

Dimpel: Gute Wörter und Level-2-Differenzen bei Delta

Level-2-Differenzen:

Differenz aus der

Z-Wert-Differenz zwischen Ratetext und Distraktortext einerseits

und der

Z-Wert-Differenz zwischen Ratetext und Autor-Vergleichstext andererseits

Auf positiven Level-2-Differenzen beruht eine funktionierende Autorerkennung.

Negative Level-2-Differenzen sind ein Störfaktor bei einer Autorerkennung.

Damit lassen sich „gute“ und „schlechte“ Wörter einzeln identifizieren!

F	Wh1	Pz2	Trist3	Mean	StDev	Z1	Wh	Z2	Pz	Trist	Gleiche Autoren		Versch. Autoren		L2Diff
											Z3	Abs Diff	Abs Diff	Z3	
1 der	3,19	2,62	2,13	2,65	0,53	1,02	-0,05	-0,98	1,07	2,00	0,93				
2 daz	2,02	1,93	2,77	2,24	0,46	-0,48	-0,67	1,15	0,19	1,63	1,44				
3 und	1,97	1,01	3,52	2,17	1,27	-0,15	-0,91	1,07	0,76	1,22	0,46				
4 ir	1,33	1,81	2,14	1,76	0,41	-1,06	0,12	0,93	1,18	1,99	0,81				
5 er	1,51	1,63	1,97	1,70	0,24	-0,80	-0,32	1,12	0,48	1,92	1,44				
6 in	1,30	1,25	1,66	1,40	0,23	-0,48	-0,67	1,15	0,19	1,63	1,44				
7 ich	1,29	1,43	1,30	1,34	0,08	-0,67	1,15	-0,48	1,82	0,18	-1,63				
...															
200 zît	0,06	0,06	0,08	0,06	0,01	-0,36	-0,77	1,13	0,41	1,49	1,09				

Zwei Texte von Wolfram: Willehalm (Wh1) und Parzival (Pz2).

Ein Text von Gottfried: Tristan (Trist3).

Delta – Schritt 1: Worthäufigkeiten prozentual ermitteln.

F	Wh1	Pz2	Trist3	Mean	StDev	Z1 Wh	Z2 Pz	Trist	Gleiche Autoren		Versch. Autoren	
									Z3	Abs Diff	Abs Diff	L2Diff
1 der	3,19	2,62	2,13	2,65	0,53	1,02	-0,05	-0,98	1,07	2,00	0,93	
2 daz	2,02	1,93	2,77	2,24	0,46	-0,48	-0,67	1,15	0,19	1,63	1,44	
3 und	1,97	1,01	3,52	2,17	1,27	-0,15	-0,91	1,07	0,76	1,22	0,46	
4 ir	1,33	1,81	2,14	1,76	0,41	-1,06	0,12	0,93	1,18	1,99	0,81	
5 er	1,51	1,63	1,97	1,70	0,24	-0,80	-0,32	1,12	0,48	1,92	1,44	
6 in	1,30	1,25	1,66	1,40	0,23	-0,48	-0,67	1,15	0,19	1,63	1,44	
7 ich	1,29	1,43	1,30	1,34	0,08	-0,67	1,15	-0,48	1,82	0,18	-1,63	
...												
200 zît	0,06	0,06	0,08	0,06	0,01	-0,36	-0,77	1,13	0,41	1,49	1,09	

Delta – Schritt 2: Mittelwerte und Standardabweichung berechnen.

F	Wh1	Pz2	Trist3	Mean	StDev	Z1 Wh	Z2 Pz	Z3 Trist	Gleiche	Versch.	L2Diff
									Autoren	Autoren	
									Abs	Abs	
									Diff	Diff	
									Z1 Z2	Z1 Z3	
1 der	3,19	2,62	2,13	2,65	0,53	1,02	-0,05	-0,98	1,07	2,00	0,93
2 daz	2,02	1,93	2,77	2,24	0,46	-0,48	-0,67	1,15	0,19	1,63	1,44
3 und	1,97	1,01	3,52	2,17	1,27	-0,15	-0,91	1,07	0,76	1,22	0,46
4 ir	1,33	1,81	2,14	1,76	0,41	-1,06	0,12	0,93	1,18	1,99	0,81
5 er	1,51	1,63	1,97	1,70	0,24	-0,80	-0,32	1,12	0,48	1,92	1,44
6 in	1,30	1,25	1,66	1,40	0,23	-0,48	-0,67	1,15	0,19	1,63	1,44
7 ich	1,29	1,43	1,30	1,34	0,08	-0,67	1,15	-0,48	1,82	0,18	-1,63
...											
200 zît	0,06	0,06	0,08	0,06	0,01	-0,36	-0,77	1,13	0,41	1,49	1,09

$$Z_{Wert} = \frac{\text{Messwert} - \text{Mittelwert}}{\text{Standardabweichung}}$$

$$Z_{Wert} = \frac{3,19 - 2,65}{0,53} = 1,02$$

Delta – Schritt 3: Z-Werte berechnen.

F	Wh1	Pz2	Trist3	Mean	StDev	Z1 Wh	Z2 Pz	Trist	Gleiche Autoren		Versch. Autoren		
									Z1	Z2	Abs	Abs	
									Z1	Z2	Z1	Z3	L2Diff
1 der	3,19	2,62	2,13	2,65	0,53	1,02	-0,05	-0,98	1,07		2,00	0,93	
2 daz	2,02	1,93	2,77	2,24	0,46	-0,48	-0,67	1,15	0,19		1,63	1,44	
3 und	1,97	1,01	3,52	2,17	1,27	-0,15	-0,91	1,07	0,76		1,22	0,46	
4 ir	1,33	1,81	2,14	1,76	0,41	-1,06	0,12	0,93	1,18		1,99	0,81	
5 er	1,51	1,63	1,97	1,70	0,24	-0,80	-0,32	1,12	0,48		1,92	1,44	
6 in	1,30	1,25	1,66	1,40	0,23	-0,48	-0,67	1,15	0,19		1,63	1,44	
7 ich	1,29	1,43	1,30	1,34	0,08	-0,67	1,15	-0,48	1,82		0,18	-1,63	
...													
200 zît	0,06	0,06	0,08	0,06	0,01	-0,36	-0,77	1,13	0,41		1,49	1,09	

MW:
Delta

$$Diff. = Z_{Wh} - Z_{Pz} = -1,02 - 0,05 = 1,07$$

Delta – Schritt 4: Differenz der positiven Z-Werte für jedes Wort berechnen.

Delta = Mittelwert all dieser positiven Z-Wert-Differenzen.

F	Wh1	Pz2	Trist3	Mean	StDev	Z1	Wh	Z2	Pz	Trist	Z3		L2Diff
											Z1	Z2	
											Gleiche Autoren Abs Diff	Versch. Autoren Abs Diff	
1 der	3,19	2,62	2,13	2,65	0,53	1,02	-0,05	-0,98	1,07	2,00	0,93	0,93	
2 daz	2,02	1,93	2,77	2,24	0,46	-0,48	-0,67	1,15	0,19	1,63	1,44	1,44	
3 und	1,97	1,01	3,52	2,17	1,27	-0,15	-0,91	1,07	0,76	1,22	0,46	0,46	
4 ir	1,33	1,81	2,14	1,76	0,41	-1,06	0,12	0,93	1,18	1,99	0,81	0,81	
5 er	1,51	1,63	1,97	1,70	0,24	-0,80	-0,32	1,12	0,48	1,92	1,44	1,44	
6 in	1,30	1,25	1,66	1,40	0,23	-0,48	-0,67	1,15	0,19	1,63	1,44	1,44	
7 ich	1,29	1,43	1,30	1,34	0,08	-0,67	1,15	-0,48	1,82	0,18	-1,63	-1,63	
...													

Level-2-Differenz für „der“: Differenz der positiven Z-Wert-Differenz von den Texten der verschiedenen Autoren und der positiven Z-Wert-Differenz von den Texten der identischen Autoren

Gleiche Autoren	Versch. Autoren
Abs	Abs
Diff	Diff
Z1 Z2	Z1 Z3
L2Diff	

F	Wh1	Pz2	Trist3	Mean	StDev	Z1	Wh	Z2	Pz	Trist	Z3	Z1 Z2	Z1 Z3	L2Diff
1 der	3,19	2,62	2,13	2,65	0,53	1,02	-0,05	-0,98				1,07	2,00	0,93
2 daz	2,02	1,93	2,77	2,24	0,46	-0,48	-0,67	1,15				0,19	1,63	1,44
...														
7 ich	1,29	1,43	1,30	1,34	0,08	-0,67	1,15	-0,48				1,82	0,18	-1,63
...														

„der“: Positive L2-Differenz: Geeignet, um Wolfram und Gottfried auseinander zu halten

„ich“: Negative L2-Differenz: Ungeeignet, um Wolfram und Gottfried auseinanderzuhalten.

In diesem einfachen Beispiel mit nur drei Texten ist das auch unmittelbar anhand der Worthäufigkeitswerte links und anhand der Z-Werte nachvollziehbar. Die Analysen in Dimpel, Delta-Rätsel, Dariah-de-Working Papers 25, zeigen jedoch, dass weder erhöhte Worthäufigkeitswerte noch erhöhte Z-Werte allein hinreichend sind, um ein Wort als distinktiv zu identifizieren.