Carter Auer
NLP Take Home Exam portion

Let me just start by saying I'm not a linguist and I'm not someone who knows much about how sentence structure works in any way really. That being said I threw things at a wall with a goal in mind and I saw what stuck. Here's what I learned, I learned that there are not a lot of articles we are processing so there are some subjects that I wasn't able to get a single parse out of for example Event and Product. However there were some that have many for example GPE and ORG and Money to a lesser degree. I also learned that passive words are free and far between and when using that in my dependency I would get at most one or 2 sentences. My goal was to find information about countries' money and companies since I believe there would be some interesting overlap. And along with that goal I wanted to find out information about what was happening with those subjects. For example, Apple released its earnings report.

In order to achieve this I added a third condition to the token.ent_type_ check:

```
        (Note there are not docs that fall into Product or Event)
        I created a third or so i ca get more data since rightnow I want to find data about money and its involvement in countries
        and adding ORG will bring in some more relevant information
...

    if token.ent_type_ == "GPE" or token.ent_type_ == "MONEY" or token.ent_type_ == "ORG" :#GPE and ORG are entitly labels
        #If the entity is the object
        #print(token)
        #print(token.head)
```

I also modified the entity labels in a different pattern compared to the example code given. This is because in my finding a lot of the combinations I was trying would yield no results at all when parsing. This lead me to fall back to a very similar combination to the one that was given with one noticeable change

```
    if token.dep_ in ("prep"): #orginally dobj
        #tried using prep since i want to gather data on what governments and comapnies are doing
        #Find the subject (who performed the action)
        subj = [w for w in token.head.lefts if w.dep_ == "nsubj"] #orginally nsubj
        #What action was taken
        verb = token.head
        if subj:
            print(subj, verb, token)
    #If the entity is the subject

    elif token.dep_ in ("nsubj"): #orginally nsubj
        #Find the object
        obj = [w for w in token.head.rights if w.dep_ == "dobj"] #orginally dobj
        #tried using prep since I want to gather data on what governments and comapnies are doing however I reverted it back based on results
        #Find the action taken
        verb = token.head
        if obj:#changed subj to obj
            print(token, verb, obj)
```

Using the NSubj and Dobj gave the best information from my observation but I felt that that was just like yours so I wanted to see if I was able to extract any more relevant data about the happenings of countries and industries in regards to money. I added the prep in order to extract just a little more information however it still wasn't as good as just using dobj in its place and I'm interested to learn why that is relative to this dataset.

Interesting observations when experimenting that was interesting is when changing where you look for prepositions for example here is two code snippets and their differences

```python
if token.dep_ in ("dobj"): #orginally dobj
    #tried using prep since i want to gather data on what governments and comapnies are doing
    #Find the subject (who performed the action)
    subj = [w for w in token.head.lefts if w.dep_ == "nsubj"] #orginally nsubj
    #What action was taken
    verb = token.head
    if subj:
        print(subj, verb, token)
#If the entity is the subject
elif token.dep_ in ("nsubj"): #orginally nsubj
    #Find the object
    obj = [w for w in token.head.rights if w.dep_ == "prep"] #orginally dobj
    #tried using prep since i want to gather data on what governments and comapnies are doing
    #Find the action taken
    verb = token.head
    if obj:#changed subj to obj
        print(token, verb, obj)
```

Output:

[it] acquired North Country Media
  Group
==
==
==
==
[a spokesman] told Reuters
==
==
==
VW confirmed [on]

>>>

```
        #print(token)
        #print(token.head)

        if token.dep_ in ("prep"): #orginally dobj
            #tried using prep since i want to gather data on what governments and comapnies are doing
            #Find the subject (who performed the action)
            subj = [w for w in token.head.lefts if w.dep_ == "nsubj"] #orginally nsubj
            #What action was taken
            verb = token.head
            if subj:
                print(subj, verb, token)
        #If the entity is the subject
        elif token.dep_ in ("nsubj"): #orginally nsubj
            #Find the object
            obj = [w for w in token.head.rights if w.dep_ == "dobj"] #orginally dobj
            #tried using prep since i want to gather data on what governments and comapnies are doing
            #Find the action taken
            verb = token.head
            if obj:#changed subj to obj
                print(token, verb, obj)
```

Output:
==
==
the Netherlands planted [a record 137,600
  hectares]
==
==
==
VW confirmed [the currency scandal]

Final findings below I have highlighted where I have been able to pull out useful information pertaining to money, companies, and government. And I'm very happy with the output I was able to get since in a lot of cases it cross referenced money in relation to companies and government.

>>> %Run 'CS 397 Take Home Code Exam.py'
==
Taiwan had [a trade trade surplus]
South Korea had [a trade surplus]
Japan selling [semiconductors]
Japan's ruling Liberal Democratic Party outlined [a package]
==
CHINA DAILY SAYS [VERMIN EAT, PCT GRAIN STOCKS]
==
JAPAN revise [its long-term energy supply/demand
  outlook]
MITI review [the breakdown]
==

==

Indonesia maintain [its exports]

==

==

The trade ministry and exchange board considering [the
  introduction]
Indonesia expand [its rubber marketing effort]

==

the U.S. Department of Agriculture approved [the Continental Grain Cosale, a tonne C]

==

==

Sumitomo paid [much]
the Federal Reserve Board forbidden [them]

==

SUBROTO SAYS [INDONESIA]

==

The Bundesbank accepted [bids]
the Bundesbank withdrawing [a net 8.1 billion marks]
The Bundesbank steering [call money]

==

BOND CORP STILL CONSIDERING [ATLAS MINING BAIL-OUT
  Bond Corp Holdings Ltd &]
Bond acquire [Atlas' total loans]
Bond Corp considering [several
  investments]

==

==

JAPAN MINISTRY SAYS [OPEN FARM TRADE]
Japan open [its farm products market]
the U.S. demand [liberalise]
Japan open [its protected rice market]
Australia increase [their wheat exports]
Japan opens [its beef imports]
Japan has [the potential]
The U.S. asked [the General Agreement on Tariffs]
Japan's
  top business group Keidanren urged [the government]

==

==

==

==

South Korea cut [its trade surplus]

==

Finland's Neste Oy has [extensive experience]

==

CRA SOLD [FORREST GOLD, Creek Consolidated NL]
CRA disclose [the price]

==

==

stg exceeded [market expectations]

==

Citibank lost [490,000 crowns]

==

Vieille Montagne reported [a 198 mln franc net loss]

==

==

Spain raised [no objections]

==

Japan has [little chance]
Japan take [steps]
Congress takes [action]

==

Thailand exported [4,842 tonnes]

==

the European Commission accepted [785,000 tonnes]

==

==

ANHEUSER-BUSCH joined [several other foreign bidders]
Anheuser-Busch told [the government]
Australian brewer Alan Bond's Bond Corp Holdings
  Ltd offered [150 pesos]
The Hong Kong Economic Journal quoted [a
  spokesman]
Anheuser-Busch made [a 150 mln dlr bid]

==

Australia take [tough
  economic decisions]

==

Kuwait has [an OPEC quota]

==

Malaysia sell [more
  palm oil]
==
==
Redland holds [about 49 pct]
==
NO INTERIM DIVIDEND
  Isuzu Motor Ltd & pay [no
  dividend]
==
U.S. Charges
  Japan violated [a pact]
The U.S. Has [excessive expectations]
==
Bell held [10 pct]
==
NIPPON STEEL DENIES [CHINA]
The Mainichi Shimbun quoted [Nippon Steel]
China's State Planning Commission asked [Japanese makers]
China made [no such official request]
Mainichi quoted [Nippon Steel officials]
==
==
Japan has [little
  chance]
Japan take [steps]
Congress takes [action]
==
The Bank of Japan bought [a small amount]
==
==
==
De los Angeles asked [the SEC]
de los Angeles amended [an earlier suit]
San Miguel Corp reached [agreement]
Diaz's Presidential Commission sequestered [the shares]
San Miguel Corp reported [sales revenue]
==
==
==

Iniziativa Meta & controls [the largest single stake]

==
==
==

the United States face [a vicious circle]
West Germany taken [steps]
the Bundesbank tolerated [rapid money supply
  growth]
West Germany faced [a difficult dilemma]
Poehl welcomed [the so-called Louvre accord]

==
==

California Biotech earned [114,000 dlrs]

==
==

Unilever indicating [plans, some smaller assets]
Unilever aquired [Chesebrough]

==
==

CONRAC <CAX started [negotiations]
Conrac Corp sait started [negotiations]

==
==
==
==
==

HOME SAVINGS AND LOAN ASSOCIATION INC & given [790,874]

==
==
==

Compaq reported [earnings]
the new COMPAQ DESKPRO [286]

==
==
==

The Federal Reserve Board has [a 50 pct margin requirement]
Congress pass [a bill]

==
==

Brazil suspended [interest payments]

==
==
==

Japan fulfilled [its
  pledges]

==
==
==
==

QTR MARCH 27 NET
  Shr primary 78 cts diluted [75 cts]

==
==
==
==
==
==
==
==

the Netherlands planted [a record 137,600
  hectares]

==
==
==

VW confirmed [the currency scandal]
>>>