

**Emmerich Kelih
(Graz)**



**Vokal- und Konsonanten-Anteil als
Charakteristikum slawischer Literatursprachen (Slowenisch,
Mazedonisch, Tschechisch und Russisch) ?
(Mit einem Ausblick auf synergetische Wechselbeziehungen)**

- Institut für Slawistik, Universität Graz
- <http://www-gewi.uni-graz.at/quanta/> [Graz Project on Quantitative Text-Analysis]
- <http://www.uni-graz.at/emmerich.kelih/> [emmerich.kelih@uni-graz.at]

Welche sprachtypologische Bedeutung hat der Vokal- und Konsonantenanteil?

1. Anteil von Vokalen/Konsonanten am Phonem-Inventar
2. Studie von Isačenko (1939/1940): Probleme und Kritik
3. Erweiterungen: Synergetischer Regelkreis
4. Zusammenhänge: Inventarumfang – Silbenlänge – Wortlänge – Vokalhäufigkeit
5. Empirische Fallstudien (Parallelkorpus)
6. Abschließende Diskussion & Perspektiven

Anteil von Vokalen (V) und Konsonanten (C) am Phoneminventar

Schleicher (1852); Budilovič (1883); Dikarev (1891/1892); Čistjakov/Kramarenko (1929), Burlakova et al. (1962), Majewicz (1989), Gil (1990), Sawicka/Grzybowski (1999) u.v.m. ...

- Anteil von V und C am Phoneminventar als sprachtypologisches Merkmal
 - V/C als das „charakteristische“ einer Sprache
 - Möglichkeit einer typologischen Gliederung von slawischen Sprachen
-
- keine systematischen Studien zu slawischen Sprachen
 - z.T. Beschränkung auf Intuition („musikalischer“ Typus, sehr konsonantenreiche Sprache)
 - methodologische Schwächen: keine Prozentangaben, kein relativer Anteil ...
 - Studie von Isačenko (1939/1940) “Versuch einer Typologie der slawischen Sprachen”

Isačenko (1939/1940): Berechnung des Prozentsatzes von V und C am Inventar:

Sprache	Konsonanten	Vokale	silbische Konsonanten	Summe	% der Konsonanten
Serbokroatisch-Štokavisch	24	20	4	48	50
Slowenisch	21	18	3	42	50
Kašubisch	27	26	0	53	50,9
Slowakisch	27	14	4	45	60
Čechisch	26	11	2	39	66,7
Ukrainisch	31	12	0	43	72,1
Bulgarisch	34	9	0	43	79,1
Ober-Sorbisch	32	7	0	39	82,1
Russisch	37	8	0	45	82,2
Nieder-Sorbisch	33	7	0	40	82,5
Polnisch	35	5	0	40	87,5

RESULTAT nach Isačenko (1939/1940):

- I. ein radikal vokalischer Typus (Serbokroatisch, Slowenisch, Kaschubisch)
- II. ein radikal konsonantischer Typus (Ostslawische Sprachen, Bulgarisch)
- III. das „Schriftslowakische“, welches in der Mitte zwischen den beiden Extremtypen liegt

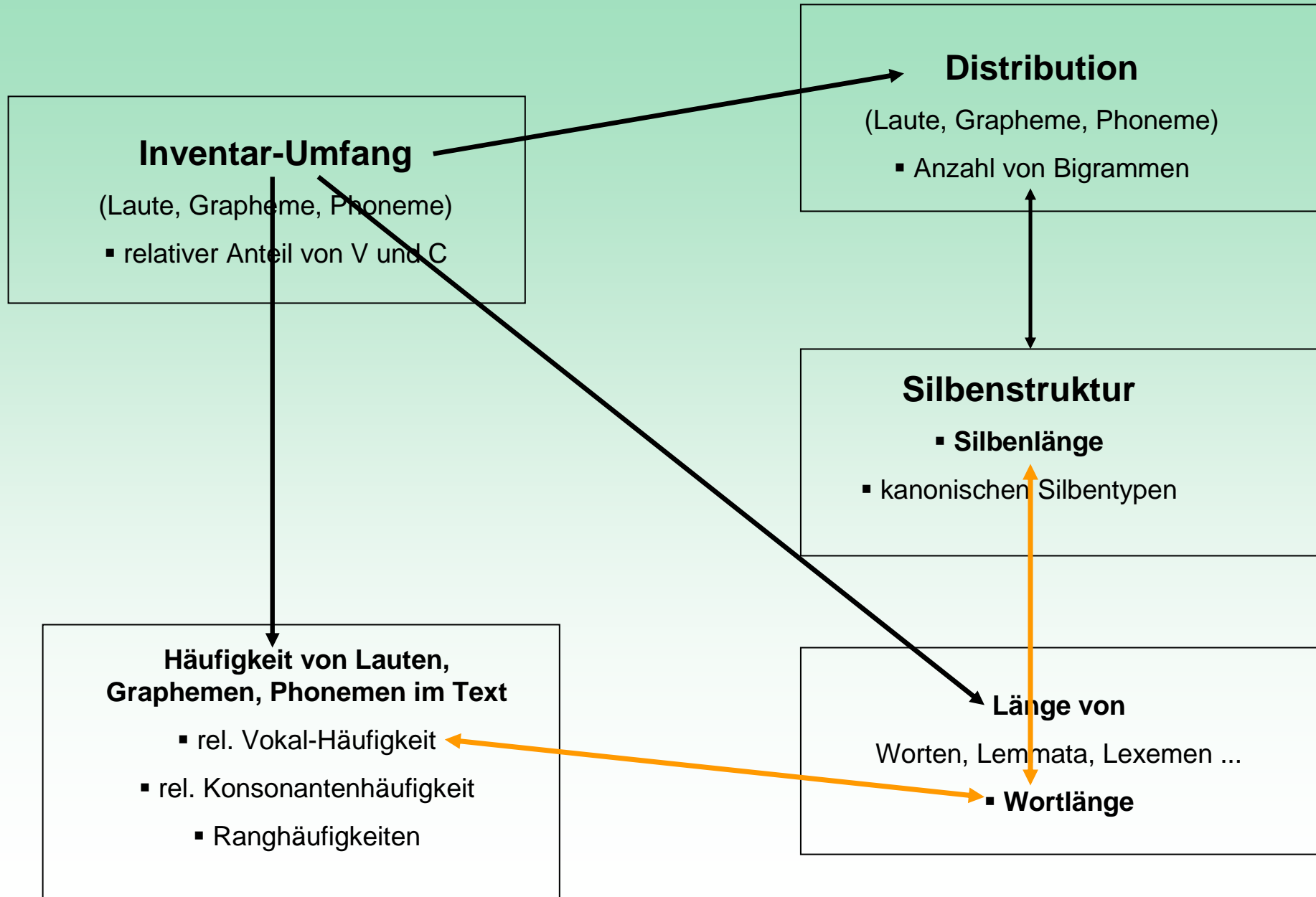
Weitere Rezeption

- methodologische Weiterentwicklungen durch V. Skalička, J. Krámský, E. Stankiewicz
- allgemeine Akzeptanz von „vokalischen“ vs. „konsonantischen“ Sprachtypus in Andersen (1978), ähnlich Stadnik (1998)
- Erklärung: Kompensationsmechanismen
- **ausführliche Kritik** durch Kempgen (1991)
 - es werden nur binäre Merkmale benutzt
 - Mittelstellung des Slowakischen ?
 - fehlende Zuordnung des Ukrainischen !
 - fehlende Trennung von segmentalen und supra-segmentalen Eigenschaften !
 - linguistische Kriterien für Bestimmung des Phoneminventars ist offen !
 - keine Begründung über Festlegung der Grenzen
 - nur der Anteil von V und C wird beachtet,
nicht aber Häufigkeit von einzelnen Phonemen !

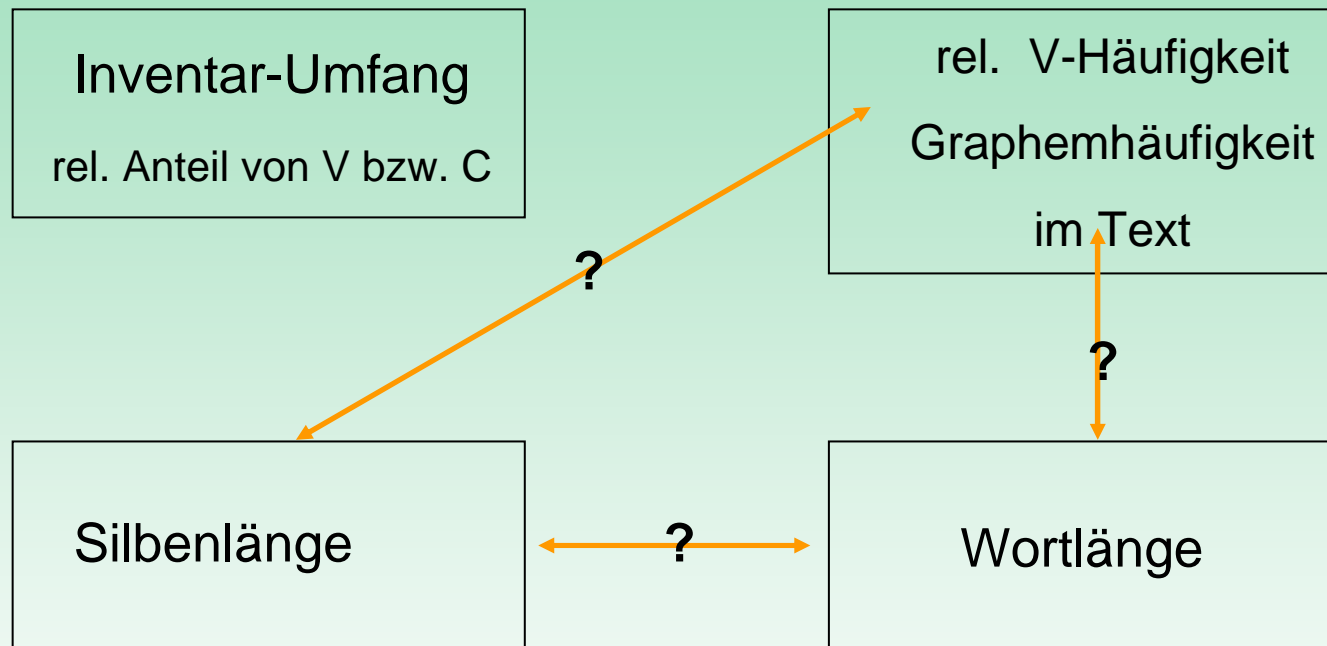
Potentielle Merkmale einer Sprachtypologie (Altmann/Lehfeldt 1973 Kempgen 1991 u.a.)

1. Zahl der Vokale im Inventar
2. Zahl der Konsonanten im Inventar
3. Zahl von silbenbildenden Konsonanten
4. Zahl der Korrelationspaare hart/weich
5. Anzahl von Phonemverbindungen (zulässige, realisierte ...)
6. Bindefähigkeit von Phonemen („Attraktivität“ „Aggressivität“)
7. positionelle Beschränkungen (initial, medial, final)
8. Silbenlänge und Silbenhäufigkeit
9. Anzahl von kanonischen Silbentypen
10. Morphemlänge – Morphem-Struktur
11.

Synergetische Sichtweise: Wechselbeziehungen



Fallstudie: 4 ausgewählte Eigenschaften



- keine Eigenschaft ist isoliert
- Aufdecken von Wechselbeziehungen (Korrelationen)
- Selektion von „informativen“ Eigenschaften

Zusammenhang von Wortlänge und Silbenlänge

Wirkung des Menzerath'schen Gesetzes (MG)

- „The longer a language construct the shorter its components (constituents).“ (Altmann 1980)
- „Mit zunehmender Wortlänge, nimmt die Silbenlänge ab“. (Altmann/Schwibbe 1989)

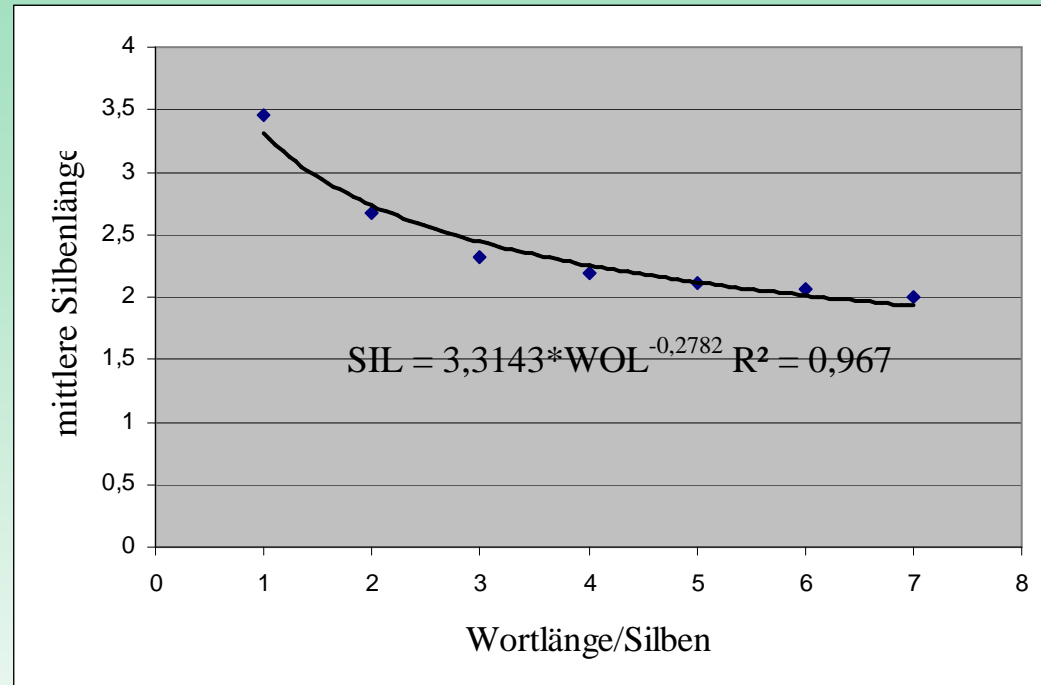
Linguistische Spezifizierungen:

- Bestimmung der „Wortlänge“ auf unterschiedlichen Ebenen (Types, Tokens, Lemma ...)
- Messung der Wortlänge: Anzahl von Silben, Phonemen, Lauten ...
- Silbenlänge: Anzahl von Phonemen/Segmenten/Graphemen/ Lauten
- Mathematische Ausformulierung: $y = a \cdot x^{-b}$

Wortlänge vs. mittlere Silbenlänge

Fallbeispiel: Serbokroatisch (nach Gajić 1950)

Wortlänge in Anzahl von Silben	mittlere Silbenlänge
1	3,46
2	2,67
3	2,32
4	2,2
5	2,11
6	2,06
7	2



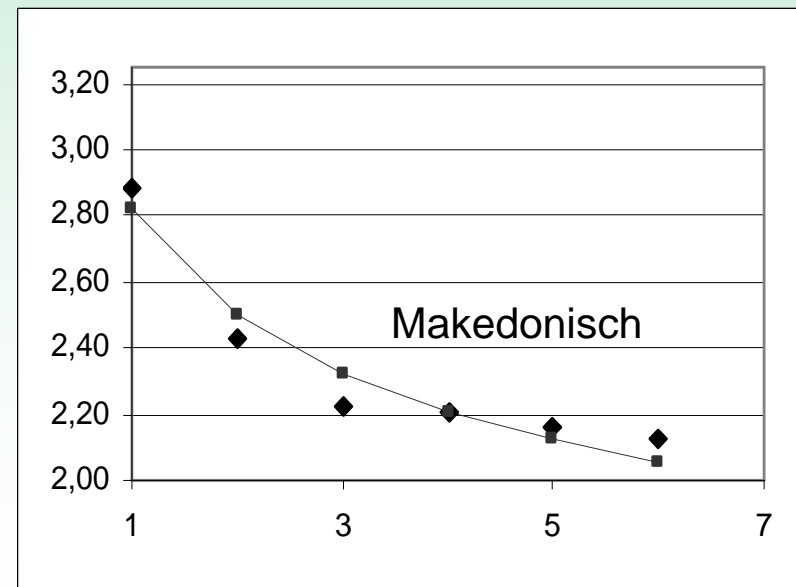
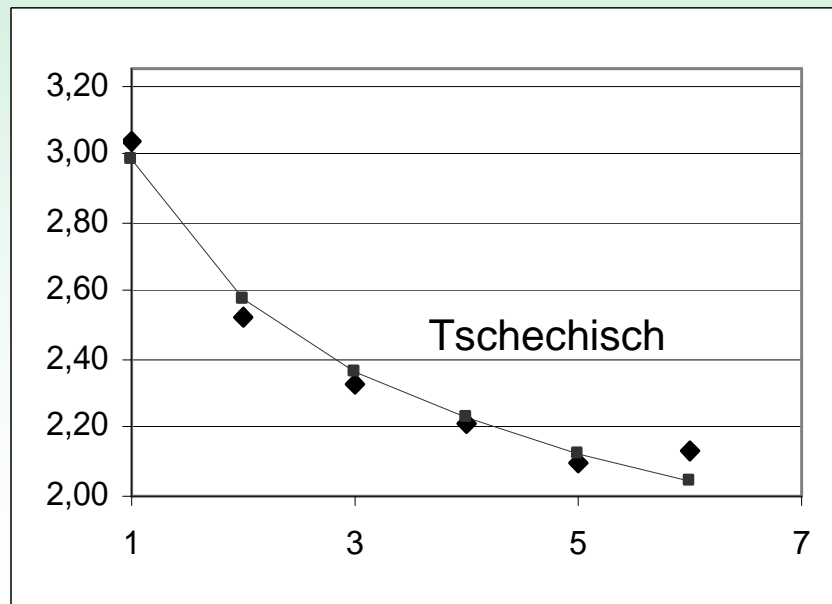
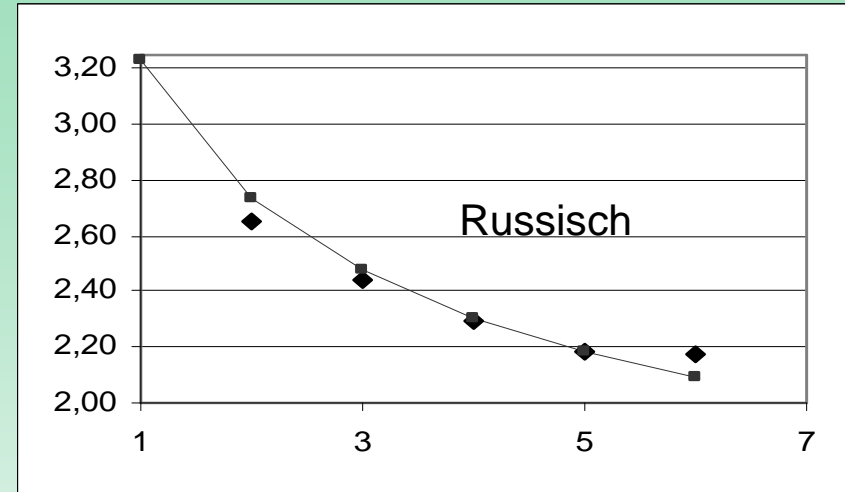
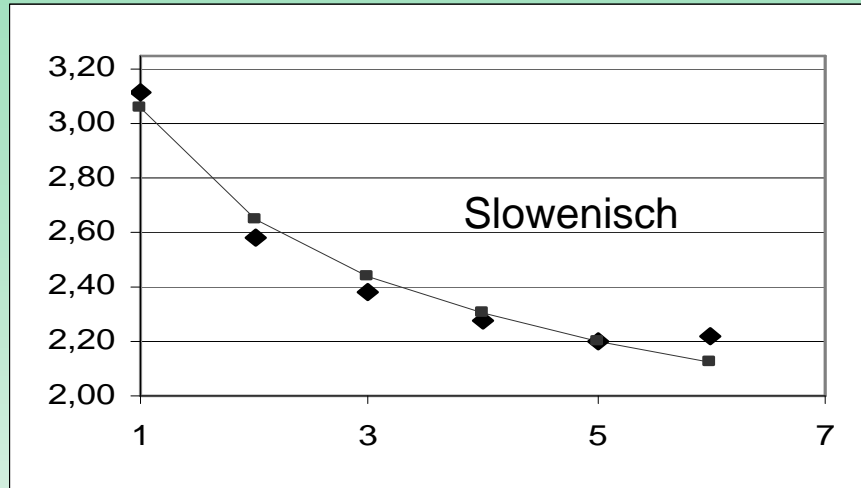
- MG allgemein akzeptiert in quantitativer Linguistik
- wenige systematische Studien für slawische Sprachen (Lehfeldt/Altmann 2006, Grzybek 1999)
- nur ansatzweise Diskussion über die Rolle von MG für sprachtypologische Fragestellungen !

Fallstudie I: Slawisches Parallel-Korpus

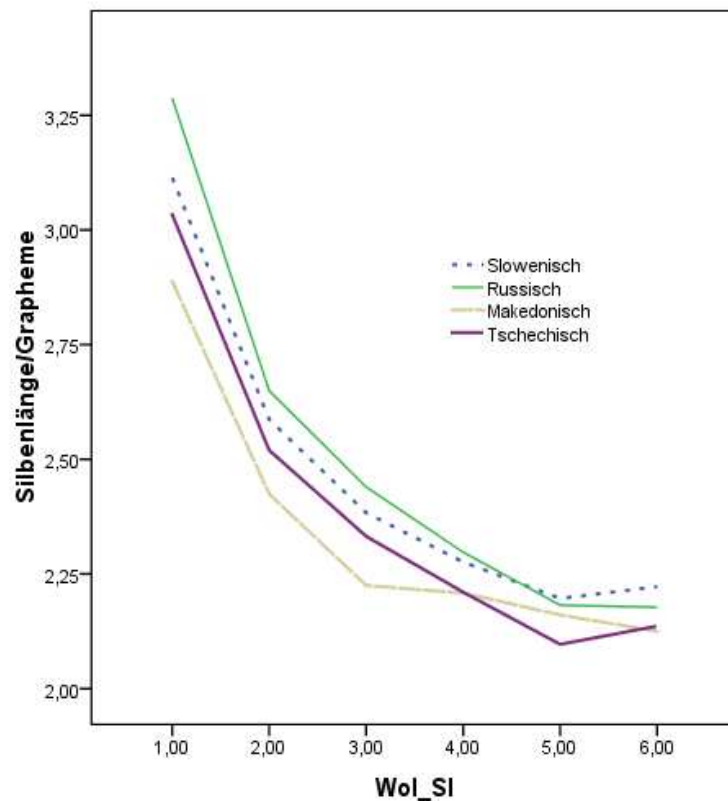
Untersuchte Sprachen: Russisch, Slowenisch, Tschechisch, Mazedonisch

- Texte: je 10 Kapitel von «Kak zakaljalas' stal'» (1932-1934) von N.A. Ostrovskij
- Analyse von Wortformen-Types
- Bestimmung der Wortlänge in der Anzahl von Silben
- Silbenlänge in der Anzahl von Graphemen
- orthographische Kodierung der Texte
- „vokalische“ bzw. „konsonantische“ Grapheme als Basis

Systemlinguistische Zusammenhänge: Wortlänge – Silbenlänge



Systemlinguistische Zusammenhänge: Wortlänge – Silbenlänge



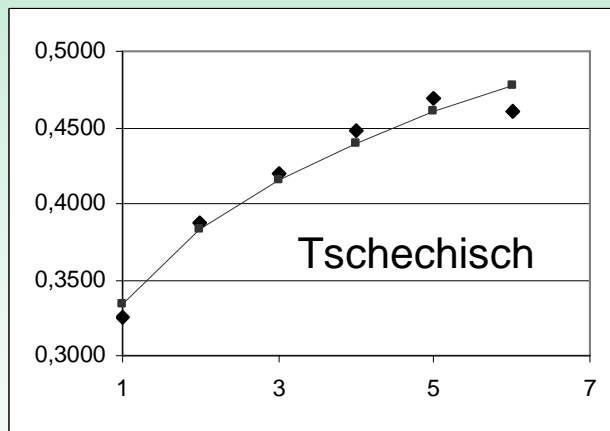
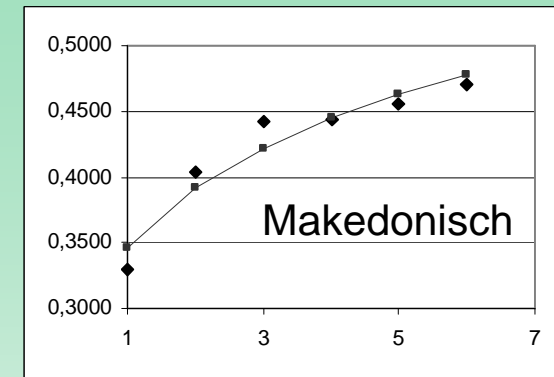
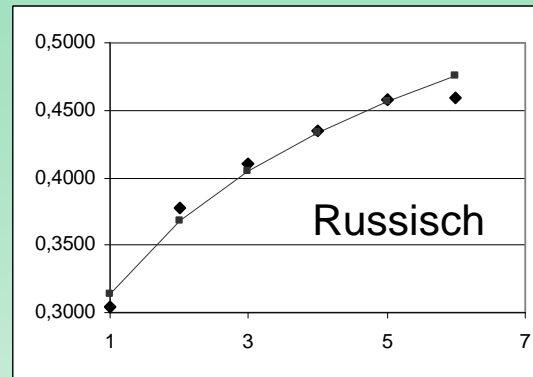
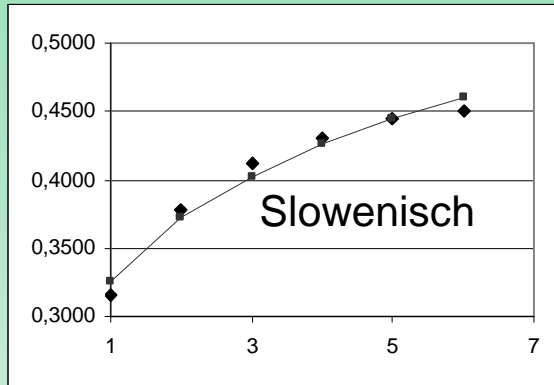
Sprache	a	b	R ²
Slowenisch	3,06	-0,20	0,95
Mazedonisch	2,82	-0,18	0,92
Russisch	3,23	-0,24	0,98
Tschechisch	2,99	-0,21	0,97

1. unterschiedliche Silbenlänge der Sprachen
2. „ähnlicher“ Verlauf der Kurve
3. Individualität der Sprachen bleibt erhalten (Parameter)

→ Interpretation eines Merkmals für sprachtypologische Belange ausreichend !

Auswirkungen des MG auf die relative Häufigkeit von Vokalen im Text?

→ Je länger das Wort, desto höher der relative Vokalanteil ?



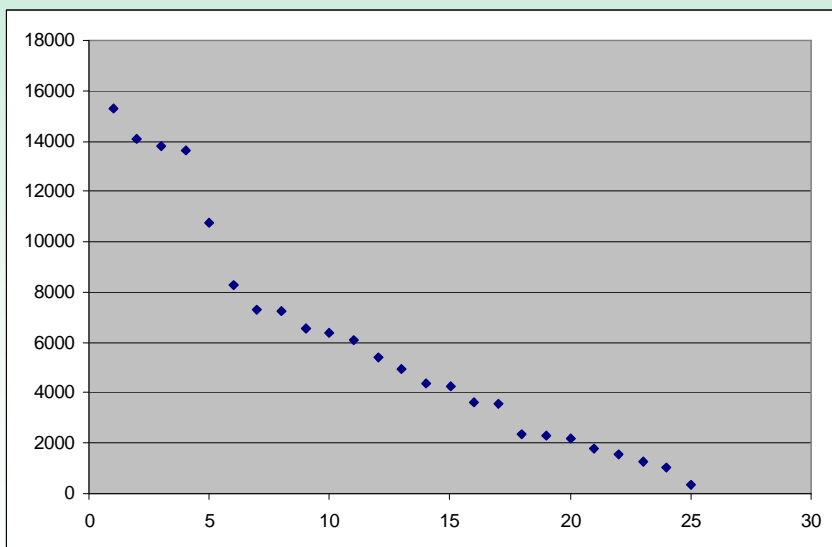
Sprache	c	d	R ²
Slowenisch	0,33	0,19	0,95
Mazedonisch	0,35	0,18	0,92
Russisch	0,31	0,23	0,98
Tschechisch	0,33	0,20	0,97

- unterschiedlicher relativer Anteil von Vokalen (siehe Parameter c)
- Regulation durch den gleichen Mechanismus wie bei Wortlänge - Silbenlänge

Auswirkungen auf Graphemfrequenz?

Fallstudie II: Slowenisch

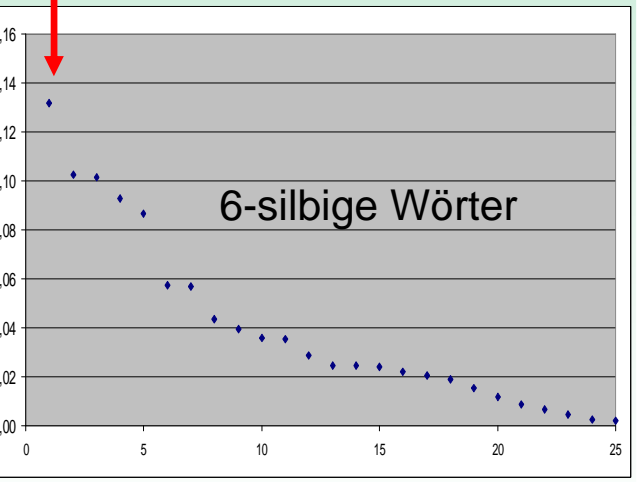
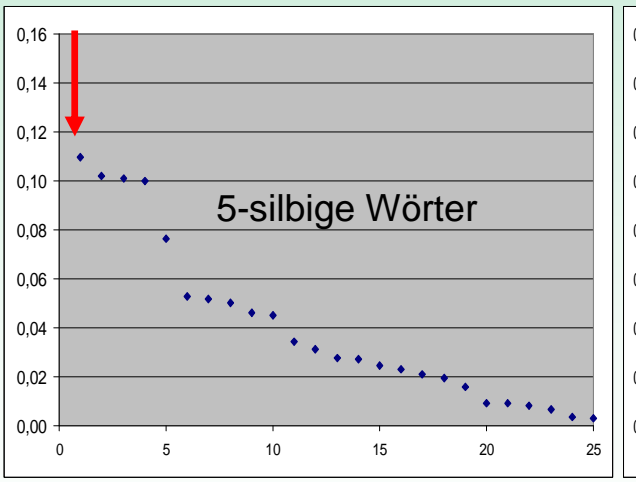
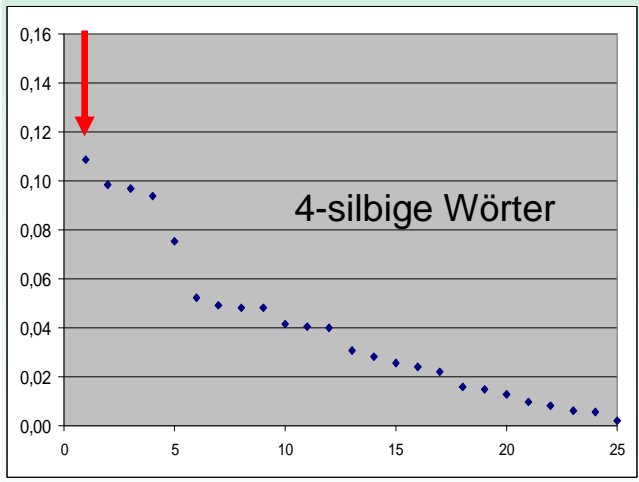
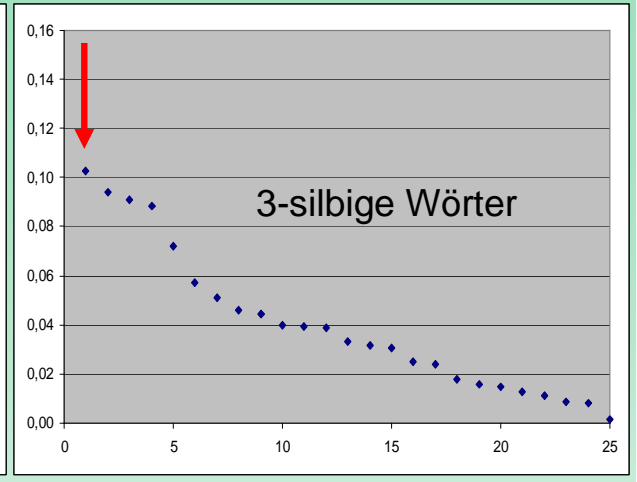
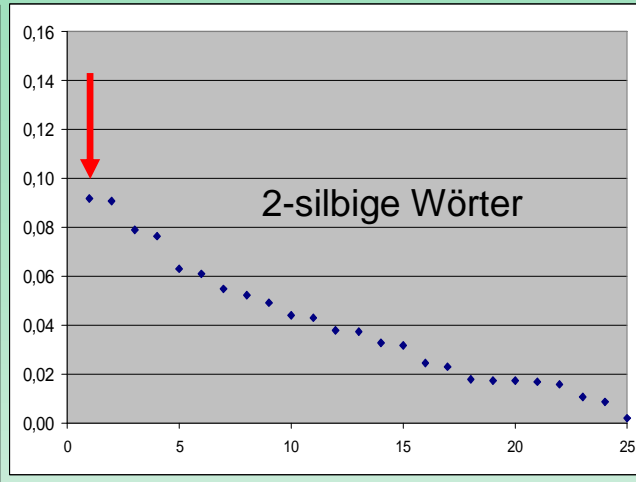
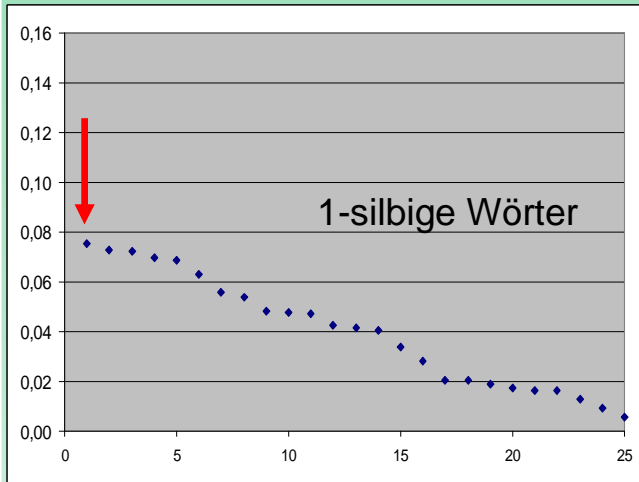
Textsorten	Anzahl	Wordformen-Types
Diplomarbeiten	5 Kap.	6144
Privatbriefe	30	5182
Predigten	32	7977
Drama	42 Akte	5616
Gesamtkorpus		24919



Graphemhäufigkeiten

Grapheme	Frequenz (abs.)	Freq. (rel.)	Rang
a	15290	0,10	1
b	2180	0,01	20
c	1279	0,01	23
č	2299	0,02	19
d	4371	0,03	14
e	14117	0,10	2
f	362	0,00	25
g	2374	0,02	18
h	1545	0,01	22
i	13776	0,09	3
j	6394	0,04	10
k	4284	0,03	15
l	7251	0,05	8
m	4941	0,03	13
n	10767	0,07	5
o	13644	0,09	4
p	5408	0,04	12
r	8289	0,06	6
s	6528	0,04	21
š	1772	0,01	21
t	7326	0,05	7
u	3622	0,02	16
v	6122	0,04	11
z	3570	0,02	17
ž	1051	0,01	24
	148562		

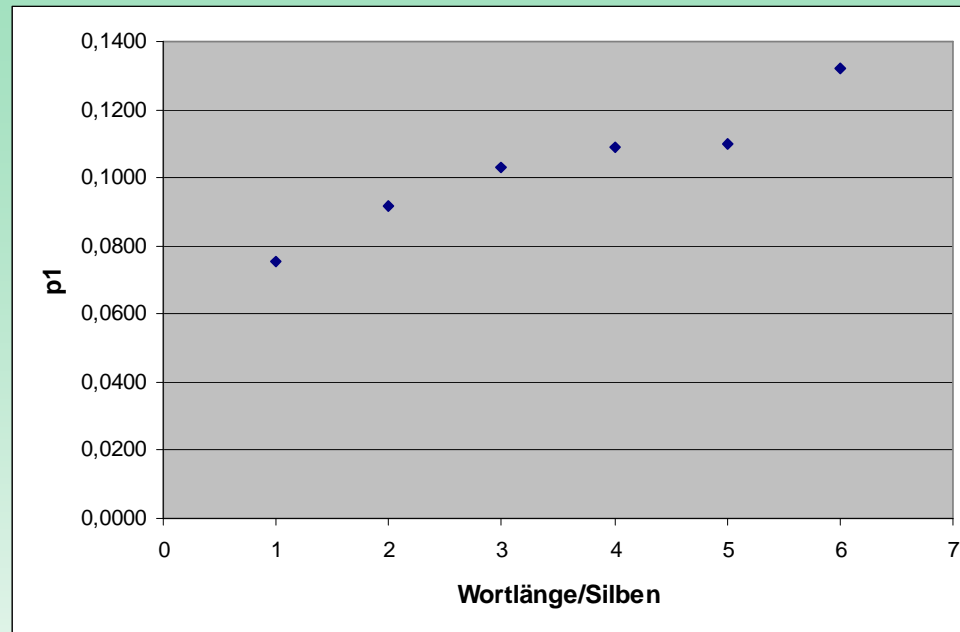
Rel. Graphemhäufigkeiten von 1, 2, 3, 4 ... x silbigen Wörtern



1. Anwachsen von p_1 (=erste Ranghäufigkeit) mit ansteigender Wortlänge
2. Gesamtbild der Verteilung ändert sich !

Zusammenhang: Wortlänge und p1?

Wortlänge Silben	p1
1	0,0755
2	0,0919
3	0,1030
4	0,1088
5	0,1098
6	0,1319



→ Je länger das Wort, desto höher der Anteil von p1.

Zusammenfassung

- Erweiterung der phonologischen Typologie slawischer Sprachen
- Textbasis (maximale Homogenität durch Paralleltex te)
- Beachtung von systematischen Interrelationen (MG)
- Graphemhäufigkeiten – relativer Vokalanteil – Silbenlänge – Wortlänge
= korrelierte Größen !
→ **Reduktion auf 1 Eigenschaft möglich und notwendig!**

Perspektiven:

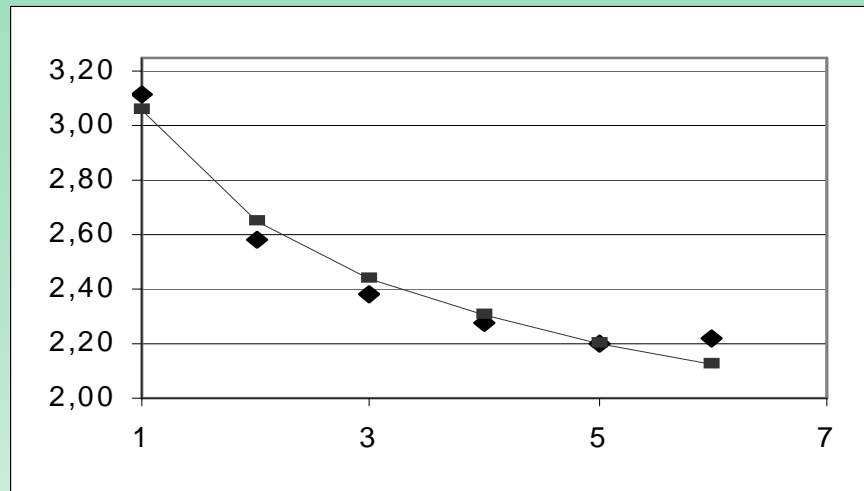
- Bedeutung des Phoneminventars muss geklärt werden
- phonologische Kodierung der Texte
- detaillierte Untersuchung der Phonemdistribution & Silbenstruktur
- Erweiterung auf weitere slawische Sprachen

Analyse von 1-silbigen Wörtern: Anzahl von kanonischen Silbentypen

SLO	MZ	CZ	RUS
CVC	CVC	CVC	CVC
CCVC	CCVC	CCVC	CCVC
CVCC	CV	CCV	CVCC
CCV	CVCC	CV	CCVCC
CV	CCV	CVCC	CCV
CCCVC	VC	CCCVC	CV
VC	CCVCC	VC	CVCCC
CCVCC	CCCVC	CCVCC	VC
V	V	CCCV	CCCVC
CCCVCC	CVCCC	V	V
VCC	CCCV	VCC	CCVCCC
CCCCVC	CCCVCC	CVCCC	VCC
CCCV	CCCCV	CCCCVC	VCCC
CVCCC	VCC	CCCVCC	CCCCVC
14	CCCCCV	14	CCCVCC
	CCVCCC		CCCV
	15		CVCCCC
			CCCVCCCC
			CCCVCCC
			CCVCCV
			20

Sprache	Wortlänge	Silbentypen
Slo	2,55	14
Mz	2,83	15
Rus	2,8	20
Cz	2,56	14

Silbenstruktur (kanonische Silbentypen) vs. Wortlänge ?



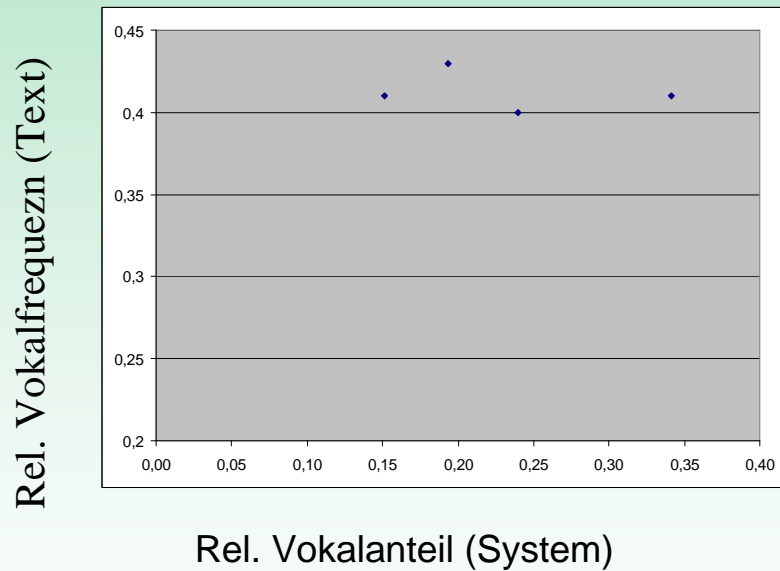
→ Je höher die Wortlänge, desto einfacher die Silbenstruktur

= Je höher die Wortlänge, desto weniger kanonische Silbentypen?

Vokalanteil (System) vs. rel. V-Anteil (Text)?

Krámský (1948): Je mehr V im Inventar, desto höher Vokalfrequenz im Text.

Sprachen	Inventar	Vokale	rel. Vokalanteil	rel. Vokalfrequenz
Slo	25	6	0,24	0,40
Mz	31	6	0,19	0,43
Ru	33	5	0,15	0,41
Cz	41	14	0,34	0,41



- kein systematischer Zusammenhang

Ziele einer quantitativen Sprachtypologie nach Altmann/Lehfeldt (1973: 15)

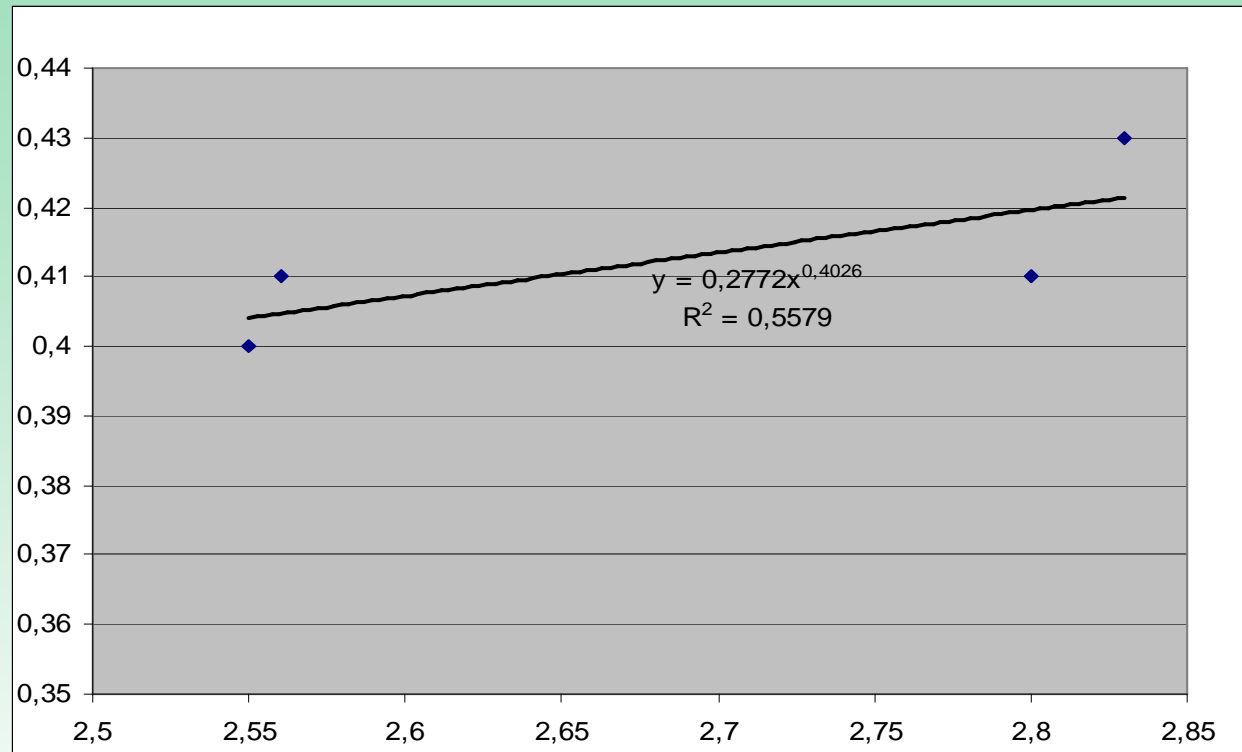
a.) Sprachklassifikation, d.h. Aufbau eines Ordnungssystems für die natürlichen Sprachen

b.) Aufdeckung von Konstruktionsmechanismen der Sprachen = Aufbau eines Beziehungssystems

- komplementär zu rein qualitativen Verfahren
- Unterschiede zur „crosslinguistisch“ orientierten Universalienforschung
- Verwendung statistischer Methoden
- theoretische Anlehnung an die synergetische Linguistik

Cross-linguistische Vergleich:

Je länger die mittlere Wortlänge, desto höher der Vokalanteil?



- nur schwache statische Bestätigung !
- weitere Faktoren beeinflussen diese Relation !

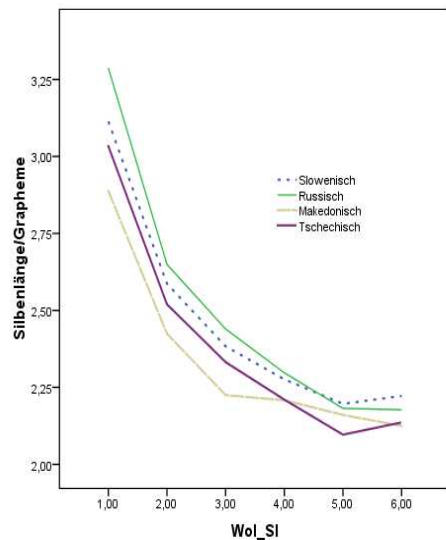
Isačenko (1939/1940: 67) unterscheidet [...] mit Rücksicht auf die phonologische Belastung, resp. auf ihren prosodischen 'Überbau' [...] innerhalb der slawischen Sprachen folgende Typen:

Bezogen auf das Vokalsystem:

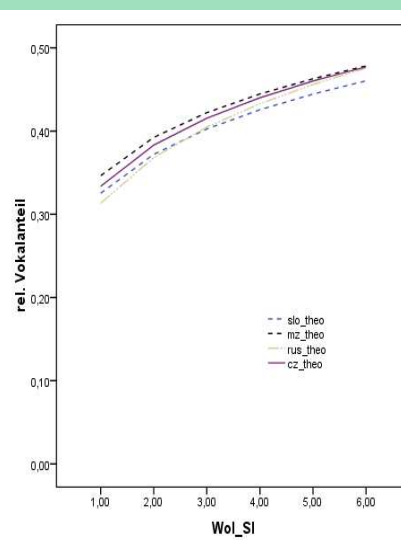
1. polytonische Sprachen (a) mit musikalischer Intonation in kurzen und langen Silben und (b) in langen Silben
2. monotonische Sprachen mit freier Quantität
3. monotonische Sprachen mit einem freien dynamischen Akzent
4. monotonische Sprachen ohne prosodische Belastung

Bezogen auf das Konsonantensystem:

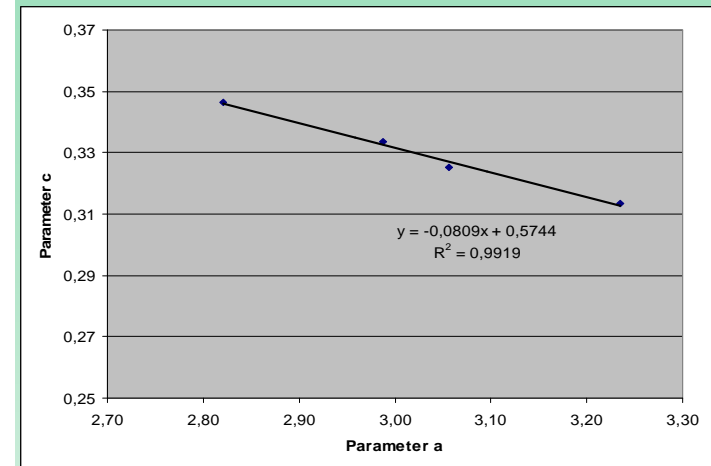
1. Sprachen mit Unterscheidung harter und weicher Konsonanten
2. hart/weich nur innerhalb der Dentalgruppe (Schriftslowakisch, Čechisch, Štokavisch)
3. Sprachen ohne weiche Konsonanten (z.B. Ljubljauer Aussprache des Slowenischen)



Wortlänge vs. Silbenlänge



Wortlänge vs. rel. V-Anteil



Sprache	a	b	R ²
Slowenisch	3,06	-0,20	0,95
Mazedonisch	2,82	-0,18	0,92
Russisch	3,23	-0,24	0,98
Tschechisch	2,99	-0,21	0,97

Sprache	c	d	R ²
Slowenisch	0,33	0,19	0,95
Mazedonisch	0,35	0,18	0,92
Russisch	0,31	0,23	0,98
Tschechisch	0,33	0,20	0,97

→ Zusammenhang von rel. Vokalhäufigkeit und Silbenlänge !

→ sprachübergreifende Gültigkeit !