

```

1 # NZZ Storytelling, script for the following article:
  https://www.nzz.ch/international/freund-und-feind-betrachtet-durch-die-brille-des-
  weissen-hauses-ld.1349175#subtitle-die-methodik-im-detail
2 # questions and comments: marie-jose.kolly@nzz.ch
3
4
5 #USA per file, corpus 1 because we also want caps for US (vs. us-pronoun)
6 cd mypath/data/sotu_corpus1
7 for file in $(ls -l |grep -i .txt ) ; do printf "$file\t" \
8 && egrep -o 'americ|Americ' $file | wc -l | tr -d '\n' && printf "\t" \
9 && egrep -o 'our republic|our Republic' $file | wc -l | tr -d '\n' && printf "\t" \
10 && egrep -o 'our union|our Union|federal union|federal Union' $file | wc -l | tr -d '\n'
  && printf "\t" \
11 && egrep -o 'united states|United States|\<US\>|\<USA\>|\<usa\>' $file | wc -l; done >
  mypath/data/datafiles/amicUSPerFile.txt
12
13 #countries per file
14 cd mypath/data/sotu_corpus2
15 for file in $(ls -l |grep -i .txt ) ; do printf "$file\t" \
16 && grep -o 'afghan' $file | wc -l | tr -d '\n' && printf "\t" \
17 && grep -o '\<aland' $file | wc -l | tr -d '\n' && printf "\t" \
18 && grep -o 'albania' $file | wc -l | tr -d '\n' && printf "\t" \
19 && egrep -o 'algeria|algerine' $file | wc -l | tr -d '\n' && printf "\t" \
20 && grep -o 'american samoa' $file | wc -l | tr -d '\n' && printf "\t" \
21 && grep -o 'andorra' $file | wc -l | tr -d '\n' && printf "\t" \
22 && grep -o '\<angol' $file | wc -l | tr -d '\n' && printf "\t" \
23 && grep -o 'anguill?a' $file | wc -l | tr -d '\n' && printf "\t" \
24 && grep -o 'antarctica' $file | wc -l | tr -d '\n' && printf "\t" \
25 && grep -o 'antigua' $file | wc -l | tr -d '\n' && printf "\t" \
26 && grep -o 'argentin' $file | wc -l | tr -d '\n' && printf "\t" \
27 && grep -o 'armenia' $file | wc -l | tr -d '\n' && printf "\t" \
28 && egrep -o 'aruba' $file | wc -l | tr -d '\n' && printf "\t" \
29 && grep -o 'australia' $file | wc -l | tr -d '\n' && printf "\t" \
30 && grep -o 'austria' $file | wc -l | tr -d '\n' && printf "\t" \
31 && grep -o 'austria-hungary' $file | wc -l | tr -d '\n' && printf "\t" \
32 && grep -o 'azerbaijan' $file | wc -l | tr -d '\n' && printf "\t" \
33 && grep -o 'baden' $file | wc -l | tr -d '\n' && printf "\t" \
34 && grep -o 'bahamas' $file | wc -l | tr -d '\n' && printf "\t" \
35 && grep -o 'bahrain' $file | wc -l | tr -d '\n' && printf "\t" \
36 && grep -o 'bangladesh' $file | wc -l | tr -d '\n' && printf "\t" \
37 && grep -o 'barbados' $file | wc -l | tr -d '\n' && printf "\t" \
38 && grep -o 'bavaria' $file | wc -l | tr -d '\n' && printf "\t" \
39 && egrep -o 'belarus|byelo' $file | wc -l | tr -d '\n' && printf "\t" \
40 && grep -o 'belgi' $file | wc -l | tr -d '\n' && printf "\t" \
41 && grep -o 'beliz' $file | wc -l | tr -d '\n' && printf "\t" \
42 && egrep -o 'benin|dahome' $file | wc -l | tr -d '\n' && printf "\t" \
43 && grep -o 'bermuda' $file | wc -l | tr -d '\n' && printf "\t" \
44 && grep -o 'bhutan' $file | wc -l | tr -d '\n' && printf "\t" \
45 && grep -o 'bolivia' $file | wc -l | tr -d '\n' && printf "\t" \
46 && egrep -o 'bonaire|eustatius|caribbean netherlands|saba' $file | wc -l | tr -d '\n' &&
  printf "\t" \
47 && egrep -o 'herzegovina|bosnia' $file | wc -l | tr -d '\n' && printf "\t" \
48 && egrep -o 'botswana|bechuana' $file | wc -l | tr -d '\n' && printf "\t" \
49 && grep -o 'bouvet' $file | wc -l | tr -d '\n' && printf "\t" \
50 && grep -o 'brazil' $file | wc -l | tr -d '\n' && printf "\t" \
51 && grep -o 'british.?indian.?ocean' $file | wc -l | tr -d '\n' && printf "\t" \

```

```

52 && grep -o 'brunei' $file | wc -l | tr -d '\n' && printf "\t" \
53 && grep -o 'bulgaria' $file | wc -l | tr -d '\n' && printf "\t" \
54 && egrep -o 'burkina|bfaso|upper.?volta' $file | wc -l | tr -d '\n' && printf "\t" \
55 && grep -o 'burundi' $file | wc -l | tr -d '\n' && printf "\t" \
56 && grep -o '<verde' $file | wc -l | tr -d '\n' && printf "\t" \
57 && egrep -o 'cambodia|kampuchea|khmer' $file | wc -l | tr -d '\n' && printf "\t" \
58 && grep -o 'cameroon' $file | wc -l | tr -d '\n' && printf "\t" \
59 && grep -o 'canada' $file | wc -l | tr -d '\n' && printf "\t" \
60 && grep -o 'cayman' $file | wc -l | tr -d '\n' && printf "\t" \
61 && grep -o '\bcentral.africa' $file | wc -l | tr -d '\n' && printf "\t" \
62 && grep -o '\bchad' $file | wc -l | tr -d '\n' && printf "\t" \
63 && grep -o '\bchile' $file | wc -l | tr -d '\n' && printf "\t" \
64 && grep -o '<china' $file | wc -l | tr -d '\n' && printf "\t" \
65 && grep -o 'christmas.isl' $file | wc -l | tr -d '\n' && printf "\t" \
66 && egrep -o '\bcocos|keeling' $file | wc -l | tr -d '\n' && printf "\t" \
67 && grep -o 'colombia' $file | wc -l | tr -d '\n' && printf "\t" \
68 && grep -o 'comoro' $file | wc -l | tr -d '\n' && printf "\t" \
69 && egrep -o 'kinshasa|zaire|l.opoldville|congo' $file | wc -l | tr -d '\n' && printf "\t" \
\
70 && grep -o '<cook' $file | wc -l | tr -d '\n' && printf "\t" \
71 && egrep -o 'costa.rica|costarica' $file | wc -l | tr -d '\n' && printf "\t" \
72 && egrep -o 'ivoire|ivory.coast' $file | wc -l | tr -d '\n' && printf "\t" \
73 && grep -o 'croatia' $file | wc -l | tr -d '\n' && printf "\t" \
74 && grep -o '<cuba' $file | wc -l | tr -d '\n' && printf "\t" \
75 && egrep -o 'bonaire|bcuracao|bcuração' $file | wc -l | tr -d '\n' && printf "\t" \
76 && grep -o 'cyprus' $file | wc -l | tr -d '\n' && printf "\t" \
77 && grep -o 'czechoslovakia' $file | wc -l | tr -d '\n' && printf "\t" \
78 && egrep -o 'czech repu|czechia|bohemia' $file | wc -l | tr -d '\n' && printf "\t" \
79 && egrep -o 'congo|kinshasa|zaire|l.opoldville|drc|rdc' $file | wc -l | tr -d '\n' &&
printf "\t" \
80 && grep -o 'denmark' $file | wc -l | tr -d '\n' && printf "\t" \
81 && grep -o 'djibouti' $file | wc -l | tr -d '\n' && printf "\t" \
82 && egrep -o 'dominica|dominican' $file | wc -l | tr -d '\n' && printf "\t" \
83 && egrep -o 'dominican.rep|dominican' $file | wc -l | tr -d '\n' && printf "\t" \
84 && grep -o 'ecuador' $file | wc -l | tr -d '\n' && printf "\t" \
85 && grep -o 'egypt' $file | wc -l | tr -d '\n' && printf "\t" \
86 && egrep -o 'el.salvador|elsalvador' $file | wc -l | tr -d '\n' && printf "\t" \
87 && grep -o 'equatorial guine' $file | wc -l | tr -d '\n' && printf "\t" \
88 && grep -o 'eritrea' $file | wc -l | tr -d '\n' && printf "\t" \
89 && grep -o 'estonia' $file | wc -l | tr -d '\n' && printf "\t" \
90 && egrep -o 'ethiopia|abyssinia' $file | wc -l | tr -d '\n' && printf "\t" \
91 && egrep -o 'falkland|malvinas' $file | wc -l | tr -d '\n' && printf "\t" \
92 && egrep -o 'faroe|faeroe' $file | wc -l | tr -d '\n' && printf "\t" \
93 && grep -o 'german' $file | wc -l | tr -d '\n' && printf "\t" \
94 && grep -o 'fiji' $file | wc -l | tr -d '\n' && printf "\t" \
95 && grep -o 'finland' $file | wc -l | tr -d '\n' && printf "\t" \
96 && egrep -o 'martinique|france|french|\bgaul' $file | wc -l | tr -d '\n' && printf "\t" \
97 && grep -o 'french.guiana' $file | wc -l | tr -d '\n' && printf "\t" \
98 && egrep -o 'french.polynesia|tahiti' $file | wc -l | tr -d '\n' && printf "\t" \
99 && grep -o 'french.southern' $file | wc -l | tr -d '\n' && printf "\t" \
100 && grep -o 'gabon' $file | wc -l | tr -d '\n' && printf "\t" \
101 && grep -o 'gambia' $file | wc -l | tr -d '\n' && printf "\t" \
102 && grep -o 'georgia' $file | wc -l | tr -d '\n' && printf "\t" \
103 && egrep -o 'german.democratic.republic|democratic.republic.germany|east.germany|eastern
germany' $file | wc -l | tr -d '\n' && printf "\t" \
104 && grep -o 'german' $file | wc -l | tr -d '\n' && printf "\t" \
105 && egrep -o 'ghana|gold.coast' $file | wc -l | tr -d '\n' && printf "\t" \

```

```

106 && grep -o 'gibraltar' $file | wc -l | tr -d '\n' && printf "\t" \
107 && egrep -o 'greece|hellenic|hellas' $file | wc -l | tr -d '\n' && printf "\t" \
108 && grep -o 'greenland' $file | wc -l | tr -d '\n' && printf "\t" \
109 && grep -o 'grenada' $file | wc -l | tr -d '\n' && printf "\t" \
110 && grep -o 'guadeloupe' $file | wc -l | tr -d '\n' && printf "\t" \
111 && grep -o '\bguam' $file | wc -l | tr -d '\n' && printf "\t" \
112 && grep -o 'guatemala' $file | wc -l | tr -d '\n' && printf "\t" \
113 && grep -o 'guernsey' $file | wc -l | tr -d '\n' && printf "\t" \
114 && grep -o 'guinea' $file | wc -l | tr -d '\n' && printf "\t" \
115 && egrep -o 'bissau|guinea' $file | wc -l | tr -d '\n' && printf "\t" \
116 && egrep -o 'guyana|british.guiana' $file | wc -l | tr -d '\n' && printf "\t" \
117 && grep -o 'haiti' $file | wc -l | tr -d '\n' && printf "\t" \
118 && grep -o 'hanover' $file | wc -l | tr -d '\n' && printf "\t" \
119 && egrep -o 'heard|mcdonald' $file | wc -l | tr -d '\n' && printf "\t" \
120 && grep -o 'hesse' $file | wc -l | tr -d '\n' && printf "\t" \
121 && grep -o 'hess.grand.ducal' $file | wc -l | tr -d '\n' && printf "\t" \
122 && egrep -o 'holy.see|vatican|papal.st' $file | wc -l | tr -d '\n' && printf "\t" \
123 && grep -o 'honduras' $file | wc -l | tr -d '\n' && printf "\t" \
124 && grep -o 'hong.kong' $file | wc -l | tr -d '\n' && printf "\t" \
125 && grep -o 'hungary' $file | wc -l | tr -d '\n' && printf "\t" \
126 && grep -o 'iceland' $file | wc -l | tr -d '\n' && printf "\t" \
127 && grep -o '\<india>' $file | wc -l | tr -d '\n' && printf "\t" \
128 && grep -o 'indonesia' $file | wc -l | tr -d '\n' && printf "\t" \
129 && egrep -o '\<iran|persia' $file | wc -l | tr -d '\n' && printf "\t" \
130 && egrep -o '\<iraq|mesopotamia|\<irak' $file | wc -l | tr -d '\n' && printf "\t" \
131 && egrep -o 'ireland|irish' $file | wc -l | tr -d '\n' && printf "\t" \
132 && grep -o 'isle of man' $file | wc -l | tr -d '\n' && printf "\t" \
133 && grep -o 'israel' $file | wc -l | tr -d '\n' && printf "\t" \
134 && grep -o 'italy' $file | wc -l | tr -d '\n' && printf "\t" \
135 && grep -o 'jamaica' $file | wc -l | tr -d '\n' && printf "\t" \
136 && grep -o 'japan' $file | wc -l | tr -d '\n' && printf "\t" \
137 && grep -o 'jersey' $file | wc -l | tr -d '\n' && printf "\t" \
138 && grep -o 'jordan' $file | wc -l | tr -d '\n' && printf "\t" \
139 && grep -o 'kazak' $file | wc -l | tr -d '\n' && printf "\t" \
140 && egrep -o 'kenya|british.east.africa|east.africa.prot' $file | wc -l | tr -d '\n' &&
printf "\t" \
141 && grep -o 'kiribati' $file | wc -l | tr -d '\n' && printf "\t" \
142 && grep -o 'korea' $file | wc -l | tr -d '\n' && printf "\t" \
143 && egrep -o 'north.\bkorea|dprk' $file | wc -l | tr -d '\n' && printf "\t" \
144 && grep -o 'kosovo' $file | wc -l | tr -d '\n' && printf "\t" \
145 && grep -o 'kuwait' $file | wc -l | tr -d '\n' && printf "\t" \
146 && egrep -o 'kyrgyz|kirghiz|kirgiz' $file | wc -l | tr -d '\n' && printf "\t" \
147 && grep -o '\blaos\b' $file | wc -l | tr -d '\n' && printf "\t" \
148 && grep -o 'latvia' $file | wc -l | tr -d '\n' && printf "\t" \
149 && grep -o 'lebanon' $file | wc -l | tr -d '\n' && printf "\t" \
150 && egrep -o 'lesotho|basuto' $file | wc -l | tr -d '\n' && printf "\t" \
151 && grep -o 'liberia' $file | wc -l | tr -d '\n' && printf "\t" \
152 && grep -o 'libya' $file | wc -l | tr -d '\n' && printf "\t" \
153 && grep -o 'liechtenstein' $file | wc -l | tr -d '\n' && printf "\t" \
154 && grep -o 'lithuania' $file | wc -l | tr -d '\n' && printf "\t" \
155 && grep -o 'luxem' $file | wc -l | tr -d '\n' && printf "\t" \
156 && egrep -o 'macao|macau' $file | wc -l | tr -d '\n' && printf "\t" \
157 && egrep -o 'madagascar|malagasy' $file | wc -l | tr -d '\n' && printf "\t" \
158 && egrep -o 'malawi|nyasa' $file | wc -l | tr -d '\n' && printf "\t" \
159 && grep -o 'malaysia' $file | wc -l | tr -d '\n' && printf "\t" \
160 && grep -o 'maldiv' $file | wc -l | tr -d '\n' && printf "\t" \
161 && grep -o '\bmali\b' $file | wc -l | tr -d '\n' && printf "\t" \

```

```
162 && grep -o '\bmalta' $file | wc -l | tr -d '\n' && printf "\t" \
163 && grep -o 'marshall.isl' $file | wc -l | tr -d '\n' && printf "\t" \
164 && grep -o 'martinique' $file | wc -l | tr -d '\n' && printf "\t" \
165 && grep -o 'mauritania' $file | wc -l | tr -d '\n' && printf "\t" \
166 && grep -o 'mauritius' $file | wc -l | tr -d '\n' && printf "\t" \
167 && grep -o '\bmayotte' $file | wc -l | tr -d '\n' && printf "\t" \
168 && grep -o 'mecklenburg.schwerin' $file | wc -l | tr -d '\n' && printf "\t" \
169 && grep -o '\bmexic' $file | wc -l | tr -d '\n' && printf "\t" \
170 && egrep -o 'micronesia|micronesia' $file | wc -l | tr -d '\n' && printf "\t" \
171 && grep -o 'modena' $file | wc -l | tr -d '\n' && printf "\t" \
172 && grep -o 'monaco' $file | wc -l | tr -d '\n' && printf "\t" \
173 && grep -o 'mongolia' $file | wc -l | tr -d '\n' && printf "\t" \
174 && grep -o 'montenegro' $file | wc -l | tr -d '\n' && printf "\t" \
175 && grep -o 'montserrat' $file | wc -l | tr -d '\n' && printf "\t" \
176 && egrep -o 'morocco|\bmaroc|moroccan' $file | wc -l | tr -d '\n' && printf "\t" \
177 && grep -o 'mozambique' $file | wc -l | tr -d '\n' && printf "\t" \
178 && egrep -o 'myanmar|burma' $file | wc -l | tr -d '\n' && printf "\t" \
179 && grep -o 'namibia' $file | wc -l | tr -d '\n' && printf "\t" \
180 && grep -o 'nauru' $file | wc -l | tr -d '\n' && printf "\t" \
181 && grep -o 'nepal' $file | wc -l | tr -d '\n' && printf "\t" \
182 && egrep -o 'netherlands|holand|dutch' $file | wc -l | tr -d '\n' && printf "\t" \
183 && grep -o 'netherlands.antilles' $file | wc -l | tr -d '\n' && printf "\t" \
184 && grep -o 'new.caledonia' $file | wc -l | tr -d '\n' && printf "\t" \
185 && grep -o 'new.zealand' $file | wc -l | tr -d '\n' && printf "\t" \
186 && grep -o 'nicaragua' $file | wc -l | tr -d '\n' && printf "\t" \
187 && grep -o '\bniger\>' $file | wc -l | tr -d '\n' && printf "\t" \
188 && grep -o 'nigeria' $file | wc -l | tr -d '\n' && printf "\t" \
189 && grep -o '\<niue' $file | wc -l | tr -d '\n' && printf "\t" \
190 && grep -o 'norfolk' $file | wc -l | tr -d '\n' && printf "\t" \
191 && grep -o 'mariana' $file | wc -l | tr -d '\n' && printf "\t" \
192 && grep -o 'norway' $file | wc -l | tr -d '\n' && printf "\t" \
193 && egrep -o '\boman\>|trucial|omanian' $file | wc -l | tr -d '\n' && printf "\t" \
194 && grep -o 'pakistan' $file | wc -l | tr -d '\n' && printf "\t" \
195 && grep -o 'palau' $file | wc -l | tr -d '\n' && printf "\t" \
196 && egrep -o 'palestin|\bgaza|west.bank' $file | wc -l | tr -d '\n' && printf "\t" \
197 && grep -o 'panama' $file | wc -l | tr -d '\n' && printf "\t" \
198 && egrep -o 'papua|new.guinea' $file | wc -l | tr -d '\n' && printf "\t" \
199 && grep -o 'paraguay' $file | wc -l | tr -d '\n' && printf "\t" \
200 && grep -o 'parma' $file | wc -l | tr -d '\n' && printf "\t" \
201 && grep -o 'peru' $file | wc -l | tr -d '\n' && printf "\t" \
202 && grep -o 'philippines' $file | wc -l | tr -d '\n' && printf "\t" \
203 && grep -o 'pitcairn' $file | wc -l | tr -d '\n' && printf "\t" \
204 && grep -o 'poland' $file | wc -l | tr -d '\n' && printf "\t" \
205 && grep -o 'portugal' $file | wc -l | tr -d '\n' && printf "\t" \
206 && grep -o 'puerto.rico' $file | wc -l | tr -d '\n' && printf "\t" \
207 && grep -o 'qatar' $file | wc -l | tr -d '\n' && printf "\t" \
208 && grep -o 'south.\bkorea' $file | wc -l | tr -d '\n' && printf "\t" \
209 && egrep -o 'moldov|bessarabia|bassarabia' $file | wc -l | tr -d '\n' && printf "\t" \
210 && grep -o 'NA' $file | wc -l | tr -d '\n' && printf "\t" \
211 && egrep -o 'reunion|r union' $file | wc -l | tr -d '\n' && printf "\t" \
212 && egrep -o 'romania|rumania|roumania' $file | wc -l | tr -d '\n' && printf "\t" \
213 && egrep -o '\brussia|soviet|u\.\?s\.\?r|socialist.?republics' $file | wc -l | tr -d '\n' && printf "\t" \
214 && grep -o 'rwanda' $file | wc -l | tr -d '\n' && printf "\t" \
215 && egrep -o 'barthelemy|barth lemy' $file | wc -l | tr -d '\n' && printf "\t" \
216 && grep -o 'helena' $file | wc -l | tr -d '\n' && printf "\t" \
217 && egrep -o 'kitts|\bnevis' $file | wc -l | tr -d '\n' && printf "\t" \
```



```

218 && grep -o '\blucia' $file | wc -l | tr -d '\n' && printf "\t" \
219 && grep -o '\<martin\>' $file | wc -l | tr -d '\n' && printf "\t" \
220 && grep -o 'miquelon' $file | wc -l | tr -d '\n' && printf "\t" \
221 && grep -o 'vincent' $file | wc -l | tr -d '\n' && printf "\t" \
222 && grep -o 'samoa' $file | wc -l | tr -d '\n' && printf "\t" \
223 && egrep -o 'san.marino|sanmarino' $file | wc -l | tr -d '\n' && printf "\t" \
224 && egrep -o '\bsao.tome|\bsão.tome|\bsao.tomé|\bsão.tomé' $file | wc -l | tr -d '\n' &&
printf "\t" \
225 && grep -o '\<saudi' $file | wc -l | tr -d '\n' && printf "\t" \
226 && grep -o 'saxony' $file | wc -l | tr -d '\n' && printf "\t" \
227 && grep -o 'senegal' $file | wc -l | tr -d '\n' && printf "\t" \
228 && grep -o 'serbia' $file | wc -l | tr -d '\n' && printf "\t" \
229 && grep -o 'seychell' $file | wc -l | tr -d '\n' && printf "\t" \
230 && grep -o 'leone' $file | wc -l | tr -d '\n' && printf "\t" \
231 && grep -o 'singapore' $file | wc -l | tr -d '\n' && printf "\t" \
232 && grep -o 'maarten' $file | wc -l | tr -d '\n' && printf "\t" \
233 && grep -o '\<slovak' $file | wc -l | tr -d '\n' && printf "\t" \
234 && grep -o 'slovenia' $file | wc -l | tr -d '\n' && printf "\t" \
235 && grep -o 'solomon.isl' $file | wc -l | tr -d '\n' && printf "\t" \
236 && grep -o 'somali' $file | wc -l | tr -d '\n' && printf "\t" \
237 && egrep -o 'south.africa|southafrica' $file | wc -l | tr -d '\n' && printf "\t" \
238 && egrep -o 'south.georgia|sandwich' $file | wc -l | tr -d '\n' && printf "\t" \
239 && grep -o '\bsouth.sudan' $file | wc -l | tr -d '\n' && printf "\t" \
240 && egrep -o 'spain|spanish' $file | wc -l | tr -d '\n' && printf "\t" \
241 && egrep -o 'sri.lanka|srilanka|ceylon' $file | wc -l | tr -d '\n' && printf "\t" \
242 && grep -o 'sudan' $file | wc -l | tr -d '\n' && printf "\t" \
243 && egrep -o 'surinam|dutch.guiana' $file | wc -l | tr -d '\n' && printf "\t" \
244 && grep -o 'svalbard' $file | wc -l | tr -d '\n' && printf "\t" \
245 && grep -o 'swazi' $file | wc -l | tr -d '\n' && printf "\t" \
246 && egrep -o 'sweden|swedish' $file | wc -l | tr -d '\n' && printf "\t" \
247 && egrep -o 'switz|swiss' $file | wc -l | tr -d '\n' && printf "\t" \
248 && grep -o 'syria' $file | wc -l | tr -d '\n' && printf "\t" \
249 && egrep -o 'taiwan|taipei|formosa' $file | wc -l | tr -d '\n' && printf "\t" \
250 && grep -o 'tajik' $file | wc -l | tr -d '\n' && printf "\t" \
251 && egrep -o 'thai|\bsiam' $file | wc -l | tr -d '\n' && printf "\t" \
252 && grep -o 'macedonia|fyrom' $file | wc -l | tr -d '\n' && printf "\t" \
253 && egrep -o 'NA' $file | wc -l | tr -d '\n' && printf "\t" \
254 && grep -o '\<togo' $file | wc -l | tr -d '\n' && printf "\t" \
255 && grep -o 'tokelau' $file | wc -l | tr -d '\n' && printf "\t" \
256 && grep -o 'tonga' $file | wc -l | tr -d '\n' && printf "\t" \
257 && egrep -o 'trinidad|tobago' $file | wc -l | tr -d '\n' && printf "\t" \
258 && grep -o 'tunisia' $file | wc -l | tr -d '\n' && printf "\t" \
259 && egrep -o '\<turkey|turki|turks' $file | wc -l | tr -d '\n' && printf "\t" \
260 && grep -o 'turkmen' $file | wc -l | tr -d '\n' && printf "\t" \
261 && grep -o 'caicos' $file | wc -l | tr -d '\n' && printf "\t" \
262 && grep -o 'tuscany' $file | wc -l | tr -d '\n' && printf "\t" \
263 && grep -o 'tuvalu' $file | wc -l | tr -d '\n' && printf "\t" \
264 && grep -o 'two.sicilies' $file | wc -l | tr -d '\n' && printf "\t" \
265 && grep -o 'uganda' $file | wc -l | tr -d '\n' && printf "\t" \
266 && grep -o 'ukrain' $file | wc -l | tr -d '\n' && printf "\t" \
267 && egrep -o 'emirates|u\.\?a\.\?e\.\?|united.arab.em' $file | wc -l | tr -d '\n' && printf
"\t" \
268 && egrep -o 'united.kingdom|britain|british|\<uk\>|england' $file | wc -l | tr -d '\n' &&
printf "\t" \
269 && grep -o 'tanzania' $file | wc -l | tr -d '\n' && printf "\t" \
270 && egrep -o 'united.states|\<usa\>|america' $file | wc -l | tr -d '\n' && printf "\t" \
271 && grep -o 'minor.outlying.is' $file | wc -l | tr -d '\n' && printf "\t" \

```

```

272 && grep -o 'uruguay' $file | wc -l | tr -d '\n' && printf "\t" \
273 && grep -o 'uzbek' $file | wc -l | tr -d '\n' && printf "\t" \
274 && egrep -o 'vanuatu|new.hebrides' $file | wc -l | tr -d '\n' && printf "\t" \
275 && grep -o 'venezuela' $file | wc -l | tr -d '\n' && printf "\t" \
276 && egrep -o 'viet.nam|vietnam' $file | wc -l | tr -d '\n' && printf "\t" \
277 && grep -o 'virgin.isl' $file | wc -l | tr -d '\n' && printf "\t" \
278 && grep -o 'virgin.isl' $file | wc -l | tr -d '\n' && printf "\t" \
279 && egrep -o 'futuna|wallis' $file | wc -l | tr -d '\n' && printf "\t" \
280 && grep -o 'western.sahara' $file | wc -l | tr -d '\n' && printf "\t" \
281 && egrep -o 'wuerttemberg|württemberg' $file | wc -l | tr -d '\n' && printf "\t" \
282 && grep -o 'yemen' $file | wc -l | tr -d '\n' && printf "\t" \
283 && grep -o 'yemen' $file | wc -l | tr -d '\n' && printf "\t" \
284 && grep -o 'yemen' $file | wc -l | tr -d '\n' && printf "\t" \
285 && grep -o 'yugoslavia' $file | wc -l | tr -d '\n' && printf "\t" \
286 && egrep -o 'zambia|rhodesia' $file | wc -l | tr -d '\n' && printf "\t" \
287 && grep -o 'zanzibar' $file | wc -l | tr -d '\n' && printf "\t" \
288 && egrep -o 'zimbabwe|rhodesia' $file | wc -l; done >
mypath/data/datafiles/countriesPerFileC2.txt

289
290 #regions per file
291 cd mypath/data/sotu_corpus2
292 for file in $(ls -l |grep -i .txt) ; do printf "$file\t" \
293 && egrep -o '\<asia\>|in the pacific|pacific ocean|pacific countr|pacific area|on the
pacific|of the pacific|atlantic and pacific|pacific coast|pacific railroad|norht pacific'
$file | wc -l | tr -d '\n' && printf "\t" \
294 && grep -o '\<europ' $file | wc -l | tr -d '\n' && printf "\t" \
295 && grep -o 'africa' $file | wc -l | tr -d '\n' && printf "\t" \
296 && grep -o 'polynesia' $file | wc -l | tr -d '\n' && printf "\t" \
297 && grep -o 'caribbean' $file | wc -l | tr -d '\n' && printf "\t" \
298 && egrep -o '\<south.america|southern.america|latin.america|western.hemisphere' $file |
wc -l | tr -d '\n' && printf "\t" \
299 && egrep -o 'north.america|northern america' $file | wc -l | tr -d '\n' && printf "\t" \
300 && egrep -o 'middle.east|near.east' $file | wc -l | tr -d '\n' && printf "\t" \
301 && grep -o 'south.asia' $file | wc -l; done > mypath/data/datafiles/regionsPerFileC2.txt
302
303

```