

a. Study of various corpus – Brown, Inaugural, Reuters, UDHR with various methods like fields, raw, words, sent, categories.

b. Study Conditional Frequency Distribution

Code and Output:

```
#Brown_corpus
```

```
import nltk
```

```
from nltk.corpus import brown
```

```
brown.categories()
```

```
['adventure', 'belles_lettres', 'editorial', 'fiction', 'government', 'hobbies', 'humor', 'learned', 'lore', 'mystery', 'news', 'religion', 'reviews', 'romance', 'science_fiction']
```

```
brown.words()
```

```
['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
```

```
brown.words(categories='news')
```

```
['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
```

```
brown.words(categories='adventure')
```

```
['Dan', 'Morgan', 'told', 'himself', 'he', 'would', ...]
```

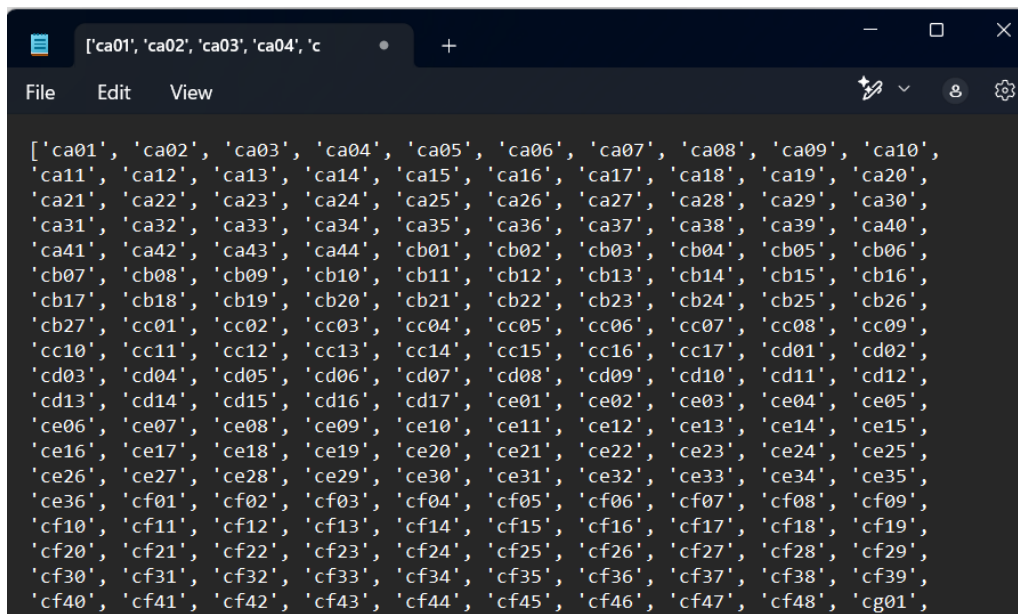
```
brown.words(categories='government')
```

```
['The', 'Office', 'of', 'Business', 'Economics', '(', ...]
```

```
brown.raw()
```

```
this/dt autumn/nn is/bez the/at reverse/nn of/in rosy/jj ./.\nThe/at inventory/nn you/ppss acquired/vbd from/in me/ppp isn't/bez* going/vbg to/to be/be easy/jj to/to move/vb ;/./.\nyou/ppss can't/md* very/ql well/rb sidle/vb up/in to/in people/nns on/in the/at street/nn and/cc ask/vb if/cs they/ppss want/vb to/to buy/vb a/at hot/jj Bodhisattva/np ./.\nAdditionally/rb ,/, since/cs you're/ppss+ber going/vbg to/to be/be hors/fw-rb de/fw-in combat/fw-nn pretty/ql soon/rb with/in sprue/nn ,/, yaws/nns ,/, Delhi/np boil/nn ,/, the/at Granville/np wilt/nn ,/, liver/nn fluke/nn ,/, bilharziasis/nn ,/, and/cc a/at host/nn of/in other/ap complications/nns of/in the/at hex/nn you've/ppss+hv aroused/vbn ,/, you/ppss mustn't/md* expect/vb to/to be/be lionized/vbn socially/rb ./.\nMy/pp$ advice/nn ,/, if/cs you/ppss live/vb long/jj enough/qlp to/to continue/vb your/pp$ vocation/nn ,/, is/bez that/cs the/at next/ap time/nn you're/ppss+ber attracted/vbn by/in the/at exotic/jj ,/, pass/vb it/ppp up/rp --/-- it's/pps+bez nothing/pn but/cc a/at headache/nn ./.\nAs/cs you/ppss can/md count/vb on/in me/ppp to/to do/do the/at same/ap ./.\n\n\n\tCompassionately/rb yours/pp$$ ,/, \n\n\tS./np J./np Perelman/np \n\n\n\trevulsion/nn-hl in/in-hl the/at-hl desert/nn-hl \nthe/at doors/nns of/in the/at D/np-tl train/nn slid/vbd shut/vbn ,/, and/cc as/cs I/ppss dropped/vbd into/in a/at seat/nn and/cc ,/, exhaling/vbg ,/, looked/vbd up/rp across/in the/at aisle/nn ,/, the/at whole/jj aviary/nn in/in my/pp$
```

```
brown.fileids()
```



```
[ 'ca01', 'ca02', 'ca03', 'ca04', 'ca05', 'ca06', 'ca07', 'ca08', 'ca09', 'ca10',
  'ca11', 'ca12', 'ca13', 'ca14', 'ca15', 'ca16', 'ca17', 'ca18', 'ca19', 'ca20',
  'ca21', 'ca22', 'ca23', 'ca24', 'ca25', 'ca26', 'ca27', 'ca28', 'ca29', 'ca30',
  'ca31', 'ca32', 'ca33', 'ca34', 'ca35', 'ca36', 'ca37', 'ca38', 'ca39', 'ca40',
  'ca41', 'ca42', 'ca43', 'ca44', 'cb01', 'cb02', 'cb03', 'cb04', 'cb05', 'cb06',
  'cb07', 'cb08', 'cb09', 'cb10', 'cb11', 'cb12', 'cb13', 'cb14', 'cb15', 'cb16',
  'cb17', 'cb18', 'cb19', 'cb20', 'cb21', 'cb22', 'cb23', 'cb24', 'cb25', 'cb26',
  'cb27', 'cc01', 'cc02', 'cc03', 'cc04', 'cc05', 'cc06', 'cc07', 'cc08', 'cc09',
  'cc10', 'cc11', 'cc12', 'cc13', 'cc14', 'cc15', 'cc16', 'cc17', 'cd01', 'cd02',
  'cd03', 'cd04', 'cd05', 'cd06', 'cd07', 'cd08', 'cd09', 'cd10', 'cd11', 'cd12',
  'cd13', 'cd14', 'cd15', 'cd16', 'cd17', 'ce01', 'ce02', 'ce03', 'ce04', 'ce05',
  'ce06', 'ce07', 'ce08', 'ce09', 'ce10', 'ce11', 'ce12', 'ce13', 'ce14', 'ce15',
  'ce16', 'ce17', 'ce18', 'ce19', 'ce20', 'ce21', 'ce22', 'ce23', 'ce24', 'ce25',
  'ce26', 'ce27', 'ce28', 'ce29', 'ce30', 'ce31', 'ce32', 'ce33', 'ce34', 'ce35',
  'ce36', 'cf01', 'cf02', 'cf03', 'cf04', 'cf05', 'cf06', 'cf07', 'cf08', 'cf09',
  'cf10', 'cf11', 'cf12', 'cf13', 'cf14', 'cf15', 'cf16', 'cf17', 'cf18', 'cf19',
  'cf20', 'cf21', 'cf22', 'cf23', 'cf24', 'cf25', 'cf26', 'cf27', 'cf28', 'cf29',
  'cf30', 'cf31', 'cf32', 'cf33', 'cf34', 'cf35', 'cf36', 'cf37', 'cf38', 'cf39',
  'cf40', 'cf41', 'cf42', 'cf43', 'cf44', 'cf45', 'cf46', 'cf47', 'cf48', 'cg01',
```

```
brown.words(fileids=['cg22'])
```

```
['Does', 'our', 'society', 'have', 'a', 'runaway', ',', '...', ...]
```

```
brown.sents()
```

```
[['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', 'Friday',
  'an', 'investigation', 'of', "Atlanta's", 'recent', 'primary',
  'election', 'produced', 'no', 'evidence', 'that',
  'any', 'irregularities', 'took', 'place', '.'], ['The', 'ju',
  'ry', 'further', 'said', 'in', 'term-end', 'presentments', 'tha',
  't', 'the', 'City', 'Executive', 'Committee', 'which', 'ha',
  'd', 'over-all', 'charge', 'of', 'the', 'election', 'deserves',
  'the', 'praise', 'and', 'thanks', 'of', 'the', 'Ci',
  'ty', 'of', 'Atlanta', 'for', 'the', 'manner', 'in', 'whi',
  'ch', 'the', 'election', 'was', 'conducted', '.'], ...]
```

```
fdist = nltk.FreqDist([w.lower() for w in news_text])
```

```
modals = ['can', 'could', 'may', 'might', 'must', 'will']
```

```
for m in modals:
```

```
    print(m + ': ' + str(fdist[m]))
```

```
can: 94
could: 87
may: 93
might: 38
must: 53
will: 389
```

```
cfd = nltk.ConditionalFreqDist(  
    (genre, word)  
    for genre in brown.categories()  
    for word in brown.words(categories=genre))  
genres = ['news', 'hobbies', 'science_fiction', 'romance']
```

```
modals = ['can', 'could']  
print(cfd.tabulate(condition=genres, samples=modals))
```

```
      can  could  
adventure    46   151  
belles_lettres 246   213  
editorial   121    56  
fiction      37   166  
government  117    38  
hobbies     268    58  
humor       16    30  
learned     365   159  
lore        170   141  
mystery      42   141  
news         93    86  
religion     82    59  
reviews      45    40  
romance      74   193  
science_fiction 16    49  
None
```

```
.>
```

```
#Reuters_corpus
```

```
from nltk.corpus import reuters
```

```
reuters.categories()
```

```
['acq', 'alum', 'barley', 'bop', 'carcass', 'castor-oil', 'coc
oa', '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', 'inc
ome', 'instal-debt', 'interest', 'ipi', 'iron-steel', 'jet', '
jobs', 'l-cattle', 'lead', 'lei', 'lin-oil', 'livestock', 'lum
ber', 'meal-feed', 'money-fx', 'money-supply', 'naphtha', 'nat
-gas', 'nickel', 'nkr', 'nzdrl', 'oat', 'oilseed', 'orange', '
palladium', 'palm-oil', 'palmkernel', 'pet-chem', 'platinum',
'potato', 'propane', 'rand', 'rape-oil', 'rapeseed', 'reserves
', 'retail', 'rice', 'rubber', 'rye', 'ship', 'silver', 'sorgh
um', 'soy-meal', 'soy-oil', 'soybean', 'strategic-metal', 'sug
ar', 'sun-meal', 'sun-oil', 'sunseed', 'tea', 'tin', 'trade',
'veg-oil', 'wheat', 'wpi', 'yen', 'zinc']
```

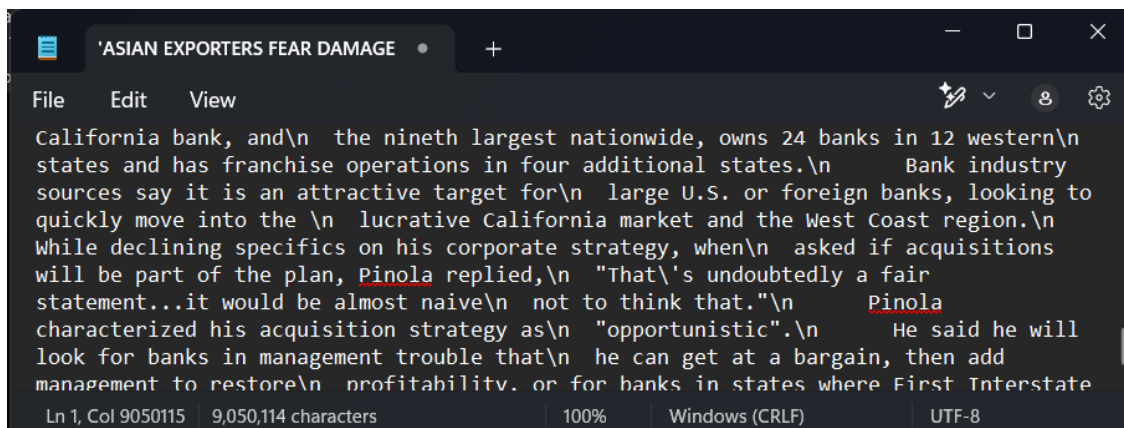
```
reuters.words()
```

```
['ASIAN', 'EXPORTERS', 'FEAR', 'DAMAGE', 'FROM', 'U', ...]
```

```
reuters.sents()
```

```
[[['ASIAN', 'EXPORTERS', 'FEAR', 'DAMAGE', 'FROM', 'U', '.', 'S
', '-.', 'JAPAN', 'RIFT', 'Mounting', 'trade', 'friction', 'be
tween', 'the', 'U', '.', 'S', '.', 'And', 'Japan', 'has', 'rai
sed', 'fears', 'among', 'many', 'of', 'Asia', '"', 's', 'expor
ting', 'nations', 'that', 'the', 'row', 'could', 'inflict', 'f
ar', '-', 'reaching', 'economic', 'damage', ',', 'businessmen'
, 'and', 'officials', 'said', '.'], ['They', 'told', 'Reuter',
'correspondents', 'in', 'Asian', 'capitals', 'a', 'U', '.', 'S
', '.', 'Move', 'against', 'Japan', 'might', 'boost', 'protect
ionist', 'sentiment', 'in', 'the', 'U', '.', 'S', '.', 'And',
'lead', 'to', 'curbs', 'on', 'American', 'imports', 'of', 'the
ir', 'products', '.'], ...]
```

```
reuters.raw()
```



```
File Edit View
California bank, and\n the ninth largest nationwide, owns 24 banks in 12 western\n
states and has franchise operations in four additional states.\n Bank industry
sources say it is an attractive target for\n large U.S. or foreign banks, looking
to quickly move into the \n lucrative California market and the West Coast region.\n
While declining specifics on his corporate strategy, when\n asked if acquisitions
will be part of the plan, Pinola replied,\n "That's undoubtedly a fair
statement...it would be almost naive\n not to think that."\n Pinola
characterized his acquisition strategy as\n "opportunistic".\n He said he will
look for banks in management trouble that\n he can get at a bargain, then add
management to restore\n profitability. or for banks in states where First Interstate
Ln 1, Col 9050115 | 9,050,114 characters | 100% | Windows (CRLF) | UTF-8
```

```
reuters.fileids()
```

```
['training/98', 'training/9801', 'training/9804', 'training/9805', 'training/9807',
'training/9809', 'training/981', 'training/9810', 'training/9812', 'training/9814',
'training/9815', 'training/9816', 'training/9818', 'training/982', 'training/9821',
'training/9822', 'training/9823', 'training/9824', 'training/9825', 'training/9827',
'training/9829', 'training/983', 'training/9833', 'training/9834', 'training/9836',
'training/9837', 'training/9839', 'training/984', 'training/9841', 'training/9844',
'training/9847', 'training/9848', 'training/9849', 'training/985', 'training/9850',
'training/9851', 'training/9852', 'training/9853', 'training/9855', 'training/9857',
'training/9858', 'training/986', 'training/9860', 'training/9861', 'training/9862',
'training/9864', 'training/9865', 'training/9866', 'training/9867', 'training/9868',
'training/9869', 'training/9871', 'training/9872', 'training/9874', 'training/9875',
'training/9878', 'training/988', 'training/9880', 'training/9884', 'training/9891',
'training/9892', 'training/9893', 'training/9894', 'training/9896', 'training/9897',
'training/99', 'training/9901', 'training/9902', 'training/9903', 'training/9904',
'training/9905', 'training/9906', 'training/9907', 'training/9909', 'training/991',
'training/9912', 'training/9913', 'training/9914', 'training/9915', 'training/9918',
'training/9919', 'training/9920', 'training/9923', 'training/9925', 'training/9926',
'training/9927', 'training/9928', 'training/9929', 'training/9930', 'training/9933',
'training/9934', 'training/9936', 'training/9937', 'training/9939', 'training/9940',
'training/9941', 'training/9942', 'training/9943', 'training/9946', 'training/9947',
'training/995', 'training/9952', 'training/9953', 'training/9954', 'training/9955',
'training/9956', 'training/9957', 'training/9958', 'training/9959', 'training/9961',
'training/9963', 'training/9964', 'training/9965', 'training/9967', 'training/9970',
'training/9971', 'training/9972', 'training/9973', 'training/9974', 'training/9975',
'training/9976', 'training/9977', 'training/9978', 'training/998', 'training/9981',
'training/9982', 'training/9984', 'training/9985', 'training/9988', 'training/9989',
'training/999', 'training/9992', 'training/9993', 'training/9994', 'training/9995']
```

```
reuters.fileids(['corn'])
```

```
['training/1385', 'training/13852', 'training/13856', 'training/1395', 'training/1399',
'training/14483', 'training/1582', 'training/1652', 'training/1777', 'training/1843',
'training/193', 'training/1952', 'training/197', 'training/2044', 'training/2172',
'training/2183', 'training/2217', 'training/2264', 'training/235', 'training/2382',
'training/2436', 'training/2456', 'training/2595', 'training/2599', 'training/2617',
'training/2727', 'training/2741', 'training/2749', 'training/2777', 'training/2848',
'training/2913', 'training/2922', 'training/2947', 'training/3138', 'training/3191',
'training/327', 'training/3282', 'training/3299', 'training/3306', 'training/3330',
'training/3337', 'training/3358', 'training/3401', 'training/3429', 'training/3847',
'training/3855', 'training/3881', 'training/3949', 'training/395', 'training/3979',
'training/3981', 'training/4047', 'training/4133', 'training/4289', 'training/4296',
'training/4382', 'training/4490', 'training/4599', 'training/4825', 'training/4905',
'training/4939', 'training/4988', 'training/5', 'training/5003', 'training/501',
'training/5017', 'training/5033', 'training/5109', 'training/516', 'training/5185',
'training/5338', 'training/5467', 'training/5518', 'training/5531', 'training/5606',
'training/5610', 'training/5636', 'training/5637', 'training/57', 'training/5847',
'training/5933', 'training/6', 'training/6142', 'training/6221', 'training/6236',
'training/6239', 'training/6259', 'training/6269', 'training/6386', 'training/6585',
'training/6588', 'training/6735', 'training/6890', 'training/6897', 'training/694',
'training/7062', 'training/7205', 'training/7215', 'training/7336', 'training/7387',
'training/7389', 'training/7390', 'training/7395', 'training/7700', 'training/7792',
'training/7917', 'training/7934', 'training/7943', 'training/8004', 'training/8140',
'training/8161', 'training/8166', 'training/8257', 'training/8273', 'training/8400',
'training/8443', 'training/8446', 'training/8535', 'training/855', 'training/8759',
'training/8941', 'training/8983', 'training/8993', 'training/9058', 'training/9093',
'training/9094', 'training/934', 'training/9470', 'training/9521', 'training/9667',
'training/97', 'training/9865', 'training/9958', 'training/9989']
```

```
reuters.words('training/9865')[:14]
```

```
['FRENCH', 'FREE', 'MARKET', 'CEREAL', 'EXPORT', 'BIDS', 'DETAILED', 'French', 'operators', 'have', 'requested', 'licences', 'to', 'export']
```

```
reuters.words('training/9865')[:3]
```

```
['FRENCH', 'FREE', 'MARKET']
```

```
cfd = nltk.ConditionalFreqDist(
```

```
    (target, fileid[:4])
```

```
    for fileid in reuters.fileids()
```

```
    for w in reuters.words(fileid)
```

```
    for target in ['barley', 'corn']
```

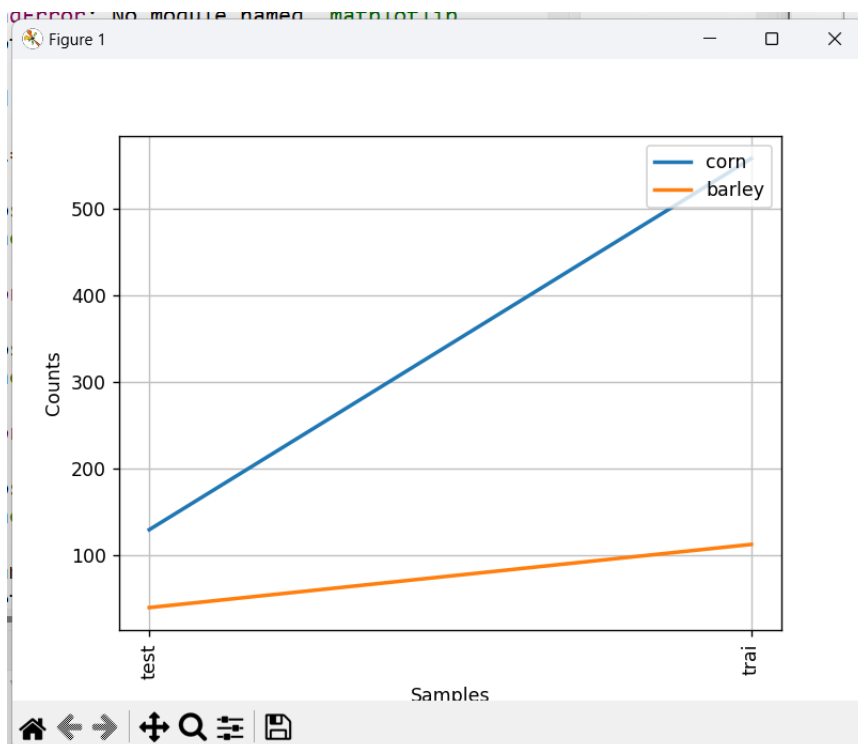
```
    if w.lower().startswith(target))
```

```
import matplotlib.pyplot as plt
```

```
cfd.plot()
```

```
<Axes: xlabel='Samples', ylabel='Counts'>
```

```
plt.show()
```





```
#Inaugural_corpus
```

```
from nltk.corpus import inaugural
```

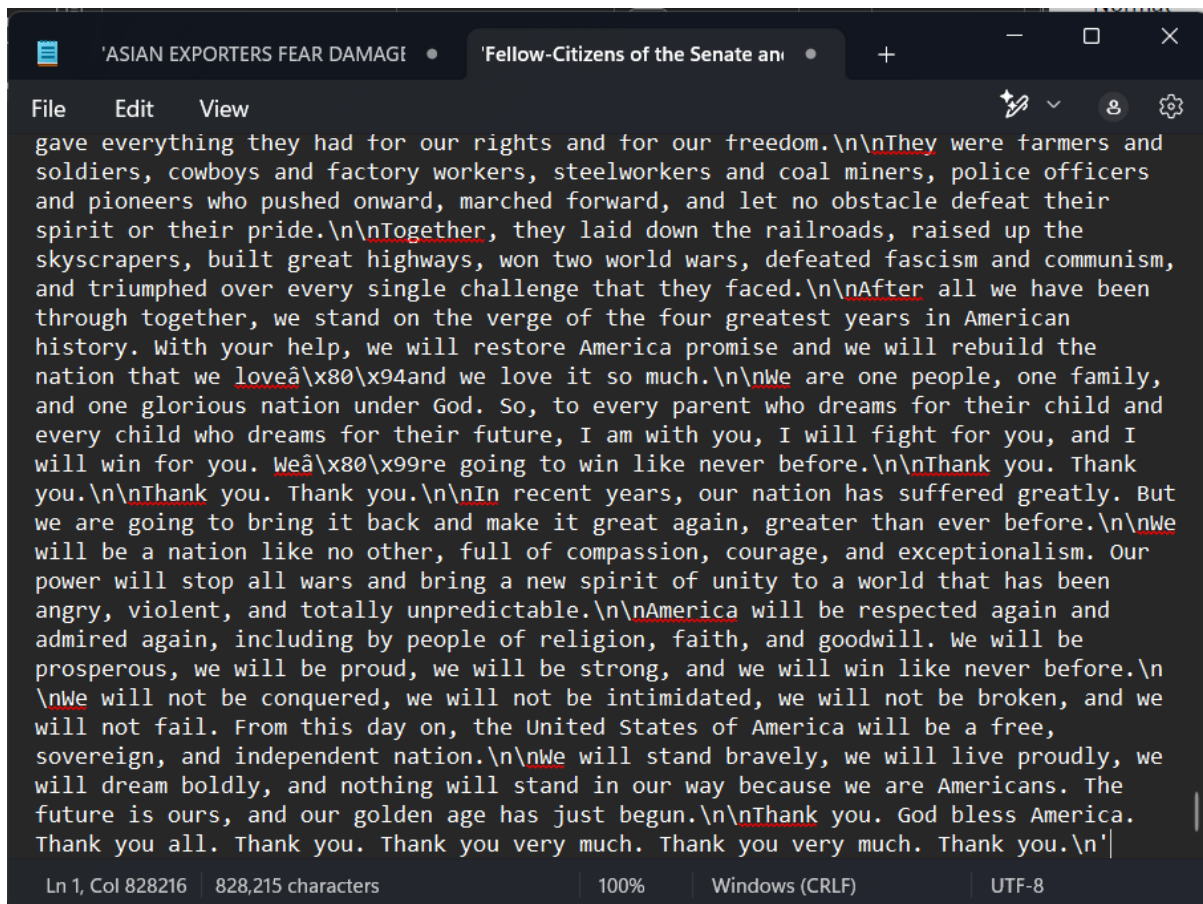
```
inaugural.words()
```

```
['Fellow', '-', 'Citizens', 'of', 'the', 'Senate', ...]
```

```
inaugural.sents()
```

```
[[['Fellow', '-', 'Citizens', 'of', 'the', 'Senate', 'a  
nd', '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',  
'.'], ...]]
```

```
inaugural.raw()
```



```
gave everything they had for our rights and for our freedom.\n\nThey were farmers and  
soldiers, cowboys and factory workers, steelworkers and coal miners, police officers  
and pioneers who pushed onward, marched forward, and let no obstacle defeat their  
spirit or their pride.\n\nTogether, they laid down the railroads, raised up the  
skyscrapers, built great highways, won two world wars, defeated fascism and communism,  
and triumphed over every single challenge that they faced.\n\nAfter all we have been  
through together, we stand on the verge of the four greatest years in American  
history. With your help, we will restore America promise and we will rebuild the  
nation that we love&x80\x94and we love it so much.\n\nWe are one people, one family,  
and one glorious nation under God. So, to every parent who dreams for their child and  
every child who dreams for their future, I am with you, I will fight for you, and I  
will win for you. We&x80\x99re going to win like never before.\n\nThank you. Thank  
you.\n\nThank you. Thank you.\n\nIn recent years, our nation has suffered greatly. But  
we are going to bring it back and make it great again, greater than ever before.\n\nWe  
will be a nation like no other, full of compassion, courage, and exceptionalism. Our  
power will stop all wars and bring a new spirit of unity to a world that has been  
angry, violent, and totally unpredictable.\n\nAmerica will be respected again and  
admired again, including by people of religion, faith, and goodwill. We will be  
prosperous, we will be proud, we will be strong, and we will win like never before.\n\nWe will not be conquered, we will not be intimidated, we will not be broken, and we  
will not fail. From this day on, the United States of America will be a free,  
sovereign, and independent nation.\n\nWe will stand bravely, we will live proudly, we  
will dream boldly, and nothing will stand in our way because we are Americans. The  
future is ours, and our golden age has just begun.\n\nThank you. God bless America.  
Thank you all. Thank you. Thank you very much. Thank you very much. Thank you.\n'
```

Ln 1, Col 828216 | 828,215 characters | 100% | Windows (CRLF) | UTF-8

```
inaugural.fileids()
```

```
['1789-Washington.txt', '1793-Washington.txt', '1797-Adams.txt',
 '1801-Jefferson.txt', '1805-Jefferson.txt', '1809-Madison.txt',
 '1813-Madison.txt', '1817-Monroe.txt', '1821-Monroe.txt',
 '1825-Adams.txt', '1829-Jackson.txt', '1833-Jackson.txt', '1837-VanBuren.txt',
 '1841-Harrison.txt', '1845-Polk.txt', '1849-Taylor.txt', '1853-Pierce.txt',
 '1857-Buchanan.txt', '1861-Lincoln.txt', '1865-Lincoln.txt', '1869-Grant.txt',
 '1873-Grant.txt', '1877-Hayes.txt', '1881-Garfield.txt', '1885-Cleveland.txt',
 '1889-Harrison.txt', '1893-Cleveland.txt', '1897-McKinley.txt', '1901-McKinley.txt',
 '1905-Roosevelt.txt', '1909-Taft.txt', '1913-Wilson.txt', '1917-Wilson.txt',
 '1921-Harding.txt', '1925-Coolidge.txt', '1929-Hoover.txt', '1933-Roosevelt.txt',
 '1937-Roosevelt.txt', '1941-Roosevelt.txt', '1945-Roosevelt.txt', '1949-Truman.txt',
 '1953-Eisenhower.txt', '1957-Eisenhower.txt', '1961-Kennedy.txt', '1965-Johnson.txt',
 '1969-Nixon.txt', '1973-Nixon.txt', '1977-Carter.txt', '1981-Reagan.txt', '1985-Reagan.txt',
 '1989-Bush.txt', '1993-Clinton.txt', '1997-Clinton.txt', '2001-Bush.txt',
 '2005-Bush.txt', '2009-Obama.txt', '2013-Obama.txt', '2017-Trump.txt', '2021-Biden.txt', '2025-Trump.txt']
```

```
[fileid[:4] for fileid in inaugural.fileids()]
```

```
['1789', '1793', '1797', '1801', '1805', '1809', '1813', '1817',
 '1821', '1825', '1829', '1833', '1837', '1841', '1845', '1849',
 '1853', '1857', '1861', '1865', '1869', '1873', '1877', '1881',
 '1885', '1889', '1893', '1897', '1901', '1905', '1909', '1913',
 '1917', '1921', '1925', '1929', '1933', '1937', '1941', '1945',
 '1949', '1953', '1957', '1961', '1965', '1969', '1973', '1977',
 '1981', '1985', '1989', '1993', '1997', '2001', '2005', '2009',
 '2013', '2017', '2021', '2025']
```

```
cdf = nltk.ConditionalFreqDist(
```

```
    (target, fileid[:4])
```

```
    for fileid in inaugural.fileids()
```

```
    for w in inaugural.words(fileid)
```

```
    for target in ['america', 'citizen']
```

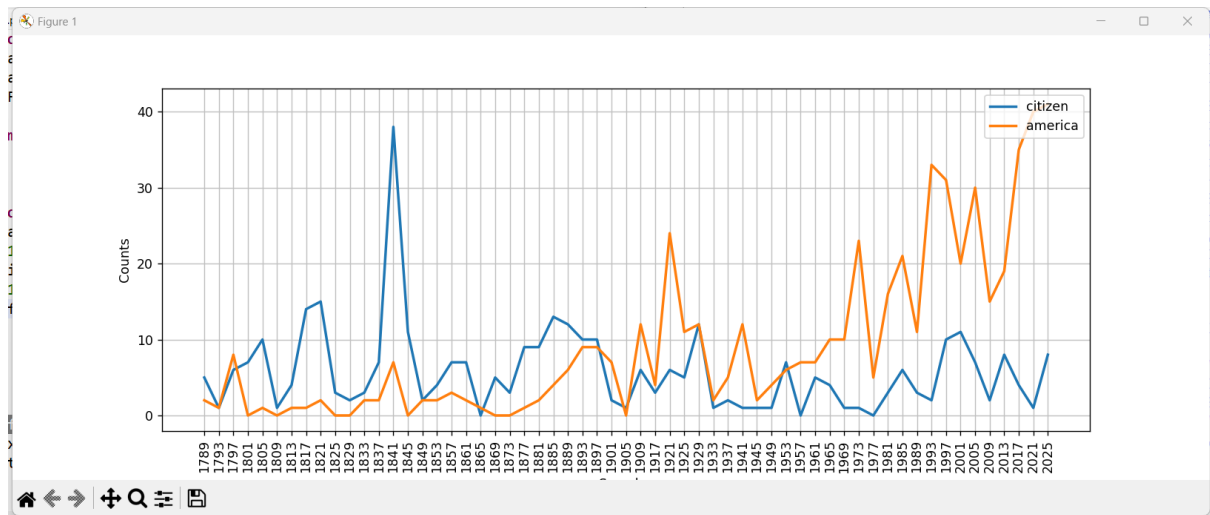
```
    if w.lower().startswith(target))
```

```
cdf.plot()
```

```
<Axes: xlabel='Samples', ylabel='Counts'>
```

```
plt.show()
```





```
#UDHR_corpus
```

```
from nltk.corpus import udhr
```

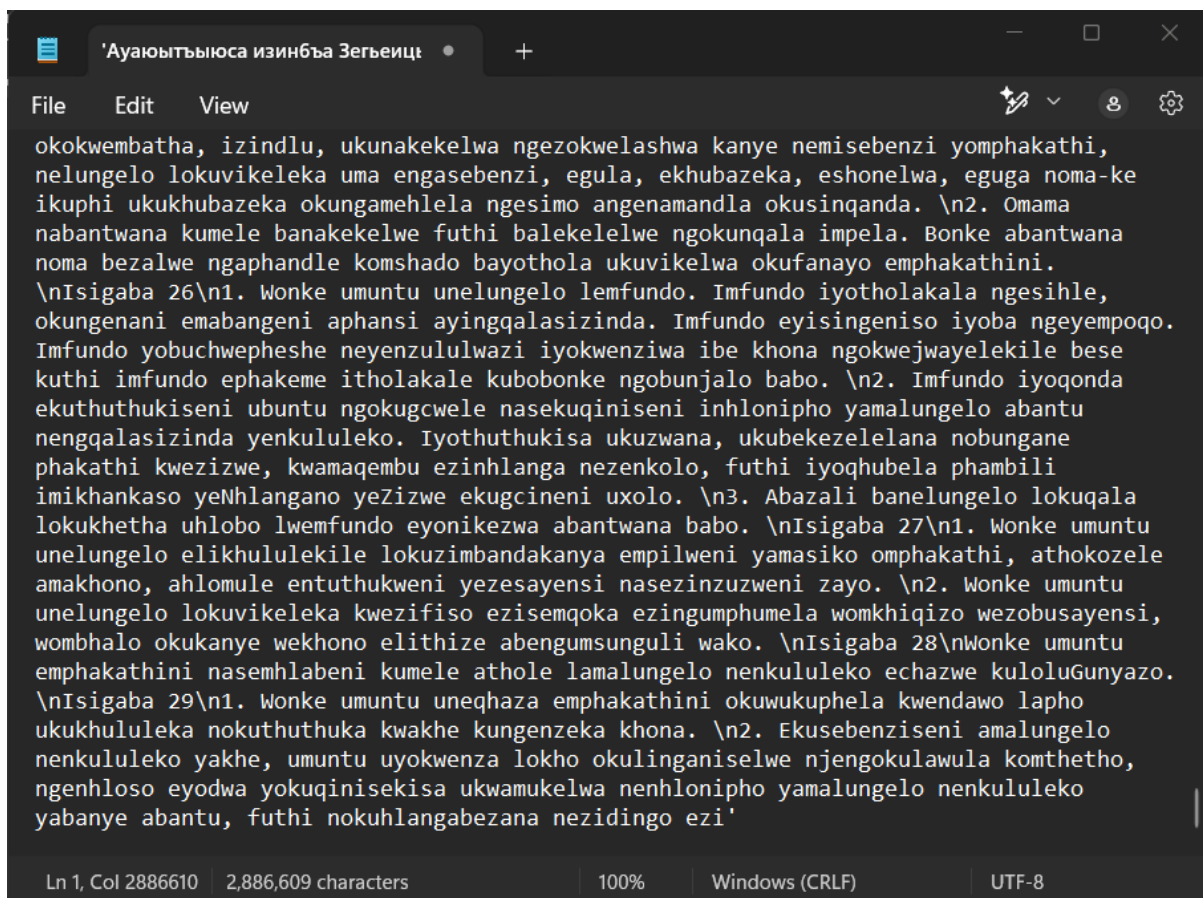
```
udhr.words()
```

```
['Ауаюыттыюса', 'изинбѣа', 'Зегъеицързеи8шу', ...]
```

```
udhr.sents()
```

```
[['Ауаюыттыюса', 'изинбѣа', 'Зегъеицързеи8шу', 'Адекла  
рация', 'Алагалажѣа', 'Дызус0заалак', ',', 'ауааттыюса  
тъ', '0аацѣара', 'иалахѣу', 'ища0ыри', ',', 'иара', 'и  
узийттым0хо', 'имоу', 'изинбѣеи', ',', 'ахабѣи0реи', ',',  
, 'аиашареи', ',', 'адунеижѣларбжѣаратѣи', 'а0ынчреи',  
, 'шьа0ас', 'ишрымоу', 'хшыюзышь0ра', 'азуа', ',', 'иа  
ра', 'убас', 'ауаюы', 'изинбѣа', 'ратѣамбареи', 'хърюа  
а', 'рым0ареи', ',', 'ламыс', 'змоу', 'дарбанзаалак',  
'изымычщаша', 'агыгшѣыгратѣ', 'хымюа8гашѣа', 'айынѣа',  
'ауаюы', 'дышнанагахѣо', 'еилкааны', ',', 'настѣы', 'а  
уаюы', 'иажѣеи', 'идунеихѣа8шышьѣеи', 'дрыбѣи0ны', ',',
```

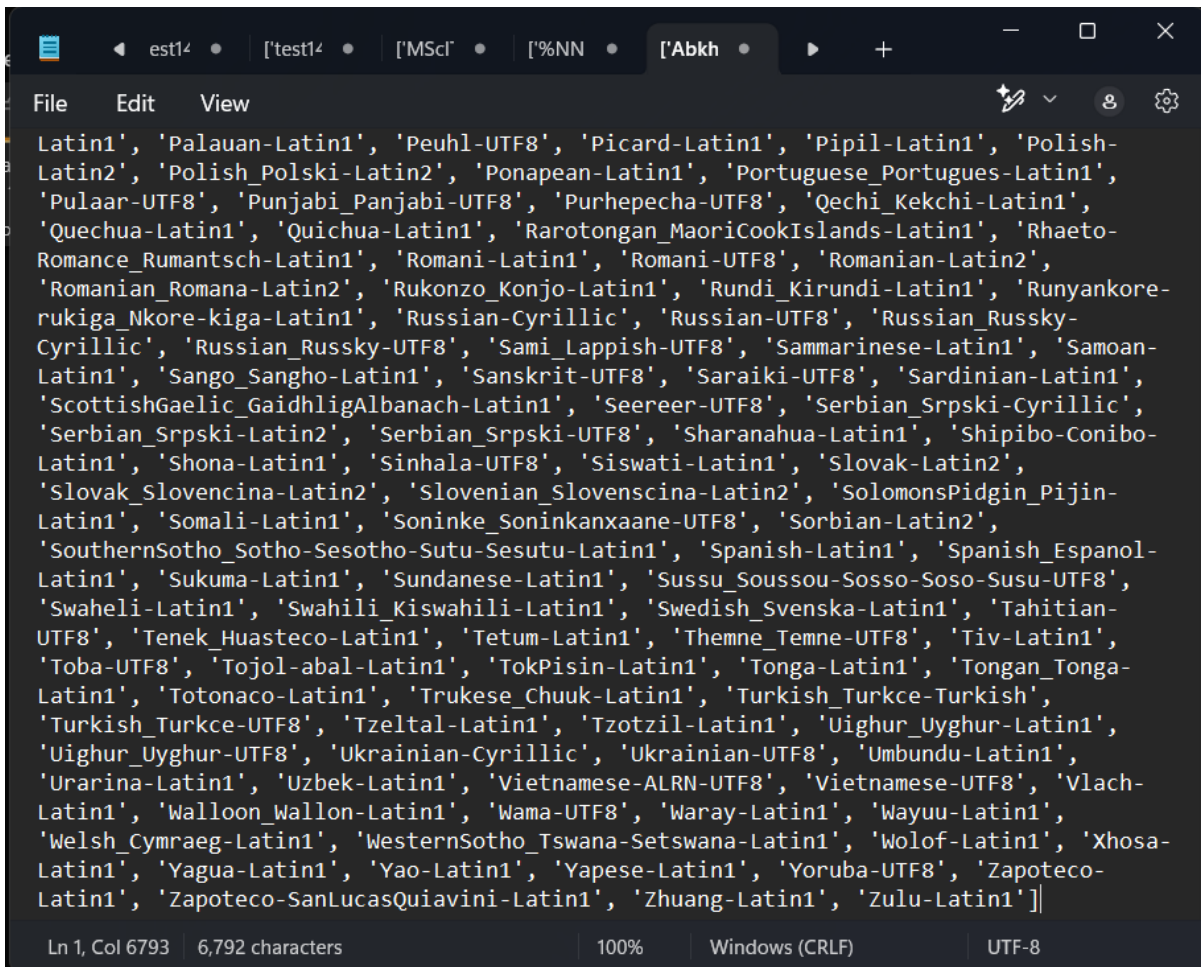
```
udhr.raw()
```



```
'Ауаюыттыюса изинбѣа Зегъеицързеи8шу'
File Edit View
okokwembatha, izindlu, ukunakekelwa ngezokwelashwa kanye nemisebenzi yomphakathi, nelungelo lokuvikeleka uma engasebenzi, egula, ekhubazeka, eshonelwa, eguga noma-ke ikuphi ukukhubazeka okungamehlela ngesimo angenamandla okusinqanda. \n2. Omama nabantwana kumele banakekelwe futhi balekelelwe ngokunqala impela. Bonke abantwana noma bezalwe ngaphandle komshado bayothola ukuvikelwa okufanayo emphakathini. \nIsigaba 26\n1. Wonke umuntu unelungelo lemfundo. Imfundo iyotholakala ngesihle, okungenani emabangeni aphansi ayingqalasizinda. Imfundo eyisingeniso iyoba ngeyempoqo. Imfundo yobuchwepheshe neyenzululwazi iyokwenziwa ibe khona ngokwejwayelekile bese kuthi imfundo ephakeme itholakale kubobonke ngobunjalo babo. \n2. Imfundo iyoqonda ekuthuthukiseni ubuntu ngokugcwele nasekuqiniseni inhlonipho yamalungelo abantu nengqalasizinda yenkululeko. Iyothuthukisa ukuzwana, ukubekezelelana nobungane phakathi kwezizwe, kwamaqembu ezinhlanga nezenkolo, futhi iyoqhubela phambili imikhankaso yeNhlango yeZizwe ekugcineni uxolo. \n3. Abazali banelungelo lokuqala lokukhetha uhlobo lwemfundo eyonikezwa abantwana babo. \nIsigaba 27\n1. Wonke umuntu unelungelo elikhululekile lokuzimbandakanya empilweni yamasiko omphakathi, athokozele amakhono, ahlomule entuthukweni yezesayensi nasezinzuzweni zayo. \n2. Wonke umuntu unelungelo lokuvikeleka kwezifiso ezisemqoka ezingumphumela womkhiqizo wezobusayensi, wombhalo okukanye wekhono elithize abengumsunguli wako. \nIsigaba 28\nWonke umuntu emphakathini nasemhlabeni kumele athole lamalungelo nenkululeko echazwe kuloluGunyazo. \nIsigaba 29\n1. Wonke umuntu uneqhaza emphakathini okuwukuphela kwendawo lapho ukukhululeka nokuthuthuka kwakhe kungenzeka khona. \n2. Ekusebenziseni amalungelo nenkululeko yakhe, umuntu uyokwenza lokho okulinganiselwe njengokulawula komthetho, ngenhloso eyodwa yokuqinisekisa ukwamukelwa nenhlonipho yamalungelo nenkululeko yabanye abantu, futhi nokuhlangabezana nezidingo ezi'
```

Ln 1, Col 2886610 2,886,609 characters 100% Windows (CRLF) UTF-8

```
udhr.fileids()
```



```

Latin1', 'Palauan-Latin1', 'Peuhl-UTF8', 'Picard-Latin1', 'Pipil-Latin1', 'Polish-
Latin2', 'Polish_Polski-Latin2', 'Ponapean-Latin1', 'Portuguese_Portugues-Latin1',
'Pulaar-UTF8', 'Punjabi_Panjabi-UTF8', 'Purhepecha-UTF8', 'Qechi_Kekchi-Latin1',
'Quechua-Latin1', 'Quichua-Latin1', 'Rarotongan_MaoriCookIslands-Latin1', 'Rhaeto-
Romance_Rumantsch-Latin1', 'Romani-Latin1', 'Romani-UTF8', 'Romanian-Latin2',
'Romanian_Romana-Latin2', 'Rukonzo_Konjo-Latin1', 'Rundi_Kirundi-Latin1', 'Runyankore-
rukiga_Nkore-kiga-Latin1', 'Russian-Cyrillic', 'Russian-UTF8', 'Russian_Russky-
Cyrillic', 'Russian_Russky-UTF8', 'Sami_Lappish-UTF8', 'Sammarinese-Latin1', 'Samoan-
Latin1', 'Sango_Sangho-Latin1', 'Sanskrit-UTF8', 'Saraiki-UTF8', 'Sardinian-Latin1',
'ScottishGaelic_GaidhligAlbanach-Latin1', 'Seereer-UTF8', 'Serbian_Srpski-Cyrillic',
'Serbian_Srpski-Latin2', 'Serbian_Srpski-UTF8', 'Sharanahua-Latin1', 'Shipibo-Conibo-
Latin1', 'Shona-Latin1', 'Sinhala-UTF8', 'Siswati-Latin1', 'Slovak-Latin2',
'Slovak_Slovenčina-Latin2', 'Slovenian_Slovenscina-Latin2', 'SolomonsPidgin_Pijin-
Latin1', 'Somali-Latin1', 'Soninke_Soninkanxaane-UTF8', 'Sorbian-Latin2',
'SouthernSotho_Sotho-Sesotho-Sutu-Sesutu-Latin1', 'Spanish-Latin1', 'Spanish_Espanol-
Latin1', 'Sukuma-Latin1', 'Sundanese-Latin1', 'Sussu_Soussou-Sosso-Soso-Susu-UTF8',
'Swaheli-Latin1', 'Swahili_Kiswahili-Latin1', 'Swedish_Svenska-Latin1', 'Tahitian-
UTF8', 'Tenek_Huasteco-Latin1', 'Tetum-Latin1', 'Themne_Temne-UTF8', 'Tiv-Latin1',
'Toba-UTF8', 'Tojol-abal-Latin1', 'TokPisin-Latin1', 'Tonga-Latin1', 'Tongan_Tonga-
Latin1', 'Totonaco-Latin1', 'Trukese_Chuuuk-Latin1', 'Turkish_Turkce-Turkish',
'Turkish_Turkce-UTF8', 'Tzeltal-Latin1', 'Tzotzil-Latin1', 'Uighur_Uyghur-Latin1',
'Uighur_Uyghur-UTF8', 'Ukrainian-Cyrillic', 'Ukrainian-UTF8', 'Umbundu-Latin1',
'Urarina-Latin1', 'Uzbek-Latin1', 'Vietnamese-ALRN-UTF8', 'Vietnamese-UTF8', 'Vlach-
Latin1', 'Walloon_Wallon-Latin1', 'Wama-UTF8', 'Waray-Latin1', 'Wayuu-Latin1',
'Welsh_Cymraeg-Latin1', 'WesternSotho_Tswana-Setswana-Latin1', 'Wolof-Latin1', 'Xhosa-
Latin1', 'Yagua-Latin1', 'Yao-Latin1', 'Yapese-Latin1', 'Yoruba-UTF8', 'Zapoteco-
Latin1', 'Zapoteco-SanLucasQuiavini-Latin1', 'Zhuang-Latin1', 'Zulu-Latin1']]

```

```
languages = ['Chickasaw', 'English', 'German_Deutsch']
```

```
cdf = nltk.ConditionalFreqDist(
```

```
    (lang, len(word))
```

```
    for lang in languages
```

```
    for word in udhr.words(lang+'-Latin1'))
```

```
cdf.plot(cumulative=True)
```

```
<Axes: xlabel='Samples', ylabel='Counts'>
```

```
plt.show()
```

