

Feature extraction stage

June 14, 2023

```
[1]: import nltk
import numpy as np
from string import punctuation
```

```
[2]: import matplotlib.pyplot as plt
%matplotlib notebook
```

```
[3]: tagset = set([k[1] for k in nltk.corpus.treebank.tagged_words()])
print(tagset)
```

```
{'-LRB-', 'CD', ',', 'VBD', '$', 'RB', 'FW', 'WDT', '"', 'JJR', 'MD', 'POS',
'LS', 'VBN', 'NNS', 'WP', '#', 'IN', 'UH', 'RP', 'NN', '`', 'TO', '.', 'RBR',
'PRP$', 'JJ', 'PDT', 'NNPS', '-RRB-', 'VBG', 'RBS', 'CC', 'VBZ', 'PRP', 'WP$',
'SYM', 'JJS', 'VBP', 'EX', 'DT', 'VB', ':', '-NONE-', 'WRB', 'NNP'}
```

```
[4]: nltk.help.upenn_tagset("VBN")
```

```
VBN: verb, past participle
    multihulled dilapidated aerosolized chaired languished panelized used
    experimented flourished imitated reunified factored condensed sheared
    unsettled primed dubbed desired ...
```

```
[5]: text = nltk.corpus.treebank.tagged_sents()
print(text[0])
```

```
[('Pierre', 'NNP'), ('Vinken', 'NNP'), (',', ',', ','), ('61', 'CD'), ('years',
'NNS'), ('old', 'JJ'), (',', ',', ','), ('will', 'MD'), ('join', 'VB'), ('the',
'DT'), ('board', 'NN'), ('as', 'IN'), ('a', 'DT'), ('nonexecutive', 'JJ'),
('director', 'NN'), ('Nov.', 'NNP'), ('29', 'CD'), ('.', '.')]


```

```
[6]: print(*[i[0] for i in text[0]])
print([i[1] for i in text[0]])
```

```
Pierre Vinken , 61 years old , will join the board as a nonexecutive director
Nov. 29 .
['NNP', 'NNP', ',', 'CD', 'NNS', 'JJ', ',', 'MD', 'VB', 'DT', 'NN', 'IN', 'DT',
'JJ', 'NN', 'NNP', 'CD', '.']
```

```
[7]: len(text)
```

```
[7]: 3914
```

```
[8]: len(nltk.corpus.treebank.words())
```

```
[8]: 100676
```

```
[9]: len(list(set(nltk.corpus.treebank.words())))
```

```
[9]: 12408
```

```
[10]: len(set([i[1] for s in text for i in s]))
```

```
[10]: 46
```

```
[11]: print(len(punctuation), punctuation)
```

```
32 !"#$%&'()*+,-./:;<=>?@[\\]^_`{|}~
```

1 Context Function

```
[12]: def consent(s):  
    """coordinates represents, in the same order:  
    preceding word (by 2 positions),  
    preceding word (by 1 position)  
    central word  
    1st following word  
    2nd following word  
    is it a number  
    starts with capital  
    is it a starting word  
    is an ending word  
    etc."""  
    contxt = {}  
    l = len(s)  
    for i, w in enumerate(s):  
        contxt[(w[0],w[1],i)] = ((lambda x : s[i-2][0] if x>=2 else 0)(i),  
↪(lambda x : s[i-1][0] if x>=1 else 0)(i),  
        s[i][0], (lambda x : s[i+1][0] if x<1-1 else 0)(i), (lambda x :  
↪s[i+2][0] if x<1-2 else 0)(i),  
                int(w[0].isdigit()), int(w[0][0].isupper()),  
                int((lambda x : x==0)(i)), int((lambda x : x==1-1)(i)))  
    return contxt
```

```
[13]: print("Tag ---> Context", end="\n\n")
print(*[str(k[1])+str(" ---> ")+str(v) for k, v in consent(text[0]).items()],
      ↪sep="\n")
```

Tag ---> Context

```
NNP ---> (0, 0, 'Pierre', 'Vinken', ',', 0, 1, 1, 0)
NNP ---> (0, 'Pierre', 'Vinken', ',', '61', 0, 1, 0, 0)
, ---> ('Pierre', 'Vinken', ',', '61', 'years', 0, 0, 0, 0)
CD ---> ('Vinken', ',', '61', 'years', 'old', 1, 0, 0, 0)
NNS ---> (',', '61', 'years', 'old', ',', 0, 0, 0, 0)
JJ ---> ('61', 'years', 'old', ',', 'will', 0, 0, 0, 0)
, ---> ('years', 'old', ',', 'will', 'join', 0, 0, 0, 0)
MD ---> ('old', ',', 'will', 'join', 'the', 0, 0, 0, 0)
VB ---> (',', 'will', 'join', 'the', 'board', 0, 0, 0, 0)
DT ---> ('will', 'join', 'the', 'board', 'as', 0, 0, 0, 0)
NN ---> ('join', 'the', 'board', 'as', 'a', 0, 0, 0, 0)
IN ---> ('the', 'board', 'as', 'a', 'nonexecutive', 0, 0, 0, 0)
DT ---> ('board', 'as', 'a', 'nonexecutive', 'director', 0, 0, 0, 0)
JJ ---> ('as', 'a', 'nonexecutive', 'director', 'Nov.', 0, 0, 0, 0)
NN ---> ('a', 'nonexecutive', 'director', 'Nov.', '29', 0, 0, 0, 0)
NNP ---> ('nonexecutive', 'director', 'Nov.', '29', '.', 0, 1, 0, 0)
CD ---> ('director', 'Nov.', '29', '.', 0, 1, 0, 0, 0)
. ---> ('Nov.', '29', '.', 0, 0, 0, 0, 0, 1)
```

1.1 Choix des features

```
[14]: freq = 10 # filtre qui sélectionne les features apparaissant plus que freq fois
```

1.1.1 Preceding word by two positions

```
[15]: tag_prec2 = [(k[1],v[0]) for s in text for k, v in consent(s).items()]
tag_prec2 = nltk.Text(tag_prec2)
tag_prec2_freq = nltk.FreqDist(tag_prec2)
selected_tag_prec2 = [couple for couple in tag_prec2_freq.keys() if
  ↪tag_prec2_freq[couple]>5]
print("La nombre de paires (tag, word-2) apparaissant plus de 5 fois est: ",
  ↪len(selected_tag_prec2))
selected_prec2 = list(set([w[1] for w in selected_tag_prec2]))
print("Le nombre de feature (preceding_word_by_2_positions) sélectionnés est:
  ↪", len(selected_prec2))
print("Exemple: ", selected_prec2[:5])
```

La nombre de paires (tag, word-2) apparaissant plus de 5 fois est: 1947
 Le nombre de feature (preceding_word_by_2_positions) sélectionnés est: 603
 Exemple: [0, 'into', 'raise', 'first', 'effect']

1.1.2 Preceding word by one position

```
[16]: tag_prec1 = [(k[1],v[1]) for s in text for k, v in consent(s).items()]
tag_prec1 = nltk.Text(tag_prec1)
tag_prec1_freq = nltk.FreqDist(tag_prec1)
selected_tag_prec1 = [couple for couple in tag_prec1_freq.keys() if
    ↪tag_prec1_freq[couple]>5]
print("La nombre de paires (tag, word-1) apparaissant plus de 5 fois est: ",
    ↪len(selected_tag_prec1))
selected_prec1 = list(set([w[1] for w in selected_tag_prec1]))
print("Le nombre de feature (preceding_word_by_1_positions) sélectionnés est:
    ↪", len(selected_prec1))
print("Exemple: ", selected_prec1[:5])
```

La nombre de paires (tag, word-1) apparaissant plus de 5 fois est: 2083
Le nombre de feature (preceding_word_by_1_positions) sélectionnés est: 999
Exemple: [0, 'ringers', 'into', 'raise', 'first']

1.1.3 Following word by one position

```
[17]: tag_follow1 = [(k[1],v[3]) for s in text for k, v in consent(s).items()]
tag_follow1 = nltk.Text(tag_follow1)
tag_follow1_freq = nltk.FreqDist(tag_follow1)
selected_tag_follow1 = [couple for couple in tag_follow1_freq.keys() if
    ↪tag_follow1_freq[couple]>5]
print("La nombre de paires (tag, word+1) apparaissant plus de 5 fois est: ",
    ↪len(selected_tag_follow1))
selected_follow1 = list(set([w[1] for w in selected_tag_follow1]))
print("Le nombre de feature (following_word_by_1_positions) sélectionnés est:
    ↪", len(selected_follow1))
print("Exemple: ", selected_follow1[:5])
```

La nombre de paires (tag, word+1) apparaissant plus de 5 fois est: 2119
Le nombre de feature (following_word_by_1_positions) sélectionnés est: 974
Exemple: [0, 'ringers', 'hour', 'into', 'raise']

1.1.4 Following word by two positions

```
[18]: tag_follow2 = [(k[1],v[4]) for s in text for k, v in consent(s).items()]
tag_follow2 = nltk.Text(tag_follow2)
tag_follow2_freq = nltk.FreqDist(tag_follow2)
selected_tag_follow2 = [couple for couple in tag_follow2_freq.keys() if
    ↪tag_follow2_freq[couple]>5]
print("La nombre de paires (tag, word+2) apparaissant plus de 5 fois est: ",
    ↪len(selected_tag_follow2))
selected_follow2 = list(set([w[1] for w in selected_tag_follow2]))
```

```
print("Le nombre de feature (following_word_by_2_positions) sélectionnés est:␣
↪", len(selected_follow2))
print("Exemple: ", selected_follow2[:5])
```

La nombre de paires (tag, word+2) apparaissant plus de 5 fois est: 1951
 Le nombre de feature (following_word_by_2_positions) sélectionnés est: 600
 Exemple: [0, 'into', 'raise', 'first', 'taken']

1.1.5 Central word

```
[19]: tag_w0 = [(k[1],v[2]) for s in text for k, v in consent(s).items()]
tag_w0 = nltk.Text(tag_w0)
tag_w0_freq = nltk.FreqDist(tag_w0)
selected_tag_w0 = [couple for couple in tag_w0_freq.keys() if␣
↪tag_w0_freq[couple]>1]
print("La nombre de paires (tag, word) apparaissant plus de 5 fois est: ",␣
↪len(selected_tag_w0))
selected_w0 = list(set([w[1] for w in selected_tag_w0]))
print("Le nombre de feature (central_word) sélectionnés est: ",␣
↪len(selected_w0))
print("Exemple: ", selected_w0[:5])
```

La nombre de paires (tag, word) apparaissant plus de 5 fois est: 6332
 Le nombre de feature (central_word) sélectionnés est: 5753
 Exemple: ['Estate', 'settled', '1\\2', 'Serial', 'increasingly']

```
[20]: def context(s):
        """coordinates : the same order as for consent"""
        #if len(s)<=1:
        #    s = nltk.word_tokenize(s)
        contxt = {}
        l = len(s)
        for i, w in enumerate(s): #(w[1],i) >> we add i to ensure having repeated␣
↪tags
            contxt[(w[1], i)] = ((lambda x : s[i-2][0] if x>=2 else 0)(i), (lambda␣
↪x : s[i-1][0] if x>=1 else 0)(i),
                s[i][0], (lambda x : s[i+1][0] if x<1-1 else 0)(i), (lambda x :␣
↪s[i+2][0] if x<1-2 else 0)(i),
                                int(w[0].isdigit()), int(w[0][0].isupper()),
                                int((lambda x : x==0)(i)), int((lambda x : x==1-1)(i)))
        return [(k[0],contxt[k]) for k in contxt.keys()]
```

```
[21]: context(text[0])
```

```
[21]: [('NNP', (0, 0, 'Pierre', 'Vinken', ',', 0, 1, 1, 0)),
        ('NNP', (0, 'Pierre', 'Vinken', ',', '61', 0, 1, 0, 0)),
        (',', ('Pierre', 'Vinken', ',', '61', 'years', 0, 0, 0, 0)),
```

```

('CD', ('Vinken', ',', '61', 'years', 'old', 1, 0, 0, 0)),
('NNS', (',', '61', 'years', 'old', ',', 0, 0, 0, 0)),
('JJ', ('61', 'years', 'old', ',', 'will', 0, 0, 0, 0)),
(',', ('years', 'old', ',', 'will', 'join', 0, 0, 0, 0)),
('MD', ('old', ',', 'will', 'join', 'the', 0, 0, 0, 0)),
('VB', (',', 'will', 'join', 'the', 'board', 0, 0, 0, 0)),
('DT', ('will', 'join', 'the', 'board', 'as', 0, 0, 0, 0)),
('NN', ('join', 'the', 'board', 'as', 'a', 0, 0, 0, 0)),
('IN', ('the', 'board', 'as', 'a', 'nonexecutive', 0, 0, 0, 0)),
('DT', ('board', 'as', 'a', 'nonexecutive', 'director', 0, 0, 0, 0)),
('JJ', ('as', 'a', 'nonexecutive', 'director', 'Nov.', 0, 0, 0, 0)),
('NN', ('a', 'nonexecutive', 'director', 'Nov.', '29', 0, 0, 0, 0)),
('NNP', ('nonexecutive', 'director', 'Nov.', '29', '.', 0, 1, 0, 0)),
('CD', ('director', 'Nov.', '29', '.', 0, 1, 0, 0, 0)),
('.', ('Nov.', '29', '.', 0, 0, 0, 0, 0, 1))

```

```

[22]: new_dataset = []
      for txt in text:
          new_dataset.extend(context(txt))

new_dataset[15:30]

```

```

[22]: [('NNP', ('nonexecutive', 'director', 'Nov.', '29', '.', 0, 1, 0, 0)),
      ('CD', ('director', 'Nov.', '29', '.', 0, 1, 0, 0, 0)),
      ('.', ('Nov.', '29', '.', 0, 0, 0, 0, 0, 1)),
      ('NNP', (0, 0, 'Mr.', 'Vinken', 'is', 0, 1, 1, 0)),
      ('NNP', (0, 'Mr.', 'Vinken', 'is', 'chairman', 0, 1, 0, 0)),
      ('VBZ', ('Mr.', 'Vinken', 'is', 'chairman', 'of', 0, 0, 0, 0)),
      ('NN', ('Vinken', 'is', 'chairman', 'of', 'Elsevier', 0, 0, 0, 0)),
      ('IN', ('is', 'chairman', 'of', 'Elsevier', 'N.V.', 0, 0, 0, 0)),
      ('NNP', ('chairman', 'of', 'Elsevier', 'N.V.', ',', 0, 1, 0, 0)),
      ('NNP', ('of', 'Elsevier', 'N.V.', ',', 'the', 0, 1, 0, 0)),
      (',', ('Elsevier', 'N.V.', ',', 'the', 'Dutch', 0, 0, 0, 0)),
      ('DT', ('N.V.', ',', 'the', 'Dutch', 'publishing', 0, 0, 0, 0)),
      ('NNP', (',', 'the', 'Dutch', 'publishing', 'group', 0, 1, 0, 0)),
      ('VBG', ('the', 'Dutch', 'publishing', 'group', '.', 0, 0, 0, 0)),
      ('NN', ('Dutch', 'publishing', 'group', '.', 0, 0, 0, 0, 0))]

```

```

[23]: # Feature vector block that looks to the preceding word by two positions
      def f0(row):
          tmp = np.zeros(len(selected_prec2), dtype=np.int8)
          x = row[1][0]
          if x in selected_prec2:
              tmp[selected_prec2.index(x)]=1
          return tmp

```

```
[24]: # Feature vector block that looks to the preceding word by two positions
def f1(row):
    tmp = np.zeros(len(selected_prec1), dtype=np.int8)
    x = row[1][1]
    if x in selected_prec1:
        tmp[selected_prec1.index(x)]=1
    return tmp
```

```
[25]: # Feature vector block that looks to the central word
def f2(row):
    tmp = np.zeros(len(selected_w0), dtype=np.int8)
    x = row[1][2]
    if x in selected_w0:
        tmp[selected_w0.index(x)]=1
    return tmp
```

```
[26]: # Feature vector block that looks to the following word by one position
def f3(row):
    tmp = np.zeros(len(selected_follow1), dtype=np.int8)
    x = row[1][3]
    if x in selected_follow1:
        tmp[selected_follow1.index(x)]=1
    return tmp
```

```
[27]: # Feature vector block that looks to the following word by two positions
def f4(row):
    tmp = np.zeros(len(selected_follow2), dtype=np.int8)
    x = row[1][4]
    if x in selected_follow2:
        tmp[selected_follow2.index(x)]=1
    return tmp
```

```
[28]: def f(row):
    feature = np.concatenate((f0(row),f1(row), f2(row), f3(row), f4(row),
    ↪row[1][-4:]))
    return feature
```

1.1.6 A small test to control

```
[29]: f(new_dataset[0])
```

```
[29]: array([1, 0, 0, ..., 1, 1, 0])
```

```
[30]: # The number of activated features for the first 35 entries (t,x)
print([sum(f(i)) for i in new_dataset[:35]])
```

```
[5, 4, 2, 5, 4, 4, 4, 3, 5, 4, 4, 4, 3, 5, 3, 4, 6, 5, 6, 6, 4, 3, 3, 4, 5, 2,
```

2, 5, 4, 4, 5, 6, 3, 3, 5]

```
[31]: print("The number of activated features is: ", sum(f(new_dataset[33])), " for_
↳the pair below.")
print('The tag is: "{}".format(new_dataset[33][0]), " and the corresponding_
↳context is: ", new_dataset[33][1])
```

The number of activated features is: 3 for the pair below.

The tag is: "," and the corresponding context is: ('Rudolph', 'Agnew', ',', '55', 'years', 0, 0, 0, 0)

```
[32]: teeest = list(map(lambda x : (x[0],sum(f(x))), new_dataset[:10]))
teeest
```

```
[32]: [('NNP', 5),
('NNP', 4),
(',', 2),
('CD', 5),
('NNS', 4),
('JJ', 4),
(',', 4),
('MD', 3),
('VB', 5),
('DT', 4)]
```

```
[33]: len(f(new_dataset[0]))
```

```
[33]: 8933
```

```
[35]: len(new_dataset)
```

```
[35]: 100676
```

```
[36]: dataset = map(lambda x : (x[0],f(x)), new_dataset)
```

```
[ ]: dataset = list(dataset)
```

```
[ ]:
```

```
[10]: #####"" Context_
↳#####
text = nltk.corpus.treebank.tagged_sents()
def consent(s):
    """coordinates represents, in the same order:
    preceding word (by 2 positions),
    preceding word (by 1 position)
    central word
```



```

1st following word
2nd following word
is it a number
starts with capital
is it a starting word
is an ending word
etc. """
contxt = {}
# First word :
contxt[(s[0][0],s[0][1],1)] = (0, 0, s[0][0], s[1][0],s[2][0],s[0][0].
↳isdigit(), s[0][0][0].isupper(), 1, 0)
# Second word :
contxt[(s[1][0],s[1][1],2)] = (0, s[0][0], s[1][0], s[2][0], s[3][0],↳
↳s[1][0].isdigit(), s[1][0][0].isupper(), 0, 0)
# For middle words
for i, w in enumerate(s[2:-2]) :
    contxt[(w[0],w[1],i+3)] = (s[i][0], s[i+1][0], s[i+2][0], s[i+3][0],↳
↳s[i+4][0], w[0].isdigit(), w[0][0].isupper(), 0, 0)
# The word before last - the penultimate one - the second word to last
contxt[(s[-2][0],s[-2][1],len(s)-1)] = (s[-4][0], s[-3][0], s[-2][0],↳
↳s[-1][0], 0, s[-2][0].isdigit(), s[-2][0][0].isupper(),0, 0)
# The last word
contxt[(s[-1][0],s[-1][1],len(s))] = (s[-3][0], s[-2][0], s[-1][0], 0, 0,↳
↳s[-1][0].isdigit(), s[-1][0][0].isupper(), 0, 1)
return [(k,v) for k,v in contxt.items()]

def context(text):
    txt = []
    for s in text :
        txt.extend(consent(s))
    return txt

```

```

[11]: context(text[:10])
[(k[0][1],k[1]) for k in context(text[650:1000]) if k[0][1]=="NNP"][:20]

```

```

[11]: [('NNP', ('0', '*T*-2', 'Richard', 'Drobnick', ',', False, True, 0, 0)),
('NNP', (*T*-2', 'Richard', 'Drobnick', ',', 'director', False, True, 0, 0)),
('NNP', ('at', 'the', 'University', 'of', 'Southern', False, True, 0, 0)),
('NNP',
('University', 'of', 'Southern', 'California', 's', False, True, 0, 0)),
('NNP',
('of', 'Southern', 'California', 's', 'Graduate', False, True, 0, 0)),
('NNP', ('California', 's', 'Graduate', 'School', 'of', False, True, 0, 0)),
('NNP', ('s', 'Graduate', 'School', 'of', 'Business', False, True, 0, 0)),
('NNP', ('School', 'of', 'Business', '.', 0, False, True, 0, 0)),
('NNP', (*-1', 'turn', 'Southeast', 'Asia', 'into', False, True, 0, 0)),
('NNP', ('turn', 'Southeast', 'Asia', 'into', 'a', False, True, 0, 0)),

```

```

('NNP',
 ('parallel', 'the', 'European', 'Common', 'Market', False, True, 0, 0)),
('NNP',
 ('the', 'European', 'Common', 'Market', 'approach', False, True, 0, 0)),
('NNP', ('European', 'Common', 'Market', 'approach', '.', False, True, 0, 0)),
('NNP', ('0', '*T*-1', 'Kent', 'Calder', ',', False, True, 0, 0)),
('NNP', (*T*-1, 'Kent', 'Calder', ',', 'a', False, True, 0, 0)),
('NNP', ('at', 'the', 'Woodrow', 'Wilson', 'School', False, True, 0, 0)),
('NNP', ('the', 'Woodrow', 'Wilson', 'School', 'for', False, True, 0, 0)),
('NNP', ('Woodrow', 'Wilson', 'School', 'for', 'Public', False, True, 0, 0)),
('NNP',
 ('School', 'for', 'Public', 'and', 'Internatonal', False, True, 0, 0)),
('NNP',
 ('Public', 'and', 'Internatonal', 'Affairs', 'at', False, True, 0, 0))]

```

```

[12]: freq = 10
def preced2(tagged_sents):
    prec2 = nltk.defaultdict(int)
    precedent2 = nltk.defaultdict(list)
    for t in tagset:
        for i in [k[1] for k in context(tagged_sents) if k[0][1]==t]:
            prec2[i[0]] += 1
        for k,v in prec2.items():
            if v > freq : precedent2[t].append(k)

    return [(k,v) for k,v in precedent2.items()]

```

```

[13]: preced2(text[:100])

```

```

[13]: [('JJ', [0]),
 ('FW', [0]),
 ('WDT', [0]),
 ('VBD', [0, 'the', ',',']),
 ('VBN', [0, 'the', ',',']),
 ('UH', [0, 'the', ',',']),
 ('VB', [0, 'the', ',',']),
 ('CC', [0, 'the', ',',']),
 ('PRP', [0, '*', 'the', ',',']),
 ('SYM', [0, '*', 'the', ',',']),
 ('NNS', ['of', 0, 'to', 'and', 'in', '*', 'the', ',',']),
 ('VBP', ['of', 0, 'to', 'and', 'in', '*', 'the', ',',']),
 ('``', [*-1, 'of', 0, 'to', 'and', 'in', '*', 'the', ',',']),
 ('RP', [*-1, 'of', 0, 'to', 'and', 'in', '*', 'the', ',',']),
 (',', [*-1, 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', ',',']),
 ('NNPS', [*-1, 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', ',',']),
 ('WP', [*-1, 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', ',',']),
 ('VBG', [*-1, 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', ',',']),

```

```

('-RRB-', ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', ',']),
('#', ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', ',']),
('VBZ', ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', 'that', ',']),
('EX', ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', 'that', ',']),
('RB', ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'the', 'that', ',']),
('MD',
 ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'The', 'the', 'that', ',']),
('WRB',
 ['*-1', 'of', 0, 'to', 'and', 'in', '*', 'for', 'The', 'the', 'that', ',']),
('NN',
 ['*-1',
  'of',
  0,
  'a',
  'to',
  'and',
  'in',
  '*',
  'for',
  'The',
  'the',
  'that',
  ',',
  "'s"]]),
('RBS',
 ['*-1',
  'of',
  0,
  'a',
  'to',
  'and',
  'in',
  '*',
  'for',
  'The',
  'the',
  'that',
  ',',
  "'s"]]),
('',
 ['*-1',
  'of',
  0,
  'a',
  'to',
  'and',
  'in',

```

```
'*',
'for',
'The',
'the',
'that',
',',
"'s"]),
('PRP$',
['*-1',
'of',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
',',
"'s"]),
('-NONE-',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
',',
"'s",
'said']),
('CD',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
```

'in',
'*',
'for',
'The',
'the',
'that',
'',
"s",
'from',
'said',
'\$']),
(':',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"s",
'from',
'said',
'\$']),
('POS',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"s",
'from',

'said',
'\$']),
('RBR',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"'s",
'from',
'said',
'\$']),
('WP\$',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"'s",
'from',
'said',
'\$']),
('-LRB-',
['*-1',
'of',
'is',
0,
'a',
'to',

'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"s",
'from',
'said',
'\$']),
('PDT',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"s",
'from',
'said',
'\$']),
('LS',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"s",

```
'from',
'said',
'$']),
('NNP',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
',',
"'s",
'from',
'said',
'at',
'$']),
('.',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
',',
"'s",
'0',
'from',
'by',
'said',
'at',
'$']),
('TO',
['*-1',
```


'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"'s",
'0',
'from',
'by',
'said',
'at',
('\$']),
('\$',
['*-1',
'of',
'is',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'The',
'the',
'that',
'',
"'s",
'0',
'from',
'by',
'said',
'at',
('\$']),
('IN',
['*-1',
'of',
'is',
'with',

0,
'a',
'to',
'and',
'in',
'*',
'for',
'on',
'The',
'the',
'that',
,',
"s",
'0',
'from',
'*T*-1',
'by',
'said',
'at',
('\$']),
('DT',
['*-1',
'of',
'is',
'with',
0,
'a',
'to',
'and',
'in',
'*',
'for',
'on',
'The',
'the',
'that',
,',
"s",
'0',
'%',
'from',
'*T*-1',
'by',
'said',
'at',
('\$']),
('JJR',

```
['*-1',  
 'of',  
 'is',  
 'with',  
 0,  
 'a',  
 'to',  
 'and',  
 'in',  
 '*',  
 'for',  
 'on',  
 'The',  
 'the',  
 'that',  
 ',',  
 "'s",  
 '0',  
 '%',  
 'from',  
 '*T*-1',  
 'by',  
 'said',  
 'at',  
 'was',  
 '$']),  
( 'JJS',  
 ['*-1',  
 'of',  
 'is',  
 'with',  
 0,  
 'a',  
 'to',  
 'and',  
 'in',  
 '*',  
 'for',  
 'on',  
 'The',  
 'the',  
 'that',  
 ',',  
 "'s",  
 '0',  
 '%',  
 'from',
```

```
'*T*-1',  
'by',  
'said',  
'at',  
'was',  
('$']])
```

2 the previous words from two positions

```
[14]: p2 = []  
for i in [v for k,v in preced2([s for s in text if len(s)>5])]:  
    p2.extend(i)  
  
p2 = list(set(p2))  
p2
```

```
[14]: [0,  
    'Bank',  
    'Judge',  
    'study',  
    'know',  
    'minimum',  
    'put',  
    'package',  
    'not',  
    '13',  
    'volatility',  
    'rate',  
    '*RNR*-1',  
    'And',  
    'Johnson',  
    'Lane',  
    'previous',  
    'gains',  
    'Union',  
    'bad',  
    'Herald',  
    'Sea',  
    'latest',  
    'It',  
    'plans',  
    '-RCB-',  
    'unit',  
    'case',  
    'Many',  
    'has',
```

'services',
'across',
'legislation',
'wanted',
'Simmons',
'known',
'age',
'failed',
'decided',
'officer',
'chairman',
'Phelan',
'due',
'far',
'risks',
'operations',
'insurance',
'authority',
'percentage',
'1989',
'development',
'Despite',
'designed',
'sign',
'free',
'Angeles',
'*-2',
'ca',
'By',
'series',
'two',
'junk',
'Big',
'marketing',
'leading',
'Cray',
'closed',
'second',
'old',
'enforcement',
'months',
'increase',
'owns',
'advanced',
'Wedtech',
'Such',
'Scoring',

'past',
'Ltd.',
'11',
'include',
'what',
'Random',
'increased',
'5,000',
'position',
'publishing',
'meet',
'face',
'their',
'least',
'period',
'There',
'month',
'Ford',
'makes',
'18',
'goes',
'Trade',
'lower',
'shareholder',
'sale',
'year',
'convertible',
'make',
'just',
'Campbell',
'food',
'While',
'action',
'declined',
'equity',
'president',
'risk',
'increases',
'generally',
'An',
'from',
'research',
'teachers',
'University',
'look',
'my',
'end',

'his',
'ban',
'London',
'above',
'Federal',
'fell',
'terms',
'earnings',
'being',
'back',
'future',
'Washington',
'financing',
'in',
'purchasing',
'news',
'alleged',
'become',
'total',
'All',
'car',
'be',
'four',
'Stock',
'Industries',
'managers',
'slowing',
'cents',
';',
'School',
'one-time',
'companies',
'number',
'countries',
'-RRB-',
'must',
'administration',
'paper',
'early',
'also',
'five',
'offset',
'holding',
'Group',
'investor',
'became',
'around',

'order',
'Department',
'possible',
'John',
'her',
'15,000',
'Greenville',
'plant',
'been',
'members',
'Although',
'trucks',
'However',
'using',
'Last',
'Dow',
'Commission',
'something',
'World',
'British',
'general',
'following',
'more',
'*?*',
'political',
'banking',
'financial',
'*U*',
'board',
'budget',
'planned',
'expects',
'United',
'return',
'City',
'continued',
'filed',
'of',
'funds',
'by',
'So',
'further',
'private',
'though',
'report',
'little',
'record',

'light',
'week',
'next',
'called',
'who',
'40',
'company',
'recent',
'yield',
'management',
'power',
'Co.',
'payments',
'',
'cut',
'3',
'makers',
'wines',
'cause',
'tax',
'off',
'nearly',
'city',
'states',
'computers',
'decline',
'raised',
'better',
'70',
'probably',
'all',
'International',
'Yeargin',
'Commonwealth',
'read',
'William',
'wine',
'Fund',
'built',
'life',
'that',
'go',
'*-5',
'an',
'A',
'clear',
'PLC',

'wants',
'-LRB-',
'would',
'acquired',
'Congress',
'went',
'lawyers',
'agreed',
'D.',
'greater',
'100',
'well',
'very',
'under',
'value',
'December',
'traded',
'ordered',
'times',
'As',
'world',
'50',
'took',
'nine',
'less',
'court',
'good',
'force',
'York',
'volume',
'appropriations',
'export',
'She',
'*T*-1',
'transactions',
'information',
'sales',
'this',
'find',
'remain',
'family',
'dropped',
'delivery',
'dollar',
'those',
'few',
'No',

'General',
'reason',
'But',
'dealers',
'own',
'problems',
'NL',
'7',
'Wall',
'safety',
'closely',
'paying',
'does',
'In',
'top',
'office',
'Nov.',
'bill',
'1990',
'bond',
'Fujitsu',
'Thomas',
'West',
'lot',
'exchange',
'ended',
'Ward',
'late',
'airline',
'continuing',
'no',
'us',
'strong',
'a',
'real',
'Jones',
'via',
'effort',
'manager',
'ability',
'ringers',
'included',
'Financial',
'issued',
'veto',
'the',
'posted',

'largely',
'local',
'3\\8',
'pressure',
'cars',
'way',
'caused',
'England',
'nation',
'Hahn',
'received',
'1987',
'drop',
'8',
'offered',
'things',
'lead',
'Inc.',
'On',
'Boston',
'Nixon',
'gained',
'Committee',
'your',
'line',
'500',
'feel',
'bid',
'Rep.',
'index',
'half',
'having',
'see',
'really',
'got',
'Richard',
'most',
'fall',
'are',
'one',
'students',
'domestic',
'indicated',
'until',
'talks',
'point',
'many',

'and',
'net',
'For',
'deal',
'fined',
'After',
'commercial',
'or',
'first',
'cases',
'results',
'after',
'American',
'women',
'Treasury',
'say',
'had',
'says',
'part',
'several',
'young',
'Its',
'earlier',
'concern',
'60',
'hard',
'abortion',
'he',
'Fed',
'notes',
'State',
'New',
'fact',
'credit',
'public',
'needed',
'*T*-4',
'if',
'S&P',
'buy-out',
'Chinese',
'steel',
'21',
'concerns',
'according',
'advertisers',
'question',

'she',
'President',
'down',
'began',
"'re",
'Sony',
'issue',
'Dr.',
'fields',
'buyers',
'1991',
'billion',
'any',
'Commerce',
'grain',
'so',
'helped',
'compared',
'were',
'big',
'although',
'improve',
'If',
'may',
'profits',
'than',
'.',
'here',
'1\\2',
'investment',
'6',
'charge',
'*ICH*-1',
'control',
'within',
'One',
'only',
'remains',
'always',
'spending',
'estimated',
'earned',
'individual',
'test',
'debt',
'average',
'country',

'OSHA',
'use',
'90',
'Association',
'apparently',
'takeover',
'oil',
'up',
'economic',
'Coleman',
'Exchange',
'different',
'working',
'March',
'He',
'yen',
'costs',
'group',
'on',
'effect',
'fiscal',
'9',
'into',
'restructuring',
'conditions',
'executive',
'April',
'key',
'view',
'Gulf',
'third',
'traders',
'give',
'particularly',
'investments',
'these',
'analyst',
'still',
'assets',
'director',
'almost',
'interest',
'USX',
'Dec.',
'priced',
'health',
'about',

'Mrs.',
'role',
'changed',
" 'm",
'production',
'institutions',
'California',
'which',
'IRS',
'They',
'as',
'already',
'certain',
'Sales',
'This',
'support',
'September',
'Under',
'data',
'special',
'trading',
'Michael',
'Bernstein',
'whether',
'committee',
'pay',
'Program',
'imports',
'without',
'prices',
'*T*-3',
'Some',
'evidence',
'nor',
'Express',
'form',
'Korea',
'today',
'based',
'To',
'dividend',
'set',
'savings',
'violations',
'law',
'16',
'industrial',

'jobs',
'wo',
'31',
'approved',
'St.',
'with',
'help',
'software',
'underlying',
'Calif.',
'business',
'they',
'increasing',
'named',
'turned',
'right',
'rejected',
's',
'Containers',
'\$',
'change',
'Texas',
'computer',
'home',
'held',
'new',
'That',
'dollars',
'there',
'day',
'transaction',
'led',
'was',
'customers',
'II',
'Shearson',
'relatively',
'auto',
'but',
'USIA',
'noted',
'fine',
'Mary',
'plan',
'orders',
'shareholders',
'system',

'sell',
'how',
'each',
'told',
'major',
'significant',
'ruling',
'James',
'once',
'said',
'Jr.',
'currently',
'given',
'like',
'while',
'days',
'Georgia',
'5\\8',
'scheduled',
'barred',
'process',
'rule',
'Carolina',
'foreign',
'should',
'holders',
'it',
'school',
'making',
'available',
'believes',
'suspension',
'seek',
'years',
'take',
'is',
'researchers',
'October',
'below',
'Securities',
'National',
'June',
'bills',
'market',
'survey',
'Buick',
'did',

'coming',
'First',
'principal',
'raise',
'away',
'review',
'materials',
'questions',
'keep',
'now',
'trying',
'manufacturing',
'our',
'operating',
'newspaper',
'come',
'finance',
'at',
'last',
'U.S.',
'figures',
'Moody',
'full',
'Tuesday',
'construction',
'bonds',
'parent',
'highly',
'problem',
'might',
'levels',
'South',
'Institute',
'10',
'loss',
'goods',
'Hearst',
'same',
'Tokyo',
'capacity',
'growing',
'You',
'Senate',
'long-term',
'another',
'What',
'provide',

'subject',
'sector',
'gives',
'At',
'chief',
'2',
'revenue',
'man',
'tests',
'others',
'show',
'David',
'during',
'rating',
'church',
'through',
'between',
'firms',
'*T*-2',
'Health',
'such',
'Chairman',
'Research',
'sold',
'to',
'workers',
'trades',
'capital',
'antitrust',
'common',
'Justice',
'among',
'found',
'--',
'used',
'stocks',
'Standard',
'agreement',
'America',
'going',
'August',
'Western',
'factory',
'thing',
'department',
'Also',
'Japan',

'#',
'activity',
'growth',
'rise',
'changes',
'serious',
'however',
'stake',
'Street',
"",
'money',
" 've",
""",
'bells',
'idea',
'12',
'Waertsilae',
'believe',
'Court',
'amount',
'anything',
'fees',
'contracts',
':',
'example',
'wage',
'we',
'consented',
'state',
'Management',
'maker',
'investors',
'meeting',
'job',
'*-1',
'We',
'China',
'defense',
'J. ',
'ad',
'Asia',
'Board',
'Ms. ',
'even',
'points',
'3\\4',
'continues',

'plants',
'200',
'best',
'long',
'when',
'much',
'them',
'including',
'pence',
'current',
'card',
'Mitsubishi',
'drug',
'&',
'boost',
'1985',
'spokesman',
'shares',
'trade',
'nations',
'buying',
'decision',
'likely',
'recently',
'time',
'5',
'against',
'vice',
'*ICH*-2',
'launched',
'30-year',
'came',
'too',
'*ICH*-3',
'losses',
'every',
'Among',
'involved',
'proposal',
'have',
'Wednesday',
'team',
'advertising',
'Mr.',
'efforts',
'low',
'High',

'kind',
'loan',
'rates',
'Dinkins',
'potential',
'25',
'Japanese',
'million',
'close',
'whose',
'corn',
'yet',
'get',
'large',
'performance',
'-LCB-',
'important',
'weeks',
'lost',
'because',
'you',
'higher',
'*-3',
'officials',
'work',
'campaign',
'employees',
'thrift',
'followed',
'saying',
'three',
'paid',
'personal',
'single',
'composite',
'able',
'contract',
'15',
'third-quarter',
'stock-index',
'think',
'annual',
'T.',
'could',
'hopes',
'program',
'watch',

'do',
'Hills',
'for',
'senior',
'despite',
'slow',
'official',
'some',
'0',
'short',
'result',
'out',
'share',
'products',
'reports',
'people',
'quarter',
'Robert',
'includes',
'need',
'can',
'Bush',
'teacher',
'international',
'often',
'dividends',
'rather',
'Francisco',
'cost',
'firm',
'seeking',
'Columbia',
'San',
'regulators',
'Valley',
'10,000',
'farmers',
'head',
'receive',
'charges',
'I',
'situation',
'former',
'agency',
'continue',
'short-term',
'course',


```
'banks',
'fully',
'departure',
'expected',
'analysts',
'magazine',
'enough',
'him',
'preferred',
'high',
'R.',
'over',
'level',
'30',
'showed',
'announced',
'UAL',
'%',
'limit',
'division',
'reduce',
'then',
'prevent',
'largest',
'rights',
'Drexel',
'May',
'bank',
'since',
'corporate',
...]
```

```
[15]: def preced1(tagged_sents):
    prec1 = nltk.defaultdict(int)
    precedent1 = nltk.defaultdict(list)
    for t in tagset:
        for i in [k[1] for k in context(tagged_sents) if k[0][1]==t]:
            prec1[i[1]] += 1
        for k,v in prec1.items() :
            if v > freq : precedent1[t].append(k)

    return [(k,v) for k,v in precedent1.items()]
```

3 Previous words from one position

```
[16]: p1 = []
      for i in [v for k,v in preced1([s for s in text if len(s)>5])]:
          p1.extend(i)

      p1 = list(set(p1))
      p1
```

```
[16]: [0,
      'Bank',
      'Judge',
      'study',
      'know',
      'minimum',
      'conference',
      'put',
      'package',
      'not',
      '13',
      'volatility',
      '*RNR*-1',
      'rate',
      'And',
      'Johnson',
      'Corp',
      'Lane',
      'previous',
      'gains',
      'Union',
      'bad',
      'Sea',
      'latest',
      'It',
      '-RCB-',
      'unit',
      'case',
      'Many',
      'has',
      'services',
      'across',
      'legislation',
      'wanted',
      'Simmons',
      'known',
      'age',
      'failed',
```

'decided',
'win',
'officer',
'chairman',
'Phelan',
'due',
'far',
'risks',
'operations',
'insurance',
'authority',
'percentage',
'1989',
'development',
'Despite',
'designed',
'bottle',
'sign',
'free',
'Angeles',
'publicly',
'*-2',
'ca',
'rather',
'series',
'two',
'junk',
'Big',
'By',
'marketing',
'leading',
'second',
'closed',
'Cray',
'old',
'enforcement',
'months',
'increase',
'owns',
'Constitution',
'Wedtech',
'advanced',
'Scoring',
'Such',
'past',
'Ltd.',
'11',

'Random',
'what',
'include',
'increased',
'5,000',
'overseas',
'member',
'position',
'publishing',
'meet',
'face',
'their',
'least',
'period',
'month',
'There',
'Ford',
'makes',
'18',
'Trade',
'goes',
'lower',
'shareholder',
'sale',
'year',
'convertible',
'make',
'just',
'measure',
'food',
'Campbell',
'action',
'While',
'equity',
'declined',
'president',
'risk',
'increases',
'generally',
'An',
'from',
'research',
'teachers',
'University',
'options',
'look',
'my',

'end',
'his',
'ban',
'soon',
'London',
'above',
'Federal',
'fell',
'terms',
'being',
'earnings',
'back',
'future',
'Washington',
'financing',
'in',
'purchasing',
'news',
'?',
'become',
'total',
'All',
'car',
'be',
'four',
'Stock',
'Industries',
'managers',
'slowing',
'cents',
';',
'School',
'one-time',
'companies',
'number',
'countries',
'-RRB-',
'must',
'administration',
'paper',
'early',
'also',
'five',
'offset',
'holding',
'Group',
'investor',

'became',
'around',
'order',
'Department',
'possible',
'Greenville',
'her',
'15,000',
'John',
'plant',
'been',
'quickly',
'Although',
'trucks',
'However',
'using',
'Last',
'Dow',
'Commission',
'World',
'something',
'British',
'general',
'following',
'more',
'*?*',
'political',
'banking',
'financial',
'*U*',
'board',
'budget',
'planned',
'expects',
'City',
'United',
'further',
'continued',
'filed',
'funds',
'of',
'by',
'So',
'Jones',
'return',
'private',
'January',

'though',
'report',
'little',
'record',
'light',
'week',
'next',
'40',
'who',
'called',
'company',
'Inc',
'exports',
'recent',
'yield',
'management',
'power',
'Co.',
'payments',
'',
'cut',
'3',
'makers',
'wines',
'telephone',
'tax',
'cause',
'off',
'nearly',
'computers',
'70',
'city',
'decline',
'states',
'better',
'raised',
'probably',
'all',
'International',
'Yeargin',
'Commonwealth',
'paying',
'read',
'wine',
'built',
'William',
'Fund',

'life',
'that',
'go',
'*-5',
'an',
'A',
'clear',
'PLC',
'agreed',
'-LRB-',
'would',
'lawyers',
'Congress',
'acquired',
'greater',
'went',
'D.',
'wants',
'100',
'well',
'very',
'under',
'December',
'value',
'traded',
'rising',
'ordered',
'times',
'As',
'world',
'50',
'took',
'nine',
'less',
'court',
'good',
'force',
'York',
'volume',
'appropriations',
'export',
'She',
'*T*-1',
'transactions',
'information',
'sales',
'this',

'find',
'region',
'family',
'remain',
'speculation',
'dropped',
'delivery',
'dollar',
'those',
'few',
'homeless',
'General',
'No',
'reason',
'But',
'dealers',
'own',
'NL',
'problems',
'closely',
'safety',
'top',
'7',
'believe',
'does',
'In',
'Wall',
'office',
'bill',
'Nov.',
'1990',
'bond',
'Fujitsu',
'Thomas',
'West',
'lot',
'7\\8',
'exchange',
'ended',
'Ward',
'late',
'airline',
'us',
'no',
'continuing',
'strong',
'a',

'real',
'issued',
'ability',
'via',
'manager',
'Herald',
'ringers',
'included',
'veto',
'Financial',
'largely',
'the',
'editor',
'posted',
'local',
'3\\8',
'pressure',
'cars',
'way',
'caused',
'received',
'1987',
'Hahn',
'offered',
'England',
'8',
'nation',
'things',
'drop',
'lead',
'Inc.',
'Boston',
'On',
'Nixon',
'gained',
'Committee',
'your',
'feel',
'500',
'line',
'bid',
'Rep.',
'index',
'half',
'having',
'see',
'really',

'got',
'history',
'Richard',
'most',
'fall',
'are',
'one',
'students',
'domestic',
'indicated',
'until',
'talks',
'point',
'many',
'and',
'net',
'For',
'fined',
'deal',
'After',
'commercial',
'or',
'first',
'cases',
'results',
'after',
'American',
'women',
'Treasury',
'say',
'had',
'says',
'part',
'several',
'young',
'Its',
'earlier',
'concern',
'members',
'60',
'hard',
'abortion',
'he',
'Fed',
'notes',
'State',
'crash',

'New',
'fact',
'credit',
'public',
'needed',
'*T*-4',
'if',
'S&P',
'buy-out',
'Chinese',
'steel',
'21',
'concerns',
'according',
'advertisers',
'question',
'she',
'President',
'down',
'began',
're",
'1\\4',
'Sony',
'issue',
'Dr.',
'fields',
'Monday',
'buyers',
'1991',
'billion',
'any',
'Commerce',
'grain',
'so',
'helped',
'strike',
'were',
'big',
'improve',
'compared',
'term',
'although',
'If',
'may',
'profits',
'than',
'.',

'here',
'1\\2',
'district',
'investment',
'6',
'charge',
'*ICH*-1',
'control',
'within',
'One',
'only',
'remains',
'always',
'spending',
'estimated',
'earned',
'individual',
'test',
'debt',
'average',
'country',
'OSHA',
'use',
'90',
'Association',
'apparently',
'oil',
'takeover',
'up',
'economic',
'Coleman',
'Exchange',
'different',
'working',
'yen',
'He',
'March',
'costs',
'group',
'on',
'effect',
'fiscal',
'9',
'into',
'restructuring',
'conditions',
'executive',

'April',
'key',
'view',
'third',
'Gulf',
'traders',
'give',
'particularly',
'investments',
'these',
'analyst',
'still',
'assets',
'director',
'almost',
'interest',
'USX',
'Dec.',
'priced',
'health',
'about',
'Mrs.',
'changed',
'role',
'm',
'production',
'institutions',
'California',
'which',
'IRS',
'They',
'as',
'already',
'certain',
'Sales',
'This',
'September',
'plans',
'support',
'data',
'Under',
'special',
'trading',
'Michael',
'committee',
'whether',
'Bernstein',

'pay',
'1\\8',
'Program',
'imports',
'without',
'prices',
'*T*-3',
'Some',
'evidence',
'answers',
'matter',
'nor',
'Express',
'returns',
'form',
'Korea',
'based',
'today',
'savings',
'dividend',
'set',
'To',
'violations',
'law',
'16',
'effort',
'industrial',
'jobs',
'wo',
'31',
'approved',
'St.',
'with',
'help',
'underlying',
'software',
'Calif.',
'business',
'they',
'increasing',
'named',
'turned',
'right',
'rejected',
's',
'\$',
'Containers',

'change',
'Texas',
'computer',
'home',
'held',
'new',
'That',
'dollars',
'there',
'day',
'transaction',
'led',
'was',
'customers',
'II',
'Shearson',
'relatively',
'auto',
'but',
'USIA',
'noted',
'fine',
'Mary',
'plan',
'orders',
'shareholders',
'system',
'sell',
'how',
'each',
'significant',
'major',
'ruling',
'told',
'James',
'once',
'said',
'Jr.',
'currently',
'given',
'days',
'while',
'like',
'scheduled',
'Georgia',
'5\\8',
'process',

'barred',
'rule',
'Carolina',
'should',
'foreign',
'gives',
'holders',
'it',
'making',
'school',
'themselves',
'available',
'believes',
'suspension',
'seek',
'years',
'take',
'is',
'researchers',
'October',
'below',
'Securities',
'National',
'June',
'bills',
'market',
'survey',
'Buick',
'did',
'coming',
'First',
'principal',
'raise',
'away',
'review',
'materials',
'questions',
'trying',
'now',
'keep',
'manufacturing',
'our',
'operating',
'newspaper',
'come',
'finance',
'at',

'last',
'U.S.',
'figures',
'Moody',
'full',
'Tuesday',
'construction',
'bonds',
'parent',
'highly',
'problem',
'might',
'levels',
'South',
'Institute',
'10',
'goods',
'loss',
'Hearst',
'Tokyo',
'same',
'capacity',
'growing',
'You',
'Senate',
'long-term',
'another',
'What',
'subject',
'provide',
'sector',
'revenue',
'chief',
'At',
'2',
'man',
'tests',
'others',
'show',
'David',
'during',
'rating',
'church',
'through',
'between',
'firms',
'*T*-2',

'Health',
'such',
'Chairman',
'Research',
'sold',
'to',
'workers',
'trades',
'capital',
'antitrust',
'common',
'Justice',
'among',
'found',
'--',
'used',
'stocks',
'Standard',
'agreement',
'selling',
'act',
'America',
'factory',
'going',
'Western',
'August',
'thing',
'department',
'Also',
'Japan',
'activity',
'#',
'growth',
'...',
'rise',
'changes',
'serious',
'however',
'stake',
'Street',
"" ,
" 've",
'money',
"" ,
'bells',
'idea',
'12',

'Waertsilae',
'Court',
'drugs',
'amount',
'anything',
'fees',
'contracts',
'similar',
':',
'example',
'wage',
'we',
'consented',
'state',
'Management',
'maker',
'investors',
'job',
'meeting',
'Co',
'*-1',
'We',
'defense',
'China',
'J.',
'ad',
'Asia',
'Board',
'points',
'even',
'3\\4',
'Ms.',
'best',
'asbestos',
'200',
'plants',
'long',
'continues',
'when',
'much',
'them',
'Georgia-Pacific',
'including',
'pence',
'current',
'card',
'Mitsubishi',

'drug',
'&',
'boost',
'1985',
'spokesman',
'shares',
'trade',
'decision',
'likely',
'nations',
'14',
'buying',
'recently',
'time',
'5',
'against',
'vice',
'*ICH*-2',
'launched',
'30-year',
'came',
'benefits',
'too',
'losses',
'*ICH*-3',
'every',
'Among',
'involved',
'proposal',
'have',
'Wednesday',
'team',
'advertising',
'Mr.',
'efforts',
'River',
'low',
'loan',
'High',
'kind',
'rates',
'Dinkins',
'potential',
'25',
'person',
'Japanese',
'million',

'close',
'whose',
'yet',
'corn',
'get',
'large',
'performance',
'-LCB-',
'important',
'weeks',
'lost',
'you',
'because',
'higher',
'*-3',
'officials',
'campaign',
'work',
'employees',
'thrift',
'followed',
'saying',
'three',
'personal',
'paid',
'single',
'composite',
'able',
'contract',
'15',
'scores',
'third-quarter',
'stock-index',
'think',
'annual',
'T.',
'could',
'hopes',
'program',
'watch',
'do',
'Hills',
'for',
'senior',
'despite',
'slow',
'official',

```
'some',
'0',
'result',
'short',
'out',
'share',
'products',
'reports',
'people',
'quarter',
'Robert',
'includes',
'need',
'can',
'Bush',
'teacher',
'international',
'often',
'dividends',
'alleged',
'Francisco',
'seeking',
...]
```

```
[17]: def current(tagged_sents):
    cur = nltk.defaultdict(int)
    curr = nltk.defaultdict(list)
    for t in tagset:
        for i in [k[1] for k in context(tagged_sents) if k[0][1]==t]:
            cur[i[2]] += 1
        for k,v in cur.items() :
            if v > freq : curr[t].append(k)

    return [(k,v) for k,v in curr.items()]
```

4 Current words

```
[18]: c = []
for i in [v for k,v in current([s for s in text if len(s)>5])]:
    c.extend(i)

c = list(set(c))
c
```

```
[18]: ['study',
'Bank',
```

'Judge',
'know',
'minimum',
'conference',
'put',
'package',
'not',
'13',
'volatility',
'rate',
'*RNR*-1',
'And',
'Johnson',
'Corp',
'Lane',
'previous',
'gains',
'Union',
'bad',
'Herald',
'-RCB-',
'Sea',
'It',
'latest',
'unit',
'case',
'Many',
'services',
'has',
'across',
'wanted',
'legislation',
'Simmons',
'known',
'age',
'failed',
'decided',
'win',
'officer',
'chairman',
'Phelan',
'due',
'far',
'risks',
'operations',
'authority',
'percentage',

'insurance',
'1989',
'development',
'Despite',
'designed',
'bottle',
'sign',
'free',
'publicly',
'Angeles',
'series',
'ca',
'*-2',
'By',
'two',
'junk',
'Big',
'marketing',
'leading',
'second',
'closed',
'Cray',
'old',
'enforcement',
'increase',
'months',
'advanced',
'owns',
'Constitution',
'Such',
'Scoring',
'Wedtech',
'past',
'Ltd.',
'11',
'include',
'what',
'Random',
'increased',
'overseas',
'5,000',
'member',
'position',
'publishing',
'meet',
'face',
'their',

'least',
'period',
'makes',
'There',
'month',
'Ford',
'18',
'goes',
'Trade',
'lower',
'shareholder',
'sale',
'convertible',
'year',
'make',
'just',
'measure',
'risk',
'food',
'declined',
'equity',
'action',
'Campbell',
'president',
'While',
'increases',
'generally',
'An',
'from',
'research',
'teachers',
'University',
'options',
'look',
'my',
'soon',
'end',
'ban',
'his',
'London',
'above',
'Federal',
'fell',
'terms',
'earnings',
'being',
'back',

'future',
'Washington',
'financing',
'in',
'purchasing',
'news',
'?',
'total',
'become',
'All',
'car',
'be',
'four',
'Stock',
'Industries',
'before',
'managers',
'slowing',
'cents',
';',
'School',
'one-time',
'companies',
'number',
'countries',
'-RRB-',
'must',
'paper',
'administration',
'early',
'also',
'offset',
'five',
'holding',
'Group',
'became',
'investor',
'around',
'order',
'Department',
'possible',
'members',
'her',
'15,000',
'John',
'quickly',
'been',

'plant',
'Greenville',
'trucks',
'However',
'Although',
'using',
'Last',
'Dow',
'Commission',
'something',
'World',
'British',
'general',
'following',
'more',
'*?*',
'political',
'banking',
'financial',
'board',
'budget',
'*U*',
'planned',
'expects',
'United',
'return',
'further',
'continued',
'filed',
'funds',
'So',
'City',
'Jones',
'by',
'of',
'private',
'January',
'though',
'report',
'little',
'record',
'light',
'week',
'next',
'called',
'exports',
'who',

'company',
'40',
'Inc',
'recent',
'yield',
'management',
'power',
'Co.',
'payments',
'cut',
'makers',
',',
'3',
'wines',
'telephone',
'cause',
'tax',
'off',
'nearly',
'computers',
'raised',
'states',
'decline',
'city',
'70',
'better',
'probably',
'all',
'International',
'Yeargin',
'read',
'Commonwealth',
'life',
'built',
'wine',
'Fund',
'*-5',
'that',
'go',
'William',
'an',
'clear',
'A',
'PLC',
'agreed',
'wants',
'acquired',

'went',
'lawyers',
'would',
'-LRB-',
'Congress',
'D.',
'greater',
'100',
'well',
'very',
'under',
'value',
'December',
'traded',
'rising',
'ordered',
'times',
'world',
'As',
'50',
'took',
'nine',
'less',
'good',
'force',
'court',
'volume',
'York',
'appropriations',
'export',
'She',
'transactions',
'*T*-1',
'information',
'remain',
'sales',
'find',
'region',
'family',
'this',
'speculation',
'dropped',
'delivery',
'dollar',
'those',
'few',
'homeless',

'General',
'No',
'reason',
'But',
'dealers',
'own',
'problems',
'NL',
'believe',
'paying',
'top',
'closely',
'safety',
'does',
'7',
'Wall',
'office',
'bill',
'Nov.',
'In',
'1990',
'bond',
'Fujitsu',
'Thomas',
'West',
'lot',
'7\\8',
'exchange',
'ended',
'Ward',
'late',
'airline',
'us',
'continuing',
'no',
'strong',
'real',
'losses',
'issued',
'fourth',
'effort',
'largely',
'manager',
'included',
'posted',
'ringers',
'veto',

'editor',
'Financial',
'the',
'local',
'3\\8',
'pressure',
'cars',
'way',
'caused',
'offered',
'received',
'nation',
'1987',
'things',
'drop',
'8',
'England',
'Hahn',
'lead',
'Inc.',
'Boston',
'On',
'Nixon',
'gained',
'Committee',
'feel',
'line',
'your',
'500',
'bid',
'Rep.',
'index',
'half',
'having',
'see',
'really',
'got',
'history',
'Richard',
'most',
'fall',
'are',
'students',
'one',
'domestic',
'indicated',
'talks',

'until',
'point',
'many',
'net',
'and',
'After',
'fined',
'deal',
'For',
'commercial',
'or',
'first',
'cases',
'results',
'after',
'American',
'women',
'say',
'Treasury',
'had',
'says',
'part',
'several',
'young',
'Its',
'earlier',
'concern',
'hard',
'60',
'abortion',
'he',
'Fed',
'notes',
'State',
'crash',
'New',
'fact',
'credit',
'public',
'needed',
'*T*-4',
'if',
'buy-out',
'S&P',
'Chinese',
'steel',
'21',

'concerns',
'according',
'advertisers',
'question',
'she',
'President',
'began',
'down',
're",
'1\\4',
'Sony',
'fields',
'issue',
'Dr.',
'Monday',
'buyers',
'1991',
'billion',
'any',
'Commerce',
'grain',
'so',
'helped',
'compared',
'were',
'big',
'improve',
'strike',
'term',
'although',
'If',
'may',
'profits',
'district',
'here',
'1\\2',
'.',
'than',
'investment',
'6',
'charge',
'control',
'*ICH*-1',
'within',
'One',
'remains',
'only',

'always',
'spending',
'estimated',
'earned',
'individual',
'test',
'debt',
'average',
'country',
'OSHA',
'use',
'90',
'Association',
'apparently',
'oil',
'takeover',
'economic',
'up',
'Coleman',
'Exchange',
'different',
'working',
'yen',
'He',
'March',
'costs',
'group',
'on',
'effect',
'fiscal',
'9',
'into',
'restructuring',
'conditions',
'executive',
'April',
'key',
'view',
'third',
'Gulf',
'give',
'traders',
'particularly',
'investments',
'these',
'analyst',
'still',

'assets',
'director',
'almost',
'interest',
'USX',
'Dec.',
'priced',
'health',
'about',
'Mrs.',
'role',
'm',
'production',
'institutions',
'California',
'which',
'plans',
'They',
'data',
'as',
'certain',
'Sales',
'already',
'support',
'September',
'IRS',
'Under',
'This',
'special',
'trading',
'Michael',
'committee',
'Bernstein',
'whether',
'pay',
'1\\8',
'Program',
'imports',
'without',
'prices',
'*T*-3',
'Some',
'nor',
'answers',
'matter',
'evidence',
'returns',

'Express',
'form',
'violations',
'based',
'today',
'savings',
'dividend',
'set',
'Korea',
'To',
'law',
'16',
'industrial',
'jobs',
'wo',
'31',
'approved',
'St.',
'with',
'turned',
'help',
'underlying',
'named',
'increasing',
'they',
'business',
'software',
'Calif.',
'right',
'rejected',
'Containers',
's",
'\$',
'change',
'Texas',
'home',
'computer',
'held',
'new',
'That',
'dollars',
'there',
'day',
'transaction',
'led',
'was',
'customers',

'II',
'Shearson',
'relatively',
'auto',
'but',
'USIA',
'noted',
'fine',
'Mary',
'plan',
'orders',
'shareholders',
'system',
'sell',
'how',
'each',
'ability',
'significant',
'major',
'told',
'ruling',
'James',
'once',
'said',
'Jr.',
'currently',
'given',
'barred',
'days',
'scheduled',
'like',
'5\\8',
'Georgia',
'process',
'while',
'rule',
'Carolina',
'foreign',
'should',
'holders',
'it',
'making',
'school',
'themselves',
'available',
'believes',
'suspension',

'seek',
'take',
'years',
'is',
'researchers',
'National',
'Securities',
'October',
'below',
'June',
'bills',
'review',
'market',
'coming',
'did',
'survey',
'First',
'principal',
'raise',
'away',
'Buick',
'materials',
'questions',
'keep',
'trying',
'now',
'manufacturing',
'operating',
'newspaper',
'our',
'come',
'finance',
'last',
'figures',
'U.S.',
'at',
'Moody',
'full',
'parent',
'construction',
'bonds',
'Tuesday',
'highly',
'problem',
'might',
'levels',
'South',

'Institute',
'loss',
'goods',
'10',
'Hearst',
'same',
'Tokyo',
'capacity',
'long-term',
'You',
'growing',
'gives',
'subject',
'What',
'provide',
'Senate',
'sector',
'revenue',
'chief',
'At',
'2',
'another',
'man',
'tests',
'others',
'show',
'David',
'during',
'rating',
'church',
'through',
'between',
'firms',
'*T*-2',
'Health',
'such',
'via',
'Chairman',
'Research',
'sold',
'to',
'workers',
'a',
'trades',
'capital',
'antitrust',
'common',

'Justice',
'found',
'among',
'--',
'used',
'stocks',
'agreement',
'selling',
'Standard',
'act',
'August',
'going',
'factory',
'Western',
'America',
'thing',
'department',
'Also',
'#',
'activity',
'Japan',
'rise',
'growth',
'...',
'changes',
'serious',
'however',
'stake',
'Street',
" """,
" 've",
'money',
" """,
'bells',
'idea',
'12',
'Waertsilae',
'drugs',
'Court',
'amount',
'anything',
'fees',
'contracts',
'similar',
'example',
':',
'wage',

'we',
'consented',
'state',
'Management',
'maker',
'investors',
'meeting',
'job',
'Co',
'*-1',
'We',
'defense',
'China',
'J.',
'ad',
'Asia',
'points',
'Board',
'even',
'plants',
'asbestos',
'continues',
'3\\4',
'200',
'Ms.',
'long',
'much',
'when',
'them',
'Georgia-Pacific',
'including',
'pence',
'current',
'card',
'Mitsubishi',
'drug',
'&',
'boost',
'1985',
'spokesman',
'nations',
'likely',
'shares',
'trade',
'buying',
'decision',
'14',

'recently',
'time',
'5',
'against',
'vice',
'launched',
'*ICH*-2',
'30-year',
'came',
'benefits',
'too',
'*ICH*-3',
'every',
'Among',
'involved',
'proposal',
'have',
'Wednesday',
'advertising',
'efforts',
'team',
'Mr.',
'River',
'low',
'Columbia',
'kind',
'loan',
'High',
'rates',
'Dinkins',
'potential',
'person',
'25',
'Japanese',
'million',
'close',
'whose',
'yet',
'corn',
'get',
'large',
'lost',
'performance',
'important',
'weeks',
'-LCB-',
'you',

'because',
'higher',
'campaign',
'officials',
'work',
'*-3',
'employees',
'thrift',
'followed',
'saying',
'three',
'personal',
'paid',
'single',
'composite',
'able',
'scores',
'contract',
'15',
'third-quarter',
'think',
'stock-index',
'annual',
'T.',
'hopes',
'could',
'program',
'watch',
'do',
'Hills',
'senior',
'for',
'despite',
'slow',
'result',
'official',
'short',
'0',
'some',
'out',
'products',
'reports',
'share',
'people',
'quarter',
'Robert',
'includes',

```
'need',
'can',
'Bush',
'teacher',
'international',
'changed',
'dividends',
'alleged',
'seeking',
'rather',
'often',
...]
```

```
[19]: def follow1(tagged_sents):
      fol = nltk.defaultdict(int)
      foll = nltk.defaultdict(list)
      for t in tagset:
          for i in [k[1] for k in context(tagged_sents) if k[0][1]==t]:
              fol[i[3]] += 1
          for k,v in fol.items() :
              if v > freq : foll[t].append(k)

      return [(k,v) for k,v in foll.items()]

f1 = []
for i in [v for k,v in follow1([s for s in text if len(s)>5])]:
    f1.extend(i)

f1 = list(set(f1))
f1
```

```
[19]: [0,
'Bank',
'Judge',
'study',
'know',
'minimum',
'conference',
'put',
'package',
'not',
'13',
'volatility',
'rate',
'*RNR*-1',
'Johnson',
'Corp',
```

'Lane',
'previous',
'gains',
'Union',
'bad',
'Herald',
'-RCB-',
'Sea',
'It',
'latest',
'unit',
'case',
'has',
'services',
'across',
'legislation',
'wanted',
'Simmons',
'known',
'age',
'failed',
'decided',
'win',
'officer',
'chairman',
'Phelan',
'due',
'far',
'risks',
'operations',
'insurance',
'authority',
'percentage',
'1989',
'development',
'designed',
'bottle',
'sign',
'free',
'Angeles',
'publicly',
'*-2',
'ca',
'series',
'two',
'junk',
'Big',

'marketing',
'leading',
'closed',
'Cray',
'second',
'old',
'enforcement',
'months',
'increase',
'owns',
'advanced',
'Constitution',
'Scoring',
'past',
'Ltd.',
'11',
'include',
'what',
'Random',
'increased',
'5,000',
'overseas',
'member',
'position',
'publishing',
'meet',
'face',
'their',
'least',
'period',
'month',
'There',
'makes',
'Ford',
'18',
'goes',
'Trade',
'lower',
'shareholder',
'sale',
'year',
'convertible',
'make',
'just',
'measure',
'food',
'Campbell',

'action',
'declined',
'equity',
'risk',
'president',
'generally',
'increases',
'from',
'research',
'teachers',
'University',
'options',
'look',
'my',
'soon',
'his',
'end',
'ban',
'above',
'London',
'Federal',
'fell',
'terms',
'earnings',
'being',
'back',
'future',
'Washington',
'financing',
'in',
'news',
'purchasing',
'?',
'alleged',
'become',
'total',
'car',
'be',
'four',
'Stock',
'Industries',
'managers',
'slowing',
'cents',
';',
'School',
'one-time',

'companies',
'number',
'countries',
'-RRB-',
'must',
'paper',
'administration',
'early',
'also',
'five',
'offset',
'holding',
'Group',
'investor',
'became',
'around',
'order',
'Department',
'possible',
'members',
'her',
'John',
'15,000',
'quickly',
'been',
'plant',
'Greenville',
'trucks',
'using',
'Dow',
'Commission',
'something',
'World',
'British',
'general',
'following',
'more',
'*?*',
'political',
'banking',
'financial',
'*U*',
'board',
'budget',
'planned',
'expects',
'City',

'return',
'further',
'filed',
'continued',
'of',
'funds',
'by',
'United',
'private',
'January',
'though',
'report',
'little',
'record',
'light',
'week',
'next',
'called',
'who',
'40',
'company',
'Inc',
'exports',
'yield',
'recent',
'management',
'power',
'Co.',
'payments',
',',
'cut',
'makers',
'3',
'wines',
'telephone',
'tax',
'cause',
'off',
'nearly',
'70',
'raised',
'states',
'decline',
'computers',
'better',
'city',
'probably',

'all',
'International',
'Yeargin',
'read',
'William',
'built',
'life',
'Fund',
'*-5',
'that',
'go',
'wine',
'an',
'A',
'clear',
'PLC',
'agreed',
'-LRB-',
'would',
'acquired',
'Congress',
'went',
'wants',
'lawyers',
'D.',
'greater',
'100',
'well',
'very',
'under',
'value',
'December',
'traded',
'rising',
'ordered',
'times',
'world',
'50',
'took',
'nine',
'less',
'good',
'court',
'force',
'volume',
'York',
'appropriations',

'export',
'transactions',
'*T*-1',
'information',
'sales',
'this',
'find',
'remain',
'family',
'region',
'speculation',
'dropped',
'delivery',
'dollar',
'those',
'few',
'homeless',
'General',
'reason',
'dealers',
'own',
'problems',
'NL',
'believe',
'safety',
'Wall',
'paying',
'7',
'does',
'closely',
'top',
'office',
'Nov.',
'bill',
'1990',
'bond',
'Thomas',
'lot',
'7\\8',
'exchange',
'ended',
'Ward',
'late',
'airline',
'us',
'no',
'via',

'strong',
'a',
'*ICH*-3',
'issued',
'Jones',
'losses',
'manager',
'largely',
'included',
'posted',
'veto',
'Financial',
'ringers',
'the',
'editor',
'local',
'3\\8',
'pressure',
'cars',
'way',
'caused',
'offered',
'received',
'1987',
'England',
'things',
'8',
'Hahn',
'drop',
'nation',
'lead',
'Inc.',
'Boston',
'Nixon',
'gained',
'Committee',
'your',
'line',
'500',
'feel',
'bid',
'Rep.',
'index',
'half',
'having',
'see',
'really',

'got',
'history',
'Richard',
'most',
'fall',
'are',
'one',
'students',
'domestic',
'indicated',
'until',
'talks',
'point',
'net',
'and',
'many',
'fined',
'deal',
'cases',
'or',
'first',
'results',
'commercial',
'after',
'American',
'women',
'say',
'Treasury',
'had',
'says',
'part',
'several',
'young',
'earlier',
'concern',
'60',
'hard',
'abortion',
'he',
'Fed',
'notes',
'State',
'crash',
'New',
'fact',
'credit',
'public',

'needed',
'*T*-4',
'if',
'buy-out',
'S&P',
'Chinese',
'steel',
'21',
'concerns',
'according',
'advertisers',
'question',
'she',
'President',
'down',
'began',
'"re",
'1\\4',
'issue',
'fields',
'Monday',
'buyers',
'1991',
'billion',
'any',
'Commerce',
'grain',
'so',
'helped',
'compared',
'were',
'although',
'term',
'big',
'improve',
'strike',
'If',
'may',
'profits',
'than',
'.',
'here',
'1\\2',
'district',
'investment',
'6',
'charge',

'*ICH*-1',
'control',
'within',
'only',
'remains',
'always',
'spending',
'estimated',
'earned',
'individual',
'test',
'debt',
'average',
'country',
'use',
'90',
'Association',
'apparently',
'oil',
'takeover',
'up',
'economic',
'Coleman',
'Exchange',
'different',
'working',
'yen',
'March',
'costs',
'group',
'on',
'effect',
'9',
'fiscal',
'into',
'restructuring',
'conditions',
'executive',
'April',
'key',
'view',
'Gulf',
'third',
'traders',
'give',
'particularly',
'investments',

'these',
'analyst',
'still',
'assets',
'director',
'almost',
'interest',
'USX',
'Dec.',
'priced',
'health',
'about',
'Mrs.',
'role',
'm',
'production',
'institutions',
'California',
'plans',
'which',
'IRS',
'as',
'already',
'certain',
'data',
'This',
'support',
'September',
'special',
'trading',
'Michael',
'Bernstein',
'whether',
'committee',
'pay',
'1\\8',
'imports',
'without',
'*T*-3',
'prices',
'nor',
'evidence',
'answers',
'matter',
'Express',
'returns',
'form',

'violations',
'based',
'today',
'To',
'dividend',
'set',
'Korea',
'savings',
'law',
'16',
'effort',
'industrial',
'jobs',
'wo',
'31',
'approved',
'St.',
'with',
'help',
'software',
'turned',
'named',
'business',
'they',
'Calif.',
'increasing',
'underlying',
'right',
'rejected',
'\$',
"'s",
'Containers',
'change',
'Texas',
'computer',
'home',
'held',
'new',
'dollars',
'there',
'day',
'transaction',
'led',
'was',
'customers',
'II',
'Shearson',

'relatively',
'auto',
'but',
'USIA',
'noted',
'fine',
'plan',
'system',
'orders',
'shareholders',
'sell',
'how',
'each',
'ability',
'told',
'major',
'significant',
'ruling',
'James',
'once',
'said',
'Jr.',
'currently',
'given',
'like',
'while',
'scheduled',
'days',
'5\\8',
'Georgia',
'process',
'barred',
'rule',
'Carolina',
'should',
'foreign',
'holders',
'it',
'making',
'school',
'themselves',
'available',
'believes',
'suspension',
'seek',
'years',
'take',

'is',
'researchers',
'October',
'below',
'Securities',
'National',
'June',
'bills',
'market',
'review',
'Buick',
'did',
'survey',
'First',
'principal',
'raise',
'away',
'coming',
'materials',
'questions',
'trying',
'now',
'keep',
'manufacturing',
'our',
'operating',
'newspaper',
'come',
'finance',
'at',
'last',
'U.S.',
'figures',
'Moody',
'full',
'Tuesday',
'construction',
'bonds',
'parent',
'highly',
'problem',
'might',
'levels',
'South',
'Institute',
'loss',
'10',

'goods',
'Hearst',
'Tokyo',
'same',
'capacity',
'growing',
'You',
'gives',
'long-term',
'another',
'Senate',
'provide',
'subject',
'sector',
'revenue',
'chief',
'2',
'man',
'tests',
'others',
'show',
'David',
'during',
'rating',
'church',
'through',
'between',
'firms',
'*T*-2',
'Health',
'such',
'Chairman',
'Research',
'sold',
'to',
'workers',
'trades',
'capital',
'antitrust',
'common',
'Justice',
'among',
'found',
'--',
'used',
'stocks',
'agreement',

'selling',
'act',
'August',
'going',
'America',
'Western',
'factory',
'thing',
'department',
'Japan',
'activity',
'#',
'growth',
'rise',
'...',
'changes',
'serious',
'however',
'stake',
'Street',
"",
'money',
" 've",
""",
'bells',
'idea',
'12',
'Court',
'drugs',
'amount',
'anything',
'fees',
'contracts',
'similar',
':',
'example',
'wage',
'we',
'consented',
'Management',
'state',
'maker',
'investors',
'job',
'meeting',
'Co',
'*-1',

'We',
'China',
'defense',
'J.',
'ad',
'Asia',
'points',
'Board',
'even',
'3\\4',
'Ms.',
'continues',
'asbestos',
'200',
'plants',
'long',
'much',
'when',
'them',
'including',
'pence',
'card',
'current',
'Mitsubishi',
'drug',
'&',
'boost',
'1985',
'spokesman',
'shares',
'trade',
'nations',
'likely',
'buying',
'14',
'decision',
'recently',
'time',
'5',
'against',
'vice',
'*ICH*-2',
'launched',
'30-year',
'came',
'benefits',
'too',

'every',
'involved',
'proposal',
'have',
'Wednesday',
'advertising',
'team',
'Mr.',
'efforts',
'low',
'loan',
'High',
'kind',
'rates',
'Dinkins',
'person',
'25',
'potential',
'Japanese',
'million',
'close',
'whose',
'yet',
'corn',
'get',
'performance',
'lost',
'-LCB-',
'important',
'weeks',
'large',
'because',
'higher',
'you',
'*-3',
'officials',
'work',
'campaign',
'employees',
'thrift',
'followed',
'saying',
'three',
'paid',
'personal',
'single',
'able',

'composite',
'contract',
'15',
'scores',
'third-quarter',
'think',
'stock-index',
'annual',
'T.',
'could',
'hopes',
'program',
'real',
'watch',
'do',
'Hills',
'for',
'despite',
'senior',
'slow',
'result',
'some',
'0',
'official',
'short',
'out',
'products',
'share',
'reports',
'people',
'quarter',
'Robert',
'includes',
'need',
'can',
'Bush',
'teacher',
'rather',
'often',
'changed',
'cost',
'seeking',
'Francisco',
'firm',
'dividends',
'Columbia',
'international',

'regulators',
'farmers',
'10,000',
'San',
'head',
'receive',
'charges',
'I',
'situation',
'reading',
'continue',
'agency',
'former',
'short-term',
'course',
'banks',
'departure',
'fully',
'expected',
'analysts',
'enough',
'him',
'magazine',
'preferred',
'high',
'R.',
'over',
'level',
'30',
'showed',
'announced',
'UAL',
'limit',
'%',
'division',
'prevent',
'then',
'reduce',
'largest',
'rights',
'May',
'bank',
'since',
'corporate',
...]

```

[20]: def follow2(tagged_sents):
    fol = nltk.defaultdict(int)
    foll = nltk.defaultdict(list)
    for t in tagset:
        for i in [k[1] for k in context(tagged_sents) if k[0][1]==t]:
            fol[i[4]] += 1
        for k,v in fol.items() :
            if v > freq : foll[t].append(k)

    return [(k,v) for k,v in foll.items()]

f2 = []
for i in [v for k,v in follow2([s for s in text if len(s)>5])]:
    f2.extend(i)

f2 = list(set(f2))
f2

```

```

[20]: [0,
'Bank',
'Judge',
'know',
'minimum',
'conference',
'put',
'package',
'not',
'13',
'volatility',
'*RNR*-1',
'rate',
'Johnson',
'Corp',
'previous',
'gains',
'Union',
'bad',
'Sea',
'-RCB-',
'latest',
'unit',
'case',
'has',
'services',
'across',
'legislation',
'wanted',

```

'known',
'age',
'failed',
'decided',
'win',
'officer',
'chairman',
'Phelan',
'due',
'far',
'risks',
'operations',
'insurance',
'authority',
'1989',
'development',
'designed',
'bottle',
'sign',
'free',
'Angeles',
'publicly',
'*-2',
'ca',
'two',
'junk',
'Big',
'marketing',
'leading',
'Cray',
'closed',
'second',
'old',
'enforcement',
'months',
'increase',
'owns',
'past',
'Ltd.',
'Random',
'include',
'what',
'11',
'increased',
'overseas',
'5,000',
'member',

'position',
'publishing',
'meet',
'face',
'their',
'least',
'Ford',
'month',
'makes',
'period',
'18',
'goes',
'Trade',
'lower',
'shareholder',
'sale',
'year',
'convertible',
'make',
'just',
'Campbell',
'food',
'action',
'declined',
'equity',
'risk',
'president',
'generally',
'increases',
'from',
'research',
'teachers',
'University',
'look',
'my',
'end',
'his',
'soon',
'ban',
'London',
'above',
'Federal',
'fell',
'terms',
'earnings',
'being',
'back',

'Washington',
'future',
'financing',
'in',
'news',
'purchasing',
'?',
'alleged',
'total',
'become',
'car',
'be',
'four',
'Stock',
'Industries',
'managers',
'slowing',
'cents',
';',
'School',
'one-time',
'companies',
'number',
'countries',
'-RRB-',
'must',
'paper',
'administration',
'early',
'also',
'five',
'offset',
'holding',
'Group',
'investor',
'became',
'around',
'order',
'Department',
'possible',
'members',
'her',
'15,000',
'John',
'Greenville',
'been',
'plant',

'quickly',
'trucks',
'using',
'Dow',
'Commission',
'something',
'World',
'general',
'following',
'more',
'*?*',
'political',
'banking',
'financial',
'*U*',
'board',
'budget',
'planned',
'expects',
'United',
'return',
'further',
'continued',
'filed',
'of',
'funds',
'by',
'City',
'Jones',
'private',
'though',
'report',
'little',
'record',
'light',
'week',
'next',
'called',
'who',
'exports',
'company',
'40',
'Inc',
'recent',
'yield',
'management',
'power',

'Co.',
'payments',
'',
'cut',
'3',
'wines',
'telephone',
'cause',
'tax',
'off',
'nearly',
'computers',
'raised',
'city',
'decline',
'states',
'better',
'70',
'probably',
'all',
'International',
'Yeargin',
'William',
'built',
'life',
'wine',
'*-5',
'that',
'go',
'Fund',
'an',
'clear',
'A',
'PLC',
'agreed',
'-LRB-',
'would',
'acquired',
'Congress',
'greater',
'lawyers',
'went',
'D.',
'wants',
'100',
'well',
'very',

'under',
'value',
'traded',
'ordered',
'times',
'world',
'50',
'took',
'nine',
'less',
'good',
'court',
'force',
'York',
'volume',
'appropriations',
'export',
'*T*-1',
'transactions',
'information',
'sales',
'this',
'find',
'remain',
'region',
'family',
'speculation',
'dropped',
'delivery',
'dollar',
'those',
'few',
'homeless',
'General',
'dealers',
'own',
'problems',
'NL',
'closely',
'Wall',
'safety',
'7',
'believe',
'does',
'top',
'paying',
'office',

'bill',
'Nov.',
'1990',
'bond',
'Thomas',
'lot',
'7\\8',
'exchange',
'ended',
'late',
'airline',
'us',
'no',
'via',
'strong',
'a',
'real',
'continuing',
'issued',
'losses',
'manager',
'largely',
'ringers',
'included',
'posted',
'fourth',
'veto',
'the',
'editor',
'Financial',
'local',
'3\\8',
'pressure',
'cars',
'way',
'caused',
'received',
'offered',
'1987',
'Hahn',
'England',
'8',
'drop',
'nation',
'things',
'lead',
'Inc.',

'Boston',
'Nixon',
'gained',
'Committee',
'your',
'line',
'500',
'bid',
'Rep.',
'index',
'half',
'having',
'see',
'really',
'got',
'history',
'Richard',
'most',
'fall',
'are',
'one',
'students',
'domestic',
'indicated',
'until',
'talks',
'point',
'many',
'and',
'net',
'fined',
'deal',
'commercial',
'or',
'first',
'results',
'cases',
'after',
'American',
'women',
'say',
'Treasury',
'had',
'says',
'part',
'several',
'young',

'earlier',
'concern',
'60',
'hard',
'abortion',
'he',
'notes',
'State',
'crash',
'New',
'fact',
'credit',
'public',
'needed',
'*T*-4',
'if',
'S&P',
'buy-out',
'steel',
'21',
'concerns',
'according',
'advertisers',
'question',
'she',
'President',
'down',
'began',
're',
'1\\4',
'issue',
'fields',
'Monday',
'buyers',
'1991',
'billion',
'any',
'Commerce',
'grain',
'so',
'helped',
'compared',
'were',
'big',
'improve',
'although',
'strike',

'term',
'may',
'profits',
'.',
'than',
'here',
'1\\2',
'district',
'investment',
'6',
'control',
'*ICH*-1',
'charge',
'within',
'only',
'remains',
'always',
'spending',
'estimated',
'individual',
'test',
'debt',
'average',
'country',
'use',
'90',
'Association',
'apparently',
'takeover',
'oil',
'up',
'economic',
'Exchange',
'different',
'working',
'yen',
'March',
'costs',
'group',
'on',
'effect',
'fiscal',
'9',
'into',
'restructuring',
'conditions',
'executive',

'April',
'view',
'Gulf',
'third',
'traders',
'give',
'particularly',
'investments',
'these',
'analyst',
'still',
'assets',
'director',
'almost',
'interest',
'USX',
'Dec.',
'priced',
'health',
'about',
'Mrs.',
'role',
'm',
'production',
'institutions',
'California',
'which',
'plans',
'IRS',
'as',
'already',
'certain',
'data',
'support',
'September',
'special',
'trading',
'Michael',
'committee',
'whether',
'pay',
'1\\8',
'imports',
'without',
'*T*-3',
'prices',
'evidence',

'answers',
'matter',
'nor',
'returns',
'Express',
'form',
'violations',
'today',
'based',
'Korea',
'dividend',
'set',
'savings',
'law',
'16',
'effort',
'jobs',
'industrial',
'wo',
'31',
'approved',
'St.',
'with',
'help',
'software',
'underlying',
'named',
'business',
'they',
'increasing',
'Calif.',
'turned',
'right',
'rejected',
's',
'\$',
'Containers',
'change',
'Texas',
'computer',
'home',
'held',
'new',
'dollars',
'there',
'day',
'transaction',

'led',
'was',
'customers',
'II',
'relatively',
'auto',
'but',
'USIA',
'noted',
'fine',
'plan',
'orders',
'shareholders',
'system',
'sell',
'each',
'how',
'ability',
'told',
'major',
'significant',
'ruling',
'James',
'once',
'said',
'Jr.',
'currently',
'given',
'like',
'while',
'days',
'scheduled',
'barred',
'Georgia',
'5\\8',
'process',
'rule',
'Carolina',
'should',
'foreign',
'holders',
'it',
'school',
'making',
'themselves',
'available',
'believes',

'suspension',
'seek',
'years',
'take',
'is',
'October',
'below',
'Securities',
'National',
'June',
'bills',
'market',
'review',
'Buick',
'did',
'coming',
'First',
'principal',
'raise',
'away',
'survey',
'materials',
'questions',
'trying',
'now',
'keep',
'manufacturing',
'operating',
'our',
'newspaper',
'come',
'finance',
'at',
'last',
'U.S.',
'Moody',
'full',
'Tuesday',
'construction',
'bonds',
'parent',
'highly',
'problem',
'might',
'levels',
'South',
'Institute',

'10',
'loss',
'goods',
'Hearst',
'same',
'Tokyo',
'capacity',
'growing',
'long-term',
'gives',
'subject',
'another',
'Senate',
'provide',
'sector',
'chief',
'2',
'tests',
'others',
'show',
'during',
'church',
'rating',
'through',
'between',
'firms',
'*T*-2',
'Health',
'such',
'Research',
'sold',
'to',
'workers',
'trades',
'capital',
'antitrust',
'common',
'Justice',
'among',
'found',
'--',
'used',
'stocks',
'agreement',
'selling',
'act',
'August',

'going',
'America',
'factory',
'Western',
'thing',
'department',
'Japan',
'#',
'activity',
'growth',
'rise',
'...',
'changes',
'serious',
'however',
'stake',
'Street',
"",
'money',
" 've",
" """,
'bells',
'12',
'drugs',
'Court',
'amount',
'anything',
'fees',
'contracts',
'similar',
':',
'example',
'wage',
'we',
'consented',
'state',
'Management',
'maker',
'investors',
'meeting',
'job',
'Co',
'*-1',
'China',
'defense',
'J.',
'ad',

'Asia',
'points',
'Board',
'even',
'3\\4',
'best',
'continues',
'plants',
'200',
'long',
'when',
'much',
'them',
'including',
'pence',
'current',
'card',
'Mitsubishi',
'drug',
'&',
'boost',
'1985',
'spokesman',
'shares',
'trade',
'likely',
'buying',
'decision',
'14',
'nations',
'recently',
'time',
'5',
'against',
'vice',
'*ICH*-2',
'launched',
'came',
'benefits',
'too',
'*ICH*-3',
'every',
'involved',
'proposal',
'have',
'Wednesday',
'advertising',

'team',
'Mr.',
'low',
'loan',
'kind',
'High',
'rates',
'potential',
'25',
'person',
'Japanese',
'million',
'close',
'whose',
'yet',
'corn',
'get',
'large',
'performance',
'lost',
'important',
'-LCB-',
'weeks',
'because',
'higher',
'you',
'*-3',
'officials',
'work',
'campaign',
'employees',
'thrift',
'saying',
'three',
'paid',
'personal',
'single',
'able',
'composite',
'contract',
'15',
'scores',
'third-quarter',
'think',
'stock-index',
'annual',
'T.',

'could',
'program',
'watch',
'do',
'Hills',
'for',
'senior',
'despite',
'slow',
'result',
'some',
'0',
'official',
'short',
'out',
'share',
'products',
'reports',
'people',
'Robert',
'quarter',
'includes',
'need',
'can',
'Bush',
'teacher',
'rather',
'often',
'dividends',
'international',
'cost',
'changed',
'firm',
'seeking',
'Columbia',
'San',
'regulators',
'farmers',
'Francisco',
'head',
'10,000',
'receive',
'charges',
'I',
'situation',
'continue',
'former',

'agency',
'short-term',
'banks',
'fully',
'departure',
'expected',
'analysts',
'enough',
'him',
'magazine',
'preferred',
'high',
'R.',
'over',
'30',
'level',
'showed',
'announced',
'UAL',
'%',
'limit',
'division',
'prevent',
'then',
'reduce',
'largest',
'rights',
'since',
'corporate',
'bank',
'disclosed',
'its',
'House',
'ago',
'Sept.',
'cash',
'unchanged',
'Oct.',
'later',
'White',
'offering',
'Soviet',
'fund',
'made',
'*EXP*-1',
'clients',
'proposed',

'sugar',
'move',
'six',
'Chicago',
'acquisition',
'account',
'federal',
'run',
'purchase',
'pretax',
'issues',
'yesterday',
'huge',
'want',
'place',
'units',
'where',
'never',
'Louis',
'legal',
'traditional',
'floor',
'product',
'gain',
'20',
'17',
'statement',
'U.K.',
'both',
'industry',
'4',
...]

- []:
- []:
- []:
- []:
- []:
- []:
- []:

[]:

[]:

[]:

```
[21]: ##### Context
      ↪ ""#####
text = nltk.corpus.treebank.tagged_sents()
def consent2(s):
    """coordinates represents, in the same order:
    preceding word (by 2 positions),
    preceding word (by 1 position)
    central word
    1st following word
    2nd following word
    is it a number
    starts with capital
    is it a starting word
    is an ending word
    etc."""
    s = nltk.word_tokenize(s)
    contxt = {}
    # First word :
    contxt[(s[0],1)] = (0, 0, s[0], s[1],s[2],s[0].isdigit(), s[0][0].
↪isupper(), 1, 0)
    # Second word :
    contxt[(s[1],2)] = (0, s[0], s[1], s[2], s[3], s[1].isdigit(), s[1][0].
↪isupper(), 0, 0)
    # For middle words
    for i, w in enumerate(s[2:-2]) :
        contxt[(w,i+3)] = (s[i], s[i+1], s[i+2], s[i+3], s[i+4], w.isdigit(),
↪w[0].isupper(), 0, 0)
    # The word before last - the penultimate one - the second word to last
    contxt[(s[-2],len(s)-1)] = (s[-4], s[-3], s[-2], s[-1], 0, s[-2].isdigit(),
↪s[-2][0].isupper(),0, 0)
    # The last word
    contxt[(s[-1],len(s))] = (s[-3], s[-2], s[-1], 0, 0, s[-1].isdigit(),
↪s[-1][0].isupper(), 0, 1)
    return [(k[0],v) for k,v in contxt.items()]

def context2(text):
    txt = []
    for s in nltk.sent_tokenize(text) :
        txt.extend(consent2(s))
    return txt
```



```
[33]: current
```

```
[33]: defaultdict(int,
    {'Pierre': 1,
     'Vinken': 2,
     'nonexecutive': 1,
     'Mr.': 1,
     'chairman': 3,
     'of': 13,
     ',': 9,
     'Rudolph': 1,
     'Agnew': 1,
     'Consolidated': 1,
     'Gold': 1,
     'to': 4,
     'Lorillard': 4,
     'Inc.': 2,
     'New': 2,
     'York-based': 1,
     '*T*-2': 1,
     'today': 1,
     "'s": 4,
     'Journal': 1,
     'said': 4,
     '*T*-1': 3,
     'James': 1,
     'Talcott': 3,
     'Boston': 1,
     'Institute': 1,
     'Dr.': 1,
     'from': 1,
     'the': 10,
     'National': 2,
     'schools': 1,
     'University': 3,
     '9.8': 1,
     'at': 3,
     'Groton': 1,
     '*-4': 1,
     'Hollingsworth': 1,
     '&': 1,
     'that': 1,
     'buildings': 1,
     'U.S.': 1,
     'according': 2,
     'Brooke': 1,
     'Medicine': 1,
```

'body': 1,
'July': 1,
'Environmental': 1,
'for': 1,
'resources': 1,
'Vose': 1,
'*': 3,
'IBC': 1,
'Money': 2,
'Tuesday': 1,
'since': 1,
'Brenda': 1,
'editor': 1,
'on': 4,
'Dreyfus': 2,
'World-Wide': 1,
'J.P.': 1,
'Bolduc': 1,
'Grace': 2,
'He': 1,
'succeeds': 1,
'Terrence': 1,
'formerly': 1,
'a': 1,
'W.R.': 1,
'three': 1,
'Pacific': 2,
'First': 2,
'acquisition': 1,
'by': 2,
'Royal': 1,
'Ltd.': 1,
'McDermott': 2,
'International': 1,
'O': 2,
'Babcock': 1,
'its': 1,
'Bailey': 2,
'Operations': 1,
'S.p': 1,
'Finmeccanica': 1,
'Controls': 1,
'Wickliffe': 1,
'sales': 1,
'bonds': 1,
'Congress': 1,
'House': 1,

'but': 1,
'Treasury': 1,
'default': 1,
'9': 1,
'Clark': 2,
'J.': 1,
'auto': 1,
'maker': 1,
'Corp': 1,
'will': 1,
'Mazda': 1,
'Previously': 1,
'manager': 1,
'Corp.': 1,
'executive': 1,
'towns': 1,
'like': 3,
'Raton': 1,
'Springs': 1,
'The': 1,
'Association': 1,
'capital': 1,
'giants': 1,
'Pont': 1,
'knowns': 1,
'and': 3,
'Valley': 1,
'Factory': 1,
'executives': 1,
'joined': 1,
'Mayor': 1,
'William': 1,
'H.': 1,
'Indianapolis': 2,
'guest': 1,
'Borge': 1,
'race': 1,
'renovated': 1,
'fed': 1,
'meeting': 1,
'February': 1,
'South': 2,
'Korea': 2,
'*U*': 1,
'Wednesday': 1,
'Trade': 1,
'deficit': 1,

```

'cloud': 1,
'Exports': 1,
'October': 1,
'forecast': 1,
'January': 2,
'Newsweek': 2,
'with': 1,
'plan': 1,
'Washington': 1,
'competition': 1,
'Time': 1,
'Mortimer': 1,
'Zuckerman': 1,
'News': 1,
'Report': 1})

```

[]:

[]:

```

[ ]: current = nltk.defaultdict(int)
precedent = nltk.defaultdict(int)
sent_begin = nltk.defaultdict(int)

for i in [k[1] for k in context3(text[:100]) if k[0][1]=="NNP"]:
    current[i[1]] += 1
    precedent[i[0]] += 1
    sent_begin[str(i[9])] += 1

""" Pour tout t:
1) Construire:
preced1 = nltk.defaultdict
..
follow2 = nltk.defaultdict

Garder les key pour lesquelles value > 3 --> exemple : prec["the"]=18

-----

2) On les stock dans un vecteur, tout en respectant l'ordre (prec21, prec22, ...
↳ prec2, prec11, ..., current1.. )
Remarque prec21 = précédé par "the" ; alors on pourra le noter "the_2"
follow2 = suivi par "the" ; alors on pourra le noter "the+2"

3) définir la fonction f_t, qui pour chaque couple (t,x) associe un np avec des_
↳ 0 ou 1 selon si

```

la condition est vérifiée. --> exemple 't' est-il précédé par 'w' (si oui 1) et
↪(0 sinon)

4) On définit la fonction vectorielle f dont les coordonnées sont f_t

!! En prenant soin de supprimer les redondances -> Par exemple si f_{NNP}
↪contient "précédé par 'the'"
et f_{ADJ} contient "précédé par 'the'" aussi, alors on en garde que 1.

Autre méthode:

1) Se concentrer sur $prec1$ d'abord (pour tout t) afin d'obtenir au final un
↪seul grand vecteur
contenant tous les mots précédant les tags

""

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

[]:

```
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
```

```
[104]: ##### Context
↳ #####
text = nltk.corpus.treebank.tagged_sents()
def con_sent(s):
    """coordinates represents, in the same order:
    preceding word (by 2 positions),
    preceding word (by 1 position)
    central word
    1st following word
    2nd following word
    is it a number
    starts with capital
    is it a starting word
    is an ending word
    etc."""
    if len(s)>=5 :
        contxt = {}
        # First word :
        contxt[(s[0][0],s[0][1],1)] = (0, 0, s[0][0], s[1][0],s[2][0],s[0][0].
↳isdigit(), s[0][0][0].isupper(), 1, 0)
        # Second word :
        contxt[(s[1][0],s[1][1],2)] = (0, s[0][0], s[1][0], s[2][0], s[3][0],
↳s[1][0].isdigit(), s[1][0][0].isupper(), 0, 0)
```

```

    # For middle words
    for i, w in enumerate(s[2:-2]) :
        contxt[(w[0],w[1],i+3)] = (s[i][0], s[i+1][0], s[i+2][0],
↪s[i+3][0], s[i+4][0], w[0].isdigit(), w[0][0].isupper(), 0, 0)
        # The word before last - the penultimate one - the second word to last
        contxt[(s[-2][0],s[-2][1],len(s)-1)] = (s[-4][0], s[-3][0], s[-2][0],
↪s[-1][0], 0, s[-2][0].isdigit(), s[-2][0][0].isupper(),0, 0)
        # The last word
        contxt[(s[-1][0],s[-1][1],len(s))] = (s[-3][0], s[-2][0], s[-1][0], 0,
↪0, s[-1][0].isdigit(), s[-1][0][0].isupper(), 0, 1)

    elif len(s)==4:
        contxt = {}
        # First word :
        contxt[(s[0][0],s[0][1],1)] = (0, 0, s[0][0], s[1][0],s[2][0],s[0][0].
↪isdigit(), s[0][0][0].isupper(), 1, 0)
        # Second word :
        contxt[(s[1][0],s[1][1],2)] = (0, s[0][0], s[1][0], s[2][0], s[3][0],
↪s[1][0].isdigit(), s[1][0][0].isupper(), 0, 0)
        # The word before last - the penultimate one - the second word to last
        contxt[(s[-2][0],s[-2][1],len(s)-1)] = (s[-4][0], s[-3][0], s[-2][0],
↪s[-1][0], 0, s[-2][0].isdigit(), s[-2][0][0].isupper(),0, 0)
        # The last word
        contxt[(s[-1][0],s[-1][1],len(s))] = (s[-3][0], s[-2][0], s[-1][0], 0,
↪0, s[-1][0].isdigit(), s[-1][0][0].isupper(), 0, 1)

    elif len(s)==3:
        contxt = {}
        # First word :
        contxt[(s[0][0],s[0][1],1)] = (0, 0, s[0][0], s[1][0],s[2][0],s[0][0].
↪isdigit(), s[0][0][0].isupper(), 1, 0)
        # Second word :
        contxt[(s[1][0],s[1][1],2)] = (0, s[0][0], s[1][0], s[2][0], 0, s[1][0].
↪isdigit(), s[1][0][0].isupper(), 0, 0)
        # The last word
        contxt[(s[-1][0],s[-1][1],len(s))] = (s[-3][0], s[-2][0], s[-1][0], 0,
↪0, s[-1][0].isdigit(), s[-1][0][0].isupper(), 0, 1)

    elif len(s)==2:
        contxt = {}
        # First word :
        contxt[(s[0][0],s[0][1],1)] = (0, 0, s[0][0], s[1][0],0,s[0][0].
↪isdigit(), s[0][0][0].isupper(), 1, 0)
        # The last word
        contxt[(s[-1][0],s[-1][1],len(s))] = (0, s[0][0], s[1][0], 0, 0,
↪s[-1][0].isdigit(), s[-1][0][0].isupper(), 0, 1)

```



```

    else: contxt={}; contxt[(s[0][0],s[0][1],1)] = (0, 0, s[0][0], 0,0,s[0][0].
↳isdigit(), s[0][0][0].isupper(), 1, 1)
    return [(k,v) for k,v in contxt.items()]

def con_text(text):
    txt = []
    for s in text :
        txt.extend(con_sent(s))
    return txt

```

```
[108]: con_text(text)
```

```

[108]: [((('Pierre', 'NNP', 1), (0, 0, 'Pierre', 'Vinken', ',', False, True, 1, 0)),
((('Vinken', 'NNP', 2), (0, 'Pierre', 'Vinken', ',', '61', False, True, 0, 0)),
(((',', ',', 3), ('Pierre', 'Vinken', ',', '61', 'years', False, False, 0, 0)),
((('61', 'CD', 4), ('Vinken', ',', '61', 'years', 'old', True, False, 0, 0)),
((('years', 'NNS', 5), (',', '61', 'years', 'old', ',', False, False, 0, 0)),
((('old', 'JJ', 6), ('61', 'years', 'old', ',', 'will', False, False, 0, 0)),
(((',', ',', 7), ('years', 'old', ',', 'will', 'join', False, False, 0, 0)),
((('will', 'MD', 8), ('old', ',', 'will', 'join', 'the', False, False, 0, 0)),
((('join', 'VB', 9),
(',', 'will', 'join', 'the', 'board', False, False, 0, 0)),
((('the', 'DT', 10),
('will', 'join', 'the', 'board', 'as', False, False, 0, 0)),
((('board', 'NN', 11),
('join', 'the', 'board', 'as', 'a', False, False, 0, 0)),
((('as', 'IN', 12),
('the', 'board', 'as', 'a', 'nonexecutive', False, False, 0, 0)),
((('a', 'DT', 13),
('board', 'as', 'a', 'nonexecutive', 'director', False, False, 0, 0)),
((('nonexecutive', 'JJ', 14),
('as', 'a', 'nonexecutive', 'director', 'Nov.', False, False, 0, 0)),
((('director', 'NN', 15),
('a', 'nonexecutive', 'director', 'Nov.', '29', False, False, 0, 0)),
((('Nov.', 'NNP', 16),
('nonexecutive', 'director', 'Nov.', '29', '.', False, True, 0, 0)),
((('29', 'CD', 17), ('director', 'Nov.', '29', '.', 0, True, False, 0, 0)),
((('.', '.', 18), ('Nov.', '29', '.', 0, 0, False, False, 0, 1)),
((('Mr.', 'NNP', 1), (0, 0, 'Mr.', 'Vinken', 'is', False, True, 1, 0)),
((('Vinken', 'NNP', 2),
(0, 'Mr.', 'Vinken', 'is', 'chairman', False, True, 0, 0)),
((('is', 'VBZ', 3),
('Mr.', 'Vinken', 'is', 'chairman', 'of', False, False, 0, 0)),
((('chairman', 'NN', 4),
('Vinken', 'is', 'chairman', 'of', 'Elsevier', False, False, 0, 0)),
((('of', 'IN', 5),

```

('is', 'chairman', 'of', 'Elsevier', 'N.V.', False, False, 0, 0)),
 (('Elsevier', 'NNP', 6),
 ('chairman', 'of', 'Elsevier', 'N.V.', ',', False, True, 0, 0)),
 (('N.V.', 'NNP', 7),
 ('of', 'Elsevier', 'N.V.', ',', 'the', False, True, 0, 0)),
 ((' ', ' ', ' ', 8),
 ('Elsevier', 'N.V.', ',', 'the', 'Dutch', False, False, 0, 0)),
 (('the', 'DT', 9),
 ('N.V.', ',', 'the', 'Dutch', 'publishing', False, False, 0, 0)),
 (('Dutch', 'NNP', 10),
 (' ', 'the', 'Dutch', 'publishing', 'group', False, True, 0, 0)),
 (('publishing', 'VBG', 11),
 ('the', 'Dutch', 'publishing', 'group', '.', False, False, 0, 0)),
 (('group', 'NN', 12),
 ('Dutch', 'publishing', 'group', '.', 0, False, False, 0, 0)),
 ((' ', ' ', 13), ('publishing', 'group', '.', 0, 0, False, False, 0, 1)),
 (('Rudolph', 'NNP', 1), (0, 0, 'Rudolph', 'Agnew', ',', False, True, 1, 0)),
 (('Agnew', 'NNP', 2), (0, 'Rudolph', 'Agnew', ',', '55', False, True, 0, 0)),
 ((' ', ' ', ' ', 3), ('Rudolph', 'Agnew', ',', '55', 'years', False, False, 0, 0)),
 (('55', 'CD', 4), ('Agnew', ',', '55', 'years', 'old', True, False, 0, 0)),
 (('years', 'NNS', 5), (' ', '55', 'years', 'old', 'and', False, False, 0, 0)),
 (('old', 'JJ', 6),
 ('55', 'years', 'old', 'and', 'former', False, False, 0, 0)),
 (('and', 'CC', 7),
 ('years', 'old', 'and', 'former', 'chairman', False, False, 0, 0)),
 (('former', 'JJ', 8),
 ('old', 'and', 'former', 'chairman', 'of', False, False, 0, 0)),
 (('chairman', 'NN', 9),
 ('and', 'former', 'chairman', 'of', 'Consolidated', False, False, 0, 0)),
 (('of', 'IN', 10),
 ('former', 'chairman', 'of', 'Consolidated', 'Gold', False, False, 0, 0)),
 (('Consolidated', 'NNP', 11),
 ('chairman', 'of', 'Consolidated', 'Gold', 'Fields', False, True, 0, 0)),
 (('Gold', 'NNP', 12),
 ('of', 'Consolidated', 'Gold', 'Fields', 'PLC', False, True, 0, 0)),
 (('Fields', 'NNP', 13),
 ('Consolidated', 'Gold', 'Fields', 'PLC', ',', False, True, 0, 0)),
 (('PLC', 'NNP', 14),
 ('Gold', 'Fields', 'PLC', ',', 'was', False, True, 0, 0)),
 ((' ', ' ', ' ', 15), ('Fields', 'PLC', ',', 'was', 'named', False, False, 0, 0)),
 (('was', 'VBD', 16), ('PLC', ',', 'was', 'named', '*-1', False, False, 0, 0)),
 (('named', 'VBN', 17), (' ', 'was', 'named', '*-1', 'a', False, False, 0, 0)),
 (('*-1', '-NONE-', 18),
 ('was', 'named', '*-1', 'a', 'nonexecutive', False, False, 0, 0)),
 (('a', 'DT', 19),
 ('named', '*-1', 'a', 'nonexecutive', 'director', False, False, 0, 0)),
 (('nonexecutive', 'JJ', 20),

(*-1', 'a', 'nonexecutive', 'director', 'of', False, False, 0, 0)),
 (('director', 'NN', 21),
 ('a', 'nonexecutive', 'director', 'of', 'this', False, False, 0, 0)),
 (('of', 'IN', 22),
 ('nonexecutive', 'director', 'of', 'this', 'British', False, False, 0, 0)),
 (('this', 'DT', 23),
 ('director', 'of', 'this', 'British', 'industrial', False, False, 0, 0)),
 (('British', 'JJ', 24),
 ('of', 'this', 'British', 'industrial', 'conglomerate', False, True, 0, 0)),
 (('industrial', 'JJ', 25),
 ('this', 'British', 'industrial', 'conglomerate', '.', False, False, 0, 0)),
 (('conglomerate', 'NN', 26),
 ('British', 'industrial', 'conglomerate', '.', 0, False, False, 0, 0)),
 (('.', '.', 27),
 ('industrial', 'conglomerate', '.', 0, 0, False, False, 0, 1)),
 (('A', 'DT', 1), (0, 0, 'A', 'form', 'of', False, True, 1, 0)),
 (('form', 'NN', 2), (0, 'A', 'form', 'of', 'asbestos', False, False, 0, 0)),
 (('of', 'IN', 3),
 ('A', 'form', 'of', 'asbestos', 'once', False, False, 0, 0)),
 (('asbestos', 'NN', 4),
 ('form', 'of', 'asbestos', 'once', 'used', False, False, 0, 0)),
 (('once', 'RB', 5),
 ('of', 'asbestos', 'once', 'used', '*', False, False, 0, 0)),
 (('used', 'VBN', 6),
 ('asbestos', 'once', 'used', '*', '*', False, False, 0, 0)),
 (('*', '-NONE-', 7), ('once', 'used', '*', '*', 'to', False, False, 0, 0)),
 (('*', '-NONE-', 8), ('used', '*', '*', 'to', 'make', False, False, 0, 0)),
 (('to', 'TO', 9), ('*', '*', 'to', 'make', 'Kent', False, False, 0, 0)),
 (('make', 'VB', 10),
 ('*', 'to', 'make', 'Kent', 'cigarette', False, False, 0, 0)),
 (('Kent', 'NNP', 11),
 ('to', 'make', 'Kent', 'cigarette', 'filters', False, True, 0, 0)),
 (('cigarette', 'NN', 12),
 ('make', 'Kent', 'cigarette', 'filters', 'has', False, False, 0, 0)),
 (('filters', 'NNS', 13),
 ('Kent', 'cigarette', 'filters', 'has', 'caused', False, False, 0, 0)),
 (('has', 'VBZ', 14),
 ('cigarette', 'filters', 'has', 'caused', 'a', False, False, 0, 0)),
 (('caused', 'VBN', 15),
 ('filters', 'has', 'caused', 'a', 'high', False, False, 0, 0)),
 (('a', 'DT', 16),
 ('has', 'caused', 'a', 'high', 'percentage', False, False, 0, 0)),
 (('high', 'JJ', 17),
 ('caused', 'a', 'high', 'percentage', 'of', False, False, 0, 0)),
 (('percentage', 'NN', 18),
 ('a', 'high', 'percentage', 'of', 'cancer', False, False, 0, 0)),
 (('of', 'IN', 19),

('high', 'percentage', 'of', 'cancer', 'deaths', False, False, 0, 0)),
 (('cancer', 'NN', 20),
 ('percentage', 'of', 'cancer', 'deaths', 'among', False, False, 0, 0)),
 (('deaths', 'NNS', 21),
 ('of', 'cancer', 'deaths', 'among', 'a', False, False, 0, 0)),
 (('among', 'IN', 22),
 ('cancer', 'deaths', 'among', 'a', 'group', False, False, 0, 0)),
 (('a', 'DT', 23),
 ('deaths', 'among', 'a', 'group', 'of', False, False, 0, 0)),
 (('group', 'NN', 24),
 ('among', 'a', 'group', 'of', 'workers', False, False, 0, 0)),
 (('of', 'IN', 25),
 ('a', 'group', 'of', 'workers', 'exposed', False, False, 0, 0)),
 (('workers', 'NNS', 26),
 ('group', 'of', 'workers', 'exposed', '*', False, False, 0, 0)),
 (('exposed', 'VBN', 27),
 ('of', 'workers', 'exposed', '*', 'to', False, False, 0, 0)),
 (('*', '-NONE-', 28),
 ('workers', 'exposed', '*', 'to', 'it', False, False, 0, 0)),
 (('to', 'TO', 29), ('exposed', '*', 'to', 'it', 'more', False, False, 0, 0)),
 (('it', 'PRP', 30), ('*', 'to', 'it', 'more', 'than', False, False, 0, 0)),
 (('more', 'RBR', 31), ('to', 'it', 'more', 'than', '30', False, False, 0, 0)),
 (('than', 'IN', 32),
 ('it', 'more', 'than', '30', 'years', False, False, 0, 0)),
 (('30', 'CD', 33), ('more', 'than', '30', 'years', 'ago', True, False, 0, 0)),
 (('years', 'NNS', 34),
 ('than', '30', 'years', 'ago', ',', False, False, 0, 0)),
 (('ago', 'IN', 35),
 ('30', 'years', 'ago', ',', 'researchers', False, False, 0, 0)),
 ((' ', ' ', 36),
 ('years', 'ago', ',', 'researchers', 'reported', False, False, 0, 0)),
 (('researchers', 'NNS', 37),
 ('ago', ',', 'researchers', 'reported', '0', False, False, 0, 0)),
 (('reported', 'VBD', 38),
 (',', 'researchers', 'reported', '0', '*T*-1', False, False, 0, 0)),
 (('0', '-NONE-', 39),
 ('researchers', 'reported', '0', '*T*-1', '.', True, False, 0, 0)),
 (('*T*-1', '-NONE-', 40),
 ('reported', '0', '*T*-1', '.', 0, False, False, 0, 0)),
 (('.', '.', 41), ('0', '*T*-1', '.', 0, 0, False, False, 0, 1)),
 (('The', 'DT', 1), (0, 0, 'The', 'asbestos', 'fiber', False, True, 1, 0)),
 (('asbestos', 'NN', 2),
 (0, 'The', 'asbestos', 'fiber', ',', False, False, 0, 0)),
 (('fiber', 'NN', 3),
 ('The', 'asbestos', 'fiber', ',', 'crocidolite', False, False, 0, 0)),
 ((' ', ' ', 4),
 ('asbestos', 'fiber', ',', 'crocidolite', ',', False, False, 0, 0)),

(('crocidolite', 'NN', 5),
 ('fiber', ',', 'crocidolite', ',', 'is', False, False, 0, 0)),
 ((',', ',', 6),
 (',', 'crocidolite', ',', 'is', 'unusually', False, False, 0, 0)),
 (('is', 'VBZ', 7),
 ('crocidolite', ',', 'is', 'unusually', 'resilient', False, False, 0, 0)),
 (('unusually', 'RB', 8),
 (',', 'is', 'unusually', 'resilient', 'once', False, False, 0, 0)),
 (('resilient', 'JJ', 9),
 ('is', 'unusually', 'resilient', 'once', 'it', False, False, 0, 0)),
 (('once', 'IN', 10),
 ('unusually', 'resilient', 'once', 'it', 'enters', False, False, 0, 0)),
 (('it', 'PRP', 11),
 ('resilient', 'once', 'it', 'enters', 'the', False, False, 0, 0)),
 (('enters', 'VBZ', 12),
 ('once', 'it', 'enters', 'the', 'lungs', False, False, 0, 0)),
 (('the', 'DT', 13),
 ('it', 'enters', 'the', 'lungs', ',', False, False, 0, 0)),
 (('lungs', 'NNS', 14),
 ('enters', 'the', 'lungs', ',', 'with', False, False, 0, 0)),
 ((',', ',', 15), ('the', 'lungs', ',', 'with', 'even', False, False, 0, 0)),
 (('with', 'IN', 16),
 ('lungs', ',', 'with', 'even', 'brief', False, False, 0, 0)),
 (('even', 'RB', 17),
 (',', 'with', 'even', 'brief', 'exposures', False, False, 0, 0)),
 (('brief', 'JJ', 18),
 ('with', 'even', 'brief', 'exposures', 'to', False, False, 0, 0)),
 (('exposures', 'NNS', 19),
 ('even', 'brief', 'exposures', 'to', 'it', False, False, 0, 0)),
 (('to', 'TO', 20),
 ('brief', 'exposures', 'to', 'it', 'causing', False, False, 0, 0)),
 (('it', 'PRP', 21),
 ('exposures', 'to', 'it', 'causing', 'symptoms', False, False, 0, 0)),
 (('causing', 'VBG', 22),
 ('to', 'it', 'causing', 'symptoms', 'that', False, False, 0, 0)),
 (('symptoms', 'NNS', 23),
 ('it', 'causing', 'symptoms', 'that', '*T*-1', False, False, 0, 0)),
 (('that', 'WDT', 24),
 ('causing', 'symptoms', 'that', '*T*-1', 'show', False, False, 0, 0)),
 ((*T*-1', '-NONE-', 25),
 ('symptoms', 'that', '*T*-1', 'show', 'up', False, False, 0, 0)),
 (('show', 'VBP', 26),
 ('that', '*T*-1', 'show', 'up', 'decades', False, False, 0, 0)),
 (('up', 'RP', 27),
 (*T*-1', 'show', 'up', 'decades', 'later', False, False, 0, 0)),
 (('decades', 'NNS', 28),
 ('show', 'up', 'decades', 'later', ',', False, False, 0, 0)),

(('later', 'JJ', 29),
 ('up', 'decades', 'later', ',', 'researchers', False, False, 0, 0)),
 (('decades', 'later', ',', 'researchers', 'said', False, False, 0, 0)),
 (('researchers', 'NNS', 31),
 ('later', ',', 'researchers', 'said', '0', False, False, 0, 0)),
 (('said', 'VBD', 32),
 (',', 'researchers', 'said', '0', '*T*-2', False, False, 0, 0)),
 (('0', '-NONE-', 33),
 ('researchers', 'said', '0', '*T*-2', '.', True, False, 0, 0)),
 ((*T*-2', '-NONE-', 34), ('said', '0', '*T*-2', '.', 0, False, False, 0, 0)),
 (('.', '.', 35), ('0', '*T*-2', '.', 0, 0, False, False, 0, 1)),
 (('Lorillard', 'NNP', 1),
 (0, 0, 'Lorillard', 'Inc.', ',', False, True, 1, 0)),
 (('Inc.', 'NNP', 2), (0, 'Lorillard', 'Inc.', ',', 'the', False, True, 0, 0)),
 (',', ',', 3),
 ('Lorillard', 'Inc.', ',', 'the', 'unit', False, False, 0, 0)),
 (('the', 'DT', 4), ('Inc.', ',', 'the', 'unit', 'of', False, False, 0, 0)),
 (('unit', 'NN', 5), (',', 'the', 'unit', 'of', 'New', False, False, 0, 0)),
 (('of', 'IN', 6),
 ('the', 'unit', 'of', 'New', 'York-based', False, False, 0, 0)),
 (('New', 'JJ', 7),
 ('unit', 'of', 'New', 'York-based', 'Loews', False, True, 0, 0)),
 (('York-based', 'JJ', 8),
 ('of', 'New', 'York-based', 'Loews', 'Corp.', False, True, 0, 0)),
 (('Loews', 'NNP', 9),
 ('New', 'York-based', 'Loews', 'Corp.', 'that', False, True, 0, 0)),
 (('Corp.', 'NNP', 10),
 ('York-based', 'Loews', 'Corp.', 'that', '*T*-2', False, True, 0, 0)),
 (('that', 'WDT', 11),
 ('Loews', 'Corp.', 'that', '*T*-2', 'makes', False, False, 0, 0)),
 ((*T*-2', '-NONE-', 12),
 ('Corp.', 'that', '*T*-2', 'makes', 'Kent', False, False, 0, 0)),
 (('makes', 'VBZ', 13),
 ('that', '*T*-2', 'makes', 'Kent', 'cigarettes', False, False, 0, 0)),
 (('Kent', 'NNP', 14),
 (*T*-2', 'makes', 'Kent', 'cigarettes', ',', False, True, 0, 0)),
 (('cigarettes', 'NNS', 15),
 ('makes', 'Kent', 'cigarettes', ',', 'stopped', False, False, 0, 0)),
 (',', ',', 16),
 ('Kent', 'cigarettes', ',', 'stopped', 'using', False, False, 0, 0)),
 (('stopped', 'VBD', 17),
 ('cigarettes', ',', 'stopped', 'using', 'crocidolite', False, False, 0, 0)),
 (('using', 'VBG', 18),
 (',', 'stopped', 'using', 'crocidolite', 'in', False, False, 0, 0)),
 (('crocidolite', 'NN', 19),
 ('stopped', 'using', 'crocidolite', 'in', 'its', False, False, 0, 0)),

```

(('in', 'IN', 20),
 ('using', 'crocidolite', 'in', 'its', 'Micronite', False, False, 0, 0)),
(('its', 'PRP$', 21),
 ('crocidolite', 'in', 'its', 'Micronite', 'cigarette', False, False, 0, 0)),
(('Micronite', 'NN', 22),
 ('in', 'its', 'Micronite', 'cigarette', 'filters', False, True, 0, 0)),
(('cigarette', 'NN', 23),
 ('its', 'Micronite', 'cigarette', 'filters', 'in', False, False, 0, 0)),
(('filters', 'NNS', 24),
 ('Micronite', 'cigarette', 'filters', 'in', '1956', False, False, 0, 0)),
(('in', 'IN', 25),
 ('cigarette', 'filters', 'in', '1956', '.', False, False, 0, 0)),
(('1956', 'CD', 26), ('filters', 'in', '1956', '.', 0, True, False, 0, 0)),
(('.', '.', 27), ('in', '1956', '.', 0, 0, False, False, 0, 1)),
(('Although', 'IN', 1),
 (0, 0, 'Although', 'preliminary', 'findings', False, True, 1, 0)),
(('preliminary', 'JJ', 2),
 (0, 'Although', 'preliminary', 'findings', 'were', False, False, 0, 0)),
(('findings', 'NNS', 3),
 ('Although',
 'preliminary',
 'findings',
 'were',
 'reported',
 False,
 False,
 0,
 0)),
(('were', 'VBD', 4),
 ('preliminary', 'findings', 'were', 'reported', '*-2', False, False, 0, 0)),
(('reported', 'VBN', 5),
 ('findings', 'were', 'reported', '*-2', 'more', False, False, 0, 0)),
(('*-2', '-NONE-', 6),
 ('were', 'reported', '*-2', 'more', 'than', False, False, 0, 0)),
(('more', 'RBR', 7),
 ('reported', '*-2', 'more', 'than', 'a', False, False, 0, 0)),
(('than', 'IN', 8), ('*-2', 'more', 'than', 'a', 'year', False, False, 0, 0)),
(('a', 'DT', 9), ('more', 'than', 'a', 'year', 'ago', False, False, 0, 0)),
(('year', 'NN', 10), ('than', 'a', 'year', 'ago', ',', False, False, 0, 0)),
(('ago', 'IN', 11), ('a', 'year', 'ago', ',', 'the', False, False, 0, 0)),
((',', ',', 12), ('year', 'ago', ',', 'the', 'latest', False, False, 0, 0)),
(('the', 'DT', 13),
 ('ago', ',', 'the', 'latest', 'results', False, False, 0, 0)),
(('latest', 'JJS', 14),
 (',', 'the', 'latest', 'results', 'appear', False, False, 0, 0)),
(('results', 'NNS', 15),
 ('the', 'latest', 'results', 'appear', 'in', False, False, 0, 0)),

```

(('appear', 'VBP', 16),
 ('latest', 'results', 'appear', 'in', 'today', False, False, 0, 0)),
 (('in', 'IN', 17),
 ('results', 'appear', 'in', 'today', "'s", False, False, 0, 0)),
 (('today', 'NN', 18),
 ('appear', 'in', 'today', "'s", 'New', False, False, 0, 0)),
 (('s', 'POS', 19),
 ('in', 'today', "'s", 'New', 'England', False, False, 0, 0)),
 (('New', 'NNP', 20),
 ('today', "'s", 'New', 'England', 'Journal', False, True, 0, 0)),
 (('England', 'NNP', 21),
 ("s", 'New', 'England', 'Journal', 'of', False, True, 0, 0)),
 (('Journal', 'NNP', 22),
 ('New', 'England', 'Journal', 'of', 'Medicine', False, True, 0, 0)),
 (('of', 'IN', 23),
 ('England', 'Journal', 'of', 'Medicine', ',', False, False, 0, 0)),
 (('Medicine', 'NNP', 24),
 ('Journal', 'of', 'Medicine', ',', 'a', False, True, 0, 0)),
 ((',', ',', ',', 25), ('of', 'Medicine', ',', 'a', 'forum', False, False, 0, 0)),
 (('a', 'DT', 26),
 ('Medicine', ',', 'a', 'forum', 'likely', False, False, 0, 0)),
 (('forum', 'NN', 27), ((',', 'a', 'forum', 'likely', '*', False, False, 0, 0)),
 (('likely', 'JJ', 28),
 ('a', 'forum', 'likely', '*', 'to', False, False, 0, 0)),
 (('*', '-NONE-', 29),
 ('forum', 'likely', '*', 'to', 'bring', False, False, 0, 0)),
 (('to', 'TO', 30), ('likely', '*', 'to', 'bring', 'new', False, False, 0, 0)),
 (('bring', 'VB', 31),
 ('*', 'to', 'bring', 'new', 'attention', False, False, 0, 0)),
 (('new', 'JJ', 32),
 ('to', 'bring', 'new', 'attention', 'to', False, False, 0, 0)),
 (('attention', 'NN', 33),
 ('bring', 'new', 'attention', 'to', 'the', False, False, 0, 0)),
 (('to', 'TO', 34),
 ('new', 'attention', 'to', 'the', 'problem', False, False, 0, 0)),
 (('the', 'DT', 35),
 ('attention', 'to', 'the', 'problem', '.', False, False, 0, 0)),
 (('problem', 'NN', 36), ('to', 'the', 'problem', '.', 0, False, False, 0, 0)),
 (('.', '.', 37), ('the', 'problem', '.', 0, 0, False, False, 0, 1)),
 (('A', 'DT', 1), (0, 0, 'A', 'Lorillard', 'spokewoman', False, True, 1, 0)),
 (('Lorillard', 'NNP', 2),
 (0, 'A', 'Lorillard', 'spokewoman', 'said', False, True, 0, 0)),
 (('spokewoman', 'NN', 3),
 ('A', 'Lorillard', 'spokewoman', 'said', ',', False, False, 0, 0)),
 (('said', 'VBD', 4),
 ('Lorillard', 'spokewoman', 'said', ',', '``', False, False, 0, 0)),
 ((',', ',', ',', 5),


```

('spokewoman', 'said', ',', '``', 'This', False, False, 0, 0)),
(('``', '``', 6), ('said', ',', '``', 'This', 'is', False, False, 0, 0)),
(('This', 'DT', 7), (',', '``', 'This', 'is', 'an', False, True, 0, 0)),
(('is', 'VBZ', 8), ('``', 'This', 'is', 'an', 'old', False, False, 0, 0)),
(('an', 'DT', 9), ('This', 'is', 'an', 'old', 'story', False, False, 0, 0)),
(('old', 'JJ', 10), ('is', 'an', 'old', 'story', '.', False, False, 0, 0)),
(('story', 'NN', 11), ('an', 'old', 'story', '.', 0, False, False, 0, 0)),
(('.', '.', 12), ('old', 'story', '.', 0, 0, False, False, 0, 1)),
(('We', 'PRP', 1), (0, 0, 'We', "'re", 'talking', False, True, 1, 0)),
(("'re", 'VBP', 2), (0, 'We', "'re", 'talking', 'about', False, False, 0, 0)),
(('talking', 'VBG', 3),
 ('We', "'re", 'talking', 'about', 'years', False, False, 0, 0)),
(('about', 'IN', 4),
 ("re", 'talking', 'about', 'years', 'ago', False, False, 0, 0)),
(('years', 'NNS', 5),
 ('talking', 'about', 'years', 'ago', 'before', False, False, 0, 0)),
(('ago', 'IN', 6),
 ('about', 'years', 'ago', 'before', 'anyone', False, False, 0, 0)),
(('before', 'IN', 7),
 ('years', 'ago', 'before', 'anyone', 'heard', False, False, 0, 0)),
(('anyone', 'NN', 8),
 ('ago', 'before', 'anyone', 'heard', 'of', False, False, 0, 0)),
(('heard', 'VBD', 9),
 ('before', 'anyone', 'heard', 'of', 'asbestos', False, False, 0, 0)),
(('of', 'IN', 10),
 ('anyone', 'heard', 'of', 'asbestos', 'having', False, False, 0, 0)),
(('asbestos', 'NN', 11),
 ('heard', 'of', 'asbestos', 'having', 'any', False, False, 0, 0)),
(('having', 'VBG', 12),
 ('of', 'asbestos', 'having', 'any', 'questionable', False, False, 0, 0)),
(('any', 'DT', 13),
 ('asbestos',
  'having',
  'any',
  'questionable',
  'properties',
  False,
  False,
  0,
  0)),
(('questionable', 'JJ', 14),
 ('having', 'any', 'questionable', 'properties', '.', False, False, 0, 0)),
(('properties', 'NNS', 15),
 ('any', 'questionable', 'properties', '.', 0, False, False, 0, 0)),
(('.', '.', 16),
 ('questionable', 'properties', '.', 0, 0, False, False, 0, 1)),
(('There', 'EX', 1), (0, 0, 'There', 'is', 'no', False, True, 1, 0)),

```

(('is', 'VBZ', 2), (0, 'There', 'is', 'no', 'asbestos', False, False, 0, 0)),
 (('no', 'DT', 3),
 ('There', 'is', 'no', 'asbestos', 'in', False, False, 0, 0)),
 (('asbestos', 'NN', 4),
 ('is', 'no', 'asbestos', 'in', 'our', False, False, 0, 0)),
 (('in', 'IN', 5),
 ('no', 'asbestos', 'in', 'our', 'products', False, False, 0, 0)),
 (('our', 'PRP\$', 6),
 ('asbestos', 'in', 'our', 'products', 'now', False, False, 0, 0)),
 (('products', 'NNS', 7),
 ('in', 'our', 'products', 'now', '.', False, False, 0, 0)),
 (('now', 'RB', 8), ('our', 'products', 'now', '.', '"', False, False, 0, 0)),
 (('.', '.', 9), ('products', 'now', '.', '"', 0, False, False, 0, 0)),
 (('"', '"', 10), ('now', '.', '"', 0, 0, False, False, 0, 1)),
 (('Neither', 'DT', 1),
 (0, 0, 'Neither', 'Lorillard', 'nor', False, True, 1, 0)),
 (('Lorillard', 'NNP', 2),
 (0, 'Neither', 'Lorillard', 'nor', 'the', False, True, 0, 0)),
 (('nor', 'CC', 3),
 ('Neither', 'Lorillard', 'nor', 'the', 'researchers', False, False, 0, 0)),
 (('the', 'DT', 4),
 ('Lorillard', 'nor', 'the', 'researchers', 'who', False, False, 0, 0)),
 (('researchers', 'NNS', 5),
 ('nor', 'the', 'researchers', 'who', '*T*-3', False, False, 0, 0)),
 (('who', 'WP', 6),
 ('the', 'researchers', 'who', '*T*-3', 'studied', False, False, 0, 0)),
 (('*T*-3', '-NONE-', 7),
 ('researchers', 'who', '*T*-3', 'studied', 'the', False, False, 0, 0)),
 (('studied', 'VBD', 8),
 ('who', '*T*-3', 'studied', 'the', 'workers', False, False, 0, 0)),
 (('the', 'DT', 9),
 ('*T*-3', 'studied', 'the', 'workers', 'were', False, False, 0, 0)),
 (('workers', 'NNS', 10),
 ('studied', 'the', 'workers', 'were', 'aware', False, False, 0, 0)),
 (('were', 'VBD', 11),
 ('the', 'workers', 'were', 'aware', 'of', False, False, 0, 0)),
 (('aware', 'JJ', 12),
 ('workers', 'were', 'aware', 'of', 'any', False, False, 0, 0)),
 (('of', 'IN', 13),
 ('were', 'aware', 'of', 'any', 'research', False, False, 0, 0)),
 (('any', 'DT', 14),
 ('aware', 'of', 'any', 'research', 'on', False, False, 0, 0)),
 (('research', 'NN', 15),
 ('of', 'any', 'research', 'on', 'smokers', False, False, 0, 0)),
 (('on', 'IN', 16),
 ('any', 'research', 'on', 'smokers', 'of', False, False, 0, 0)),
 (('smokers', 'NNS', 17),

('research', 'on', 'smokers', 'of', 'the', False, False, 0, 0)),
 (('of', 'IN', 18),
 ('on', 'smokers', 'of', 'the', 'Kent', False, False, 0, 0)),
 (('the', 'DT', 19),
 ('smokers', 'of', 'the', 'Kent', 'cigarettes', False, False, 0, 0)),
 (('Kent', 'NNP', 20),
 ('of', 'the', 'Kent', 'cigarettes', '.', False, True, 0, 0)),
 (('cigarettes', 'NNS', 21),
 ('the', 'Kent', 'cigarettes', '.', 0, False, False, 0, 0)),
 (('.', '.', 22), ('Kent', 'cigarettes', '.', 0, 0, False, False, 0, 1)),
 (('``', '``', 1), (0, 0, '``', 'We', 'have', False, False, 1, 0)),
 (('We', 'PRP', 2), (0, '``', 'We', 'have', 'no', False, True, 0, 0)),
 (('have', 'VBP', 3),
 ('``', 'We', 'have', 'no', 'useful', False, False, 0, 0)),
 (('no', 'DT', 4),
 ('We', 'have', 'no', 'useful', 'information', False, False, 0, 0)),
 (('useful', 'JJ', 5),
 ('have', 'no', 'useful', 'information', 'on', False, False, 0, 0)),
 (('information', 'NN', 6),
 ('no', 'useful', 'information', 'on', 'whether', False, False, 0, 0)),
 (('on', 'IN', 7),
 ('useful', 'information', 'on', 'whether', 'users', False, False, 0, 0)),
 (('whether', 'IN', 8),
 ('information', 'on', 'whether', 'users', 'are', False, False, 0, 0)),
 (('users', 'NNS', 9),
 ('on', 'whether', 'users', 'are', 'at', False, False, 0, 0)),
 (('are', 'VBP', 10),
 ('whether', 'users', 'are', 'at', 'risk', False, False, 0, 0)),
 (('at', 'IN', 11), ('users', 'are', 'at', 'risk', ',', False, False, 0, 0)),
 (('risk', 'NN', 12), ('are', 'at', 'risk', ',', '``', False, False, 0, 0)),
 ((',' , ',', 13), ('at', 'risk', ',', '``', 'said', False, False, 0, 0)),
 (('``', '``', 14), ('risk', ',', '``', 'said', '*T*-1', False, False, 0, 0)),
 (('said', 'VBD', 15),
 (',', '``', 'said', '*T*-1', 'James', False, False, 0, 0)),
 (('*T*-1', '-NONE-', 16),
 ('``', 'said', '*T*-1', 'James', 'A.', False, False, 0, 0)),
 (('James', 'NNP', 17),
 ('said', '*T*-1', 'James', 'A.', 'Talcott', False, True, 0, 0)),
 (('A.', 'NNP', 18),
 ('*T*-1', 'James', 'A.', 'Talcott', 'of', False, True, 0, 0)),
 (('Talcott', 'NNP', 19),
 ('James', 'A.', 'Talcott', 'of', 'Boston', False, True, 0, 0)),
 (('of', 'IN', 20),
 ('A.', 'Talcott', 'of', 'Boston', "'s", False, False, 0, 0)),
 (('Boston', 'NNP', 21),
 ('Talcott', 'of', 'Boston', "'s", 'Dana-Farber', False, True, 0, 0)),
 (('s", 'POS', 22),

('of', 'Boston', "'s", 'Dana-Farber', 'Cancer', False, False, 0, 0)),
 (('Dana-Farber', 'NNP', 23),
 ('Boston', "'s", 'Dana-Farber', 'Cancer', 'Institute', False, True, 0, 0)),
 (('Cancer', 'NNP', 24),
 ("'s", 'Dana-Farber', 'Cancer', 'Institute', '.', False, True, 0, 0)),
 (('Institute', 'NNP', 25),
 ('Dana-Farber', 'Cancer', 'Institute', '.', 0, False, True, 0, 0)),
 (('.', '.', 26), ('Cancer', 'Institute', '.', 0, 0, False, False, 0, 1)),
 (('Dr.', 'NNP', 1), (0, 0, 'Dr.', 'Talcott', 'led', False, True, 1, 0)),
 (('Talcott', 'NNP', 2), (0, 'Dr.', 'Talcott', 'led', 'a', False, True, 0, 0)),
 (('led', 'VBD', 3),
 ('Dr.', 'Talcott', 'led', 'a', 'team', False, False, 0, 0)),
 (('a', 'DT', 4), ('Talcott', 'led', 'a', 'team', 'of', False, False, 0, 0)),
 (('team', 'NN', 5),
 ('led', 'a', 'team', 'of', 'researchers', False, False, 0, 0)),
 (('of', 'IN', 6),
 ('a', 'team', 'of', 'researchers', 'from', False, False, 0, 0)),
 (('researchers', 'NNS', 7),
 ('team', 'of', 'researchers', 'from', 'the', False, False, 0, 0)),
 (('from', 'IN', 8),
 ('of', 'researchers', 'from', 'the', 'National', False, False, 0, 0)),
 (('the', 'DT', 9),
 ('researchers', 'from', 'the', 'National', 'Cancer', False, False, 0, 0)),
 (('National', 'NNP', 10),
 ('from', 'the', 'National', 'Cancer', 'Institute', False, True, 0, 0)),
 (('Cancer', 'NNP', 11),
 ('the', 'National', 'Cancer', 'Institute', 'and', False, True, 0, 0)),
 (('Institute', 'NNP', 12),
 ('National', 'Cancer', 'Institute', 'and', 'the', False, True, 0, 0)),
 (('and', 'CC', 13),
 ('Cancer', 'Institute', 'and', 'the', 'medical', False, False, 0, 0)),
 (('the', 'DT', 14),
 ('Institute', 'and', 'the', 'medical', 'schools', False, False, 0, 0)),
 (('medical', 'JJ', 15),
 ('and', 'the', 'medical', 'schools', 'of', False, False, 0, 0)),
 (('schools', 'NNS', 16),
 ('the', 'medical', 'schools', 'of', 'Harvard', False, False, 0, 0)),
 (('of', 'IN', 17),
 ('medical', 'schools', 'of', 'Harvard', 'University', False, False, 0, 0)),
 (('Harvard', 'NNP', 18),
 ('schools', 'of', 'Harvard', 'University', 'and', False, True, 0, 0)),
 (('University', 'NNP', 19),
 ('of', 'Harvard', 'University', 'and', 'Boston', False, True, 0, 0)),
 (('and', 'CC', 20),
 ('Harvard',
 'University',
 'and',

```

    'Boston',
    'University',
    False,
    False,
    0,
    0)),
(('Boston', 'NNP', 21),
 ('University', 'and', 'Boston', 'University', '.', False, True, 0, 0)),
(('University', 'NNP', 22),
 ('and', 'Boston', 'University', '.', 0, False, True, 0, 0)),
(('.', '.', 23), ('Boston', 'University', '.', 0, 0, False, False, 0, 1)),
(('The', 'DT', 1),
 (0, 0, 'The', 'Lorillard', 'spokeswoman', False, True, 1, 0)),
(('Lorillard', 'NNP', 2),
 (0, 'The', 'Lorillard', 'spokeswoman', 'said', False, True, 0, 0)),
(('spokeswoman', 'NN', 3),
 ('The', 'Lorillard', 'spokeswoman', 'said', '0', False, False, 0, 0)),
(('said', 'VBD', 4),
 ('Lorillard', 'spokeswoman', 'said', '0', 'asbestos', False, False, 0, 0)),
(('0', '-NONE-', 5),
 ('spokeswoman', 'said', '0', 'asbestos', 'was', True, False, 0, 0)),
(('asbestos', 'NN', 6),
 ('said', '0', 'asbestos', 'was', 'used', False, False, 0, 0)),
(('was', 'VBD', 7),
 ('0', 'asbestos', 'was', 'used', '*-1', False, False, 0, 0)),
(('used', 'VBN', 8),
 ('asbestos', 'was', 'used', '*-1', 'in', False, False, 0, 0)),
(('*-1', '-NONE-', 9),
 ('was', 'used', '*-1', 'in', '```', False, False, 0, 0)),
(('in', 'IN', 10), ('used', '*-1', 'in', '```', 'very', False, False, 0, 0)),
(('```', '```', 11), ('*-1', 'in', '```', 'very', 'modest', False, False, 0, 0)),
(('very', 'RB', 12),
 ('in', '```', 'very', 'modest', 'amounts', False, False, 0, 0)),
(('modest', 'JJ', 13),
 ('```', 'very', 'modest', 'amounts', '""', False, False, 0, 0)),
(('amounts', 'NNS', 14),
 ('very', 'modest', 'amounts', '""', 'in', False, False, 0, 0)),
(('""', '""', 15),
 ('modest', 'amounts', '""', 'in', '*', False, False, 0, 0)),
(('in', 'IN', 16),
 ('amounts', '""', 'in', '*', 'making', False, False, 0, 0)),
(('*', '-NONE-', 17),
 ('""', 'in', '*', 'making', 'paper', False, False, 0, 0)),
(('making', 'VBG', 18),
 ('in', '*', 'making', 'paper', 'for', False, False, 0, 0)),
(('paper', 'NN', 19),
 ('*', 'making', 'paper', 'for', 'the', False, False, 0, 0)),

```

```

(('for', 'IN', 20),
 ('making', 'paper', 'for', 'the', 'filters', False, False, 0, 0)),
(('the', 'DT', 21),
 ('paper', 'for', 'the', 'filters', 'in', False, False, 0, 0)),
(('filters', 'NNS', 22),
 ('for', 'the', 'filters', 'in', 'the', False, False, 0, 0)),
(('in', 'IN', 23),
 ('the', 'filters', 'in', 'the', 'early', False, False, 0, 0)),
(('the', 'DT', 24),
 ('filters', 'in', 'the', 'early', '1950s', False, False, 0, 0)),
(('early', 'JJ', 25),
 ('in', 'the', 'early', '1950s', 'and', False, False, 0, 0)),
(('1950s', 'CD', 26),
 ('the', 'early', '1950s', 'and', 'replaced', False, False, 0, 0)),
(('and', 'CC', 27),
 ('early', '1950s', 'and', 'replaced', '*-1', False, False, 0, 0)),
(('replaced', 'VBN', 28),
 ('1950s', 'and', 'replaced', '*-1', 'with', False, False, 0, 0)),
(('*-1', '-NONE-', 29),
 ('and', 'replaced', '*-1', 'with', 'a', False, False, 0, 0)),
(('with', 'IN', 30),
 ('replaced', '*-1', 'with', 'a', 'different', False, False, 0, 0)),
(('a', 'DT', 31),
 ('*-1', 'with', 'a', 'different', 'type', False, False, 0, 0)),
(('different', 'JJ', 32),
 ('with', 'a', 'different', 'type', 'of', False, False, 0, 0)),
(('type', 'NN', 33),
 ('a', 'different', 'type', 'of', 'filter', False, False, 0, 0)),
(('of', 'IN', 34),
 ('different', 'type', 'of', 'filter', 'in', False, False, 0, 0)),
(('filter', 'NN', 35),
 ('type', 'of', 'filter', 'in', '1956', False, False, 0, 0)),
(('in', 'IN', 36), ('of', 'filter', 'in', '1956', '.', False, False, 0, 0)),
(('1956', 'CD', 37), ('filter', 'in', '1956', '.', 0, True, False, 0, 0)),
(('.', '.', 38), ('in', '1956', '.', 0, 0, False, False, 0, 1)),
(('From', 'IN', 1), (0, 0, 'From', '1953', 'to', False, True, 1, 0)),
(('1953', 'CD', 2), (0, 'From', '1953', 'to', '1955', True, False, 0, 0)),
(('to', 'TO', 3), ('From', '1953', 'to', '1955', ',', False, False, 0, 0)),
(('1955', 'CD', 4), ('1953', 'to', '1955', ',', '9.8', True, False, 0, 0)),
(('', ',', 5), ('to', '1955', ',', '9.8', 'billion', False, False, 0, 0)),
(('9.8', 'CD', 6),
 ('1955', ',', '9.8', 'billion', 'Kent', False, False, 0, 0)),
(('billion', 'CD', 7),
 ('', '9.8', 'billion', 'Kent', 'cigarettes', False, False, 0, 0)),
(('Kent', 'NNP', 8),
 ('9.8', 'billion', 'Kent', 'cigarettes', 'with', False, True, 0, 0)),
(('cigarettes', 'NNS', 9),

```

('billion', 'Kent', 'cigarettes', 'with', 'the', False, False, 0, 0)),
 (('with', 'IN', 10),
 ('Kent', 'cigarettes', 'with', 'the', 'filters', False, False, 0, 0)),
 (('the', 'DT', 11),
 ('cigarettes', 'with', 'the', 'filters', 'were', False, False, 0, 0)),
 (('filters', 'NNS', 12),
 ('with', 'the', 'filters', 'were', 'sold', False, False, 0, 0)),
 (('were', 'VBD', 13),
 ('the', 'filters', 'were', 'sold', '*-3', False, False, 0, 0)),
 (('sold', 'VBN', 14),
 ('filters', 'were', 'sold', '*-3', ',', False, False, 0, 0)),
 (('*-3', '-NONE-', 15),
 ('were', 'sold', '*-3', ',', 'the', False, False, 0, 0)),
 ((' ', ',', ',', 16), ('sold', '*-3', ',', 'the', 'company', False, False, 0, 0)),
 (('the', 'DT', 17),
 ('*-3', ',', 'the', 'company', 'said', False, False, 0, 0)),
 (('company', 'NN', 18),
 (' ', 'the', 'company', 'said', '0', False, False, 0, 0)),
 (('said', 'VBD', 19),
 ('the', 'company', 'said', '0', '*T*-1', False, False, 0, 0)),
 (('0', '-NONE-', 20),
 ('company', 'said', '0', '*T*-1', '.', True, False, 0, 0)),
 (('*T*-1', '-NONE-', 21), ('said', '0', '*T*-1', '.', 0, False, False, 0, 0)),
 (('.', '.', 22), ('0', '*T*-1', '.', 0, 0, False, False, 0, 1)),
 (('Among', 'IN', 1), (0, 0, 'Among', '33', 'men', False, True, 1, 0)),
 (('33', 'CD', 2), (0, 'Among', '33', 'men', 'who', True, False, 0, 0)),
 (('men', 'NNS', 3),
 ('Among', '33', 'men', 'who', '*T*-4', False, False, 0, 0)),
 (('who', 'WP', 4),
 ('33', 'men', 'who', '*T*-4', 'worked', False, False, 0, 0)),
 (('*T*-4', '-NONE-', 5),
 ('men', 'who', '*T*-4', 'worked', 'closely', False, False, 0, 0)),
 (('worked', 'VBD', 6),
 ('who', '*T*-4', 'worked', 'closely', 'with', False, False, 0, 0)),
 (('closely', 'RB', 7),
 ('*T*-4', 'worked', 'closely', 'with', 'the', False, False, 0, 0)),
 (('with', 'IN', 8),
 ('worked', 'closely', 'with', 'the', 'substance', False, False, 0, 0)),
 (('the', 'DT', 9),
 ('closely', 'with', 'the', 'substance', ',', False, False, 0, 0)),
 (('substance', 'NN', 10),
 ('with', 'the', 'substance', ',', '28', False, False, 0, 0)),
 ((' ', ',', ',', 11),
 ('the', 'substance', ',', '28', '*ICH*-1', False, False, 0, 0)),
 (('28', 'CD', 12),
 ('substance', ',', '28', '*ICH*-1', 'have', True, False, 0, 0)),
 (('*ICH*-1', '-NONE-', 13),

('', '28', '*ICH*-1', 'have', 'died', False, False, 0, 0)),
 (('have', 'VBP', 14),
 ('28', '*ICH*-1', 'have', 'died', '--', False, False, 0, 0)),
 (('died', 'VBN', 15),
 ('*ICH*-1', 'have', 'died', '--', 'more', False, False, 0, 0)),
 (('--', ':', 16), ('have', 'died', '--', 'more', 'than', False, False, 0, 0)),
 (('more', 'JJ', 17),
 ('died', '--', 'more', 'than', 'three', False, False, 0, 0)),
 (('than', 'IN', 18),
 ('--', 'more', 'than', 'three', 'times', False, False, 0, 0)),
 (('three', 'CD', 19),
 ('more', 'than', 'three', 'times', 'the', False, False, 0, 0)),
 (('times', 'NNS', 20),
 ('than', 'three', 'times', 'the', 'expected', False, False, 0, 0)),
 (('the', 'DT', 21),
 ('three', 'times', 'the', 'expected', 'number', False, False, 0, 0)),
 (('expected', 'VBN', 22),
 ('times', 'the', 'expected', 'number', '.', False, False, 0, 0)),
 (('number', 'NN', 23),
 ('the', 'expected', 'number', '.', 0, False, False, 0, 0)),
 (('.', '.', 24), ('expected', 'number', '.', 0, 0, False, False, 0, 1)),
 (('Four', 'CD', 1), (0, 0, 'Four', 'of', 'the', False, True, 1, 0)),
 (('of', 'IN', 2), (0, 'Four', 'of', 'the', 'five', False, False, 0, 0)),
 (('the', 'DT', 3),
 ('Four', 'of', 'the', 'five', 'surviving', False, False, 0, 0)),
 (('five', 'CD', 4),
 ('of', 'the', 'five', 'surviving', 'workers', False, False, 0, 0)),
 (('surviving', 'VBG', 5),
 ('the', 'five', 'surviving', 'workers', 'have', False, False, 0, 0)),
 (('workers', 'NNS', 6),
 ('five',
 'surviving',
 'workers',
 'have',
 'asbestos-related',
 False,
 False,
 0,
 0)),
 (('have', 'VBP', 7),
 ('surviving',
 'workers',
 'have',
 'asbestos-related',
 'diseases',
 False,
 False,


```

0,
0)),
(('asbestos-related', 'JJ', 8),
 ('workers',
  'have',
  'asbestos-related',
  'diseases',
  ',',
  False,
  False,
  0,
  0)),
(('diseases', 'NNS', 9),
 ('have',
  'asbestos-related',
  'diseases',
  ',',
  'including',
  False,
  False,
  0,
  0)),
((',', ',', 10),
 ('asbestos-related',
  'diseases',
  ',',
  'including',
  'three',
  False,
  False,
  0,
  0)),
(('including', 'VBG', 11),
 ('diseases', ',', 'including', 'three', 'with', False, False, 0, 0)),
(('three', 'CD', 12),
 (',', 'including', 'three', 'with', 'recently', False, False, 0, 0)),
(('with', 'IN', 13),
 ('including', 'three', 'with', 'recently', 'diagnosed', False, False, 0, 0)),
(('recently', 'RB', 14),
 ('three', 'with', 'recently', 'diagnosed', 'cancer', False, False, 0, 0)),
(('diagnosed', 'VBN', 15),
 ('with', 'recently', 'diagnosed', 'cancer', '.', False, False, 0, 0)),
(('cancer', 'NN', 16),
 ('recently', 'diagnosed', 'cancer', '.', 0, False, False, 0, 0)),
(('.', '.', 17), ('diagnosed', 'cancer', '.', 0, 0, False, False, 0, 1)),
(('The', 'DT', 1), (0, 0, 'The', 'total', 'of', False, True, 1, 0)),
(('total', 'NN', 2), (0, 'The', 'total', 'of', '18', False, False, 0, 0)),

```

(('of', 'IN', 3), ('The', 'total', 'of', '18', 'deaths', False, False, 0, 0)),
 (('18', 'CD', 4), ('total', 'of', '18', 'deaths', 'from', True, False, 0, 0)),
 (('deaths', 'NNS', 5),
 ('of', '18', 'deaths', 'from', 'malignant', False, False, 0, 0)),
 (('from', 'IN', 6),
 ('18', 'deaths', 'from', 'malignant', 'mesothelioma', False, False, 0, 0)),
 (('malignant', 'JJ', 7),
 ('deaths', 'from', 'malignant', 'mesothelioma', ',', False, False, 0, 0)),
 (('mesothelioma', 'NN', 8),
 ('from', 'malignant', 'mesothelioma', ',', 'lung', False, False, 0, 0)),
 ((',', ',', ',', 9),
 ('malignant', 'mesothelioma', ',', 'lung', 'cancer', False, False, 0, 0)),
 (('lung', 'NN', 10),
 ('mesothelioma', ',', 'lung', 'cancer', 'and', False, False, 0, 0)),
 (('cancer', 'NN', 11),
 (',', 'lung', 'cancer', 'and', 'asbestosis', False, False, 0, 0)),
 (('and', 'CC', 12),
 ('lung', 'cancer', 'and', 'asbestosis', 'was', False, False, 0, 0)),
 (('asbestosis', 'NN', 13),
 ('cancer', 'and', 'asbestosis', 'was', 'far', False, False, 0, 0)),
 (('was', 'VBD', 14),
 ('and', 'asbestosis', 'was', 'far', 'higher', False, False, 0, 0)),
 (('far', 'RB', 15),
 ('asbestosis', 'was', 'far', 'higher', 'than', False, False, 0, 0)),
 (('higher', 'JJR', 16),
 ('was', 'far', 'higher', 'than', '*', False, False, 0, 0)),
 (('than', 'IN', 17),
 ('far', 'higher', 'than', '*', 'expected', False, False, 0, 0)),
 (('*', '-NONE-', 18),
 ('higher', 'than', '*', 'expected', '*?* ', False, False, 0, 0)),
 (('expected', 'VBN', 19),
 ('than', '*', 'expected', '*?* ', ',', False, False, 0, 0)),
 (('*?* ', '-NONE-', 20),
 ('*', 'expected', '*?* ', ',', 'the', False, False, 0, 0)),
 ((',', ',', ',', 21),
 ('expected', '*?* ', ',', 'the', 'researchers', False, False, 0, 0)),
 (('the', 'DT', 22),
 ('*?* ', ',', 'the', 'researchers', 'said', False, False, 0, 0)),
 (('researchers', 'NNS', 23),
 (',', 'the', 'researchers', 'said', '0', False, False, 0, 0)),
 (('said', 'VBD', 24),
 ('the', 'researchers', 'said', '0', '*T*-1', False, False, 0, 0)),
 (('0', '-NONE-', 25),
 ('researchers', 'said', '0', '*T*-1', '.', True, False, 0, 0)),
 (('*T*-1', '-NONE-', 26), ('said', '0', '*T*-1', '.', 0, False, False, 0, 0)),
 (('.', '.', 27), ('0', '*T*-1', '.', 0, 0, False, False, 0, 1)),
 (('`', '`', 1), (0, 0, '`', 'The', 'morbidity', False, False, 1, 0)),

(('The', 'DT', 2), (0, '', 'The', 'morbidity', 'rate', False, True, 0, 0)),
 (('morbidity', 'NN', 3),
 ('', 'The', 'morbidity', 'rate', 'is', False, False, 0, 0)),
 (('rate', 'NN', 4),
 ('The', 'morbidity', 'rate', 'is', 'a', False, False, 0, 0)),
 (('is', 'VBZ', 5),
 ('morbidity', 'rate', 'is', 'a', 'striking', False, False, 0, 0)),
 (('a', 'DT', 6),
 ('rate', 'is', 'a', 'striking', 'finding', False, False, 0, 0)),
 (('striking', 'JJ', 7),
 ('is', 'a', 'striking', 'finding', 'among', False, False, 0, 0)),
 (('finding', 'NN', 8),
 ('a', 'striking', 'finding', 'among', 'those', False, False, 0, 0)),
 (('among', 'IN', 9),
 ('striking', 'finding', 'among', 'those', 'of', False, False, 0, 0)),
 (('those', 'DT', 10),
 ('finding', 'among', 'those', 'of', 'us', False, False, 0, 0)),
 (('of', 'IN', 11), ('among', 'those', 'of', 'us', 'who', False, False, 0, 0)),
 (('us', 'PRP', 12),
 ('those', 'of', 'us', 'who', '*T*-5', False, False, 0, 0)),
 (('who', 'WP', 13),
 ('of', 'us', 'who', '*T*-5', 'study', False, False, 0, 0)),
 ((*T*-5', '-NONE-', 14),
 ('us', 'who', '*T*-5', 'study', 'asbestos-related', False, False, 0, 0)),
 (('study', 'VBP', 15),
 ('who',
 '*T*-5',
 'study',
 'asbestos-related',
 'diseases',
 False,
 False,
 0,
 0)),
 (('asbestos-related', 'JJ', 16),
 (*T*-5', 'study', 'asbestos-related', 'diseases', ',', False, False, 0, 0)),
 (('diseases', 'NNS', 17),
 ('study', 'asbestos-related', 'diseases', ',', '"', False, False, 0, 0)),
 ((' ', ', ', 18),
 ('asbestos-related', 'diseases', ',', '"', 'said', False, False, 0, 0)),
 (('"', '"', 19),
 ('diseases', ', ', '"', 'said', '*T*-1', False, False, 0, 0)),
 (('said', 'VBD', 20),
 (' ', '"', 'said', '*T*-1', 'Dr.', False, False, 0, 0)),
 ((*T*-1', '-NONE-', 21),
 ('"', 'said', '*T*-1', 'Dr.', 'Talcott', False, False, 0, 0)),
 (('Dr.', 'NNP', 22),

('said', '*T*-1', 'Dr.', 'Talcott', '.', False, True, 0, 0)),
 (('Talcott', 'NNP', 23),
 ('*T*-1', 'Dr.', 'Talcott', '.', 0, False, True, 0, 0)),
 (('.', '.', 24), ('Dr.', 'Talcott', '.', 0, 0, False, False, 0, 1)),
 (('The', 'DT', 1), (0, 0, 'The', 'percentage', 'of', False, True, 1, 0)),
 (('percentage', 'NN', 2),
 (0, 'The', 'percentage', 'of', 'lung', False, False, 0, 0)),
 (('of', 'IN', 3),
 ('The', 'percentage', 'of', 'lung', 'cancer', False, False, 0, 0)),
 (('lung', 'NN', 4),
 ('percentage', 'of', 'lung', 'cancer', 'deaths', False, False, 0, 0)),
 (('cancer', 'NN', 5),
 ('of', 'lung', 'cancer', 'deaths', 'among', False, False, 0, 0)),
 (('deaths', 'NNS', 6),
 ('lung', 'cancer', 'deaths', 'among', 'the', False, False, 0, 0)),
 (('among', 'IN', 7),
 ('cancer', 'deaths', 'among', 'the', 'workers', False, False, 0, 0)),
 (('the', 'DT', 8),
 ('deaths', 'among', 'the', 'workers', 'at', False, False, 0, 0)),
 (('workers', 'NNS', 9),
 ('among', 'the', 'workers', 'at', 'the', False, False, 0, 0)),
 (('at', 'IN', 10),
 ('the', 'workers', 'at', 'the', 'West', False, False, 0, 0)),
 (('the', 'DT', 11),
 ('workers', 'at', 'the', 'West', 'Groton', False, False, 0, 0)),
 (('West', 'NNP', 12),
 ('at', 'the', 'West', 'Groton', ',', False, True, 0, 0)),
 (('Groton', 'NNP', 13),
 ('the', 'West', 'Groton', ',', 'Mass.', False, True, 0, 0)),
 ((',', ',', 14), ('West', 'Groton', ',', 'Mass.', ',', False, False, 0, 0)),
 (('Mass.', 'NNP', 15),
 ('Groton', ',', 'Mass.', ',', 'paper', False, True, 0, 0)),
 ((',', ',', 16), ((',', 'Mass.', ',', 'paper', 'factory', False, False, 0, 0)),
 (('paper', 'NN', 17),
 ('Mass.', ',', 'paper', 'factory', 'appears', False, False, 0, 0)),
 (('factory', 'NN', 18),
 ((',', 'paper', 'factory', 'appears', '*-1', False, False, 0, 0)),
 (('appears', 'VBZ', 19),
 ('paper', 'factory', 'appears', '*-1', 'to', False, False, 0, 0)),
 ((*-1', '-NONE-', 20),
 ('factory', 'appears', '*-1', 'to', 'be', False, False, 0, 0)),
 (('to', 'TO', 21), ('appears', '*-1', 'to', 'be', 'the', False, False, 0, 0)),
 (('be', 'VB', 22), (*-1', 'to', 'be', 'the', 'highest', False, False, 0, 0)),
 (('the', 'DT', 23),
 ('to', 'be', 'the', 'highest', 'for', False, False, 0, 0)),
 (('highest', 'JJS', 24),
 ('be', 'the', 'highest', 'for', 'any', False, False, 0, 0)),

(('for', 'IN', 25),
 ('the', 'highest', 'for', 'any', 'asbestos', False, False, 0, 0)),
 (('any', 'DT', 26),
 ('highest', 'for', 'any', 'asbestos', 'workers', False, False, 0, 0)),
 (('asbestos', 'NN', 27),
 ('for', 'any', 'asbestos', 'workers', 'studied', False, False, 0, 0)),
 (('workers', 'NNS', 28),
 ('any', 'asbestos', 'workers', 'studied', '*', False, False, 0, 0)),
 (('studied', 'VBN', 29),
 ('asbestos', 'workers', 'studied', '*', 'in', False, False, 0, 0)),
 (('*', '-NONE-', 30),
 ('workers', 'studied', '*', 'in', 'Western', False, False, 0, 0)),
 (('in', 'IN', 31),
 ('studied', '*', 'in', 'Western', 'industrialized', False, False, 0, 0)),
 (('Western', 'JJ', 32),
 ('*', 'in', 'Western', 'industrialized', 'countries', False, True, 0, 0)),
 (('industrialized', 'VBN', 33),
 ('in', 'Western', 'industrialized', 'countries', ',', False, False, 0, 0)),
 (('countries', 'NNS', 34),
 ('Western', 'industrialized', 'countries', ',', 'he', False, False, 0, 0)),
 ((' ', ' ', ' ', 35),
 ('industrialized', 'countries', ',', 'he', 'said', False, False, 0, 0)),
 (('he', 'PRP', 36),
 ('countries', ',', 'he', 'said', '0', False, False, 0, 0)),
 (('said', 'VBD', 37), (' ', 'he', 'said', '0', '*T*-2', False, False, 0, 0)),
 (('0', '-NONE-', 38), ('he', 'said', '0', '*T*-2', '.', True, False, 0, 0)),
 ((' '*T*-2', '-NONE-', 39), ('said', '0', '*T*-2', '.', 0, False, False, 0, 0)),
 (('.', '.', 40), ('0', '*T*-2', '.', 0, 0, False, False, 0, 1)),
 (('The', 'DT', 1), (0, 0, 'The', 'plant', ',', False, True, 1, 0)),
 (('plant', 'NN', 2), (0, 'The', 'plant', ',', 'which', False, False, 0, 0)),
 ((' ', ' ', ' ', 3), ('The', 'plant', ',', 'which', '*T*-1', False, False, 0, 0)),
 (('which', 'WDT', 4),
 ('plant', ',', 'which', '*T*-1', 'is', False, False, 0, 0)),
 ((' '*T*-1', '-NONE-', 5),
 (' ', 'which', '*T*-1', 'is', 'owned', False, False, 0, 0)),
 (('is', 'VBZ', 6),
 ('which', '*T*-1', 'is', 'owned', '*-4', False, False, 0, 0)),
 (('owned', 'VBN', 7),
 (' '*T*-1', 'is', 'owned', '*-4', 'by', False, False, 0, 0)),
 ((' '*-4', '-NONE-', 8),
 ('is', 'owned', '*-4', 'by', 'Hollingsworth', False, False, 0, 0)),
 (('by', 'IN', 9),
 ('owned', '*-4', 'by', 'Hollingsworth', '&', False, False, 0, 0)),
 (('Hollingsworth', 'NNP', 10),
 (' *-4', 'by', 'Hollingsworth', '&', 'Vose', False, True, 0, 0)),
 (('&', 'CC', 11),
 ('by', 'Hollingsworth', '&', 'Vose', 'Co.', False, False, 0, 0)),

(('Vose', 'NNP', 12),
 ('Hollingsworth', '&', 'Vose', 'Co.', ',', False, True, 0, 0)),
 (('Co.', 'NNP', 13), ('&', 'Vose', 'Co.', ',', 'was', False, True, 0, 0)),
 ((',', ',', 14), ('Vose', 'Co.', ',', 'was', 'under', False, False, 0, 0)),
 (('was', 'VBD', 15),
 ('Co.', ',', 'was', 'under', 'contract', False, False, 0, 0)),
 (('under', 'IN', 16),
 ((',', 'was', 'under', 'contract', '*ICH*-2', False, False, 0, 0)),
 (('contract', 'NN', 17),
 ('was', 'under', 'contract', '*ICH*-2', 'with', False, False, 0, 0)),
 ((*ICH*-2', '-NONE-', 18),
 ('under', 'contract', '*ICH*-2', 'with', 'Lorillard', False, False, 0, 0)),
 (('with', 'IN', 19),
 ('contract', '*ICH*-2', 'with', 'Lorillard', '*', False, False, 0, 0)),
 (('Lorillard', 'NN', 20),
 (*ICH*-2', 'with', 'Lorillard', '*', 'to', False, True, 0, 0)),
 ((*', '-NONE-', 21),
 ('with', 'Lorillard', '*', 'to', 'make', False, False, 0, 0)),
 (('to', 'TO', 22),
 ('Lorillard', '*', 'to', 'make', 'the', False, False, 0, 0)),
 (('make', 'VB', 23),
 (*', 'to', 'make', 'the', 'cigarette', False, False, 0, 0)),
 (('the', 'DT', 24),
 ('to', 'make', 'the', 'cigarette', 'filters', False, False, 0, 0)),
 (('cigarette', 'NN', 25),
 ('make', 'the', 'cigarette', 'filters', '.', False, False, 0, 0)),
 (('filters', 'NNS', 26),
 ('the', 'cigarette', 'filters', '.', 0, False, False, 0, 0)),
 (('.', '.', 27), ('cigarette', 'filters', '.', 0, 0, False, False, 0, 1)),
 (('The', 'DT', 1), (0, 0, 'The', 'finding', 'probably', False, True, 1, 0)),
 (('finding', 'NN', 2),
 (0, 'The', 'finding', 'probably', 'will', False, False, 0, 0)),
 (('probably', 'RB', 3),
 ('The', 'finding', 'probably', 'will', 'support', False, False, 0, 0)),
 (('will', 'MD', 4),
 ('finding', 'probably', 'will', 'support', 'those', False, False, 0, 0)),
 (('support', 'VB', 5),
 ('probably', 'will', 'support', 'those', 'who', False, False, 0, 0)),
 (('those', 'DT', 6),
 ('will', 'support', 'those', 'who', '*T*-6', False, False, 0, 0)),
 (('who', 'WP', 7),
 ('support', 'those', 'who', '*T*-6', 'argue', False, False, 0, 0)),
 ((*T*-6', '-NONE-', 8),
 ('those', 'who', '*T*-6', 'argue', 'that', False, False, 0, 0)),
 (('argue', 'VBP', 9),
 ('who', '*T*-6', 'argue', 'that', 'the', False, False, 0, 0)),
 (('that', 'IN', 10),

(*T*-6, 'argue', 'that', 'the', 'U.S.', False, False, 0, 0)),
 (('the', 'DT', 11),
 ('argue', 'that', 'the', 'U.S.', 'should', False, False, 0, 0)),
 (('U.S.', 'NNP', 12),
 ('that', 'the', 'U.S.', 'should', 'regulate', False, True, 0, 0)),
 (('should', 'MD', 13),
 ('the', 'U.S.', 'should', 'regulate', 'the', False, False, 0, 0)),
 (('regulate', 'VB', 14),
 ('U.S.', 'should', 'regulate', 'the', 'class', False, False, 0, 0)),
 (('the', 'DT', 15),
 ('should', 'regulate', 'the', 'class', 'of', False, False, 0, 0)),
 (('class', 'NN', 16),
 ('regulate', 'the', 'class', 'of', 'asbestos', False, False, 0, 0)),
 (('of', 'IN', 17),
 ('the', 'class', 'of', 'asbestos', 'including', False, False, 0, 0)),
 (('asbestos', 'NN', 18),
 ('class', 'of', 'asbestos', 'including', 'crocidolite', False, False, 0, 0)),
 (('including', 'VBG', 19),
 ('of', 'asbestos', 'including', 'crocidolite', 'more', False, False, 0, 0)),
 (('crocidolite', 'NN', 20),
 ('asbestos',
 'including',
 'crocidolite',
 'more',
 'stringently',
 False,
 False,
 0,
 0)),
 (('more', 'RBR', 21),
 ('including',
 'crocidolite',
 'more',
 'stringently',
 'than',
 False,
 False,
 0,
 0)),
 (('stringently', 'RB', 22),
 ('crocidolite', 'more', 'stringently', 'than', 'the', False, False, 0, 0)),
 (('than', 'IN', 23),
 ('more', 'stringently', 'than', 'the', 'common', False, False, 0, 0)),
 (('the', 'DT', 24),
 ('stringently', 'than', 'the', 'common', 'kind', False, False, 0, 0)),
 (('common', 'JJ', 25),
 ('than', 'the', 'common', 'kind', 'of', False, False, 0, 0)),

(('kind', 'NN', 26),
 ('the', 'common', 'kind', 'of', 'asbestos', False, False, 0, 0)),
 (('of', 'IN', 27),
 ('common', 'kind', 'of', 'asbestos', ',', False, False, 0, 0)),
 (('asbestos', 'NN', 28),
 ('kind', 'of', 'asbestos', ',', 'chrysotile', False, False, 0, 0)),
 ((',', ',', 29),
 ('of', 'asbestos', ',', 'chrysotile', ',', False, False, 0, 0)),
 (('chrysotile', 'NN', 30),
 ('asbestos', ',', 'chrysotile', ',', 'found', False, False, 0, 0)),
 ((',', ',', 31), ((',', 'chrysotile', ',', 'found', '*', False, False, 0, 0)),
 (('found', 'VBN', 32),
 ('chrysotile', ',', 'found', '*', 'in', False, False, 0, 0)),
 (('*', '-NONE-', 33), ((',', 'found', '*', 'in', 'most', False, False, 0, 0)),
 (('in', 'IN', 34),
 ('found', '*', 'in', 'most', 'schools', False, False, 0, 0)),
 (('most', 'JJS', 35),
 ('*', 'in', 'most', 'schools', 'and', False, False, 0, 0)),
 (('schools', 'NNS', 36),
 ('in', 'most', 'schools', 'and', 'other', False, False, 0, 0)),
 (('and', 'CC', 37),
 ('most', 'schools', 'and', 'other', 'buildings', False, False, 0, 0)),
 (('other', 'JJ', 38),
 ('schools', 'and', 'other', 'buildings', ',', False, False, 0, 0)),
 (('buildings', 'NNS', 39),
 ('and', 'other', 'buildings', ',', 'Dr.', False, False, 0, 0)),
 ((',', ',', 40),
 ('other', 'buildings', ',', 'Dr.', 'Talcott', False, False, 0, 0)),
 (('Dr.', 'NNP', 41),
 ('buildings', ',', 'Dr.', 'Talcott', 'said', False, True, 0, 0)),
 (('Talcott', 'NNP', 42),
 ((',', 'Dr.', 'Talcott', 'said', '0', False, True, 0, 0)),
 (('said', 'VBD', 43),
 ('Dr.', 'Talcott', 'said', '0', '*T*-1', False, False, 0, 0)),
 (('0', '-NONE-', 44),
 ('Talcott', 'said', '0', '*T*-1', '.', True, False, 0, 0)),
 (('*T*-1', '-NONE-', 45), ('said', '0', '*T*-1', '.', 0, False, False, 0, 0)),
 (('.', '.', 46), ('0', '*T*-1', '.', 0, 0, False, False, 0, 1)),
 (('The', 'DT', 1), (0, 0, 'The', 'U.S.', 'is', False, True, 1, 0)),
 (('U.S.', 'NNP', 2), (0, 'The', 'U.S.', 'is', 'one', False, True, 0, 0)),
 (('is', 'VBZ', 3), ('The', 'U.S.', 'is', 'one', 'of', False, False, 0, 0)),
 (('one', 'CD', 4), ('U.S.', 'is', 'one', 'of', 'the', False, False, 0, 0)),
 (('of', 'IN', 5), ('is', 'one', 'of', 'the', 'few', False, False, 0, 0)),
 (('the', 'DT', 6),
 ('one', 'of', 'the', 'few', 'industrialized', False, False, 0, 0)),
 (('few', 'JJ', 7),
 ('of', 'the', 'few', 'industrialized', 'nations', False, False, 0, 0)),

(('industrialized', 'VBN', 8),
 ('the', 'few', 'industrialized', 'nations', 'that', False, False, 0, 0)),
 (('nations', 'NNS', 9),
 ('few', 'industrialized', 'nations', 'that', '*T*-7', False, False, 0, 0)),
 (('that', 'WDT', 10),
 ('industrialized', 'nations', 'that', '*T*-7', 'does', False, False, 0, 0)),
 ((*T*-7', '-NONE-', 11),
 ('nations', 'that', '*T*-7', 'does', "n't", False, False, 0, 0)),
 (('does', 'VBZ', 12),
 ('that', '*T*-7', 'does', "n't", 'have', False, False, 0, 0)),
 (('n't", 'RB', 13),
 (*T*-7', 'does', "n't", 'have', 'a', False, False, 0, 0)),
 (('have', 'VB', 14),
 ('does', "n't", 'have', 'a', 'higher', False, False, 0, 0)),
 (('a', 'DT', 15),
 ("n't", 'have', 'a', 'higher', 'standard', False, False, 0, 0)),
 (('higher', 'JJR', 16),
 ('have', 'a', 'higher', 'standard', 'of', False, False, 0, 0)),
 (('standard', 'NN', 17),
 ('a', 'higher', 'standard', 'of', 'regulation', False, False, 0, 0)),
 (('of', 'IN', 18),
 ('higher', 'standard', 'of', 'regulation', 'for', False, False, 0, 0)),
 (('regulation', 'NN', 19),
 ('standard', 'of', 'regulation', 'for', 'the', False, False, 0, 0)),
 (('for', 'IN', 20),
 ('of', 'regulation', 'for', 'the', 'smooth', False, False, 0, 0)),
 (('the', 'DT', 21),
 ('regulation', 'for', 'the', 'smooth', ',', False, False, 0, 0)),
 (('smooth', 'JJ', 22),
 ('for', 'the', 'smooth', ',', 'needle-like', False, False, 0, 0)),
 (('needle-like', 'JJ', 24),
 ('smooth', ',', 'needle-like', 'fibers', 'such', False, False, 0, 0)),
 (('fibers', 'NNS', 25),
 (',', 'needle-like', 'fibers', 'such', 'as', False, False, 0, 0)),
 (('such', 'JJ', 26),
 ('needle-like', 'fibers', 'such', 'as', 'crocidolite', False, False, 0, 0)),
 (('as', 'IN', 27),
 ('fibers', 'such', 'as', 'crocidolite', 'that', False, False, 0, 0)),
 (('crocidolite', 'NN', 28),
 ('such', 'as', 'crocidolite', 'that', '*T*-1', False, False, 0, 0)),
 (('that', 'WDT', 29),
 ('as', 'crocidolite', 'that', '*T*-1', 'are', False, False, 0, 0)),
 ((*T*-1', '-NONE-', 30),
 ('crocidolite', 'that', '*T*-1', 'are', 'classified', False, False, 0, 0)),
 (('are', 'VBP', 31),

('that', '*T*-1', 'are', 'classified', '*-5', False, False, 0, 0)),
 (('classified', 'VBN', 32),
 ('*T*-1', 'are', 'classified', '*-5', 'as', False, False, 0, 0)),
 (('*-5', '-NONE-', 33),
 ('are', 'classified', '*-5', 'as', 'amphobiles', False, False, 0, 0)),
 (('as', 'IN', 34),
 ('classified', '*-5', 'as', 'amphobiles', ',', False, False, 0, 0)),
 (('amphobiles', 'NNS', 35),
 ('*-5', 'as', 'amphobiles', ',', 'according', False, False, 0, 0)),
 ((' ', ' ', ' ', 36),
 ('as', 'amphobiles', ',', 'according', 'to', False, False, 0, 0)),
 (('according', 'VBG', 37),
 ('amphobiles', ',', 'according', 'to', 'Brooke', False, False, 0, 0)),
 (('to', 'TO', 38),
 (' ', 'according', 'to', 'Brooke', 'T.', False, False, 0, 0)),
 (('Brooke', 'NNP', 39),
 ('according', 'to', 'Brooke', 'T.', 'Mossman', False, True, 0, 0)),
 (('T.', 'NNP', 40),
 ('to', 'Brooke', 'T.', 'Mossman', ',', False, True, 0, 0)),
 (('Mossman', 'NNP', 41),
 ('Brooke', 'T.', 'Mossman', ',', 'a', False, True, 0, 0)),
 ((' ', ' ', ' ', 42),
 ('T.', 'Mossman', ',', 'a', 'professor', False, False, 0, 0)),
 (('a', 'DT', 43),
 ('Mossman', ',', 'a', 'professor', 'of', False, False, 0, 0)),
 (('professor', 'NN', 44),
 (' ', 'a', 'professor', 'of', 'pathlogy', False, False, 0, 0)),
 (('of', 'IN', 45),
 ('a', 'professor', 'of', 'pathlogy', 'at', False, False, 0, 0)),
 (('pathlogy', 'NN', 46),
 ('professor', 'of', 'pathlogy', 'at', 'the', False, False, 0, 0)),
 (('at', 'IN', 47),
 ('of', 'pathlogy', 'at', 'the', 'University', False, False, 0, 0)),
 (('the', 'DT', 48),
 ('pathlogy', 'at', 'the', 'University', 'of', False, False, 0, 0)),
 (('University', 'NNP', 49),
 ('at', 'the', 'University', 'of', 'Vermont', False, True, 0, 0)),
 (('of', 'IN', 50),
 ('the', 'University', 'of', 'Vermont', 'College', False, False, 0, 0)),
 (('Vermont', 'NNP', 51),
 ('University', 'of', 'Vermont', 'College', 'of', False, True, 0, 0)),
 (('College', 'NNP', 52),
 ('of', 'Vermont', 'College', 'of', 'Medicine', False, True, 0, 0)),
 (('of', 'IN', 53),
 ('Vermont', 'College', 'of', 'Medicine', '.', False, False, 0, 0)),
 (('Medicine', 'NNP', 54),
 ('College', 'of', 'Medicine', '.', 0, False, True, 0, 0)),

(('.', '.', 55), ('of', 'Medicine', '.', 0, 0, False, False, 0, 1)),
 (('More', 'RBR', 1),
 (0, 0, 'More', 'common', 'chrysotile', False, True, 1, 0)),
 (('common', 'JJ', 2),
 (0, 'More', 'common', 'chrysotile', 'fibers', False, False, 0, 0)),
 (('chrysotile', 'NN', 3),
 ('More', 'common', 'chrysotile', 'fibers', 'are', False, False, 0, 0)),
 (('fibers', 'NNS', 4),
 ('common', 'chrysotile', 'fibers', 'are', 'curly', False, False, 0, 0)),
 (('are', 'VBP', 5),
 ('chrysotile', 'fibers', 'are', 'curly', 'and', False, False, 0, 0)),
 (('curly', 'JJ', 6),
 ('fibers', 'are', 'curly', 'and', 'are', False, False, 0, 0)),
 (('and', 'CC', 7),
 ('are', 'curly', 'and', 'are', 'more', False, False, 0, 0)),
 (('are', 'VBP', 8),
 ('curly', 'and', 'are', 'more', 'easily', False, False, 0, 0)),
 (('more', 'RBR', 9),
 ('and', 'are', 'more', 'easily', 'rejected', False, False, 0, 0)),
 (('easily', 'RB', 10),
 ('are', 'more', 'easily', 'rejected', '*-1', False, False, 0, 0)),
 (('rejected', 'VBN', 11),
 ('more', 'easily', 'rejected', '*-1', 'by', False, False, 0, 0)),
 (('*-1', '-NONE-', 12),
 ('easily', 'rejected', '*-1', 'by', 'the', False, False, 0, 0)),
 (('by', 'IN', 13),
 ('rejected', '*-1', 'by', 'the', 'body', False, False, 0, 0)),
 (('the', 'DT', 14), ('*-1', 'by', 'the', 'body', ',', False, False, 0, 0)),
 (('body', 'NN', 15), ('by', 'the', 'body', ',', 'Dr.', False, False, 0, 0)),
 ((' ', ' ', 16), ('the', 'body', ',', 'Dr.', 'Mossman', False, False, 0, 0)),
 (('Dr.', 'NNP', 17),
 ('body', ',', 'Dr.', 'Mossman', 'explained', False, True, 0, 0)),
 (('Mossman', 'NNP', 18),
 (' ', 'Dr.', 'Mossman', 'explained', '0', False, True, 0, 0)),
 (('explained', 'VBD', 19),
 ('Dr.', 'Mossman', 'explained', '0', '*T*-2', False, False, 0, 0)),
 (('0', '-NONE-', 20),
 ('Mossman', 'explained', '0', '*T*-2', '.', True, False, 0, 0)),
 (('*T*-2', '-NONE-', 21),
 ('explained', '0', '*T*-2', '.', 0, False, False, 0, 0)),
 (('.', '.', 22), ('0', '*T*-2', '.', 0, 0, False, False, 0, 1)),
 (('In', 'IN', 1), (0, 0, 'In', 'July', ',', False, True, 1, 0)),
 (('July', 'NNP', 2), (0, 'In', 'July', ',', 'the', False, True, 0, 0)),
 ((' ', ' ', 3),
 ('In', 'July', ',', 'the', 'Environmental', False, False, 0, 0)),
 (('the', 'DT', 4),
 ('July', ',', 'the', 'Environmental', 'Protection', False, False, 0, 0)),

```

(('Environmental', 'NNP', 5),
 ('', 'the', 'Environmental', 'Protection', 'Agency', False, True, 0, 0)),
(('Protection', 'NNP', 6),
 ('the',
 'Environmental',
 'Protection',
 'Agency',
 'imposed',
 False,
 True,
 0,
 0)),
(('Agency', 'NNP', 7),
 ('Environmental',
 'Protection',
 'Agency',
 'imposed',
 'a',
 False,
 True,
 0,
 0)),
(('imposed', 'VBD', 8),
 ('Protection', 'Agency', 'imposed', 'a', 'gradual', False, False, 0, 0)),
(('a', 'DT', 9),
 ('Agency', 'imposed', 'a', 'gradual', 'ban', False, False, 0, 0)),
(('gradual', 'JJ', 10),
 ('imposed', 'a', 'gradual', 'ban', 'on', False, False, 0, 0)),
(('ban', 'NN', 11),
 ('a', 'gradual', 'ban', 'on', 'virtually', False, False, 0, 0)),
(('on', 'IN', 12),
 ('gradual', 'ban', 'on', 'virtually', 'all', False, False, 0, 0)),
(('virtually', 'RB', 13),
 ('ban', 'on', 'virtually', 'all', 'uses', False, False, 0, 0)),
(('all', 'DT', 14),
 ('on', 'virtually', 'all', 'uses', 'of', False, False, 0, 0)),
(('uses', 'NNS', 15),
 ('virtually', 'all', 'uses', 'of', 'asbestos', False, False, 0, 0)),
(('of', 'IN', 16),
 ('all', 'uses', 'of', 'asbestos', '.', False, False, 0, 0)),
(('asbestos', 'NN', 17),
 ('uses', 'of', 'asbestos', '.', 0, False, False, 0, 0)),
(('.', '.', 18), ('of', 'asbestos', '.', 0, 0, False, False, 0, 1)),
(('By', 'IN', 1), (0, 0, 'By', '1997', ',', False, True, 1, 0)),
(('1997', 'CD', 2), (0, 'By', '1997', ',', 'almost', True, False, 0, 0)),
(('', ',', ',', 3), ('By', '1997', ',', 'almost', 'all', False, False, 0, 0)),
(('almost', 'RB', 4),

```

```

('1997', ',', 'almost', 'all', 'remaining', False, False, 0, 0)),
(('all', 'DT', 5),
 ('', 'almost', 'all', 'remaining', 'uses', False, False, 0, 0)),
(('remaining', 'VBG', 6),
 ('almost', 'all', 'remaining', 'uses', 'of', False, False, 0, 0)),
(('uses', 'NNS', 7),
 ('all', 'remaining', 'uses', 'of', 'cancer-causing', False, False, 0, 0)),
(('of', 'IN', 8),
 ('remaining',
  'uses',
  'of',
  'cancer-causing',
  'asbestos',
  False,
  False,
  0,
  0)),
(('cancer-causing', 'JJ', 9),
 ('uses', 'of', 'cancer-causing', 'asbestos', 'will', False, False, 0, 0)),
(('asbestos', 'NN', 10),
 ('of', 'cancer-causing', 'asbestos', 'will', 'be', False, False, 0, 0)),
(('will', 'MD', 11),
 ('cancer-causing',
  'asbestos',
  'will',
  'be',
  'outlawed',
  False,
  False,
  0,
  0)),
(('be', 'VB', 12),
 ('asbestos', 'will', 'be', 'outlawed', '*-6', False, False, 0, 0)),
(('outlawed', 'VBN', 13),
 ('will', 'be', 'outlawed', '*-6', '.', False, False, 0, 0)),
(('*-6', '-NONE-', 14),
 ('be', 'outlawed', '*-6', '.', 0, False, False, 0, 0)),
(('.', '.', 15), ('outlawed', '*-6', '.', 0, 0, False, False, 0, 1)),
(('About', 'IN', 1), (0, 0, 'About', '160', 'workers', False, True, 1, 0)),
(('160', 'CD', 2), (0, 'About', '160', 'workers', 'at', True, False, 0, 0)),
(('workers', 'NNS', 3),
 ('About', '160', 'workers', 'at', 'a', False, False, 0, 0)),
(('at', 'IN', 4),
 ('160', 'workers', 'at', 'a', 'factory', False, False, 0, 0)),
(('a', 'DT', 5),
 ('workers', 'at', 'a', 'factory', 'that', False, False, 0, 0)),
(('factory', 'NN', 6),

```

('at', 'a', 'factory', 'that', '*T*-8', False, False, 0, 0)),
 (('that', 'WDT', 7),
 ('a', 'factory', 'that', '*T*-8', 'made', False, False, 0, 0)),
 (('T*-8', '-NONE-', 8),
 ('factory', 'that', '*T*-8', 'made', 'paper', False, False, 0, 0)),
 (('made', 'VBD', 9),
 ('that', '*T*-8', 'made', 'paper', 'for', False, False, 0, 0)),
 (('paper', 'NN', 10),
 ('T*-8', 'made', 'paper', 'for', 'the', False, False, 0, 0)),
 (('for', 'IN', 11),
 ('made', 'paper', 'for', 'the', 'Kent', False, False, 0, 0)),
 (('the', 'DT', 12),
 ('paper', 'for', 'the', 'Kent', 'filters', False, False, 0, 0)),
 (('Kent', 'NNP', 13),
 ('for', 'the', 'Kent', 'filters', 'were', False, True, 0, 0)),
 (('filters', 'NNS', 14),
 ('the', 'Kent', 'filters', 'were', 'exposed', False, False, 0, 0)),
 (('were', 'VBD', 15),
 ('Kent', 'filters', 'were', 'exposed', '*-7', False, False, 0, 0)),
 (('exposed', 'VBN', 16),
 ('filters', 'were', 'exposed', '*-7', 'to', False, False, 0, 0)),
 (('*-7', '-NONE-', 17),
 ('were', 'exposed', '*-7', 'to', 'asbestos', False, False, 0, 0)),
 (('to', 'TO', 18),
 ('exposed', '*-7', 'to', 'asbestos', 'in', False, False, 0, 0)),
 (('asbestos', 'NN', 19),
 ('*-7', 'to', 'asbestos', 'in', 'the', False, False, 0, 0)),
 (('in', 'IN', 20),
 ('to', 'asbestos', 'in', 'the', '1950s', False, False, 0, 0)),
 (('the', 'DT', 21),
 ('asbestos', 'in', 'the', '1950s', '.', False, False, 0, 0)),
 (('1950s', 'CD', 22), ('in', 'the', '1950s', '.', 0, False, False, 0, 0)),
 (('.', '.', 23), ('the', '1950s', '.', 0, 0, False, False, 0, 1)),
 (('Areas', 'NNS', 1), (0, 0, 'Areas', 'of', 'the', False, True, 1, 0)),
 (('of', 'IN', 2), (0, 'Areas', 'of', 'the', 'factory', False, False, 0, 0)),
 (('the', 'DT', 3),
 ('Areas', 'of', 'the', 'factory', '*ICH*-2', False, False, 0, 0)),
 (('factory', 'NN', 4),
 ('of', 'the', 'factory', '*ICH*-2', 'were', False, False, 0, 0)),
 (('ICH*-2', '-NONE-', 5),
 ('the', 'factory', '*ICH*-2', 'were', 'particularly', False, False, 0, 0)),
 (('were', 'VBD', 6),
 ('factory', '*ICH*-2', 'were', 'particularly', 'dusty', False, False, 0, 0)),
 (('particularly', 'RB', 7),
 ('ICH*-2', 'were', 'particularly', 'dusty', 'where', False, False, 0, 0)),
 (('dusty', 'JJ', 8),
 ('were', 'particularly', 'dusty', 'where', 'the', False, False, 0, 0)),

```

(('where', 'WRB', 9),
 ('particularly',
  'dusty',
  'where',
  'the',
  'crocidolite',
  False,
  False,
  0,
  0)),
(('the', 'DT', 10),
 ('dusty', 'where', 'the', 'crocidolite', 'was', False, False, 0, 0)),
(('crocidolite', 'NN', 11),
 ('where', 'the', 'crocidolite', 'was', 'used', False, False, 0, 0)),
(('was', 'VBD', 12),
 ('the', 'crocidolite', 'was', 'used', '*-8', False, False, 0, 0)),
(('used', 'VBN', 13),
 ('crocidolite', 'was', 'used', '*-8', '*T*-1', False, False, 0, 0)),
((*-8', '-NONE-', 14),
 ('was', 'used', '*-8', '*T*-1', '.', False, False, 0, 0)),
(*T*-1', '-NONE-', 15),
 ('used', '*-8', '*T*-1', '.', 0, False, False, 0, 0)),
(('.', '.', 16), (*-8', '*T*-1', '.', 0, 0, False, False, 0, 1)),
(('Workers', 'NNS', 1),
 (0, 0, 'Workers', 'dumped', 'large', False, True, 1, 0)),
(('dumped', 'VBD', 2),
 (0, 'Workers', 'dumped', 'large', 'burlap', False, False, 0, 0)),
(('large', 'JJ', 3),
 ('Workers', 'dumped', 'large', 'burlap', 'sacks', False, False, 0, 0)),
(('burlap', 'NN', 4),
 ('dumped', 'large', 'burlap', 'sacks', 'of', False, False, 0, 0)),
(('sacks', 'NNS', 5),
 ('large', 'burlap', 'sacks', 'of', 'the', False, False, 0, 0)),
(('of', 'IN', 6),
 ('burlap', 'sacks', 'of', 'the', 'imported', False, False, 0, 0)),
(('the', 'DT', 7),
 ('sacks', 'of', 'the', 'imported', 'material', False, False, 0, 0)),
(('imported', 'VBN', 8),
 ('of', 'the', 'imported', 'material', 'into', False, False, 0, 0)),
(('material', 'NN', 9),
 ('the', 'imported', 'material', 'into', 'a', False, False, 0, 0)),
(('into', 'IN', 10),
 ('imported', 'material', 'into', 'a', 'huge', False, False, 0, 0)),
(('a', 'DT', 11),
 ('material', 'into', 'a', 'huge', 'bin', False, False, 0, 0)),
(('huge', 'JJ', 12), ('into', 'a', 'huge', 'bin', ',', False, False, 0, 0)),
(('bin', 'NN', 13), ('a', 'huge', 'bin', ',', 'poured', False, False, 0, 0)),

```

((' ', ' ', ' ', ' ', 14), ('huge', 'bin', ' ', ' ', 'poured', 'in', False, False, 0, 0)),
 (('poured', 'VBD', 15),
 ('bin', ' ', ' ', 'poured', 'in', 'cotton', False, False, 0, 0)),
 (('in', 'RP', 16),
 (' ', ' ', 'poured', 'in', 'cotton', 'and', False, False, 0, 0)),
 (('cotton', 'NN', 17),
 ('poured', 'in', 'cotton', 'and', 'acetate', False, False, 0, 0)),
 (('and', 'CC', 18),
 ('in', 'cotton', 'and', 'acetate', 'fibers', False, False, 0, 0)),
 (('acetate', 'NN', 19),
 ('cotton', 'and', 'acetate', 'fibers', 'and', False, False, 0, 0)),
 (('fibers', 'NNS', 20),
 ('and', 'acetate', 'fibers', 'and', 'mechanically', False, False, 0, 0)),
 (('and', 'CC', 21),
 ('acetate', 'fibers', 'and', 'mechanically', 'mixed', False, False, 0, 0)),
 (('mechanically', 'RB', 22),
 ('fibers', 'and', 'mechanically', 'mixed', 'the', False, False, 0, 0)),
 (('mixed', 'VBD', 23),
 ('and', 'mechanically', 'mixed', 'the', 'dry', False, False, 0, 0)),
 (('the', 'DT', 24),
 ('mechanically', 'mixed', 'the', 'dry', 'fibers', False, False, 0, 0)),
 (('dry', 'JJ', 25),
 ('mixed', 'the', 'dry', 'fibers', 'in', False, False, 0, 0)),
 (('fibers', 'NNS', 26),
 ('the', 'dry', 'fibers', 'in', 'a', False, False, 0, 0)),
 (('in', 'IN', 27),
 ('dry', 'fibers', 'in', 'a', 'process', False, False, 0, 0)),
 (('a', 'DT', 28),
 ('fibers', 'in', 'a', 'process', 'used', False, False, 0, 0)),
 (('process', 'NN', 29),
 ('in', 'a', 'process', 'used', '*', False, False, 0, 0)),
 (('used', 'VBN', 30), ('a', 'process', 'used', '*', '*', False, False, 0, 0)),
 (('*', '-NONE-', 31),
 ('process', 'used', '*', '*', 'to', False, False, 0, 0)),
 (('*', '-NONE-', 32), ('used', '*', '*', 'to', 'make', False, False, 0, 0)),
 (('to', 'TO', 33), ('*', '*', 'to', 'make', 'filters', False, False, 0, 0)),
 (('make', 'VB', 34), ('*', 'to', 'make', 'filters', '.', False, False, 0, 0)),
 (('filters', 'NNS', 35),
 ('to', 'make', 'filters', '.', 0, False, False, 0, 0)),
 (('.', '.', 36), ('make', 'filters', '.', 0, 0, False, False, 0, 1)),
 (('Workers', 'NNS', 1),
 (0, 0, 'Workers', 'described', '`', False, True, 1, 0)),
 (('described', 'VBD', 2),
 (0, 'Workers', 'described', '`', 'clouds', False, False, 0, 0)),
 (('`', '`', 3),
 ('Workers', 'described', '`', 'clouds', 'of', False, False, 0, 0)),
 (('clouds', 'NNS', 4),

('described', '``', 'clouds', 'of', 'blue', False, False, 0, 0)),
 (('of', 'IN', 5), ('``', 'clouds', 'of', 'blue', 'dust', False, False, 0, 0)),
 (('blue', 'JJ', 6),
 ('clouds', 'of', 'blue', 'dust', '""', False, False, 0, 0)),
 (('dust', 'NN', 7), ('of', 'blue', 'dust', '""', 'that', False, False, 0, 0)),
 (('""', '""', 8),
 ('blue', 'dust', '""', 'that', '*T*-1', False, False, 0, 0)),
 (('that', 'WDT', 9),
 ('dust', '""', 'that', '*T*-1', 'hung', False, False, 0, 0)),
 (('*T*-1', '-NONE-', 10),
 ('""', 'that', '*T*-1', 'hung', 'over', False, False, 0, 0)),
 (('hung', 'VBD', 11),
 ('that', '*T*-1', 'hung', 'over', 'parts', False, False, 0, 0)),
 (('over', 'IN', 12),
 ('*T*-1', 'hung', 'over', 'parts', 'of', False, False, 0, 0)),
 (('parts', 'NNS', 13),
 ('hung', 'over', 'parts', 'of', 'the', False, False, 0, 0)),
 (('of', 'IN', 14),
 ('over', 'parts', 'of', 'the', 'factory', False, False, 0, 0)),
 (('the', 'DT', 15),
 ('parts', 'of', 'the', 'factory', ',', False, False, 0, 0)),
 (('factory', 'NN', 16),
 ('of', 'the', 'factory', ',', 'even', False, False, 0, 0)),
 ((',', ',', 17),
 ('the', 'factory', ',', 'even', 'though', False, False, 0, 0)),
 (('even', 'RB', 18),
 ('factory', ',', 'even', 'though', 'exhaust', False, False, 0, 0)),
 (('though', 'IN', 19),
 (',', 'even', 'though', 'exhaust', 'fans', False, False, 0, 0)),
 (('exhaust', 'NN', 20),
 ('even', 'though', 'exhaust', 'fans', 'ventilated', False, False, 0, 0)),
 (('fans', 'NNS', 21),
 ('though', 'exhaust', 'fans', 'ventilated', 'the', False, False, 0, 0)),
 (('ventilated', 'VBD', 22),
 ('exhaust', 'fans', 'ventilated', 'the', 'area', False, False, 0, 0)),
 (('the', 'DT', 23),
 ('fans', 'ventilated', 'the', 'area', '.', False, False, 0, 0)),
 (('area', 'NN', 24),
 ('ventilated', 'the', 'area', '.', 0, False, False, 0, 0)),
 (('.', '.', 25), ('the', 'area', '.', 0, 0, False, False, 0, 1)),
 (('``', '``', 1), (0, 0, '``', 'There', "'s", False, False, 1, 0)),
 (('There', 'EX', 2), (0, '``', 'There', "'s", 'no', False, True, 0, 0)),
 (('s", 'VBZ', 3),
 ('``', 'There', "'s", 'no', 'question', False, False, 0, 0)),
 (('no', 'DT', 4),
 ('There', "'s", 'no', 'question', 'that', False, False, 0, 0)),
 (('question', 'NN', 5),

```

('s', 'no', 'question', 'that', 'some', False, False, 0, 0)),
(('that', 'IN', 6),
 ('no', 'question', 'that', 'some', 'of', False, False, 0, 0)),
(('some', 'DT', 7),
 ('question', 'that', 'some', 'of', 'those', False, False, 0, 0)),
(('of', 'IN', 8),
 ('that', 'some', 'of', 'those', 'workers', False, False, 0, 0)),
(('those', 'DT', 9),
 ('some', 'of', 'those', 'workers', 'and', False, False, 0, 0)),
(('workers', 'NNS', 10),
 ('of', 'those', 'workers', 'and', 'managers', False, False, 0, 0)),
(('and', 'CC', 11),
 ('those', 'workers', 'and', 'managers', 'contracted', False, False, 0, 0)),
(('managers', 'NNS', 12),
 ('workers',
  'and',
  'managers',
  'contracted',
  'asbestos-related',
  False,
  False,
  0,
  0)),
(('contracted', 'VBD', 13),
 ('and',
  'managers',
  'contracted',
  'asbestos-related',
  'diseases',
  False,
  False,
  0,
  0)),
(('asbestos-related', 'JJ', 14),
 ('managers',
  'contracted',
  'asbestos-related',
  'diseases',
  ',',
  False,
  False,
  0,
  0)),
(('diseases', 'NNS', 15),
 ('contracted',
  'asbestos-related',
  'diseases',

```

```

',',
''''',
False,
False,
0,
0)),
(('',', ',', 16),
 ('asbestos-related', 'diseases', ',,', ''', 'said', False, False, 0, 0)),
('''', ''', 17),
 ('diseases', ',,', ''', 'said', '*T*-1', False, False, 0, 0)),
('said', 'VBD', 18),
 ('',', ''', 'said', '*T*-1', 'Darrell', False, False, 0, 0)),
('*T*-1', '-NONE-', 19),
 ('''', 'said', '*T*-1', 'Darrell', 'Phillips', False, False, 0, 0)),
('Darrell', 'NNP', 20),
 ('said', '*T*-1', 'Darrell', 'Phillips', ',,', False, True, 0, 0)),
('Phillips', 'NNP', 21),
 (*T*-1', 'Darrell', 'Phillips', ',,', 'vice', False, True, 0, 0)),
(',',', ',', 22),
 ('Darrell', 'Phillips', ',,', 'vice', 'president', False, False, 0, 0)),
('vice', 'NN', 23),
 ('Phillips', ',,', 'vice', 'president', 'of', False, False, 0, 0)),
('president', 'NN', 24),
 ('',', 'vice', 'president', 'of', 'human', False, False, 0, 0)),
('of', 'IN', 25),
 ('vice', 'president', 'of', 'human', 'resources', False, False, 0, 0)),
('human', 'JJ', 26),
 ('president', 'of', 'human', 'resources', 'for', False, False, 0, 0)),
('resources', 'NNS', 27),
 ('of', 'human', 'resources', 'for', 'Hollingsworth', False, False, 0, 0)),
('for', 'IN', 28),
 ('human', 'resources', 'for', 'Hollingsworth', '&', False, False, 0, 0)),
('Hollingsworth', 'NNP', 29),
 ('resources', 'for', 'Hollingsworth', '&', 'Vose', False, True, 0, 0)),
('&', 'CC', 30),
 ('for', 'Hollingsworth', '&', 'Vose', '.', False, False, 0, 0)),
('Vose', 'NNP', 31),
 ('Hollingsworth', '&', 'Vose', '.', 0, False, True, 0, 0)),
('.',', '.', 32), ('&', 'Vose', '.', 0, 0, False, False, 0, 1)),
('``', '``', 1), (0, 0, '``', 'But', 'you', False, False, 1, 0)),
('But', 'CC', 2), (0, '``', 'But', 'you', 'have', False, True, 0, 0)),
('you', 'PRP', 3), ('``', 'But', 'you', 'have', '*-1', False, False, 0, 0)),
('have', 'VBP', 4), ('But', 'you', 'have', '*-1', 'to', False, False, 0, 0)),
('*-1', '-NONE-', 5),
 ('you', 'have', '*-1', 'to', 'recognize', False, False, 0, 0)),
('to', 'TO', 6),
 ('have', '*-1', 'to', 'recognize', 'that', False, False, 0, 0)),

```

```

(('recognize', 'VB', 7),
 ('*-1', 'to', 'recognize', 'that', 'these', False, False, 0, 0)),
(('that', 'IN', 8),
 ('to', 'recognize', 'that', 'these', 'events', False, False, 0, 0)),
(('these', 'DT', 9),
 ('recognize', 'that', 'these', 'events', 'took', False, False, 0, 0)),
(('events', 'NNS', 10),
 ('that', 'these', 'events', 'took', 'place', False, False, 0, 0)),
(('took', 'VBD', 11),
 ('these', 'events', 'took', 'place', '35', False, False, 0, 0)),
(('place', 'NN', 12),
 ('events', 'took', 'place', '35', 'years', False, False, 0, 0)),
(('35', 'CD', 13),
 ('took', 'place', '35', 'years', 'ago', True, False, 0, 0)),
(('years', 'NNS', 14),
 ('place', '35', 'years', 'ago', '.', False, False, 0, 0)),
(('ago', 'IN', 15), ('35', 'years', 'ago', '.', 0, False, False, 0, 0)),
(('.', '.', 16), ('years', 'ago', '.', 0, 0, False, False, 0, 1)),
(('It', 'PRP', 1), (0, 0, 'It', 'has', 'no', False, True, 1, 0)),
(('has', 'VBZ', 2), (0, 'It', 'has', 'no', 'bearing', False, False, 0, 0)),
(('no', 'DT', 3), ('It', 'has', 'no', 'bearing', 'on', False, False, 0, 0)),
(('bearing', 'NN', 4),
 ('has', 'no', 'bearing', 'on', 'our', False, False, 0, 0)),
(('on', 'IN', 5), ('no', 'bearing', 'on', 'our', 'work', False, False, 0, 0)),
(('our', 'PRP$', 6),
 ('bearing', 'on', 'our', 'work', 'force', False, False, 0, 0)),
(('work', 'NN', 7),
 ('on', 'our', 'work', 'force', 'today', False, False, 0, 0)),
(('force', 'NN', 8),
 ('our', 'work', 'force', 'today', '.', False, False, 0, 0)),
(('today', 'NN', 9), ('work', 'force', 'today', '.', 0, False, False, 0, 0)),
(('.', '.', 10), ('force', 'today', '.', 0, 0, False, False, 0, 1)),
(('Yields', 'NNS', 1),
 (0, 0, 'Yields', 'on', 'money-market', False, True, 1, 0)),
(('on', 'IN', 2),
 (0, 'Yields', 'on', 'money-market', 'mutual', False, False, 0, 0)),
(('money-market', 'JJ', 3),
 ('Yields', 'on', 'money-market', 'mutual', 'funds', False, False, 0, 0)),
(('mutual', 'JJ', 4),
 ('on', 'money-market', 'mutual', 'funds', 'continued', False, False, 0, 0)),
(('funds', 'NNS', 5),
 ('money-market', 'mutual', 'funds', 'continued', '*-1', False, False, 0, 0)),
(('continued', 'VBD', 6),
 ('mutual', 'funds', 'continued', '*-1', 'to', False, False, 0, 0)),
(('*-1', '-NONE-', 7),
 ('funds', 'continued', '*-1', 'to', 'slide', False, False, 0, 0)),
(('to', 'TO', 8),

```

```

('continued', '*-1', 'to', 'slide', ',', False, False, 0, 0)),
(('slide', 'VB', 9), (*-1, 'to', 'slide', ',', 'amid', False, False, 0, 0)),
((',', ',', 10), ('to', 'slide', ',', 'amid', 'signs', False, False, 0, 0)),
(('amid', 'IN', 11),
 ('slide', ',', 'amid', 'signs', 'that', False, False, 0, 0)),
(('signs', 'NNS', 12),
 (',', 'amid', 'signs', 'that', 'portfolio', False, False, 0, 0)),
(('that', 'IN', 13),
 ('amid', 'signs', 'that', 'portfolio', 'managers', False, False, 0, 0)),
(('portfolio', 'NN', 14),
 ('signs', 'that', 'portfolio', 'managers', 'expect', False, False, 0, 0)),
(('managers', 'NNS', 15),
 ('that', 'portfolio', 'managers', 'expect', 'further', False, False, 0, 0)),
(('expect', 'VBP', 16),
 ('portfolio',
  'managers',
  'expect',
  'further',
  'declines',
  False,
  False,
  0,
  0)),
(('further', 'JJ', 17),
 ('managers', 'expect', 'further', 'declines', 'in', False, False, 0, 0)),
(('declines', 'NNS', 18),
 ('expect', 'further', 'declines', 'in', 'interest', False, False, 0, 0)),
(('in', 'IN', 19),
 ('further', 'declines', 'in', 'interest', 'rates', False, False, 0, 0)),
(('interest', 'NN', 20),
 ('declines', 'in', 'interest', 'rates', '.', False, False, 0, 0)),
(('rates', 'NNS', 21),
 ('in', 'interest', 'rates', '.', 0, False, False, 0, 0)),
(('.', '.', 22), ('interest', 'rates', '.', 0, 0, False, False, 0, 1)),
(('The', 'DT', 1), (0, 0, 'The', 'average', 'seven-day', False, True, 1, 0)),
(('average', 'JJ', 2),
 (0, 'The', 'average', 'seven-day', 'compound', False, False, 0, 0)),
(('seven-day', 'JJ', 3),
 ('The', 'average', 'seven-day', 'compound', 'yield', False, False, 0, 0)),
(('compound', 'NN', 4),
 ('average', 'seven-day', 'compound', 'yield', 'of', False, False, 0, 0)),
(('yield', 'NN', 5),
 ('seven-day', 'compound', 'yield', 'of', 'the', False, False, 0, 0)),
(('of', 'IN', 6),
 ('compound', 'yield', 'of', 'the', '400', False, False, 0, 0)),
(('the', 'DT', 7),
 ('yield', 'of', 'the', '400', 'taxable', False, False, 0, 0)),

```

(('400', 'CD', 8),
 ('of', 'the', '400', 'taxable', 'funds', True, False, 0, 0)),
 (('taxable', 'JJ', 9),
 ('the', '400', 'taxable', 'funds', 'tracked', False, False, 0, 0)),
 (('funds', 'NNS', 10),
 ('400', 'taxable', 'funds', 'tracked', '*', False, False, 0, 0)),
 (('tracked', 'VBN', 11),
 ('taxable', 'funds', 'tracked', '*', 'by', False, False, 0, 0)),
 (('*', '-NONE-', 12),
 ('funds', 'tracked', '*', 'by', 'IBC', False, False, 0, 0)),
 (('by', 'IN', 13), ('tracked', '*', 'by', 'IBC', "'s", False, False, 0, 0)),
 (('IBC', 'NNP', 14), ('*', 'by', 'IBC', "'s", 'Money', False, True, 0, 0)),
 (('s', 'POS', 15), ('by', 'IBC', "'s", 'Money', 'Fund', False, False, 0, 0)),
 (('Money', 'NNP', 16),
 ('IBC', "'s", 'Money', 'Fund', 'Report', False, True, 0, 0)),
 (('Fund', 'NNP', 17),
 ('s', 'Money', 'Fund', 'Report', 'eased', False, True, 0, 0)),
 (('Report', 'NNP', 18),
 ('Money', 'Fund', 'Report', 'eased', 'a', False, True, 0, 0)),
 (('eased', 'VBD', 19),
 ('Fund', 'Report', 'eased', 'a', 'fraction', False, False, 0, 0)),
 (('a', 'DT', 20),
 ('Report', 'eased', 'a', 'fraction', 'of', False, False, 0, 0)),
 (('fraction', 'NN', 21),
 ('eased', 'a', 'fraction', 'of', 'a', False, False, 0, 0)),
 (('of', 'IN', 22),
 ('a', 'fraction', 'of', 'a', 'percentage', False, False, 0, 0)),
 (('a', 'DT', 23),
 ('fraction', 'of', 'a', 'percentage', 'point', False, False, 0, 0)),
 (('percentage', 'NN', 24),
 ('of', 'a', 'percentage', 'point', 'to', False, False, 0, 0)),
 (('point', 'NN', 25),
 ('a', 'percentage', 'point', 'to', '8.45', False, False, 0, 0)),
 (('to', 'TO', 26),
 ('percentage', 'point', 'to', '8.45', '%', False, False, 0, 0)),
 (('8.45', 'CD', 27),
 ('point', 'to', '8.45', '%', 'from', False, False, 0, 0)),
 (('%', 'NN', 28), ('to', '8.45', '%', 'from', '8.47', False, False, 0, 0)),
 (('from', 'IN', 29), ('8.45', '%', 'from', '8.47', '%', False, False, 0, 0)),
 (('8.47', 'CD', 30), ('%', 'from', '8.47', '%', 'for', False, False, 0, 0)),
 (('%', 'NN', 31), ('from', '8.47', '%', 'for', 'the', False, False, 0, 0)),
 (('for', 'IN', 32), ('8.47', '%', 'for', 'the', 'week', False, False, 0, 0)),
 (('the', 'DT', 33), ('%', 'for', 'the', 'week', 'ended', False, False, 0, 0)),
 (('week', 'NN', 34),
 ('for', 'the', 'week', 'ended', 'Tuesday', False, False, 0, 0)),
 (('ended', 'VBD', 35),
 ('the', 'week', 'ended', 'Tuesday', '.', False, False, 0, 0)),

(('Tuesday', 'NNP', 36),
 ('week', 'ended', 'Tuesday', '.', 0, False, True, 0, 0)),
 (('.', '.', 37), ('ended', 'Tuesday', '.', 0, 0, False, False, 0, 1)),
 (('Compound', 'NN', 1),
 (0, 0, 'Compound', 'yields', 'assume', False, True, 1, 0)),
 (('yields', 'NNS', 2),
 (0, 'Compound', 'yields', 'assume', 'reinvestment', False, False, 0, 0)),
 (('assume', 'VBP', 3),
 ('Compound', 'yields', 'assume', 'reinvestment', 'of', False, False, 0, 0)),
 (('reinvestment', 'NN', 4),
 ('yields', 'assume', 'reinvestment', 'of', 'dividends', False, False, 0, 0)),
 (('of', 'IN', 5),
 ('assume', 'reinvestment', 'of', 'dividends', 'and', False, False, 0, 0)),
 (('dividends', 'NNS', 6),
 ('reinvestment', 'of', 'dividends', 'and', 'that', False, False, 0, 0)),
 (('and', 'CC', 7),
 ('of', 'dividends', 'and', 'that', 'the', False, False, 0, 0)),
 (('that', 'IN', 8),
 ('dividends', 'and', 'that', 'the', 'current', False, False, 0, 0)),
 (('the', 'DT', 9),
 ('and', 'that', 'the', 'current', 'yield', False, False, 0, 0)),
 (('current', 'JJ', 10),
 ('that', 'the', 'current', 'yield', 'continues', False, False, 0, 0)),
 (('yield', 'NN', 11),
 ('the', 'current', 'yield', 'continues', 'for', False, False, 0, 0)),
 (('continues', 'VBZ', 12),
 ('current', 'yield', 'continues', 'for', 'a', False, False, 0, 0)),
 (('for', 'IN', 13),
 ('yield', 'continues', 'for', 'a', 'year', False, False, 0, 0)),
 (('a', 'DT', 14), ('continues', 'for', 'a', 'year', '.', False, False, 0, 0)),
 (('year', 'NN', 15), ('for', 'a', 'year', '.', 0, False, False, 0, 0)),
 (('.', '.', 16), ('a', 'year', '.', 0, 0, False, False, 0, 1)),
 (('Average', 'JJ', 1),
 (0, 0, 'Average', 'maturity', 'of', False, True, 1, 0)),
 (('maturity', 'NN', 2),
 (0, 'Average', 'maturity', 'of', 'the', False, False, 0, 0)),
 (('of', 'IN', 3),
 ('Average', 'maturity', 'of', 'the', 'funds', False, False, 0, 0)),
 (('the', 'DT', 4),
 ('maturity', 'of', 'the', 'funds', '"', False, False, 0, 0)),
 (('funds', 'NNS', 5),
 ('of', 'the', 'funds', '"', 'investments', False, False, 0, 0)),
 (('"', 'POS', 6),
 ('the', 'funds', '"', 'investments', 'lengthened', False, False, 0, 0)),
 (('investments', 'NNS', 7),
 ('funds', '"', 'investments', 'lengthened', 'by', False, False, 0, 0)),
 (('lengthened', 'VBD', 8),

("", 'investments', 'lengthened', 'by', 'a', False, False, 0, 0)),
 (('by', 'IN', 9),
 ('investments', 'lengthened', 'by', 'a', 'day', False, False, 0, 0)),
 (('a', 'DT', 10), ('lengthened', 'by', 'a', 'day', 'to', False, False, 0, 0)),
 (('day', 'NN', 11), ('by', 'a', 'day', 'to', '41', False, False, 0, 0)),
 (('to', 'TO', 12), ('a', 'day', 'to', '41', 'days', False, False, 0, 0)),
 (('41', 'CD', 13), ('day', 'to', '41', 'days', ',', True, False, 0, 0)),
 (('days', 'NNS', 14), ('to', '41', 'days', ',', 'the', False, False, 0, 0)),
 (('the', 'DT', 16), ('days', ',', 'the', 'longest', 'since', False, False, 0, 0)),
 (('longest', 'JJ', 17), ('the', 'longest', 'since', 'early', False, False, 0, 0)),
 (('since', 'IN', 18), ('the', 'longest', 'since', 'early', 'August', False, False, 0, 0)),
 (('early', 'JJ', 19), ('longest', 'since', 'early', 'August', ',', False, False, 0, 0)),
 (('August', 'NNP', 20), ('since', 'early', 'August', ',', 'according', False, True, 0, 0)),
 (('according', 'VBG', 22), ('early', 'August', ',', 'according', 'to', False, False, 0, 0)),
 (('to', 'TO', 23), ('August', ',', 'according', 'to', 'Donoghue', False, False, 0, 0)),
 (('Donoghue', 'NNP', 24), ('according', 'to', 'Donoghue', '"s"', False, True, 0, 0)),
 (('s"', 'POS', 25), ('to', 'Donoghue', '"s"', '.', 0, False, False, 0, 0)),
 (('Longer', 'JJR', 1), (0, 0, 'Longer', 'maturities', 'are', False, True, 1, 0)),
 (('maturities', 'NNS', 2), (0, 'Longer', 'maturities', 'are', 'thought', False, False, 0, 0)),
 (('are', 'VBP', 3), ('Longer', 'maturities', 'are', 'thought', '*-1', False, False, 0, 0)),
 (('thought', 'VBN', 4), ('maturities', 'are', 'thought', '*-1', 'to', False, False, 0, 0)),
 (('*-1', '-NONE-', 5), ('are', 'thought', '*-1', 'to', 'indicate', False, False, 0, 0)),
 (('to', 'TO', 6), ('thought', '*-1', 'to', 'indicate', 'declining', False, False, 0, 0)),
 (('indicate', 'VB', 7), ('*-1', 'to', 'indicate', 'declining', 'interest', False, False, 0, 0)),
 (('declining', 'VBG', 8), ('to', 'indicate', 'declining', 'interest', 'rates', False, False, 0, 0)),
 (('interest', 'NN', 9), ('indicate',


```

'declining',
'interest',
'rates',
'because',
False,
False,
0,
0)),
(('rates', 'NNS', 10),
 ('declining', 'interest', 'rates', 'because', 'they', False, False, 0, 0)),
(('because', 'IN', 11),
 ('interest', 'rates', 'because', 'they', 'permit', False, False, 0, 0)),
(('they', 'PRP', 12),
 ('rates', 'because', 'they', 'permit', 'portfolio', False, False, 0, 0)),
(('permit', 'VBP', 13),
 ('because', 'they', 'permit', 'portfolio', 'managers', False, False, 0, 0)),
(('portfolio', 'NN', 14),
 ('they', 'permit', 'portfolio', 'managers', 'to', False, False, 0, 0)),
(('managers', 'NNS', 15),
 ('permit', 'portfolio', 'managers', 'to', 'retain', False, False, 0, 0)),
(('to', 'TO', 16),
 ('portfolio', 'managers', 'to', 'retain', 'relatively', False, False, 0, 0)),
(('retain', 'VB', 17),
 ('managers', 'to', 'retain', 'relatively', 'higher', False, False, 0, 0)),
(('relatively', 'RB', 18),
 ('to', 'retain', 'relatively', 'higher', 'rates', False, False, 0, 0)),
(('higher', 'JJR', 19),
 ('retain', 'relatively', 'higher', 'rates', 'for', False, False, 0, 0)),
(('rates', 'NNS', 20),
 ('relatively', 'higher', 'rates', 'for', 'a', False, False, 0, 0)),
(('for', 'IN', 21),
 ('higher', 'rates', 'for', 'a', 'longer', False, False, 0, 0)),
(('a', 'DT', 22),
 ('rates', 'for', 'a', 'longer', 'period', False, False, 0, 0)),
(('longer', 'JJR', 23),
 ('for', 'a', 'longer', 'period', '.', False, False, 0, 0)),
(('period', 'NN', 24), ('a', 'longer', 'period', '.', 0, False, False, 0, 0)),
((('.', '.', 25), ('longer', 'period', '.', 0, 0, False, False, 0, 1)),
(('Shorter', 'JJR', 1),
 (0, 0, 'Shorter', 'maturities', 'are', False, True, 1, 0)),
(('maturities', 'NNS', 2),
 (0, 'Shorter', 'maturities', 'are', 'considered', False, False, 0, 0)),
(('are', 'VBP', 3),
 ('Shorter', 'maturities', 'are', 'considered', '*-9', False, False, 0, 0)),
(('considered', 'VBN', 4),
 ('maturities', 'are', 'considered', '*-9', 'a', False, False, 0, 0)),
(('*-9', '-NONE-', 5),

```

('are', 'considered', '*-9', 'a', 'sign', False, False, 0, 0)),
 (('a', 'DT', 6),
 ('considered', '*-9', 'a', 'sign', 'of', False, False, 0, 0)),
 (('sign', 'NN', 7), ('*-9', 'a', 'sign', 'of', 'rising', False, False, 0, 0)),
 (('of', 'IN', 8), ('a', 'sign', 'of', 'rising', 'rates', False, False, 0, 0)),
 (('rising', 'VBG', 9),
 ('sign', 'of', 'rising', 'rates', 'because', False, False, 0, 0)),
 (('rates', 'NNS', 10),
 ('of', 'rising', 'rates', 'because', 'portfolio', False, False, 0, 0)),
 (('because', 'IN', 11),
 ('rising', 'rates', 'because', 'portfolio', 'managers', False, False, 0, 0)),
 (('portfolio', 'NN', 12),
 ('rates', 'because', 'portfolio', 'managers', 'can', False, False, 0, 0)),
 (('managers', 'NNS', 13),
 ('because', 'portfolio', 'managers', 'can', 'capture', False, False, 0, 0)),
 (('can', 'MD', 14),
 ('portfolio', 'managers', 'can', 'capture', 'higher', False, False, 0, 0)),
 (('capture', 'VB', 15),
 ('managers', 'can', 'capture', 'higher', 'rates', False, False, 0, 0)),
 (('higher', 'JJR', 16),
 ('can', 'capture', 'higher', 'rates', 'sooner', False, False, 0, 0)),
 (('rates', 'NNS', 17),
 ('capture', 'higher', 'rates', 'sooner', '.', False, False, 0, 0)),
 (('sooner', 'RB', 18),
 ('higher', 'rates', 'sooner', '.', 0, False, False, 0, 0)),
 (('.', '.', 19), ('rates', 'sooner', '.', 0, 0, False, False, 0, 1)),
 (('The', 'DT', 1), (0, 0, 'The', 'average', 'maturity', False, True, 1, 0)),
 (('average', 'JJ', 2),
 (0, 'The', 'average', 'maturity', 'for', False, False, 0, 0)),
 (('maturity', 'NN', 3),
 ('The', 'average', 'maturity', 'for', 'funds', False, False, 0, 0)),
 (('for', 'IN', 4),
 ('average', 'maturity', 'for', 'funds', 'open', False, False, 0, 0)),
 (('funds', 'NNS', 5),
 ('maturity', 'for', 'funds', 'open', 'only', False, False, 0, 0)),
 (('open', 'JJ', 6),
 ('for', 'funds', 'open', 'only', 'to', False, False, 0, 0)),
 (('only', 'RB', 7),
 ('funds', 'open', 'only', 'to', 'institutions', False, False, 0, 0)),
 (('to', 'TO', 8),
 ('open', 'only', 'to', 'institutions', ',', False, False, 0, 0)),
 (('institutions', 'NNS', 9),
 ('only', 'to', 'institutions', ',', 'considered', False, False, 0, 0)),
 ((' ', ' ', 10),
 ('to', 'institutions', ',', 'considered', 'by', False, False, 0, 0)),
 (('considered', 'VBN', 11),
 ('institutions', ',', 'considered', 'by', 'some', False, False, 0, 0)),

```
(('by', 'IN', 12),
 ('', 'considered', 'by', 'some', '*', False, False, 0, 0)),
 (('some', 'DT', 13),
 ('considered', 'by', 'some', '*', 'to', False, False, 0, 0)),
 (('*', '-NONE-', 14), ('by', 'some', '*', 'to', 'be', False, False, 0, 0)),
 (('to', 'TO', 15), ('some', '*', 'to', 'be', 'a', False, False, 0, 0)),
 ...]
```

[]: