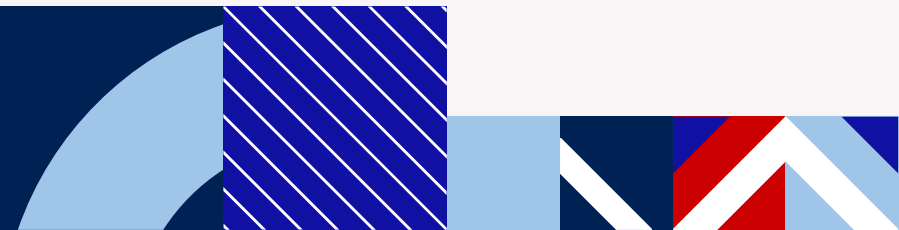




# Wortlänge und Vokal-/Konsonantenhäufigkeit

## Evidenz aus slowenischen, mazedonischen, tschechischen und russischen Paralleltexten

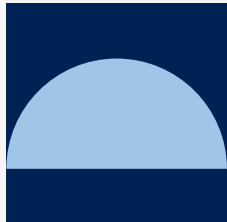
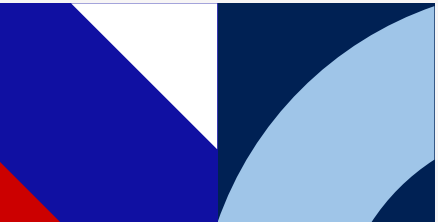
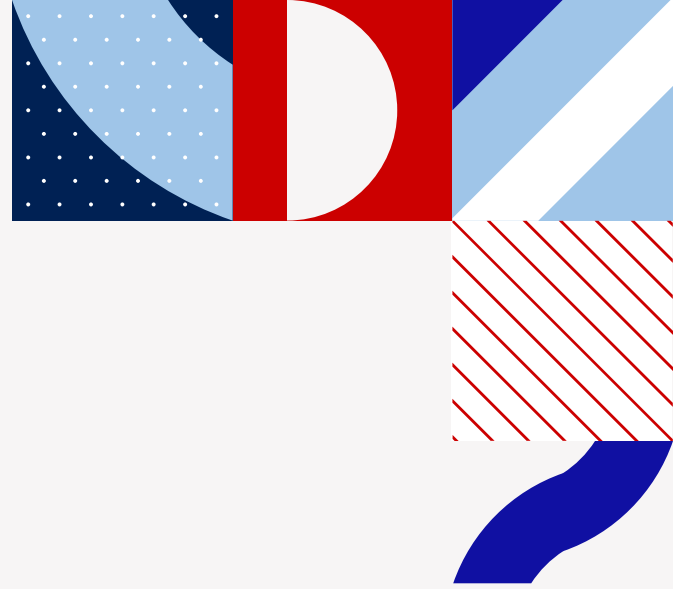
Emmerich Kelih (Graz)



Vorbereitet von: Ivan Boyanov

01

# Einleitung



# Synergetische Linguistik

Wortlänge •



• Vokalhäufigkeit

# Menzerath'sches Gesetz:

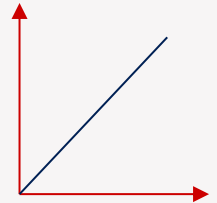
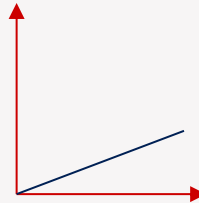
- Paul Menzerath (1883–1954)
- „Die relative Lautzahl nimmt mit steigender Silbenzahl ab, oder mit anderer Formel gesagt: je mehr Silben ein Wort hat, um so (relativ) kürzer (lautärmer) ist es.“
- „Je größer das Ganze, desto kleiner die Teile“
- nicht nur auf eine bestimmte Sprachebene (d.h. hier Silbenlänge und Lautdauer)

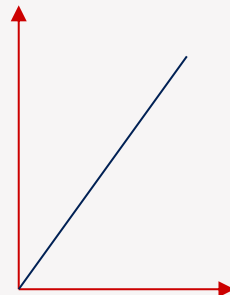
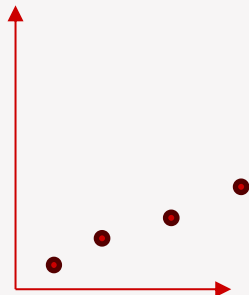
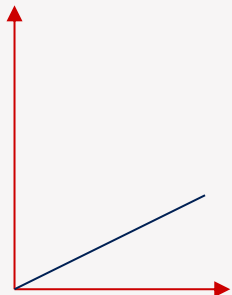
# Menzerath'sches Gesetz:

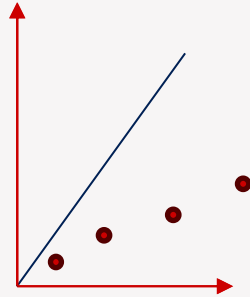
## Ausformulierung des Gesetzes von Altmann:

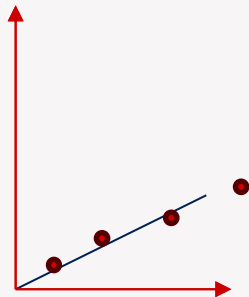
- $SLG$  - durchschnittliche Silbenlänge in der Anzahl von Graphemen
- $\left(\frac{SI}{WO}\right)$  - Wortlänge in Silben
- $a$  und  $b$  als Kurvenmodulatoren - sprachspezifisch

$$SLG = a \left( \frac{SI}{WO} \right)^{-b}$$







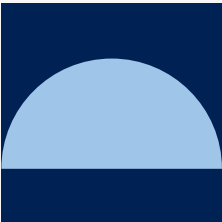






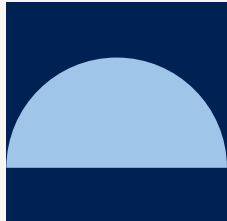
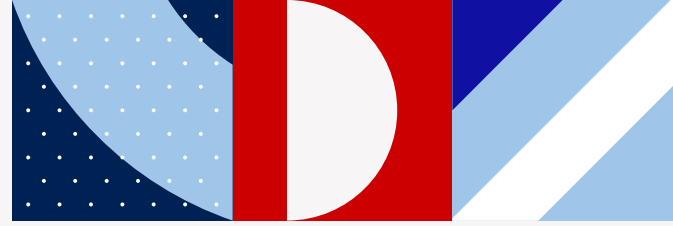
02

# Zentrale Hypothesen

- 1) Mit zunehmender Wortlänge werden Silben kürzer (Menzerath).
  - 2) Mit zunehmender Wortlänge steigt die relative Vokalhäufigkeit.
- 

03

# Methode





# Methode

- Korpus:
  - Russischer Roman „Wie der Stahl gehärtet wurde“
  - + Übersetzungen (Slowenisch, Mazedonisch, Tschechisch)

Für die Bestimmung der Vokalhäufigkeit wird das erste Kapitel von „Wie der Stahl gehärtet wurde“ verwendet

# Methode

**möglichst ähnliche Textsorten und Funktionalstilen**

- Korpus:
  - Russischer Roman „Wie der Stahl gehärtet wurde“
  - + Übersetzungen (Slowenisch, Mazedonisch, Tschechisch)

Für die Bestimmung der Vokalhäufigkeit wird das erste Kapitel von „Wie der Stahl gehärtet wurde“ verwendet

**Die Wortlänge hängt stark vom Stil ab!**




# Methode

- Analyseebene: Orthographisch (nicht phonologisch)
  - Graphemsequenz zwischen zwei Leerstellen
  - Pro- und Enklitika werden als eigenständige Wörter gezählt
- Messgrößen:
  - Wortlänge (in Silben)
  - Durchschnittliche Silbenlänge (in Graphemen)
  - Relative Vokalhäufigkeit



# Methode

- Slowenischen Halbvokals /ə/
  - Das silbenbildende /r/ im Mazedonischen
  - Die Diphthonge im Tschechischen
  - Die Interpretation des /ɤ/ und /ɘ/ im Russischen
- 

# Methode

- Slowenischen Halbvokals /ə/
- Das silbenbildende /r/ im Mazedonischen
- Die Diphthonge im Tschechischen
- Die Interpretation des /ь/ und /ъ/ im Russischen

	Vokale [V]
Slowenisch	i, e, a, o, u
Mazedonisch	i, e, a, o, u
Russisch	и, е, ё, ю, я, ы, э, о, у, а
Tschechisch	a, á, e, é, ě, i, í, o, ó, u, ú, ů

# Welche Werte hat der Autor bekommen?

Sprache	V (abs.)	C (abs)	V (rel.)	C (rel.)	Types	mittlere Types- länge in Silben
Slowenisch	4968	7380	0,40	0,60	1955	2,55
Mazedonisch	4630	6083	0,43	0,57	1637	2,83
Russisch	5340	7738	0,41	0,59	1904	2,80
Tschechisch	4536	6446	0,41	0,59	1769	2,56

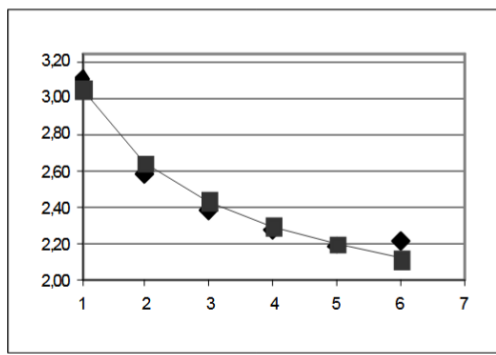


# Was hat Emmerich Kelih noch gemacht?

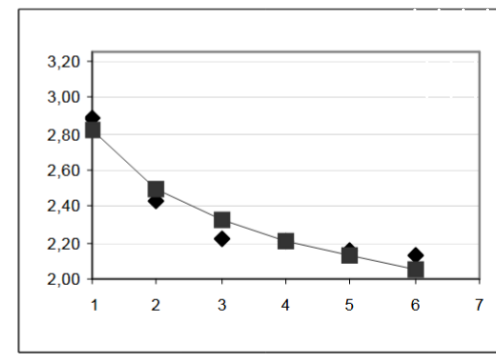
	Slowenisch	Mazedonisch	Russisch	Tschechisch
Si/WO	SLG	SLG	SLG	SLG
1	3,1143	2,8879	3,2881	3,0338
2	2,5866	2,4245	2,6496	2,5198
3	2,3834	2,225	2,4396	2,3321
4	2,2762	2,2085	2,2974	2,2103
5	2,1966	2,1607	2,182	2,0966
6	2,2222	2,125	2,1774	2,1364

Die durchschnittliche  
Silbenlänge  
(Anzahl von Graphemen  
pro Silbe)  
in dieser  
Wortlängenklasse

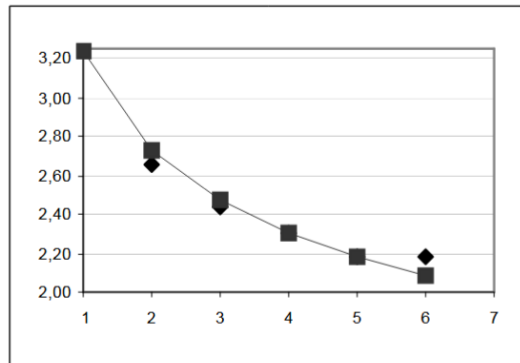
1-, 2-, 3-, ... x-silbigen Wortformen



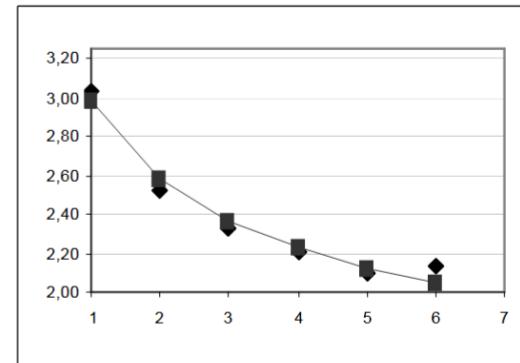
**Abb. 1a:**  
Wortlänge vs. mittlere Silbenlänge:  
Slowenisch



**Abb. 1b:**  
Wortlänge vs. mittlere Silbenlänge:  
Makedonisch



**Abb. 1c:**  
Wortlänge vs. mittlere Silbenlänge:  
Russisch

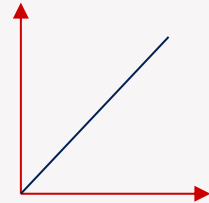
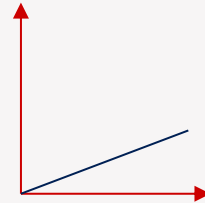


**Abb. 1d:**  
Wortlänge vs. mittlere Silbenlänge:  
Tschechisch

# Menzerath'sches Gesetz:

- Der Determinationskoeffizient  $R^2$  → Indiz für den Grad der Gültigkeit
- Übereinstimmung: empirische / theoretische Werte
- $R^2 > 0.85$  → sehr gute Übereinstimmung

$$SLG = a \left( \frac{SI}{WO} \right)^{-b}$$



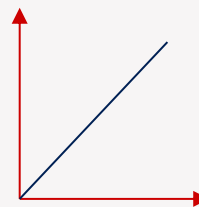
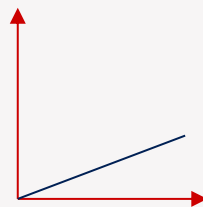
**Tabelle 4:** Anpassungswerte und Parameter

	<b>Slowenisch</b>		<b>Mazedonisch</b>		<b>Russisch</b>		<b>Tschechisch</b>	
Si/WO	SLG	theo.	SLG	theo.	SLG	theo.	SLG	theo.
1	3,11	3,06	2,89	2,82	3,29	3,23	3,03	2,99
2	2,59	2,65	2,42	2,50	2,65	2,73	2,52	2,58
3	2,38	2,44	2,23	2,32	2,44	2,47	2,33	2,37
4	2,28	2,30	2,21	2,21	2,30	2,31	2,21	2,23
5	2,20	2,20	2,16	2,12	2,18	2,18	2,10	2,12
6	2,22	2,12	2,13	2,06	2,18	2,09	2,14	2,04
<b><math>R^2</math></b>	<b>0,96</b>		<b>0,92</b>		<b>0,98</b>		<b>0,97</b>	
<b><math>a</math></b>	3,0556		2,8203		3,2347		2,9871	
<b><math>b</math></b>	-0,2038		-0,1762		-0,2440		-0,2119	

# Menzerath'sches Gesetz:

$a$  und  $b$  als Kurvenmodulatoren – sprachspezifisch

$$SLG = a \left( \frac{SI}{WO} \right)^{-b}$$



	Slowenisch		Mazedonisch		Russisch		Tschechisch	
$R^2$	0,96		0,92		0,98		0,97	
$a$	3,0556		2,8203		3,2347		2,9871	
$b$	-0,2038		-0,1762		-0,2440		-0,2119	

# Gibt es signifikante Unterschiede bei $b$ als Kurvenmodulator in den untersuchten Sprachen?

**Tabelle 5:** Ergebnisse für  $t$ -verteilte Prüfgröße

Vergleichspaare		t-Wert	FG	p
Russisch	Slowenisch	0,1142	8	0,9119
	Mazedonisch	0,2006	8	0,8460
	Tschechisch	0,0896	8	0,9308
Mazedonisch	Slowenisch	0,0808	8	0,9376
	Russisch	0,1886	8	0,8551
	Tschechisch	0,1049	8	0,9191

# Gibt es signifikante Unterschiede bei *b* als Kurvenmodulator in den untersuchten Sprachen?

$p > 0.05$   
bedeutet  
in den Geisteswissenschaften  
(und in vielen anderen Fächern)  
nicht statistisch signifikant

p
0,9119
0,8460
0,9308
0,9376
0,8551
0,9191

# ein Potenzmodell

$$\textit{rel. V-Freq.} = c * SW^d$$

## Was haben wir in diesem Model:

- *rel. V – Freq.* – die relative Vokalfrequenz
- *SW* – die Anzahl von Silben
- *c* und *d* als Kurvenmodulatoren
- Der Exponent *d* als Kurvenmodulator – muss positiv sein



**Tabelle 6:**

Wortlänge und relative Vokalhäufigkeit in den slawischen Sprachen

Si/WO	Slowenisch		Mazedonisch		Russisch		Tschechisch	
	V rel.	theo.	V rel.	theo.	V rel.	theo.	V rel.	theo.
1	0,3150	0,3253	0,3301	0,3464	0,3041	0,3133	0,3254	0,3336
2	0,3783	0,3721	0,4038	0,3925	0,3774	0,3683	0,3873	0,3832
3	0,4117	0,4026	0,4427	0,4222	0,4099	0,4048	0,4193	0,4155
4	0,4311	0,4257	0,4436	0,4447	0,4353	0,4329	0,4479	0,4401
5	0,4443	0,4445	0,4559	0,4629	0,4583	0,4560	0,4688	0,4602
6	0,4500	0,4605	0,4706	0,4784	0,4593	0,4759	0,4610	0,4773
$R^2$	0,96		0,91		0,97		0,96	
<i>c</i>	0,3253		0,3464		0,3133		0,3336	
<i>d</i>	0,1940		0,1802		0,2332		0,1998	



Von den Berechnungen werden die 7-silbigen Wörter ausgeschlossen.

Diese haben eine sehr geringe Frequenz  
(max. rel. Anteil von ca. 0.00079)  
und treten als Ausreißer auf.

d.h. sie passen nicht in das Gesamtbild der  
Menzerath'schen Kurven.



**Tabelle 6:**

Wortlänge und relative Vokalhäufigkeit in den slawischen Sprachen

	Slowenisch		Mazedonisch		Russisch		Tschechisch	
Si/WO	V rel.	theo.	V rel.	theo.	V rel.	theo.	V rel.	theo.
1	0,3150	0,3253	0,3301	0,3464	0,3041	0,3133	0,3254	0,3336
2	0,3783	0,3721	0,4038	0,3925	0,3774	0,3683	0,3873	0,3832
3	0,4117	0,4026	0,4427	0,4222	0,4099	0,4048	0,4193	0,4155
4	0,4311	0,4257	0,4436	0,4447	0,4353	0,4329	0,4479	0,4401
5	0,4443	0,4445	0,4559	0,4629	0,4583	0,4560	0,4688	0,4602
6	0,4500	0,4605	0,4706	0,4784	0,4593	0,4759	0,4610	0,4773
$R^2$	0,96		0,91		0,97		0,96	
$c$	0,3253		0,3464		0,3133		0,3336	
$d$	0,1940		0,1802		0,2332		0,1998	

 $\approx 30\%$

**Tabelle 6:**

Wortlänge und relative Vokalhäufigkeit in den slawischen Sprachen

Si/WO	Slowenisch		Mazedonisch		Russisch		Tschechisch	
	V rel.	theo.	V rel.	theo.	V rel.	theo.	V rel.	theo.
1	0,3150	0,3253	0,3301	0,3464	0,3041	0,3133	0,3254	0,3336
2	0,3783	0,3721	0,4038	0,3925	0,3774	0,3683	0,3873	0,3832
3	0,4117	0,4026	0,4427	0,4222	0,4099	0,4048	0,4193	0,4155
4	0,4311	0,4257	0,4436	0,4447	0,4353	0,4329	0,4479	0,4401
5	0,4443	0,4445	0,4559	0,4629	0,4583	0,4560	0,4688	0,4602
6	0,4500	0,4605	0,4706	0,4784	0,4593	0,4759	0,4610	0,4773
$R^2$	0,96		0,91		0,97		0,96	
$c$	0,3253		0,3464		0,3133		0,3336	
$d$	0,1940		0,1802		0,2332		0,1998	

# Gibt es signifikante Unterschiede bei $d$ als Kurvenmodulator in den untersuchten Sprachen?

$$\text{rel. } V\text{-Freq.} = c * SW^d$$



$$\log(\text{rel}V) = \log(c) + d * \log(SW)$$

$p > 0.05$   
bedeutet

in den Geisteswissenschaften  
(und in vielen anderen Fächern)  
nicht statistisch signifikant

# Zusätzliche Beobachtungen

Das Menzerath'sche Gesetz reguliert sowohl **die relative Vokalhäufigkeit** als auch **die Wortlänge** in allen untersuchten slawischen Sprachen.

**Wortlänge ↔ Silbenlänge**

$$SLG = a \left( \frac{SI}{WO} \right)^{-b}$$

**Wortlänge ↔ Vokalhäufigkeit**

$$rel. V-Freq. = c * SW^d$$

# Zusätzliche Beobachtungen

Das Menzerath'sche Gesetz reguliert sowohl **die relative Vokalhäufigkeit** als auch **die Wortlänge** in allen untersuchten slawischen Sprachen.

**Wortlänge ↔ Silbenlänge**

$$SLG = a \left( \frac{SI}{WO} \right)^{-b}$$

**Wortlänge ↔ Vokalhäufigkeit**

$$rel. V-Freq. = c * SW^d$$

$$c = -0,0809 * a + 0,5744$$

# Zusätzliche Beobachtungen

Das Menzerath'sche Gesetz reguliert sowohl **die relative Vokalhäufigkeit** als auch **die Wortlänge** in allen untersuchten slawischen Sprachen.

**Wortlänge ↔ Silbenlänge**

$$SLG = a \left( \frac{SI}{WO} \right)^{-b}$$

**Wortlänge ↔ Vokalhäufigkeit**

$$rel. V-Freq. = c * SW^d$$

$$c = -0,0809 * a + 0,5744$$

$$R^2 = 0.99$$

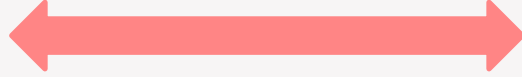




**Silbenlänge**

**Vokalhäufigkeit**

**Wortlänge**





# Zusammenfassung

- Der Zusammenhang zwischen Wortlänge und Silbenlänge (Das Menzerath'sche Gesetz) gilt für Slowenisch, Mazedonisch, Russisch und Tschechisch.
- Die relative Vokal-/Konsonantenhäufigkeit steht in einer klaren, mathematisch modellierbaren Beziehung zur Wortlänge.



**Haben Sie Fragen?**

**Danke für die Aufmerksamkeit!**