

# spaCy

Candace Edwards  
ICS 691C  
Presented: 2/2/24

# NLP Libraries/Tools

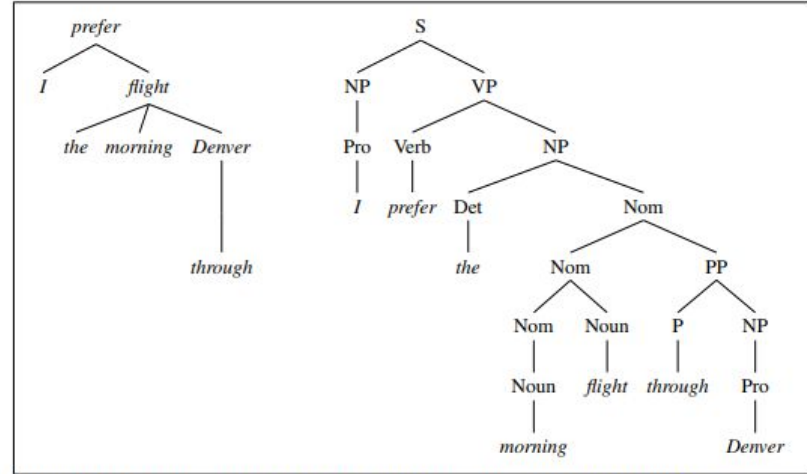
- spaCy is a free open-source library for [Natural Language Processing](#) in Python. It features NER, POS tagging, dependency parsing, word vectors and more.
- Other popular libraries:
  - NLTK: Natural Language Toolkit

SYSTEM	ABSOLUTE (MS PER DOC)			RELATIVE (TO SPACY)		
	TOKENIZE	TAG	PARSE	TOKENIZE	TAG	PARSE
spaCy	0.2ms	1ms	19ms	1x	1x	1x
CoreNLP	0.18ms	10ms	49ms	0.9x	10x	2.6x
ZPar	1ms	8ms	850ms	5x	8x	44.7x
NLTK	4ms	443ms	n/a	20x	443x	n/a

Source: [Medium](#)

# spaCy: Dependency Parsing and NER

- Explore spaCy library and documentation
- Perform a task using tool:
  - **Name Entity Recognition (NER)**  
An information extraction task that seeks to locate and classify named entities mentioned in unstructured text into predefined categories such as person names, organizations, locations...



**Figure 18.1** Dependency and constituent analyses for *I prefer the morning flight through Denver*.

# spaCy Demo + Documentation

## Assigned Attributes

- [Notebook](#)

Dependency predictions are assigned to the `Token.dep` and `Token.head` fields. Beside the dependencies themselves, the parser decides sentence boundaries, which are saved in `Token.is_sent_start` and accessible via `Doc.sents`.

LOCATION	VALUE
<code>Token.dep</code>	The type of dependency relation (hash). TYPE: <code>int</code>
<code>Token.dep_</code>	The type of dependency relation. TYPE: <code>str</code>
<code>Token.head</code>	The syntactic parent, or "governor", of this token. TYPE: <code>Token</code>
<code>Token.is_sent_start</code>	A boolean value indicating whether the token starts a sentence. After the parser runs this will be <code>True</code> or <code>False</code> for all tokens. TYPE: <code>bool</code>
<code>Doc.sents</code>	An iterator over sentences in the <code>Doc</code> , determined by <code>Token.is_sent_start</code> values. TYPE: <code>Iterator[Span]</code>

# spaCy Visualizer: displaCy

- Viz Demo: <https://demos.explosion.ai/displacy>

