

# **Unterschiedliche Textprofile aufgrund Wortdistribution**

- ◆ **Wortschatzzunahme** (Markus Piller)
- ◆ **Wortwiederholungen** (Christoph Wolter)
- ◆ **Binnenwortschatz** (Markus Widl)
- ◆ **Cluster** (Max Hütter)

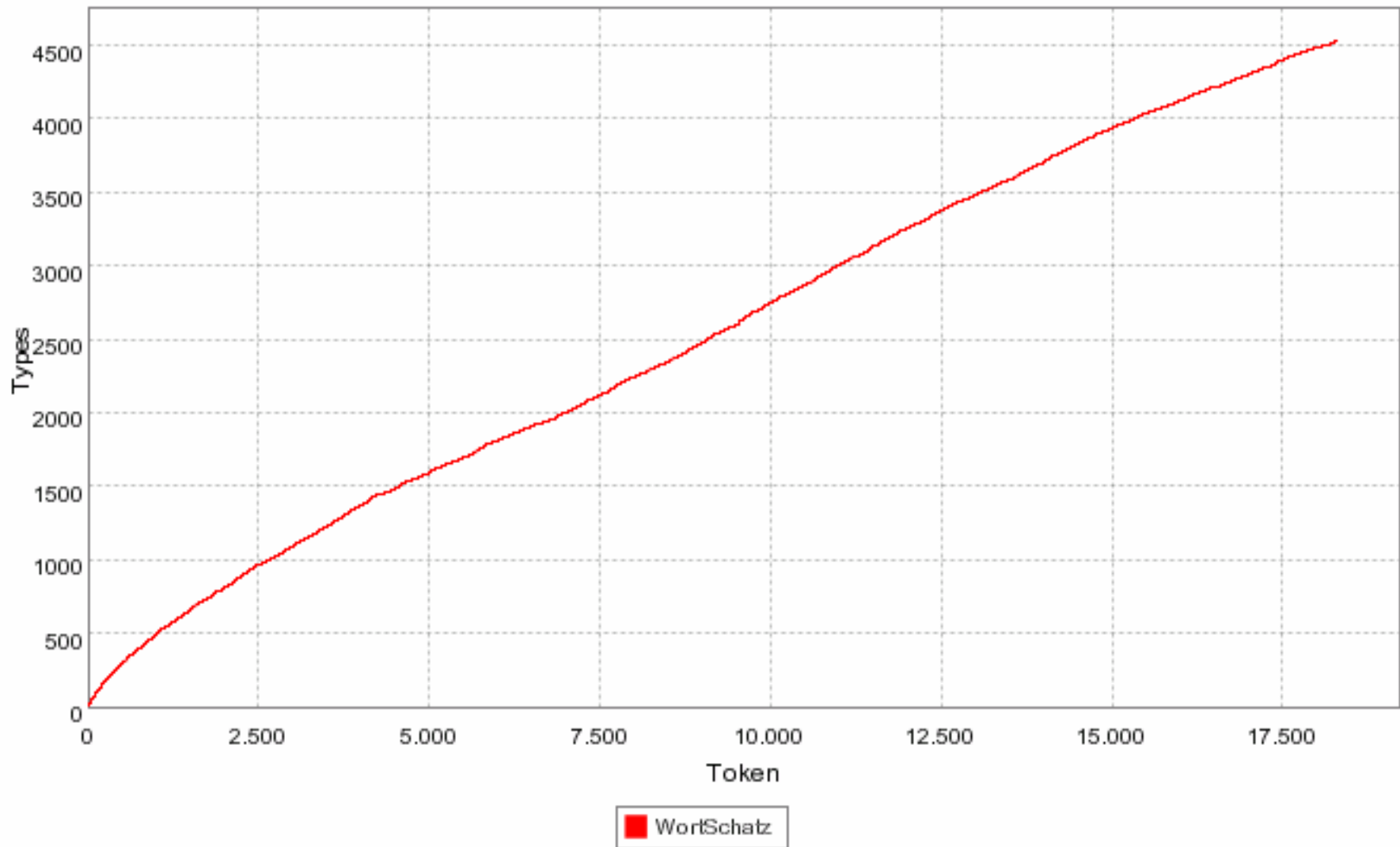
# Vier bearbeitete Texte

- ◆ **„Der Findling“**  
von Kleist                      5992 Token
- ◆ **„Novelle“**  
von Goethe                      7091 Token
- ◆ **„Der Mann von fünfzig Jahren“**  
von Goethe                      18287 Token
- ◆ **„Aus dem Leben eines Taugenichts“**  
von Eichendorff                31146 Token

# **I. Bereichsabgrenzung aufgrund erhöhter Wortschatzzunahme**

# **Neuer Algorithmus über die Steigung der Wortschatzkurve**

# Beispiel einer Wortschatzkurve



# Strikte Wortschatzzunahme

(Der Taugenichts)

von	bis	# Tokens
1	24	24
26	31	6
35	39	5
46	52	7
95	100	6
113	117	5
180	184	5
230	234	5
289	293	5
435	439	5
462	466	5
542	547	6
733	737	5
957	961	5
1500	1504	5
2401	2405	5
4095	4099	5
9426	9431	6
23379	23383	5
25499	25503	5
27935	27942	8
27985	27993	9
30098	30103	6

5% Token = 66% Types  
der Liste

(Der Mann von 50 Jahren)

von	bis	# Tokens
1	19	19
21	30	10
48	53	6
81	87	7
89	94	6
129	134	6
161	166	6
225	231	7
238	244	7
338	346	9
712	717	6
1024	1032	9
1481	1487	7
9530	9540	11
9686	9700	15

5% Token = 69,6% Types  
der Liste

(Der Findling)

von	bis	# Tokens
1	9	9
11	34	24
36	46	11
50	59	10
65	70	6
78	83	6
500	504	5
541	545	5
981	987	7
1056	1060	5

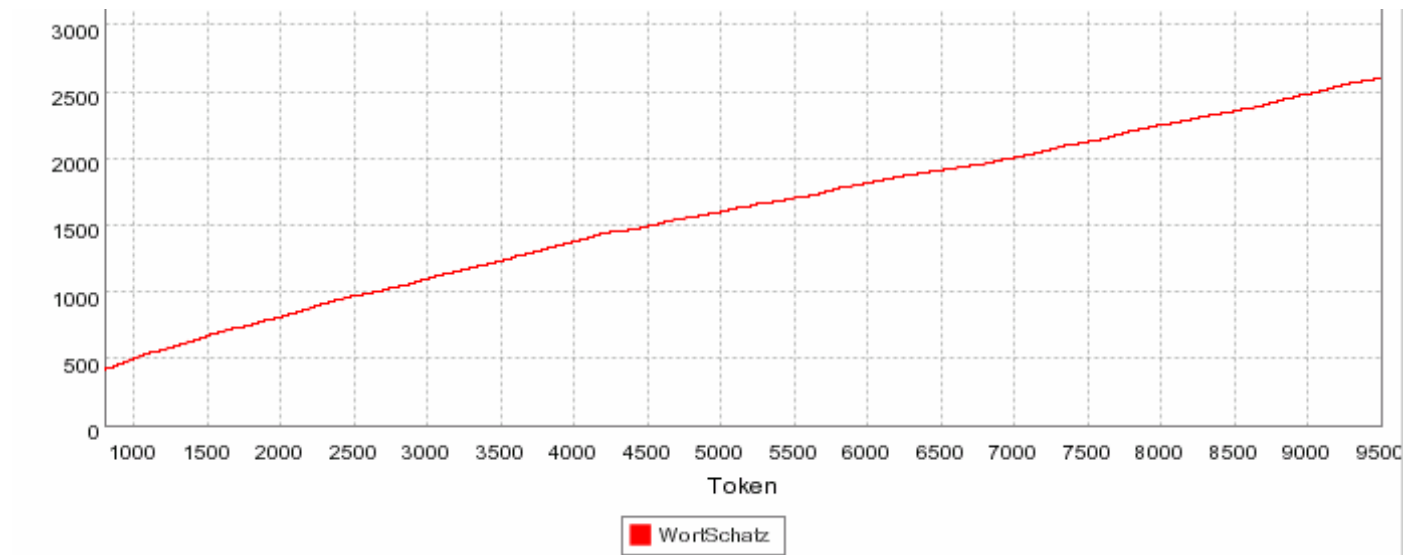
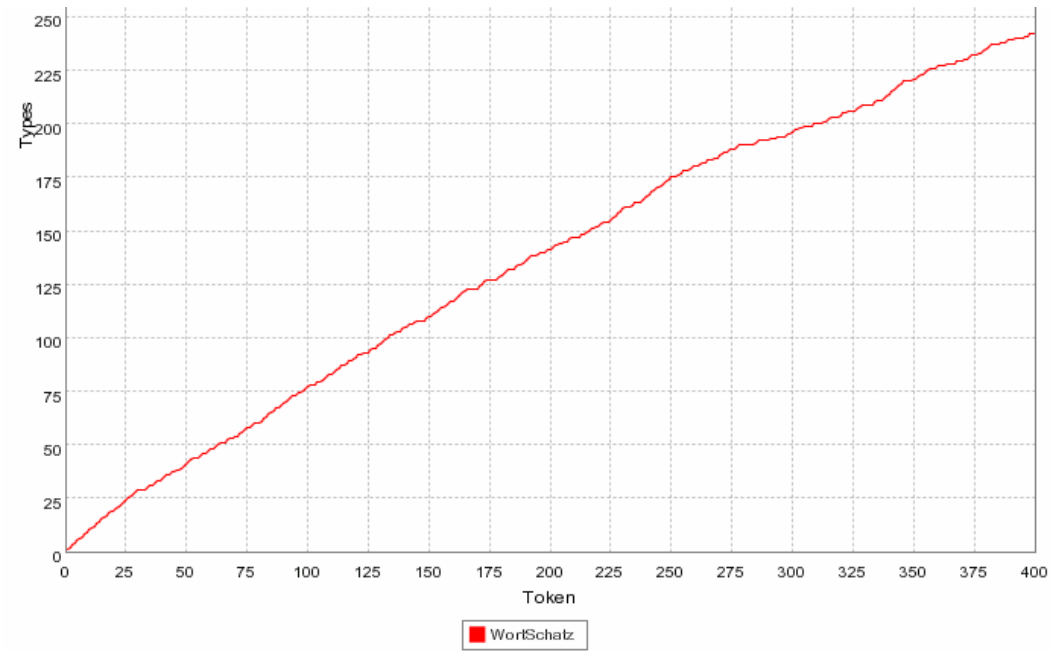
5% Token = 75% Types  
der Liste

5%

5%

5%

# Strikte Wortschatzzunahme



# Strikte Wortschatzzunahme

(Der Taugenichts)

Datei		
von	bis	# Tokens
1	24	24
26	31	6
35	39	5
46	52	7
95	100	6
113	117	5
180	184	5
230	234	5
289	293	5
435	439	5
462	466	5
542	547	6
733	737	5
957	961	5
1500	1504	5
2401	2405	5
4095	4099	5
9426	9431	6
23379	23383	5
25499	25503	5
27935	27942	8
27985	27993	9
30098	30103	6

(Der Mann von 50 Jahren)

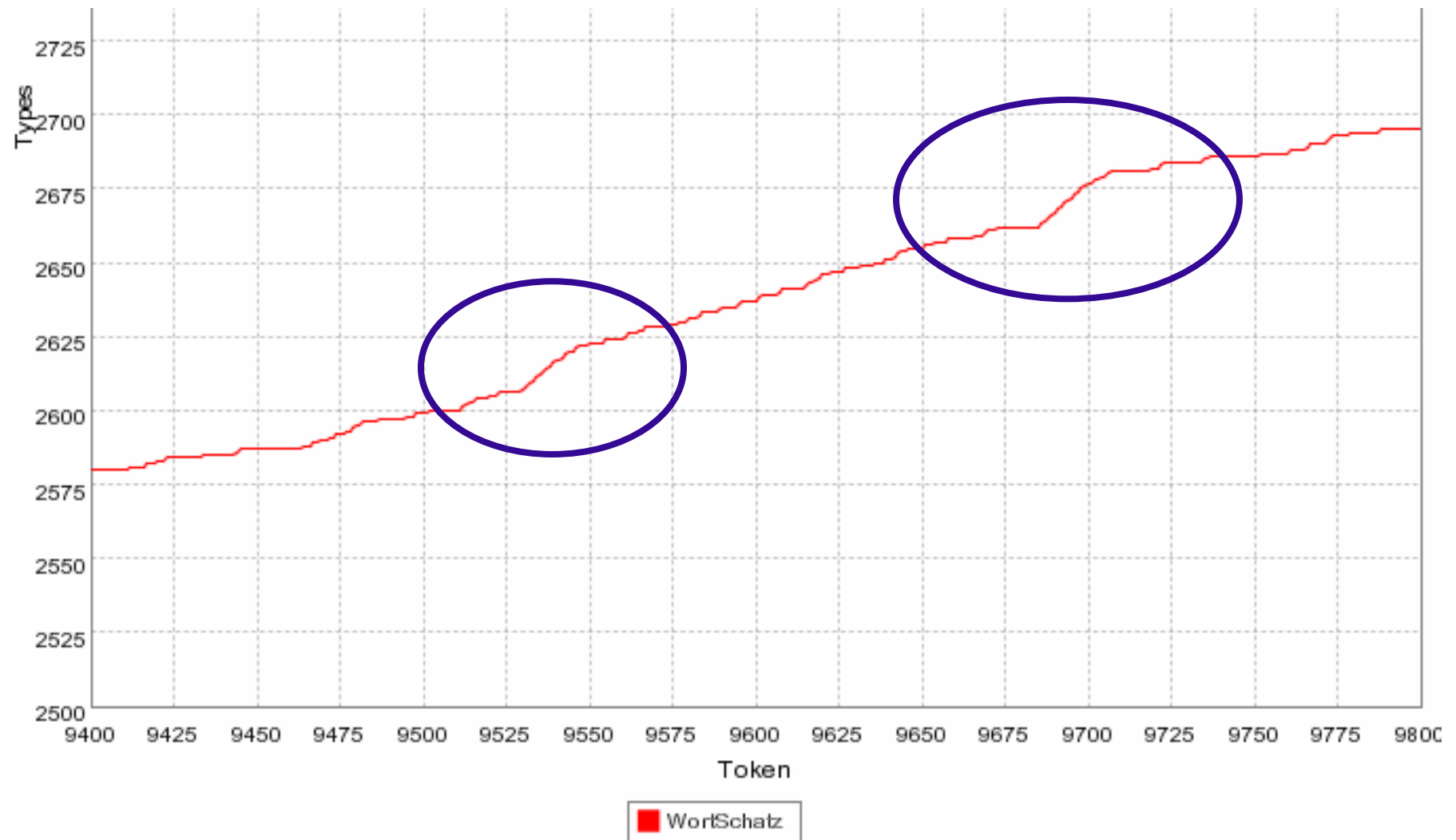
Datei		
von	bis	# Tokens
1	19	19
21	30	10
48	53	6
81	87	7
89	94	6
129	134	6
161	166	6
225	231	7
238	244	7
338	346	9
712	717	6
1024	1032	9
1481	1487	7
9530	9540	11
9686	9700	15

(Der Findling)

Datei		
von	bis	# Tokens
1	9	9
11	34	24
36	46	11
50	59	10
65	70	6
78	83	6
500	504	5
541	545	5
981	987	7
1056	1060	5

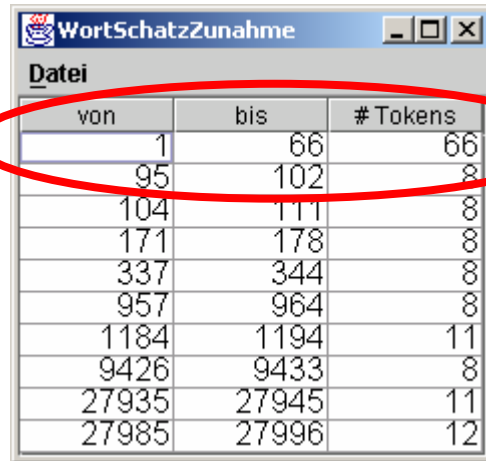


# Strikte Wortschatzzunahme



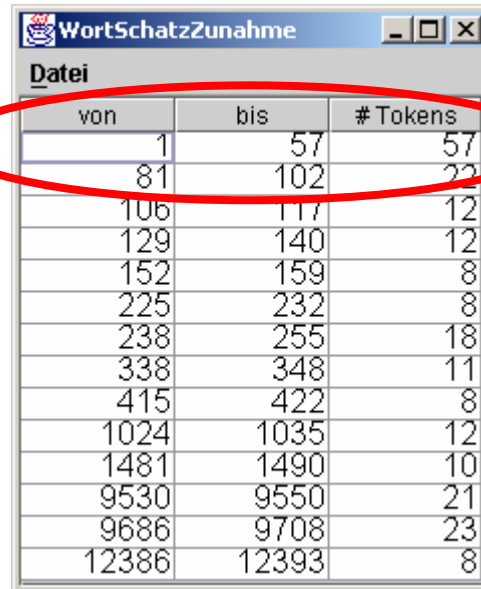
# Relative Wortschatzzunahme mit 80%

(Der Taugenichts)



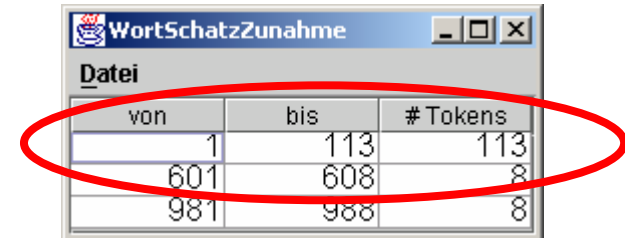
von	bis	# Tokens
1	66	66
95	102	8
104	111	8
171	178	8
337	344	8
957	964	8
1184	1194	11
9426	9433	8
27935	27945	11
27985	27996	12

(Der Mann von 50 Jahren)



von	bis	# Tokens
1	57	57
81	102	22
106	117	12
129	140	12
152	159	8
225	232	8
238	255	18
338	348	11
415	422	8
1024	1035	12
1481	1490	10
9530	9550	21
9686	9708	23
12386	12393	8

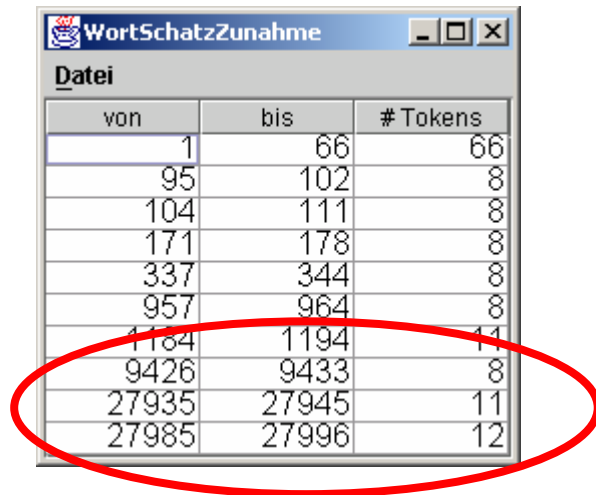
(Der Findling)



von	bis	# Tokens
1	113	113
601	608	8
981	988	8

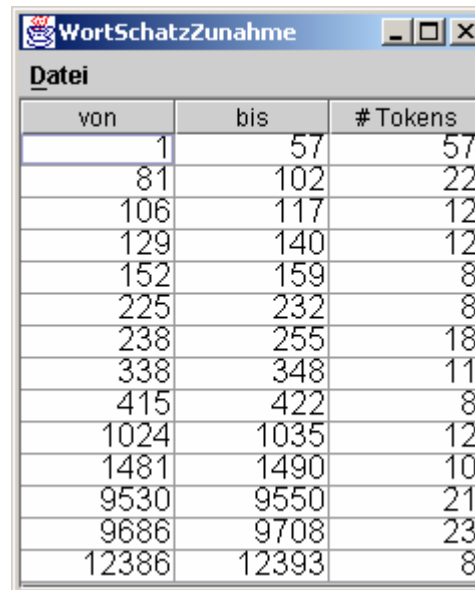
# Relative Wortschatzzunahme mit 80%

(Der Taugenichts)



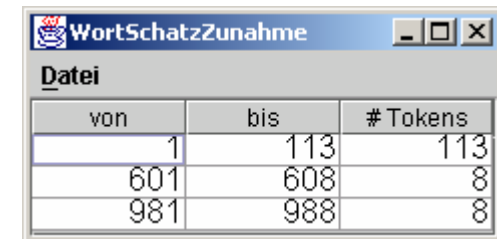
von	bis	# Tokens
1	66	66
95	102	8
104	111	8
171	178	8
337	344	8
957	964	8
1184	1194	11
9426	9433	8
27935	27945	11
27985	27996	12

(Der Mann von 50 Jahren)



von	bis	# Tokens
1	57	57
81	102	22
106	117	12
129	140	12
152	159	8
225	232	8
238	255	18
338	348	11
415	422	8
1024	1035	12
1481	1490	10
9530	9550	21
9686	9708	23
12386	12393	8

(Der Findling)



von	bis	# Tokens
1	113	113
601	608	8
981	988	8

# Relative Wortschatzzunahme mit 80%

