# Incremental Parsing for Semantically-Sensitive Psycholinguistic Predictors

Byung-Doh Oh

Department of Linguistics
The Ohio State University

April 17, 2020
OSU Linguistics Colloquiumfest

# Motivation

Patterns of processing difficulty shed light on the mechanism behind language processing

# Motivation

Patterns of processing difficulty shed light on the mechanism behind language processing

Processing difficulty can be observed in behavioral responses

- Reading times
- Eye movements
- Event-related potentials

# Motivation

Patterns of processing difficulty shed light on the mechanism behind language processing

Processing difficulty can be observed in behavioral responses

- Reading times
- Eye movements
- Event-related potentials

Statistical modeling approaches try to account for these dependent variables by regressing various predictors

# Motivation

Expectation-based theories of sentence processing

- Processing difficulty is determined by predictability in context

# Motivation

Expectation-based theories of sentence processing

- Processing difficulty is determined by predictability in context
- Can be quantified via *surprisal* (Shannon, 1948)

# Motivation

Expectation-based theories of sentence processing

- Processing difficulty is determined by predictability in context
- Can be quantified via *surprisal* (Shannon, 1948)

## Surprisal

$$S(w_i) \stackrel{\text{def}}{=} -\log P(w_i | w_1, w_2, ..., w_{i-1})$$

# Motivation

Expectation-based theories of sentence processing

- Processing difficulty is determined by predictability in context
- Can be quantified via *surprisal* (Shannon, 1948)

## Surprisal

$$S(w_i) \stackrel{\text{def}}{=} -\log P(w_i|w_1, w_2, ..., w_{i-1})$$

- Can be calculated from any probability model over words

# Motivation

Expectation-based theories of sentence processing

- Processing difficulty is determined by predictability in context
- Can be quantified via *surprisal* (Shannon, 1948)

## Surprisal

$$S(w_i) \stackrel{\text{def}}{=} -\log P(w_i | w_1, w_2, ..., w_{i-1})$$

- Can be calculated from any probability model over words
- Empirical support for surprisal based on n-gram, probabilistic context-free grammar (PCFG), and long short-term memory (LSTM) (Goodkind & Bicknell, 2018; Hale, 2001; R. Levy, 2008; Smith & Levy, 2013)

# Motivation

Open question as to how to best estimate the *human language comprehender's* probability model

# Motivation

Open question as to how to best estimate the *human language comprehender's* probability model

In other words, what should we consider in our estimate of predictability?

# Motivation

Open question as to how to best estimate the *human language comprehender's* probability model

In other words, what should we consider in our estimate of predictability?

- n-gram surprisal: sequence of previous $n-1$ words
- PCFG surprisal: syntactic parse of previous words
- LSTM surprisal: sequence of arbitrarily many previous words

# Motivation

Open question as to how to best estimate the *human language comprehender's* probability model

In other words, what should we consider in our estimate of predictability?

- n-gram surprisal: sequence of previous $n - 1$ words
- PCFG surprisal: syntactic parse of previous words
- LSTM surprisal: sequence of arbitrarily many previous words

This work presents a parser that incorporates both *syntactic structure* and *propositional content* in estimating the predictability of a given word

# Motivation

Open question as to how to best estimate the *human language comprehender's* probability model

In other words, what should we consider in our estimate of predictability?

- n-gram surprisal: sequence of previous $n-1$ words
- PCFG surprisal: syntactic parse of previous words
- LSTM surprisal: sequence of arbitrarily many previous words

This work presents a parser that incorporates both *syntactic structure* and *propositional content* in estimating the predictability of a given word

Explicitly incorporating propositional content into the probability model allows further experiments that manipulate access to this knowledge

# Motivation

Why propositional content?

# Motivation

Why propositional content?

- Comprehension $\supset$ building a coherent mental representation of propositional content (Kintsch, 1988)

# Motivation

Why propositional content?

- Comprehension $\supset$ building a coherent mental representation of propositional content (Kintsch, 1988)
- Propositional content rather than surface form stored during processing (Bransford & Franks, 1971; Jarvella, 1971)

# Motivation

Why propositional content?

- Comprehension $\supset$ building a coherent mental representation of propositional content (Kintsch, 1988)
- Propositional content rather than surface form stored during processing (Bransford & Franks, 1971; Jarvella, 1971)
- Parsing decisions are informed by semantic interpretation (Brown-Schmidt et al., 2002; Tanenhaus et al., 1995)

# Motivation

Why propositional content?

- Comprehension $\supset$ building a coherent mental representation of propositional content (Kintsch, 1988)
- Propositional content rather than surface form stored during processing (Bransford & Franks, 1971; Jarvella, 1971)
- Parsing decisions are informed by semantic interpretation (Brown-Schmidt et al., 2002; Tanenhaus et al., 1995)

How?

- Train a left-corner parser (Johnson-Laird, 1983) to make decisions based on propositional content as well as syntactic structure

# Incorporating Propositional Content

Each node in the parse tree has a *semantic context vector*

(O. Levy & Goldberg, 2014)

# Incorporating Propositional Content

Each node in the parse tree has a *semantic context vector*

(O. Levy & Goldberg, 2014)

- Each element has the form of $predicate_{role}$, representing argument structure (e.g. $pour_2$)

# Incorporating Propositional Content

Each node in the parse tree has a *semantic context vector*

(O. Levy & Goldberg, 2014)

- Each element has the form of $predicate_{role}$, representing argument structure (e.g. $pour_2$)
- Argument structure derived from generalized categorial grammar reannotation (Bach, 1981; Nguyen et al., 2012), which reflects syntactic valency

# Incorporating Propositional Content

Each node in the parse tree has a *semantic context vector* (O. Levy & Goldberg, 2014)

- Each element has the form of $predicate_{role}$, representing argument structure (e.g. $pour_2$)
- Argument structure derived from generalized categorial grammar reannotation (Bach, 1981; Nguyen et al., 2012), which reflects syntactic valency
- Predicates and arguments derived from lemmatized sentences normalized to declarative form

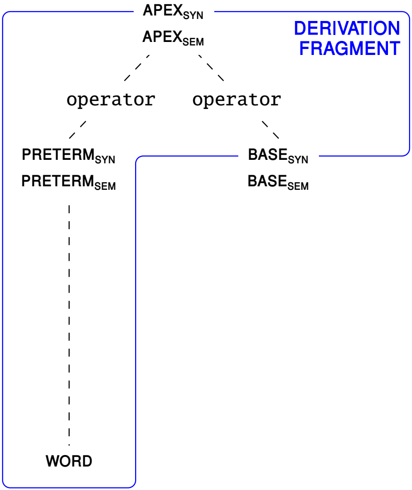
# Incorporating Propositional Content

Each node in the parse tree has a *semantic context vector*
(O. Levy & Goldberg, 2014)

- Each element has the form of $predicate_{role}$, representing argument structure (e.g. $pour_2$)
- Argument structure derived from generalized categorial grammar reannotation (Bach, 1981; Nguyen et al., 2012), which reflects syntactic valency
- Predicates and arguments derived from lemmatized sentences normalized to declarative form

The left-corner parser generates a semantic context vector for each word and propagates it along the parse tree

# Practice Parse: *Horses pull carts*




### Lexical phase

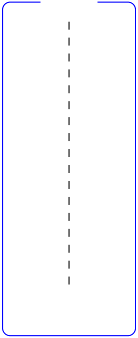
- Attach?
- Preterminal?
- Word?

### Grammatical phase

- Attach?
- Operators?
- Apex?
- Base?

- The parser assumes that a word sequence is generated through these decisions
- For an observed word sequence, the parser returns the sequence of decisions that most likely generated it

# Practice Parse: *Horses pull carts*



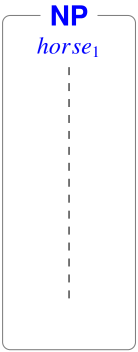
## Lexical phase

- Attach? **No**
- Preterminal?
- Word?

## Grammatical phase

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



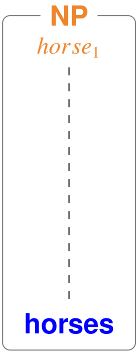

**Lexical phase**

- Attach? **No**
- Preterminal? **NP** $horse_1$
- Word?

**Grammatical phase**

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



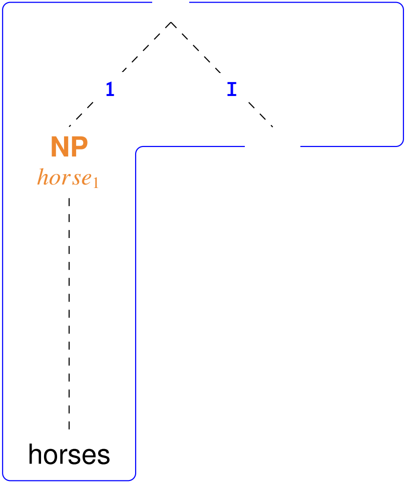

## Lexical phase

- Attach? **No**
- Preterminal? **NP** $horse_1$
- Word? **horses**

## Grammatical phase

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



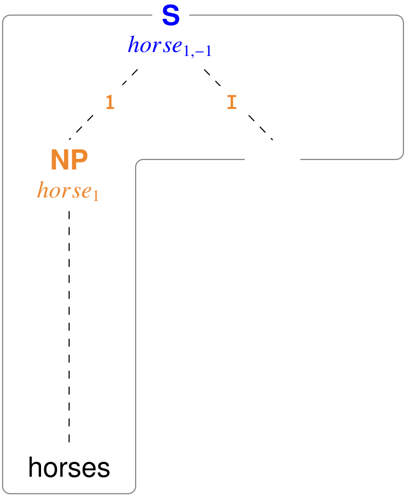

**Lexical phase**

- Attach? **No**
- Preterminal? **NP** $horse_1$
- Word? **horses**

**Grammatical phase**

- Attach? **No**
- Operators? **1 I**
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



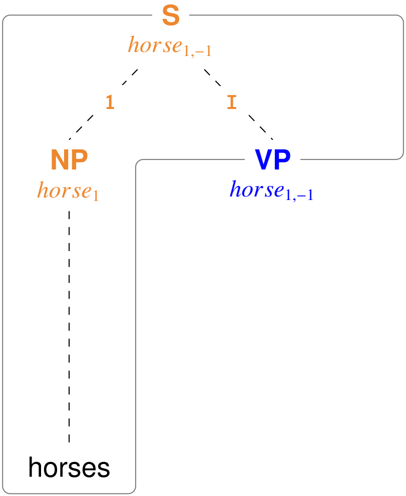

**Lexical phase**

- Attach? **No**
- Preterminal? **NP** $horse_1$
- Word? **horses**

**Grammatical phase**

- Attach? **No**
- Operators? 1 I
- Apex? **S** $horse_{1,-1}$
- Base?

# Practice Parse: *Horses pull carts*



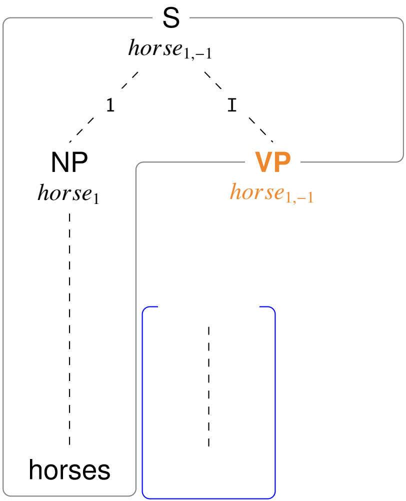

## Lexical phase

- Attach? **No**
- Preterminal? **NP** $horse_1$
- Word? **horses**

## Grammatical phase

- Attach? **No**
- Operators? `1 I`
- Apex? **S** $horse_{1,-1}$
- Base? **VP** $horse_{1,-1}$

# Practice Parse: *Horses pull carts*



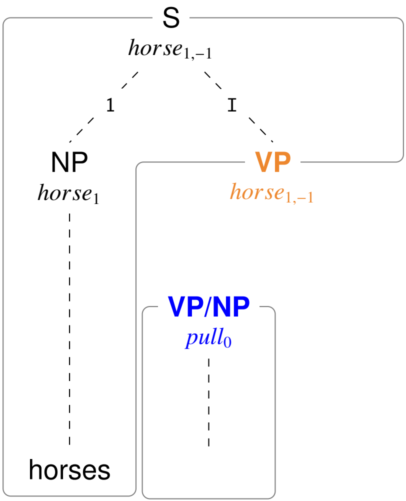

**Lexical phase**

- Attach? **No**
- Preterminal?
- Word?

**Grammatical phase**

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



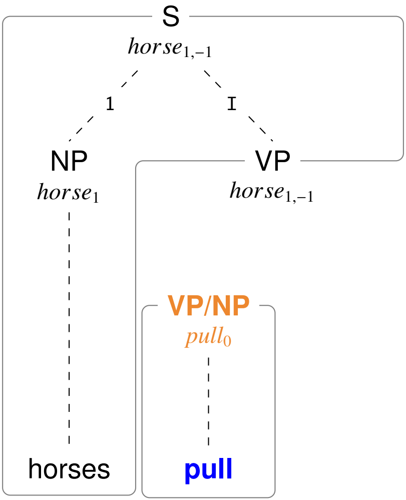

**Lexical phase**

- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word?

**Grammatical phase**

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



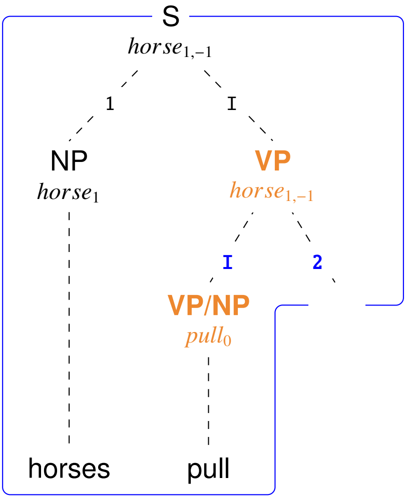

**Lexical phase**

- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word? **pull**

**Grammatical phase**

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



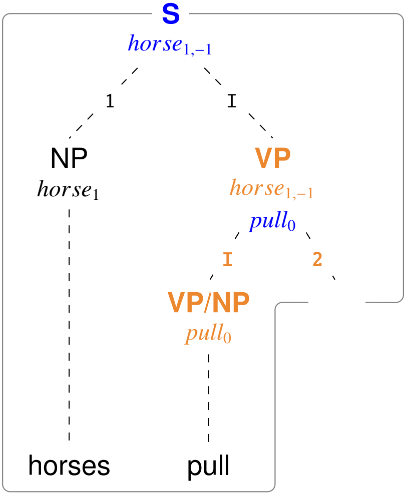

## Lexical phase

- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word? **pull**

## Grammatical phase

- Attach? **Yes**
- Operators? **I 2**
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



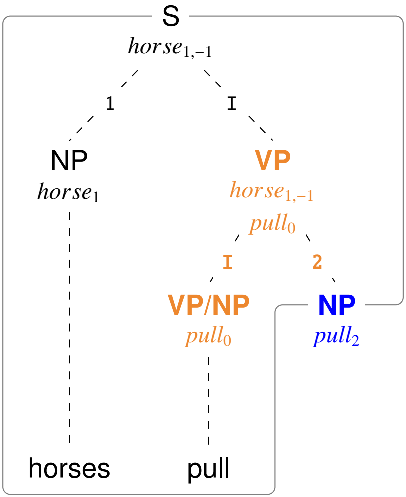

## Lexical phase

- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word? **pull**

## Grammatical phase

- Attach? **Yes**
- Operators? `I` `2`
- Apex? **S** $horse_{1,-1}$ / $pull_0$
- Base?

# Practice Parse: *Horses pull carts*



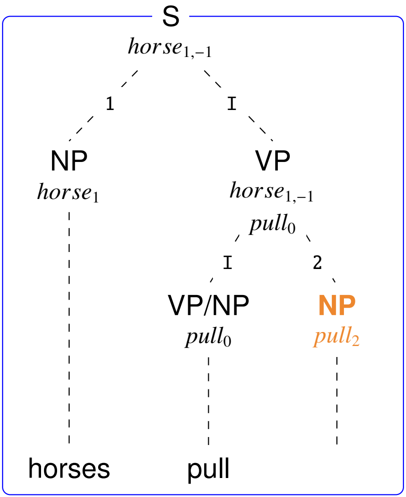

## Lexical phase

- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word? **pull**

## Grammatical phase

- Attach? **Yes**
- Operators? `I 2`
- Apex? **S** $horse_{1,-1}$ / $pull_0$
- Base? **NP** $pull_2$

# Practice Parse: *Horses pull carts*



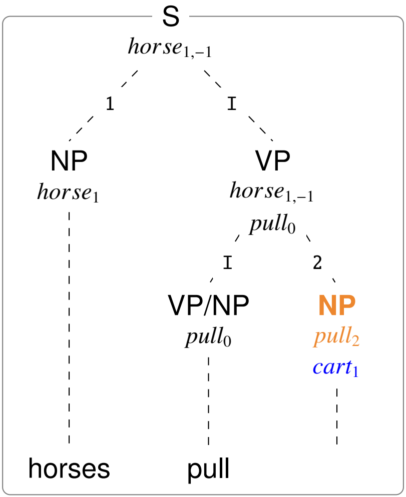

**Lexical phase**

- Attach? **Yes**
- Preterminal?
- Word?

**Grammatical phase**

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*



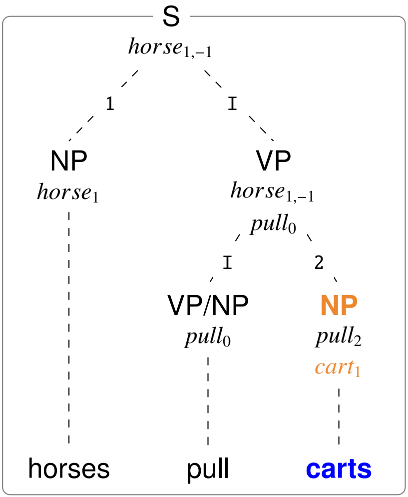

## Lexical phase

- Attach? **Yes**
- Preterminal? **NP** $pull_2$ $cart_1$
- Word?

## Grammatical phase

- Attach?
- Operators?
- Apex?
- Base?

# Practice Parse: *Horses pull carts*




## Lexical phase

- Attach? **Yes**
- Preterminal? **NP** $pull_2$ $cart_1$
- Word? **carts**

## Parse complete

- No derivation fragments
- No more words left to process

# Training Setup

Estimate a probability distribution for each individual parsing decision

# Training Setup

Estimate a probability distribution for each individual parsing decision

Data: WSJ02-21 (Marcus et al., 1993)

- 39,832 sentences
- 950,028 words
- Reannotated to generalized categorial grammar

# Linguistic Evaluation

How well does the parser recover syntactic constituents?

# Linguistic Evaluation

How well does the parser recover syntactic constituents?

Evaluation on: WSJ22 (Marcus et al., 1993)

- 1,700 sentences
- 40,118 words
- Metric: bracketing F1 score (sentences with $<$40 words)

# Linguistic Evaluation

How well does the parser recover syntactic constituents?

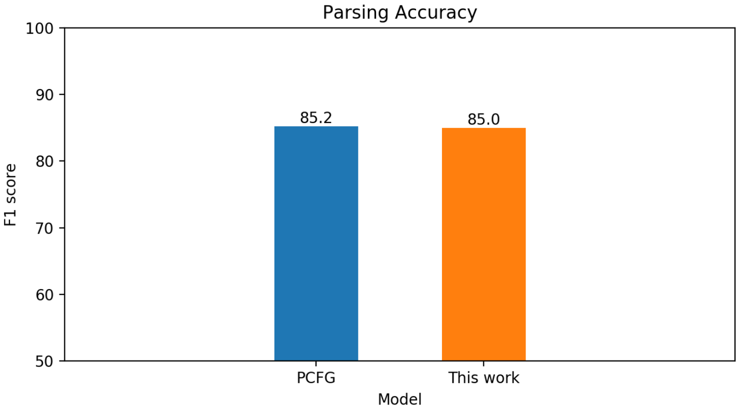
Evaluation on: WSJ22 (Marcus et al., 1993)

- 1,700 sentences
- 40,118 words
- Metric: bracketing F1 score (sentences with <40 words)

Comparing to: "PCFG" model (van Schijndel et al., 2013)

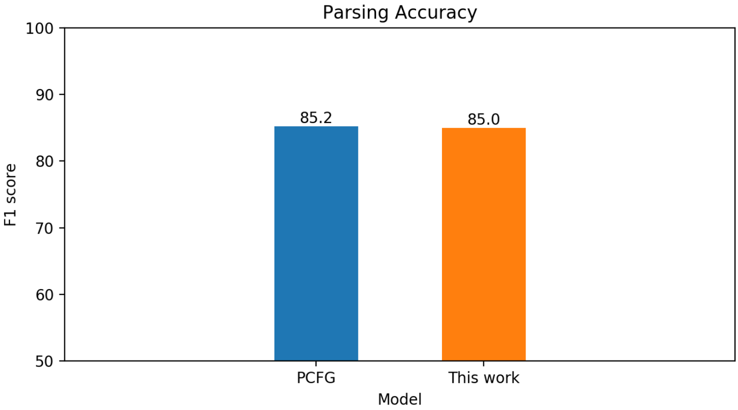
- PCFG-based incremental parser
- Learns to sub-divide each syntactic category according to distributional similarity (Petrov et al., 2006)

# Results



Parsing Accuracy
100
90
80
70
60
50
F1 score
85.2
85.0
PCFG
This work
Model

# Results




- Comparable performance indicates it is a reasonable model of syntactic parsing

# Linguistic Evaluation

Does the incorporation of propositional content influence parsing accuracy?

# Linguistic Evaluation

Does the incorporation of propositional content influence parsing accuracy?

Evaluation on: WSJ22 (Marcus et al., 1993)

- 1,700 sentences
- 40,118 words
- Metric: bracketing F1 score (sentences with $<$40 words)

# Linguistic Evaluation

Does the incorporation of propositional content influence parsing accuracy?

Evaluation on: WSJ22 (Marcus et al., 1993)

- 1,700 sentences
- 40,118 words
- Metric: bracketing F1 score (sentences with $<$40 words)

Comparing to: “semantically-ablated” model

- Train the same parser to make decisions without depending on propositional content of ongoing parse
- Allows us to isolate the contribution of propositional content

# Ablation of Propositional Content

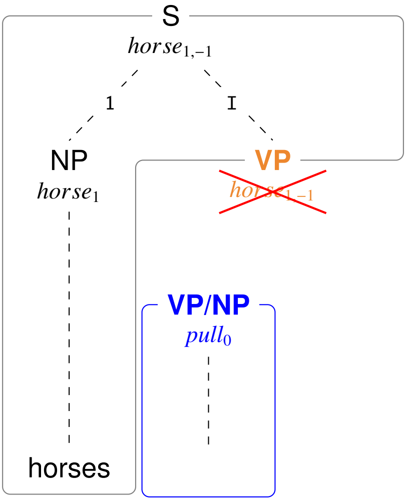
- Structural decisions depend only on syntactic category

# Ablation of Propositional Content

- Structural decisions depend only on syntactic category
- Semantic context vectors still generated as part of word generation

# Ablation of Propositional Content

- Structural decisions depend only on syntactic category
- Semantic context vectors still generated as part of word generation




**Lexical phase**

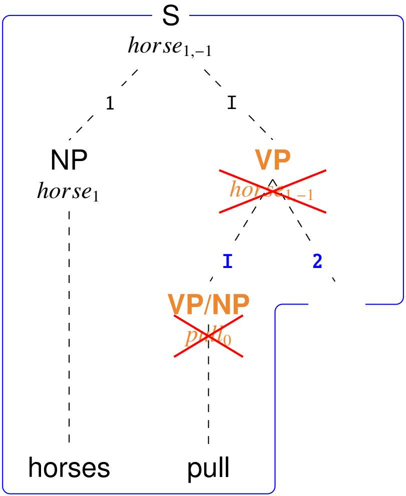
- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word?

**Grammatical phase**

- Attach?
- Operators?
- Apex?
- Base?

# Ablation of Propositional Content

- Structural decisions depend only on syntactic category
- Semantic context vectors still generated as part of word generation



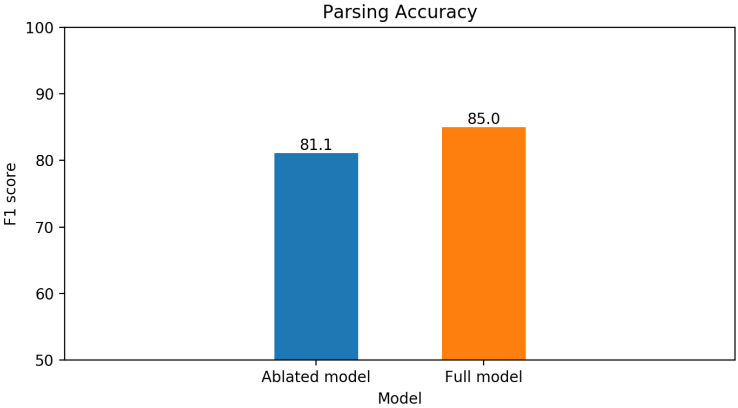

### Lexical phase

- Attach? **No**
- Preterminal? **VP/NP** $pull_0$
- Word? **pull**

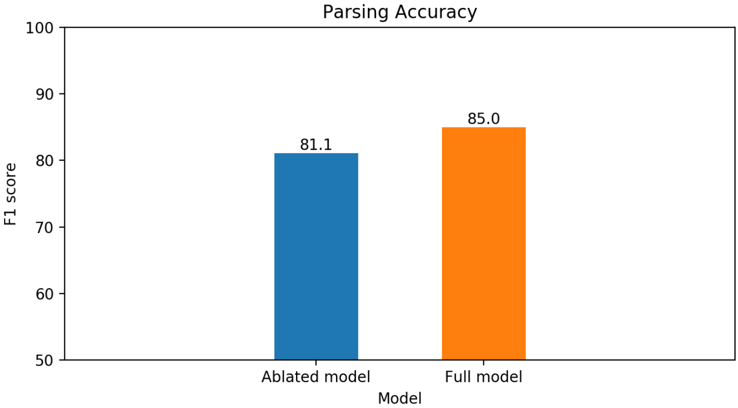
### Grammatical phase

- Attach? **Yes**
- Operators? **I 2**
- Apex?
- Base?

# Results

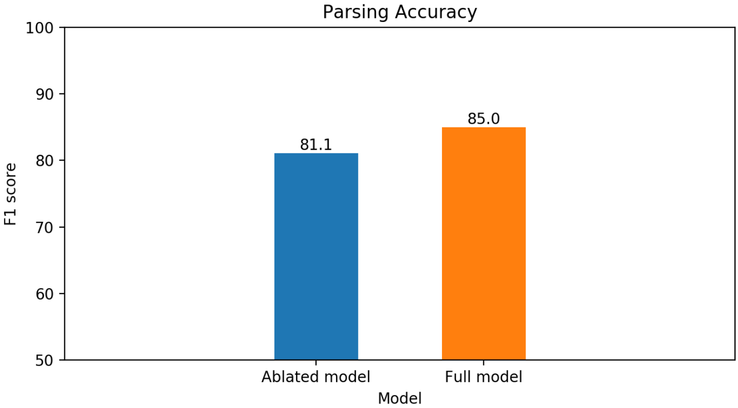

Parsing Accuracy
100
90
80
70
60
50
F1 score
81.1
85.0
Ablated model
Full model
Model

# Results




- Incorporating propositional content contributes to producing more accurate syntactic parses

# Results




- Incorporating propositional content contributes to producing more accurate syntactic parses
- The discrepancy offers a testbed for investigating the role of propositional content in syntactic processing

# Psycholinguistic Evaluation

How well does the surprisal measure from parser explain human behavioral responses?

# Psycholinguistic Evaluation

How well does the surprisal measure from parser explain human behavioral responses?

Evaluation on: Natural Stories Corpus (Futrell et al., 2018)

- Self-paced reading times from 181 participants
- 485 sentences
- 10,245 words

# Psycholinguistic Evaluation

How well does the surprisal measure from parser explain human behavioral responses?

Evaluation on: Natural Stories Corpus (Futrell et al., 2018)

- Self-paced reading times from 181 participants
- 485 sentences
- 10,245 words

Comparing to: “semantically-ablated” model

- Train the same parser to make decisions without depending on propositional content of ongoing parse
- Allows us to isolate the contribution of propositional content

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

Common baseline predictors

- Word length in characters
- Position within sentence

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

Common baseline predictors

- Word length in characters
- Position within sentence

Predictors of interest

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

Common baseline predictors

- Word length in characters
- Position within sentence

Predictors of interest

- Model A: surprisal from ablated model

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

Common baseline predictors

- Word length in characters
- Position within sentence

Predictors of interest

- Model A: surprisal from ablated model
- Model B: surprisal from ablated model, surprisal from full model

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

Common baseline predictors

- Word length in characters
- Position within sentence

Predictors of interest

- Model A: surprisal from ablated model
- Model B: surprisal from ablated model, surprisal from full model
- Both surprisal predictors spilled-over by one position

# Psycholinguistic Evaluation

Fitting two nested linear mixed-effects models

Common baseline predictors

- Word length in characters
- Position within sentence

Predictors of interest

- Model A: surprisal from ablated model
- Model B: surprisal from ablated model, surprisal from full model
- Both surprisal predictors spilled-over by one position

Dependent variable

- Log-transformed self-paced reading times (383,906 data points)

# Results

Comparison of goodness-of-fit (LRT)

| Comparison | $\chi^2$ | df | p-value |
|---|---|---|---|
| Model B over Model A | 13.568 | 1 | **0.00023*** ** |

# Results

Comparison of goodness-of-fit (LRT)

| Comparison | $\chi^2$ | df | p-value |
|---|---|---|---|
| Model B over Model A | 13.568 | 1 | **0.00023*** ** |

- Surprisal from full model makes independent contribution to predicting reading times over surprisal from ablated model

# Results

Comparison of goodness-of-fit (LRT)

| Comparison | $\chi^2$ | df | p-value |
|---|---|---|---|
| Model B over Model A | 13.568 | 1 | **0.00023*** ** |

- Surprisal from full model makes independent contribution to predicting reading times over surprisal from ablated model
- Incorporating propositional content into the probability model results in surprisal measures that are more predictive of human behavioral responses

# Results

Comparison of goodness-of-fit (LRT)

| Comparison | $\chi^2$ | df | p-value |
|---|---|---|---|
| Model B over Model A | 13.568 | 1 | **0.00023*** ** |

- Surprisal from full model makes independent contribution to predicting reading times over surprisal from ablated model
- Incorporating propositional content into the probability model results in surprisal measures that are more predictive of human behavioral responses
- Propositional content that the model has access to can be manipulated to further study its influence on surprisal measures

# Conclusion

- We present an incremental parser that incorporates local propositional content into a probabilistic language model

# Conclusion

- We present an incremental parser that incorporates local propositional content into a probabilistic language model
- As parsing decisions explicitly depend on local propositional content, its contribution to the probability model can be manipulated

# Conclusion

- We present an incremental parser that incorporates local propositional content into a probabilistic language model
- As parsing decisions explicitly depend on local propositional content, its contribution to the probability model can be manipulated
- Analyses show independent contribution of propositional content in producing accurate parses and predicting reading times, suggesting its role in sentence processing

# Future Directions

Experiment using semantically-sensitive predictors

- Look for evidence of memory formation in behavioral measures
- Identify brain responses (e.g. Shain et al., 2019) to semantic processing

# Future Directions

Experiment using semantically-sensitive predictors

- Look for evidence of memory formation in behavioral measures
- Identify brain responses (e.g. Shain et al., 2019) to semantic processing

Further optimize processing model

- Incorporate other psycholinguistic phenomena such as coreference resolution (Jaffe et al., 2018)
- Relax independence assumptions between parsing decisions for higher accuracy

*Thank you for listening!*

Thanks to Clippers, CaCL, and Cory Shain for constructive feedback
Additional thanks to William Schuler for his kind patience

# References I


Bach, E. (1981). Discontinuous constituents in generalized categorial grammars. *Proceedings of the Annual Meeting of the Northeast Linguistic Society (NELS)*, *11*, 1–12.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*, 331–350.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002). Reference resolution in the wild: Online circumscription of referential domains in a natural interactive problem-solving task. In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*.

Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S., & Fedorenko, E. (2018). The Natural Stories Corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*.

Jaffe, E., Shain, C., & Schuler, W. (2018). Coreference and focus in reading times. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*, 409–416.

# References II


Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA, USA, Harvard University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313–330.

Nguyen, L., van Schijndel, M., & Schuler, W. (2012). Accurate unbounded dependency recovery using generalized categorial grammars. In *Proceedings of COLING 2012*.

Petrov, S., Barrett, L., Thibaux, R., & Klein, D. (2006). Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*.

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2019). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

# References III


Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

van Schijndel, M., Exley, A., & Schuler, W. (2013). A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, *5*(3), 522–540.