Finding top 10 words that are most different in usage...

1)Word(s)='to'. Native count=16770 vs Non-native count=23452. Diff=6682

2)Word(s)='in'. Native count=10226 vs Non-native count=14372. Diff=4146

3)Word(s)='is'. Native count=7381 vs Non-native count=10852. Diff=3471

4)Word(s)='that'. Native count=9192 vs Non-native count=12590. Diff=3398

5)Word(s)='will'. Native count=1474 vs Non-native count=4674. Diff=3200

6)Word(s)='more'. Native count=1747 vs Non-native count=4470. Diff=2723

7)Word(s)='n't'. Native count=3856 vs Non-native count=1435. Diff=2421

8)Word(s)='are'. Native count=3788 vs Non-native count=6191. Diff=2403

9)Word(s)='and'. Native count=15245 vs Non-native count=17640. Diff=2395

10)Word(s)='can'. Native count=1552 vs Non-native count=3915. Diff=2363

['to', 'in', 'is', 'that', 'will', 'more', "n't", 'are', 'and', 'can']

-------------FUNC\_WORDS\_END-------------------------

top\_words\_list size is 343

Reading parsed data

There are 11040 Native samples and 11044 Non-Native

Train set size 17667 - native=8842, non-native=8825

Test set size 4417 - native=2198, non-native=2219

Train data 'all feature vector' size is 343x17667

Train data size is 17667

Test data 'all feature vector' size is 343x4417

Test data size is 4417

------------------TOP\_WORDS\_START--------------------

native

0)'s 4726

1)people 2341

2)has 1614

3)just 1609

4)eu 1378

5)were 1245

6)think 1181

7)'re 1129

8)even 1081

9)get 1055

Total count for 10 top f\_words=17359

non-native

0)people 7239

1)life 3015

2)time 2513

3)young 2380

4)think 2262

5)new 2224

6)cars 2089

7)things 2013

8)better 1888

9)knowledge 1866

Total count for 10 top f\_words=27489

Finding top 10 words that are most different in usage...

1)Word(s)='people'. Native count=2341 vs Non-native count=7239. Diff=4898

2)Word(s)=''s'. Native count=4726 vs Non-native count=1255. Diff=3471

3)Word(s)='life'. Native count=255 vs Non-native count=3015. Diff=2760

4)Word(s)='young'. Native count=65 vs Non-native count=2380. Diff=2315

5)Word(s)='cars'. Native count=36 vs Non-native count=2089. Diff=2053

6)Word(s)='knowledge'. Native count=55 vs Non-native count=1866. Diff=1811

7)Word(s)='new'. Native count=475 vs Non-native count=2224. Diff=1749

8)Word(s)='time'. Native count=841 vs Non-native count=2513. Diff=1672

9)Word(s)='facts'. Native count=49 vs Non-native count=1578. Diff=1529

10)Word(s)='things'. Native count=496 vs Non-native count=2013. Diff=1517

['people', "'s", 'life', 'young', 'cars', 'knowledge', 'new', 'time', 'facts', 'things']

---------------TOP\_WORDS\_END-----------------------