Let's use FP_to_PB.py to generate PB instances.

```
In [92]: from FB_to_PB import *
```

```
In [110]: f2p = fb_to_pb_map('FB_to_PB.csv')
          fb_dir = 'tmp.business_Freebase'
          for fbr in ['acquisition']:
              ctr(fbr)
              pb = pb_instances(fb_dir,fbr, f2p)
              print pb[:10]
```

```
_____acquisition_____
[{'ARG0': 'Art.com', 'ARG1': 'AllPosters.com'}, {'ARG0': 'Hewlett-Packard', 'ARG1':
'VoodooPC'}, {'ARG0': 'InBev', 'ARG1': 'Anheuser-Busch'}, {'ARG0': 'Sugar, Inc.', 'ARG1':
'Coutorture'}, {'ARG0': 'Scripps Network', 'ARG1': 'Incando Corporation'}, {'ARG0': 'Vast',
'ARG1': 'Edgeio'}, {'ARG0': 'Comcast', 'ARG1': 'TechTV'}, {'ARG0': 'Microsoft', 'ARG1':
'Jellyfish'}, {'ARG0': 'MSNBC Interactive', 'ARG1': 'Newsvine'}, {'ARG0': 'Pearson PLC',
'ARG1': 'Longman Publishing Group'}]
```

Treebank 2 lets us build a list of sentences so we can look for the instances from FB.

```
In [94]: from nltk.corpus import treebank as tb
```

Here's one way to get out information about the words.

```
In [97]: words = [(fileid,sentnum,wordnum,word)
                  for fileid in tb.fileids()[:1]
                  for sentnum,sent in enumerate(tb.tagged_sents(fileid))
                  for wordnum,(word,pos) in enumerate(sent)
                  if pos in ['NNP','NNPS']]
         words[:3]
```

```
Out[97]: [('wsj_0001.mrg', 0, 0, 'Pierre'),
          ('wsj_0001.mrg', 0, 1, 'Vinken'),
          ('wsj_0001.mrg', 0, 15, 'Nov.')]
```

```
In [65]: from sets import Set
         a1 = Set(['a','b'])
         a2 = a1 | Set(['c'])
         print a1,a2
         a1.issubset(a2)
```

```
Set(['a', 'b']) Set(['a', 'c', 'b'])
```

```
Out[65]: True
```

We need a way to find all the sentences that contain all the words from each new instance. This is a significant simplification over how we might do this. We'll start by building a helpful list of sentences from TB.

```
In [141]: sents = [
                  {'fileid':fileid,
                   'sentnum':sentnum,
                   'nnp':Set([word for word,pos in sent if pos in ['NNP','NNPS']]),
                   'nnp_loc':{word:wordnum for wordnum,(word,pos) in enumerate(sent) if pos in ['NNP'
                   'word_loc':{word:wordnum for wordnum,(word,pos) in enumerate(sent)}
                   'sent':
                  }
                  for fileid in tb.fileids()
                  for sentnum,sent in enumerate(tb.tagged_sents(fileid))
                  ]
```

An example:

```
In [146]: print len([sent for fileid in tb.fileids() for sent in tb.tagged_sents(fileid)])
          print len([word for fileid in tb.fileids() for sent in tb.tagged_sents(fileid) for word in s

          49208
          1253013
```

```
In [147]: len(tb.fileids())
```

```
Out[147]: 2312
```

```
In [151]: print len(sents)
          sents[:1]

          49208
```

```
Out[151]: [{'fileid': 'wsj_0001.mrg',
            'nnp': Set(['Nov.', 'Pierre', 'Vinken']),
            'nnp_loc': {'Nov.': 15, 'Pierre': 0, 'Vinken': 1},
            'sentnum': 0,
            'word_loc': {',': 6,
             '.': 17,
             '29': 16,
             '61': 3,
             'Nov.': 15,
             'Pierre': 0,
             'Vinken': 1,
             'a': 12,
             'as': 11,
             'board': 10,
             'director': 14,
             'join': 8,
             'nonexecutive': 13,
             'old': 5,
             'the': 9,
             'will': 7,
             'years': 4}}]
```

Get some FB instances, and look for TB sentences that include all of the tokens among all the args in the FB instance.

```
In [160]: from nltk.tokenize import word_tokenize
          for fbr in ['acquisition']:
              ctr(fbr)
              pb = pb_instances(fb_dir,fbr,f2p)
              print 'FB relation %s (%d instances)'%(fbr,len(pb))
              for pbi in pb:

                  if len([word for arg,word in pbi.iteritems() if word.strip() == '']) > 0:
                      print 'SKIPPING (blank argument): %s'% str(pbi)
                      continue

                  pbi_words = Set([token for arg,word in pbi.iteritems() for token in word_tokenize(wo
                  #print pbi_words
                  #print ''

                  tb_sents = [sent for sent in sents if pbi_words.issubset(sent['nnp'])]
                  if(len(tb_sents) > 0):
                      ctr('MATCH!!!')
                      print pbi
                      for sent in tb_sents:
```

```
                    ctrs('TB sentence')
                    print sent
                    vprt('sentence',' '.join(tb.sents(sent['fileid'])[sent['sentnum']]))
              ctr('')
```

_____acquisition_____
FB relation acquisition (2085 instances)
SKIPPING (blank argument): {'ARG0': 'Volcom', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Lanscroft Mine'}
SKIPPING (blank argument): {'ARG0': 'Thomson Corporation', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Parakey'}
SKIPPING (blank argument): {'ARG0': 'Infor Global Solutions', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'British Aerospace', 'ARG1': 'Rover Group'}
_____TB sentence_____
{'word_loc': {'and': 29, 'Rover': 9, 'U.K.': 27, 'Group': 10, 'Industriels': 36, '*-1': 17,
'France': 42, 'through': 22, 'have': 38, 'Under': 0, '.': 43, 'its': 23, 'also': 14,
'offer': 19, 'able': 16, ',': 28, 'arrangement': 3, 'to': 18, 'truck-building': 32,
'Aerospace': 6, 'be': 15, "'s": 31, 'Renault': 34, 'separate': 2, 'Vehicles': 35,
'British': 5, 'subsidiary': 33, 'a': 1, 'in': 41, 'dealers': 24, 'rights': 40, 'vehicles':
21, 'will': 37, 'PLC': 11, 'the': 26, 'similar': 39}, 'nnp': Set(['Rover', 'U.K.', 'Group',
'Renault', 'Industriels', 'Vehicles', 'British', 'France', 'PLC', 'Aerospace']), 'nnp_loc':
{'Rover': 9, 'U.K.': 27, 'Group': 10, 'Renault': 34, 'Industriels': 36, 'Vehicles': 35,
'British': 5, 'France': 42, 'PLC': 11, 'Aerospace': 6}, 'sentnum': 11, 'fileid':
'wsj_1794.mrg'}
sentence: Under a separate arrangement , British Aerospace PLC 's Rover Group PLC
subsidiary will also be able *-1 to offer the vehicles through its dealers in the U.K. ,
and Renault 's truck-building subsidiary Renault Vehicles Industriels will have similar
rights in France .
_____
SKIPPING (blank argument): {'ARG0': 'Google', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'InviteShare'}
SKIPPING (blank argument): {'ARG0': 'Hearst Corporation', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'British Broadcasting Corporation', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Grand Junction Railway', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Merseyrail', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'IAC/InterActiveCorp', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'Philip Morris USA', 'ARG1': 'Kraft Foods'}
_____TB sentence_____
{'word_loc': {'and': 19, 'says': 28, 'USA': 38, '*T*-1': 29, 'is': 13, 'Dick': 30,
'coffee': 17, "'''": 27, 'committed': 14, 'Mayer': 31, 'business': 18, "'": 9, ',': 32, '.':
40, 'to': 20, 'Philip': 3, 'Kraft': 6, '-RCB-': 12, 'division': 39, '``': 0, '-LCB-': 5,
'increased': 21, 'parent': 10, 'company': 11, 'of': 34, 'Morris': 4, 'president': 33,
'Now': 1, 'for': 23, 'Maxwell': 24, 'General': 36, 'Foods': 37, 'advertising': 22, 'House':
25, 'the': 35}, 'nnp': Set(['Mayer', 'USA', 'Maxwell', 'Dick', 'Morris', 'General',
'Foods', 'House', 'Philip', 'Kraft']), 'nnp_loc': {'Mayer': 31, 'USA': 38, 'Maxwell': 24,
'Dick': 30, 'Morris': 4, 'General': 36, 'Foods': 37, 'House': 25, 'Philip': 3, 'Kraft': 6},
'sentnum': 72, 'fileid': 'wsj_1856.mrg'}
sentence: `` Now , Philip Morris -LCB- Kraft General Foods ' parent company -RCB- is
committed to the coffee business and to increased advertising for Maxwell House , '' says
T-1 Dick Mayer , president of the General Foods USA division .
_____
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Wheal Fortune'}
SKIPPING (blank argument): {'ARG0': 'Ramb\xc3\xb8ll Group', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'Random House', 'ARG1': 'Alfred A. Knopf'}
_____TB sentence_____
{'word_loc': {'Alfred': 42, 'family': 58, '*T*-1': 19, '*-2': 45, 'is': 1, 'There': 0,
'owned': 53, 'bring': 8, 'Mr.': 5, 'another': 14, 'in': 26, "'s": 41, 'A.': 43, 'by': 55,
'combination': 27, 'also': 52, '*': 54, ',': 51, 'powerhouse': 11, 'member': 17, '0': 18,
'Robert': 33, '*T*-38': 37, 'Knopf': 44, 'executive': 31, 'run': 47, 'to': 46, 'business':
23, 'that': 4, 'who': 36, 'Newhouse': 57, 'Yorker': 50, 'with': 28, 'speculation': 3,

'publishing': 30, 'a': 29, 'like': 32, 'New': 49, 'House': 40, 'could': 7, 'Random': 39,
'Gottlieb': 34, 'side': 24, '.': 59, 'businessman': 12, 'the': 56, 'or': 13, 'left': 38},
'nnp': Set(['Knopf', 'Yorker', 'Newhouse', 'House', 'Robert', 'Random', 'Gottlieb', 'A.',
'Mr.', 'Alfred', 'New']), 'nnp_loc': {'Knopf': 44, 'Yorker': 50, 'Newhouse': 57, 'House':
40, 'Robert': 33, 'Random': 39, 'Gottlieb': 34, 'A.': 43, 'Mr.': 5, 'Alfred': 42, 'New':
49}, 'sentnum': 19, 'fileid': 'wsj_0111.mrg'}
sentence: There is also speculation that Mr. Newhouse could bring in a powerhouse
businessman or another Newhouse family member 0 *T*-1 to run the business side , in
combination with a publishing executive like Robert Gottlieb , who *T*-38 left Random House
's Alfred A. Knopf *-2 to run the New Yorker , also owned * by the Newhouse family .
_____TB sentence_____
{'word_loc': {'and': 27, 'Sonny': 28, 'Knopf': 37, 'trade': 21, 'prestigious': 34, 'books':
22, 'are': 6, 'Alfred': 35, 'Joni': 7, '.': 39, 'Mehta': 29, 'for': 23, '*': 15, ',': 30,
'A.': 36, 'to': 16, 'unit': 38, 'Not': 0, 'recruited': 10, 'most-likely-successor': 4,
'be': 17, 'publisher': 18, 'two': 12, 'included': 1, 'president': 31, 'years': 13, 'ago':
14, 'on': 2, 'Evans': 8, 'of': 32, 'Random': 24, 'list': 5, 'House': 25, 'the': 33,
'adult': 20}, 'nnp': Set(['Knopf', 'Sonny', 'Evans', 'House', 'Random', 'A.', 'Alfred',
'Joni', 'Mehta']), 'nnp_loc': {'Knopf': 37, 'Sonny': 28, 'Evans': 8, 'House': 25, 'Random':
24, 'A.': 36, 'Alfred': 35, 'Joni': 7, 'Mehta': 29}, 'sentnum': 20, 'fileid':
'wsj_0111.mrg'}
sentence: Not included on the most-likely-successor list are Joni Evans , recruited *
two years ago * to be publisher of adult trade books for Random House , and Sonny Mehta ,
president of the prestigious Alfred A. Knopf unit .
_____
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'JDS Uniphase'}
SKIPPING (blank argument): {'ARG0': 'Hill International, Inc.', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Unilever', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'Electronic Data Systems', 'ARG1': 'National Advanced Systems'}
_____TB sentence_____
{'word_loc': {'and': 13, 'is': 22, '*-1': 24, 'powerful': 30, 'Systems': 20, 'mainframes':
32, ',': 21, '.': 36, 'to': 25, 'IBM-compatible': 31, 'expected': 23, 'Corp.': 16, 'Data':
19, 'Advanced': 2, 'a': 27, "'s": 17, 'year': 35, 'unveil': 26, 'joint': 6, 'Ltd.': 12,
'line': 28, 'And': 0, 'this': 34, 'of': 29, 'National': 1, 'later': 33, 'Motors': 15,
'General': 14, 'Hitachi': 11, 'venture': 7, 'Japan': 9, 'Electronic': 18}, 'nnp':
Set(['Data', 'National', 'Motors', 'General', 'Hitachi', 'Systems', 'Ltd.', 'Japan',
'Corp.', 'Electronic', 'Advanced']), 'nnp_loc': {'Data': 19, 'National': 1, 'Motors': 15,
'General': 14, 'Hitachi': 11, 'Systems': 20, 'Ltd.': 12, 'Japan': 9, 'Corp.': 16,
'Electronic': 18, 'Advanced': 2}, 'sentnum': 55, 'fileid': 'wsj_1809.mrg'}
sentence: And National Advanced Systems , a joint venture of Japan 's Hitachi Ltd. and
General Motors Corp. 's Electronic Data Systems , is expected *-1 to unveil a line of
powerful IBM-compatible mainframes later this year .
_____
SKIPPING (blank argument): {'ARG0': 'News International', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'The Boston Patriot'}
_____MATCH!!!_____
{'ARG0': 'Daimler-Benz', 'ARG1': 'AEG'}
_____TB sentence_____
{'word_loc': {'the': 10, 'industrial': 14, '%': 3, 'AG': 8, 'country': 11, 'is': 1, '*-1':
5, ',': 9, 'owned': 4, '.': 16, 'Daimler-Benz': 7, 'AEG': 0, 'biggest': 13, "'s": 12, '80':
2, 'by': 6, 'concern': 15}, 'nnp': Set(['AG', 'Daimler-Benz', 'AEG']), 'nnp_loc': {'AG': 8,
'Daimler-Benz': 7, 'AEG': 0}, 'sentnum': 6, 'fileid': 'wsj_1324.mrg'}
sentence: AEG is 80 % owned *-1 by Daimler-Benz AG , the country 's biggest industrial
concern .
_____
SKIPPING (blank argument): {'ARG0': 'Fox Interactive Media', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'International Paper', 'ARG1': 'Champion International Paper'}
_____TB sentence_____
{'word_loc': {'and': 40, 'Louisiana-Pacific': 41, 'Federal': 7, '3\\/4': 2, 'Board': 9,
'International': 34, '3\\/8': 36, '24': 13, '7\\/8': 31, 'Champion': 33, '48': 30, 'rose':
1, ',': 32, '40': 45, '.': 47, 'to': 44, 'added': 10, 'fell': 27, 'Mead': 0, 'gained': 18,
'39': 4, 'lost': 35, '31': 38, '1\\/2': 39, 'while': 24, 'Paper': 26, 'Scott': 16, '1\\/4':
46, '1\\/8': 43, 'dropped': 42}, 'nnp': Set(['Louisiana-Pacific', 'Champion', 'Federal',
'Paper', 'Scott', 'Mead', 'International', 'Board']), 'nnp_loc': {'Louisiana-Pacific': 41,

```
'Champion': 33, 'Federal': 7, 'Paper': 26, 'Scott': 16, 'Mead': 0, 'International': 34,
'Board': 9}, 'sentnum': 30, 'fileid': 'wsj_0162.mrg'}
sentence:        Mead rose 3\/4 to 39 1\/2 , Federal Paper Board added 1\/2 to 24 3\/8 and
Scott Paper gained 1\/2 to 48 3\/8 , while International Paper fell 7\/8 to 48 7\/8 ,
Champion International lost 3\/8 to 31 1\/2 and Louisiana-Pacific dropped 1\/8 to 40 1\/4 .
_____TB sentence_____
{'word_loc': {'and': 29, 'rating': 6, 'Northern': 21, 'Great': 20, "'": 12, 'Mr.': 2,
'International': 24, 'Weyerhaeuser': 30, 'Louisiana': 27, 'from': 7, 'Champion': 17,
'avoid': 11, ',': 26, 'Boise': 14, 'to': 9, 'Adler': 3, 'week': 1, '``': 10, 'his': 5,
'Last': 0, 'Cascade': 15, 'lowered': 4, 'Nekoosa': 22, 'hold': 8, 'on': 13, 'Pacific': 28,
'Paper': 25, '.': 31}, 'nnp': Set(['Great', 'Champion', 'Northern', 'Pacific', 'Boise',
'Cascade', 'Mr.', 'Paper', 'Nekoosa', 'International', 'Weyerhaeuser', 'Louisiana',
'Adler']), 'nnp_loc': {'Great': 20, 'Champion': 17, 'Northern': 21, 'Pacific': 28, 'Boise':
14, 'Cascade': 15, 'Mr.': 2, 'Paper': 25, 'Nekoosa': 22, 'International': 24,
'Weyerhaeuser': 30, 'Louisiana': 27, 'Adler': 3}, 'sentnum': 13, 'fileid': 'wsj_0666.mrg'}
sentence:        Last week Mr. Adler lowered his rating from hold to `` avoid '' on Boise
Cascade , Champion International , Great Northern Nekoosa , International Paper , Louisiana
Pacific and Weyerhaeuser .
_____
SKIPPING (blank argument): {'ARG0': 'HCC Insurance Holdings', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'Kraft Foods', 'ARG1': 'General Foods'}
_____TB sentence_____
{'word_loc': {'157': 29, 'executive': 5, 'for': 27, '.': 32, 'unit': 16, 'Sept.': 25, '$':
28, "'": 12, ',': 17, 'shares': 20, 'A.': 1, 'Philip': 8, 'Kraft': 13, 'bought': 18, 'Cos':
10, 'company': 23, 'Michael': 0, '6,000': 19, 'Morris': 9, 'Miles': 2, 'Foods': 15, 'on':
24, '22': 26, '*U*': 30, 'of': 21, 'General': 14, 'chief': 4, 'officer': 6, 'each': 31,
'the': 22}, 'nnp': Set(['Cos', 'Sept.', 'Michael', 'Foods', 'General', 'Morris', 'Miles',
'Philip', 'Kraft']), 'nnp_loc': {'Cos': 10, 'Sept.': 25, 'Michael': 0, 'Foods': 15,
'General': 14, 'Morris': 9, 'Miles': 2, 'Philip': 8, 'Kraft': 13}, 'sentnum': 31, 'fileid':
'wsj_1157.mrg'}
sentence:        Michael A. Miles , chief executive officer of Philip Morris Cos . ' Kraft
General Foods unit , bought 6,000 shares of the company on Sept. 22 for $ 157 *U* each .
_____TB sentence_____
{'word_loc': {'purchases': 4, 'named': 9, 'formerly': 20, 'being': 8, '*-1': 10, 'before':
6, "'": 26, '*': 7, 'Foods': 25, '.': 30, 'current': 13, 'to': 11, ';': 19, 'Kraft': 23,
'was': 22, 'executive': 15, 'his': 12, "'s": 17, 'operating': 28, 'The': 0, 'he': 21,
'made': 2, 'shortly': 5, 'General': 24, 'chief': 27, 'officer': 29, 'position': 18}, 'nnp':
Set(['Foods', 'Kraft', 'General']), 'nnp_loc': {'Foods': 25, 'Kraft': 23, 'General': 24},
'sentnum': 36, 'fileid': 'wsj_1157.mrg'}
sentence:        The executive made his purchases shortly before * being named *-1 to his
current chief executive officer 's position ; formerly he was Kraft General Foods ' chief
operating officer .
_____TB sentence_____
{'word_loc': {'and': 7, 'Express': 1, 'ad': 21, 'move': 13, 'relationships': 18, 'said':
10, 'their': 17, 'agency': 22, ',': 6, '.': 23, '0': 11, 'Kraft': 3, 'wo': 14, 'Mattel': 8,
'American': 0, 'affect': 16, 'with': 19, 'General': 4, "n't": 15, 'Foods': 5, 'the': 20,
'executives': 9}, 'nnp': Set(['Express', 'Foods', 'Mattel', 'American', 'General',
'Kraft']), 'nnp_loc': {'Express': 1, 'Foods': 5, 'Mattel': 8, 'American': 0, 'General': 4,
'Kraft': 3}, 'sentnum': 42, 'fileid': 'wsj_1317.mrg'}
sentence:        American Express , Kraft General Foods , and Mattel executives said 0 the
move wo n't affect their relationships with the ad agency .
_____TB sentence_____
{'word_loc': {'and': 17, '*T*-2': 27, '*-1': 20, "'": 25, 'are': 8, 'its': 6,
'relationships': 4, 'said': 26, 'based': 9, "'": 14, '*?*': 23, ',': 30, '.': 38, 'to': 21,
'Kraft': 35, 'be': 22, '``': 0, "'s": 3, 'agencies': 13, 'David': 28, 'president': 33,
'with': 5, 'a': 31, 'on': 11, 'Hurwitt': 29, 'vice': 32, 'of': 34, 'work': 15, 'General':
36, 'will': 18, 'Foods': 37, 'continue': 19, 'the': 12}, 'nnp': Set(['Kraft', 'Foods',
'David', 'Hurwitt', 'General']), 'nnp_loc': {'Kraft': 35, 'Foods': 37, 'David': 28,
'Hurwitt': 29, 'General': 36}, 'sentnum': 43, 'fileid': 'wsj_1317.mrg'}
sentence:        `` General Foods 's relationships with its agencies are based *-1 on the
agencies ' work , and will continue *-1 to be *?* , '' said *T*-2 David Hurwitt , a vice
president of Kraft General Foods .
_____TB sentence_____
{'word_loc': {'and': 21, '*-1': 6, 'share': 19, 'in': 25, 'its': 22, 'market': 18, '1987':
```

13, 'for': 8, '.': 29, 'several': 17, 'gave': 15, 'coffee': 27, 'business': 28, 'year': 11,
'brand': 4, 'leadership': 23, 'The': 0, 'a': 10, 'about': 9, 'Kraft': 1, 'up': 16,
'General': 2, 'Foods': 3, 'points': 20, 'stopped': 5, 'advertising': 7, 'position': 24,
'the': 26}, 'nnp': Set(['Foods', 'Kraft', 'General']), 'nnp_loc': {'Foods': 3, 'Kraft': 1,
'General': 2}, 'sentnum': 70, 'fileid': 'wsj_1856.mrg'}
sentence: The Kraft General Foods brand stopped *-1 advertising for about a year in
1987 and gave up several market share points and its leadership position in the coffee
business .
_____TB sentence_____
{'word_loc': {'and': 19, 'says': 28, 'USA': 38, '*T*-1': 29, 'is': 13, 'Dick': 30,
'coffee': 17, "'''": 27, 'committed': 14, 'Mayer': 31, 'business': 18, "'": 9, ',': 32, '.':
40, 'to': 20, 'Philip': 3, 'Kraft': 6, '-RCB-': 12, 'division': 39, '``': 0, '-LCB-': 5,
'increased': 21, 'parent': 10, 'company': 11, 'of': 34, 'Morris': 4, 'president': 33,
'Now': 1, 'for': 23, 'Maxwell': 24, 'General': 36, 'Foods': 37, 'advertising': 22, 'House':
25, 'the': 35}, 'nnp': Set(['Mayer', 'USA', 'Maxwell', 'Dick', 'Morris', 'General',
'Foods', 'House', 'Philip', 'Kraft']), 'nnp_loc': {'Mayer': 31, 'USA': 38, 'Maxwell': 24,
'Dick': 30, 'Morris': 4, 'General': 36, 'Foods': 37, 'House': 25, 'Philip': 3, 'Kraft': 6},
'sentnum': 72, 'fileid': 'wsj_1856.mrg'}
sentence: `` Now , Philip Morris -LCB- Kraft General Foods ' parent company -RCB- is
committed to the coffee business and to increased advertising for Maxwell House , '' says
T-1 Dick Mayer , president of the General Foods USA division .
_____TB sentence_____
{'word_loc': {'about': 14, 'made': 1, '%': 16, 'share': 20, 'of': 17, '*': 2, 'had': 13,
'Morris': 11, '.': 21, 'General': 6, 'Philip': 10, 'Foods': 7, 'market': 19, 'Cos.': 12,
'Brands': 0, 'the': 18, 'Kraft': 5, '95': 15, 'by': 3, 'unit': 8}, 'nnp': Set(['Cos.',
'Foods', 'General', 'Morris', 'Philip', 'Kraft']), 'nnp_loc': {'Cos.': 12, 'Foods': 7,
'General': 6, 'Morris': 11, 'Philip': 10, 'Kraft': 5}, 'sentnum': 112, 'fileid':
'wsj_1875.mrg'}
sentence: Brands made * by the Kraft General Foods unit of Philip Morris Cos. had
about 95 % of the market share .
_____TB sentence_____
{'word_loc': {'detailed': 4, 'and': 27, 'major': 9, '*-1': 13, 'except': 12, 'say': 15,
'including': 19, 'Co.': 30, 'information': 5, 'little': 3, 'Brewing': 29, ',': 31, '.': 35,
'to': 14, 'disclosed': 2, 'performance': 7, 'Philip': 20, 'Kraft': 24, 'increased': 33,
'business': 10, 'that': 16, 'most': 17, 'Morris': 21, 'by': 8, 'posted': 32, 'about': 6,
'U.S.A.': 22, 'lines': 11, 'General': 25, 'Foods': 26, 'Miller': 28, 'revenues': 34},
'nnp': Set(['U.S.A.', 'Brewing', 'Morris', 'General', 'Foods', 'Miller', 'Philip', 'Kraft',
'Co.']), 'nnp_loc': {'U.S.A.': 22, 'Brewing': 29, 'Morris': 21, 'General': 25, 'Foods': 26,
'Miller': 28, 'Philip': 20, 'Kraft': 24, 'Co.': 30}, 'sentnum': 5, 'fileid':
'wsj_1885.mrg'}
sentence: Philip Morris disclosed little detailed information about performance by
major business lines except *-1 to say that most , including Philip Morris U.S.A. , Kraft
General Foods and Miller Brewing Co. , posted increased revenues .
_____
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'SevenSpace'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Massa'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'ON Technology'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Rochester Gas and Electric Corporation'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Vauxhall Motors'}
_____MATCH!!!_____
{'ARG0': 'American Airlines', 'ARG1': 'Trans World Airlines'}
_____TB sentence_____
{'word_loc': {'and': 21, 'Airlines': 24, 'as': 6, 'additions': 1, 'Pan': 17, 'Airways': 20,
',': 12, '.': 25, 'behind': 13, 'World': 23, 'Corp.': 15, "'s": 16, 'U.S.': 9,
'transatlantic': 10, 'American': 18, 'The': 0, 'Trans': 22, 'PanAm': 14, 'third-largest':
8, 'carrier': 11, 'solidify': 2, 'position': 5, 'the': 7}, 'nnp': Set(['Trans', 'Airways',
'PanAm', 'Airlines', 'American', 'World', 'Corp.', 'U.S.', 'Pan']), 'nnp_loc': {'Trans':
22, 'Airways': 20, 'PanAm': 14, 'Airlines': 24, 'American': 18, 'World': 23, 'Corp.': 15,
'U.S.': 9, 'Pan': 17}, 'sentnum': 5, 'fileid': 'wsj_0502.mrg'}
sentence: The additions solidify American 's position as the third-largest U.S.
transatlantic carrier , behind PanAm Corp. 's Pan American World Airways and Trans World
Airlines .
_____TB sentence_____
{'word_loc': {'and': 10, 'certain': 39, '*-1': 23, 'Airlines': 9, 'frequent-flier': 40,
'as': 22, 'in': 45, 'Pan': 15, '$': 35, 'for': 4, 'Airways': 18, ',': 6, '.': 51, 'to': 38,

'flights': 44, 'World': 17, 'Corp.': 13, 'checks': 30, 'week': 2, '200': 33, "'s": 14,
'business': 46, 'offer': 25, 'far': 21, 'rebates': 27, 'American': 16, 'members': 41,
'Trans': 7, 'class': 50, 'trans-Atlantic': 43, 'last': 1, 'gift': 29, 'Just': 0, '*U*': 37,
'of': 31, 'Am': 12, 'cash': 26, 'or': 48, '1,000': 36, 'so': 20, 'making': 42, 'went': 19,
'example': 5, 'first': 49}, 'nnp': Set(['Corp.', 'Airways', 'Am', 'Airlines', 'American',
'World', 'Trans', 'Pan']), 'nnp_loc': {'Corp.': 13, 'Airways': 18, 'Am': 12, 'Airlines': 9,
'American': 16, 'World': 17, 'Trans': 7, 'Pan': 15}, 'sentnum': 13, 'fileid':
'wsj_1192.mrg'}
sentence: Just last week , for example , Trans World Airlines and Pan Am Corp. 's Pan
American World Airways went so far as *-1 to offer cash rebates or gift checks of $ 200 to
$ 1,000 *U* to certain frequent-flier members making trans-Atlantic flights in business
class or first class .

_____
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Wheal Chance'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Biomet'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Wheal Treasure'}
_____MATCH!!!_____
{'ARG0': 'Commerzbank', 'ARG1': 'Dresdner Bank'}
_____TB sentence_____
{'word_loc': {'and': 5, 'Dresdner': 3, 'all': 7, '1992': 18, 'eagerness': 11, 'to': 13,
'.': 19, '*-1': 12, 'Deutsche': 0, 'expressed': 10, ',': 2, 'also': 8, 'expand': 14,
'have': 9, 'in': 15, 'France': 16, 'Commerzbank': 6, 'Bank': 4, 'before': 17}, 'nnp':
Set(['Dresdner', 'Commerzbank', 'Deutsche', 'Bank', 'France']), 'nnp_loc': {'Dresdner': 3,
'Commerzbank': 6, 'Deutsche': 0, 'Bank': 4, 'France': 16}, 'sentnum': 9, 'fileid':
'wsj_0477.mrg'}
sentence: Deutsche Bank , Dresdner Bank and Commerzbank all also have expressed
eagerness *-1 to expand in France before 1992 .
_____TB sentence_____
{'word_loc': {'and': 15, 'Dresdner': 8, 'shedding': 10, 'to': 19, 'with': 7, ',': 14,
'2.5': 18, '4.8': 11, 'resilient': 5, '320': 13, 'Other': 0, 'slipping': 17, 'were': 2,
'banks': 1, 'slightly': 3, 'Commerzbank': 16, 'Bank': 9, 'more': 4}, 'nnp':
Set(['Dresdner', 'Commerzbank', 'Bank']), 'nnp_loc': {'Dresdner': 8, 'Commerzbank': 16,
'Bank': 9}, 'sentnum': 48, 'fileid': 'wsj_1015.mrg'}
sentence: Other banks were slightly more resilient , with Dresdner Bank shedding 4.8
to 320 , and Commerzbank slipping 2.5 to

_____
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Newsweek Inc'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Central Maine Power Company'}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'OKCupid'}
SKIPPING (blank argument): {'ARG0': 'SK Telecom', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Activision Blizzard', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Multinational Underwriters LLC', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'UMNO Singapore'}
SKIPPING (blank argument): {'ARG0': 'WorldCom', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Google', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'Advance Publications', 'ARG1': 'Random House'}
_____TB sentence_____
{'word_loc': {'named': 4, '*T*-2': 30, 'family': 23, '*-1': 5, 'Mr.': 12, 'have': 15,
'Bernstein': 13, 'whose': 22, 'clashed': 16, 'Jr.': 20, ',': 29, '.': 34, '*T*-35': 8,
'which': 7, 'was': 2, 'A': 0, 'Advance': 26, 'that': 11, 'may': 14, 'company': 24, 'S.I.':
18, 'fueled': 9, 'Newhouse': 19, 'Publications': 27, 'with': 17, 'Inc.': 28, 'successor':
1, 'House': 33, 'Random': 32, "n't": 3, 'owns': 31, 'speculation': 10}, 'nnp':
Set(['Advance', 'Jr.', 'Random', 'Newhouse', 'S.I.', 'Mr.', 'Publications', 'House',
'Bernstein', 'Inc.']), 'nnp_loc': {'Advance': 26, 'Jr.': 20, 'Random': 32, 'Newhouse': 19,
'S.I.': 18, 'Mr.': 12, 'Publications': 27, 'House': 33, 'Bernstein': 13, 'Inc.': 28},
'sentnum': 1, 'fileid': 'wsj_0111.mrg'}
sentence: A successor was n't named *-1 , which *T*-35 fueled speculation that Mr.
Bernstein may have clashed with S.I. Newhouse Jr. , whose family company , Advance
Publications Inc. , *T*-2 owns Random House .

_____
SKIPPING (blank argument): {'ARG0': 'Cond\xc3\xa9 Nast Publications', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Shopping.com', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Sony Computer Entertainment', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': '', 'ARG1': 'Pure Software'}

SKIPPING (blank argument): {'ARG0': 'News Corporation', 'ARG1': ''}
_____MATCH!!!_____
{'ARG0': 'United Airlines', 'ARG1': 'Continental Airlines'}
_____TB sentence_____
{'word_loc': {'and': 52, 'United': 50, 'Ridge': 8, 'USAir': 53, '...': 59, 'Airlines': 51,
'West': 28, 'it': 14, 'Hertz': 4, 'month': 1, 'Continental': 34, 'ended': 57, 'sometime':
60, 'at': 21, 'year': 22, 'its': 17, 'Texas': 30, 'Dec.': 62, 'said': 12, 'end': 23,
'would': 55, 'Eastern': 37, 'Park': 7, ',': 46, '.': 64, 'also': 54, 'agreements': 19,
'Delta': 25, 'N.J.': 10, 'America': 27, 'UAL': 47, 'A': 0, 'be': 56, "'s": 49, 'that': 41,
'marketing': 18, 'after': 61, 'Air': 31, 'American': 44, 'pacts': 42, 'with': 43, 'ago': 2,
'Corp.': 32, '*-2': 58, 'of': 6, '31': 63, 'drop': 16, 'Inc': 48, '*ICH*-1': 20}, 'nnp':
Set(['Dec.', 'Corp.', 'Ridge', 'USAir', 'Eastern', 'West', 'Park', 'Airlines', 'Hertz',
'Air', 'Continental', 'American', 'United', 'UAL', 'Delta', 'N.J.', 'America', 'Texas',
'Inc']), 'nnp_loc': {'Dec.': 62, 'Corp.': 32, 'Ridge': 8, 'USAir': 53, 'Eastern': 37,
'West': 28, 'Park': 7, 'Airlines': 51, 'Hertz': 4, 'Air': 31, 'Continental': 34,
'American': 44, 'United': 50, 'UAL': 47, 'Delta': 25, 'N.J.': 10, 'America': 27, 'Texas':
30, 'Inc': 48}, 'sentnum': 10, 'fileid': 'wsj_2394.mrg'}
sentence: A month ago , Hertz , of Park Ridge , N.J. , said that it would drop its
marketing agreements *ICH*-1 at year end with Delta , America West and Texas Air Corp. 's
Continental Airlines and Eastern Airlines , and that pacts with American Airlines , UAL Inc
's United Airlines and USAir also would be ended *-2 ... sometime after Dec. 31 .

_____
SKIPPING (blank argument): {'ARG0': 'Infochimps', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Bear Stearns', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Skotos', 'ARG1': ''}
SKIPPING (blank argument): {'ARG0': 'Bank One Corporation', 'ARG1': ''}

In [ ]: