

1. Evaluate on all of the sentences from the Brown corpus with the category lore:

0.7978585481282695

Evaluate on all of the sentences from the Brown corpus with the category news:

0.9349006503968017

Output of tagger on the 200th sentence of the lore category of the Brown Corpus : [(('I', 'PPSS'), ('can't', 'MD\*'), ('tell', 'VB'), ('when', 'WRB'), (',', ','), ('but', 'CC'), ('I'm', 'PPSS+BEM'), ('positive', 'JJ'), ('I', 'PPSS'), ('witnessed', 'VBN'), ('this', 'DT'), ('same', 'AP'), ('scene', 'NN'), ('of', 'IN'), ('this', 'DT'), ('particular', 'JJ'), ('gathering', 'NN'), ('at', 'IN'), ('some', 'DTI'), ('time', 'NN'), ('in', 'IN'), ('the', 'AT'), ('past', 'NN'), ('', ''), ('!', '!'), ('!', '!')]

2. In this exercise I have answered the questions a-d for given categories humor, romance, and government and on total brown corpus.

## a. Government

Distinct words tagged with "JJ" : 1010

First five words tagged with "JJ" : ['10-gallon', '10-year', '150-milliampere', '25-liter', '4-cell']

Humor

Distinct words tagged with "JJ" : 649

First five words tagged with "JJ" : ['able', 'absent', 'absurd', 'academic', 'acid']

Romance

Distinct words tagged with "JJ" : 1104

First five words tagged with "JJ" : ['able', 'abrupt', 'absent', 'absent-minded', 'absolute']

The tag JJ denotes adjective or numerical, ordinal. It is obvious that government genre has numerical words as first five. Since humor and romance is more English oriented it has adjectives in first five list.

- b. For plural noun the tag is "NNS" and for singular verb "VBZ". I have created 2 sets for words with tag "NNS" and for tag "VBZ" and I found the unknown of these two sets.

Government

Distinct words tagged with "Plural nouns or singular verbs" : 1343

First ten words tagged with "Plural nouns or singular verbs" : ['\$0.027', '\$.07', '\$.076', '\$.09', '\$0.9', '\$1', '\$1,000', '\$1,276', '\$1,390', '\$1,500']

Humor

Distinct words tagged with "Plural nouns or singular verbs" : 609

First ten words tagged with "Plural nouns or singular verbs" : ['\$85,000', 'absences', 'accents', 'actions', 'activities', 'actors', 'actresses', 'acts', 'ads', 'afternoons']

Romance

Distinct words tagged with "Plural nouns or singular verbs" : 960

First ten words tagged with "Plural nouns or singular verbs" : ['accounts', 'acknowledgments', 'acres', 'actions', 'activities', 'actresses', 'ads', 'adults', 'adventures', 'affairs']

Since government genre uses more numerical values it has amount values as plural nouns or singular verbs. Whereas humor and romance has similar kinds of words.

- c. For this I have constructed trigrams for the text and check if each trigram is of given format or not.

government

Most common 3 phrases in this genre is : [('of the act', 14), ('of the department', 16), ('of the state', 35)]

humor

Most common 3 phrases in this genre is : [('at the studio', 3), ('by the time', 3), ('in the world', 5)]

romance

Most common 3 phrases in this genre is : [('for a moment', 8), ('in the morning', 12), ('in the world', 8)]

(In total Brown Corpus) Most common word phrases in entire brown corpus is : [('in the world', 90), ('at the end', 72), ('of the world', 71)]

Since government genre deals with acts, departments and state there are phrases like 'of the act'. 'of the department' mostly used. Since humor and romance is mostly imaginative they have similar kind of phrases.

- d. The ratio of masculine pronouns is

government

masculine ratio in this genre is : 87.5

humor

masculine ratio in this genre is : 2.576923076923077

romance

masculine ratio in this genre is : 1.4066666666666667

(In total Brown Corpus) ratio of masculine pronouns over brown corpus is : 3.2476066850559793

Since there will be more men involved in government it is obvious that it has more masculine ratio. Since humor and romance has much more even gender their values are less.

3. a. Number of words which has exactly 5 possible tags : 176

Few words have similar world classes like 'news': {'NN-NC', 'NP', 'NN', 'NN-HL', 'NN-TL'}

b. Word tags which has different prefix has most distinct tags. I have taken an example word "that".

the distinct word is : that

Sentence 1: The Fulton County Grand Jury said Friday an investigation of Atlanta's recent primary election produced `` no evidence " that any irregularities took place .

Sentence 2: Regarding Atlanta's new multi-million-dollar airport , the jury recommended `` that when the new management takes charge Jan. 1 the airport be operated in a manner that will eliminate political influences " .

Sentence 3: `` Actually , the abuse of the process may have constituted a contempt of the Criminal court of Cook county , altho vindication of the authority of that court is not the function of this court " , said Karns , who is a City judge in East St. Louis sitting in Cook County court .

Sentence 4: While the city council suggested that the Legislative Council might perform the review , Mr. Notte said that instead he will take up the matter with Atty. Gen. J. Joseph Nugent to get `` the benefit of his views " .

Sentence 5: Thus , as a development program is being launched , commitments and obligations must be entered into in a given year which may exceed by twofold or threefold the expenditures to be made in that year .