

Lab 2

Steven Bird, Ewan Klein, and Edward Loper

5 Categorizing and Tagging Words

14. Use `sorted()` and `set()` to get a sorted list of tags used in the Brown corpus, removing duplicates.

Answer:

```
Grabbing tagged Text...
Finding unique tags...
Printing 472 tags...
["'",
 "'",
 '(',
 '(-HL',
 ')',
 ')-HL',
 '* ',
 '*-HL',
 '*-NC',
 '*-TL',
 ', ',
 ', -HL',
 ', -NC',
 ', -TL',
 '- ',
 '---HL',
 '. ',
 '. -HL',
 '. -NC',
 '. -TL',
 ': ',
 ': -HL',
 ': -TL',
 'ABL',
 'ABN',
 'ABN-HL',
```

'ABN-NC',
'ABN-TL',
'ABX',
'AP',
'AP\$',
'AP+AP-NC',
'AP-HL',
'AP-NC',
'AP-TL',
'AT',
'AT-HL',
'AT-NC',
'AT-TL',
'AT-TL-HL',
'BE',
'BE-HL',
'BE-TL',
'BED',
'BED*',
'BED-NC',
'BEDZ',
'BEDZ*',
'BEDZ-HL',
'BEDZ-NC',
'BEG',
'BEM',
'BEM*',
'BEM-NC',
'BEN',
'BEN-TL',
'BER',
'BER*',
'BER*-NC',
'BER-HL',
'BER-NC',
'BER-TL',
'BEZ',
'BEZ*',
'BEZ-HL',
'BEZ-NC',
'BEZ-TL',
'CC',
'CC-HL',
'CC-NC',
'CC-TL',
'CC-TL-HL',
'CD',

'CD\$ ',
'CD-HL ',
'CD-NC ',
'CD-TL ',
'CD-TL-HL ',
'CS ',
'CS-HL ',
'CS-NC ',
'CS-TL ',
'DO ',
'DO* ',
'DO*-HL ',
'DO+PPSS ',
'DO-HL ',
'DO-NC ',
'DO-TL ',
'DOD ',
'DOD* ',
'DOD*-TL ',
'DOD-NC ',
'DOZ ',
'DOZ* ',
'DOZ*-TL ',
'DOZ-HL ',
'DOZ-TL ',
'DT ',
'DT\$ ',
'DT+BEZ ',
'DT+BEZ-NC ',
'DT+MD ',
'DT-HL ',
'DT-NC ',
'DT-TL ',
'DTI ',
'DTI-HL ',
'DTI-TL ',
'DTS ',
'DTS+BEZ ',
'DTS-HL ',
'DTX ',
'EX ',
'EX+BEZ ',
'EX+HVD ',
'EX+HVZ ',
'EX+MD ',
'EX-HL ',
'EX-NC ',

'FW-*',
 'FW*-TL',
 'FW-AT',
 'FW-AT+NN-TL',
 'FW-AT+NP-TL',
 'FW-AT-HL',
 'FW-AT-TL',
 'FW-BE',
 'FW-BER',
 'FW-BEZ',
 'FW-CC',
 'FW-CC-TL',
 'FW-CD',
 'FW-CD-TL',
 'FW-CS',
 'FW-DT',
 'FW-DT+BEZ',
 'FW-DTS',
 'FW-HV',
 'FW-IN',
 'FW-IN+AT',
 'FW-IN+AT-T',
 'FW-IN+AT-TL',
 'FW-IN+NN',
 'FW-IN+NN-TL',
 'FW-IN+NP-TL',
 'FW-IN-TL',
 'FW-JJ',
 'FW-JJ-NC',
 'FW-JJ-TL',
 'FW-JJR',
 'FW-JJT',
 'FW-NN',
 'FW-NN\$',
 'FW-NN\$-TL',
 'FW-NN-NC',
 'FW-NN-TL',
 'FW-NN-TL-NC',
 'FW-NNS',
 'FW-NNS-NC',
 'FW-NNS-TL',
 'FW-NP',
 'FW-NP-TL',
 'FW-NPS',
 'FW-NPS-TL',
 'FW-NR',
 'FW-NR-TL',

'FW-OD-NC',
 'FW-OD-TL',
 'FW-PN',
 'FW-PP\$',
 'FW-PP\$-NC',
 'FW-PP\$-TL',
 'FW-PPL',
 'FW-PPL+VBZ',
 'FW-PPO',
 'FW-PPO+IN',
 'FW-PPS',
 'FW-PPSS',
 'FW-PPSS+HV',
 'FW-QL',
 'FW-RB',
 'FW-RB+CC',
 'FW-RB-TL',
 'FW-TO+VB',
 'FW-UH',
 'FW-UH-NC',
 'FW-UH-TL',
 'FW-VB',
 'FW-VB-NC',
 'FW-VB-TL',
 'FW-VBD',
 'FW-VBD-TL',
 'FW-VBG',
 'FW-VBG-TL',
 'FW-VBN',
 'FW-VBZ',
 'FW-WDT',
 'FW-WPO',
 'FW-WPS',
 'HV',
 'HV*',
 'HV+TO',
 'HV-HL',
 'HV-NC',
 'HV-TL',
 'HVD',
 'HVD*',
 'HVD-HL',
 'HVG',
 'HVG-HL',
 'HVN',
 'HVZ',
 'HVZ*',

'HVZ-NC',
 'HVZ-TL',
 'IN',
 'IN+IN',
 'IN+PPO',
 'IN-HL',
 'IN-NC',
 'IN-TL',
 'IN-TL-HL',
 'JJ',
 'JJ\$-TL',
 'JJ+JJ-NC',
 'JJ-HL',
 'JJ-NC',
 'JJ-TL',
 'JJ-TL-HL',
 'JJ-TL-NC',
 'JJR',
 'JJR+CS',
 'JJR-HL',
 'JJR-NC',
 'JJR-TL',
 'JJS',
 'JJS-HL',
 'JJS-TL',
 'JJT',
 'JJT-HL',
 'JJT-NC',
 'JJT-TL',
 'MD',
 'MD*',
 'MD*-HL',
 'MD+HV',
 'MD+PPSS',
 'MD+TO',
 'MD-HL',
 'MD-NC',
 'MD-TL',
 'NIL',
 'NN',
 'NN\$',
 'NN\$-HL',
 'NN\$-TL',
 'NN+BEZ',
 'NN+BEZ-TL',
 'NN+HVD-TL',
 'NN+HVZ',

'NN+HVZ-TL',
 'NN+IN',
 'NN+MD',
 'NN+NN-NC',
 'NN-HL',
 'NN-NC',
 'NN-TL',
 'NN-TL-HL',
 'NN-TL-NC',
 'NNS',
 'NNS\$',
 'NNS\$-HL',
 'NNS\$-NC',
 'NNS\$-TL',
 'NNS\$-TL-HL',
 'NNS+MD',
 'NNS-HL',
 'NNS-NC',
 'NNS-TL',
 'NNS-TL-HL',
 'NNS-TL-NC',
 'NP',
 'NP\$',
 'NP\$-HL',
 'NP\$-TL',
 'NP+BEZ',
 'NP+BEZ-NC',
 'NP+HVZ',
 'NP+HVZ-NC',
 'NP+MD',
 'NP-HL',
 'NP-NC',
 'NP-TL',
 'NP-TL-HL',
 'NPS',
 'NPS\$',
 'NPS\$-HL',
 'NPS\$-TL',
 'NPS-HL',
 'NPS-NC',
 'NPS-TL',
 'NR',
 'NR\$',
 'NR\$-TL',
 'NR+MD',
 'NR-HL',
 'NR-NC',

'NR-TL ',
 'NR-TL-HL ',
 'NRS ',
 'NRS-TL ',
 'OD ',
 'OD-HL ',
 'OD-NC ',
 'OD-TL ',
 'PN ',
 'PN\$ ',
 'PN+BEZ ',
 'PN+HVD ',
 'PN+HVZ ',
 'PN+MD ',
 'PN-HL ',
 'PN-NC ',
 'PN-TL ',
 'PP\$ ',
 'PP\$\$ ',
 'PP\$-HL ',
 'PP\$-NC ',
 'PP\$-TL ',
 'PPL ',
 'PPL-HL ',
 'PPL-NC ',
 'PPL-TL ',
 'PPLS ',
 'PPO ',
 'PPO-HL ',
 'PPO-NC ',
 'PPO-TL ',
 'PPS ',
 'PPS+BEZ ',
 'PPS+BEZ-HL ',
 'PPS+BEZ-NC ',
 'PPS+HVD ',
 'PPS+HVZ ',
 'PPS+MD ',
 'PPS-HL ',
 'PPS-NC ',
 'PPS-TL ',
 'PPSS ',
 'PPSS+BEM ',
 'PPSS+BER ',
 'PPSS+BER-N ',
 'PPSS+BER-NC ',
 'PPSS+BER-TL ',

'PPSS+BEZ',
 'PPSS+BEZ*',
 'PPSS+HV',
 'PPSS+HV-TL',
 'PPSS+HVD',
 'PPSS+MD',
 'PPSS+MD-NC',
 'PPSS+VB',
 'PPSS-HL',
 'PPSS-NC',
 'PPSS-TL',
 'QL',
 'QL-HL',
 'QL-NC',
 'QL-TL',
 'QLP',
 'RB',
 'RB\$',
 'RB+BEZ',
 'RB+BEZ-HL',
 'RB+BEZ-NC',
 'RB+CS',
 'RB-HL',
 'RB-NC',
 'RB-TL',
 'RBR',
 'RBR+CS',
 'RBR-NC',
 'RBT',
 'RN',
 'RP',
 'RP+IN',
 'RP-HL',
 'RP-NC',
 'RP-TL',
 'TO',
 'TO+VB',
 'TO-HL',
 'TO-NC',
 'TO-TL',
 'UH',
 'UH-HL',
 'UH-NC',
 'UH-TL',
 'VB',
 'VB+AT',
 'VB+IN',

'VB+JJ-NC',
 'VB+PPO',
 'VB+RP',
 'VB+TO',
 'VB+VB-NC',
 'VB-HL',
 'VB-NC',
 'VB-TL',
 'VBD',
 'VBD-HL',
 'VBD-NC',
 'VBD-TL',
 'VBG',
 'VBG+TO',
 'VBG-HL',
 'VBG-NC',
 'VBG-TL',
 'VBN',
 'VBN+TO',
 'VBN-HL',
 'VBN-NC',
 'VBN-TL',
 'VBN-TL-HL',
 'VBN-TL-NC',
 'VBZ',
 'VBZ-HL',
 'VBZ-NC',
 'VBZ-TL',
 'WDT',
 'WDT+BER',
 'WDT+BER+PP',
 'WDT+BEZ',
 'WDT+BEZ-HL',
 'WDT+BEZ-NC',
 'WDT+BEZ-TL',
 'WDT+DO+PPS',
 'WDT+DOD',
 'WDT+HVZ',
 'WDT-HL',
 'WDT-NC',
 'WP\$',
 'WPO',
 'WPO-NC',
 'WPO-TL',
 'WPS',
 'WPS+BEZ',
 'WPS+BEZ-NC',

```

'WPS+BEZ-TL',
'WPS+HVD',
'WPS+HVZ',
'WPS+MD',
'WPS-HL',
'WPS-NC',
'WPS-TL',
'WQL',
'WQL-TL',
'WRB',
'WRB+BER',
'WRB+BEZ',
'WRB+BEZ-TL',
'WRB+DO',
'WRB+DOD',
'WRB+DOD*',
'WRB+DOZ',
'WRB+IN',
'WRB+MD',
'WRB-HL',
'WRB-NC',
'WRB-TL',
',',',',']

```

15. Write programs to process the Brown Corpus and find answers to the following questions:

1. Which nouns are more common in their plural form, rather than their singular form? (Only consider regular plurals, formed with the **-s** suffix.)
2. Which word has the greatest number of distinct tags. What are they, and what do they represent?
3. List tags in order of decreasing frequency. What do the 20 most frequent tags represent?
4. Which tags are nouns most commonly found after? What do these tags represent?

Answer:

```

Grabbing tagged Text...
1. Finding mostly plural nouns...
{('barrack', 'barracks'): (1, 2),
 ('bicep', 'biceps'): (1, 3),
 ('ethic', 'ethics'): (4, 6),
 ('headquarter', 'headquarters'): (1, 37),
 ('mean', 'means'): (12, 74),
 ('metaphysic', 'metaphysics'): (3, 11)}
2. Finding word with most distinct tags...
('that',
 {'CS',

```

'CS-HL',
 'CS-NC',
 'DT',
 'DT-NC',
 'NIL',
 'QL',
 'WPO',
 'WPO-NC',
 'WPS',
 'WPS-HL',
 'WPS-NC'})

3. Listing most common tags...

```
[('NN', 152470),
 ('IN', 120557),
 ('AT', 97959),
 ('JJ', 64028),
 ('.', 60638),
 (',', 58156),
 ('NNS', 55110),
 ('CC', 37718),
 ('RB', 36464),
 ('NP', 34476),
 ('VB', 33693),
 ('VBN', 29186),
 ('VBD', 26167),
 ('CS', 22143),
 ('PPS', 18253),
 ('VBG', 17893),
 ('PP$', 16872),
 ('TO', 14918),
 ('PPSS', 13802),
 ('CD', 13510)]
```

4. Finding noun collocations...

```
22449 ('JJ', 'NN') [ (('fiscal', 'year'), 56), (('high', 'school'),
    , 54), (('old', 'man'), 52), (('young', 'man'), 47), (('great',
    , 'deal'), 43)]
12288 ('AT', 'NN') [ (('the', 'world'), 334), (('a', 'man'), 249),
    (('the', 'time'), 242), (('the', 'way'), 224), (('the', 'end',
    , 203)]
9850 ('JJ', 'NNS') [ (('new', 'members'), 30), (('young', 'men'),
    29), (('recent', 'years'), 25), (('American', 'people'), 22),
    (('human', 'beings'), 22)]
8592 ('NN', 'NN') [ (('wage', 'rate'), 24), (('index', 'word'),
    21), (('information', 'cell'), 18), (('carbon', 'tetrachloride',
    , 18), (('anode', 'holder'), 18)]
7621 ('IN', 'NN') [ (('of', 'course'), 237), (('per', 'cent'),
    146), (('in', 'order'), 129), (('for', 'example'), 109), (('of
```

, 'life'), 106)]

4765 ('NN', 'NNS') [(('home', 'runs'), 25), (('index', 'words'), 21), (('school', 'districts'), 16), (('interest', 'rates'), 15), (('faculty', 'members'), 14)]

4342 ('PP\$', 'NN') [(('his', 'head'), 122), (('his', 'wife'), 93), (('his', 'face'), 72), (('his', 'mind'), 65), (('his', 'life'), 63)]

3429 ('IN', 'NNS') [(('of', 'men'), 72), (('in', 'terms'), 72), (('of', 'people'), 59), (('of', 'years'), 36), (('at', 'times'), 27)]

2762 ('AT', 'NNS') [(('the', 'people'), 113), (('the', 'men'), 75), (('the', 'others'), 60), (('the', 'children'), 56), (('the', 'years'), 48)]

2612 ('CC', 'NN') [(('and', 'Af'), 78), (('and', 'development'), 23), (('and', 'equipment'), 22), (('and', 'water'), 17), (('and', 'wife'), 17)]

2356 ('VBG', 'NN') [(('living', 'room'), 26), (('dining', 'room'), 17), (('locking', 'bar'), 10), (('cutting', 'force'), 7), (('starting', 'buffer'), 7)]

2069 ('DT', 'NN') [(('this', 'time'), 92), (('that', 'time'), 65), (('this', 'country'), 63), (('this', 'way'), 50), (('this', 'year'), 49)]

2067 ('VBN', 'NN') [(('taken', 'place'), 11), (('barbed', 'wire'), 9), (('aerated', 'lagoon'), 7), (('mixed', 'liquor'), 7), (('married', 'life'), 6)]

2035 ('CD', 'NNS') [(('two', 'years'), 58), (('ten', 'years'), 34), (('two', 'men'), 33), (('five', 'years'), 31), (('two', 'weeks'), 31)]

1819 ('PP\$', 'NNS') [(('his', 'eyes'), 76), (('his', 'hands'), 53), (('his', 'feet'), 44), (('her', 'eyes'), 35), (('his', 'lips'), 27)]

1553 ('AP', 'NN') [(('same', 'time'), 94), (('last', 'year'), 68), (('other', 'hand'), 60), (('last', 'night'), 55), (('last', 'week'), 43)]

1503 ('', 'NN') [(('', 'Af'), 46), (('', 'sir'), 29), (('', 'president'), 21), (('', 'p'), 19), (('', 'state'), 16)]

1393 ('AP', 'NNS') [(('few', 'years'), 49), (('few', 'days'), 41), (('many', 'years'), 36), (('few', 'minutes'), 34), (('other', 'words'), 26)]

1374 ('VBG', 'NNS') [(('locking', 'bars'), 9), (('operating', 'variables'), 8), (('training', 'programs'), 8), (('operating', 'costs'), 7), (('burning', 'arcs'), 6)]

1346 ('CC', 'NNS') [(('and', 'others

18. Generate some statistics for tagged data to answer the following questions:

1. What proportion of word types are always assigned the same part-of-speech tag?

2. How many words are ambiguous, in the sense that they appear with at least two tags?
3. What proportion of word *tokens* in the Brown Corpus involve these ambiguous words?

Answer:

Grabbing tagged Text...

1. Finding proportion of word types(?) always tagged same...

Proportion Unambiguous: 0.456

2. Finding ambiguous words...

Number Ambiguous: 30498

3. Calculating percentage of ambiguous tokens...

Proportion Ambiguous Tokens: 0.978

35. Write a program to classify contexts involving the word *must* according to the tag of the following word. Can this be used to discriminate between the epistemic and deontic uses of *must*?

Answer: Deontic and Epistemic definitions need to be well-defined first. It appears that many of the categories below might be able to be interpreted into one of the other with some degree of reliability.

Grabbing tagged Text...

Categorizing...

```
AT:{
    only must the : RB MD AT
    so must the : RB MD AT
}
ABX:{
    Af must both : NN MD ABX
}
PPSS:{
    times must we : NNS MD PPSS
}
BE-HL:{
    Life must be : NN-HL MD-HL BE-HL
    There must be : EX-HL MD-HL BE-HL
}
VBZ:{
    fans must needs : NNS MD VBZ
}
,:{
    Wagner must , : NP MD ,
    organization must , : NN MD ,
    We must , : PPSS MD ,
    I must , : PPSS MD ,
    and must , : CC MD ,
    course must , : NN MD ,
```

```

degrees must , : NNS MD ,
wine must , : NN MD ,
He must , : PPS MD ,
He must , : PPS MD ,
It must , : PPS MD ,
We must , : PPSS MD ,
they must , : PPSS MD ,
he must , : PPS MD ,
They must , : PPSS MD ,
generators must , : NNS MD ,
gentleman must , : NN MD ,
}
HV-TL:{
    You Must Have : PPSS-TL MD-TL HV-TL
}
NN:{
    South must guarantee : NR-TL MD NN
}
VB-HL:{
    ‘‘ Must solve : ‘‘ MD-HL VB-HL
    We must use : PPSS-HL MD-HL VB-HL
}
NIL:{
    obligations must be : NIL NIL NIL
}
VB-TL:{
    Knight Must Fall : NP-TL MD-TL VB-TL
    We Must Join : PPSS-TL MD-TL VB-TL
}
BE:{
    courses must be : NNS MD BE
    sheet must be : NN MD BE
    and must be : CC MD BE
    ‘‘ must be : ‘‘ MD BE
    they must be : PPSS MD BE
    offenses must be : NNS MD BE
    we must be : PPSS MD BE
    ye must be : PPSS MD BE
    lay-offs must be : NNS MD BE
    We must be : PPSS MD BE
    deduction must be : NN MD BE
    Requests must be : NNS MD BE
    it must be : PPS MD BE
    there must be : EX MD BE
    method must be : NN MD BE
    problem must be : NN MD BE
    but must be : CC MD BE

```

nevertheless must be : RB MD BE
leader must be : NN MD BE
crash must be : NN MD BE
boost must be : NN MD BE
there must be : EX MD BE
we must be : PPSS MD BE
we must be : PPSS MD BE
man must be : NN MD BE
it must be : PPS MD BE
public must be : NN MD BE
course must be : NN MD BE
Moloch must be : NP MD BE
this must be : DT MD BE
there must be : EX MD BE
, must be : , MD BE
it must be : PPS MD BE
it must be : PPS MD BE
error must be : NN MD BE
two must be : CD MD BE
States must be : NNS-TL MD BE
man must be : NN MD BE
leader must be : NN MD BE
babies must be : NNS MD BE
thought must be : NN MD BE
it must be : PPS MD BE
demand must be : NN MD BE
This must be : DT MD BE
demythologization must be : NN MD BE
so must be : RB MD BE
it must be : PPS MD BE
it must be : PPS MD BE
It must be : PPS MD BE
You must be : PPSS MD BE
It must be : PPS MD BE
church must be : NN MD BE
Care must be : NN MD BE
effect must be : NN MD BE
decision must be : NN MD BE
decisions must be : NNS MD BE
thing must be : NN MD BE
, must be : , MD BE
You must be : PPSS MD BE
You must be : PPSS MD BE
You must be : PPSS MD BE
target must be : NN MD BE
attack must be : NN MD BE
forces must be : NNS MD BE

questions must be : NNS MD BE
moves must be : NNS MD BE
lever must be : NN MD BE
bars must be : NNS MD BE
pieces must be : NNS MD BE
subtraction must be : NN MD BE
volume must be : NN MD BE
gasket must be : NN MD BE
that must be : WPS MD BE
tire must be : NN MD BE
car must be : NN MD BE
measurement must be : NN MD BE
measurement must be : NN MD BE
that must be : DT MD BE
they must be : PPSS MD BE
) must be :) MD BE
divers must be : NNS MD BE
bottles must be : NNS MD BE
clarity must be : NN MD BE
clarity must be : NN MD BE
It must be : PPS MD BE
and must be : CC MD BE
lenses must be : NNS MD BE
chest must be : NN MD BE
bunk must be : NN MD BE
cost must be : NN MD BE
roads must be : NNS MD BE
material must be : NN MD BE
He must be : PPS MD BE
work must be : NN MD BE
furniture must be : NN MD BE
subject must be : NN MD BE
also must be : RB MD BE
deceased must be : NN MD BE
naturally must be : RB MD BE
There must be : EX MD BE
message must be : NN MD BE
journeys must be : NNS MD BE
it must be : PPS MD BE
Privacy must be : NN MD BE
This must be : DT MD BE
drama must be : NN MD BE
life must be : NN MD BE
Livestock must be : NN MD BE
that must be : WPS MD BE
rent must be : NN MD BE
taxes must be : NNS MD BE

water must be : NN MD BE
 system must be : NN MD BE
 therefore must be : RB MD BE
 writer must be : NN MD BE
 sea must be : NN MD BE
 freedom must be : NN MD BE
 Bantus must be : NPS MD BE
 lethargy must be : NN MD BE
 and must be : CC MD BE
 this must be : DT MD BE
) must be :) MD BE
 it must be : PPS MD BE
 job must be : NN MD BE
 cities must be : NNS MD BE
 movement must be : NN MD BE
 which must be : WDT MD BE
 Engages must be : FW-NNS MD BE
 traders must be : NNS MD BE
 , must be : , MD BE
 , must be : , MD BE
 , must be : , MD BE
 pieces must be : NNS MD BE
 work must be : NN MD BE
 right must be : NN MD BE
 it must be : PPS MD BE
 action must be : NN MD BE
 man must be : NN MD BE
 murder must be : NN MD BE
 only must be : RB MD BE
 also must be : RB MD BE
 sending must be : NN MD BE
 beings must be : NNS MD BE
 proprietors must be : NNS MD BE
 and must be : CC MD BE
 there must be : EX MD BE
 protected must be : VBN MD BE
 it must be : PPO MD BE
 personality must be : NN MD BE
 It must be : PPS MD BE
 she must be : PPS MD BE
 We must be : PPSS MD BE
 it must be : PPS MD BE
 monacle must be : NN MD BE
 factors must be : NNS MD BE
 literature must be : NN MD BE
 it must be : PPS MD BE
 There must be : EX MD BE

, must be : , MD BE
learning must be : VBG MD BE
it must be : PPS MD BE
they must be : PPSS MD BE
it must be : PPS MD BE
that must be : WPS MD BE
revolts must be : NNS MD BE
them must be : PPO MD BE
There must be : EX MD BE
community must be : NN MD BE
, must be : , MD BE
unavailable must be : JJ MD BE
applicant must be : NN MD BE
loan must be : NN MD BE
cooperative must be : NN MD BE
It must be : PPS MD BE
recipients must be : NNS MD BE
that must be : WPS MD BE
population must be : NN MD BE
program must be : NN MD BE
It must be : PPS MD BE
work must be : NN MD BE
countries must be : NNS MD BE
They must be : PPSS MD BE
They must be : PPSS MD BE
, must be : , MD BE
it must be : PPS MD BE
situation must be : NN MD BE
It must be : PPS MD BE
force must be : NN MD BE
it must be : PPS MD BE
percentage must be : NN MD BE
, must be : , MD BE
share must be : NN MD BE
check must be : NN MD BE
, must be : , MD BE
reports must be : NNS MD BE
Means must be : NNS MD BE
aim must be : NN MD BE
this must be : DT MD BE
programs must be : NNS MD BE
response must be : NN MD BE
warfare must be : NN MD BE
forces must be : NNS MD BE
care must be : NN MD BE
they must be : PPSS MD BE
which must be : WDT MD BE

course must be : NN MD BE
 which must be : WDT MD BE
 due must be : JJ MD BE
 request must be : NN MD BE
 year must be : NN MD BE
 interest must be : NN MD BE
 Production must be : NN MD BE
 it must be : PPS MD BE
 others must be : NNS MD BE
 accommodations must be : NNS MD BE
 Budgeting must be : VBG MD BE
 he must be : PPS MD BE
 which must be : WDT MD BE
 group must be : NN MD BE
 soil must be : NN MD BE
 cloth must be : NN MD BE
 which must be : WDT MD BE
 there must be : EX MD BE
 it must be : PPS MD BE
 parameters must be : NNS MD BE
 and must be : CC MD BE
 It must be : PPS MD BE
 There must be : EX MD BE
 This must be : DT MD BE
 anaconda must be : NN MD BE
 data must be : NNS MD BE
 di-iodotyrosine must be : NN MD BE
 , must be : , MD BE
 there must be : EX MD BE
 trials must be : NNS MD BE
 them must be : PPO MD BE
 f{t} must be : NN MD BE
 Af must be : NN MD BE
 F must be : NP MD BE
 T must be : NP MD BE
 There must be : EX MD BE
 bundle must be : NN MD BE
 regulus must be : NN MD BE
 **zg must be : NN MD BE
 they must be : PPSS MD BE
 communities must be : NNS MD BE
 tasks must be : NNS MD BE
 sources must be : NNS MD BE
 flows must be : NNS MD BE
 balance must be : NN MD BE
 It must be : PPS MD BE
 there must be : EX MD BE

stress must be : NN MD BE
 interview must be : NN MD BE
 worker must be : NN MD BE
 -- must be : -- MD BE
 entries must be : NNS MD BE
 which must be : WDT MD BE
 sections must be : NNS MD BE
 arm-levitation must be : NN MD BE
 response must be : NN MD BE
 she must be : PPS MD BE
 cells must be : NNS MD BE
 chain must be : NN MD BE
 text must be : NN MD BE
 that must be : WPS MD BE
 comments must be : NNS MD BE
 This must be : DT MD BE
 2 must be : CD-TL MD BE
 usually must be : RB MD BE
 field must be : NN MD BE
 configuration must be : NN MD BE
 rate must be : NN MD BE
 action must be : NN MD BE
 claim must be : NN MD BE
 action must be : NN MD BE
 claim must be : NN MD BE
 tax must be : NN MD BE
 still must be : RB MD BE
 teacher must be : NN MD BE
 this must be : DT MD BE
 rates must be : NNS MD BE
 , must be : , MD BE
 it must be : PPS MD BE
 justice must be : NN MD BE
 example must be : NN MD BE
 fanaticism must be : NN MD BE
 representations must be : NNS MD BE
 white must be : NN MD BE
 therefore must be : RB MD BE
 alienation must be : NN MD BE
 author must be : NN MD BE
 components must be : NNS MD BE
 columns must be : NNS MD BE
 it must be : PPS MD BE
 care must be : NN MD BE
 file must be : NN MD BE
 entry must be : NN MD BE
 statement must be : NN MD BE

DIOCS must be : NN MD BE
 plan must be : NN MD BE
 plan must be : NN MD BE
 plan must be : NN MD BE
 plan must be : NN MD BE
 principle must be : NN MD BE
 plan must be : NN MD BE
 it must be : PPS MD BE
 it must be : PPS MD BE
) must be :) MD BE
 capabilities must be : NNS MD BE
 values must be : NNS MD BE
 which must be : WDT MD BE
 which must be : WDT MD BE
 and must be : CC MD BE
 It must be : PPS MD BE
 states must be : NNS MD BE
 It must be : PPS MD BE
 compromise must be : NN MD BE
 platform must be : NN MD BE
 accelerometers must be : NNS MD BE
 axis must be : NN MD BE
 north must be : NN MD BE
 itself must be : PPL MD BE
 that must be : DT MD BE
 you must be : PPSS MD BE
 he must be : PPS MD BE
 , must be : , MD BE
 You must be : PPSS MD BE
 these must be : DTS MD BE
 who must be : WPS MD BE
 there must be : EX MD BE
 what must be : WDT MD BE
 lists must be : NNS MD BE
 this must be : DT MD BE
 Nancy must be : NP MD BE
 it must be : PPS MD BE
 it must be : PPS MD BE
 It must be : PPS MD BE
 it must be : PPS MD BE
 it must be : PPS MD BE
 it must be : PPS MD BE
 something must be : PN MD BE
 there must be : EX MD BE
 who must be : WPS MD BE
 and must be : CC MD BE
 She must be : PPS MD BE

```

    hers must be : PP$$ MD BE
    she must be : PPS MD BE
    What must be : WDT MD BE
    , must be : , MD BE
    he must be : PPS MD BE
    There must be : EX MD BE
    he must be : PPS MD BE
    marines must be : NNS MD BE
    she must be : PPS MD BE
    It must be : PPS MD BE
    They must be : PPSS MD BE
    now must be : RB MD BE
    what must be : WDT MD BE
    He must be : PPS MD BE
    house must be : NN MD BE
    it must be : PPS MD BE
    mate must be : NN MD BE
    Susan must be : NP MD BE
    money must be : NN MD BE
    which must be : WDT MD BE
    snapshots must be : NNS MD BE
    bodies must be : NNS MD BE
    It must be : PPS MD BE
}
'':{
    and must '' : CC MD ''
    '' must '' : '' NN ''
    a must '' : AT NN ''
}
CC:{
    ; must and : . MD CC
    . Must or : . MD CC
}
PPO:{
    Or must it : CC MD PPO
    Jan must you : NP MD PPO
}
ABN:{
    we must all : PPSS MD ABN
}
NP-HL:{
    . Must Berlin : . MD-HL NP-HL
}
DO:{
    we must do : PPSS MD DO
    government must do : NN MD DO
    and must do : CC MD DO
}

```

```

they must do : PPSS MD DO
they must do : PPSS MD DO
They must do : PPSS MD DO
we must do : PPSS MD DO
}
RB:{
You must also : PPSS MD RB
children must first : NNS MD RB
it must truthfully : PPS MD RB
end must somehow : NN MD RB
person must also : NN MD RB
, must now : , MD RB
three must indeed : CD MD RB
measure must inevitably : NN MD RB
discussed must surely : VBN MD RB
We must also : PPSS MD RB
One must first : PN MD RB
rodger must necessarily : NN MD RB
, must first : , MD RB
'' must also : '' MD RB
linguist must always : NN MD RB
attention must also : NN MD RB
jaws must necessarily : NNS MD RB
man must first : NN MD RB
what must always : WDT MD RB
we must never : PPSS MD RB
they must always : PPSS MD RB
There must first : EX MD RB
congressman must constantly : NN MD RB
we must certainly : PPSS MD RB
love-in-action must first : NN MD RB
There must also : EX MD RB
'' must surely : '' MD RB
what must still : WDT MD RB
perception must immediately : NN MD RB
religion must also : NN MD RB
ambitions must always : NNS MD RB
he must now : PPS MD RB
we must first : PPSS MD RB
we must squarely : PPSS MD RB
Salem must frequently : NP MD RB
'' must always : '' MD RB
I must then : PPSS MD RB
list must naturally : NN MD RB
bishops must never : NNS MD RB
Shakespeare must often : NP MD RB
historian must never : NN MD RB

```


history must always : NN MD RB
 merchants must constantly : NNS MD RB
 they must also : PPSS MD RB
 and must often : CC MD RB
 areas must necessarily : NNS MD RB
 judgment must therefore : NN MD RB
 he must also : PPS MD RB
 we must always : PPSS MD RB
 productivity must inevitably : NN MD RB
 president must also : NN MD RB
 plan must therefore : NN MD RB
 we must still : PPSS MD RB
 We must now : PPSS MD RB
 we must first : PPSS MD RB
 civilization must formally : NN MD RB
 succession must also : NN MD RB
 he must also : PPS MD RB
 which must therefore : WDT MD RB
 costs must somehow : NNS MD RB
 then must somehow : RB MD RB
 we must first : PPSS MD RB
 we must now : PPSS MD RB
 , must necessarily : , MD RB
 epithets must reasonably : NNS MD RB
 humorous must therefore : JJ MD RB
 variations must eventually : NNS MD RB
 operator must continually : NN MD RB
 He must mentally : PPS MD RB
 Music must now : NN-TL MD RB
 Purdew must once : NP MD RB
 Central must eventually : JJ-TL MD RB
 town must still : NN MD RB
 hair must still : NN MD RB

}
 NNS:{
 . Must records : . MD NNS
 }
 PPS:{
 neither must she : CC MD PPS
 , must she : , MD PPS
 }
 IN:{
 a must for : AT NN IN
 hearings must at : NNS MD IN
 a must for : AT NN IN
 machinery must in : NN MD IN
 They must in : PPSS MD IN

```

        we must at : PPSS MD IN
    }
    RB-HL:{
        minors must also : NNS-HL MD-HL RB-HL
    }
    PPLS:{
        we must ourselves : PPSS MD PPLS
    }
    VB:{
        county must sign : NN MD VB
        Republicans must hold : NPS MD VB
        We must solve : PPSS MD VB
        council must establish : NN MD VB
        ‘‘ Must share : ‘‘ MD VB
        Democrats must resolve : NPS MD VB
        We must keep : PPSS MD VB
        apparently must make : RB MD VB
        We must persuade : PPSS MD VB
        States must plan : NNS-TL MD VB
        You must play : PPSS MD VB
        it must accomplish : PPS MD VB
        we must stay : PPSS MD VB
        agency must train : NN MD VB
        We must justify : PPSS MD VB
        we must get : PPSS MD VB
        company must apply : NN MD VB
        you must wear : PPSS MD VB
        law must treat : NN MD VB
        equality must forbid : NN MD VB
        it must compete : PPS MD VB
        it must play : PPS MD VB
        libraries must face : NNS MD VB
        one must cherish : PN MD VB
        world must acquiesce : NN MD VB
        He must show : PPS MD VB
        it must deal : PPS MD VB
        labor must learn : NN MD VB
        , must develop : , MD VB
        we must confess : PPSS MD VB
        project must undergo : NN MD VB
        person must pass : NN MD VB
        community must stand : NN MD VB
        DuPont must divest : NP MD VB
        who must transfer : WPS MD VB
        ads must get : NNS MD VB
        who must take : WPS MD VB
        we must embark : PPSS MD VB
    }

```

escalation must end : NN MD VB
too must start : RB MD VB
States must choose : NNS-TL MD VB
question must arise : NN MD VB
yourself must seek : PPL MD VB
they must depend : PPSS MD VB
this must sound : DT MD VB
I must say : PPSS MD VB
song must contribute : NN MD VB
one must go : PN MD VB
one must build : PN MD VB
one must wonder : PN MD VB
prestige must lead : NN MD VB
leader must strengthen : NN MD VB
world must become : NN MD VB
we must number : PPSS MD VB
ourselves must strive : PPLS MD VB
man must recognize : NN MD VB
I must speak : PPSS MD VB
one must belong : PN MD VB
we must keep : PPSS MD VB
He must learn : PPS MD VB
Committee must follow : NN-TL MD VB
they must find : PPSS MD VB
it must rule : PPS MD VB
we must learn : PPSS MD VB
We must claim : PPSS MD VB
prominent must serve : JJ MD VB
thou must give : PPSS MD VB
we must use : PPSS MD VB
we must take : PPSS MD VB
We must determine : PPSS MD VB
We must make : PPSS MD VB
which must dominate : WDT MD VB
junior must win : NN MD VB
you must apply : PPSS MD VB
you must present : PPSS MD VB
you must file : PPSS MD VB
You must know : PPSS MD VB
they must study : PPSS MD VB
gymnast must develop : NN MD VB
that must come : WPS MD VB
man must revel : NN MD VB
he must understand : PPS MD VB
We must build : PPSS MD VB
he must select : PPS MD VB
They must learn : PPSS MD VB

He must undertake : PPS MD VB
 man must read : NN MD VB
 dealers must earn : NNS MD VB
 therapist must know : NN MD VB
 else must know : RB MD VB
 theory must seem : NN MD VB
 fiction must travel : NN MD VB
 care must take : NN MD VB
 Everything must balance : PN MD VB
 you must get : PPSS MD VB
 We must realize : PPSS MD VB
 vagina must dilate : NN MD VB
 generally must lie : RB MD VB
 husband must take : NN MD VB
 He must begin : PPS MD VB
 He must become : PPS MD VB
 he must spend : PPS MD VB
 sitter must decide : NN MD VB
) must mark :) MD VB
 farmer must get : NN MD VB
 he must work : PPS MD VB
 generally must pay : RB MD VB
 you must keep : PPSS MD VB
 settlement must depend : NN MD VB
 you must forgive : PPSS MD VB
 he must realize : PPS MD VB
 , must defer : , MD VB
 we must choose : PPSS MD VB
 we must fight : PPSS MD VB
 we must remember : PPSS MD VB
 who must deal : WPS MD VB
 man must attend : NN MD VB
 he must go : PPS MD VB
 growth must go : NN MD VB
 one must admit : PN MD VB
 education must start : NN MD VB
 ecumenical must meet : JJ MD VB
 ecumenical must become : JJ MD VB
 we must emphasize : PPSS MD VB
 something must seep : PN MD VB
 and must conform : CC MD VB
 , must follow : , MD VB
 and must serve : CC MD VB
 Indians must bring : NPS MD VB
 He must take : PPS MD VB
 we must become : PPSS MD VB
 Each must match : DT MD VB

three must compare : CD MD VB
 who must run : WPS MD VB
 nose must deny : NN MD VB
 Negro must discover : NP MD VB
 South must grapple : NR-TL MD VB
 farmer must perform : NN MD VB
 yearly must readjust : RB MD VB
 composer must select : NN MD VB
 I must reach : PPSS MD VB
 else must follow : RB MD VB
 metamorphosis must come : NN MD VB
 choreographer must arbitrate : NN MD VB
 He must rearrange : PPS MD VB
 He must construct : PPS MD VB
 that must fall : WPS MD VB
 one must let : PN MD VB
 one must speak : PN MD VB
 she must justify : PPS MD VB
 they must move : PPSS MD VB
 I must confess : PPSS MD VB
 ye must stagger : PPSS MD VB
 detective must save : NN MD VB
 he must abandon : PPS MD VB
 We must believe : PPSS MD VB
 we must employ : PPSS MD VB
 Businesses must develop : NNS MD VB
 they must become : PPSS MD VB
 government must establish : NN MD VB
 we must turn : PPSS MD VB
 it must grow : PPS MD VB
 , must make : , MD VB
 , must say : , MD VB
 I must stand : PPSS MD VB
 I must tell : PPSS MD VB
 I must make : PPSS MD VB
 we must watch : PPSS MD VB
 I must devote : PPSS MD VB
 nations must extend : NNS MD VB
 We must meet : PPSS MD VB
 we must induce : PPSS MD VB
 we must co-operate : PPSS MD VB
 we must act : PPSS MD VB
 we must maintain : PPSS MD VB
 makers must make : NNS MD VB
 whole must decide : NN MD VB
 this must stand : DT MD VB
 we must think : PPSS MD VB

it must sever : PPS MD VB
 we must admit : PPSS MD VB
 , must center : , MD VB
 it must seek : PPS MD VB
 we must remember : PPSS MD VB
 We must avoid : PPSS MD VB
 I must hurry : PPSS MD VB
 we must go : PPSS MD VB
 I must go : PPSS MD VB
 man must take : NN MD VB
 we must examine : PPSS MD VB
 theater must stand : NN MD VB
 , must create : , MD VB
 we must prove : PPSS MD VB
 we must construe : PPSS MD VB
 we must find : PPSS MD VB
 historian must consider : NN MD VB
 he must regard : PPS MD VB
 he must recognize : PPS MD VB
 he must expect : PPS MD VB
 we must add : PPSS MD VB
 , must set : , MD VB
 one must act : PN MD VB
 we must attend : PPSS MD VB
 we must carry : PPSS MD VB
 , must develop : , MD VB
 community must lead : NN MD VB
 we must develop : PPSS MD VB
 he must battle : PPS MD VB
 detective must rely : NN MD VB
 He must become : PPS MD VB
 and must meet : CC MD VB
 future must indicate : NN MD VB
 and must meet : CC MD VB
 it must express : PPS MD VB
 development must come : NN MD VB
 we must encourage : PPSS MD VB
 they must plan : PPSS MD VB
 world must rise : NN MD VB
 state must raise : NN MD VB
 , must give : , MD VB
 We must ask : PPSS MD VB
 enforcement must accommodate : NN MD VB
 Department must think : NN-TL MD VB
 we must think : PPSS MD VB
 we must act : PPSS MD VB
 he must give : PPS MD VB

He must think : PPS MD VB
 principles must take : NNS MD VB
 He must deal : PPS MD VB
 He must think : PPS MD VB
 objectives must conform : NNS MD VB
 readiness must take : NN MD VB
 you must place : PPSS MD VB
 representative must file : NN MD VB
 application must include : NN MD VB
 You must attach : PPSS MD VB
 he must file : PPS MD VB
 -- must remain : -- MD VB
 artist must cross : NN MD VB
 which must compete : WDT MD VB
 it must concentrate : PPS MD VB
 student must attend : NN MD VB
 president must involve : NN MD VB
 university must know : NN MD VB
 universities must consider : NNS MD VB
 , must account : , MD VB
 gas must come : NN MD VB
 , must sputter : , MD VB
 one must expect : PN MD VB
 One must include : PN MD VB
 I must add : PPSS MD VB
 I must plead : PPSS MD VB
 mechanism must exist : NN MD VB
 one must examine : PN MD VB
 one must know : PN MD VB
 one must know : PN MD VB
 trial must result : NN MD VB
 Af must cross : NN MD VB
 set must consist : NN MD VB
 Af must coincide : NN MD VB
 we must reject : PPSS MD VB
 They must continue : PPSS MD VB
 countries must formulate : NNS MD VB
 developments must take : NNS MD VB
 and must become : CC MD VB
 society must generate : NN MD VB
 worker must demonstrate : NN MD VB
 worker must communicate : NN MD VB
 knowledge must determine : NN MD VB
 analysis must take : NN MD VB
 we must define : PPSS MD VB
 which must support : WDT MD VB
 districts must pay : NNS MD VB

they must pay : PPSS MD VB
 opposition must persist : NN MD VB
 few must win : AP MD VB
 we must relate : PPSS MD VB
 profession must work : NN MD VB
 one must distinguish : PN MD VB
 we must distinguish : PPSS MD VB
 company must incur : NN MD VB
 we must keep : PPSS MD VB
 and must hope : CC MD VB
 I must postpone : PPSS MD VB
 we must consider : PPSS MD VB
 which must move : WDT MD VB
 one must work : PN MD VB
 there must remain : EX MD VB
 we must make : PPSS MD VB
 who must suffer : WPS MD VB
 whole must remain : NN MD VB
 he must improvise : PPS MD VB
 Pip must learn : NP MD VB
 operand must begin : NN MD VB
 exchange must result : NN MD VB
 targets must come : NNS MD VB
 preservation must accomplish : NN MD VB
 tests must determine : NNS MD VB
 stages must use : NNS MD VB
 decisions must constitute : NNS MD VB
 , must use : , MD VB
 we must wait : PPSS MD VB
 they must look : PPSS MD VB
 he must leave : PPS MD VB
 he must receive : PPS MD VB
 place must go : NN MD VB
 he must go : PPS MD VB
 he must follow : PPS MD VB
 he must pray : PPS MD VB
 You must make : PPSS MD VB
 he must seem : PPS MD VB
 I must pretend : PPSS MD VB
 I must state : PPSS MD VB
 city must rid : NN MD VB
 Geneva must remain : NP MD VB
 , must belong : , MD VB
 she must hold : PPS MD VB
 sculpture must take : NN MD VB
 peace must come : NN MD VB
 I must admit : PPSS MD VB

I must know : PPSS MD VB
 Kayabashi must feel : NP MD VB
 girls must take : NNS MD VB
 he must take : PPS MD VB
 who must take : WPS MD VB
 People must think : NNS MD VB
 one must make : PN MD VB
 it must link : PPS MD VB
 you must forgive : PPSS MD VB
 You must know : PPSS MD VB
 I must tell : PPSS MD VB
 I must tell : PPSS MD VB
 you must know : PPSS MD VB
 You must understand : PPSS MD VB
 we must know : PPSS MD VB
 he must express : PPS MD VB
 , must envy : , MD VB
 he must chance : PPS MD VB
 You must forgive : PPSS MD VB
 I must say : PPSS MD VB
 he must observe : PPS MD VB
 Everything must wait : PN MD VB
 ourselves must stand : PPLS MD VB
 he must act : PPS MD VB
 I must talk : PPSS MD VB
 but must stand : CC MD VB
 which must follow : WDT MD VB
 everyone must justify : PN MD VB
 you must flog : PPSS MD VB
 you must tell : PPSS MD VB
 God must look : NP MD VB
 I must remember : PPSS MD VB
 I must answer : PPSS MD VB
 ours must look : PP\$\$ MD VB

 }
 .:{

 you must . : PPSS MD .

 }
 *:{

 we must not : PPSS MD *
 Education must not : NN MD *
 We must not : PPSS MD *
 warnings must not : NNS MD *
 and must not : CC MD *
 one must not : PN MD *
 it must not : PPS MD *
 we must not : PPSS MD *

```

There must not : EX MD *
it must not : PPS MD *
We must not : PPSS MD *
living must not : VBG MD *
it must not : PPS MD *
citizens must not : NNS MD *
he must not : PPS MD *
He must not : PPS MD *
I must not : PPSS MD *
I must not : PPSS MD *
Israelites must not : NPS MD *
She must not : PPS MD *
we must not : PPSS MD *
}
HV:{
people must have : NNS MD HV
We must have : PPSS MD HV
coeds must have : NNS MD HV
they must have : PPSS MD HV
both must have : ABX MD HV
it must have : PPS MD HV
General must have : JJ-TL MD HV
we must have : PPSS MD HV
what must have : WDT MD HV
himself must have : PPL MD HV
start must have : NN MD HV
it must have : PPS MD HV
you must have : PPSS MD HV
some must have : DTI MD HV
Centrality must have : NN-TL MD HV
family must have : NN MD HV
it must have : PPS MD HV
we must have : PPSS MD HV
we must have : PPSS MD HV
them must have : PPO MD HV
it must have : PPS MD HV
reporter must have : NN MD HV
boards must have : NNS MD HV
you must have : PPSS MD HV
teacher must have : NN MD HV
course must have : NN MD HV
school must have : NN MD HV
she must have : PPS MD HV
it must have : PPS MD HV
what must have : WDT MD HV
, must have : , MD HV
This must have : DT MD HV

```

feels must have : VBZ MD HV
who must have : WPS MD HV
She must have : PPS MD HV
which must have : WDT MD HV
what must have : WDT MD HV
You must have : PPSS MD HV
imagery must have : NN MD HV
we must have : PPSS MD HV
men must have : NNS MD HV
men must have : NNS MD HV
they must have : PPSS MD HV
Bang-Jensen must have : NP MD HV
work must have : NN MD HV
he must have : PPS MD HV
dispatch must have : NN MD HV
silence must have : NN MD HV
she must have : PPS MD HV
, must have : , MD HV
magistrates must have : NNS MD HV
it must have : PPS MD HV
There must have : EX MD HV
troops must have : NNS MD HV
I must have : PPSS MD HV
town must have : NN MD HV
generation must have : NN MD HV
applicant must have : NN MD HV
Central must have : JJ-TL MD HV
entrance must have : NN MD HV
It must have : PPS MD HV
team must have : NN MD HV
which must have : WDT MD HV
snake must have : NN MD HV
space must have : NN MD HV
it must have : PPS MD HV
trials must have : NNS MD HV
Af must have : NN MD HV
locus must have : NN MD HV
**zg must have : NN MD HV
countries must have : NNS MD HV
there must have : EX MD HV
it must have : PPS MD HV
These must have : DTS MD HV
these must have : DTS MD HV
commander must have : NN MD HV
he must have : PPS MD HV
they must have : PPSS MD HV
you must have : PPSS MD HV

Juanita must have : NP MD HV
 Eileen must have : NP MD HV
 she must have : PPS MD HV
 thoughts must have : NNS MD HV
 she must have : PPS MD HV
 that must have : WPS MD HV
 it must have : PPS MD HV
 You must have : PPSS MD HV
 voice must have : NN MD HV
 soldiers must have : NNS MD HV
 cars must have : NNS MD HV
 and must have : CC MD HV
 , must have : , MD HV
 who must have : WPS MD HV
 . Must have : . MD HV
 He must have : PPS MD HV
 certainly must have : RB MD HV
 It must have : PPS MD HV
 skepticism must have : NN MD HV
 he must have : PPS MD HV
 himself must have : PPL MD HV
 we must have : PPSS MD HV
 and must have : CC MD HV
 It must have : PPS MD HV
 It must have : PPS MD HV
 Carmer must have : NP MD HV
 man must have : NN MD HV
 I must have : PPSS MD HV
 he must have : PPS MD HV
 He must have : PPS MD HV
 He must have : PPS MD HV
 Ruffians must have : NNS MD HV
 she must have : PPS MD HV
 He must have : PPS MD HV
 who must have : WPS MD HV
 she must have : PPS MD HV
 he must have : PPS MD HV
 he must have : PPS MD HV
 Via must have : NP MD HV
 She must have : PPS MD HV
 what must have : WDT MD HV
 You must have : PPSS MD HV
 he must have : PPS MD HV
 She must have : PPS MD HV
 John must have : NP MD HV
 You must have : PPSS MD HV
 and must have : CC MD HV

```

I must have : PPSS MD HV
She must have : PPS MD HV
she must have : PPS MD HV
He must have : PPS MD HV
. Must have : . MD HV
factor must have : NN MD HV
actor must have : NN MD HV
You must have : PPSS MD HV
you must have : PPSS MD HV
which must have : WDT MD HV
}

```

6 Learning to Classify Text

3. The Senseval 2 Corpus contains data intended to train word-sense disambiguation classifiers. It contains data for four words: hard, interest, line, and serve. Choose one of these four words, and load the corresponding data:

```

>>> from nltk.corpus import senseval
>>> instances = senseval.instances('hard.pos')
>>> size = int(len(instances) * 0.1)
>>> train_set, test_set = instances[size:], instances[:size]

```

Using this dataset, build a classifier that predicts the correct sense tag for a given instance. See the corpus HOWTO at <http://nltk.org/howto> for information on using the instance objects returned by the Senseval 2 Corpus.

Answer:

```

Training...
Testing...
Accuracy: 0.589

```

6. The synonyms *strong* and *powerful* pattern differently (try combining them with *chip* and *sales*). What features are relevant in this distinction? Build a classifier that predicts when each word should be used.

Answer: It may be obvious, but the surrounding words and their POS are used as features in this dataset.

```

Grabbing tagged Text...
Finding contexts...
Building datasets...
Training...
Accuracy: 0.930
Most Informative Features

```

right_tag = 'CS'	powerf : strong =
5.1 : 1.0	
left_word = 'and'	powerf : strong =
4.6 : 1.0	
left_tag = 'CC'	powerf : strong =
3.8 : 1.0	
left_word = 'The'	powerf : strong =
3.5 : 1.0	
right_word = ','	strong : powerf =
3.3 : 1.0	

Jurafsky and Martin

4 *N*-Grams

2. Write a program to compute unsmoothed unigrams and bigrams.

Answer:

```
590:DEBUG:Tokenizing...
2016-10-26 23:47:52,590:DEBUG:Loading text...
2016-10-26 23:47:52,590:DEBUG:Loading JSON as dictionary:'
    trumpTextRaw.json'
2016-10-26 23:47:55,020:DEBUG:Unigrams...
2016-10-26 23:47:55,163:DEBUG:Bigrams...
Num Words: 344325
[('Our',),
 ('campaign',),
 ('is',),
 ('powered',),
 ('by',),
 ('our',),
 ('love',),
 ('for',),
 ('this',),
 ('country',),
 ('and',),
 ('our',),
 ('love',),
 ('for',),
 ('our',),
 ('fellow',),
 ('citizens',),
 ('.',),
 ('And',),
 ('is',)]
```

```
[('Our', 'campaign'),
 ('campaign', 'is'),
 ('is', 'powered'),
 ('powered', 'by'),
 ('by', 'our'),
 ('our', 'love'),
 ('love', 'for'),
 ('for', 'this'),
 ('this', 'country'),
 ('country', 'and'),
 ('and', 'our'),
 ('our', 'love'),
 ('love', 'for'),
 ('for', 'our'),
 ('our', 'fellow'),
 ('fellow', 'citizens'),
 ('citizens', '.'),
 ('.', 'And'),
 ('And', 'is'),
 ('is', 'there')]
```

3. Run your *N*-gram program on two different small corpora of your choice (you might use email text or newsgroups). Now compare the statistics of the two corpora. What are the differences in the most common unigrams between the two? How about interesting differences in bigrams?

Answer:

```
25 Most Common for Austen's Emma      : [ (('to', 'be'), 595),
 (('of', 'the'), 557),
 (('in', 'the'), 434),
 (('I', 'am'), 395),
 (('had', 'been'), 308),
 (('Mr', 'Knightley'), 299),
 (('it', 'was'), 287),
 (('I', 'have'), 281),
 (('could', 'not'), 277),
 (('of', 'her'), 262),
 (('Mrs', 'Weston'), 256),
 (('she', 'had'), 254),
 (('--', 'and'), 245),
 (('have', 'been'), 243),
 (('to', 'the'), 236),
 (('Mr', 'Elton'), 229),
 ((',"', 'said'), 225),
 (('she', 'was'), 218),
 (('and', 'the'), 218),
 (('would', 'be'), 216),
 (('do', 'not'), 203),
```

```

(('it', 'is'), 194),
(('of', 'his'), 191),
(('a', 'very'), 185),
(('to', 'her'), 185)]
25 Most Common for Shakespear's Macbeth: [(('I', 'haue'), 41),
(('I', 'am'), 29),
(('of', 'the'), 24),
(('to', 'the'), 24),
(('my', 'Lord'), 20),
(('in', 'the'), 20),
(('Thane', 'of'), 18),
(('and', 'the'), 17),
(('ha', 's'), 17),
(('do', 's'), 15),
(('to', 'be'), 15),
(('Exeunt', 'Scena'), 15),
(('I', 'will'), 15),
(('Enter', 'Macbeth'), 15),
(('no', 'more'), 15),
(('of', 'my'), 15),
(('i', 'th'), 14),
(('of', 'Cawdor'), 14),
(('in', 'his'), 13),
(('Macb', 'I'), 13),
(('o', 'th'), 13),
(('Thane', 'of', 'Cawdor'), 13),
(('the', 'time'), 12),
(('What', 'is'), 12),
(('d', 'with'),

```

4. Add an option to your program to generate random sentences.

Answer:

```

[jmcgover@MAINFRAME 2lab]$ for i in {1..5}; do python lab2-jm4.4.
py ; done
It is so resolutely and Miss Taylor away with her own pleasure
the same, and till a simple style of Mrs.
Weston was such a longer -- Who cried Emma was prone to be heard
." He gave her immediate feeling was in half an act of poor
girl who were the only could catch myself, at -- tacitly
brought up her own wish, and his attentions and then, that Mr.
What was not the story, I come with the advantage of Mrs.
I am convinced that in fact by this sentiment which produced the
closest agreement with Mr.
I need not poignant enough to Donwell for her to you -- but she
replied, did not be subject as your forbearance.
[jmcgover@MAINFRAME 2lab]$

```


9. (Advanced) Suppose someone took all the words in a sentence and reordered them randomly. Write a program that takes as input such a **bag of words** and produces as output a guess at the original order. You will need to use an N -gram grammar produced by your N -gram program (on some corpus), and you will need to use the Viterbi algorithm introduced in the next chapter. This task is sometimes called **bag generation**.

Answer: