

2019_1_mcpp_taller_9_mateo_lancheros

April 26, 2019

1 Taller 9

Métodos Computacionales para Políticas Públicas - UROSARIO

Entrega: viernes 26-abr-2019 11:59 PM

César Mateo Lancheros Cañón cesar.lancheros@urosario.edu.co

1.1 Instrucciones:

- Guarde una copia de este *Jupyter Notebook* en su computador, idealmente en una carpeta destinada al material del curso.
- Modifique el nombre del archivo del *notebook*, agregando al final un guión inferior y su nombre y apellido, separados estos últimos por otro guión inferior. Por ejemplo, mi *notebook* se llamaría: mcpp_taller9_santiago_matallana
- Marque el *notebook* con su nombre y e-mail en el bloque verde arriba. Reemplace el texto "[Su nombre acá]" con su nombre y apellido. Similar para su e-mail.
- Desarrolle la totalidad del taller sobre este *notebook*, insertando las celdas que sea necesario debajo de cada pregunta. Haga buen uso de las celdas para código y de las celdas tipo *markdown* según el caso.
- Recuerde salvar periódicamente sus avances.
- Cuando termine el taller:
 1. Descárguelo en PDF. Si tiene algún problema con la conversión, descárguelo en HTML.
 2. Suba todos los archivos a su repositorio en GitHub, en una carpeta destinada exclusivamente para este taller, antes de la fecha y hora límites.

2 Natural Language Processing with Python

2.1 Analyzing Text with the Natural Language Toolkit

NLTK Book (<http://www.nltk.org/book/>), Exercises: - Chapter 1: 22, 26, 28 - Chapter 2: 2, 4, 11

```
In [73]: %matplotlib inline
import matplotlib.pyplot as plt
plt.rcParams["figure.figsize"] = [18.0, 8.0]
```

```
In [3]: import nltk
```

```
In [4]: nltk.download('state_union')
        nltk.download('brown')
        nltk.download('inaugural')
        nltk.download('nps_chat')
        nltk.download('webtext')
        nltk.download('treebank')
```

```
[nltk_data] Downloading package state_union to
[nltk_data] /Users/mateolancheros/nltk_data...
[nltk_data] Package state_union is already up-to-date!
[nltk_data] Downloading package brown to
[nltk_data] /Users/mateolancheros/nltk_data...
[nltk_data] Package brown is already up-to-date!
[nltk_data] Downloading package inaugural to
[nltk_data] /Users/mateolancheros/nltk_data...
[nltk_data] Package inaugural is already up-to-date!
[nltk_data] Downloading package nps_chat to
[nltk_data] /Users/mateolancheros/nltk_data...
[nltk_data] Package nps_chat is already up-to-date!
[nltk_data] Downloading package webtext to
[nltk_data] /Users/mateolancheros/nltk_data...
[nltk_data] Package webtext is already up-to-date!
[nltk_data] Downloading package treebank to
[nltk_data] /Users/mateolancheros/nltk_data...
[nltk_data] Package treebank is already up-to-date!
```

```
Out[4]: True
```

```
In [5]: nltk.download()
```

```
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

```
Out[5]: True
```

```
In [84]: from nltk.book import text1
         from nltk.book import text4
         from nltk.book import text5
         from nltk import FreqDist
```

```
In [77]: text1
```

```
Out[77]: <Text: Moby Dick by Herman Melville 1851>
```

```
In [85]: text4
```

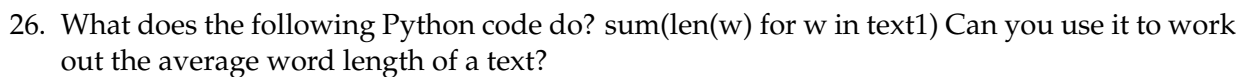
```
Out[85]: <Text: Inaugural Address Corpus>
```

```
In [9]: text5
```

```
Out[9]: <Text: Chat Corpus>
```

22. Find all the four-letter words in the Chat Corpus (text5). With the help of a frequency distribution (FreqDist), show these words in decreasing order of frequency.

['AKDT', 'AKST', 'AWAY', 'Ahhh', 'Away', 'Awww', 'Back', 'Been', 'Bone', 'Born', 'Boyz', 'CALI



3

999044

```
In [30]: print(sum(len(w) for w in text1)/len(text1))
```

3.830411128023649

28. Define a function `percent(word, text)` that calculates how often a given word occurs in a text, and expresses the result as a percentage.

```
In [67]: def percent(word, text):  
         return FreqDist(text)[word]/float(len(text))*100  
         percent("JOIN",text5)
```

Out[67]: 2.2683848033770273

2.3 Chapter 2. Accessing Text Corpora and Lexical Resources

2. Use the `corpus` module to explore `austen-persuasion.txt`. How many word tokens does this book have? How many word types?

```
In [69]: from nltk.corpus import gutenber
```

```
In [70]: austen_persuasion = gutenber.words('austen-persuasion.txt')  
         print("Number of word tokens = ",len(austen_persuasion))  
         print("Number of word types = ",len(set(austen_persuasion)))
```

Number of word tokens = 98171

Number of word types = 6132

4. Read in the texts of the State of the Union addresses, using the `state_union` corpus reader. Count occurrences of `men`, `women`, and `people` in each document. What has happened to the usage of these words over time?

```
In [86]: text4
```

Out[86]: <Text: Inaugural Address Corpus>

```
In [112]: list(text4)
```

```
Out[112]: ['Fellow',  
          '-',  
          'Citizens',  
          'of',  
          'the',  
          'Senate',  
          'and',  
          'of',
```

'the',
'House',
'of',
'Representatives',
':',
'Among',
'the',
'vicissitudes',
'incident',
'to',
'life',
'no',
'event',
'could',
'have',
'filled',
'me',
'with',
'greater',
'anxieties',
'than',
'that',
'of',
'which',
'the',
'notification',
'was',
'transmitted',
'by',
'your',
'order',
'',
'and',
'received',
'on',
'the',
'14th',
'day',
'of',
'the',
'present',
'month',
'.',
'On',
'the',
'one',
'hand',
'',

'I',
'was',
'summoned',
'by',
'my',
'Country',
'',
'whose',
'voice',
'I',
'can',
'never',
'hear',
'but',
'with',
'veneration',
'and',
'love',
'',
'from',
'a',
'retreat',
'which',
'I',
'had',
'chosen',
'with',
'the',
'fondest',
'predilection',
'',
'and',
'',
'in',
'my',
'flattering',
'hopes',
'',
'with',
'an',
'immutable',
'decision',
'',
'as',
'the',
'asylum',
'of',
'my',

'declining',
'years',
'--',
'a',
'retreat',
'which',
'was',
'rendered',
'every',
'day',
'more',
'necessary',
'as',
'well',
'as',
'more',
'dear',
'to',
'me',
'by',
'the',
'addition',
'of',
'habit',
'to',
'inclination',
'',
'and',
'of',
'frequent',
'interruptions',
'in',
'my',
'health',
'to',
'the',
'gradual',
'waste',
'committed',
'on',
'it',
'by',
'time',
'.',
'On',
'the',
'other',
'hand',

'',
'the',
'magnitude',
'and',
'difficulty',
'of',
'the',
'trust',
'to',
'which',
'the',
'voice',
'of',
'my',
'country',
'called',
'me',
'',
'being',
'sufficient',
'to',
'awaken',
'in',
'the',
'wisest',
'and',
'most',
'experienced',
'of',
'her',
'citizens',
'a',
'distrustful',
'scrutiny',
'into',
'his',
'qualifications',
'',
'could',
'not',
'but',
'overwhelm',
'with',
'despondence',
'one',
'who',
'(',
'inheriting',

'inferior',
'endowments',
'from',
'nature',
'and',
'unpracticed',
'in',
'the',
'duties',
'of',
'civil',
'administration',
)',
'ought',
'to',
'be',
'peculiarly',
'conscious',
'of',
'his',
'own',
'deficiencies',
'.',
'In',
'this',
'conflict',
'of',
'emotions',
'all',
'I',
'dare',
'aver',
'is',
'that',
'it',
'has',
'been',
'my',
'faithful',
'study',
'to',
'collect',
'my',
'duty',
'from',
'a',
'just',
'appreciation',

'of',
'every',
'circumstance',
'by',
'which',
'it',
'might',
'be',
'affected',
'.',
'All',
'I',
'dare',
'hope',
'is',
'that',
'if',
',',
'in',
'executing',
'this',
'task',
',',
'I',
'have',
'been',
'too',
'much',
'swayed',
'by',
'a',
'grateful',
'remembrance',
'of',
'former',
'instances',
',',
'or',
'by',
'an',
'affectionate',
'sensibility',
'to',
'this',
'transcendent',
'proof',
'of',
'the',

'confidence',
'of',
'my',
'fellow',
'citizens',
'',
'and',
'have',
'thence',
'too',
'little',
'consulted',
'my',
'incapacity',
'as',
'well',
'as',
'disinclination',
'for',
'the',
'weighty',
'and',
'untried',
'cares',
'before',
'me',
'',
'my',
'error',
'will',
'be',
'palliated',
'by',
'the',
'motives',
'which',
'mislead',
'me',
'',
'and',
'its',
'consequences',
'be',
'judged',
'by',
'my',
'country',
'with',

'some',
'share',
'of',
'the',
'partiality',
'in',
'which',
'they',
'originated',
'.',
'Such',
'being',
'the',
'impressions',
'under',
'which',
'I',
'have',
',',
'in',
'obedience',
'to',
'the',
'public',
'summons',
',',
'repaired',
'to',
'the',
'present',
'station',
',',
'it',
'would',
'be',
'peculiarly',
'improper',
'to',
'omit',
'in',
'this',
'first',
'official',
'act',
'my',
'fervent',
'supplications',
'to',

'that',
'Almighty',
'Being',
'who',
'rules',
'over',
'the',
'universe',
,',
'who',
'presides',
'in',
'the',
'councils',
'of',
'nations',
,',
'and',
'whose',
'providential',
'aids',
'can',
'supply',
'every',
'human',
'defect',
,',
'that',
'His',
'benediction',
'may',
'consecrate',
'to',
'the',
'liberties',
'and',
'happiness',
'of',
'the',
'people',
'of',
'the',
'United',
'States',
'a',
'Government',
'instituted',
'by',

'themselves',
'for',
'these',
'essential',
'purposes',
'',
'and',
'may',
'enable',
'every',
'instrument',
'employed',
'in',
'its',
'administration',
'to',
'execute',
'with',
'success',
'the',
'functions',
'allotted',
'to',
'his',
'charge',
'.',
'In',
'tendering',
'this',
'homage',
'to',
'the',
'Great',
'Author',
'of',
'every',
'public',
'and',
'private',
'good',
'',
'I',
'assure',
'myself',
'that',
'it',
'expresses',
'your',

'sentiments',
'not',
'less',
'than',
'my',
'own',
'',
'nor',
'those',
'of',
'my',
'fellow',
'citizens',
'at',
'large',
'less',
'than',
'either',
'.',
'No',
'people',
'can',
'be',
'bound',
'to',
'acknowledge',
'and',
'adore',
'the',
'Invisible',
'Hand',
'which',
'conducts',
'the',
'affairs',
'of',
'men',
'more',
'than',
'those',
'of',
'the',
'United',
'States',
'.',
'Every',
'step',
'by',

'which',
'they',
'have',
'advanced',
'to',
'the',
'character',
'of',
'an',
'independent',
'nation',
'seems',
'to',
'have',
'been',
'distinguished',
'by',
'some',
'token',
'of',
'providential',
'agency',
';',
'and',
'in',
'the',
'important',
'revolution',
'just',
'accomplished',
'in',
'the',
'system',
'of',
'their',
'united',
'government',
'the',
'tranquil',
'deliberations',
'and',
'voluntary',
'consent',
'of',
'so',
'many',
'distinct',
'communities',

'from',
'which',
'the',
'event',
'has',
'resulted',
'can',
'not',
'be',
'compared',
'with',
'the',
'means',
'by',
'which',
'most',
'governments',
'have',
'been',
'established',
'without',
'some',
'return',
'of',
'pious',
'gratitude',
'',
'along',
'with',
'an',
'humble',
'anticipation',
'of',
'the',
'future',
'blessings',
'which',
'the',
'past',
'seem',
'to',
'presage',
'.',
'These',
'reflections',
'',
'arising',
'out',

'of',
'the',
'present',
'crisis',
,',
'have',
'forced',
'themselves',
'too',
'strongly',
'on',
'my',
'mind',
'to',
'be',
'suppressed',
,',
'You',
'will',
'join',
'with',
'me',
,',
'I',
'trust',
,',
'in',
'thinking',
'that',
'there',
'are',
'none',
'under',
'the',
'influence',
'of',
'which',
'the',
'proceedings',
'of',
'a',
'new',
'and',
'free',
'government',
'can',
'more',
'auspiciously',

'commence',
'.',
'By',
'the',
'article',
'establishing',
'the',
'executive',
'department',
'it',
'is',
'made',
'the',
'duty',
'of',
'the',
'President',
'",
'to',
'recommend',
'to',
'your',
'consideration',
'such',
'measures',
'as',
'he',
'shall',
'judge',
'necessary',
'and',
'expedient',
'."',
'The',
'circumstances',
'under',
'which',
'I',
'now',
'meet',
'you',
'will',
'acquit',
'me',
'from',
'entering',
'into',
'that',

'subject',
'further',
'than',
'to',
'refer',
'to',
'the',
'great',
'constitutional',
'charter',
'under',
'which',
'you',
'are',
'assembled',
,',
'and',
'which',
,',
'in',
'defining',
'your',
'powers',
,',
'designates',
'the',
'objects',
'to',
'which',
'your',
'attention',
'is',
'to',
'be',
'given',
,',
'It',
'will',
'be',
'more',
'consistent',
'with',
'those',
'circumstances',
,',
'and',
'far',
'more',

'congenial',
'with',
'the',
'feelings',
'which',
'actuate',
'me',
'',
'to',
'substitute',
'',
'in',
'place',
'of',
'a',
'recommendation',
'of',
'particular',
'measures',
'',
'the',
'tribute',
'that',
'is',
'due',
'to',
'the',
'talents',
'',
'the',
'rectitude',
'',
'and',
'the',
'patriotism',
'which',
'adorn',
'the',
'characters',
'selected',
'to',
'devise',
'and',
'adopt',
'them',
'.',
'In',
'these',

'honorable',
'qualifications',
'I',
'behold',
'the',
'surest',
'pledges',
'that',
'as',
'on',
'one',
'side',
'no',
'local',
'prejudices',
'or',
'attachments',
'',
'no',
'separate',
'views',
'nor',
'party',
'animosities',
'',
'will',
'misdirect',
'the',
'comprehensive',
'and',
'equal',
'eye',
'which',
'ought',
'to',
'watch',
'over',
'this',
'great',
'assemblage',
'of',
'communities',
'and',
'interests',
'',
'so',
'',
'on',

'another',
,',
'that',
'the',
'foundation',
'of',
'our',
'national',
'policy',
'will',
'be',
'laid',
'in',
'the',
'pure',
'and',
'immutable',
'principles',
'of',
'private',
'morality',
,',
'and',
'the',
'preeminence',
'of',
'free',
'government',
'be',
'exemplified',
'by',
'all',
'the',
'attributes',
'which',
'can',
'win',
'the',
'affections',
'of',
'its',
'citizens',
'and',
'command',
'the',
'respect',
'of',
'the',

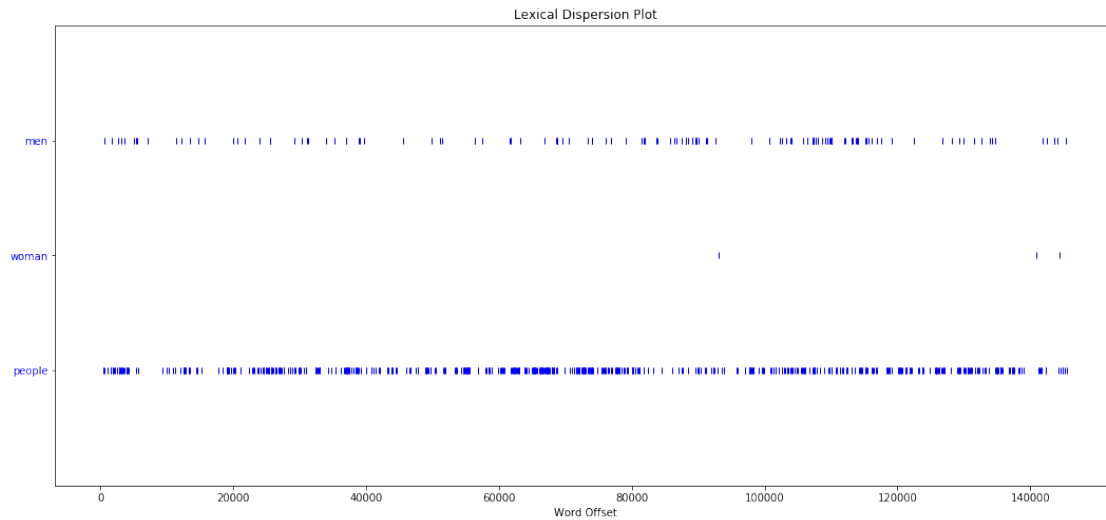
'world',
'.',
'I',
'dwell',
'on',
'this',
'prospect',
'with',
'every',
'satisfaction',
'which',
'an',
'ardent',
'love',
'for',
'my',
'country',
'can',
'inspire',
',',
'since',
'there',
'is',
'no',
'truth',
'more',
'thoroughly',
'established',
'than',
'that',
'there',
'exists',
'in',
'the',
'economy',
'and',
'course',
'of',
'nature',
'an',
'indissoluble',
'union',
'between',
'virtue',
'and',
'happiness',
';',
'between',


```
'duty',  
'and',  
'advantage',  
';',  
'between',  
'the',  
'genuine',  
'maxims',  
'of',  
'an',  
'honest',  
'and',  
'magnanimous',  
'policy',  
'and',  
'the',  
'solid',  
'rewards',  
'of',  
'public',  
'prosperity',  
'and',  
'felicity',  
';',  
'since',  
'we',  
'ought',  
'to',  
'be',  
'no',  
'less',  
'persuaded',  
...]
```

```
In [117]: len(set(text4))
```

```
Out[117]: 9754
```

```
In [119]: text4.dispersion_plot(["men", "woman", "people"]);
```



```
In [124]: fdist1 = FreqDist(text4)
          fdist1
```

```
Out[124]: FreqDist({'the': 9281, 'of': 6970, ',': 6840, 'and': 4991, '.': 4676, 'to': 4311, 'in': 4281, 'a': 4271, 'is': 4261, 'that': 4251, 'are': 4241, 'on': 4231, 'the': 4221, 'at': 4211, 'by': 4201, 'from': 4191, 'with': 4181, 'this': 4171, 'for': 4161, 'as': 4151, 'but': 4141, 'not': 4131, 'they': 4121, 'it': 4111, 'he': 4101, 'she': 4091, 'we': 4081, 'you': 4071, 'he': 4061, 'she': 4051, 'we': 4041, 'you': 4031, 'he': 4021, 'she': 4011, 'we': 4001, 'you': 3991, 'he': 3981, 'she': 3971, 'we': 3961, 'you': 3951, 'he': 3941, 'she': 3931, 'we': 3921, 'you': 3911, 'he': 3901, 'she': 3891, 'we': 3881, 'you': 3871, 'he': 3861, 'she': 3851, 'we': 3841, 'you': 3831, 'he': 3821, 'she': 3811, 'we': 3801, 'you': 3791, 'he': 3781, 'she': 3771, 'we': 3761, 'you': 3751, 'he': 3741, 'she': 3731, 'we': 3721, 'you': 3711, 'he': 3701, 'she': 3691, 'we': 3681, 'you': 3671, 'he': 3661, 'she': 3651, 'we': 3641, 'you': 3631, 'he': 3621, 'she': 3611, 'we': 3601, 'you': 3591, 'he': 3581, 'she': 3571, 'we': 3561, 'you': 3551, 'he': 3541, 'she': 3531, 'we': 3521, 'you': 3511, 'he': 3501, 'she': 3491, 'we': 3481, 'you': 3471, 'he': 3461, 'she': 3451, 'we': 3441, 'you': 3431, 'he': 3421, 'she': 3411, 'we': 3401, 'you': 3391, 'he': 3381, 'she': 3371, 'we': 3361, 'you': 3351, 'he': 3341, 'she': 3331, 'we': 3321, 'you': 3311, 'he': 3301, 'she': 3291, 'we': 3281, 'you': 3271, 'he': 3261, 'she': 3251, 'we': 3241, 'you': 3231, 'he': 3221, 'she': 3211, 'we': 3201, 'you': 3191, 'he': 3181, 'she': 3171, 'we': 3161, 'you': 3151, 'he': 3141, 'she': 3131, 'we': 3121, 'you': 3111, 'he': 3101, 'she': 3091, 'we': 3081, 'you': 3071, 'he': 3061, 'she': 3051, 'we': 3041, 'you': 3031, 'he': 3021, 'she': 3011, 'we': 3001, 'you': 2991, 'he': 2981, 'she': 2971, 'we': 2961, 'you': 2951, 'he': 2941, 'she': 2931, 'we': 2921, 'you': 2911, 'he': 2901, 'she': 2891, 'we': 2881, 'you': 2871, 'he': 2861, 'she': 2851, 'we': 2841, 'you': 2831, 'he': 2821, 'she': 2811, 'we': 2801, 'you': 2791, 'he': 2781, 'she': 2771, 'we': 2761, 'you': 2751, 'he': 2741, 'she': 2731, 'we': 2721, 'you': 2711, 'he': 2701, 'she': 2691, 'we': 2681, 'you': 2671, 'he': 2661, 'she': 2651, 'we': 2641, 'you': 2631, 'he': 2621, 'she': 2611, 'we': 2601, 'you': 2591, 'he': 2581, 'she': 2571, 'we': 2561, 'you': 2551, 'he': 2541, 'she': 2531, 'we': 2521, 'you': 2511, 'he': 2501, 'she': 2491, 'we': 2481, 'you': 2471, 'he': 2461, 'she': 2451, 'we': 2441, 'you': 2431, 'he': 2421, 'she': 2411, 'we': 2401, 'you': 2391, 'he': 2381, 'she': 2371, 'we': 2361, 'you': 2351, 'he': 2341, 'she': 2331, 'we': 2321, 'you': 2311, 'he': 2301, 'she': 2291, 'we': 2281, 'you': 2271, 'he': 2261, 'she': 2251, 'we': 2241, 'you': 2231, 'he': 2221, 'she': 2211, 'we': 2201, 'you': 2191, 'he': 2181, 'she': 2171, 'we': 2161, 'you': 2151, 'he': 2141, 'she': 2131, 'we': 2121, 'you': 2111, 'he': 2101, 'she': 2091, 'we': 2081, 'you': 2071, 'he': 2061, 'she': 2051, 'we': 2041, 'you': 2031, 'he': 2021, 'she': 2011, 'we': 2001, 'you': 1991, 'he': 1981, 'she': 1971, 'we': 1961, 'you': 1951, 'he': 1941, 'she': 1931, 'we': 1921, 'you': 1911, 'he': 1901, 'she': 1891, 'we': 1881, 'you': 1871, 'he': 1861, 'she': 1851, 'we': 1841, 'you': 1831, 'he': 1821, 'she': 1811, 'we': 1801, 'you': 1791, 'he': 1781, 'she': 1771, 'we': 1761, 'you': 1751, 'he': 1741, 'she': 1731, 'we': 1721, 'you': 1711, 'he': 1701, 'she': 1691, 'we': 1681, 'you': 1671, 'he': 1661, 'she': 1651, 'we': 1641, 'you': 1631, 'he': 1621, 'she': 1611, 'we': 1601, 'you': 1591, 'he': 1581, 'she': 1571, 'we': 1561, 'you': 1551, 'he': 1541, 'she': 1531, 'we': 1521, 'you': 1511, 'he': 1501, 'she': 1491, 'we': 1481, 'you': 1471, 'he': 1461, 'she': 1451, 'we': 1441, 'you': 1431, 'he': 1421, 'she': 1411, 'we': 1401, 'you': 1391, 'he': 1381, 'she': 1371, 'we': 1361, 'you': 1351, 'he': 1341, 'she': 1331, 'we': 1321, 'you': 1311, 'he': 1301, 'she': 1291, 'we': 1281, 'you': 1271, 'he': 1261, 'she': 1251, 'we': 1241, 'you': 1231, 'he': 1221, 'she': 1211, 'we': 1201, 'you': 1191, 'he': 1181, 'she': 1171, 'we': 1161, 'you': 1151, 'he': 1141, 'she': 1131, 'we': 1121, 'you': 1111, 'he': 1101, 'she': 1091, 'we': 1081, 'you': 1071, 'he': 1061, 'she': 1051, 'we': 1041, 'you': 1031, 'he': 1021, 'she': 1011, 'we': 1001, 'you': 991, 'he': 981, 'she': 971, 'we': 961, 'you': 951, 'he': 941, 'she': 931, 'we': 921, 'you': 911, 'he': 901, 'she': 891, 'we': 881, 'you': 871, 'he': 861, 'she': 851, 'we': 841, 'you': 831, 'he': 821, 'she': 811, 'we': 801, 'you': 791, 'he': 781, 'she': 771, 'we': 761, 'you': 751, 'he': 741, 'she': 731, 'we': 721, 'you': 711, 'he': 701, 'she': 691, 'we': 681, 'you': 671, 'he': 661, 'she': 651, 'we': 641, 'you': 631, 'he': 621, 'she': 611, 'we': 601, 'you': 591, 'he': 581, 'she': 571, 'we': 561, 'you': 551, 'he': 541, 'she': 531, 'we': 521, 'you': 511, 'he': 501, 'she': 491, 'we': 481, 'you': 471, 'he': 461, 'she': 451, 'we': 441, 'you': 431, 'he': 421, 'she': 411, 'we': 401, 'you': 391, 'he': 381, 'she': 371, 'we': 361, 'you': 351, 'he': 341, 'she': 331, 'we': 321, 'you': 311, 'he': 301, 'she': 291, 'we': 281, 'you': 271, 'he': 261, 'she': 251, 'we': 241, 'you': 231, 'he': 221, 'she': 211, 'we': 201, 'you': 191, 'he': 181, 'she': 171, 'we': 161, 'you': 151, 'he': 141, 'she': 131, 'we': 121, 'you': 111, 'he': 101, 'she': 91, 'we': 81, 'you': 71, 'he': 61, 'she': 51, 'we': 41, 'you': 31, 'he': 21, 'she': 11, 'we': 1, 'you': 0, 'he': 0, 'she': 0})
```

```
In [126]: fdist1['men']
```

Out [126]: 139

```
In [127]: fdist1['woman']
```

Out[127]: 3

```
In [128]: fdist1['people']
```

Out [128]: 559

11. Investigate the table of modal distributions and look for other patterns. Try to explain them in terms of your own impressionistic understanding of the different genres. Can you find other closed classes of words that exhibit significant differences across different genres?

```
In [133]: from nltk.corpus import brown
brown.categories()
pronouns = ['I', 'you', 'he', 'she', 'it', 'we', 'they']
cfd = nltk.ConditionalFreqDist(
    (genre, word)
    for genre in brown.categories()
    for word in brown.words(categories=genre))
genres = ['news', 'religion', 'hobbies', 'romance', 'humor', 'editorial', 'government']
cfd.tabulate(conditions=genres, samples=pronouns)
```

	I	you	he	she	it	we	they
news	179	55	451	42	363	77	205
religion	155	100	137	10	264	176	115
hobbies	154	383	155	21	476	100	177
romance	951	456	702	496	573	78	168
humor	239	131	146	58	162	32	70
editorial	201	83	268	41	386	167	148
government	97	74	120	0	218	112	92

```
In [134]: wh = ['what', 'when', 'who', 'why', 'where']
          cfd.tabulate(conditions=genres, samples=wh)
```

	what	when	who	why	where
news	76	128	268	9	58
religion	64	53	100	14	20
hobbies	78	119	103	10	72
romance	121	126	89	34	54
humor	36	52	48	9	15
editorial	84	103	172	10	40
government	43	56	74	6	46
