



Parsing and Scoring English Sentences

Emily Coleman

College of Charleston
Department of Computer Science
Department of Linguistics

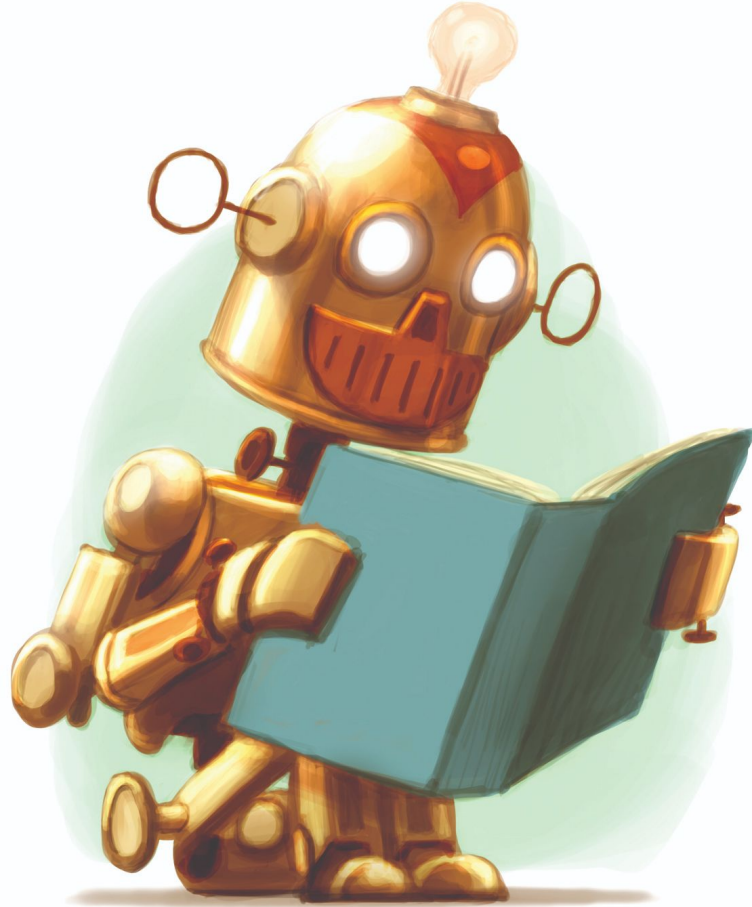
Teaching Grammar to a Computer

How?

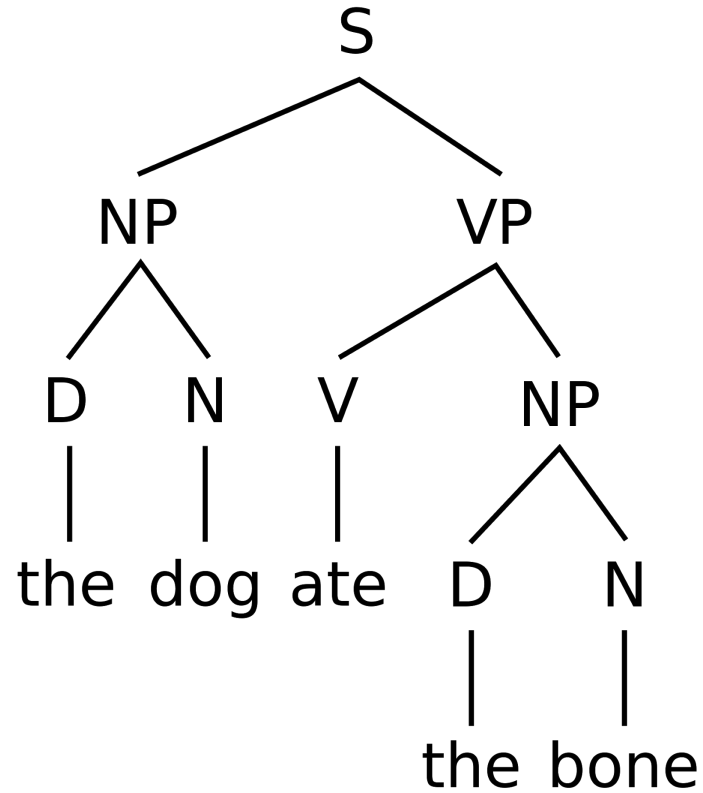
- ◎ Treat English like a formal language

Why?

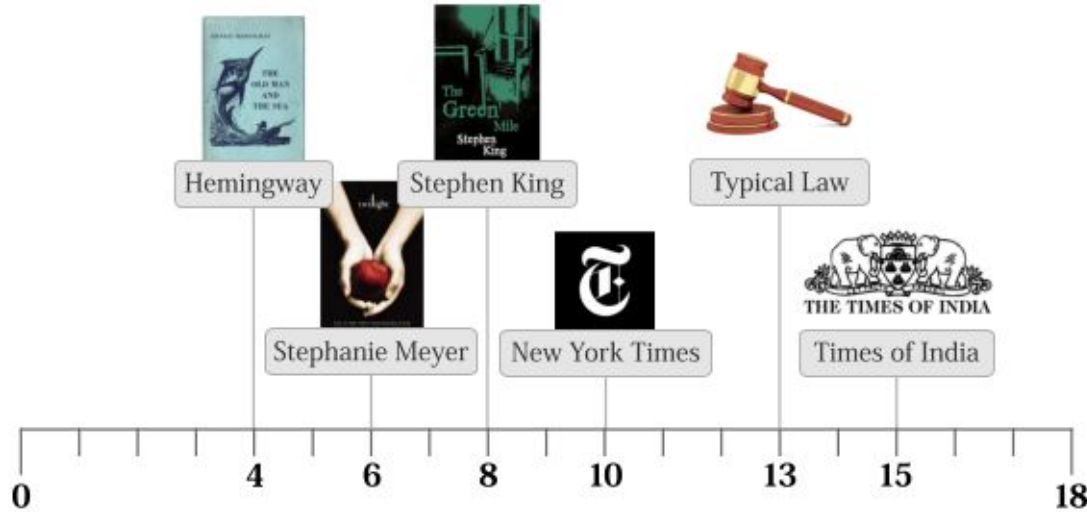
- ◎ Improve how we measure sentence readability



Generate a syntax tree



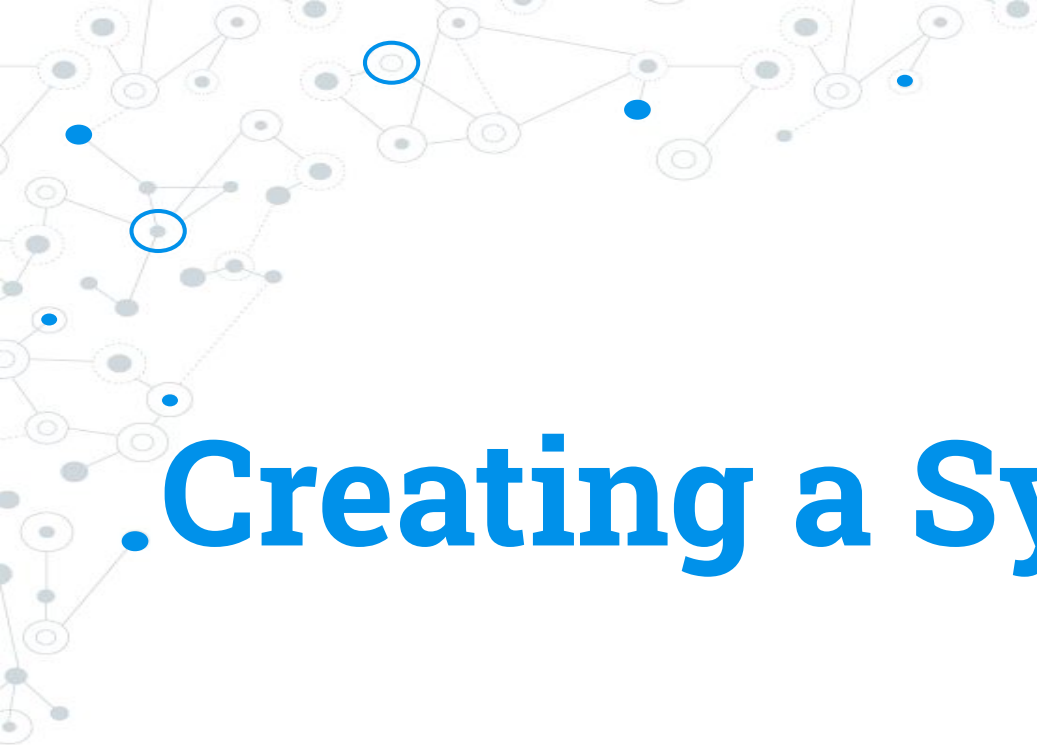
Adjust the readability score



<https://readable.io/content/the-flesch-reading-ease-and-flesch-kincaid-grade-level/>

Flesch Reading Ease Score =

$206.835 - (1.015 \times \text{Average Sentence Length}) - (84.6 \times \text{Average Syllables Per Word})$



.Creating a Syntax Tree



Creating a Syntax Tree

Part 1: Tagging parts of speech

- ◎ The Stanford Natural Language Processing Group
- ◎ Automatic tagging:
 - “I wrote a simple sentence.”
 - “I_PRP wrote_VDB a_DT simple_JJ sentence_NN”



<https://nlp.stanford.edu/static/img/logos/nlp-logo-small.gif>

Creating a Syntax Tree

Part 2: Writing a grammar

What is a formal language?

- ⦿ Consistent rules
- ⦿ Unambiguous

What are some examples?

- ⦿ Mathematical expressions
- ⦿ Programming languages



Creating a Syntax Tree

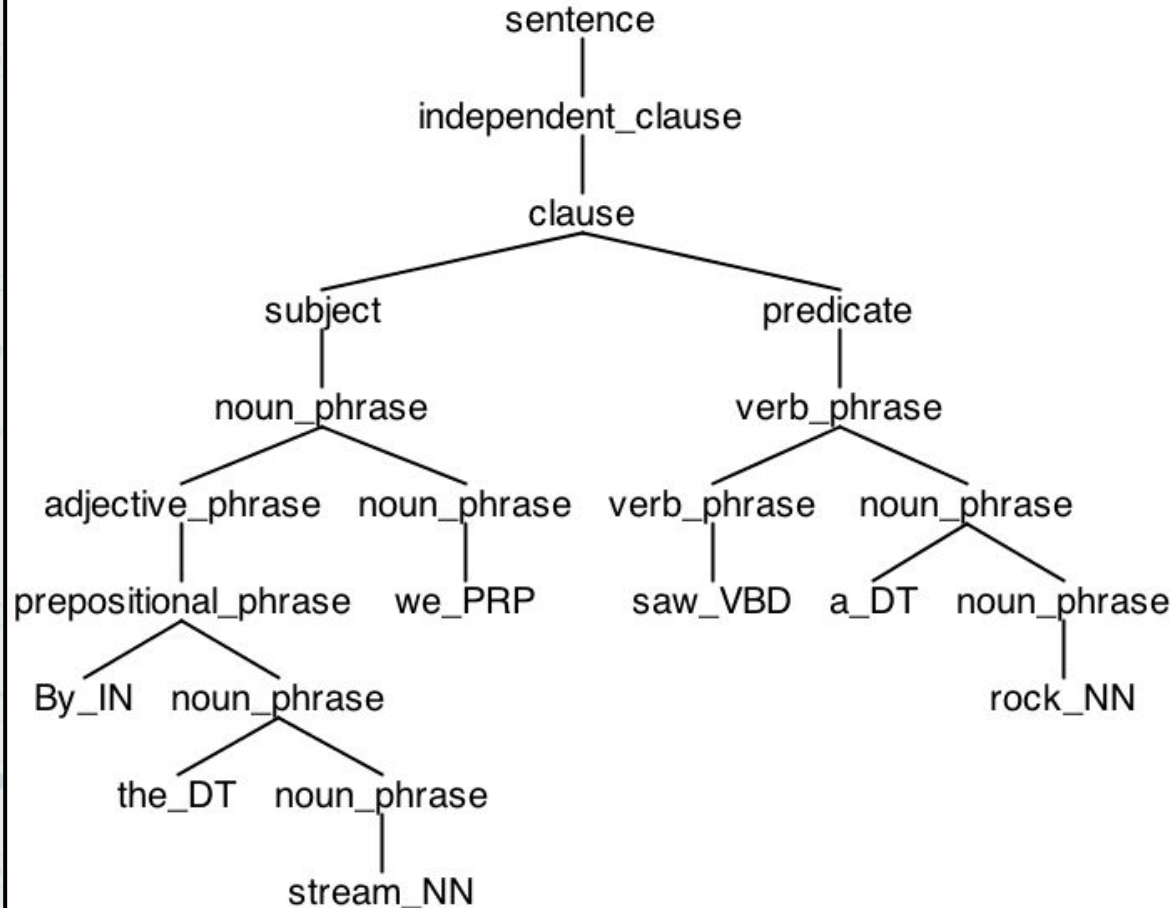
Part 3: Combining part of speech and grammar


“by the stream” \Rightarrow by_PREP the_DT stream_NN

```
PREP : WORD '_PREP';
```

```
prepositional_phrase :  
    PREP noun_phrase
```


By the stream we saw a rock.



A decorative background featuring a network diagram with nodes and connecting lines. The nodes are represented by circles of varying sizes and colors (blue, grey, and white), and the lines are thin and grey. The network is spread across the top-left and bottom-right corners of the slide.

Improving Readability Scores

Improving Readability Scores

Part 1: Score each independent clause

“The doorbell rang and I stood up to answer it.” **Flesch Score = 95.17**



“The doorbell rang.” “I stood up to answer it.” **Adjusted Score = 97.52**

Improving Readability Scores

Part 2: Penalize the number of dependent clauses

“The child who fell when she sneezed stood back up.”

Flesch Score = 103.63

Adjusted Score = 92.63

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with blue dots.

Conclusions

A decorative network diagram in the bottom-right corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with blue dots.




Conclusions: Sentence Parsing

Achievements

- ◎ All ten basic sentence patterns
- ◎ Any number of independent or dependent clauses
- ◎ Some custom part-of-speech tagging

Future Improvements

- ◎ Inverted syntax
 - ◎ Implied referents
 - ◎ Ambiguous phrases
- 




Conclusions: Readability Score

Achievements

- ⊙ Allow for stylistic compound sentences
- ⊙ Adjust score for structural complexity

Future Improvements

- ⊙ Only adjust for two variables
 - ⊙ Formula still untested
- 

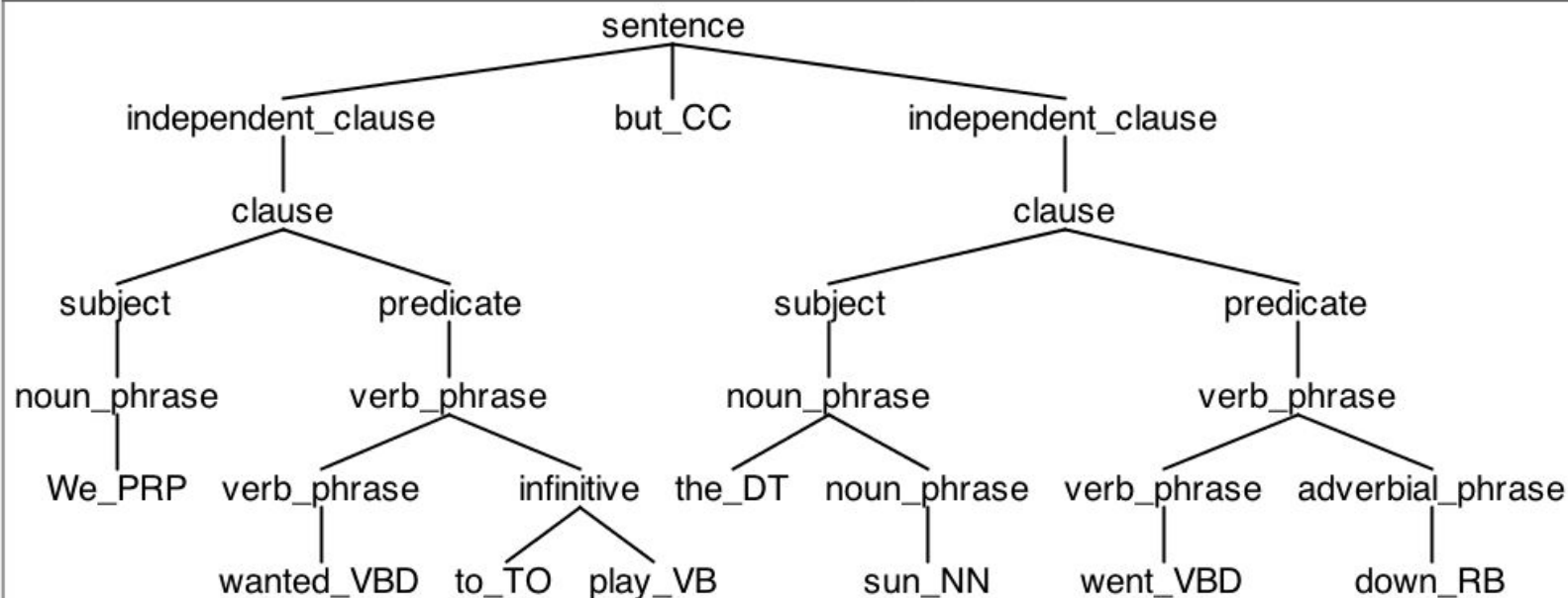
We wanted to play but the sun went down.



"We wanted to play but the sun went down"

Original Flesch Readability Ease Score: 103.70

Adjusted readability score: 108.60



Acknowledgements

- ◎ The Department of Computer Science
 - Dr. George Pothering
- ◎ The Department of Linguistics
 - Dr. Ricard Viñas de Puig



COLLEGE *of*
CHARLESTON

COMPUTER SCIENCE