incremental Sentence Processing

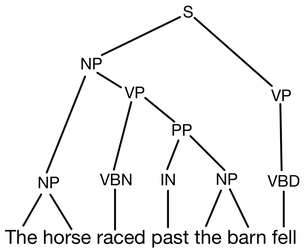
Sarah Payne[1], Peng Qian[2], Ethan Wilcox[2,3], Roger Levy[2]

[1] Departments of Linguistics and Computer & Information Science, University of Pennsylvania, [2] Department of Brain & Cognitive Science and Center for Brains, Minds & Machines, MIT, [3] Department of Linguistics, Harvard University

Background

- Sentence processing in humans is **incremental and constrained** by memory
- **Language is ambiguous:** in **"garden path" sentences,** a locally likely structural hypothesis becomes implausible due to disambiguating evidence:





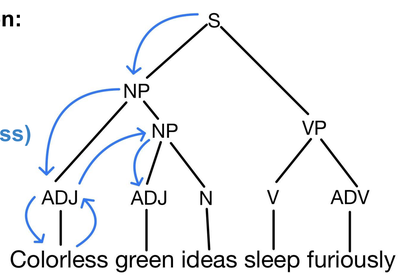
- Incremental processing difficulty can be measured via **eye tracking and maze tasks** in humans
- **Surprisal** (log inverse probability) is used to model processing difficulty, but **underpredicts the magnitude of the garden path effects**

Recurrent Neural Network Grammars (RNNGs)

- **Probabilistic model of generating top-down structural hypotheses** (Dyer et al 2016)

Actions Taken:

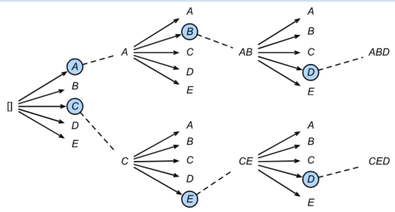
NT(S)
NT(NP)
NT(ADJ)
SHIFT(colorless)
REDUCE
NT(NP)
NT(ADJ)



- Three action types used to create trees:
  - **NT:** open a non-terminal (e.g. NP)
  - **SHIFT:** add the next terminal (i.e. word)
  - **REDUCE:** close the current non-terminal
- We train on **BLLIP** (1.75 million parsed sentences)

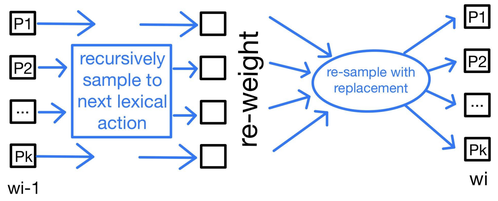
Models of Memory Limitations

We keep ***k*** hypotheses in parallel using **three models of working memory limitations:**



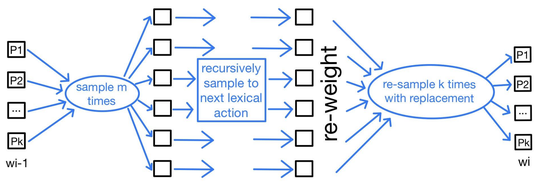
Word-Synchronous Beam Search:

- Recursively enumerate and apply actions until **enough states reach the next SHIFT action.**
- Take the top *k* of the states that reach SHIFT



Particle Filtering:

- Sample *k* times from the action distribution
- Extend each sample to the next lexical action
- Re-weight by probability of the next word given the hypothesized structure and resample *k* times



Particle Filtering with Resampling:

- **Better approximation of the probability distribution** while limiting working memory.
- Sample *m, m* > *k* times from the *k* structures and extend and re-weight each.
- Choose *k* out of *m* structures during resampling.

References: **Dyer,** Chris et al. "Recurrent Neural Network Grammars." NAACL (2016). **Hale**, John et al. "Finding Syntax in Human Encephalography with Beam Search." ACL (2018). **Hu,** Jennifer et al. "A Systematic Assessment of Syntactic Generalization in Neural Language Models." (2020), **Levy**, R. et al. "Modeling the effects of memory on human online sentence processing with particle filters." NIPS (2008). **Wilcox,** Ethan Gotlieb et al. "A Targeted Assessment of Incremental Processing in Neural Language Models and Humans." ACL (2021).

Acknowledgements: SP is grateful for funding by the Center for Brains, Minds and Machines via NSF STC award CCF-1231216

Main Verb-Reduced Relative (MV/RR) Garden Paths

- Cause garden paths by **leading the reader to interpret the start of a relative clause as a main verb.** We manipulate 2 conditions:

1. Ambiguity of the verb:

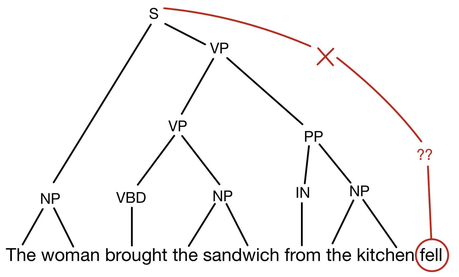
"The woman **brought** the sandwich from the kitchen fell"

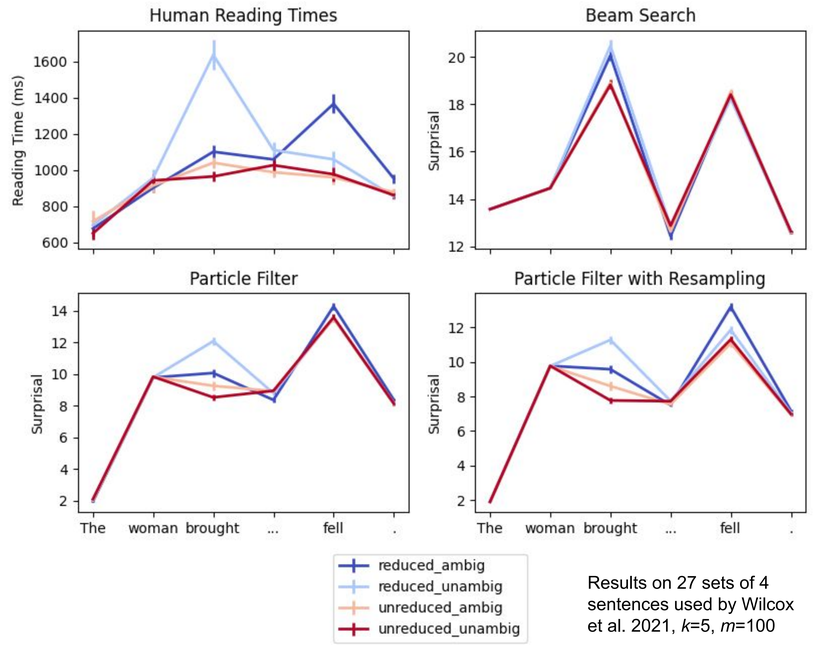
"The woman **given** the sandwich from the kitchen fell"

2. Reduction of the relative clause:

"The woman **brought** the sandwich from the kitchen fell"

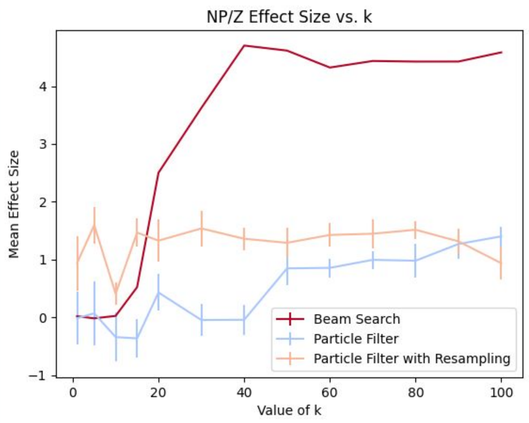
"The woman **who was brought** the sandwich from the kitchen fell"





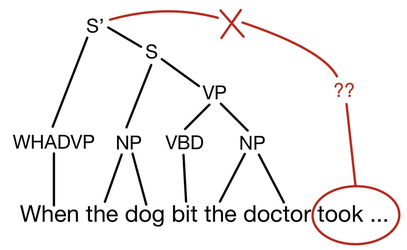
Results on 27 sets of 4 sentences used by Wilcox et al. 2021, *k*=5, *m*=100

Noun Phrase-Zero (NP/Z) Garden Paths



Results on 24 sets of 4 sentences from Hu et al. 2020, *m*=100

- Cause garden paths by **leading the reader to interpret the subject of the second clause as the object of the first clause.** We manipulate 2 conditions:



1. Transitivity of the verb:

"When the dog **bit** the doctor took off the restraint"

"When the dog **struggled** the doctor took off the restraint"

2. Comma between clauses:

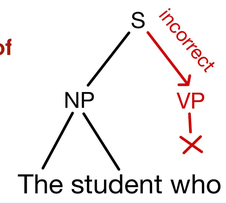
"When the dog **bit** the doctor took off the restraint"

"When the dog **bit,** the doctor took off the restraint"

- In humans, the difference in surprisal between comma & no-comma is larger for transitive than intransitive verbs.
  - We measure effect size as transitive difference - intransitive difference

Discussion & Future Directions

- **For smaller values of *k*, a better approximation of the action distribution yields larger garden path effects.**
  - Particle filtering with resampling combines small *k* and accurate approximation



- If the model makes an incorrect top-down prediction, it **cannot recover when it encounters the next word**.
  - Future work: explore **other parsing orders**, such as left corner