NLP

Find your favorite news source and grab the article text.

1. Show the most common words in the article.
2. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})
3. Find a subject/object relationship through the dependency parser in any sentence.
4. Show the most common Entities and their types.
5. Find Entites and their dependency (hint: entity.root.head)
6. Find the most similar words in the article

Note: Yes, the notebook from the video is not provided, I leave it to you to make your own :) it's your final assignment for the semester. Enjoy!

In [1]:
```
# Installing Spacy

!pip3 install spacy
```

```
Collecting spacy
  Downloading spacy-3.6.0-cp310-cp310-macosx_10_9_x86_64.whl (6.9 MB)
                                                    6.9/6.9 MB 21.3 MB/s
eta 0:00:0000:0100:01
Collecting typer<0.10.0,>=0.3.0
  Downloading typer-0.9.0-py3-none-any.whl (45 kB)
                                                    45.9/45.9 kB 1.5 MB/s
eta 0:00:00
Collecting wasabi<1.2.0,>=0.9.1
  Downloading wasabi-1.1.2-py3-none-any.whl (27 kB)
Requirement already satisfied: setuptools in /Users/obelisk/anaconda3
/lib/python3.10/site-packages (from spacy) (65.6.3)
Collecting spacy-legacy<3.1.0,>=3.0.11
  Downloading spacy_legacy-3.0.12-py2.py3-none-any.whl (29 kB)
Requirement already satisfied: numpy>=1.15.0 in /Users/obelisk/anacon
da3/lib/python3.10/site-packages (from spacy) (1.23.5)
Collecting srsly<3.0.0,>=2.4.3
  Downloading srsly-2.4.7-cp310-cp310-macosx_10_9_x86_64.whl (492 kB)
                                                    492.7/492.7 kB 12.7 MB/s
eta 0:00:00
Collecting spacy-loggers<2.0.0,>=1.0.0
  Downloading spacy_loggers-1.0.4-py3-none-any.whl (11 kB)
Collecting pathy>=0.10.0
  Downloading pathy-0.10.2-py3-none-any.whl (48 kB)
                                                    48.9/48.9 kB 1.6 MB/s
```

```
eta 0:00:00
Collecting pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4
  Downloading pydantic-1.10.12-cp310-cp310-macosx_10_9_x86_64.whl (2.
9 MB)
                                           ━━━━━━━━━ 2.9/2.9 MB 23.8 MB/s
eta 0:00:0000:0100:01
Collecting murmurhash<1.1.0,>=0.28.0
  Downloading murmurhash-1.0.9-cp310-cp310-macosx_10_9_x86_64.whl (18
kB)
Collecting cymem<2.1.0,>=2.0.2
  Downloading cymem-2.0.7-cp310-cp310-macosx_10_9_x86_64.whl (32 kB)
Collecting preshed<3.1.0,>=3.0.2
  Downloading preshed-3.0.8-cp310-cp310-macosx_10_9_x86_64.whl (106 k
B)
                                           ━━━━━━━━━ 107.0/107.0 kB 4.0 MB/s
eta 0:00:00
Collecting thinc<8.2.0,>=8.1.8
  Downloading thinc-8.1.11-cp310-cp310-macosx_10_9_x86_64.whl (859 kB
)
                                           ━━━━━━━━━ 859.7/859.7 kB 15.4 MB/s
eta 0:00:0000:01
Collecting langcodes<4.0.0,>=3.2.0
  Downloading langcodes-3.3.0-py3-none-any.whl (181 kB)
                                           ━━━━━━━━━ 181.6/181.6 kB 6.2 MB/s
eta 0:00:00
Requirement already satisfied: packaging>=20.0 in /Users/obelisk/anac
onda3/lib/python3.10/site-packages (from spacy) (22.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy) (4.64.1)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from spacy) (5.2.1)
Requirement already satisfied: jinja2 in /Users/obelisk/anaconda3/lib
/python3.10/site-packages (from spacy) (3.1.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy) (2.28.1)
Collecting catalogue<2.1.0,>=2.0.6
  Downloading catalogue-2.0.9-py3-none-any.whl (17 kB)
Requirement already satisfied: typing-extensions>=4.2.0 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from pydantic!=1.8,!=1.8
.1,<1.11.0,>=1.7.4->spacy) (4.4.0)
Requirement already satisfied: idna<4,>=2.5 in /Users/obelisk/anacond
a3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy)
(3.4)
Requirement already satisfied: charset-normalizer<3,>=2 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.
13.0->spacy) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Users/obelis
k/anaconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.
0->spacy) (1.26.14)
Requirement already satisfied: certifi>=2017.4.17 in /Users/obelisk/a
```

```
naconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->
spacy) (2022.12.7)
Collecting blis<0.8.0,>=0.7.8
  Downloading blis-0.7.10-cp310-cp310-macosx_10_9_x86_64.whl (6.1 MB)
                                                    6.1/6.1 MB 26.4 MB/s
eta 0:00:0000:0100:01
Collecting confection<1.0.0,>=0.0.1
  Downloading confection-0.1.1-py3-none-any.whl (34 kB)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from typer<0.10.0,>=0.3.0->sp
acy) (8.0.4)
Requirement already satisfied: MarkupSafe>=2.0 in /Users/obelisk/anac
onda3/lib/python3.10/site-packages (from jinja2->spacy) (2.1.1)
Installing collected packages: cymem, wasabi, typer, spacy-loggers, s
pacy-legacy, pydantic, murmurhash, langcodes, catalogue, blis, srsly,
preshed, pathy, confection, thinc, spacy
Successfully installed blis-0.7.10 catalogue-2.0.9 confection-0.1.1 c
ymem-2.0.7 langcodes-3.3.0 murmurhash-1.0.9 pathy-0.10.2 preshed-3.0.
8 pydantic-1.10.12 spacy-3.6.0 spacy-legacy-3.0.12 spacy-loggers-1.0.
4 srsly-2.4.7 thinc-8.1.11 typer-0.9.0 wasabi-1.1.2
```

In [2]:
```python
# Downloading Model

!python3 -m spacy download en
```

```
2023-08-07 19:16:34.481344: I tensorflow/core/platform/cpu_feature_gu
ard.cc:182] This TensorFlow binary is optimized to use available CPU
instructions in performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations,
rebuild TensorFlow with the appropriate compiler flags.
⚠ As of spaCy v3.0, shortcuts like 'en' are deprecated. Please use th
e
full pipeline package name 'en_core_web_sm' instead.
Collecting en-core-web-sm==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/down
load/en_core_web_sm-3.6.0/en_core_web_sm-3.6.0-py3-none-any.whl
(https://github.com/explosion/spacy-models/releases/download/en_core_
web_sm-3.6.0/en_core_web_sm-3.6.0-py3-none-any.whl) (12.8 MB)
                                                    12.8/12.8 MB 24.2 MB/s
eta 0:00:0000:0100:01
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from en-core-web-sm==3.6.0) (
3.6.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0
->en-core-web-sm==3.6.0) (5.2.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0-
>en-core-web-sm==3.6.0) (3.3.0)
```

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-sm==3.6.0) (2.4.7)
Requirement already satisfied: packaging>=20.0 in /Users/obelisk/anac
onda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core
-web-sm==3.6.0) (22.0)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /Users/obelisk
/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en
-core-web-sm==3.6.0) (1.1.2)
Requirement already satisfied: jinja2 in /Users/obelisk/anaconda3/lib
/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-web-sm==
3.6.0) (3.1.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0-
>en-core-web-sm==3.6.0) (2.0.9)
Requirement already satisfied: pathy>=0.10.0 in /Users/obelisk/anacon
da3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-w
eb-sm==3.6.0) (0.10.2)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-sm==3.6.0) (4.64.1)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0-
>en-core-web-sm==3.6.0) (2.28.1)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-sm==3.6.0) (2.0.7)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /Users/
obelisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.
6.0->en-core-web-sm==3.6.0) (3.0.12)
Requirement already satisfied: numpy>=1.15.0 in /Users/obelisk/anacon
da3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-w
eb-sm==3.6.0) (1.23.5)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /Users/obelis
k/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->e
n-core-web-sm==3.6.0) (3.0.8)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /Users/
obelisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.
6.0->en-core-web-sm==3.6.0) (1.0.4)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-sm==3.6.0) (8.1.11)
Requirement already satisfied: setuptools in /Users/obelisk/anaconda3
/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-web-
sm==3.6.0) (65.6.3)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /Users/ob
elisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.
0->en-core-web-sm==3.6.0) (1.0.9)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4
in /Users/obelisk/anaconda3/lib/python3.10/site-packages (from spacy<

```
3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.10.12)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /Users/obelisk
/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en
-core-web-sm==3.6.0) (0.9.0)
Requirement already satisfied: typing-extensions>=4.2.0 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from pydantic!=1.8,!=1.8
.1,<1.11.0,>=1.7.4->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (4.4.
0)
Requirement already satisfied: idna<4,>=2.5 in /Users/obelisk/anacond
a3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy<
3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Users/obelis
k/anaconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.
0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (1.26.14)
Requirement already satisfied: charset-normalizer<3,>=2 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.
13.0->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /Users/obelisk/a
naconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->
spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (2022.12.7)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from thinc<8.2.0,>=8.1.8
->spacy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.1.1)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /Users/obelisk/a
naconda3/lib/python3.10/site-packages (from thinc<8.2.0,>=8.1.8->spac
y<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (0.7.10)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from typer<0.10.0,>=0.3.0->sp
acy<3.7.0,>=3.6.0->en-core-web-sm==3.6.0) (8.0.4)
Requirement already satisfied: MarkupSafe>=2.0 in /Users/obelisk/anac
onda3/lib/python3.10/site-packages (from jinja2->spacy<3.7.0,>=3.6.0-
>en-core-web-sm==3.6.0) (2.1.1)
Installing collected packages: en-core-web-sm
Successfully installed en-core-web-sm-3.6.0
✔ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

In [3]:
```python
# Importing Spacy

import spacy as sy
```

```
2023-08-07 19:17:05.887324: I tensorflow/core/platform/cpu_feature_gu
ard.cc:182] This TensorFlow binary is optimized to use available CPU
instructions in performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations,
rebuild TensorFlow with the appropriate compiler flags.
```

In [138]:
```python
# Build Model

nlp = sy.load('en_core_web_sm')
```

In [31]:
```python
# Load Article

# I tried loading both .docx and .txt files and both generated a "Extr

wapo = """FACING THE SURGE
SCROLL DOWN TO CONTINUE

With a fast burst of sea level rise, New Orleans's outer defenses face

ByChris Mooney,
Zoeann Murphy,
Ricky Carioti and
John Muyskens
Aug. 7 at 6:00 a.m.


Louisiana armed itself against the seas in the years after Hurricane K
But since 2010, the U.S. Gulf Coast has seen a sudden burst of rapid s
At Shell Beach, just two miles from commercial fisherman Charles Robin

Scientists with the U.S. Geological Survey and the state have been clo
Chart shows estuarine marshes in the Terrebonne, Barataria, Breton Sou

Chart shows saltwater and brackish marshes in the Terrebonne, Baratari
Source: Brady Couvillion
"This is becoming a trend that we've seen for six or seven years now,"
A group of scientists at Tulane University have also been investigatin
"The number of the 'drowning' sites is much more than I thought before


This spring, a group of Louisianans who have seen the rapid changes ga
Polly Glover, whose life and family history are entwined with the stat
Her friend, oyster farmer and shrimper George Barisich, has also seen

Where wetlands are disappearing
The wetlands losses are especially worrying in regions crucial to prot
The wetlands of middle Barataria are shrinking faster than 75 percent
In the lower Breton Sound region, losses are happening even faster. Mo
Map comparing the extent of wetlands in the Breton Sound and Mid-Barat

Louisiana's wetlands shrank dramatically after being torn up by Katrin
In middle Barataria, wetlands up against the ocean are receding, leadi

Sea level rise is driven both by land sinking — or subsidence — and th
```

The faster seas rise, the less effective the state's widely praised pl
In 2012, Louisiana projected that, if sea level rise and other environ
"Over the period of 10 years, the state has gone from potentially bein
The science behind the coastal planning process has gotten better, sai
Louisiana still thinks its plans will make a difference. One of the la
Many commercial fishers — including Barisich and Robin — believe that

Harnessing the Mississippi
Glover traveled toward the tip of the toe of the boot that is Louisian
The trip had personal significance — she was heading to Bayou Lamoque,

Glover's father died in 2007. She knew her dad, engineer John Edgar La
In an old box she found in her brother's house, Glover discovered plan
Locator map that shows the location of the following map within Louisi

New
Orleans
Detail
Map showing the location of a freshwater diversion built in the 1950s
BRETON
SOUND
Freshwater diversion
built in the 1950s
1 MILE
Source: Satellite imagery via Planet Labs PBC
Today, the diversion is in disrepair, apparently clogged with so much
Land's name is not on the plans, but Glover says she recognizes the ha

Louisiana is planning two far more massive diversions, one channeling
The mid–Barataria diversion will cost $ 2.9 billion and channel enormo

The question now is, with seas rising faster than expected and wetland
A linked fate
Today, New Orleans's coastal barrier system is designed to protect aga
But rising seas mean levees and barriers might have to be reworked. As

Austin Feldbaum heads up hazard mitigation at New Orleans's Office of
The city has built a comprehensive strategy against storms that includ
"The preservation and restoration of coastal wetlands is a critical pi
And as seas continue to rise, the impacts will not be borne equally, s

About this story
Design and development by Hailey Haymond. Editing by Katie Zezima, Ang
Data on Louisiana wetland changes is from Brady Couvillion's work on b
The map comparing wetlands in 1985 and 2020 uses cloud–free mosaics of

In [32]:
```python
# Process Article

wapo_proc = nlp(wapo)
```

In [48]:
```python
# Word Counting

from collections import Counter

words = [token.text
         for token in wapo_proc
         if not token.is_stop and not token.is_punct and not token.is_


word_freq = Counter(words)
common_words = word_freq.most_common(10)
print(common_words)
```

[('wetlands', 20), ('Louisiana', 17), ('sea', 9), ('rise', 9), ('stat
e', 9), ('Barataria', 9), ('water', 9), ('New', 8), ('Orleans', 8), (
'marshes', 8)]

The top 10 most common words in the article are 'wetlands' wit
h 20, 'Louisiana' with 17, 'sea' with 9, 'rise' with 9, 'state
' with 9, 'New' with 8, 'Orleans' with 8, and 'marshes' with 8
. I excluded what Spacy calls 'stop words' ie 'the', 'a' etc.

In [53]:
```python
# Type Counting

nouns = [token.text
         for token in wapo_proc
         if (not token.is_stop and
             not token.is_punct and
             token.pos_ == "NOUN")]

noun_freq = Counter(nouns)
common_nouns = noun_freq.most_common(5)

verbs = [token.text
         for token in wapo_proc
         if (not token.is_stop and
             not token.is_punct and
             token.pos_ == "VERB")]

verb_freq = Counter(verbs)
```

```python
common_verbs = verb_freq.most_common(5)

adjectives = [token.text
                for token in wapo_proc
                if (not token.is_stop and
                    not token.is_punct and
                    token.pos_ == "ADJ")]

adj_freq = Counter(adjectives)
common_adj = adj_freq.most_common(5)

adverbs = [token.text
            for token in wapo_proc
            if (not token.is_stop and
                not token.is_punct and
                token.pos_ == "ADV")]

adv_freq = Counter(adverbs)
common_adv = adv_freq.most_common(5)

propers = [token.text
            for token in wapo_proc
            if (not token.is_stop and
                not token.is_punct and
                token.pos_ == "PROPN")]

propn_freq = Counter(propers)
common_propn = propn_freq.most_common(5)

print(common_nouns, common_verbs, common_adj, common_adv, common_propn
```

```
[('wetlands', 19), ('sea', 9), ('state', 9), ('water', 9), ('level',
7)] [('said', 7), ('seen', 4), ('shows', 4), ('built', 4), ('rebuild'
, 3)] [('coastal', 7), ('major', 4), ('middle', 3), ('current', 3), (
'environmental', 3)] [('faster', 4), ('farther', 2), ('potentially',
2), ('apparently', 2), ('equally', 2)] [('Louisiana', 17), ('New', 8)
, ('Orleans', 8), ('Barataria', 8), ('Breton', 7)]
```

The top 5 most common nouns are 'wetlands' with 19 instances, 'sea' with 9, 'water' with 9, and 'level' with 7.

The top 5 most common verbs are 'said' with 7 instances, 'seen' with 4, 'shows' with 4, 'built' with 4, and 'rebuild' with 3.

The top 5 most common adjectives are 'coastal' with 7 instances, 'major' with 4, 'middle' with 3, 'current' with 3, and 'environmental' with 3.

The top 5 most common adverbs are 'faster' with 4 instances, 'farther' with 2, 'potentially' with 2, 'apparently' with 2, and 'equally' with 2.

The top 5 most common proper nouns are 'Louisiana' with 17 instances, 'New' and 'Orleans' with 8, 'Barataria' with 8, and 'Breton' with 7.

In [82]:
```python
# Dependencies Parser

def pr_tree(word, level):
    if word.is_punct:
        return
    for child in word.lefts:
        pr_tree (child, level+1)
    print('\t'* level + word.text + ' - ' + word.dep_)
    for child in word.rights:
        pr_tree (child, level+1)
```

In [99]:
```python
# Parsing

sentence_spans = list(wapo_proc.sents)

for sentence in sentence_spans[3]:
    pr_tree(sentence, 0)
```

```
At - prep
                its - poss
        center - pobj
its - poss
        its - poss
center - pobj
        At - prep
```

```
                                                    its — poss
                                        center — pobj
                        lie — ROOT
                                        the — det
                                wetlands — dobj
                                                    that — nsubj
                                        make — relcl
                                                up — prt
                                                            the — det
                                                            first — amod
                                                    line — dobj
                                                            of — prep
                                                                    defense — pobj
                                                            for — prep
                                                                            New — compound
                                                                    Orleans — pobj
                                                    buffering — advcl
                                                                    the — det
                                                            levees — dobj
                                                                    and — cc
                                                                    barriers — conj
                                                                    behind — prep
                                                                            them — pobj
                        the — det
                                the — det
                        wetlands — dobj
                                                    that — nsubj
                                        make — relcl
                                                up — prt
                                                            the — det
                                                            first — amod
                                                    line — dobj
                                                            of — prep
                                                                    defense — pobj
                                                            for — prep
                                                                            New — compound
                                                                    Orleans — pobj
                                                    buffering — advcl
                                                                    the — det
                                                            levees — dobj
                                                                    and — cc
                                                                    barriers — conj
                                                                    behind — prep
                                                                            them — pobj
                        that — nsubj
                                that — nsubj
                        make — relcl
                                up — prt
                                                    the — det
                                                    first — amod
```

```
                               line — dobj
                                    of — prep
                                            defense — pobj
                                    for — prep
                                                    New — compound
                                            Orleans — pobj
                          buffering — advcl
                                            the — det
                                    levees — dobj
                                            and — cc
                                            barriers — conj
                                            behind — prep
                                                    them — pobj
            up — prt
            the — det
            first — amod
                    the — det
                    first — amod
            line — dobj
                    of — prep
                            defense — pobj
                    for — prep
                                    New — compound
                            Orleans — pobj
            of — prep
                    defense — pobj
            defense — pobj
            for — prep
                            New — compound
                    Orleans — pobj
            New — compound
                    New — compound
            Orleans — pobj
            buffering — advcl
                            the — det
                    levees — dobj
                            and — cc
                            barriers — conj
                            behind — prep
                                    them — pobj
            the — det
                    the — det
            levees — dobj
                    and — cc
                    barriers — conj
                    behind — prep
                            them — pobj
            and — cc
            barriers — conj
            behind — prep
```
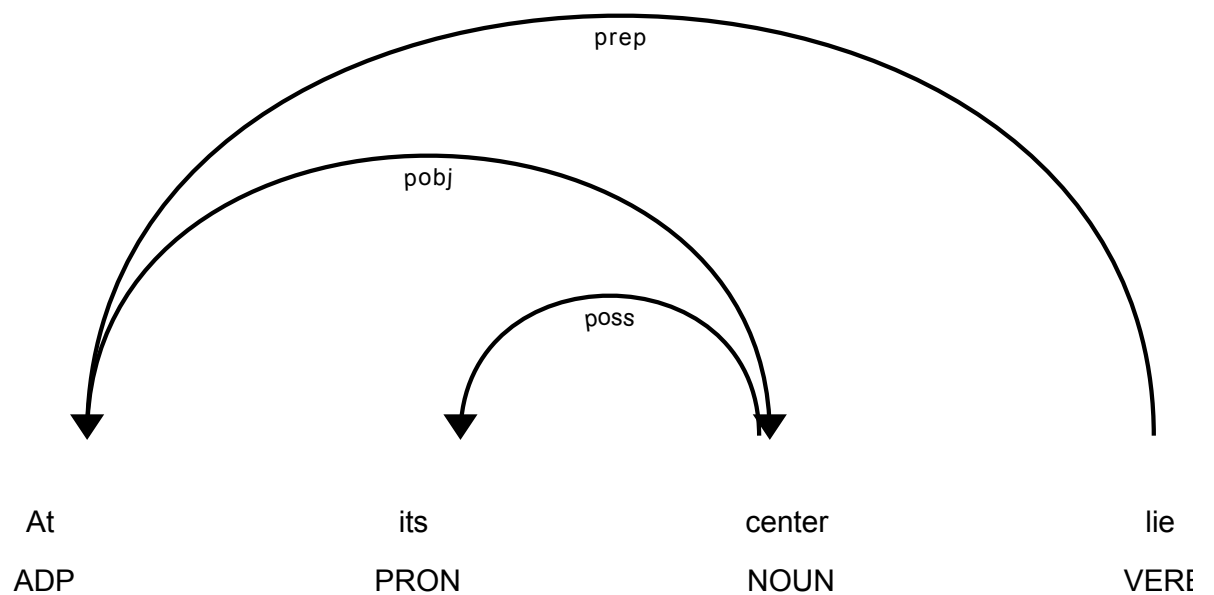
```
                them – pobj
       them – pobj
```

In [61]:
```python
# Displacy Dendencies

from spacy import display

sentence_spans = list(wapo_proc.sents)
display.render(sentence_spans[3], style="dep")
```

Below is one of the sentences from the Displacy render. You ca
n see the relations within the sentence next to the word. Refe
rence the render for the full arrows.

At ADP its PRON center NOUN lie VERB the DET wetlands NOUN that PRON make VERB
up ADP the DET first ADJ line NOUN of ADP defense NOUN for ADP New PROPN Orleans,
PROPN buffering VERB the DET levees NOUN and CCONJ barriers NOUN behind ADP
them. PRON

In [80]:
```python
# Finding Entities & Labels

entities = [(entity.text, entity.label_)
            for entity in wapo_proc.ents]

entities_freq = Counter(entities)
common_entities = entities_freq.most_common(10)

print(common_entities)
```

[(('Louisiana', 'GPE'), 14), (('Barataria', 'GPE'), 6), (('New Orlean
s', 'GPE'), 4), (('Breton Sound', 'GPE'), 4), (('the Mississippi Rive
r', 'LOC'), 4), (('first', 'ORDINAL'), 3), (('1985', 'DATE'), 3), (('
2020', 'DATE'), 3), (('Glover', 'PERSON'), 3), (('New Orleans's', 'GP
E'), 2)]

The top 10 most common entities in this article are 'Louisiana
' with 14 instances, 'Barataria' with 6, 'Breton Sound' with 5
, 'New Orleans' with 4, 'the Mississippi River' with 4, 'Glove
r' with 4, 'first' with 3, '1985' with 3, '2020' with 3, and '
New Orleans's' with 2. The model has distinguished between 'Ne
w Orleans' and 'New Orleans's' which is fascinating. I am also
uncertain why 'Louisiana' has 17 instances when looking at pro
per nouns but only 14 when looking at entities.
The entity types are as follows: Louisiana is a geopolitical e
ntity (GPE), Barataria is a GPE, Breton Sound is a GPE, New Or
leans is a GPE, the Mississippi River is a location (LOC), Glo
ver is a person, first is ordinal, 1985 is a date, 2020 is a d
ate, and New Orleans's is a GPE.

```
In [114]:  # Entity Dependencies

           for entity in wapo_proc.ents:
               pr_tree(entity.root.head, 0)
```

```
                                   Hill — conj
                                               Amanda — compound
                                       Voisard — conj
                                               Joe — compound
                                       Moore — conj
                                               and — cc
                                                     Monica — compound
                                               Ulmanu — conj
                   Katie — compound
           Zezima — pobj
                               Angela — compound
                               M. — compound
                   Hill — conj
                                       Amanda — compound
                               Voisard — conj
                                       Joe — compound
                               Moore — conj
                                       and — cc
                                               Monica — compound
                               Ulmanu — coni
```
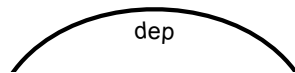
In [115]:
```python
# Displacy for Entity Dependencies

entity_spans = list(wapo_proc.ents)
displacy.render(entity_spans, style="dep")
```

Breton   Sound

PROPN   PROPN

first

ADJ

dep

In [140]:
```python
# Download larger language model

!python -m spacy download en_core_web_md
```

```
2023-08-08 14:15:30.297944: I tensorflow/core/platform/cpu_feature_gu
ard.cc:182] This TensorFlow binary is optimized to use available CPU
instructions in performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations,
rebuild TensorFlow with the appropriate compiler flags.
Collecting en-core-web-md==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/down
load/en_core_web_md-3.6.0/en_core_web_md-3.6.0-py3-none-any.whl
(https://github.com/explosion/spacy-models/releases/download/en_core_
web_md-3.6.0/en_core_web_md-3.6.0-py3-none-any.whl) (42.8 MB)
                                                   42.8/42.8 MB 22.9 MB/s
eta 0:00:0000:0100:01
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from en-core-web-md==3.6.0) (
3.6.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4
in /Users/obelisk/anaconda3/lib/python3.10/site-packages (from spacy<
3.7.0,>=3.6.0->en-core-web-md==3.6.0) (1.10.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /Users/
obelisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.
6.0->en-core-web-md==3.6.0) (1.0.4)
```

Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /Users/
obelisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.
6.0->en-core-web-md==3.6.0) (3.0.12)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0-
>en-core-web-md==3.6.0) (3.3.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-md==3.6.0) (4.64.1)
Requirement already satisfied: packaging>=20.0 in /Users/obelisk/anac
onda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core
-web-md==3.6.0) (22.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0-
>en-core-web-md==3.6.0) (2.28.1)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0
->en-core-web-md==3.6.0) (5.2.1)
Requirement already satisfied: pathy>=0.10.0 in /Users/obelisk/anacon
da3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-w
eb-md==3.6.0) (0.10.2)
Requirement already satisfied: setuptools in /Users/obelisk/anaconda3
/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-web-
md==3.6.0) (65.6.3)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /Users/obelisk
/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en
-core-web-md==3.6.0) (0.9.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /Users/obelis
k/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->e
n-core-web-md==3.6.0) (3.0.8)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /Users/obelisk
/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en
-core-web-md==3.6.0) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-md==3.6.0) (2.4.7)
Requirement already satisfied: jinja2 in /Users/obelisk/anaconda3/lib
/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-web-md==
3.6.0) (3.1.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /Users/obel
isk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0-
>en-core-web-md==3.6.0) (2.0.9)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-md==3.6.0) (8.1.11)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-
core-web-md==3.6.0) (2.0.7)
Requirement already satisfied: numpy>=1.15.0 in /Users/obelisk/anacon
da3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.0->en-core-w

```
eb-md==3.6.0) (1.23.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /Users/ob
elisk/anaconda3/lib/python3.10/site-packages (from spacy<3.7.0,>=3.6.
0->en-core-web-md==3.6.0) (1.0.9)
Requirement already satisfied: typing-extensions>=4.2.0 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from pydantic!=1.8,!=1.8
.1,<1.11.0,>=1.7.4->spacy<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (4.4.
0)
Requirement already satisfied: certifi>=2017.4.17 in /Users/obelisk/a
naconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->
spacy<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (2022.12.7)
Requirement already satisfied: idna<4,>=2.5 in /Users/obelisk/anacond
a3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.0->spacy<
3.7.0,>=3.6.0->en-core-web-md==3.6.0) (3.4)
Requirement already satisfied: charset-normalizer<3,>=2 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.
13.0->spacy<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Users/obelis
k/anaconda3/lib/python3.10/site-packages (from requests<3.0.0,>=2.13.
0->spacy<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (1.26.14)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /Users/obe
lisk/anaconda3/lib/python3.10/site-packages (from thinc<8.2.0,>=8.1.8
->spacy<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (0.1.1)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /Users/obelisk/a
naconda3/lib/python3.10/site-packages (from thinc<8.2.0,>=8.1.8->spac
y<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (0.7.10)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /Users/obelisk/
anaconda3/lib/python3.10/site-packages (from typer<0.10.0,>=0.3.0->sp
acy<3.7.0,>=3.6.0->en-core-web-md==3.6.0) (8.0.4)
Requirement already satisfied: MarkupSafe>=2.0 in /Users/obelisk/anac
onda3/lib/python3.10/site-packages (from jinja2->spacy<3.7.0,>=3.6.0-
>en-core-web-md==3.6.0) (2.1.1)
Installing collected packages: en-core-web-md
Successfully installed en-core-web-md-3.6.0
✔ Download and installation successful
You can now load the package via spacy.load('en_core_web_md')
```

In [154]:
```python
# Building Larger model

nlp_md = sy.load('en_core_web_md')  # most_similar() requires word vec

wapo_proc_md = nlp_md(wapo)
```

In [257]:
```python
# Testing my similarity code

for token in wapo_proc_md:
    if not token.is_stop and not token.is_punct and not token.is_space
        wapo_proc_md.vocab['token']
        print(token, wapo_proc_md.similarity(token))
```

```
FACING -0.0635669715220956
SURGE 0.33996460301493026
SCROLL -0.04787773373474156
CONTINUE -0.00046587897177378923
fast 0.23261411230251836
burst 0.07155603006615872
sea 0.2764892415463895
level 0.47269041937660966
rise 0.3871019456881518
New 0.14639698243437788
Orleans 0.22123043609458123
outer 0.31221468488907966
defenses 0.5262883414954734
face 0.3334105307036669
major 0.5707226999475583
test 0.3240830161677376
ByChris 0.0
Mooney -0.10311825334040431
Zoeann 0.0
Murphy -0.0620284707887364
```

In [249]:
```python
# Similarities as a list

similar = [(token.text, wapo_proc_md.similarity(token)) # It throws an
              for token in wapo_proc_md
                  if not token.is_stop and not token.is_punct and no
```

```
/var/folders/h0/3gzsn_d548l8bqt1z1_nhch00000gn/T/ipykernel_72035/1829
591939.py:1: UserWarning: [W008] Evaluating Doc.similarity based on e
mpty vectors.
  similar = [(token.text, wapo_proc_md.similarity(token))
```

In [256]:
```python
# Top 10 most similar words

def Sort_Tuple(tup):

    tup.sort(reverse=True, key = lambda x: x[1])
    return tup

print(Sort_Tuple(similar[0:10]))
```

[('considerable', 0.802863169306438), ('surrounding', 0.7833047686043
321), ('buffering', 0.7823169808875561), ('especially', 0.77895933716
68303), ('dramatically', 0.7765756515225591), ('significance', 0.7723
68379941853), ('suggesting', 0.7580252619466579), ('potentially', 0.7
558103332039718), ('potentially', 0.7558103332039718), ('Diversion',
0.7520432849106711)]

The top 10 most similar words in the article are 'considerable
' with a score of 0.803, 'surrounding' with 0.783, 'buffering'
with 0.782, 'especially' with 0.779, 'dramatically', with 0.77
7, 'siginifiance' with 0.772, 'suggesting' with 0.758, 'potent
ially' with 0.756, 'potentially' again with the same score, an
d 'Diversion' with a score of 0.752.