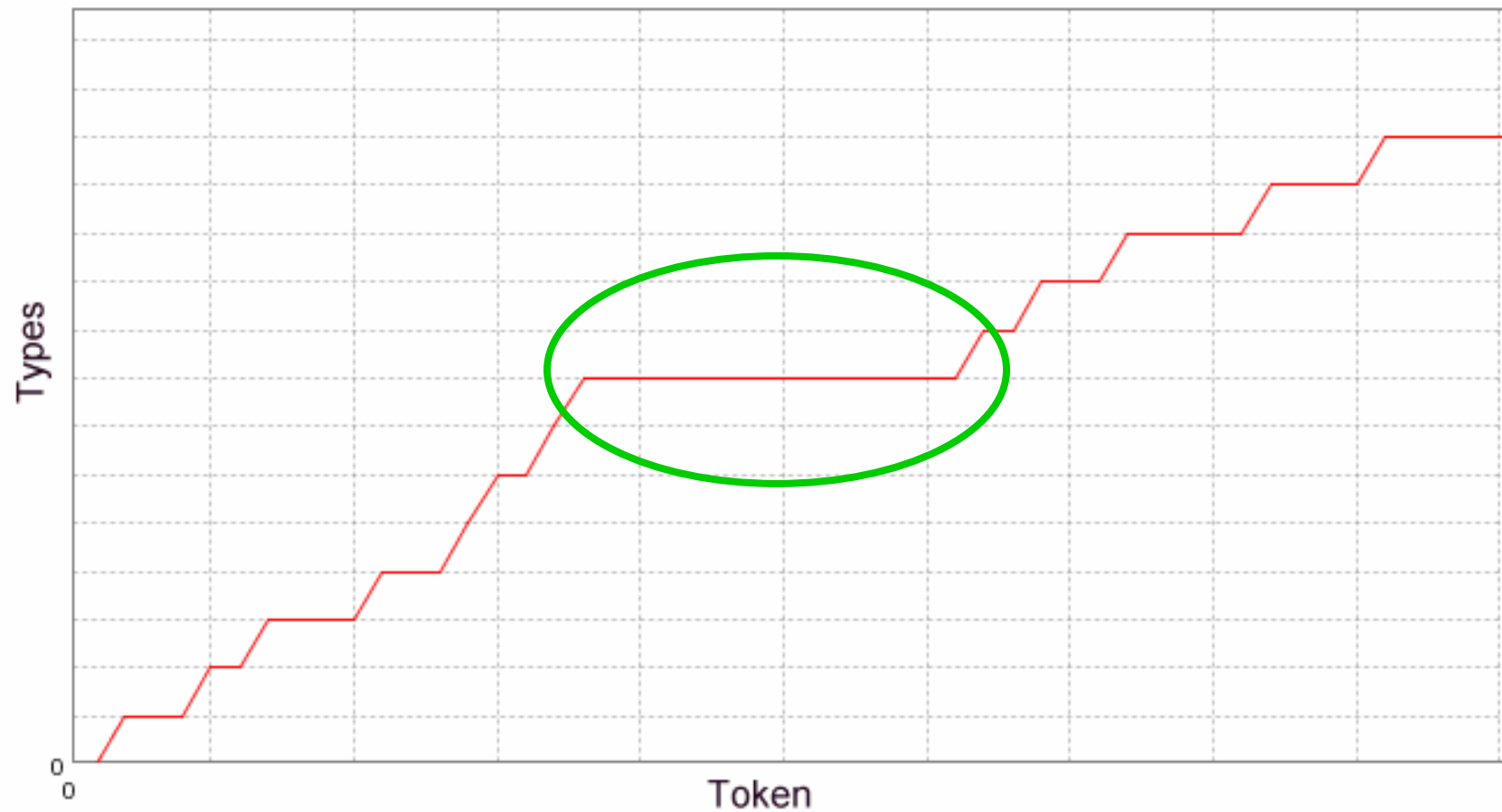


II. Wortwiederholungen:

Textbereiche mit vielen Wortwiederholungen

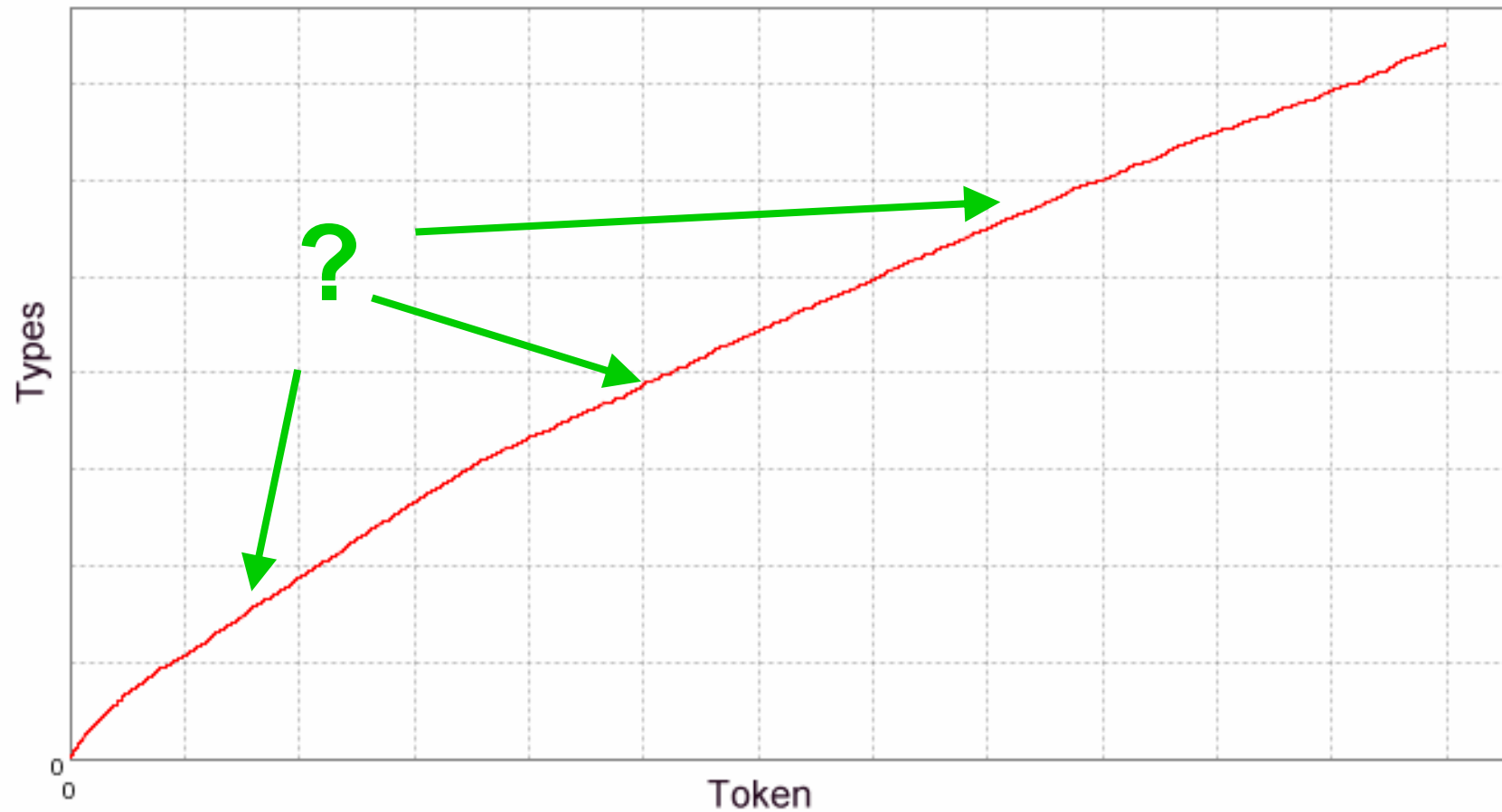
Wortwiederholung

bedeutet Stagnation der Wortschatzkurve



Problem

Auffinden „interessanter“ Bereiche



Lösung

computergestützte Suche:

- ◆ Algorithmus mit Steigungsbetrachtung
(absolute und relative Stagnation)
- ◆ Algorithmus nach W. Bader
(arbeitet mit Gewichtung, relative Stagnation)

Bereiche strikter Stagnation

„Der Findling“ (5.992 Token)

Die 15 längsten Bereiche:

erster Bereich knapp vor
der Mitte des Textes

Großteil der Bereiche im
letzten Drittel des Textes

von	bis	# Tokens
2329	2345	17
2384	2396	13
2883	2896	14
3032	3046	15
3331	3344	14
3574	3586	13
4537	4549	13
4644	4661	18
4940	4952	13
4977	4990	14
5325	5337	13
5370	5383	14
5391	5413	23
5594	5609	16
5699	5716	18
5853	5865	13

Bereiche relativer Stagnation

„Der Findling“ (5.992 Token)

längsten Bereiche bei mind.
75% Stagnation

erste Bereiche bereits in den
ersten 5% des Textes

insgesamt deutlich größere
Bereiche

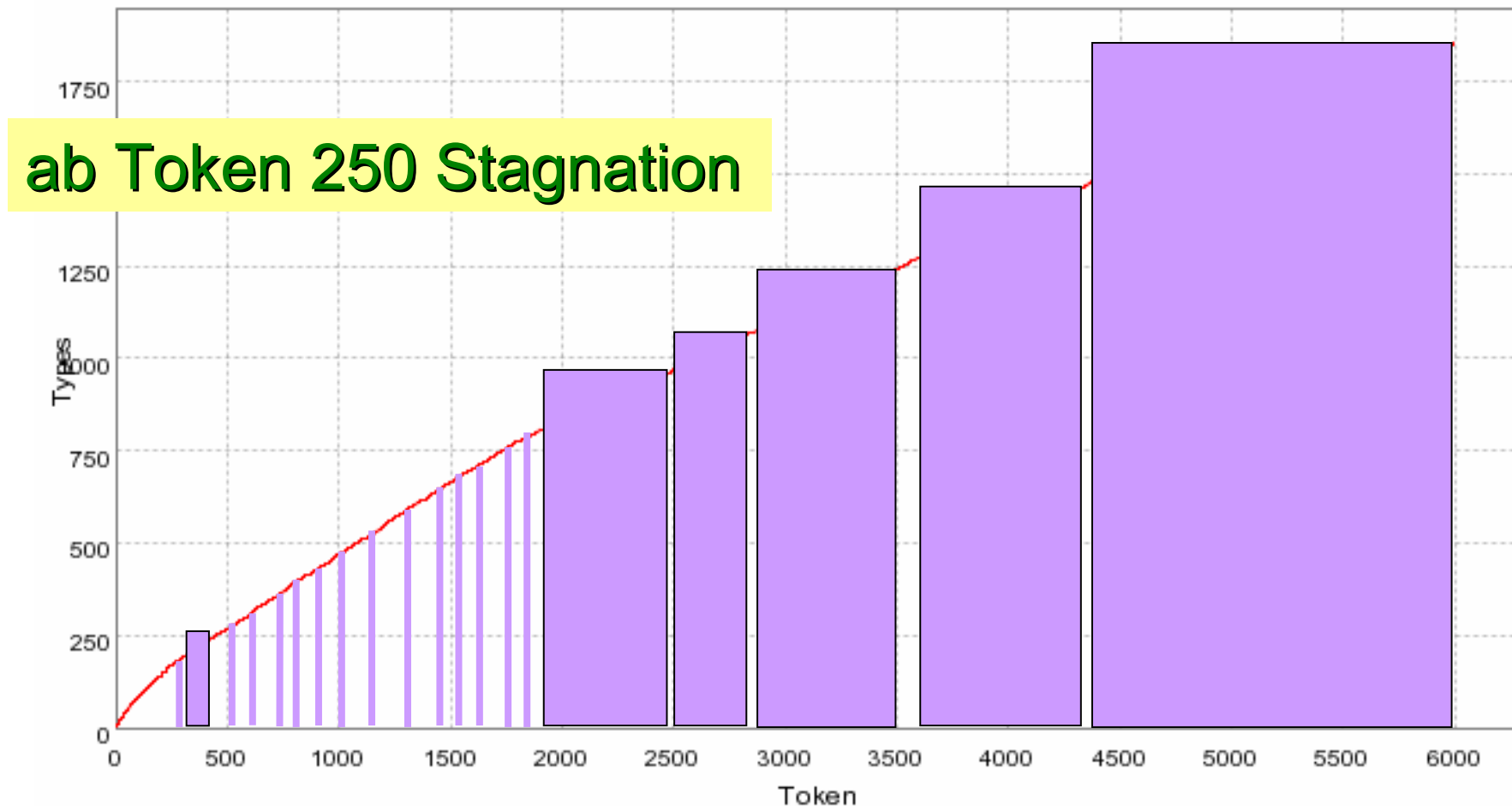
das letzte Drittel des Textes
bildet einen großen, letzten
Bereich

Datei		
von	bis	# Tokens
393	441	49
467	500	34
1892	1976	85
2001	2138	138
2195	2258	64
2260	2305	46
2322	2467	146
2513	2554	42
2571	2647	77
2649	2705	57
2741	2778	38
2835	3003	169
3027	3067	41
3071	3144	74
3187	3272	86
3274	3323	50
3331	3512	182
3574	3659	86
3661	3809	149
3811	3939	129
3954	4062	109
4094	4317	224
4375	5991	1.617

Bereiche relativer Stagnation

„Der Findling“ (5.992 Token)

Wortschatzzunahme



Vergleich: strikte Stagnation

Der Findling

WortSchatzStagnation			
Datei			
	von	bis	# Tokens
40 %	2329	2345	17
	2384	2396	13
50 %	2883	2896	14
	3032	3046	15
	3331	3344	14
	3574	3586	13
75 %	4537	4549	13
	4644	4661	18
	4940	4952	13
	4977	4990	14
90 %	5325	5337	13
	5370	5383	14
	5391	5413	23
	5594	5609	16
	5699	5716	18
	5853	5865	13

98 %

- kurze Bereiche
- ab Mitte d. Textes
- bis zum Ende

Der Mann v. 50 J.

WortSchatzStagnation			
Datei			
	von	bis	# Tokens
25 %	4241	4272	32
	5847	5881	35
	7984	8009	26
50 %	8960	8984	25
	12858	12881	24
80 %	14580	14604	25
	15219	15243	25
	15609	15633	25
90 %	16312	16339	28
	16478	16502	25
	16521	16551	31
	17629	17652	24
	17850	17878	29
	18080	18105	26
	18107	18140	34

99 %

- längere Bereiche
- ab ¼ d. Textes
- bis zum Ende

Der Taugenichts

WortSchatzStagnation			
Datei			
	von	bis	# Tokens
22 %	6867	6912	46
	17571	17615	45
	19166	19212	47
	21156	21206	51
75 %	23082	23140	59
	23708	23760	53
	23775	23828	54
	24195	24244	50
85 %	26594	26640	47
	26787	26832	46
	27265	27317	53
	28568	28633	66
	29396	29452	57

95 %

- lange Bereiche
- ab. 22% d. Textes
- nicht bis zum Ende

Vergleich: relative Stagnation

Der Findling

Datei		
von	bis	# Tokens
393	441	49
467	500	34
1892	1976	5
2001	2138	
2195	2258	
2260	2300	
2322		
2513		42
2571		77
2648		57
27		38
		169
	3067	41
	3144	74
	3272	86
	3323	50
331	3512	182
3574	3659	86
3661	3809	149
3811	3939	129
3954	4062	109
4094	4317	224
4375	5991	1.617

Ø 69% Stagn.

Der Mann v. 50 J.

Datei		
von	bis	# Tokens
1084	1193	110
1197	1250	54
1536	1584	9
1629	1822	
1919	1980	
2021		
2261		1
2366		90
247		477
		69
		37
	3267	154
	3448	166
	3553	89
	3706	104
3779	3826	48
3864	3900	37
3903	3956	54
3994	4094	101
4148	18286	14.139

Ø 75% Stagn.

Der Taugenichts

Datei		
von	bis	# Tokens
764	799	36
872	911	40
1067	1100	34
1224	1272	
1351		
1446		28
1530		73
161		166
		26
	1909	30
	1990	46
	2041	50
	2080	25
	2186	105
2204	2321	118
2432	2513	82
2550	2595	46
2603	2635	33
2655	2683	29
2739	31145	28.407

Ø 84% Stagn.

Findling:

7% 32% 42% 48% 60% 73%

Mann v. 50J:

6% 8% 12% 23%

Taugenichts:

2% 4% 5% 8% 9%

Bedeutung von Stagnation

Was können Stagnationsbereiche bedeuten?

- ◆ Ruhepunkte
- ◆ kontrastierend oder retardierend
- ◆ Betonung / Einprägung von Gesagtem
- ◆ „roter Faden“ durch die Handlung
- ◆ ...

Zusammenfassung

„Der Findling“ (5.992 Token):

- ◆ kurzer, „energischer“ Text
- ◆ „starke“ Stagnation erst ab Mitte des Textes

„Der Mann von 50 Jahren“ (18.287 Token):

- ◆ längerer, weniger energischer Text
- ◆ „starke“ Stagnation bereits ab $\frac{1}{4}$ des Textes

„Der Taugenichts“ (31.146 Token):

- ◆ sehr langer Text, energielos
- ◆ sehr früh „starke“ Stagnation