

---

## Aufgabe 4

Folgender Befehl druckt die siebte und elfte Spalte unseres durch Tabs getrennten Korpus in die Datei ouput.txt für weitere Verwendung:

```
awk -F '\t' '{print $7 "\t" $11}' tiger_merkmale.txt > output.txt
```

Die Datei output.txt beschreibt also jeweils die Beziehung zwischen grammatischer Funktion und dem entsprechenden Muster. Im nächsten Schritt gilt es diejenigen Zeilen auszufiltern, die die gewünschten grammatischen Funktionen SB und AO mit Merkmalsmustern in Verbindung setzen.

Hierfür erweist sich das grep Tool als hilfreich. Mit folgenden Befehlen werden aus der Datei output.txt zu den beiden Funktionen SB und AO Merkmalsmusterlisten gewonnen und in gleichnamigen Dateien abgelegt:

```
grep -E "SB" output.txt > tmp.txt; awk -F '\t' '{print $1}' tmp.txt > SB.txt
grep -E "OA" output.txt > tmp.txt; awk -F '\t' '{print $1}' tmp.txt > OA.txt
```

Nun interessieren uns jedoch nur Merkmalsmuster, die mindestens zwanzig mal auftreten. Zu diesem Zweck sortieren wir nun zunächst die Musterlisten:

```
sort -o SB.txt SB.txt
sort -o OA.txt OA.txt
```

Nun interessiert uns die Anzahl der Vorkommen jedes Merkmalsmusters. Über das uniq Tool können wir die sortierten Listen nun kompakt darstellen. Im Anschluss werden die Listen erneut sortiert, dieses mal nach Anzahl der Vorkommen der Muster:

```
uniq -c SB.txt tmp.txt; cat tmp.txt > SB.txt; sort -o SB.txt SB.txt
uniq -c OA.txt tmp.txt; cat tmp.txt > OA.txt; sort -o OA.txt OA.txt
```

Nun gilt es lediglich alle unwichtigen Merkmalsmuster zu ignorieren. Das kann man über reguläre Ausdrücke realisieren. `\s[0-1]?[0-9]\s` matched alle separaten Vorkommen von Zahlen, die geringer als zwanzig sind. Diese Zeilen interessieren uns aber gerade nicht, weswegen wir die `-v` (invert match) Option von grep nutzen.

```
grep -v -E "\s[0-1]?[0-9]\s" SB.txt
grep -v -E "\s[0-1]?[0-9]\s" OA.txt
```

Das Ergebnis der Befehle sind nun Listen aller mehr als zwanzig Mal vorkommenden Merkmalsmuster, die SB bzw. OA zuzuordnen sind.

SB:

20 number=pllperson=3ltense=pastlmood=subj  
22 case=nomlnumber=sglgender=neutldegree=pos  
29 case=nomlnumber=sglgender=femlperson=1  
30 case=\*lnumber=\*lgender=fem  
31 number=sglperson=3ltense=pastlmood=subj  
35 case=nomlnumber=sglgender=mascldegree=pos  
43 case=nomlnumber=pllgender=\*lperson=2  
49 number=pllperson=3ltense=pastlmood=ind  
57 case=nomlnumber=sglgender=\*lperson=2  
67 case=nomlnumber=pllgender=neutlperson=3  
89 case=datlnumber=sglgender=neut  
89 case=nomlnumber=sglgender=masclperson=1  
105 number=sglperson=3ltense=preslmood=subj  
111 case=nomlnumber=pllgender=femlperson=3  
111 number=sglperson=3ltense=pastlmood=ind  
158 case=\*lnumber=\*lgender=\*  
159 case=datlnumber=sglgender=masc  
163 number=pllperson=3ltense=preslmood=ind  
263 case=nomlnumber=pllgender=masclperson=3  
313 case=\*lnumber=\*lgender=neut  
558 number=sglperson=3ltense=preslmood=ind  
594 case=nomlnumber=sglgender=\*lperson=1  
926 case=nomlnumber=pllgender=\*lperson=1  
1033 case=nomlnumber=sglgender=femlperson=3  
12402 case=nomlnumber=sglgender=fem  
1258 case=nomlnumber=pllgender=\*lperson=3  
13897 case=nomlnumber=sglgender=masc  
1500 case=nomlnumber=sglgender=neutlperson=3  
1793 case=nomlnumber=pllgender=\*  
1838 \_  
2132 case=nomlnumber=pllgender=neut  
3006 case=nomlnumber=sglgender=masclperson=3  
4264 case=nomlnumber=pllgender=fem  
4478 case=nomlnumber=sglgender=\*  
6620 case=nomlnumber=pllgender=masc  
7647 case=nomlnumber=sglgender=neut

OA:

21 case=\*lnumber=\*lgender=masc  
21 case=acclnumber=\*lgender=neut  
28 case=acclnumber=pllgender=neutlperson=3  
28 case=nomlnumber=sglgender=neutlperson=3  
29 case=acclnumber=\*lgender=masc  
34 case=acclnumber=pllgender=masclperson=3  
35 case=acclnumber=pllgender=femlperson=3  
38 case=acclnumber=sglperson=1  
39 case=nomlnumber=sglgender=masc  
42 case=acclnumber=pllperson=1  
49 case=acclnumber=sglgender=\*lperson=1  
51 case=acclnumber=pllgender=\*lperson=3  
71 case=acclnumber=\*lgender=\*lperson=3  
77 case=acclnumber=pllgender=\*lperson=1  
88 case=acclnumber=sglgender=femlperson=3

94 case=\*lnumber=\*lgender=fem  
126 \_  
200 case=\*lnumber=\*lgender=\*  
231 case=acclnumber=sglgender=masc|person=3  
238 case=acclnumber=sglgender=\*  
376 case=acclnumber=pllgender=\*  
404 case=acclnumber=sglgender=neut|person=3  
468 case=\*lnumber=\*lgender=neut  
1473 case=acclnumber=pllgender=neut  
1478 case=acclnumber=pllperson=3  
2565 case=acclnumber=pllgender=masc  
3220 case=acclnumber=sglperson=3  
3437 case=acclnumber=pllgender=fem  
4628 case=acclnumber=sglgender=neut  
5496 case=acclnumber=sglgender=masc  
7562 case=acclnumber=sglgender=fem