

# Labeling the Semantic Roles of Commas

Naveen Arivazhagan, Christos Christodoulopoulos, Dan Roth



## Problem Description

### Motivation:

Commas play an important role in our understanding of text. They often **broadcast a relation** between the segments of the sentence they divide. This can be used as a **cue to extract information or relations** from the corresponding segments.

### Task Definition:

**Label commas** with functional tags that can help extract relations and information from text

## A Corpus of Comma Relations

Refined the corpus created by Srikumar et al. by adding more labels and getting broader coverage

**#Labels:** 4 → 8

**coverage:** 57% → 99%

**#Commas:** 1928

**#Sentences:** 1000

### Comma Classes:

- **Substitute (15%)**  
Mary , the ML scientist , was having a nice day.
- **Locative (5%)**  
She worked at Google , New York.
- **List (17%)**  
She was carrying her bag , her jacket , and her phone
- **Attribute (20%)**  
Her friend Ava , also a scientist , was walking home.
- **Introductory (17%)**  
Without thinking , Mary ran down the street.
- **Complementary (9%)**  
She ran down the street , yelling at Susan to stop.
- **Interrupter (11%)**  
Susan , finally , stopped.
- **Quotation (6%)**  
"Sorry, I couldn't hear you" , she said.
- **Other (1%)**

## Approach

### Classifiers:

- **Independent Classifier**
- **Sequence Classifier:**  
HMM-like approach to label a sequence of commas. Better performance when sequence was only those commas that shared a parent rather than all the commas in a sentence.
- **Baseline (Bayraktar et al.):**  
Rule-based system mapping from the local constituent parse structure of a comma to a comma label

### Results:

Classifier	Accuracy
Independent	<b>83.6</b>
Sequence	79.8
Baseline	80.5

## Features

### Features:

- **Surrounding Words:** window size 4
- **Constituency Parse:** 1,2-grams of local constituency labels augmented with first child's label
- **POS tags:** 2,3,4,5-grams in window size 10
- **Shallow Parse:** 1,2 grams in window size 2
- **Bayraktar-Label:** rule-based label
- **NER:** NER tags of local chunks

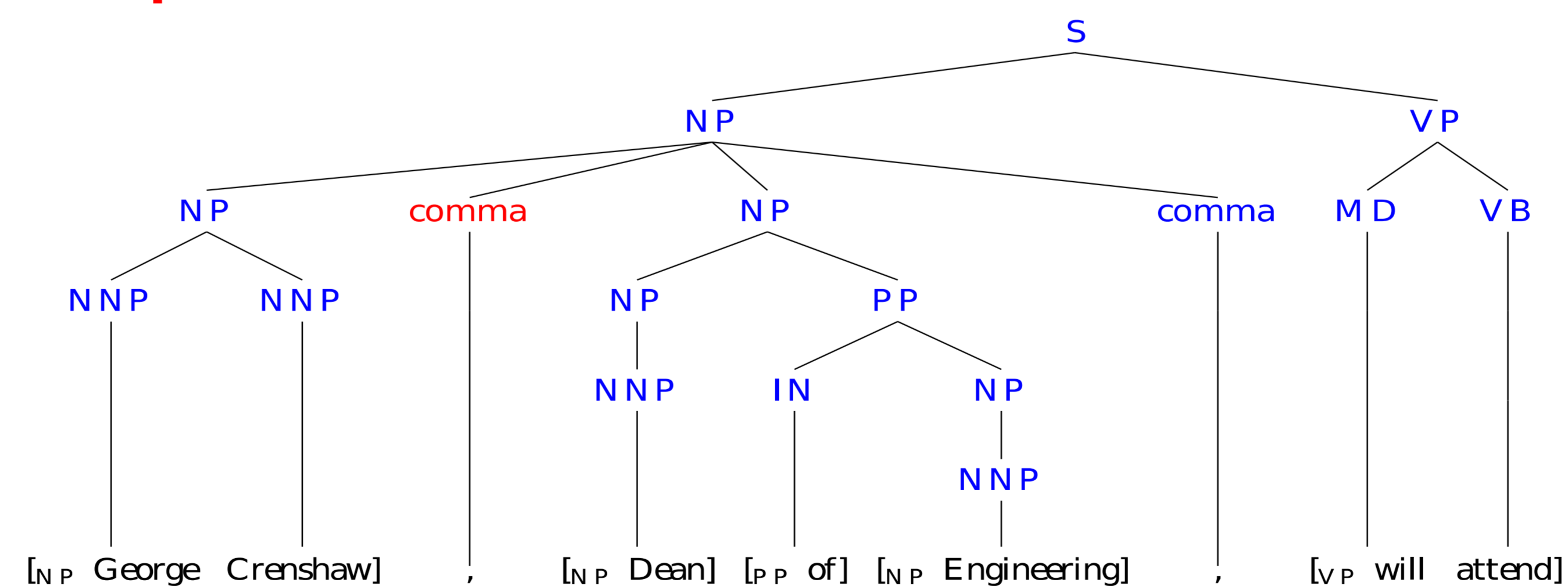
### Ablation:

Features	Accuracy
All	83.6
All - NER	83.4
All - POS	82.3
All - Shallow Parse	83.2
All - Constituency Parse	74.1

## Applications

- **Coreference Resolution:** Systems naively use the existence of a comma separating two mentions as a feature. **Apposition relations** can be better identified using Substitute and Attribute commas.
- **Discourse Parsing: Quotation** can be used for phrase attribution. Introductory is indicative of implicit discourse makers.
- **Parse Reranking:** Commas provide important information regarding **constituent boundaries** and **subordinating relations** which can be used to improve constituent parsers.
- **Textual Entailment:** The type of comma can help us weigh the importance of pieces of information during our analysis. For example, Interrupter commas are **not directly related** to the idea at hand. Similarly, List commas indicate **equal weight** and participation of the entities in an event.

## Example Parse with Features used



**Word:** George, Crenshaw, Dean, of

**Chunk:** NP, NP, PP, null-NP, NP-NP, NP-PP

**Bayraktar-Label:** Substitute

**NER:** PER, ORG

**POS tags:** NNP-NNP, NNP-NNP, NNP-IN...

**Constituent labels:** null-NP+NNP, NP+NNP-NP+NP, NP+NP-comma, NP+NP

## Bibliography

- Bayraktar, M.; Say, B.; and Akman, V. 1998. An analysis of english punctuation: The special case of comma. International Journal of Corpus Linguistics 3(1):33-57.
- Srikumar, V.; Reichart, R.; Sammons, M.; Rappoport, A.; and Roth, D. 2008. Extraction of entailed semantic relations through syntax-based comma resolution. In ACL. Columbus, OH, USA: Association for Computational Linguistics.