

# mcpp\_taller9\_john\_caro

October 28, 2016

## 1 Taller 9

Métodos Computacionales para Políticas Públicas - URSario

**Entrega: viernes 28-oct-2016 11:59 PM**

**[John Alexander Caro Becerra]** [jhonalexbc@gmail.com]

### 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\_mataallana
- 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.

---

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

```
In [1]: %matplotlib inline
import matplotlib.pyplot as plt
plt.rcParams["figure.figsize"] = [18.0, 8.0]

In [2]: import nltk

In [4]: nltk.download()
```

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

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

```
In [6]: from nltk.book import *
```

## 1.2 Chapter 1

**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.**

```
In [7]: L = list(text5)
```

```
In [8]: cuatro_letras = [w for w in L if len(w) == 4]
        cuatro_letras
```

```
Out[8]: ['left',
        'with',
        'this',
        'name',
        'PART',
        'well',
        'NICK',
        'name',
        'U121',
        'golf',
        'clap',
        'JOIN',
        'that',
        'nice',
        'JOIN',
        'PART',
        'golf',
        'clap',
        'fuck',
        'U121',
        'PART',
        'PART',
        'clap',
        'your',
        'PART',
        'dont',
        'even',
        'know',
        'what',
        'that',
        'that',
        'chat',
```

'JOIN',  
'drew',  
'cast',  
'PART',  
'sexy',  
'U115',  
'JOIN',  
'PART',  
'drew',  
'girl',  
'with',  
'legs',  
'hope',  
'draw',  
'PART',  
'head',  
'legs',  
'JOIN',  
'JOIN',  
'good',  
'JOIN',  
'PART',  
'take',  
'have',  
'docs',  
'Slip',  
'away',  
'Fade',  
'away',  
'Days',  
'away',  
'feel',  
'have',  
'back',  
'U115',  
'U129',  
'U115',  
'chat',  
'with',  
'PART',  
'JOIN',  
'JOIN',  
'fast',  
'U116',  
'bowl',  
'bong',  
'JOIN',  
'well',

'glad',  
'hard',  
'from',  
'here',  
'back',  
'PART',  
'PART',  
'JOIN',  
'U121',  
'name',  
'hard',  
'very',  
'fire',  
'from',  
'here',  
'JOIN',  
'PART',  
'itch',  
'JOIN',  
'U133',  
'ogan',  
'male',  
'JOIN',  
'JOIN',  
'show',  
'will',  
'talk',  
'PART',  
'haha',  
'opps',  
'JOIN',  
'PART',  
'U115',  
'nice',  
'warm',  
'guys',  
'with',  
'cams',  
'play',  
'sits',  
'JOIN',  
'JOIN',  
'guyz',  
'chat',  
'U126',  
'PART',  
'chat',  
'PART',

'gooo',  
'sure',  
'U126',  
'JOIN',  
'what',  
'feel',  
'like',  
'room',  
'yeee',  
'JOIN',  
'want',  
'pics',  
'look',  
'U139',  
'PART',  
'PART',  
'JOIN',  
'here',  
'JOIN',  
'PART',  
'JOIN',  
'U139',  
'PART',  
'JOIN',  
'U138',  
'U139',  
'make',  
'U139',  
'that',  
'U126',  
'late',  
'lmao',  
'ahah',  
'PART',  
'U121',  
'U121',  
'does',  
'like',  
'that',  
'guys',  
'male',  
'JOIN',  
'U139',  
'well',  
'what',  
'yeah',  
'know',  
'U136',

'hell',  
'with',  
'U139',  
'U101',  
'like',  
'when',  
'plan',  
'PART',  
'JOIN',  
'gold',  
'jeep',  
'make',  
'sure',  
'nice',  
'ring',  
'U115',  
'isnt',  
'that',  
'U136',  
'hell',  
'have',  
'have',  
'doin',  
'U139',  
'U121',  
'many',  
'Just',  
'fine',  
'that',  
'like',  
'PART',  
'hiya',  
'room',  
'lmao',  
'doin',  
'Deep',  
'Show',  
'that',  
'love',  
'that',  
'Turn',  
'take',  
'Hand',  
'just',  
'even',  
'look',  
'hang',  
'PART',

'that',  
'such',  
'word',  
'U141',  
'hear',  
'!!!!',  
'PART',  
'JOIN',  
'PART',  
'deaf',  
'here',  
'dont',  
'U115',  
'U115',  
'...',  
'hugs',  
'chat',  
'with',  
'baby',  
'Only',  
'U121',  
'U121',  
'PART',  
'have',  
'away',  
'from',  
'U121',  
'what',  
'read',  
'here',  
'with',  
'JOIN',  
'read',  
'have',  
'here',  
'JOIN',  
'want',  
'chat',  
'talk',  
'U121',  
'JOIN',  
'U121',  
'VBox',  
'PART',  
'take',  
'that',  
'JOIN',  
'PART',

'hate',  
'when',  
'U121',  
'U115',  
'lmao',  
'PART',  
'your',  
'know',  
'what',  
'your',  
'what',  
'JOIN',  
'love',  
'more',  
'than',  
'ELSE',  
'serg',  
'well',  
'most',  
'love',  
'JOIN',  
'know',  
'that',  
'what',  
'lmao',  
'well',  
'have',  
'eyes',  
'lmao',  
'know',  
'JOIN',  
'girl',  
'jerk',  
'kids',  
'guys',  
'type',  
'much',  
'shut',  
'fuck',  
'girl',  
'nice',  
'shut',  
'fuck',  
'PART',  
'dont',  
'want',  
'JOIN',  
'want',



'U115',  
'what',  
'miss',  
'much',  
'work',  
'nice',  
'U116',  
'PART',  
'PART',  
'hey',  
'U148',  
'hate',  
'boys',  
'JOIN',  
'U148',  
'hate',  
'what',  
'PART',  
'hate',  
'U121',  
'fuck',  
'your',  
'ugly',  
'JOIN',  
'bein',  
'PART',  
'What',  
'U115',  
'whys',  
'that',  
'deep',  
'U121',  
'what',  
'JOIN',  
'tape',  
'Your',  
'sexs',  
'best',  
'phil',  
'said',  
'ugly',  
'PART',  
'date',  
'feel',  
'your',  
'they',  
'form',  
'PART',

'sits',  
'JOIN',  
'sits',  
'with',  
'hmp',  
'hate',  
'does',  
'that',  
'mean',  
'want',  
'room',  
'this',  
'been',  
'PART',  
'JOIN',  
'U115',  
'U116',  
'your',  
'here',  
'talk',  
'wait',  
'that',  
'perv',  
'lets',  
'hope',  
'U121',  
'PART',  
'U115',  
'!!!!',  
'lets',  
'chat',  
'JOIN',  
'rule',  
'land',  
'wont',  
'then',  
'find',  
'need',  
'this',  
'HUGE',  
'perv',  
'that',  
'deal',  
'????',  
'JOIN',  
'shit',  
'hell',  
'lmao',

'PART',  
'hell',  
'JOIN',  
'here',  
'guys',  
'have',  
'U121',  
'JOIN',  
'U155',  
'only',  
'Poor',  
'U121',  
'love',  
'pick',  
'much',  
'that',  
'PART',  
'PART',  
'sits',  
'with',  
'U121',  
'nads',  
'JOIN',  
'from',  
'pick',  
'your',  
'pick',  
'your',  
'nose',  
'pick',  
'your',  
'nose',  
'JOIN',  
'face',  
'with',  
'PART',  
'U115',  
'owww',  
'PART',  
'JOIN',  
'U116',  
'PART',  
'does',  
'want',  
'talk',  
'head',  
'gags',  
'even',

'U121',  
'neck',  
'Meep',  
'U115',  
'LAsT',  
'time',  
'that',  
'wash',  
'your',  
'dude',  
'gets',  
'JOIN',  
'U121',  
'dang',  
'just',  
'pm's',  
'that',  
'1.99',  
'...',  
'yeah',  
'nice',  
'neck',  
'U115',  
'like',  
'shut',  
'free',  
'JOIN',  
'goes',  
'wash',  
'lmao',  
'Lies',  
'lmao',  
'U115',  
'lick',  
'very',  
'lmao',  
'U115',  
'ummm',  
'U109',  
'dont',  
'dead',  
'more',  
'than',  
'call',  
'just',  
'case',  
'dead',  
'good',

'neck',  
'talk',  
'what',  
'ummm',  
'else',  
'wont',  
'bite',  
'U115',  
'yeah',  
'wait',  
'yeah',  
'PART',  
'your',  
'want',  
'have',  
'sexy',  
'bite',  
'lmao',  
'call',  
'have',  
'free',  
'call',  
'mins',  
'JOIN',  
'nite',  
'lool',  
'know',  
'that',  
'kina',  
'give',  
'away',  
'then',  
'room',  
'call',  
'yeah',  
'U155',  
'PART',  
'U115',  
'more',  
'U115',  
'guys',  
'baby',  
'U109',  
'fuck',  
'case',  
'know',  
'were',  
'girl',

'JOIN',  
'baby',  
'what',  
'U109',  
'guys',  
'chat',  
'have',  
'have',  
'sext',  
'piff',  
'dont',  
'talk',  
'read',  
'dang',  
'lazy',  
'dont',  
'read',  
'PART',  
'JOIN',  
'mean',  
'fine',  
'...',  
'busy',  
'work',  
'okay',  
'dont',  
'talk',  
'calm',  
'down',  
'busy',  
'busy',  
'want',  
'chat',  
'arms',  
'kids',  
'name',  
'PART',  
'sits',  
'down',  
'eats',  
'JOIN',  
'hugs',  
'want',  
'U121',  
'near',  
'just',  
'PART',  
'JOIN',

'PART',  
'hell',  
'yeah',  
'U115',  
'near',  
'near',  
'good',  
'smax',  
'JOIN',  
'haha',  
'only',  
'>:->',  
'near',  
'PART',  
'piff',  
'Vvil',  
'JOIN',  
'free',  
'wont',  
'cold',  
'U121',  
'cell',  
'runs',  
'thru',  
'back',  
'hair',  
'eyes',  
'neck',  
'yeah',  
'caps',  
'PART',  
'JOIN',  
'PART',  
'PART',  
'U165',  
'jump',  
'U165',  
'baby',  
'here',  
'over',  
'your',  
'good',  
'PART',  
'that',  
'left',  
'room',  
'este',  
'U115',

'will',  
'PART',  
'U121',  
'U165',  
'lmao',  
'PART',  
'PART',  
'very',  
'guys',  
'wana',  
'chat',  
'chik',  
'from',  
'mean',  
'chat',  
'well',  
'PART',  
'that',  
'dont',  
'shit',  
'U165',  
'U165',  
'left',  
'room',  
'with',  
'your',  
'JOIN',  
'U115',  
'what',  
'list',  
'wish',  
'cmon',  
'U128',  
'nice',  
'JOIN',  
'list',  
'PART',  
'list',  
'U115',  
'good',  
'lmao',  
'U128',  
'hehe',  
'hows',  
'bout',  
'good',  
'hear',  
'U165',



'have',  
'JOIN',  
'good',  
'PART',  
'JOIN',  
'wats',  
'they',  
'PART',  
'piff',  
'aint',  
'know',  
'shut',  
'much',  
'good',  
'PART',  
'JOIN',  
'JOIN',  
'yeah',  
'lost',  
'like',  
'same',  
'well',  
'work',  
'what',  
'Boyz',  
'rock',  
'what',  
'hehe',  
'went',  
'back',  
'some',  
'then',  
'came',  
'back',  
'home',  
'PART',  
'what',  
'they',  
'coat',  
'nice',  
'read',  
'many',  
'nice',  
'hehe',  
'lmao',  
'even',  
'JOIN',  
'well',

'talk',  
'nite',  
'what',  
'very',  
'time',  
'What',  
'kind',  
'...',  
'nite',  
'PART',  
'Eyes',  
'Dawn',  
'last',  
'song',  
'LIVE',  
'cool',  
'good',  
'nite',  
'mauh',  
'nite',  
'mike',  
'keep',  
'must',  
'girl',  
'seem',  
'pick',  
'else',  
'...',  
'take',  
'your',  
'your',  
'JOIN',  
'lmao',  
'just',  
'days',  
'late',  
'with',  
'room',  
'JOIN',  
'PART',  
'good',  
'ques',  
'lmao',  
'JOIN',  
'that',  
'like',  
'dont',  
'quit',

'what',  
'your',  
'4.20',  
'PART',  
'like',  
'mine',  
'over',  
'cali',  
'good',  
'this',  
'year',  
'NICK',  
'whoa',  
'have',  
'have',  
'have',  
'what',  
'boys',  
'gosh',  
'that',  
'ruff',  
'what',  
'PART',  
'hell',  
'rock',  
'roll',  
'PART',  
'with',  
'like',  
'that',  
'nope',  
'...',  
'rest',  
'rock',  
'roll',  
'...',  
'...',  
'sing',  
'from',  
'kids',  
'...',  
'mame',  
'nada',  
'cali',  
'here',  
'that',  
'cool',  
'kids',

'cool',  
'JOIN',  
'with',  
'from',  
'said',  
'alot',  
'JOIN',  
'year',  
'band',  
'JOIN',  
'NICK',  
'cool',  
'nice',  
'here',  
'hair',  
'hard',  
'what',  
'type',  
'does',  
'your',  
'band',  
'play',  
'hair',  
'yeah',  
'what',  
'doin',  
'...',  
'with',  
'hand',  
'what',  
'JOIN',  
'room',  
'...',  
'sexy',  
'PART',  
'dumb',  
'they',  
'wont',  
'chat',  
'with',  
'lmao',  
'damn',  
'what',  
'orgy',  
'orgy',  
'lmao',  
'what',  
'word',

'some',  
'easy',  
'PART',  
'JOIN',  
'back',  
'orgy',  
'lmao',  
'PART',  
'lets',  
'play',  
'room',  
'PART',  
'JOIN',  
'JOIN',  
'were',  
'damn',  
'PART',  
'good',  
'call',  
'what',  
'like',  
'just',  
'late',  
'date',  
'know',  
'push',  
'PART',  
'lose',  
'name',  
'shit',  
'head',  
'long',  
'time',  
'said',  
'shit',  
'lost',  
'baby',  
'then',  
'JOIN',  
'PART',  
'JOIN',  
'What',  
'sure',  
'JOIN',  
'sure',  
'lmao',  
'this',  
'room',

'...',  
'...',  
'your',  
'...',  
'lmao',  
'that',  
'yeah',  
'have',  
'many',  
'lmao',  
'sexy',  
'stay',  
'keep',  
'lmao',  
'with',  
'hair',  
'like',  
'that',  
'...',  
'down',  
'door',  
'prob',  
'...',  
'hair',  
'lmao',  
'JOIN',  
'sexy',  
'just',  
'this',  
'life',  
'just',  
'PART',  
'room',  
'with',  
'hair',  
'like',  
'what',  
'from',  
'here',  
'said',  
'wild',  
'...',  
'even',  
'from',  
'JOIN',  
'cool',  
'sexy',  
'JOIN',

```
'sexy',  
'only',  
'none',  
'sexy',  
'whew',  
'sexy',  
'hell',  
'have',  
...]
```

```
In [11]: frecuencia_palabras = FreqDist(cuatro_letras)  
         frecuencia_palabras
```

```
Out[11]: FreqDist({'o.k.': 1,  
                  'caca': 1,  
                  'U102': 12,  
                  'MUAH': 1,  
                  'ally': 1,  
                  'Save': 1,  
                  'holy': 4,  
                  '39.3': 1,  
                  'mike': 2,  
                  'U181': 1,  
                  'tere': 1,  
                  'club': 2,  
                  'male': 12,  
                  'free': 13,  
                  'peel': 1,  
                  'yw's': 1,  
                  'U110': 25,  
                  'lady': 8,  
                  'newp': 2,  
                  'This': 12,  
                  'ahem': 3,  
                  'side': 2,  
                  'aime': 1,  
                  '; ..': 9,  
                  'hell': 29,  
                  '100%': 1,  
                  'boot': 1,  
                  'john': 2,  
                  'meds': 5,  
                  'fine': 16,  
                  '18ST': 1,  
                  'wish': 5,  
                  'Mono': 1,  
                  'alot': 7,  
                  'doin': 11,
```

'Matt': 1,  
'base': 1,  
'3:45': 1,  
'este': 1,  
'calm': 1,  
'U112': 2,  
'Lion': 1,  
'slam': 1,  
'ohwa': 1,  
'U172': 2,  
'yout': 1,  
'64.8': 1,  
'Look': 1,  
'land': 5,  
'shop': 2,  
'dotn': 1,  
'HERE': 1,  
'mass': 2,  
'1985': 1,  
'AKST': 1,  
'ummm': 4,  
'teck': 1,  
'line': 4,  
'poem': 1,  
'cuss': 1,  
'huge': 4,  
'ewww': 2,  
'sips': 1,  
'Only': 3,  
'DOES': 2,  
'yoll': 1,  
' :o \*': 1,  
'cant': 18,  
'cell': 2,  
'runs': 9,  
'wins': 1,  
'home': 11,  
'pigs': 1,  
'wuts': 1,  
'nana': 3,  
'soda': 1,  
'vent': 1,  
'6:41': 1,  
'Good': 1,  
'Chop': 1,  
'n9ne': 2,  
'poor': 6,  
'hots': 1,



'NAME': 1,  
'worl': 1,  
'whou': 1,  
'kool': 5,  
'case': 8,  
'sure': 26,  
'work': 38,  
'boys': 5,  
'shup': 1,  
'tisk': 2,  
'cold': 12,  
'hick': 3,  
'hits': 2,  
'tooo': 1,  
'vamp': 1,  
'puts': 1,  
'most': 6,  
'gone': 6,  
'want': 71,  
'nick': 9,  
'dick': 1,  
'head': 12,  
'Tina': 1,  
'poop': 1,  
'wore': 1,  
'left': 17,  
'send': 6,  
'Oops': 1,  
'9.53': 1,  
'them': 28,  
'pork': 1,  
'U145': 3,  
'find': 18,  
'wean': 1,  
'late': 5,  
'sits': 15,  
'7:45': 1,  
'salt': 1,  
'exit': 1,  
'twin': 2,  
'Well': 10,  
'Holy': 1,  
'Lmao': 2,  
'LMAO': 19,  
'swim': 3,  
'giva': 1,  
'nawp': 1,  
'puff': 4,

'twit': 1,  
'ltns': 2,  
'temp': 2,  
'hiom': 1,  
'been': 57,  
'dman': 1,  
'part': 6,  
'wrap': 1,  
'pour': 1,  
'Drew': 2,  
'1.98': 1,  
'dyed': 1,  
'poll': 1,  
'moms': 1,  
'girl': 43,  
'golf': 2,  
'Wyte': 1,  
'Care': 1,  
'slap': 3,  
'yada': 3,  
'damn': 34,  
'howl': 1,  
'98.6': 1,  
'come': 33,  
'nerd': 1,  
'list': 6,  
'feel': 19,  
'Wind': 3,  
'bacl': 1,  
'hear': 21,  
'full': 9,  
'Ahhh': 2,  
'wack': 3,  
'same': 19,  
'high': 4,  
'legs': 5,  
'Hott': 1,  
'pool': 2,  
'Fade': 1,  
'deal': 3,  
'flaw': 2,  
'grew': 1,  
'2:55': 1,  
'bird': 1,  
'wooo': 2,  
'4.20': 1,  
'hurt': 3,  
'gear': 1,

'slip': 1,  
'dont': 75,  
'Rofl': 1,  
'grrl': 2,  
'thje': 1,  
'U168': 13,  
'HAHA': 1,  
'WITH': 2,  
'word': 8,  
'hope': 12,  
'Need': 1,  
'heyy': 8,  
'hummm': 2,  
'LIVE': 1,  
'Home': 1,  
'pmsl': 1,  
'suck': 8,  
'Tiff': 1,  
'must': 19,  
'shes': 4,  
'scar': 1,  
'west': 1,  
'scum': 1,  
'U121': 36,  
'U107': 11,  
'face': 11,  
'read': 14,  
'fuck': 15,  
'deaf': 2,  
'ohio': 2,  
'last': 22,  
'HUGE': 1,  
'orgy': 3,  
'hank': 3,  
'spot': 2,  
'bong': 1,  
'lube': 1,  
'like': 156,  
'>:->': 2,  
'uyes': 1,  
'byes': 1,  
'hiya': 78,  
'dang': 9,  
'OOPS': 1,  
'door': 4,  
'dojn': 1,  
'jeep': 1,  
'itch': 3,

'dust': 1,  
'mono': 1,  
'U115': 36,  
'nuff': 1,  
'coem': 1,  
'Prof': 1,  
'luck': 5,  
'park': 2,  
'keep': 14,  
'take': 37,  
'eyes': 15,  
'Time': 1,  
'benz': 1,  
'kick': 7,  
'busy': 17,  
'Damn': 1,  
'warm': 5,  
'told': 12,  
'real': 16,  
'just': 125,  
'U164': 1,  
'sort': 2,  
'HALO': 1,  
'hate': 23,  
'U163': 3,  
'bloe': 1,  
'tiff': 1,  
'even': 35,  
'GIRL': 1,  
'more': 34,  
'mary': 3,  
'bomb': 1,  
'Kick': 1,  
'LONG': 1,  
' , , , , ': 4,  
'lawl': 2,  
'<<<<': 2,  
'roll': 5,  
'rose': 1,  
'tape': 1,  
'love': 60,  
'beat': 4,  
'jail': 1,  
'babe': 11,  
'Elev': 3,  
'febe': 1,  
'fair': 1,  
'bear': 2,

'pink': 6,  
'stay': 12,  
'root': 2,  
'post': 1,  
'cali': 5,  
'moon': 2,  
'pope': 1,  
'wana': 8,  
'caps': 5,  
'pass': 4,  
'cars': 2,  
'Jess': 1,  
'Song': 6,  
'then': 40,  
'sori': 1,  
'Kewl': 1,  
'easy': 5,  
'tend': 1,  
'also': 5,  
'outa': 1,  
'shot': 4,  
'comp': 6,  
'typo': 2,  
'Kids': 1,  
'Mary': 1,  
'Sure': 2,  
'away': 26,  
'ciao': 2,  
'dint': 1,  
'Live': 2,  
'Cute': 1,  
'plan': 2,  
'week': 10,  
'allo': 1,  
'back': 78,  
'lala': 1,  
'U146': 4,  
'wash': 3,  
'lisa': 1,  
'Down': 2,  
'baby': 26,  
'nads': 1,  
'pics': 9,  
'roof': 3,  
'took': 7,  
'TIME': 1,  
'tthe': 1,  
'Hold': 1,

'Hugs': 1,  
'http': 1,  
'Yoko': 1,  
'tart': 1,  
'Room': 1,  
'bust': 1,  
'None': 1,  
'burp': 2,  
'loss': 1,  
'Ohio': 2,  
'High': 2,  
'heee': 1,  
'pm's': 1,  
'kiss': 9,  
'beer': 5,  
'U123': 4,  
'Then': 1,  
'disc': 1,  
'gosh': 1,  
'made': 8,  
'trip': 2,  
'evah': 1,  
'band': 3,  
'anal': 1,  
'what': 183,  
'ssid': 1,  
'Back': 1,  
'pine': 1,  
'ello': 3,  
'Will': 1,  
'next': 6,  
'ltnc': 1,  
'...': 170,  
'Ohhh': 1,  
'mkay': 1,  
'Turn': 1,  
'fear': 1,  
'main': 6,  
'Kiss': 1,  
'Werd': 1,  
'tock': 2,  
'glad': 4,  
'blah': 6,  
'bout': 12,  
'spit': 1,  
'outs': 1,  
'lord': 4,  
'gawd': 3,

'dear': 7,  
'News': 1,  
'jerk': 4,  
'z-ro': 1,  
'heya': 15,  
'gays': 2,  
'joke': 5,  
'ones': 4,  
'wats': 2,  
'pull': 1,  
'yell': 1,  
'open': 4,  
'hows': 8,  
'tlak': 1,  
'vega': 1,  
'Jane': 1,  
'U141': 3,  
'DAMN': 1,  
'goof': 1,  
'U139': 8,  
'else': 13,  
'Same': 3,  
'hgey': 1,  
'poot': 1,  
'when': 48,  
'CALI': 1,  
'uses': 2,  
'Food': 1,  
'Bone': 1,  
'xmas': 1,  
'ahhh': 7,  
'whoo': 1,  
'CHAT': 3,  
'U196': 4,  
'sign': 1,  
'lame': 4,  
'amen': 3,  
'clap': 3,  
'Nice': 2,  
'spin': 2,  
'serg': 1,  
'dead': 9,  
'U190': 2,  
'yeah': 75,  
'ebay': 1,  
'gags': 1,  
'drew': 2,  
'whoa': 2,

'form': 1,  
'card': 1,  
'mofo': 1,  
'dogs': 1,  
'chat': 142,  
'time': 50,  
'buff': 1,  
'U134': 1,  
'vote': 3,  
'1200': 1,  
'sooo': 2,  
'Rang': 1,  
'U132': 10,  
'oooh': 1,  
'jude': 1,  
'said': 23,  
'YOUR': 2,  
'pwns': 1,  
'woot': 4,  
'U114': 25,  
'fock': 1,  
'skin': 3,  
'U138': 2,  
'York': 1,  
'walk': 3,  
'fake': 1,  
'corn': 2,  
'((((': 3,  
'ussy': 1,  
'jeff': 1,  
'mame': 1,  
'ques': 1,  
'any1': 2,  
'Hail': 1,  
'!!!!': 18,  
'ghet': 1,  
'push': 1,  
'flow': 2,  
'miss': 5,  
'seee': 1,  
'ride': 6,  
'Ummm': 2,  
'ROOM': 4,  
'?!?!': 2,  
'gimp': 2,  
'Chat': 1,  
'U122': 42,  
'herd': 1,



'Take': 1,  
'clay': 1,  
'army': 3,  
'HOTT': 1,  
'U156': 35,  
'LOUD': 1,  
'xbox': 5,  
'menu': 1,  
'guts': 1,  
'cock': 1,  
'crop': 1,  
'felt': 5,  
'mauh': 1,  
'Tide': 1,  
'NICK': 24,  
'U101': 10,  
'WHOA': 1,  
'scuk': 1,  
'Paul': 1,  
'east': 1,  
'feat': 1,  
'blue': 8,  
'STOP': 2,  
'none': 4,  
'tail': 1,  
'seen': 35,  
'samn': 1,  
'rain': 3,  
'cams': 1,  
'sean': 1,  
'!!!!.': 2,  
'bred': 1,  
'chik': 1,  
'<~~~': 1,  
'guyz': 3,  
'U126': 4,  
'Deep': 1,  
'CAPS': 1,  
'drop': 2,  
'soon': 9,  
'rofl': 2,  
'once': 10,  
'plus': 1,  
'U219': 4,  
'KNOW': 1,  
'Born': 1,  
'date': 4,  
'LoVe': 1,

'rich': 2,  
'porn': 2,  
'thot': 1,  
'U129': 6,  
'VBox': 1,  
'does': 35,  
'near': 13,  
'Mine': 1,  
'able': 1,  
'kmph': 1,  
'king': 1,  
'TEXT': 1,  
'went': 8,  
'bugs': 1,  
'toop': 1,  
'seat': 1,  
'laid': 1,  
'urls': 1,  
'seth': 1,  
'wher': 1,  
'4:03': 1,  
'fall': 5,  
'safe': 1,  
'fire': 5,  
'sang': 8,  
'Lies': 2,  
'help': 10,  
'know': 103,  
'haha': 44,  
'Heyy': 2,  
'Uhhh': 1,  
'both': 5,  
'GOOD': 1,  
'foot': 2,  
'wood': 1,  
'crib': 1,  
'icky': 1,  
'hazy': 1,  
'U117': 4,  
'gooo': 1,  
'grrr': 4,  
'seem': 6,  
'They': 6,  
'ruff': 1,  
'Rock': 1,  
'wall': 5,  
'knee': 1,  
'dumb': 2,

'only': 33,  
'U197': 6,  
'ROFL': 1,  
'numb': 1,  
'hmpH': 2,  
'plow': 1,  
'rubs': 3,  
'guys': 58,  
'with': 152,  
'mmmm': 4,  
'pair': 1,  
'city': 2,  
'sink': 1,  
'U154': 4,  
'U137': 1,  
'Evil': 1,  
'talk': 56,  
'thru': 6,  
'neck': 9,  
'wife': 8,  
'crap': 9,  
'... ': 11,  
'gift': 1,  
'nose': 5,  
'Nova': 1,  
'2DAY': 1,  
'WILL': 1,  
'knew': 6,  
'cook': 1,  
'make': 44,  
'lcos': 1,  
'U113': 1,  
'sexy': 23,  
'ugly': 4,  
'Hard': 1,  
'ther': 1,  
'goes': 14,  
'wont': 11,  
'gray': 1,  
'past': 2,  
'rest': 4,  
'size': 1,  
'than': 12,  
'woah': 1,  
'note': 3,  
'from': 92,  
'saME': 1,  
'Elle': 1,

'ladz': 1,  
'luvs': 2,  
'cepn': 1,  
'mami': 1,  
'Love': 2,  
'tory': 1,  
'U103': 6,  
'phil': 2,  
'Ctrl': 1,  
'GUYS': 1,  
'dude': 7,  
'THEY': 1,  
'shut': 14,  
'1996': 2,  
'mang': 1,  
'herE': 1,  
'City': 1,  
'Very': 1,  
'each': 4,  
'waht': 1,  
'lapd': 1,  
'Pour': 1,  
'ribs': 1,  
'quit': 4,  
'U148': 3,  
'Even': 1,  
'Male': 1,  
'Away': 1,  
'sext': 1,  
'Dude': 1,  
'frst': 1,  
'bell': 1,  
'docs': 1,  
'lool': 1,  
'hint': 2,  
'Sexy': 1,  
'!???': 1,  
'TALK': 1,  
'nods': 1,  
'book': 9,  
'orta': 1,  
'tenn': 1,  
'??!!': 2,  
'hair': 17,  
'chit': 1,  
'whos': 6,  
'acid': 1,  
'1.99': 1,

'here': 181,  
'wild': 1,  
'DING': 3,  
'U109': 3,  
'Tisk': 2,  
'meet': 5,  
'hawt': 3,  
'road': 3,  
'well': 81,  
'Slip': 1,  
'U142': 6,  
'much': 40,  
'deep': 2,  
'rang': 1,  
'anti': 1,  
'need': 43,  
'yess': 1,  
'bare': 3,  
'mine': 9,  
'goin': 6,  
'boed': 1,  
'hill': 1,  
'wait': 14,  
'U147': 1,  
'U153': 3,  
'U165': 6,  
'lois': 1,  
'this': 86,  
'yard': 2,  
'Judy': 1,  
'FACE': 1,  
'MODE': 41,  
'What': 18,  
'kind': 2,  
'were': 38,  
'wind': 1,  
'addy': 1,  
'mahn': 1,  
'####': 5,  
'Have': 5,  
'ogan': 1,  
'limp': 2,  
'babi': 2,  
'Here': 2,  
'U106': 3,  
'hour': 9,  
'gals': 1,  
'O.k.': 2,

'yeee': 1,  
'bowl': 1,  
'fool': 5,  
'town': 3,  
'!...': 1,  
'PART': 1016,  
'very': 19,  
'sexs': 1,  
'cyas': 1,  
'asss': 1,  
'JUST': 1,  
'lick': 5,  
'dark': 1,  
'cops': 1,  
'kent': 4,  
'When': 5,  
'U108': 7,  
'Help': 1,  
'pfft': 4,  
'U105': 35,  
'cute': 4,  
'soul': 5,  
'mode': 1,  
'FINE': 1,  
'))))': 6,  
'Awww': 1,  
'guns': 1,  
'dies': 1,  
'Gosh': 2,  
'hooo': 1,  
'cost': 2,  
'weed': 1,  
'MRIs': 1,  
'U144': 8,  
'cure': 1,  
'stop': 19,  
'ages': 2,  
'abou': 1,  
'puke': 1,  
'live': 20,  
'U130': 4,  
'name': 27,  
'SSRI': 1,  
'hard': 10,  
'argh': 2,  
'fish': 1,  
'fawk': 1,  
'Type': 1,

'kids': 11,  
'tell': 27,  
'SEEN': 1,  
'yes.': 1,  
'Maps': 1,  
'ears': 2,  
'otay': 1,  
'died': 3,  
'cash': 2,  
'five': 2,  
'JOIN': 1021,  
'used': 12,  
'hola': 3,  
'Days': 2,  
'room': 98,  
'Show': 1,  
'sayn': 1,  
'whys': 1,  
'SExy': 1,  
'Rick': 1,  
'opps': 2,  
'PMSL': 1,  
'givs': 1,  
'four': 1,  
'duet': 1,  
'Does': 1,  
'45.5': 1,  
'Last': 6,  
'ways': 4,  
'Phil': 1,  
'into': 11,  
'till': 5,  
'fast': 7,  
'will': 40,  
'out.': 1,  
'Fort': 1,  
'half': 3,  
'nice': 52,  
'ex's': 2,  
'Just': 2,  
'raed': 1,  
'piff': 3,  
'song': 36,  
'sent': 1,  
'Swim': 1,  
'"...': 1,  
'pies': 2,  
'move': 3,

'LAsT': 1,  
'wOOt': 2,  
'Sat.': 1,  
'ahah': 1,  
'turn': 4,  
'such': 9,  
'many': 23,  
'Were': 1,  
'firs': 1,  
'hugs': 18,  
'RN's': 1,  
'haaa': 1,  
'fits': 2,  
'mins': 2,  
'yesh': 1,  
'Over': 1,  
'ever': 22,  
'feet': 5,  
'1900': 1,  
'ok'd': 1,  
'hold': 2,  
'<3's': 1,  
'perk': 1,  
'Your': 3,  
'brwn': 1,  
'syck': 1,  
'quiz': 1,  
'Dood': 1,  
'dork': 1,  
'year': 11,  
'U542': 1,  
'nope': 13,  
'THAT': 3,  
'soft': 3,  
'????': 17,  
'whud': 2,  
'eeww': 1,  
'rape': 1,  
'bois': 1,  
'Drop': 1,  
'U133': 4,  
'lies': 2,  
'gets': 12,  
'6:38': 1,  
'HAVE': 2,  
'drug': 1,  
'KoOL': 2,  
'idnt': 1,



'adds': 2,  
'U136': 3,  
'sore': 2,  
'thnx': 1,  
'Stop': 2,  
'1980': 1,  
'sum1': 1,  
'Cool': 2,  
'Hiya': 7,  
'sand': 2,  
'zone': 2,  
'Haha': 1,  
'ohhh': 5,  
'DONT': 2,  
'Rush': 1,  
'U104': 17,  
'barn': 1,  
'days': 9,  
'wubs': 1,  
'kewl': 2,  
'hott': 2,  
'Lord': 1,  
'Okay': 2,  
'aunt': 2,  
'U155': 2,  
'1299': 1,  
'2006': 3,  
'idea': 5,  
'QUIT': 1,  
'U100': 2,  
'aint': 9,  
'bied': 1,  
'asks': 1,  
'brat': 1,  
'rock': 9,  
'offa': 1,  
'lust': 1,  
'<---': 6,  
'lung': 1,  
'tyvm': 2,  
'U170': 2,  
'bite': 2,  
'U128': 5,  
'your': 137,  
'lyin': 1,  
'best': 9,  
'hide': 1,  
'deop': 3,

'bein': 1,  
'YALL': 1,  
'lead': 3,  
'they': 77,  
'slow': 3,  
'brad': 1,  
'prob': 1,  
'rent': 2,  
'Seee': 1,  
'cool': 24,  
'GrlZ': 1,  
'lake': 1,  
'Long': 1,  
'down': 24,  
'site': 1,  
'mite': 1,  
'hurr': 1,  
'show': 10,  
'hall': 2,  
'smax': 1,  
'soup': 1,  
'rush': 1,  
'wrek': 1,  
'done': 6,  
'shit': 17,  
'wazz': 3,  
'sell': 2,  
'choc': 1,  
'mena': 1,  
'COME': 1,  
'dump': 1,  
'keys': 1,  
'yall': 11,  
'star': 1,  
'U111': 2,  
'pray': 1,  
'grin': 1,  
'kina': 1,  
'give': 19,  
'kept': 1,  
'span': 1,  
'U119': 11,  
'lots': 1,  
'hand': 7,  
'lets': 15,  
'hang': 5,  
'thah': 1,  
'spat': 1,

'Hero': 1,  
'peek': 1,  
'U116': 6,  
'beam': 1,  
'wear': 7,  
'doll': 2,  
'Heya': 1,  
'sets': 1,  
'enuf': 1,  
'caan': 1,  
'tick': 2,  
'SIZE': 1,  
'cast': 2,  
'Kold': 1,  
'U520': 6,  
'tjhe': 1,  
'have': 164,  
'pasa': 1,  
'bone': 1,  
'yawn': 3,  
'howz': 2,  
'Boyz': 1,  
'surf': 1,  
'rats': 1,  
'tall': 1,  
'hong': 1,  
'blew': 2,  
'gold': 3,  
'sing': 6,  
'eeek': 1,  
'From': 2,  
'EVEN': 1,  
'toes': 3,  
'prep': 1,  
'Girl': 1,  
'Tell': 2,  
'toss': 3,  
'10th': 1,  
'Liam': 10,  
'U150': 1,  
'okey': 1,  
'jush': 1,  
'Iowa': 1,  
'sigh': 4,  
'food': 6,  
'game': 16,  
'grea': 1,  
'wide': 1,

```

        'Life': 1,
        'snow': 3,
        'geez': 1,
        'chip': 2,
        'heal': 2,
        ...})

```

```

In [14]: import operator
         frecuencia_ordenada_palabras = sorted(frecuencia_palabras.items(), key=ope
         frecuencia_ordenada_palabras

```

```

Out[14]: [('o.k.', 1),
          ('U137', 1),
          ('U149', 1),
          ('rang', 1),
          ('U158', 1),
          ('ally', 1),
          ('gift', 1),
          ('z-ro', 1),
          ('2DAY', 1),
          ('mauh', 1),
          ('tere', 1),
          ('givs', 1),
          ('Oops', 1),
          ('pull', 1),
          ('cook', 1),
          ('Tide', 1),
          ('peel', 1),
          ('yw's", 1),
          ('yess', 1),
          ('18ST', 1),
          ('Hard', 1),
          ('ther', 1),
          ('base', 1),
          ('MUAH', 1),
          ('offa', 1),
          ('aime', 1),
          ('size', 1),
          ('1985', 1),
          ('hong', 1),
          ('boot', 1),
          ('Save', 1),
          ('woah', 1),
          ('tenn', 1),
          ('saME', 1),
          ('Elle', 1),
          ('ladz', 1),
          ('cepn', 1),

```

('vent', 1),  
('mami', 1),  
('Mono', 1),  
('duet', 1),  
('ribs', 1),  
('Ctrl', 1),  
('Matt', 1),  
('10th', 1),  
('FACE', 1),  
('este', 1),  
('calm', 1),  
('U147', 1),  
('gray', 1),  
('Lion', 1),  
('herE', 1),  
('Very', 1),  
('waht', 1),  
('yout', 1),  
('lapd', 1),  
('Pour', 1),  
('Look', 1),  
('HERE', 1),  
('AKST', 1),  
('teck', 1),  
('Room', 1),  
('poem', 1),  
('cuss', 1),  
('tlak', 1),  
('Even', 1),  
('sips', 1),  
('vega', 1),  
('sext', 1),  
('yoll', 1),  
('Dude', 1),  
(':o \*', 1),  
('twit', 1),  
('frst', 1),  
('bell', 1),  
('lool', 1),  
('wins', 1),  
('pigs', 1),  
('!???', 1),  
('wuts', 1),  
('nods', 1),  
('orta', 1),  
('push', 1),  
('Kewl', 1),  
('chit', 1),

('6:41', 1),  
('acid', 1),  
('Good', 1),  
('Chop', 1),  
('see', 1),  
('wind', 1),  
('1.99', 1),  
('goof', 1),  
('hots', 1),  
('daft', 1),  
('NAME', 1),  
('worl', 1),  
('whou', 1),  
('paid', 1),  
('addy', 1),  
('shup', 1),  
('Home', 1),  
('tooo', 1),  
('Teck', 1),  
('Evil', 1),  
('anti', 1),  
('puts', 1),  
('mahn', 1),  
('boed', 1),  
('hill', 1),  
('dick', 1),  
('sean', 1),  
('Tina', 1),  
('Take', 1),  
('lois', 1),  
('poop', 1),  
('Judy', 1),  
('Male', 1),  
('wore', 1),  
('Swim', 1),  
('ogan', 1),  
('9.53', 1),  
('See', 1),  
('soup', 1),  
('pork', 1),  
('Over', 1),  
('geez', 1),  
('wean', 1),  
('7:45', 1),  
('salt', 1),  
('exit', 1),  
('gals', 1),  
('U181', 1),

('Holy', 1),  
('giva', 1),  
('nawp', 1),  
('yeee', 1),  
('bowl', 1),  
('hgey', 1),  
('!...', 1),  
('Haha', 1),  
('Does', 1),  
('hiom', 1),  
('poot', 1),  
('sexs', 1),  
('cyas', 1),  
('U164', 1),  
('dman', 1),  
('JUST', 1),  
('dark', 1),  
('pour', 1),  
('cops', 1),  
('1.98', 1),  
('dyed', 1),  
('poll', 1),  
('moms', 1),  
('scum', 1),  
('Wyte', 1),  
('howl', 1),  
('Help', 1),  
('98.6', 1),  
('nerd', 1),  
('ussy', 1),  
('bacl', 1),  
('mode', 1),  
('FINE', 1),  
('Awww', 1),  
('dies', 1),  
('Bone', 1),  
('Fade', 1),  
('nuff', 1),  
('grew', 1),  
('weed', 1),  
('MRIs', 1),  
('bird', 1),  
('cure', 1),  
('sori', 1),  
('4.20', 1),  
('abou', 1),  
('slip', 1),  
('puke', 1),

('Were', 1),  
('Rofl', 1),  
('cams', 1),  
('SSRI', 1),  
('fish', 1),  
('fawk', 1),  
('Type', 1),  
('Tiff', 1),  
('HAHA', 1),  
('yes.', 1),  
('Maps', 1),  
('Need', 1),  
('otay', 1),  
('100%', 1),  
('GUYS', 1),  
('LIVE', 1),  
('Hill', 1),  
('WILL', 1),  
('pmsl', 1),  
('bull', 1),  
('Show', 1),  
('sayn', 1),  
('west', 1),  
('whys', 1),  
('SExy', 1),  
('Rick', 1),  
('dojn', 1),  
('Prof', 1),  
('four', 1),  
('tory', 1),  
('45.5', 1),  
('byes', 1),  
('HUGE', 1),  
('asss', 1),  
('bong', 1),  
('lube', 1),  
('uyes', 1),  
('rush', 1),  
('TYPR', 1),  
('wrap', 1),  
('junk', 1),  
('OOPS', 1),  
('raed', 1),  
('sent', 1),  
('jeep', 1),  
('dust', 1),  
('mono', 1),  
('lala', 1),



('Sat.', 1),  
('ahah', 1),  
('ebay', 1),  
('cock', 1),  
('grin', 1),  
('coem', 1),  
('RN's', 1),  
('haaa', 1),  
('mena', 1),  
('rose', 1),  
('jeff', 1),  
('Time', 1),  
('benz', 1),  
('Road', 1),  
('pure', 1),  
('<3's', 1),  
('firs', 1),  
('syck', 1),  
('Damn', 1),  
('1900', 1),  
('dork', 1),  
('Care', 1),  
('U542', 1),  
('ok'd', 1),  
('HALO', 1),  
('eeww', 1),  
('THEY', 1),  
('bois', 1),  
('tits', 1),  
('Drop', 1),  
('bloee', 1),  
('Hott', 1),  
('tiff', 1),  
('6:38', 1),  
('bomb', 1),  
('drug', 1),  
('Kick', 1),  
('LONG', 1),  
('Phil', 1),  
('idnt', 1),  
('thnx', 1),  
('tape', 1),  
('1980', 1),  
('sum1', 1),  
('Mine', 1),  
('prep', 1),  
('lcos', 1),  
('able', 1),

('jail', 1),  
('quiz', 1),  
('Rush', 1),  
('barn', 1),  
('U113', 1),  
('Lord', 1),  
('1299', 1),  
('febe', 1),  
('U134', 1),  
('QUIT', 1),  
('mang', 1),  
('post', 1),  
('asks', 1),  
('brat', 1),  
('samn', 1),  
('pope', 1),  
('slam', 1),  
('lust', 1),  
('City', 1),  
('lyin', 1),  
('hide', 1),  
('Jess', 1),  
('docs', 1),  
('prob', 1),  
('DAMN', 1),  
('bein', 1),  
('YALL', 1),  
('Slip', 1),  
('tend', 1),  
('brad', 1),  
('Sexy', 1),  
('64.8', 1),  
('outa', 1),  
('span', 1),  
('body', 1),  
('GrlZ', 1),  
('lake', 1),  
('Long', 1),  
('Mary', 1),  
('guns', 1),  
('site', 1),  
('mite', 1),  
('dint', 1),  
('hurr', 1),  
('smax', 1),  
('Cute', 1),  
('allo', 1),  
('wrek', 1),

('4:03', 1),  
('LAsT', 1),  
('COME', 1),  
('lisa', 1),  
('None', 1),  
('Dood', 1),  
('KNOW', 1),  
('dump', 1),  
('keys', 1),  
('star', 1),  
('TIME', 1),  
('Hold', 1),  
('Hugs', 1),  
('http', 1),  
('Yoko', 1),  
('pray', 1),  
('kina', 1),  
('tart', 1),  
('kept', 1),  
('lots', 1),  
('Troy', 1),  
('spat', 1),  
('Hero', 1),  
('peek', 1),  
('dotn', 1),  
('beam', 1),  
("pm's", 1),  
('heee', 1),  
('sets', 1),  
('enuf', 1),  
('Then', 1),  
('caan', 1),  
('disc', 1),  
('gosh', 1),  
('evah', 1),  
('ROFL', 1),  
('anal', 1),  
('Heya', 1),  
('ssid', 1),  
('Kold', 1),  
('tjhe', 1),  
('numb', 1),  
('Will', 1),  
('crib', 1),  
('pasa', 1),  
('seat', 1),  
('bone', 1),  
('Ohhh', 1),

('mkay', 1),  
('Turn', 1),  
('SIZE', 1),  
('Boyz', 1),  
('surf', 1),  
('fear', 1),  
('rats', 1),  
('tall', 1),  
('Kiss', 1),  
('Werd', 1),  
('bike', 1),  
('Away', 1),  
('Fort', 1),  
('spit', 1),  
('eeek', 1),  
('EVEN', 1),  
('ohwa', 1),  
('News', 1),  
('Ruth', 1),  
('39.3', 1),  
('Girl', 1),  
('gear', 1),  
('toke', 1),  
('3:45', 1),  
('yell', 1),  
('U150', 1),  
('okey', 1),  
('jush', 1),  
('Jane', 1),  
('TALK', 1),  
('Iowa', 1),  
('grea', 1),  
('wide', 1),  
('soda', 1),  
('Life', 1),  
('WHEN', 1),  
('lung', 1),  
('9:10', 1),  
('bred', 1),  
('Vvil', 1),  
('CALI', 1),  
('Food', 1),  
('1200', 1),  
('coat', 1),  
('Nooo', 1),  
('whoo', 1),  
('thje', 1),  
('sign', 1),

('Reub', 1),  
('"...', 1),  
('mame', 1),  
('pimp', 1),  
('akon', 1),  
('serg', 1),  
('test', 1),  
('whew', 1),  
('Uhhh', 1),  
('loud', 1),  
('gags', 1),  
('PM's', 1),  
('U118', 1),  
('sexi', 1),  
('form', 1),  
('card', 1),  
('icky', 1),  
('mofo', 1),  
('noth', 1),  
('dogs', 1),  
('mark', 1),  
('2Pac', 1),  
('buff', 1),  
('fair', 1),  
('wubs', 1),  
('Rang', 1),  
('rape', 1),  
('oooh', 1),  
('jude', 1),  
('pwns', 1),  
('hogs', 1),  
('MORE', 1),  
('Been', 1),  
('vamp', 1),  
('fock', 1),  
('2:55', 1),  
('Back', 1),  
('heat', 1),  
('ooer', 1),  
('nada', 1),  
('kong', 1),  
('fake', 1),  
('pine', 1),  
('xmas', 1),  
('Nope', 1),  
('lazy', 1),  
('Hail', 1),  
('Rock', 1),

('ghet', 1),  
('nude', 1),  
('SEEN', 1),  
('York', 1),  
('jack', 1),  
('Rule', 1),  
('ques', 1),  
('guts', 1),  
('caca', 1),  
('woof', 1),  
('Meep', 1),  
('Chat', 1),  
('LATE', 1),  
('draw', 1),  
('wire', 1),  
('tips', 1),  
('herd', 1),  
('gees', 1),  
('Eyes', 1),  
('98.5', 1),  
('clay', 1),  
('loss', 1),  
('ltnc', 1),  
('yesh', 1),  
('boom', 1),  
('LOUD', 1),  
('cums', 1),  
('menu', 1),  
('crop', 1),  
('bust', 1),  
('t he', 1),  
('thah', 1),  
('WHOA', 1),  
('scuk', 1),  
('Paul', 1),  
('east', 1),  
('bied', 1),  
('feat', 1),  
('owww', 1),  
('tail', 1),  
('dirt', 1),  
('hooo', 1),  
('wild', 1),  
('kold', 1),  
('GIRL', 1),  
('Talk', 1),  
('mess', 1),  
('halo', 1),

('chik', 1),  
('<~~~', 1),  
('Deep', 1),  
('CAPS', 1),  
('6:53', 1),  
('ELSE', 1),  
('dawg', 1),  
('West', 1),  
('plus', 1),  
('nads', 1),  
('Nova', 1),  
('Kids', 1),  
('Born', 1),  
('waaa', 1),  
('King', 1),  
('out.', 1),  
('LoVe', 1),  
('6:51', 1),  
('seth', 1),  
('poof', 1),  
('.op.', 1),  
('perk', 1),  
('ruff', 1),  
('thot', 1),  
('hawT', 1),  
('VBox', 1),  
('pm'n', 1),  
('inch', 1),  
('Heys', 1),  
('kmph', 1),  
('king', 1),  
('TEXT', 1),  
('bugs', 1),  
('toop', 1),  
('laid', 1),  
('Eggs', 1),  
('Kent', 1),  
('urls', 1),  
('wher', 1),  
('1930', 1),  
('Joey', 1),  
('choc', 1),  
('safe', 1),  
('SOME', 1),  
('Hand', 1),  
('brwn', 1),  
('page', 1),  
('tthe', 1),

('GOOD', 1),  
('scar', 1),  
('outs', 1),  
('wood', 1),  
('hazy', 1),  
('gooo', 1),  
('term', 1),  
('gret', 1),  
('Dawn', 1),  
('nawt', 1),  
('Came', 1),  
('knee', 1),  
('arms', 1),  
('U143', 1),  
('HOTT', 1),  
('NTMN', 1),  
('plow', 1),  
("it's", 1),  
('PMSL', 1),  
('pair', 1),  
('blow', 1),  
('sink', 1),  
('AWAY', 1),  
('lol.', 1),  
('mike', 2),  
('wOOt', 2),  
('past', 2),  
('side', 2),  
('john', 2),  
('phil', 2),  
('1996', 2),  
('U112', 2),  
('cars', 2),  
('ciao', 2),  
('mass', 2),  
('>:->', 2),  
('club', 2),  
('ewww', 2),  
('spin', 2),  
('cell', 2),  
('hint', 2),  
('n9ne', 2),  
("ex's", 2),  
('deep', 2),  
('hits', 2),  
('kind', 2),  
('limp', 2),  
('babi', 2),



('Here', 2),  
('twin', 2),  
('O.k.', 2),  
('Lmao', 2),  
('ltns', 2),  
('temp', 2),  
('Drew', 2),  
('golf', 2),  
('Ahhh', 2),  
('pool', 2),  
('flaw', 2),  
('cost', 2),  
('ages', 2),  
('grrl', 2),  
('argh', 2),  
('WITH', 2),  
('ears', 2),  
('cash', 2),  
('five', 2),  
('hummm', 2),  
('rofl', 2),  
('Days', 2),  
('Dang', 2),  
('opps', 2),  
('deaf', 2),  
('porn', 2),  
('ohio', 2),  
('spot', 2),  
('lawl', 2),  
('Just', 2),  
('pies', 2),  
('newp', 2),  
('U155', 2),  
('mins', 2),  
('park', 2),  
('hold', 2),  
('Gosh', 2),  
('yard', 2),  
('sort', 2),  
('whud', 2),  
('HAVE', 2),  
('adds', 2),  
('<<<<', 2),  
('sore', 2),  
('Stop', 2),  
('sand', 2),  
('zone', 2),  
('DONT', 2),

('kewl', 2),  
('hott', 2),  
('Okay', 2),  
('bear', 2),  
('luvs', 2),  
('root', 2),  
('moon', 2),  
('tyvm', 2),  
('U170', 2),  
('bite', 2),  
('typo', 2),  
('Sure', 2),  
('shop', 2),  
('hall', 2),  
('Live', 2),  
('plan', 2),  
('sell', 2),  
('born', 2),  
('Ohio', 2),  
('burp', 2),  
('mama', 2),  
('doll', 2),  
('tick', 2),  
('trip', 2),  
('cast', 2),  
('wooo', 2),  
('howz', 2),  
('tock', 2),  
('From', 2),  
('gays', 2),  
('Tell', 2),  
('wats', 2),  
('DOES', 2),  
('blew', 2),  
('heal', 2),  
('yeas', 2),  
('Nice', 2),  
('drew', 2),  
('whoa', 2),  
('sooo', 2),  
('clue', 2),  
('YOUR', 2),  
('High', 2),  
('whip', 2),  
('U138', 2),  
('John', 2),  
('corn', 2),  
('tisk', 2),

('Down', 2),  
('any1', 2),  
('flow', 2),  
('???!', 2),  
('rent', 2),  
('Ummm', 2),  
('haze', 2),  
('gimp', 2),  
('Poor', 2),  
('cmon', 2),  
('city', 2),  
('?!?!', 2),  
('Love', 2),  
('Lets', 2),  
('STOP', 2),  
('Come', 2),  
('uses', 2),  
('eats', 2),  
('!!!.', 2),  
('U175', 2),  
('drop', 2),  
('FROM', 2),  
('fits', 2),  
('Tisk', 2),  
('rich', 2),  
('U172', 2),  
('aunt', 2),  
('meat', 2),  
('Cool', 2),  
('area', 2),  
('KoOL', 2),  
('Lies', 2),  
('Heyy', 2),  
('U111', 2),  
('U190', 2),  
('foot', 2),  
('chip', 2),  
('<333', 2),  
('dumb', 2),  
('U100', 2),  
('eric', 2),  
('NONE', 2),  
('hmpH', 2),  
('lies', 2),  
('slap', 3),  
('ahem', 3),  
('note', 3),  
('imma', 3),

('gawd', 3),  
('U148', 3),  
('Only', 3),  
('nana', 3),  
('yawn', 3),  
('DING', 3),  
('U109', 3),  
('hawt', 3),  
('road', 3),  
('lead', 3),  
('army', 3),  
('hick', 3),  
('U153', 3),  
('U145', 3),  
('bare', 3),  
('hail', 3),  
('swim', 3),  
('U106', 3),  
('town', 3),  
('yada', 3),  
('Wind', 3),  
('rubs', 3),  
('hurt', 3),  
('CHAT', 3),  
('skin', 3),  
('died', 3),  
('hola', 3),  
('orgy', 3),  
('half', 3),  
('piff', 3),  
('itch', 3),  
('move', 3),  
('deal', 3),  
('Your', 3),  
('THAT', 3),  
('soft', 3),  
('mary', 3),  
('U136', 3),  
('Elev', 3),  
('2006', 3),  
('deop', 3),  
('slow', 3),  
('wash', 3),  
('butt', 3),  
('band', 3),  
('ello', 3),  
('gold', 3),  
('toes', 3),

('toss', 3),  
('U141', 3),  
('Same', 3),  
('bend', 3),  
('jump', 3),  
('clap', 3),  
('isnt', 3),  
('hiii', 3),  
('elle', 3),  
('roof', 3),  
('walk', 3),  
('ring', 3),  
('((((', 3),  
('amen', 3),  
('hank', 3),  
('vote', 3),  
('rain', 3),  
('guyz', 3),  
('wazz', 3),  
('wine', 3),  
('snow', 3),  
('tune', 3),  
('U163', 3),  
('AKDT', 3),  
('hump', 3),  
('wack', 3),  
('holy', 4),  
('ugly', 4),  
('rest', 4),  
('pass', 4),  
('quit', 4),  
('ummm', 4),  
('line', 4),  
('puff', 4),  
('cute', 4),  
('high', 4),  
('U130', 4),  
('shes', 4),  
('ways', 4),  
('door', 4),  
('kent', 4),  
('U133', 4),  
('turn', 4),  
('mmmm', 4),  
(',,,,', 4),  
('pfft', 4),  
('beat', 4),  
('each', 4),

('shot', 4),  
('U146', 4),  
('U123', 4),  
('lord', 4),  
('jerk', 4),  
('ones', 4),  
('open', 4),  
('huge', 4),  
('sigh', 4),  
('hook', 4),  
('U196', 4),  
('lame', 4),  
('Like', 4),  
('team', 4),  
('woot', 4),  
('U820', 4),  
('U989', 4),  
('ouch', 4),  
('ROOM', 4),  
('U988', 4),  
('U819', 4),  
('none', 4),  
('fart', 4),  
('U126', 4),  
('U219', 4),  
('date', 4),  
('glad', 4),  
('self', 4),  
('U117', 4),  
('grrr', 4),  
('pain', 4),  
('U154', 4),  
('evil', 4),  
('nose', 5),  
('wish', 5),  
('land', 5),  
('kool', 5),  
('boys', 5),  
('####', 5),  
('Have', 5),  
('late', 5),  
('fool', 5),  
('lick', 5),  
('When', 5),  
('soul', 5),  
('legs', 5),  
('till', 5),  
('luck', 5),

('feet', 5),  
('warm', 5),  
('ohhh', 5),  
('fall', 5),  
('idea', 5),  
('cali', 5),  
('caps', 5),  
('U128', 5),  
('easy', 5),  
('meds', 5),  
('hang', 5),  
('beer', 5),  
('joke', 5),  
('roll', 5),  
('also', 5),  
('Lime', 5),  
('miss', 5),  
('heck', 5),  
('xbox', 5),  
('boss', 5),  
('yoko', 5),  
('felt', 5),  
('came', 5),  
('lose', 5),  
('meet', 5),  
('fire', 5),  
('both', 5),  
('wall', 5),  
('thru', 6),  
('knew', 6),  
('most', 6),  
('gone', 6),  
('Last', 6),  
('U103', 6),  
('whos', 6),  
('U142', 6),  
('goin', 6),  
('U165', 6),  
('list', 6),  
(''))))', 6),  
('part', 6),  
('oops', 6),  
('send', 6),  
('Song', 6),  
('pink', 6),  
('<---', 6),  
('comp', 6),  
('done', 6),

('U116', 6),  
('U520', 6),  
('next', 6),  
('blah', 6),  
('sing', 6),  
('poor', 6),  
('food', 6),  
('U120', 6),  
('seem', 6),  
('ride', 6),  
('sock', 6),  
('U129', 6),  
('ball', 6),  
('U197', 6),  
('main', 6),  
('They', 6),  
('dude', 7),  
('alot', 7),  
('U108', 7),  
('fast', 7),  
('kick', 7),  
('Hiya', 7),  
('took', 7),  
('hand', 7),  
('wear', 7),  
('U169', 7),  
('dear', 7),  
('ahhh', 7),  
('rule', 7),  
('okay', 7),  
('That', 7),  
('wife', 8),  
('lady', 8),  
('made', 8),  
('case', 8),  
('U139', 8),  
('U144', 8),  
('word', 8),  
('heyy', 8),  
('suck', 8),  
('went', 8),  
('wana', 8),  
('hows', 8),  
('sang', 8),  
('blue', 8),  
('says', 8),  
('neck', 9),  
('crap', 9),



```
('; ..', 9),
('runs', 9),
('mine', 9),
...]
```

26. What does the following Python code do? `sum(len(w) for w in text1)` Can you use it to work out the average word length of a text?

```
In [15]: a = sum(len(w) for w in text1)
a
```

```
Out[15]: 999044
```

El total de palabras en text1

```
In [23]: def prome_word_length(text):
          return sum([len(w) for w in text]) / len(text)

          print (prome_word_length(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 [32]: def percent(palabra, texto):
          num_ocurrencias = float(texto.count(palabra))
          num_total = float(len(texto))
          return str(round(100 * (num_ocurrencias / num_total), 2)) + ' %'

          print (percent("whale", text1))
```

```
0.35 %
```

```
In [33]: def percent(palabra, texto):
          num_ocurrencias = float(texto.count(palabra))
          num_total = float(len(texto))
          return str(round(100 * (num_ocurrencias / num_total), 2)) + ' %'

          print (percent("the", text1))
```

```
5.26 %
```

## 1.3 Chapter 2

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

```
In [34]: import nltk
         nltk.corpus.gutenberg.fileids()
```

```
Out [34]: ['austen-emma.txt',
           'austen-persuasion.txt',
           'austen-sense.txt',
           'bible-kjv.txt',
           'blake-poems.txt',
           'bryant-stories.txt',
           'burgess-busterbrown.txt',
           'carroll-alice.txt',
           'chesterton-ball.txt',
           'chesterton-brown.txt',
           'chesterton-thursday.txt',
           'edgeworth-parents.txt',
           'melville-moby_dick.txt',
           'milton-paradise.txt',
           'shakespeare-caesar.txt',
           'shakespeare-hamlet.txt',
           'shakespeare-macbeth.txt',
           'whitman-leaves.txt']
```

```
In [14]: persuasion = nltk.corpus.gutenberg.words('austen-persuasion.txt')
         len(persuasion)
```

```
Out [14]: 98171
```

```
In [15]: len(set(persuasion))
```

```
Out [15]: 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 [16]: nltk.corpus.state_union.fileids()
```

```
Out [16]: ['1945-Truman.txt',
           '1946-Truman.txt',
           '1947-Truman.txt',
           '1948-Truman.txt',
           '1949-Truman.txt',
           '1950-Truman.txt',
           '1951-Truman.txt',
           '1953-Eisenhower.txt',
```

'1954-Eisenhower.txt',  
'1955-Eisenhower.txt',  
'1956-Eisenhower.txt',  
'1957-Eisenhower.txt',  
'1958-Eisenhower.txt',  
'1959-Eisenhower.txt',  
'1960-Eisenhower.txt',  
'1961-Kennedy.txt',  
'1962-Kennedy.txt',  
'1963-Johnson.txt',  
'1963-Kennedy.txt',  
'1964-Johnson.txt',  
'1965-Johnson-1.txt',  
'1965-Johnson-2.txt',  
'1966-Johnson.txt',  
'1967-Johnson.txt',  
'1968-Johnson.txt',  
'1969-Johnson.txt',  
'1970-Nixon.txt',  
'1971-Nixon.txt',  
'1972-Nixon.txt',  
'1973-Nixon.txt',  
'1974-Nixon.txt',  
'1975-Ford.txt',  
'1976-Ford.txt',  
'1977-Ford.txt',  
'1978-Carter.txt',  
'1979-Carter.txt',  
'1980-Carter.txt',  
'1981-Reagan.txt',  
'1982-Reagan.txt',  
'1983-Reagan.txt',  
'1984-Reagan.txt',  
'1985-Reagan.txt',  
'1986-Reagan.txt',  
'1987-Reagan.txt',  
'1988-Reagan.txt',  
'1989-Bush.txt',  
'1990-Bush.txt',  
'1991-Bush-1.txt',  
'1991-Bush-2.txt',  
'1992-Bush.txt',  
'1993-Clinton.txt',  
'1994-Clinton.txt',  
'1995-Clinton.txt',  
'1996-Clinton.txt',  
'1997-Clinton.txt',  
'1998-Clinton.txt',

```

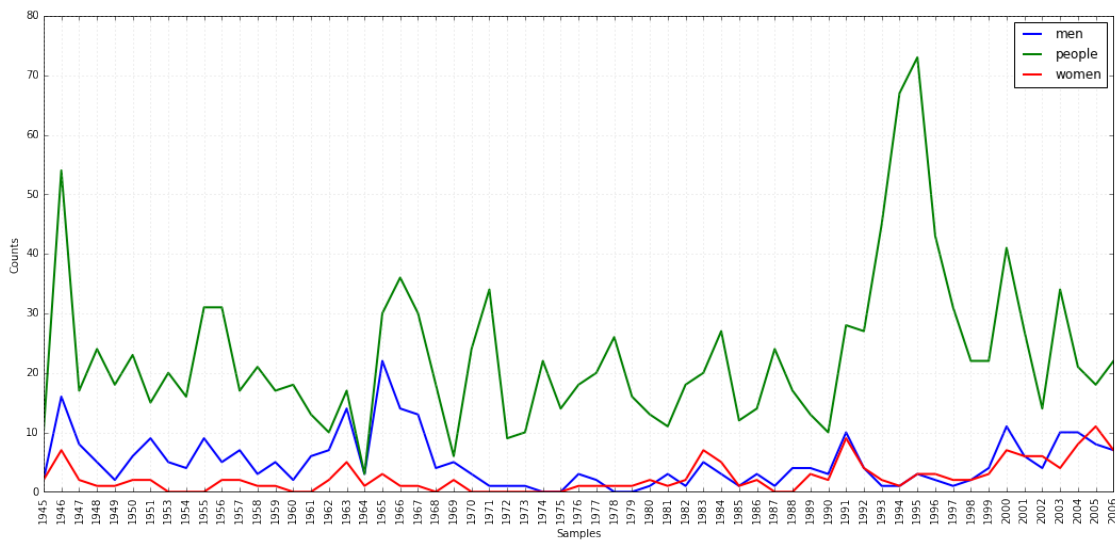
'1999-Clinton.txt',
'2000-Clinton.txt',
'2001-GWBush-1.txt',
'2001-GWBush-2.txt',
'2002-GWBush.txt',
'2003-GWBush.txt',
'2004-GWBush.txt',
'2005-GWBush.txt',
'2006-GWBush.txt']

```

```

In [17]: import nltk
         from nltk.corpus import state_union
         cfd = nltk.ConditionalFreqDist(
             (target, fileid[:4])
             for fileid in state_union.fileids()
             for w in state_union.words(fileid)
             for target in ['men', 'women', 'people']
             if w.lower().startswith(target))
         cfd.plot()

```



```

In [18]: cfd

```

```

Out[18]: ConditionalFreqDist(nltk.probability.FreqDist,
                             {'men': FreqDist({'1945': 2,
                                                '1946': 16,
                                                '1947': 8,
                                                '1948': 5,
                                                '1949': 2,
                                                '1950': 6,

```

'1951': 9,  
'1953': 5,  
'1954': 4,  
'1955': 9,  
'1956': 5,  
'1957': 7,  
'1958': 3,  
'1959': 5,  
'1960': 2,  
'1961': 6,  
'1962': 7,  
'1963': 14,  
'1964': 3,  
'1965': 22,  
'1966': 14,  
'1967': 13,  
'1968': 4,  
'1969': 5,  
'1970': 3,  
'1971': 1,  
'1972': 1,  
'1973': 1,  
'1976': 3,  
'1977': 2,  
'1980': 1,  
'1981': 3,  
'1982': 1,  
'1983': 5,  
'1984': 3,  
'1985': 1,  
'1986': 3,  
'1987': 1,  
'1988': 4,  
'1989': 4,  
'1990': 3,  
'1991': 10,  
'1992': 4,  
'1993': 1,  
'1994': 1,  
'1995': 3,  
'1996': 2,  
'1997': 1,  
'1998': 2,  
'1999': 4,  
'2000': 11,  
'2001': 6,  
'2002': 4,  
'2003': 10,

```

        '2004': 10,
        '2005': 8,
        '2006': 7}},
'people': FreqDist({'1945': 10,
    '1946': 54,
    '1947': 17,
    '1948': 24,
    '1949': 18,
    '1950': 23,
    '1951': 15,
    '1953': 20,
    '1954': 16,
    '1955': 31,
    '1956': 31,
    '1957': 17,
    '1958': 21,
    '1959': 17,
    '1960': 18,
    '1961': 13,
    '1962': 10,
    '1963': 17,
    '1964': 3,
    '1965': 30,
    '1966': 36,
    '1967': 30,
    '1968': 18,
    '1969': 6,
    '1970': 24,
    '1971': 34,
    '1972': 9,
    '1973': 10,
    '1974': 22,
    '1975': 14,
    '1976': 18,
    '1977': 20,
    '1978': 26,
    '1979': 16,
    '1980': 13,
    '1981': 11,
    '1982': 18,
    '1983': 20,
    '1984': 27,
    '1985': 12,
    '1986': 14,
    '1987': 24,
    '1988': 17,
    '1989': 13,
    '1990': 10,

```

```

'1991': 28,
'1992': 27,
'1993': 45,
'1994': 67,
'1995': 73,
'1996': 43,
'1997': 31,
'1998': 22,
'1999': 22,
'2000': 41,
'2001': 27,
'2002': 14,
'2003': 34,
'2004': 21,
'2005': 18,
'2006': 22}),
'women': FreqDist({'1945': 2,
'1946': 7,
'1947': 2,
'1948': 1,
'1949': 1,
'1950': 2,
'1951': 2,
'1956': 2,
'1957': 2,
'1958': 1,
'1959': 1,
'1962': 2,
'1963': 5,
'1964': 1,
'1965': 3,
'1966': 1,
'1967': 1,
'1969': 2,
'1976': 1,
'1977': 1,
'1978': 1,
'1979': 1,
'1980': 2,
'1981': 1,
'1982': 2,
'1983': 7,
'1984': 5,
'1985': 1,
'1986': 2,
'1989': 3,
'1990': 2,
'1991': 9,

```

```

'1992': 4,
'1993': 2,
'1994': 1,
'1995': 3,
'1996': 3,
'1997': 2,
'1998': 2,
'1999': 3,
'2000': 7,
'2001': 6,
'2002': 6,
'2003': 4,
'2004': 8,
'2005': 11,
'2006': 7}}))

```

**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 [25]: from nltk.corpus import reuters
         reuters.categories()

```

```

Out[25]: ['acq',
          'alum',
          'barley',
          'bop',
          'carcass',
          'castor-oil',
          'cocoa',
          'coconut',
          'coconut-oil',
          'coffee',
          'copper',
          'copra-cake',
          'corn',
          'cotton',
          'cotton-oil',
          'cpi',
          'cpu',
          'crude',
          'dfl',
          'dlr',
          'dmk',
          'earn',
          'fuel',
          'gas',
          'gnp',

```



'gold',  
'grain',  
'groundnut',  
'groundnut-oil',  
'heat',  
'hog',  
'housing',  
'income',  
'instal-debt',  
'interest',  
'ipi',  
'iron-steel',  
'jet',  
'jobs',  
'l-cattle',  
'lead',  
'lei',  
'lin-oil',  
'livestock',  
'lumber',  
'meal-feed',  
'money-fx',  
'money-supply',  
'naphtha',  
'nat-gas',  
'nickel',  
'nkr',  
'nzdlr',  
'oat',  
'oilseed',  
'orange',  
'palladium',  
'palm-oil',  
'palmkernel',  
'pet-chem',  
'platinum',  
'potato',  
'propane',  
'rand',  
'rape-oil',  
'rapeseed',  
'reserves',  
'retail',  
'rice',  
'rubber',  
'rye',  
'ship',  
'silver',

```

'sorghum',
'soy-meal',
'soy-oil',
'soybean',
'strategic-metal',
'sugar',
'sun-meal',
'sun-oil',
'sunseed',
'tea',
'tin',
'trade',
'veg-oil',
'wheat',
'wpi',
'yen',
'zinc']

```

```

In [28]: cfd = nltk.ConditionalFreqDist(
          (genre, word)
          for genre in reuters.categories()
          for word in reuters.words(categories=genre))
genres = ['fuel', 'cotton', 'coconut', 'alum']
modals = ['can', 'could', 'may', 'might']
cfd.tabulate(conditions=genres, samples=modals)

```

	can	could	may	might
fuel	2	2	1	1
cotton	3	14	15	8
coconut	3	3	1	0
alum	11	3	9	1