

Chapter 5

```
In [1]: import nltk
        from nltk import *
        from nltk.corpus import brown
        from collections import defaultdict
        from operator import itemgetter
```

1. What are the most common adverbs in the brown corpus (categories="news")? Please output the 10 most frequent ones (Please use the universal tagset).

```
In [24]: brown_news_tagged = brown.tagged_words(categories = 'news', tagset = 'universal')
        cfd_1 = nltk.ConditionalFreqDist((tag,word) for (word,tag) in brown_news_tagged)
        cfd_1['ADV'].most_common(10)
```

```
Out[24]: [('not', 254),
          ('when', 128),
          ('also', 120),
          ('now', 76),
          ('as', 75),
          ('here', 67),
          ('where', 58),
          ('then', 56),
          ('back', 55),
          ('about', 49)]
```

2. What are the part-of-speech tags before the word "religion" in the brown corpus (categories="religion")? (Please use the universal tagset)

```
In [36]: religion_tagged = brown.tagged_words(categories='religion', tagset='universal')
        rel_tags = [a[1] for (a,b) in nltk.bigrams(religion_tagged) if b[0]=='religion']
        fd_rel = nltk.FreqDist(rel_tags)
        fd_rel.tabulate()
```

NOUN	.	ADJ	ADP	CONJ	DET
3	2	2	1	1	1

3. What are the words that are highly ambiguous as to their part-of-speech tags ((i.e. the word has more than 3 pos tags) in the brown corpus (categories="adventure") (Please use the universal tagset).

```
In [13]: brown_adv_tagged = brown.tagged_words(categories = 'adventure', tagset = 'universal')
        data_3 = nltk.ConditionalFreqDist((word.lower(), tag)
                                           for (word,tag) in brown_adv_tagged)

        for word in sorted(data_3.conditions()):
            if len(data_3[word]) > 3:
                tags = [tag for (tag, _) in data_3[word].most_common()]
                print(word, ' '.join(tags))
```

```
back ADV NOUN VERB ADJ
last ADJ ADV VERB NOUN
outside ADV ADP ADJ NOUN
past ADP ADV ADJ NOUN
that ADP DET PRON ADV
```

4. Train a unigram tagger on the brown corpus (categories="humor"). a) Split the data into training and testing dataset- training on the 95% of data and testing on the remaining 5%. b) Evaluate the performance of this tagger on the test dataset. c) Use this tagger to tag some new text ['this','is','a','NLP','class']. d) Observe that some words are not assigned a tag. Explain why not? (Please do not use the universal tagset)

```
In [27]: #a)
```

```

brown_humor_tagged_sents = brown.tagged_sents(categories = 'humor', tagset = 'universal')
size_4 = int(len(brown_humor_tagged_sents) *.95)
train_sents_4 = brown_humor_tagged_sents[:size_4]
test_sents_4 = brown_humor_tagged_sents[size_4:]
unigram_tagger = nltk.UnigramTagger(train_sents_4)
#b)
unigram_tagger.evaluate(test_sents_4)

```

Out[27]: 0.7271062271062271

```

In [62]: string_4 = 'this is a string diddly doo da'
tokens_4 = word_tokenize(string_4)
(unigram_tagger.tag(tokens_4))

```

Out[62]:

```

[('this', 'DET'),
 ('is', 'VERB'),
 ('a', 'DET'),
 ('string', None),
 ('diddly', None),
 ('doo', None),
 ('da', None)]

```

Some are given a None tag because the word was not encountered in the training data set

5. Explore the nps_chat corpus and find out what part-of-speech tags occur before a noun, with the most frequent ones first (Please use the universal tagset).

```

In [14]: chat_tagged_words = nltk.corpus.nps_chat.tagged_words(tagset = 'universal')
chat_tag_pairs=nltk.bigrams(chat_tagged_words)
noun_preceders = [a[1] for (a, b) in chat_tag_pairs if b[1] == 'NOUN']
fdist = nltk.FreqDist(noun_preceders)
[tag for (tag, _) in fdist.most_common()]

```

Out[14]:

```

['X',
 'DET',
 'NOUN',
 'VERB',
 'ADJ',
 'PRON',
 '.',
 'ADP',
 'CONJ',
 'NUM',
 'ADV',
 'PRT']

```

6. Explore the brown corpus (categories="romance") to find out all tags starting with VB and its associated (word, frequency) pairs (no more than 6 pairs). (Please do not use the universal tagset)

For example, one of the outputs should look like:

VBG [('going', 59), ('looking', 36), ('trying', 23), ('thinking', 21), ('watching', 20), ('taking', 19)]

```

In [16]: def findtags(tag_prefix, tagged_text):
          cfd = nltk.ConditionalFreqDist((tag, word) for (word, tag) in tagged_text
          if tag.startswith(tag_prefix))
          return dict((tag, cfd[tag].most_common(6)) for tag in cfd.conditions())

```

```

In [17]: tagdict = findtags('VB', nltk.corpus.brown.tagged_words(categories='romance'))

```

```

In [18]: for tag in sorted(tagdict):

```

```
print(tag, tagdict[tag])
```

```
VB [('get', 92), ('know', 88), ('go', 76), ('see', 74), ('take', 62), ('say', 59)]
VB+PO [('"Let's"', 10), ('let's', 5)]
VBD [('said', 318), ('went', 82), ('thought', 80), ('came', 75), ('knew', 69), ('looked', 68)]
VBG [('going', 59), ('looking', 36), ('trying', 23), ('thinking', 21), ('watching', 20), ('taking', 19)]
VBG+TO [('gonna', 4)]
VBG-TL [('"Racin"', 1), ('Dancing', 1), ('Surviving', 1)]
VBN [('got', 36), ('come', 29), ('done', 29), ('gone', 25), ('seen', 20), ('made', 20)]
VBN+TO [('gotta', 1)]
VBN-TL [('"United"', 3), ('Armed', 1), ('Forked', 1)]
VBZ [('says', 7), ('wants', 7), ('goes', 5), ('gets', 4), ('thinks', 4), ('makes', 4)]
```

7. Write programs to process the brown corpus (categories="editorial") and find answers to the following questions (Please do not use the universal tagset):

a. Which nouns are more common in their plural form (e.g. tag='NNS'), rather than their singular form (e.g. tag='NN')? (Only consider regular plurals, formed with the -s suffix.)

b. What do the 10 most frequent tags represent in the Brown Corpus? Please output the tags and explain the meaning for each tag.

```
In [10]: #A
brown_ed_tagged = brown.tagged_words(categories = 'editorial')
cfd_7 = nltk.ConditionalFreqDist(brown_ed_tagged)
```

```
In [11]: conditions = cfd_7.conditions()
```

```
In [13]: for condition in conditions:
          if cfd_7[condition]['NNS'] > cfd_7[condition]['NN']:
              print(condition)
```

```
schools
appropriations
powers
procedures
legislators
programs
ways
musts
years
bills
pensions
roads
municipalities
limits
problems
fireworks
sales
questions
resolutions
minds
answers
heroics
stands
contributions
women
affairs
assemblies
judges
governors
wolves
Newspapermen
politicians
workers
pros
cons
```

issues
Cities
counties
months
eyes
members
\$83,750
\$30,000
neighbors
results
aggressions
forces
wars
shows
jobs
efforts
expenditures
services
taxes
commissioners
officials
difficulties
signs
imperialists
backs
dollars
forecasts
demands
\$1,750,000
\$1,250,000
conditions
needs
employes
meals
attendants
hours
nurses
doctors
Patients
funds
\$3.15
\$4
experts
\$9
reforms
areas
men
means
people
scores
deaths
schemes
leaders
countries
arms
things
facilities
jails
prisoners
lobbies
relations
confines
lines
slogans
institutions
buses
plans
projects
stations
barricades

tanks
zones
troops
effects
laws
offices
governments
notes
taunts
replies
authorities
remains
facts
fusillades
shots
residents
incidents
misgivings
tactics
indignities
Thousands
canyons
millions
seas
valleys
specimens
folds
Grasslands
settlers
barriers
bars
commercial
polls
ads
firms
groups
codes
grounds
competitors
soft-drinks
candidates
posts
points
opponents
citizens
ethics
policies
links
ills
shortcomings
supporters
regulars
calls
bosses
unions
sticks
marines
arrangements
dealings
bluffs
alternatives
Casualties
numbers
fighters
allies
planners
negotiations
observers
measures
prospects

communications
states
\$1,000,000
buildings
communities
sponsors
expenses
ones
partisans
qualities
secrets
others
audiences
dramas
hands
flourishes
generations
presidents
attacks
campaigns
airmen
friends
murderers
bodies
mercenaries
words
atrocities
savages
victims
weeks
rapes
mutinies
adventures
occurrences
terms
nations
telegraphers
personnel
situations
industries
delegates
internationalists
lads
writers
tons
heads
oafs
dimensions
times
kicks
Fools
shoes
domes
conventions
documents
contents
spokesmen
moralities
papers
planes
interests
methods
rations
railroads
rates
costs
truckers
grips
talks
bullets

rights
triumphs
conflicts
solutions
outbreaks
beliefs
developments
moves
tasks
days
\$15.5
\$500,000
pennies
ventures
\$8.5
levels
enterprises
accounts
paragraphs
complaints
charges
Problems
cases
tests
parties
seats
changes
trends
approaches
votes
gains
joiners
organizations
churches
accusations
statesmen
troubles
principles
instances
extremists
confusions
apprehensions
fleas
islands
owners
developers
hotels
sands
doors
objections
deliberations
committees
commissions
statutes
boards
councils
agencies
actions
servants
sessions
guarantees
lawmakers
loopholes
reservations
rationalizations
operations
aims
decisions
elections
warships

elements
crowds
lives
pressures
factories
disputes
feuds
surpluses
successes
taxpayers
supports
growers
farmers
rows
\$1.1
conclusions
layers
mistakes
assistants
subordinates
lists
statistics
orders
goods
strikes
odds
marauders
convictions
forecasters
Weeks
colleagues
figures
stores
payments
claims
benefits
miles
spaces
acres
herons
shorelines
preserves
cattle
egrets
hopes
places
dunes
applicants
recruits
youths
parents
findings
parts
plants
specialists
sides
youngsters
reformers
primaries
headquarters
badges
sachems
marks
cartoonists
stripes
steps
refugees
associates
agents
prerogatives

reports
matters
reporters
weapons
Ex-Presidents
discourses
obligations
Reports
loans
grants
\$50,400,000
\$84,000,000
passengers
fares
flights
airlines
profits
cruises
chances
remarks
resources
guerrillas
activities
rebels
floods
ends
soldiers
missions
divisions
\$1,450,000,000
recommendations
lawyers
fees
individuals
corporations
controversies
axes
talents
responsibilities
instructions
deputies
duties
advisers
beings
budgets
departments
respects
franchises
leases
functions
items
two-thirds
Highways
sewers
transfers
improvements
promises
injustices
publishers
editors
correspondents
decades
imbalances
freedoms
winds
clues
dangers
children
provinces
cities

views
colleges
hypocrisies
slums
streets
hundreds
thousands
villains
opportunities
employers
students
implications
warnings
positions
Representatives
constituents
stamps
names
addresses
newsletters
speeches
postmen
loads
letters
companies
seeds
vegetables
cans
rivers
fields
\$3,000
checks
beans
potatoes
tomatoes
couples
feet
grudges
trespasses
disciples
debts
requirements
sins
squeals
girls
lovers
jitters
signposts
relationships
wonders
periods
Jokes
cartoons
cynics
mothers-in-law
endowments
parishioners
believers
riches
standards
congregations
conceptions
satisfactions
bucks
stresses
tries
saints
sinners
ages
pieces

hooks
motions
voices
asters
reasons
sports
pains
butterflies
minutes
bidders
lots
crickets
pictures
classrooms
examples
processes
whites
threats
grades
schedules
restaurants
counters
establishments
extensions
Hotels
bombings
neighborhoods
depots
causes
businessmen
objects
freeways
m.p.h.
cars
choices
lanes
intentions
natives
drivers
occasions
managers
spots
Birds
rookies
grabs
circumstances
clubs
runs
indications
seasons
minors
platoons
fielders
experiences
mothers
comments
fingers
plays
Parents
rituals
colors
trees
traits
crayons
suggestions
circles
rooms
versions
stems
forms

patterns
imports
covers
chairs
plates
greens
panels
homes
shops
handicrafts
evenings
wraps
clubrooms
backyards
dissensions
ranks
officeholders
Plans
Blueprints
Arrangements
proposals
districts
missiles
chairmen
volunteers
boundaries
units
favorites
contests
hopefuls
races
organizers
incumbents
aspirants
strategists
echoes
committeemen
directors
representatives
works
shapes
sizes
motives
enemies
crimes
excerpts
mouthpieces
munitions
dealers
shares
prices
Quotations
exercises
predecessors
quarters
armaments
partners
listeners
readers
attempts
chestnuts
currents
opinions
rockets
suspensions
columns
essays
weeklies
fortunes
\$1,000

subscribers
\$1
continents
dreams
staircases
grandmothers
glasses
hoosegows
Inmates
holidays
occupants
defendants
convicts
officers
locks
inmates
escapees
brothers
operators
concessions
loyalties
\$2,300,000
\$750,000
\$590,000
Apologies
carpets
archives
conferences
breakfasts
dinners
messages
congratulations
stories
translations
scholars
inaccuracies
manuscripts
revisions
languages
studies
readings
passages
phrases
verses
heavens
teachings
sentences
persons
blessings
chapters
followers
Discoveries
writings
fathers
temperatures
pavements
types
leathers
styles
perforations
textures
Styles
toes
Heels
heels
semi-heights
pumps
hues
fabrics
throats

straws
weaves
sandals
tones
pastels
touches
Spectators
Flats
teeth
buttons
brushes
dentists
crevices
margins
anterior
bristles
gums
patients
disorders
muscles
tendons
bones
baths
pores
glands
possibilities
discs
tastes
records
games
tracks
conversations
sages
albums
presentations
recordings
vocabularies
Thanks
buffs
poets
releases
guides
upsets
fifties
complexes
brains
frauds
phonies
earnings
\$2,000,000
losses
Investors
People
suckers
achievements
holes
obstacles
explanations
desires
temptations
knick-knacks
nos
substitutes
magazines
books
bounds
Toys
consequences
snatches
Actions

universities
scenes
lions
uncertainties
homecomings
paths
leaves
moments
undergraduates
faces
glances
bifocals
legs
yards
persimmons
classmates
families
professors
hairs
eyeglasses
pants
sleeves
kick-offs
tackles
populations
majorities
gentlemen
towns
Bridges
errors
1930s
Farmers
trucks
carts
Smokers
makings
allegiances
invitations
spenders
contracts
oaths
visions
debates
qualifications
Congressmen
boats
stages
locations
migrants
sites
exemptions
\$3
Governors
newspapers
additions
accommodations
editorials
sums
Feelers
pupils
volumes
concepts
courts
thanks
classes
exceptions
Words
kinds
veterans
disabilities

hospitals
arguments
wishes
injuries
critics
passes
feelings
shelters
scientists
shorts
ladies
lessons
minorities
riders
policemen
firemen
week-ends
cutters
freezers
wives
husbands
chores
customers
markets
supermarkets
Judges
proprietors
premises
charities
teachers
articles
entertainers
graduates
salaries
\$550
\$3.22
\$580
stockholders
holdings
Copies
judgments
considerations
drinks
regulations
bonds
regions
loaders
colonies
cohorts
oppressors
burdens
trustees
freeholders
moneys
purposes
rules
properties
trades
muggers
models
booths
viewers
stumps
sediments
grains
spores
microfossils
dinosaurs
survivors
apartments

dogs
burglars
landlords
cafes
terraces
tourists
benches
fountains
gardens
dissents
statements
bombs
fall-outs
ducks
lands
facets
assessments
arrears
acts
tens
lights
trains
crafts
Visitors
elephants
collections
branches
ordinances
features
hymns
subjects
adults
Writers
automobiles
highways
kids
dropouts
wrongs
rocks
merits
dictates
flowers
insignificances
tunnels
ferries
crossings
curves
bridges
dictators
birds
sentiments
Starlings
blackbirds
gangs
delinquents
thugs
rapists
subversives
creatures
masses
chimes
calisthenics
hearts
ideals
drifts
scandals
hearings
details
targets
predictions

soothsayers
megatons
contestants
directions
guardians
ranges
footsteps
satellites
paeans
publicists
departures
applications
arrivals
blocks
Tanks
materials
commentators
punishments
morticians
exhumations
rehabilitations
misdeeds
neutralists
squabbles
compromises
precedents
pessimists
mentalities
connotations
employees
invaders
uniforms
carbines
pistols
rifles
accessories
guns
traitors
gods
phases
careers
priests
media
awnings
guests
ambassadors
practices
adherents
appeals
clouds
blasts
reactions
devices
balloons
channels
depths
substances
particles
bets
defects
stillbirths
diseases
generalizations
aspects
doses
announcements
sidewalks
Things
realities
dilemmas

'50's
ecumenicists
responses
theologians
estimates
provisions
projections
crises
factors
intransigents
apologies
Sessions
participants
Speakers
systems
differences
insights
discussions
thoughts
visitors
sympathies
relatives
generals
admirals
cousins
aunts
nephews
privations
idiocies
scars
frontiers
gestures
rags
emergencies
blacks
auspices
events
peaks
cycles
\$133
ideas
episodes
unoriginals
premieres
adaptations
originals
musicals
novels
sources
honors
farces
movies
Originals
Adaptations
defenders
adapters
masters
producers
classics
playwrights
nightclubs
joints
saloons
Customers
performances
graces
orchestras
guitars
violins
instruments

musicians
dresses
shifts
mechanics
shoulders
ripples
fingertips
speeds
strings
greetings
dancers
improvisations
bends
crawls
tips
costumes
\$150
\$200
veils
origins
picnics
miniatures
ankles
seamen
ships
sweaters
handkerchiefs
sailors
painters
paintings
landscapes
cityscapes
drawings
tantrums
impressionists
inhibitions
erotica
suits
collars
butts
gutters
portraits
nudes
self-portraits
ears
journals
formulas
voters
farms
peasants
plots
admissions
doubts
surprises
rents
quarrels
whipping-boys
symptoms
prophecies
suspicions
Indications
anti-Communists
maneuvers
helmets
sheriffs
shells
trails
paces
grenades
rounds

casualties
buyers
increases
sectors
ramifications
nationalisms
rivalries
customs
wounds
libertarians
pages
disruptions
deferments
brigades
corners
taverns
broadcasts
1770's
foreigners
caps
insurgents
tractors
Congratulations
Reprints
journalists
liabilities
restrictions
networks
tapes
transcripts
re-runs
shades
techniques
instructors
practitioners
courses
academies
commentaries
lectures
falsehoods
historians

```
In [14]: #B
tags_7 = [tag for word, tag in brown_ed_tagged]
```

```
In [15]: fd_7 = nltk.FreqDist(tags_7)
```

```
In [18]: common_tags = fd_7.most_common(10)
```

```
In [19]: common_tags
```

```
Out[19]: [('NN', 7675),
          ('IN', 6204),
          ('AT', 5311),
          ('JJ', 3593),
          ('.', 2988),
          ('NNS', 2972),
          (',', 2741),
          ('VB', 2129),
          ('NP', 1884),
          ('CC', 1835)]
```

NN - Noun, common, singular or mass

IN - Preposition or conjunction, subordinating

AT - Article

JJ - Adjective or numeral, ordinal

. - Period

NNS - Noun, common, plural

, - Comma

VB - Verb

NP - Proper Noun

CC - Conjunction, coordinating

8. Write code to search the brown corpus (categories="hobbies") for particular words and phrases according to tags, to answer the following questions (please do not use the universal tagset):

a. Produce an alphabetically sorted list of the distinct words tagged as MD.

b. Identify three-word prepositional phrases of the form IN + AT + NN (eg. in the lab).

```
In [45]: brown_hobbies_tagged_words = brown.tagged_words(categories = 'hobbies')
hobbies_tag_fd = nltk.FreqDist(brown_hobbies_tagged_words)
sorted(set([wt[0].lower() for (wt, _) in hobbies_tag_fd.most_common() if wt[1] == 'MD']))
```

```
Out[45]: ['can',
'could',
'dare',
'may',
'might',
'must',
'need',
'shall',
'should',
'will',
'would']
```

```
In [25]: brown_hobbies_tagged_sents = brown.tagged_sents(categories = 'hobbies')
def process(sentence):
    for (w1,t1), (w2,t2), (w3,t3) in nltk.trigrams(sentence):
        if (t1 == ('IN') and t2 == 'AT' and t3 == 'NN'):
            print(w1,w2,w3)

for tagged_sent in brown_hobbies_tagged_sents:
    process(tagged_sent)
```

```
at the time
from the sterno-cleido
of the neck
of the leg
to the value
of a muscle
of the wide-grip
on the chest
with the barbell
in the pecs
to the Aj
of the pin-point
with a bit
to the pecs
into the serratus
of every muscle
with the knowledge
in the limbo
```

at the hipline
of the champion
under the skin
on the leg
for the bodybuilder
to the height
against the back
to the rear
to the front
in a nutshell
at the back
of the neck
with the bar
of the neck
to the front
from a pansy
from a dealer
in a week
of the year
with a fog
with a board
with the plant
over the bed
of a compost
to the color
through the winter
in a flat
but a mat
over the glass
in a border
at the expense
to the earth
for the rest
of the season
from the mother
at a time
over the winter
in the year
to a meal
in the light
throughout the world
to the season
in the world
on the tree
into a sort
of the fruit
of the avocado
to the consumer
of the blood
at the pit
of a weapon
With a nation
to the advantage
to the missile
in the air
against the target
for the flight
to the target
with a load
to the conflict
in the future
to the hunter-killer
by the missile
of the missile
on the ground
on the ground
on the ground
by a factor
in the air
for the majority

by the missile
at the expense
for the aircraft
for the submarine
with a sense
on an enemy
in the case
in the dark
of a missile
in the air
of the missile
for a missile
with the chance
of the target
into the category
with the contrabass
in the background
of a bull
in a china
in the course
at the outset
by the individuality
into the spirit
to the performance
with the disc
with a group
by the magnificence
by no means
by the fact
From the beginning
of the original
On the basis
to the degree
on the basis
than a copy
of the recording
with the frequency
to the score
on the intention
of the symphony
of the scherzo
of the finale
In the end
to a melody
to the depth
of the orchestration
into the ensemble
in the score
in the history
at the show
from the dog
for a number
of the ring
into the ring
in the ring
on the subject
at the show
around the ring
over the public-address
for the year
from the group
in the center
of the ring
of the year
for the founder-originator
throughout the year
in the horse
in the horse
in the dog
in a dog

in the dog
for a room
of the way
in the room
In the morning
to the fact
near a body
in every state
of the government
over the country
to the water
to a lake
into the car
to the water
to the boat
to the gear
at the door
at the waterside
around the block
in the world
from a boat
of the country
of a boat
of the automobile
to the state
across the boundary
of the state
for the numbering
for a number
with no power
regarding the age
of the law
on the water
of the sport
in the success
of the safety
for the bureau
for an industry
for the boatman
In the frame
behind the plane
to the side
to a train
with no chance
to the tappet
for the prototype
at the model
for a junction
with the right-hand
from an end
with a Af
to the 10''
to the frame
with a Af
with a file
in the frame
for the spring
of the frame
in the insert
in the frame
into the hole
with the underside
of the bar
into the frame
against the frame
across the bar
of the tappet
from the frame
through the tappet
at the position

of the tappet
at the bottom
in the frame
of the bar
into a drill
with the bar
above the bar
through the center
of the pin
across the face
for the shape
of the notch
on the bar
in a hand
with a carborundum
on a 1/8''
in the frame
in a drill
with a lathe
in the problem
in the formula
in a row
of the order
of the symbol
from the number
of the symbol
for the problem
by the +
of the equation
in the equation
of the equation
of the equation
with a conversion
on the accuracy
in the formula
of the sport
in the displacement
of the piston
in the engine
of a cylinder
in the engine
by the stroke
of the length
by the number
in the engine
at the answer
of the group
with the total
of the cylinder
with a liquid
for the displacement
of the combustion
in the cylinder
in the chamber
into the chamber
above the surface
around the chamber
of the chamber
into the chamber
in the beaker
of the cylinder
with the formula
in the formula
of the cylinder
in the head
of an inch
of the opening
of the opening
to the combustion
of the area

between the top
of the piston
of the cylinder
from the top
of the block
to the piston
with a depth
of the cylinder
by the depth
to the total
of the combustion
For an example
to the formula
on the piston
by the piston
to the cylinder
of the formula
of a wheel
from the surface
to the center
of an inch
in the measurement
to a decimal
in the formula
For an engine
for a combination
for a couple
on the twice-around
at the center
of a colt
on the trot
around the half-mile
of the stable
with a mile
on the half-mile
in the stable
at the start
with the master
on the half-mile
during the week
during the season
of the attention
at the track
for the week
on the lookout
in a caliber
in a rifle
of the group
on an opening-day
through a rhododendron
through a sapling
of the whitetail
with the ease
with a mid-range
into a flurry
into a small-game
in the point
with every shot
to the harvest
for an auto
with an insert
with a smoothbore
to the barrel
on the clay
from the spring
with the addition
to the list
by the popularity
with a short-barrel
in a saddle

of the slide-lock
for a youngster
for a saddle
for a boy
with a side
than a top-tang
to the field
in the fog
of the table
through the redwood
along the coast
of the tallyho
in the morning
of the tariff
but the tobacco
in the mountain
at the scenery
on the box
over a spirit
for a group
on the morning
with a captain
of the period
on the stove
with a tray
on a bed
for the day
on the parsley
for the underbedding
After the meal
in the reflection
to the jaw
for the start
on the bill
with a cupful
with a tablespoonful
of a magnate
on the maelstrom
at the entrance
of the court
in the tallyho
at a curve
by the force
of a gale
down the peninsula
of the region
on the road
of the tallyho
on the dashboard
at a clump
of the road
with a list
to the end
into a saucepan
at the end
of a row
in the cellar
with the agility
of the pie
of the oven
from the baker
in the pie
In a saucepan
on the basis
of the area
of the nation
of the city
with a week
for a visit
for a whirlwind

to the public
of a second
for the tourist
of the war
at every turn
of the profusion
to the type
of the subject
on the side
of a mountain
For a resort
from the sky
of the area
with a visit
in the world
At the moment
in the world
with the development
from the center
of the city
over the entrance
Across the bridge
from the entrance
to the bridge
At the top
of the hill
to a park
in a moment
to the entrance
on the site
to a mosque
With the birth
Beneath the dome
through the flower
Across the street
At the beginning
with a glass
to the mosque
on a pedestal
of the emperor
Near the end
of a monument
with a gallery
in the middle
around the gallery
to the streetcar
up the hill
at the top
in the middle
on the car
to the university
on the side
opposite the gate
of the city
in the arcade
of the university
down the street
beyond the corner
with a red-tile
past the entrance
at the end
toward a gate
in the wall
in the wall
along a road
in the distance
with a tap
in the wall
through the gate
on the wall

In the court
through a door
to the building
at the end
through a garden
at the back
of the back
of the barbecue
of the hood
by the use
about a steak
on the family
of the refrigerator
without a dozen
into the kitchen
in the world
above the fire
to the marina
In the winter
in the kitchen
of the fireplace
in the winter
on a boat
of a meal
on the road
in the beginning
in a restaurant
at a restaurant
in the car
on a skewer
upon a compromise
over a bonfire
on the beach
of the way
to the success
of the frank
in the crowd
to the picnic
of the juice
to a boil
for the frankfurter
to a can
with a mixture
by a foot
on the underside
for the texture
under the glaze
in the glaze
For an antique
into a cylinder
on a disk
with a coil
with the clay
at the back
of a sock
near the stern
into a turn
on the outside
over the boat
of the motor
in a couple
for the hull
on a table
for the keelson
for the keelson
in the boat
with the plywood
in the hull
on the jig
below the line

of the keelson
down the center
of the thickness
on the jig
of the keelson
to the brace
of the keelson
of the keelson
on the jig
With the exception
to the stem
to the base
of the planer
with a pair
with a weight
in the fastening
on the inside
with the transom
at the contact
on the outside
from the inside
to the spray
of the bottom
at a time
with a pigment
of the bottom
with the resin
with a paint
of the cloth
at the keel
near the spray
along the keel
from the transom
to the point
of the stem
along the centerline
with the help
on the inside
of the stem
of every frame
to the jig
of the side
of the motor
from the bottom
to the top
to the transom
to the transom
to the flooring
to the floor
in the cabin
of the port
above the bunk
into the head
on the bulkhead
above the sink
to the wood
in every detail
by the architect
of the way
with the family
of the handyman
for a minimum
to the site
in a weekend
in a matter
to the fact
throughout the country
in the use
in a site
on the property

after a thaw
for the shell
of the house
for the shell
of a framework
on the quality
to the frame
in a mastic
in the detail
for the gable
for the glass
of the press
between the spindle
to the center
of the work
to the floor
with a piece
of the table
of the column
in the lubrication
in the quill
of the press
on the motor
with a pair
for the chuck
on the press
to a length
to the bottom
of the motor
of the way
on the material
on the step-cone
in the instruction
with the tool
of the tool
to the tool
of the way
through a workpiece
for the hole
of the wood
on the side
of the stock
with the mark
in the order
for an act
from the Land
of the year
for an Act
into a Body
to the town
under the direction
by an island
between the island
of an arch
above the high-water
for the invention
to a builder
of the arch
for the statement
in an engraving
with a draw
in a tavern
by the corporation
on the island
on the table
of the occasion
on the bridge
of the evening
in the beginning
by a chain

of the work
to the sale
than a century
down the steep
across the bridge
in the war
in the gasoline
than a relic
of the horse
of the war
of the bridge
against the competition
During the war
by the wagon
from the railhead
in the interior
from the rain
across the bridge
up the valley
in the case
of a pool
near the house
from the house
in a field
on a hill
in the picture
of the pool
in the forest
to the well-being
of a pool
in the ground
in the pool
for a pool
by the shape
of the house
around the pool
by the vision
around the pool
with a buffet
for the lord
on a state
by the pool
in the family
without an invitation
to the pool
to an island
of the pool
near the pool
between the pool
onto a pool
for the benefit
to the water
from the pool
to the house
from the pool
of the filter
in the water
in the fall
of the dirt
on the water
in a tree
to the letter
in the family
of the job
with no chance
in the know
across the country
at the latitude
in a mortgage
with a minimum

in a day
in the size
in the cost
of the house
in a closet
From the coil
in the yard
in a mild-winter
above the cost
in the basement
in the attic
to a point
than the price
on a variety
besides the nature
from the outside
of a conditioner
in an hour
to the cooling
for the horsepower
of the compressor
of the unit
With a unit
on the outside
of the house
in the roof
of the house
of a gas
to the moisture
in the ceiling
in the side
to a minimum
in the installation
on the basis
above a bedroom
of a site
through the work
of the site
from the county
in the field
during the time
of the year
of the climate
to the sun
in the field
at the site
at the office
in the area
regarding the site
to the site
by a group
on the character
of the site
of the investigator
of the area
for the future
to the public
of the recreation
at the site
on a body
for a park
of a valley
on the coast
of the surf
on a beach
through the beauty
for the public
for a recreation
on an ocean
of the water

in the springtime
on a reservoir
of the water
of the water
to the site
to the park
at the site
at the site
for a check
of the species
of the fish
at the site
about the area
of the period
to the recreation
of the land
of the area
of the project
for the use
of the area
of the minimum
for the future
of the area
to the use
of the development
within the park
of the site
by the appearance
of the land
in the cost
to the site
of the ground
for a well
to the use
of the area
for the development
of a site
of the area
for the use
of the site
in the foyer
of a composer
for the cause
of the stage
by the force
of the lure
into a stylist
by the life
by the compass
of a house
for a home
In the process
to the World
of the effectiveness
during the depression
in the stock
of the out-of-doors
in the field
in the stream
on the wing
of a kind
to the painting
On the death
from the shop
with the painter
in the vicinity
of a rigger
to the design
in the scene
in the painting

of the composition
to the subject
in a mat
on the sketch
in the field
to the watercolor
on a watercolor
in the field
at a tilt-top
for the watercolorist
in the studio
against the sky
for the impression
for a watercolorist
for a watercolor
in the oil
with a somersault
in the school
of the body
at the age
on the side
on a program
for the team
for the gymnastics
for the job
through the kind
considering the amount
over a door
on the chair
at the top
of the chin
with the Horse
of the room
on the floor
to a point
to the stand
at the top
of the push-up
for the letting
with the class
in a circle
into the circle
to the floor
on the floor
at a time
of the head
at the floor
toward the ceiling
with the handstand
on the back
of the chest
At the peak
of the arch
in an effort
on a mat
of the head
on the chest
into the air
into the air
of the pelvis
off the neck
for the defense
in the form
in the application
From the wealth
with a color-TV
through the specimen
by the microscope
in the TV
of a second

for the electronics
of the specimen
into a color
to the study
into a mouthpiece
in the heartbeat
of the flow
through the aorta
of the heart
of the heart
around the heart
of the heart
of the heart
to the heart
into the blood
of the heart
of the heart
into the bloodstream
by the bloodstream
of the velocity
of the blood
at the end
of a catheter
in the heart
of the instrumentation
by the patient
around the patient
after the burst
by the FM
in the patient
in the exploration
of the retina
to the body
through a water
of an Af
with a description
with an epoxy
in a sector
of the eye
by the camera
in a tub
of the neck
of the spine
in the center
on the island
in a town
by a year
to the home
From the town
from a university
of the town
in the village
in the apothecary
at the university
for the autumn
in the dormitory
at the home
at the university
at the university
with a prize
for an essay
at a time
of an apothecary
at the university
by the announcement
into an essay
of the ecliptic
to the invention
of the electrophorus
in the year

at the university
with a battery
with the management
of the pharmacy
at the university
of the chemical
under the title
to the identity
through a rise
on the magnet
in the expedition
with the assistance
on a lead
to the relationship
through a piece
across the compass
at the university
to the publication
before a group
in the spring
of the trade
on the feed
on the feed
in the prevention
in the prevention
For the treatment
For the prevention
at the onset
of the disease
in the prevention
in the prevention
in the prevention
in the reduction
in the prevention
in the reduction
at the level
at a rate
At the rate
of the cow
for a period
at a rate
of the female
in the control
with a maximum
For the prevention
For the prevention
For the treatment
of the fence
within the feed
with an elevator
to the fence
on the edge
of the feed
in the way
along the chain
of every phase
from the standpoint
of the product
for a production
for a service
To the manufacturer
in the food
In the area
by the fact
of a number
with a high-quality
from the standpoint
on the horizon
in the market
of an error

of the market
from the field
In the future
of the expansion
of the economy
to the need
for the future
at the district
in the future
in the market
of the architect
of a building
of the matter
on the subject
on the spot
of a number
on the opinion
on the market
to the client
with the aim
with the wording
in the quality
of the teaching
with a floor
of the client
in a vacuum
with the essence
at the expense
of the work
from the realization
in the trade
in the work
above the rank
for the educator
on the conduct
of the group
in the college
of the educator
of an organization
in the continuance
on the job
in the admission
for a teacher
At the end
in the profession
of the art
with the academic
in the country
in the course
to the question
in the subject
by an art
in the union
with an attendance
after the holiday
with the number
from an insurance
of the country
at a savings
to the deceased
despite the fact
to the cafeteria
during the break
at the cafeteria
for the coffee
on the job
for the convenience
of the company
on the payroll
of a coin

of a plan
for the worker
on the job
with an eye
to the operation
of the plant
after the holiday
at a profit
with a change
with the shutdown
to a vacation
on the number
including the devotion
for the enjoyment
of the director
since the start
of the program
of the fact
in every school
with the exception
in the adult
of the average
to the success
through the participation
under the direction
in the need
in the development
since the adoption
into the development
at the equipment
at the moment
from the weather
for the migration
of the country
of a vacation
with an average
of the tip
along the way
of the time
on the idea
of every camper
in the world
in the operation
for the engineer
in a management
in the design
through a number
In the GE
in the policy
on the part
by the vice-president
to the product
in the job
for the company
through the engineering-management
despite the creation
of a salary
for a research
of a salary
than the average
of a research
of the engineer
without the company
to the award
to the vice-president
in the treatment
of an individual
within the group
of the research-staff
to the difficulty

in the business
to the authority
in the management
in the country
from a line
in the law
over the fact
on the basis
in the family
with the house
by a lawyer
to the community
for the oil
to the town
by the oil
for the use
of the oilheating
for the life
of the lease
of the capital
in the deed
for the oil
of the system
on the basement
in the idea
at the time
at the site
by a revival
of the water
for the summer
down the storm
in the fluid
in the drain
around the freeze-out
of the radiator
during the summer
to an accumulation
of the coolant
in a truck
in a health
In the period
since the end
in the field
of the sign
to the adaptation
by the sign
to the field
by the photo
of the market
with the remainder
in the stronghold
from the end
in the field
From the volume
by the sign
in the field
with a handful
in the sign
of the plastic
of the designer
for the industry
around the clock
of the total
for a comparison
for the sign
to the field
of the laminate
of the unit
With the source
behind the copy

of the sign
by a photo
on the back
onto the acrylic
in the field
to the sign
on the way
at a time
in a cycle
of the sign
on the back
to the building
of the acrylic
to the face
of the bathroom
for the window
to the attic
between the house
under the roof
on the engine
in the block
from the inside
at the site
with a dealer
inside the house
on the roof
for the builder
of the total
on the problem
to the new-house
With a 15,500-lb.
at a time
in a 50'
than the average
of the load
of the load
in the open
from the yard
to the job
by a truckdriver
in a fraction
of the time
at a fraction
of the cost
on the roof
inside the house
of the gain
of the time-&-motion
on the job
on a job
from the market
of the champagne
at a sidewalk
in the heel
to the sea
behind the wheel
of a car
of the car
up a car
with the car
with the car
on the term
for a week
to the gallon
by the number
With the exception
of the agency
at the airport
in the group
In the majority

with the rate
than a month
on the road
on the road
in the charge
by the rate
for the rental
to the number
of an average

9. Use a default dictionary and itemgetter (n) to sort the most frequent tags used in the brown corpus (categories="reviews"). Please first convert the tags into the universal tags.

```
In [26]: brown_reviews_tagged_words = brown.tagged_words(categories = 'reviews', tagset = 'universal')
counts = defaultdict(int)
for (word, tag) in brown_reviews_tagged_words:
    counts[tag] += 1

[t for t, c in sorted(counts.items(), key=itemgetter(1), reverse = True)]
```

```
Out[26]: ['NOUN',
          'VERB',
          '.',
          'ADP',
          'DET',
          'ADJ',
          'ADV',
          'CONJ',
          'PRON',
          'PRT',
          'NUM',
          'X']
```

10. Explore the brown corpus (categories="learned") to find out the most 200 frequent words and store their most likely tags. We can then use this information as the model for a "lookup tagger" (an NLTK UnigramTagger). If the words are not among the 200 most frequent words, we would like to assign the default tag of "NN" to them. Then use this lookup tagger to tag a new sentence of your own.

```
In [27]: brown_tagged_sents = brown.tagged_sents(categories = 'learned')
learned_fd = nltk.FreqDist(brown.words(categories='learned'))
learned_cfd = nltk.ConditionalFreqDist(brown.tagged_words(categories='learned'))
most_freq_10 = learned_fd.most_common(200)
likely_tags = dict((word, learned_cfd[word].max()) for (word, _) in most_freq_10)
baseline_tagger_10 = nltk.UnigramTagger(model=likely_tags)
baseline_tagger_10.evaluate(brown_tagged_sents)
```

```
Out[27]: 0.5305627638986629
```

```
In [28]: baseline_tagger_10 = nltk.UnigramTagger(model=likely_tags, backoff=nltk.DefaultTagger('NN'))
```

```
In [31]: string_10 = 'yes indeed this is a very fine and dandy sentence indeed'
tokens_10 = word_tokenize(string_10)
(baseline_tagger_10.tag(tokens_10))
```

```
Out[31]: [('yes', 'NN'),
          ('indeed', 'NN'),
          ('this', 'DT'),
          ('is', 'BEZ'),
          ('a', 'AT'),
          ('very', 'QL'),
          ('fine', 'NN'),
          ('and', 'CC'),
          ('dandy', 'NN'),
          ('sentence', 'NN'),
          ('indeed', 'NN')]
```

