# LING/C SC 581:
## Advanced Computational Linguistics

Lecture 15

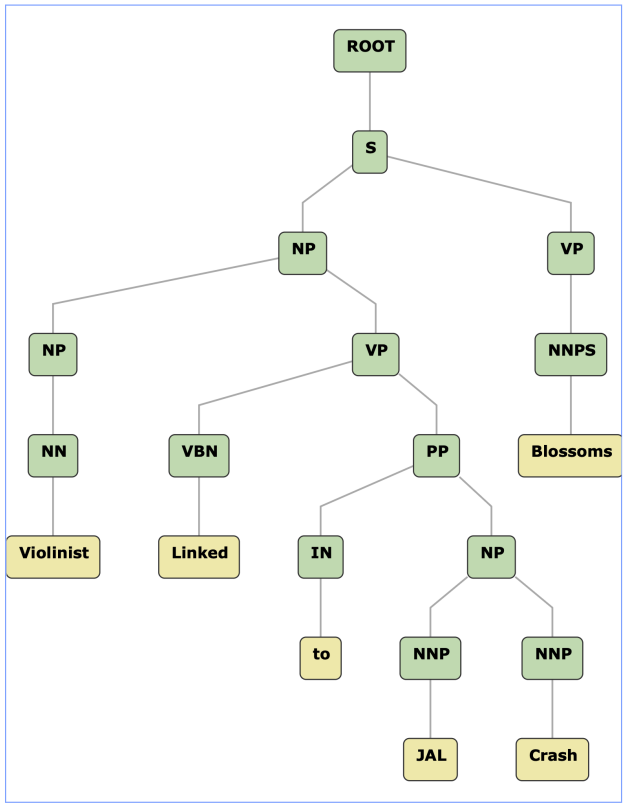
Prof. Sandiway Fong

# Today's Topics

- Install Stanza on your laptop (Homework 8: Part 1)
  - *described last time*
  - *how to use dependency and constituency parsing*
- *Crash Blossoms* (Homework 8: Part 2)

# *Violinist Linked to JAL Crash Blossoms*

Constituency Parse:

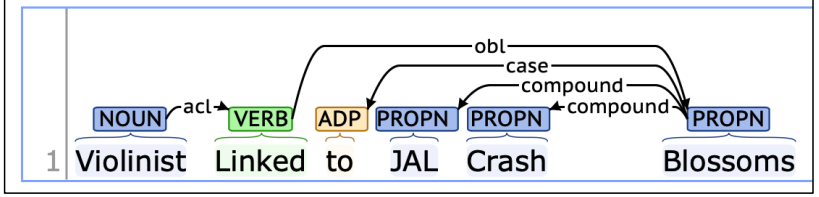



Stanza: two different pipelines, two different "parses" of the same data!

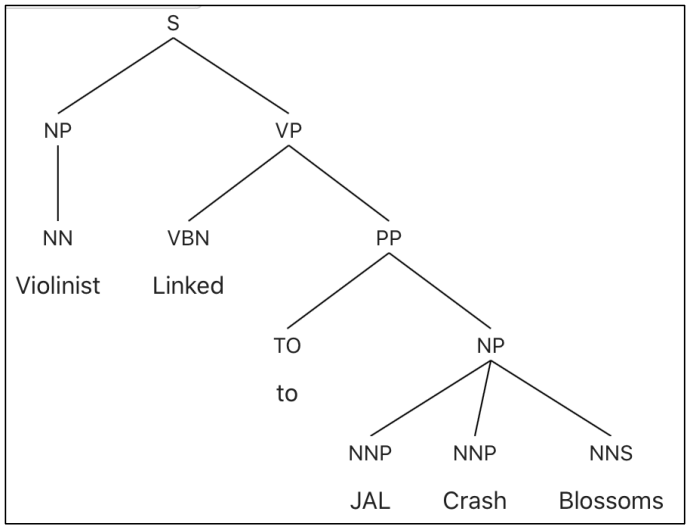
```
>>> nlp = stanza.Pipeline('en')
>>> nlp = stanza.Pipeline(lang='en',
processors='tokenize,pos,constituency')

doc = nlp("Violinist Linked to JAL Crash
Blossoms")
```

*Blossoms* = pl. noun or 3rd pers. sg. present verb

**Universal Dependencies:**

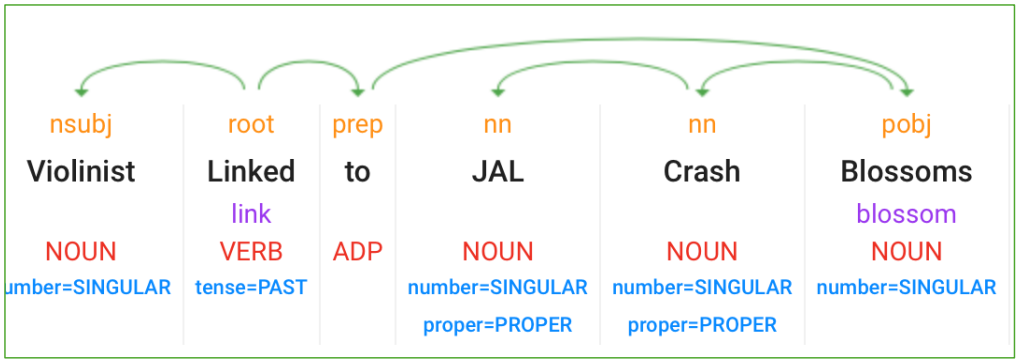


# *Violinist Linked to JAL Crash Blossoms*




Berkeley Neural Parser
https://parser.kitaev.io

Google (*no longer available*)



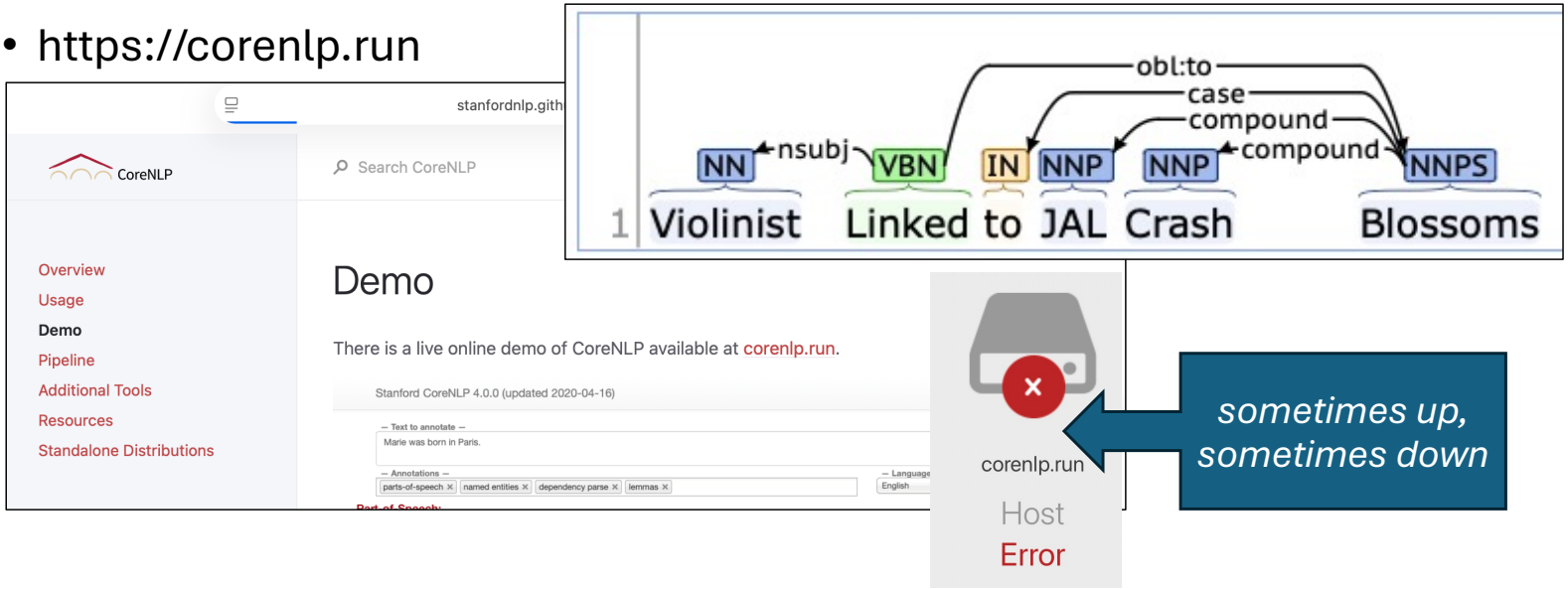

**pobj : nmod**

The `pobj` relation in SD has been renamed to `nmod` in UD.

Note that prepositional phrases are treated differently in UD a
Prepositional Phrases for more details on the new treatment

# CoreNLP

- *Violinist Linked to JAL Crash Blossoms*
- CoreNLP:
  - https://corenlp.run



# Newspapers could check their headlines?

# Was fixed!

- I went to the Daily News website:

## Shaun King column from 2016: North Carolina police kill unarmed deaf man who was using sign language

Aug 22, 2016 at 10:49 am

This is as bad as it gets.

A North Carolina state trooper shot and killed 29-year-old Daniel Harris — who was not only unarmed, but deaf — just feet from his home, over a speeding violation. According to early reports from neighbors who witnessed the shooting this past Thursday night, Harris was shot and killed "almost immediately" after exiting his vehicle.

He appeared to be trying to communicate with the officer via sign language.

# Class Exercise: dependency pipeline

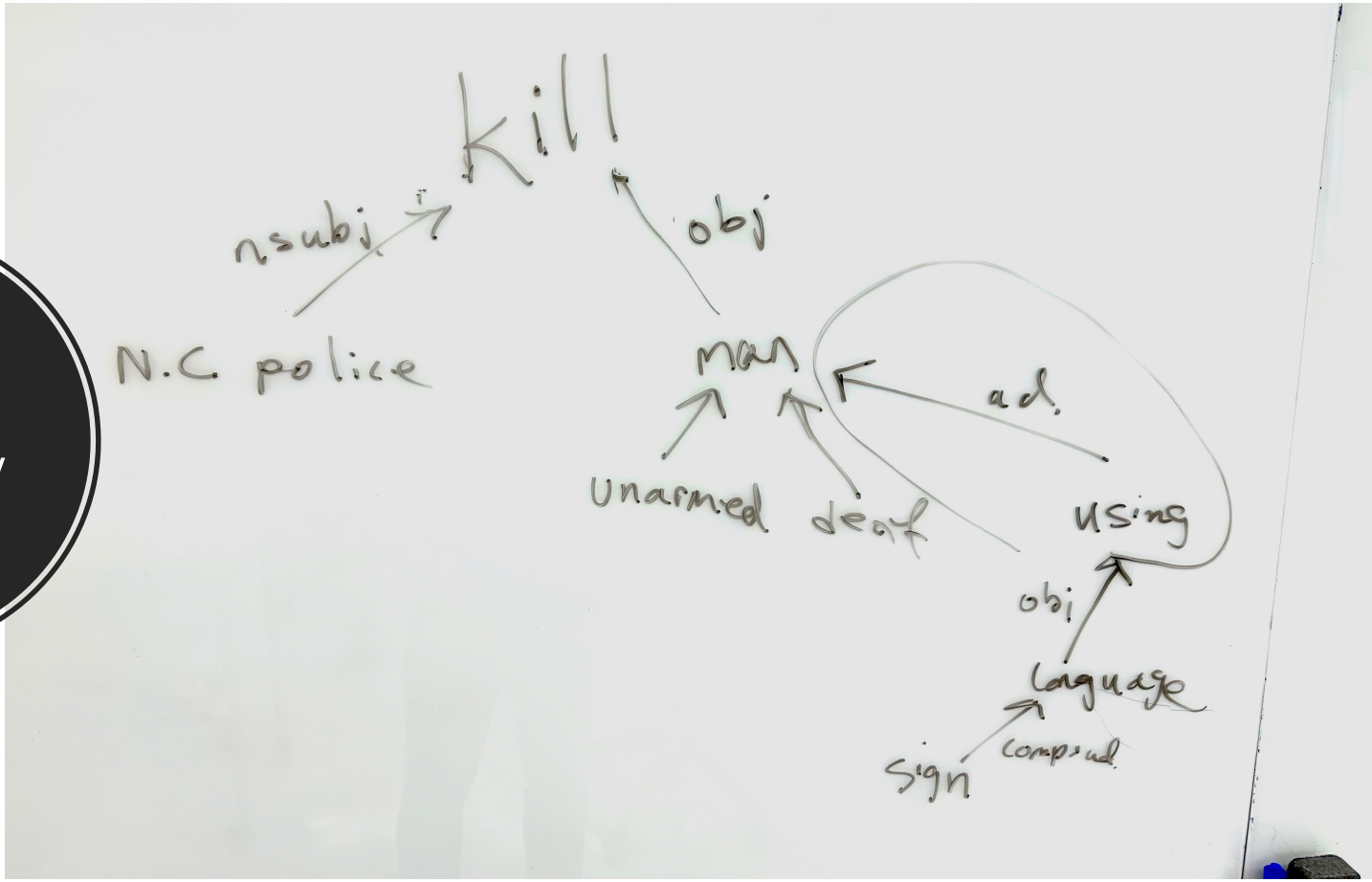
- *N.C. police kill unarmed deaf man using sign language*

```
~$ python3
>>> import stanza
>>> nlp = stanza.Pipeline('en')
>>> doc = nlp('N.C. police kill unarmed
  deaf man using sign language')
>>> for word in doc.sentences[0].words:
...     print(f"{word.id}:
  {word.text}/{word.pos} -
  {word.deprel} -> {word.head}")
...
```

```
1: N.C./PROPN - compound -> 2
2: police/NOUN - nsubj -> 3
3: kill/VERB - root -> 0
4: unarmed/ADJ - amod -> 6
5: deaf/ADJ - amod -> 6
6: man/NOUN - obj -> 3
7: using/VERB - acl -> 6
8: sign/NOUN - compound -> 9
9: language/NOUN - obj -> 7
```

key relation

Class Exercise: dependency pipeline

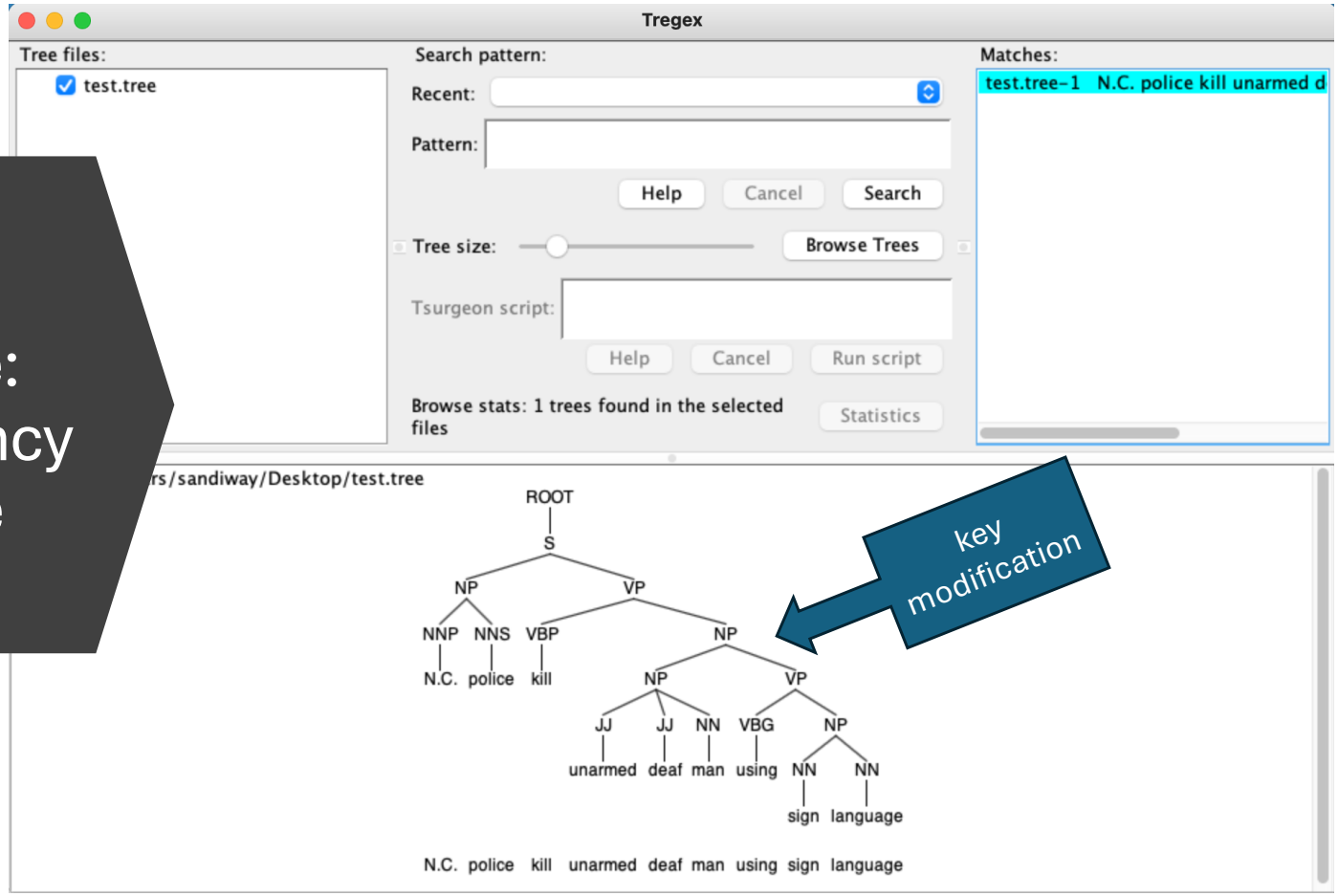


# Class Exercise: constituency pipeline

- *N.C. police kill unarmed deaf man using sign language*

```
>>>  nlp = stanza.Pipeline(lang='en',
 processors='tokenize,pos,constituency')
>>> doc = nlp('N.C. police kill unarmed deaf man using sign
language')
>>> for sentence in doc.sentences:
...     print(sentence.constituency)
...
(ROOT (S (NP (NNP N.C.) (NNS police)) (VP (VBP kill) (NP (NP (JJ
unarmed) (JJ deaf) (NN man)) (VP (VBG using) (NP (NN sign) (NN
language)))))))
```

# Class Exercise: constituency pipeline



# Crash Blossom Archive

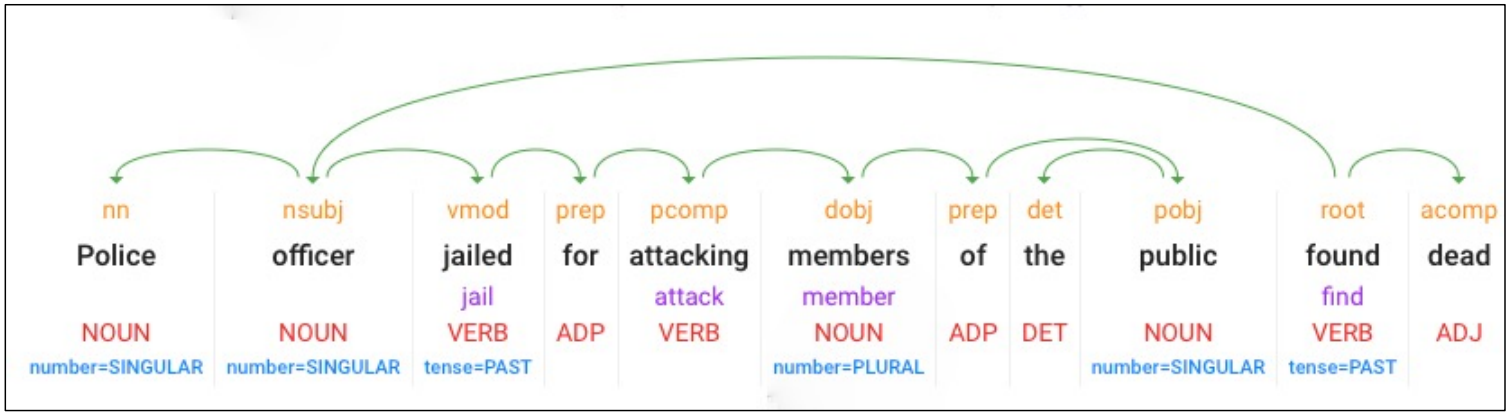
- https://languagelog.ldc.upenn.edu/nll/?cat=118

**Police officer jailed for attacking members of the public found dead**

https://www.theguardian.com/uk-news/2021/dec/29/pc-jailed-for-attacking-members-of-the-public-found-dead-at-home

- *vmod*: reduced non-finite verbal modifier
  - *jailed* modifies *police officer*
- *pcomp*: prepositional complement,
- *for* clausal complement

  *acomp*: adjectival complement



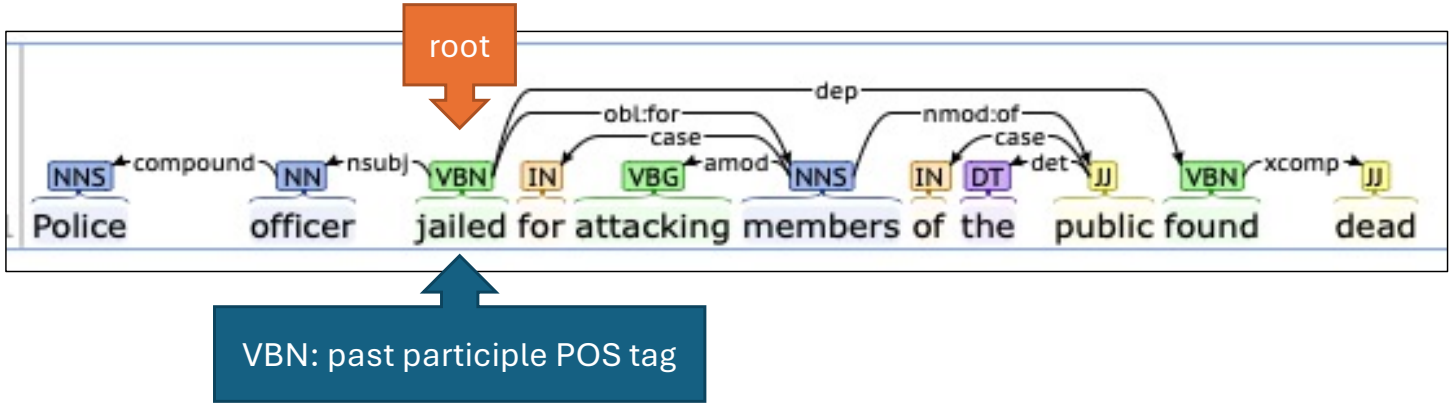
# Crash Blossom Archive

- https://languagelog.ldc.upenn.edu/nll/?cat=118

CoreNLP is your one stop shop for natural language processing in Java!
CoreNLP currently supports 8 languages: Arabic, Chinese, English, French, German, Hungarian, Italian, and Spanish.

**Police officer jailed for attacking members of the public found dead**

Stanford CoreNLP: https://corenlp.run

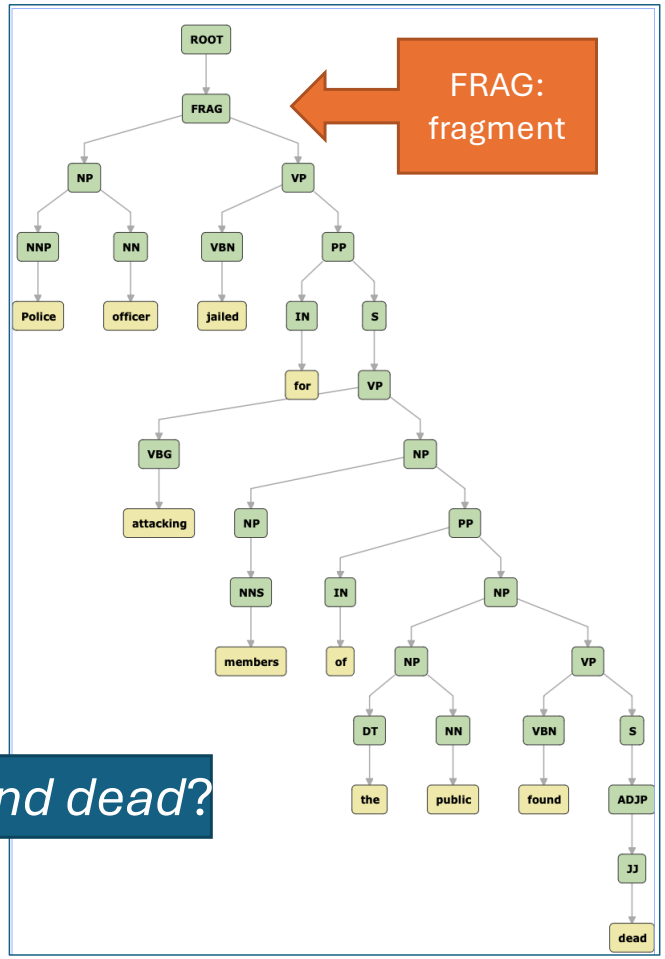


# Crash Blossom Archive

- https://languagelog.ldc.upenn.edu/nll/?cat=118

**Police officer jailed for attacking members of the public found dead**

Stanford CoreNLP: https://corenlp.run




*Who is found dead?*

# Crash Blossom Archive

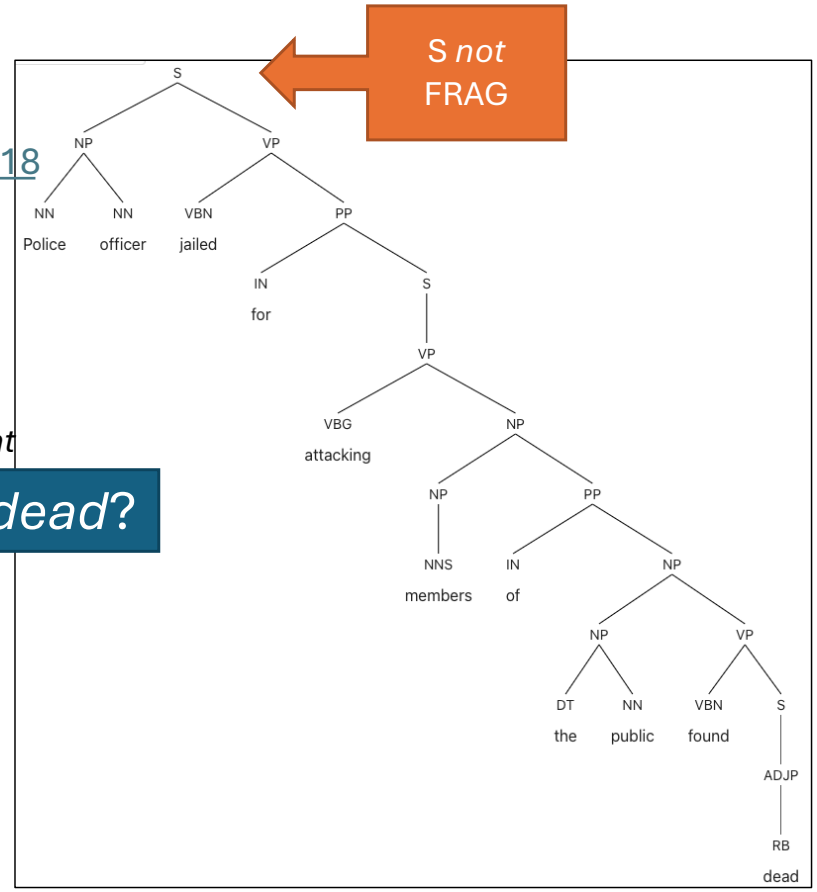
- https://languagelog.ldc.upenn.edu/nll/?cat=118

**Police officer jailed for attacking members of the public found dead**

Berkeley Neural Parser: https://parser.kitaev.io

*see the VP attachment*

*Who is found dead?*



# NYTimes Article

https://www.nytimes.com/2010/01/31/magazine/31FOB-onlanguage-t.html

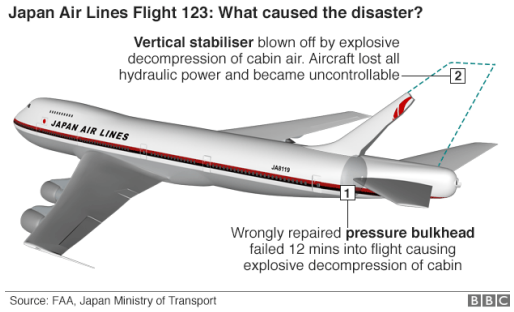
## Crash Blossoms

Jan. 27, 2010

On Language

**Elizabeth Barrett Browning** once gave the poetry of her husband, Robert, a harsh assessment, criticizing his habit of excessively paring down his syntax with opaque results. “You sometimes make a dust, a dark dust,” she wrote him, “by sweeping away your little words.”

In their quest for concision, writers of newspaper headlines are, like Robert Browning, inveterate sweepers away of little words, and the dust they kick up can lead to some amusing ambiguities. Legendary headlines from years past (some of which verge on the mythical) include “Giant Waves Down Queen Mary’s Funnel,” “MacArthur Flies Back to Front” and “Eighth Army Push Bottles Up Germans.” The Columbia Journalism Review even published two anthologies of ambiguous headlinese in the 1980s, with the classic titles “Squad Helps Dog Bite Victim” and “Red Tape Holds Up New Bridge.”




For years, there was no good name for these double-take headlines. Last August, however, one emerged in the Testy Copy Editors online discussion forum. Mike O’Connell, an American editor based in Sapporo, Japan, spotted the headline “Violinist Linked to JAL Crash Blossoms” and wondered, “What’s a crash blossom?” (The article, from the newspaper Japan Today, described the successful musical career of Diana Yukawa, whose father died in a 1985 Japan Airlines plane crash.) Another participant in the forum, Dan Bloom, suggested that “crash blossoms” could be used as a label for such infelicitous headlines that encourage alternate readings, and news of the neologism quickly spread.

# Homework 8

- Step 1: Read the New York Times article on *Crash Blossoms* (**CB**)
- Step 2: Install Stanza into nltk on your machine
- Step 3: Find two headlines that you consider to be **CBs**
  - either from the internet
  - or look at the Crash Blossom Archive
    - https://languagelog.ldc.upenn.edu/nll/?cat=118
  - (Don't use one I've already highlighted in class.)
  - Explain why they are a CB,
    - e.g. *what is the misread*, *why they may be misinterpreted*.

# Homework 8

- Step 4: Now run the headlines you chose on **all** of the following parsers:
  1. CoreNLP https://corenlp.run *(if it's running)*
  2. Stanza, on your computer, dependency and constituency, and
  3. Berkeley Neural Parser https://parser.kitaev.io
  - Explain each parse, i.e. whether it got
    - (A) the intended reading,
    - (B) your **CB** reading (*funny*), or
    - (C) some other reading (*incorrect*).

# Homework 8

- Submit to sandiway@arizona.edu
- SUBJECT: 581 Homework 8 *YOUR NAME*
- One PDF file (for grading)
  - include your screenshots in your answer
  - put a web link to each headline (*if you found it online*)
- Usual Deadline:
  - Saturday March 21st
  - we will review the homework on Monday March 23rd