universität wien

Dr. Emmerich Kelih

Institut für Slawistik

Empirische Methoden in der Sprachwissenschaft mit dem Fokus auf sprachlicher Komplexität

Peer-Mentoring-Projekte an der Fakultät für Kulturwissenschaften

Alpe-Adria Universität Klagenfurt/Celovec

21.06. 2012

Übersicht:

Kurzvorstellung: Quantitative Linguistik (QL)

Grundzüge der QL

Kernbereich (1): Zipf‘sche Gesetz
Kernbereich (2): Menzerath‘sche Gesetz

Einsatzbereich statistischer Methoden in der SPW

Thema: Sprachliche Komplexität ?
Fokus: Vergleich von Texten (Schultexte, wissenschaftliche Texte vs. studentische Arbeiten)

Fallbeispiele:

Lexikalische Struktur: TTR (Type-Token Ratio)

Wortlänge als Gradmesser für sprachliche Komplexität

Ausblick: Perspektiven und Grenzen der quantitativen Textanalyse

Kurzvorstellung: Quantitative Linguistik (QL)

Bußmann (1990: 623, 734)

Quantitative Linguistik → Statistische Linguistik [auch QL, Sprachstatistik]
experimentell orientierter Teilbereich der Mathematischen Linguistik. Die S.L. beschäftigt sich unter Verwendung statistischer Methoden mit der kontrollierten Untersuchungen sprachlicher Regularitäten unter quantitativen Aspekten. Ihren Verfahren dienen u.a. der Herstellung von → Häufigkeitswörterbüchern und zu stilistischen Textanalysen. Vgl. → Lexikostatistik.

Anmerkungen:

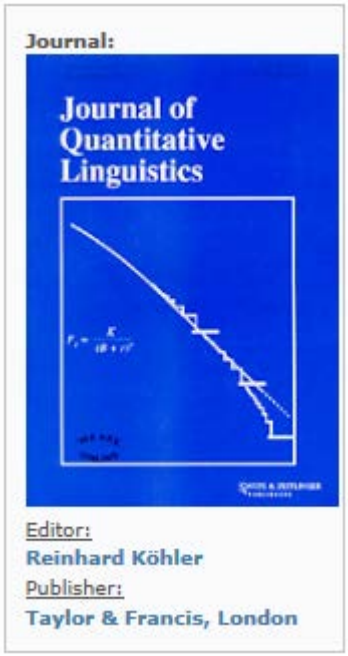
- sehr eingeschränkte, veraltete Auffassung von QL

- heute als eigenständige interdisziplinäre Richtung der SPW anzusehen
- Unterschiede vor allem epistemologischer Natur zwischen

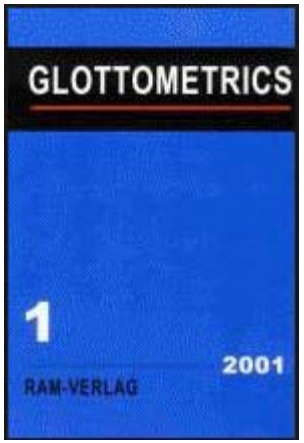
Sprachstatistik/Lexikostatistik ≠ Quantitative Linguistik

Anzeichen der Etablierung: QL als Disziplin

seit 1994



seit 2001



seit 2008

IQLA
International Quantitative Linguistics Association

Qualico 2014 (29. Mai – 1. Juni 2014
in Olomouc/CZ)

Buchreihen, Bibliographien, wichtige Monographien

Quantitative Linguistics bei Brockmeyer (Bochum) 1978-1992

Fortsetzung *Quantitative Linguistics* bei de Gruyter (Berlin u.a.)

Studies in *Quantitative Linguistics* (RAM-Verlag, Lüdenscheid)

Bibliographie:

Köhler, R. (1995): *Bibliography of Quantitative Linguistics.* Amsterdam/Philadelphia: Benjamins.

Köhler, R.; Altmann, G.; Piotrowski, R.G. (eds.) (2005): *Quantitative Linguistik. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook.* Berlin, New York: de Gruyter. [= Handbücher zur Sprach- und Kommunikationswissenschaft; 27]

in Vorbereitung:
Köhler, R.; Grzybek, P.; Naumann, S. (eds.) (2013ff): *Quantitative und Formale Linguistik.* Berlin u.a.: de Gruyter. [= Wörterbücher zur Sprach- und Kommunikationswissenschaft; 9] [im Aufbau, über 2000 Lemma]

Epistemologische und methodologische Grundsätze:

- Vielzahl von quantitativen Eigenschaften von Sprache/Texte (Forschungsobjekt)
- Untersuchung auf die Entwicklung und Funktionsweise von sprachlichen Systemen

- keine Einschränkung auf Diachronie/Synchronie, sondern genereller Fokus auf messbaren Eigenschaften von Sprache/Texten, insbesondere die

Häufigkeit & Länge

sprachlicher Phänomene stehen im Mittelpunkt des Interesses.

- in allen Bereichen und auf allen linguistischen Analyseebenen, Phonologie, Morphologie, Syntax, Textstruktur, Lexik, Semantik, Pragmatik, Dialektologie, Sprachwandelforschung, Textlinguistik, Psycho- und Soziolinguistik usw.

- Einsatz statistischer Verfahren (deskriptive Statistik, explorative Statistik, mathematische Modelle (Funktionen, Wahrscheinlichkeitsmodelle)

- besondere Bedeutung haben quantitative Methoden in der Korpuslinguistik, Computerlinguistik, forensische Linguistik, Stilometrie, Autorenschaftsbestimmung u.ä.

Einbettung in unterschiedliche Felder der Sprachwissenschaft

Zentrale Konzepte: **Häufigkeit** und **Längen**

- Sprachökonomie – Effizienz
- usage based approach: Frequentismus u.a. Joan Bybee, Paul Hopper mit einer Vielzahl von Publikationen
- natürlichkeitstheoretische Richtungen (Phonologie, Morphologie, Konzept der Markiertheit, Irregularitäten)
- Universalienforschung (statistische U.)
- Psycholinguistik (kognitive Faktoren, subjektive Wahrscheinlichkeit)
- Sprachwandel
- Spracherwerb – Fremdsprachendidaktik
- linguistische Komplexität
- synergetische Linguistik (Selbstorganisation sprachlicher Strukturen)

Erkenntnisziele der QL:

Finden und Untersuchungen von statistischen Gesetzmäßigkeiten und Regularitäten in Sprache/Texten

(1) Verteilungsgesetze (Zipf'sche Gesetz)

(2) funktionelle Gesetze (Menzerath'sche Gesetz)

(3) Entwicklungsgesetze (Piotrovskij Gesetz zum Sprachwandel)

- QL geht über deskriptive/explorative Anwendung statistischer Methoden hinaus
- deduktives Postulieren von adäquaten mathematischen Modellen
- linguistische Einbettung der Relevanz von Häufigkeiten & Längen

Fallbeispiel 1: Zipf‘sche Gesetz

- geht auf G.K. Zipf (1902-1950) zurück
- Autor von *Psychobiology of Language* (1935) und
  Human Behavior and the Principle of Least Effort (1949)

- Sammelbegriff für unterschiedliche Arten von Potenzgesetzen („power laws“)

- unterschiedliche mathematische Formulierung als stetige Funktionen oder diskrete Verteilung

- Anwendung nicht nur in der Linguistik, sondern auch in vielen anderen Wissenschaften (Biologie, Soziologie, Ökonomie usw.)

- grundlegendes Verhalten nichtlinearer dynamischer Systeme

- Verteilungsverlauf ist modellierbar

Was ist der linguistische Hintergrund der Zipf‘schen Gesetze?

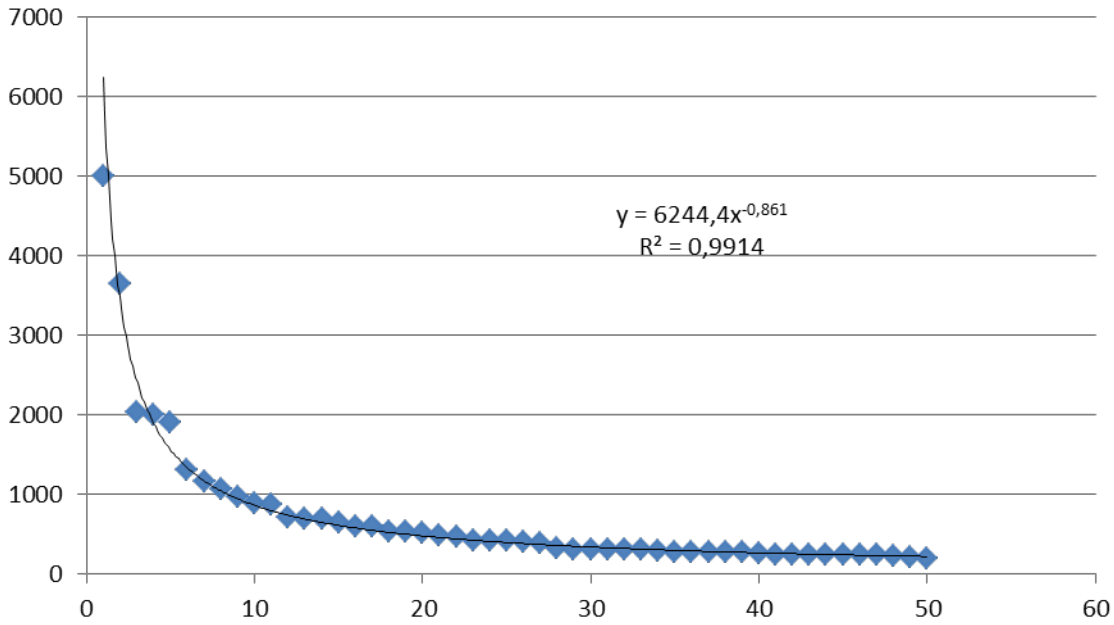
Häufigkeit von Wortformen

1. Text (running text)
2. Lemmatisierung (ja/nein)
3. Bestimmung der Anzahl von Wortformen (Tokens)
4. Sortierung der Wortformen nach ihrer Häufigkeit
5. Erstellen einer Ranghäufigkeitsverteilung

Russischer Text: Master i Margarita

Rang	Word	Freq.	Rang	Word	Freq.
1	И	5006	16	ИЗ	602
2	В	3640	17	ЖЕ	594
3	НЕ	2025	18	ПО	534
4	НА	2003	19	У	527
5	ЧТО	1905	20	ЗА	515
6	С	1307	21	ВСЕ	477
7	ОН	1153	22	БЫЛО	465
8	ТО	1071	23	МАРГАРИТА	419
9	А	974	24	ТАК	414
10	КАК	883	25	ВЫ	411
11	Я	867	26	ОНА	396
12	НО	713	27	ОТ	382
13	К	699	28	О	326
14	ЕГО	688	29	ТУТ	308
15	ЭТО	647	30	ДА	306

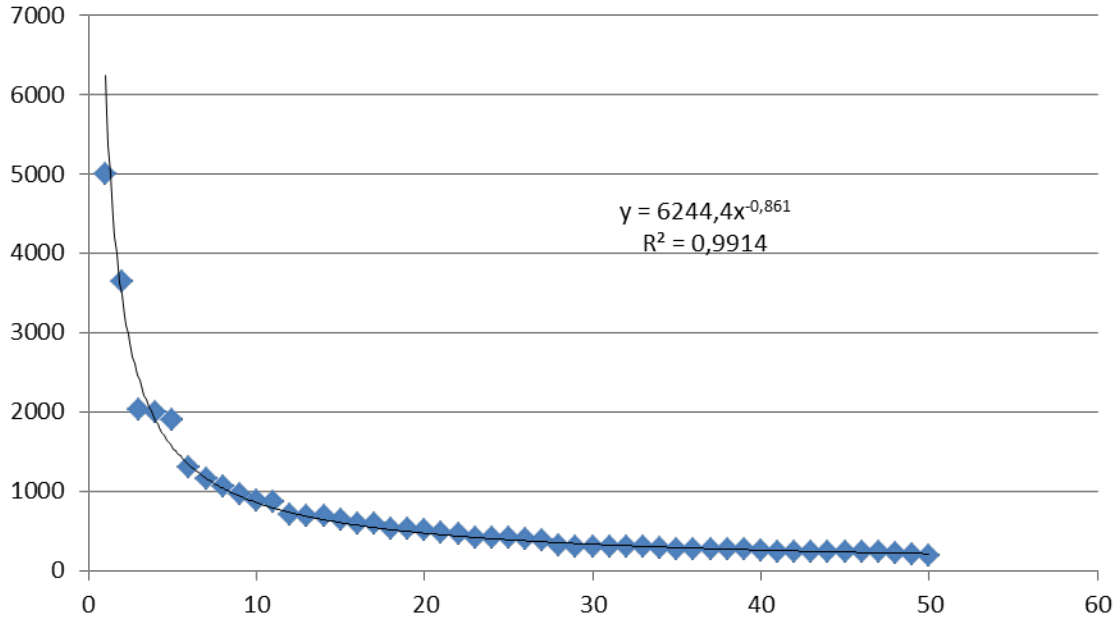
Graphische Darstellung: Ranghäufigkeitsverteilung



Welche linguistische Bedeutung?

- modellierbares Wechselspiel von häufigen und weniger häufigen Wortformen
- wenige Wortformen kommen mit sehr hoher Häufigkeit vor
- nichtlineare Verlaufsform
- Entropie & Wiederholungsrate von Wortformen

Graphische Darstellung: Ranghäufigkeitsverteilung



Verlaufsform der Ranghäufigkeitsverteilung ist abhängig von

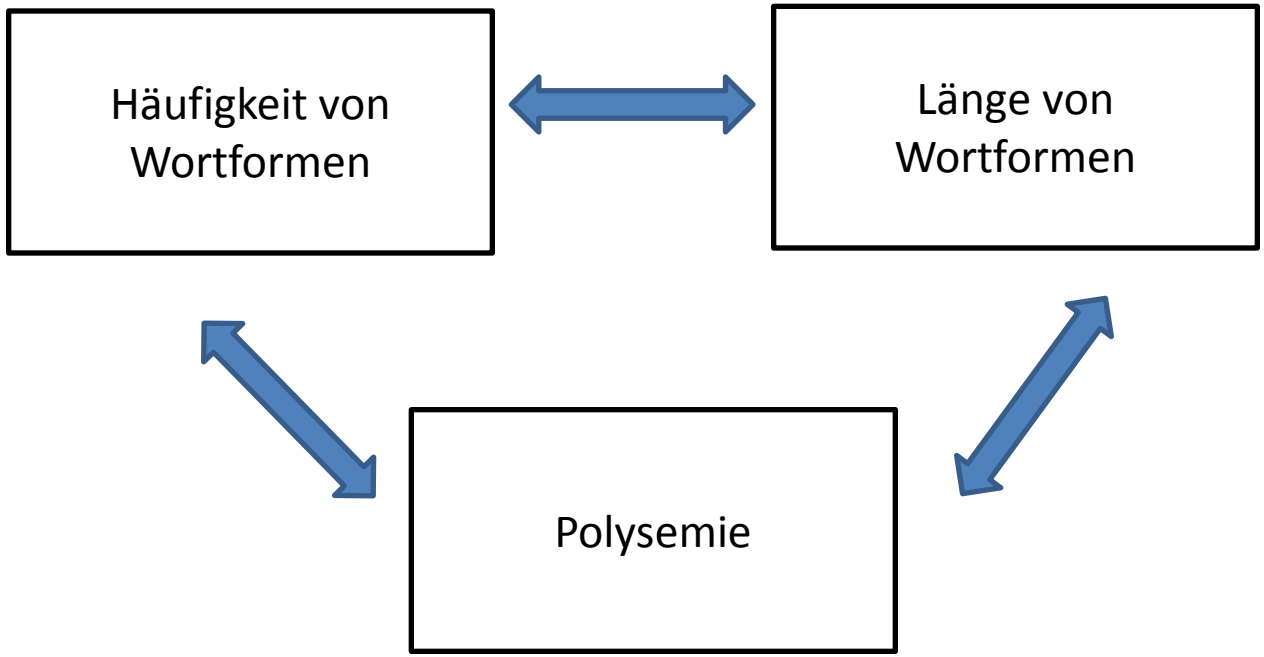
- Autor
- Textsorte
- Funktionalstil
- Sprache

+ gilt auch der Ebene von Lexemen, Lemmata bzw. generell für die Ranghäufigkeitsverteilung von linguistischen Einheiten

+ typischer Verlauf für komplexe, nichtlineare Systeme

Zipf‘sche Gesetz: Erweiterungen

1. Häufigkeit von Wortformen – Rang von Wortformen
2. Häufigkeit von Wortformen – Kürze/Länge von Wortformen
3. Häufigkeit von Wortformen – Querbezüge zur Ausprägung von Polysemie



--> Finden von gegenseitigen Abhängigkeiten und Wechselbeziehungen zwischen Merkmalen/Eigenschaften

Funktionelle Gesetze (Menzerath'sche Gesetz)

Paul Menzerath (1883-1954)

Institut für Kommunikationsforschung und Phonetik in Bonn

u.a. zentrale Arbeit relevant für die QL:

Menzerath, Paul (1954): *Die Architektonik des deutschen Wortschatzes.* Bonn: Dümmler. [= Phonetische Studien; 3]

Untersuchung der Silbenstruktur im Deutschen und anderen Sprachen

Es tritt eine "Sparsamkeitsregel" in Erscheinung, die sich psychologisch auf eine Ganzheitsregel dieser Art gründet: je größer das Ganze, um so kleiner die Teile! Diese Regel wird hier zum ersten Mal abgeleitet; sie wird aus der Tatsache verständlich, daß das Ganze jeweils "übersehbar" bleiben muß. (Menzerath 1954:

Quantitative Organisation der Länge von sprachlichen Einheiten

Gibt es einen Zusammenhang zwischen der Wortlänge und der Silbenlänge?

Spezifizierung von „Wortlänge“:

Zentrales Problem der Definition von „Wort“ (orthographische, phonologische, prosodische, morphologische, morphosyntaktische Kriterien)

Bestimmung + Definition der Silbe bzw. der Silbengrenze

<zver> f.sg.nom. ‘Raubtier‘
Länge in Graphemen: 4
Länge in Silben: **1**
durchschnittliche Silbenlänge = 4 Grapheme/Silbe

<delam> 1.p.sg.prä. ‘arbeiten‘
Länge in Graphemen: 5
Länge in Silben: **2**
durchschnittliche Silbenlänge = 5/2 = 2,5 Grapheme/Silbe

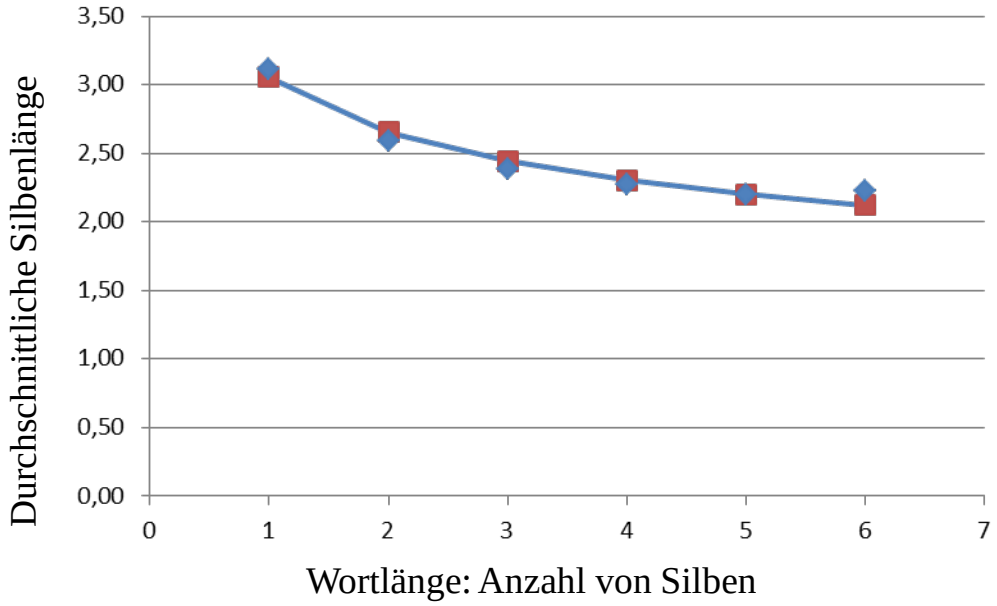
<dobrega> - m.sg.gen./acc. ‘gut‘
Länge in Graphemen: 7
Länge in Silben: **3**
durchschnittliche Silbenlänge = 7/3 = 2,33 Grapheme/Silbe

<b**e**s**e**dil**o**> - n.sg.nom. ‘Text‘
Länge in Graphemen: 8
Länge in Silben: 4
durchschnittliche Silbenlänge = 8/4 = 2 Grapheme/Silbe

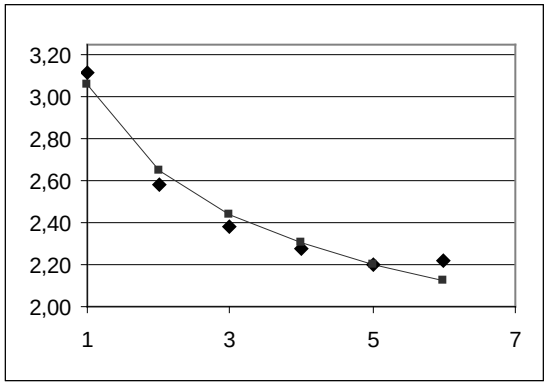
→ Je länger das Wort, desto kürzer die Silben.

Menzerath‘sche Gesetz (1)

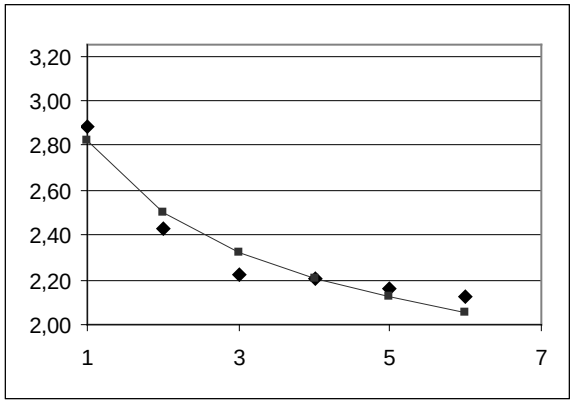
Je länger das Wort, desto kürzer die Silben.



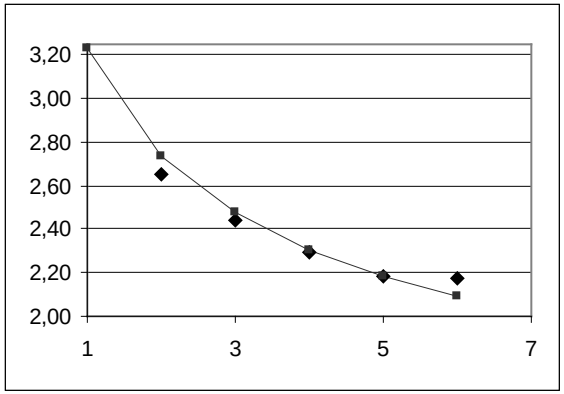
- in einer Vielzahl von Sprachen untersucht und belegt
- Was ist mit Sprachen ohne „eurozentrierten“ Wortbegriff ?



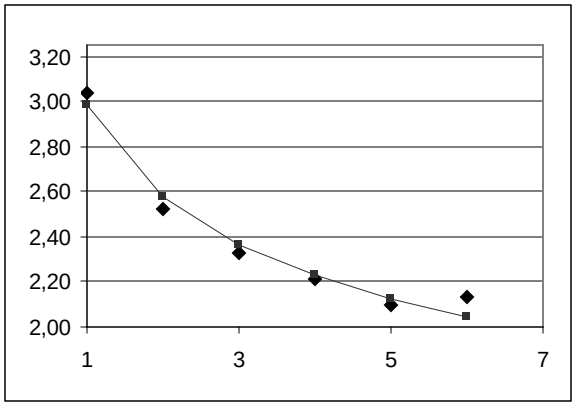
Slowenisch



Mazedonisch



Russisch



Tschechisch

- Zusammenhang von Wort- und Silbenlänge in slawischen Sprachen (Paralleltextkorpus)
- kann als eine Universalie betrachtet werden

Menzerath‘sche Gesetz (2)

(1) Mit zunehmender Wortlänge nimmt die Lautdauer ab.

(2) Mit zunehmender Wortlänge nimmt die Silbenlänge ab.

(3) Mit zunehmender Wortlänge nimmt die Morphlänge ab.

(4) Mit zunehmender Phrasenlänge nimmt die Wortlänge .

(5) Mit zunehmender Satzlänge nimmt die Teilsatzlänge (Clause) ab.

Je länger der Satz, d.h. je kürzer die Teilsätze.

→ Erweiterung auf andere sprachliche Ebenen

→ weitgehend unbekannt ist die Wirkkraft des MG bei über den Satz hinausgehenden Einheiten (Struktur von Texten)

→ zentraler Mechanismus der Steuerung: Konstruktgröße <--> Konstituente

„The longer a language construct, the shorter its components“ (Altmann 1980, S. 1)

Überleitung: QL und sprachliche Komplexität

Komplexität wird in unterschiedlichen Bereichen der Linguistik momentan intensiv diskutiert:

Givón, Talmy (2009): *The genesis of syntactic complexity: diachrony, ontogeny, neuro-cognition, evolution.* Amsterdam u.a.: Benjamins.

Hawkins, John A. (2004): *Efficiency and Complexity in Grammars.* Oxford u.a.: Oxford University Press.

Kortmann, Bernd; Szmrecsanyi, Benedikt (ed.) (2012): *Linguistic complexity. Second Language Acquisiton, Indigenization, contact.* Berlin: de Gruyter. [= Linguae & litterae; 13]

Kusters, Wouter (2003): *Linguistic complexity. The influence of social change on verbal inflection.* Utrecht: LOT.

Miestamo, Matti; Sinnemäki, Kaius; Karlsson, Fred (ed.) (2008): *Language complexity. Typology, contact, change.* Amsterdam/Philadelphia: Benjamins [= Studies in Language Companion Series; 94]

Sampson, Geoffrey; Gil, David; Trudgill, Peter (ed.) (2009): *Language Complexity as an Evolving Variable.* Oxford u.a.: Oxford University Press. [= Studies in the evolution of language; 13]

Unterschiedliche Arten der Komplexität

systemtheoretischer Hintergrund

a. deskriptive Komplexität
b. Konstituenten-Komplexität (Anzahl von Konstituenten, Varianz/Varietät von Konstituenten)
c. strukturelle Komplexität (Frage der hierarchischen Organisation, Wechselspiel der Einheiten bzw. Teilkomponenten)

→ Komplexität als eine quantifizierbare/messbare Eigenschaft

+ enge Anknüpfung an Probleme der phonologischen, morphologischen, syntaktischen Komplexität

+ Methodenapparat der QL bzw. allgemeinen Statistik steht zur Verfügung

+ Überlappung zur Verständlichkeitsforschung – sprachliche Kompetenz auf der Basis linguistischer Eigenschaften

sprachliche Kompetenz von Schülern (L1, L2)

Zweisprachige Schule: Hermagoras/Mohorjeva

Vergleich der sprachlichen „Kompetenz“

Vergleich von Deutsch – Slowenisch

Korpus von Schularbeiten = Texten (pre-processing, Lemmatisierung)

======
Vergleich: wissenschaftliche Texte und studentische wissenschaftliche Texte

Welche Eigenschaften werden diskutiert?

Wortschatz: Textlänge

- Anzahl von Wortformen (Types, Tokens)
- Anzahl von Lemma
- häufige/seltente Wortformen (Häufigkeitswörterbücher)

Hier besonders von Relevanz:

- Häufigkeitsverteilung: Zipf‘sche Verteilung
- Berechnung der **Type-Token-Ratio**

weitere (häufig) diskutierte Eigenschaften

- **Wortlänge**

- Satzlänge

- Häufigkeiten von Substantiven (S), Verben (V), Adjektiven (A) und Adverben (Adv), Konjunktionen, Pronomen,
- Bildung von Quotienten aus S, V, A, Adv.
- Subgruppen von S: (konkret vs. abstrakt)
- Häufigkeit von Pronomen (insbesondere bei Schulkindern, aber in Abhängigkeit von Aufgabe)
- Häufigkeit von Verben + Häufigkeit morphologischer Kategorien beim Verbum
- Häufigkeit von Adjektiven + Verben (Wechselbeziehungen)
- Untersuchung von als „komplex“ geltenden Formen:
- Passiv
- Konjunktiv
- „Freiheitsgrade“ im Kasus-System: z.B. Genitiv im Slow.
- Kollokationen – Kookkurrenzen
- … … ….

Lexikalischer Reichtum: Type-Token-Ratio

Untersuchung des Leistungsstandes von Schülerinnen und Schülern der VS Hermagoras Seminararbeit von Gerald Robatsch (SoSe 2012)

Textlänge
Anzahl von Wortformen (Types, Tokens)
(Perspektive: Anzahl von Lemma)

→Hier besonders von Relevanz:
Häufigkeitsverteilung
Messung des lexikalischen Reichtums?

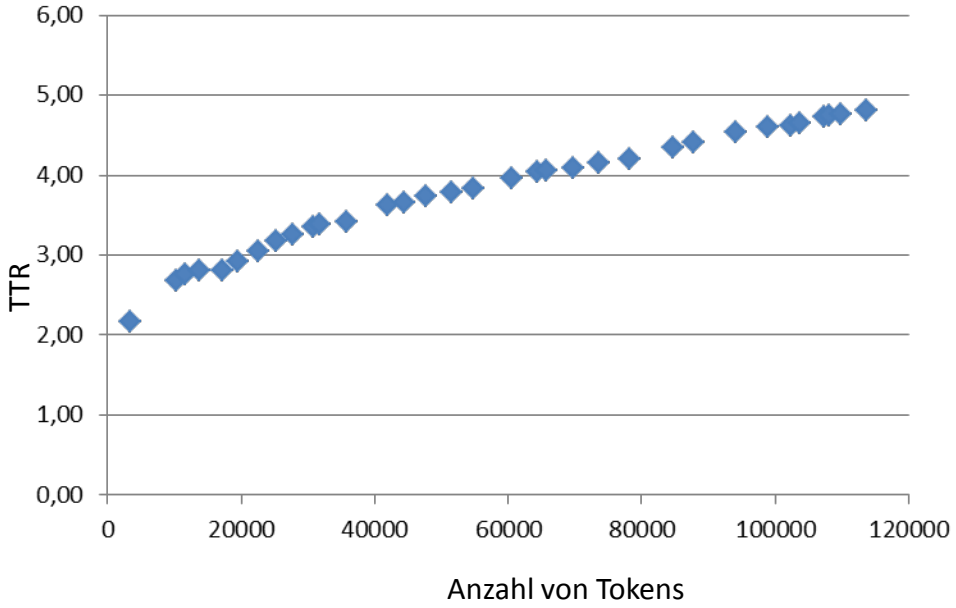
Vorschläge: Type-Token-Ratio (TTR)

- Auskunft über die lexikalische Diversifikation ?
- Entwicklung der TTR im Text ?
- besondere Bedeutung von hapax legomena ?

Berechnung der TTR:
Nicht lemmatisierte Texte:
TTR = Anzahl von Wortformen-Types/Anzahl von Wortformen-Tokens

Entwicklung in einem Text (Master i Margarita, russisch, 33 Kapitel)
Kapitel 1; Kapitel 1,2; Kapitel 1,2,3 …; Kapitel 1,2,3, 4 …

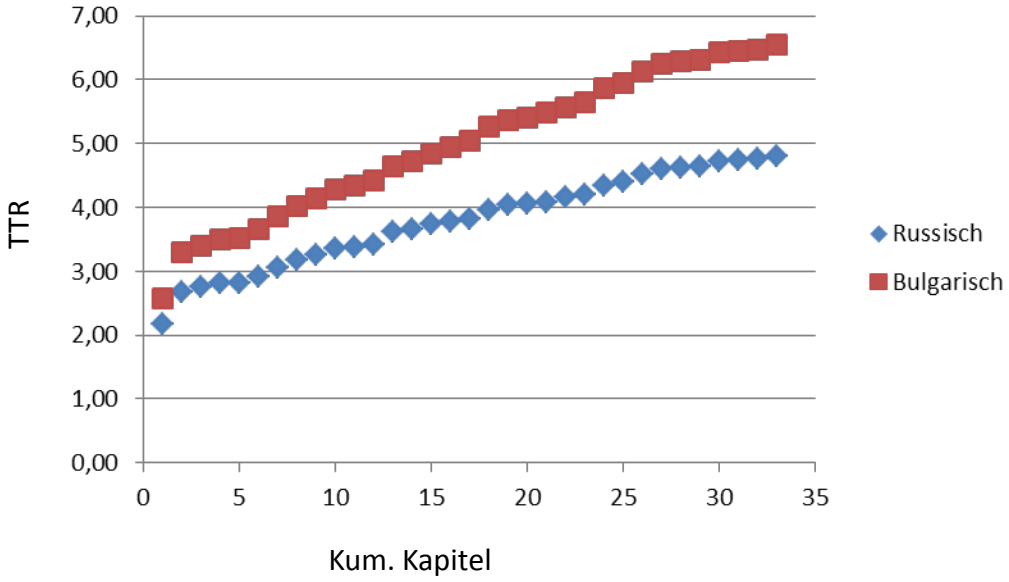
Kum. Kap.	Tokens	Types	TTR
1	3405	1573	2,16
2	10234	3822	2,68
3	11600	4215	2,75
4	13763	4901	2,81
5	17171	6121	2,81
…	…	…	…
28	102414	22157	4,62
29	103733	22306	4,65
30	107275	22669	4,73
31	108058	22789	4,74
32	109926	23056	4,77
33	113748	23640	4,81

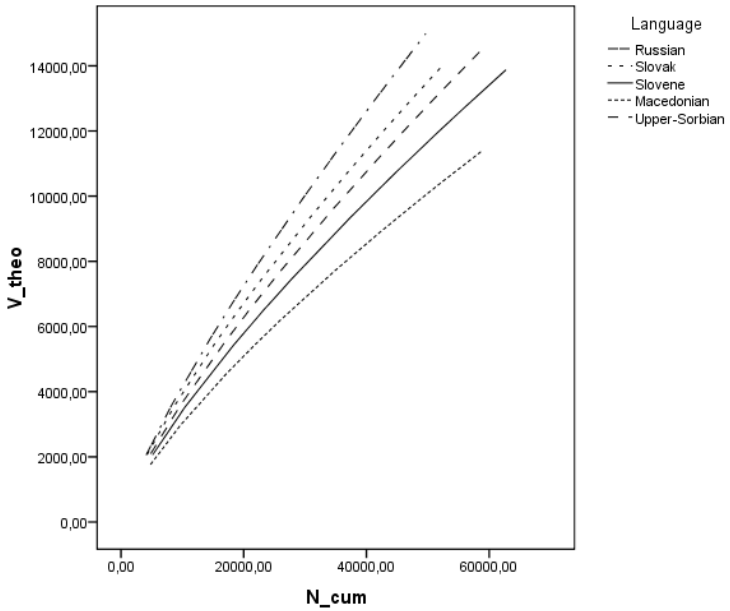


→ TTR hängt von der Textlänge ab
→ Vergleich von Texten nur eingeschränkt möglich
→ wenn Vergleich, dann in ca. gleich lange Texte
→ Testen von statistisch signifikanten Unterschieden?

Vergleich der TTR in unterschiedlichen Sprachen

Text A: Russisch (Master i Margarita)
Text B: Bulgarisch (Master i Margarita)





Bedeutung der TTR für kontrastive Untersuchungen?

Morphologische Komplexität: Wortlänge

- morphologische Natürlichkeitsforschung
- Greenberg‘schen Sprachtypologie (analytisch – synthetisch)
- Verständlichkeitsforschung (reading ease, Flesh-Formel)
- Psycholinguistik
- QL
- Stilometrie
- Autorenschaftsbestimmung

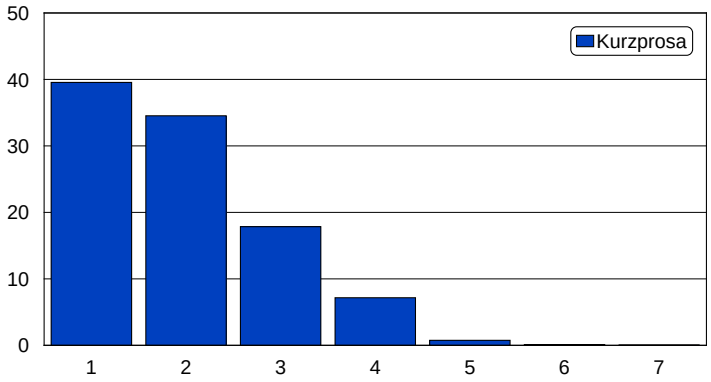
Untersuchung der WL in unterschiedlichen Textsorten

(Russisch, Kroatisch, Slowenisch)

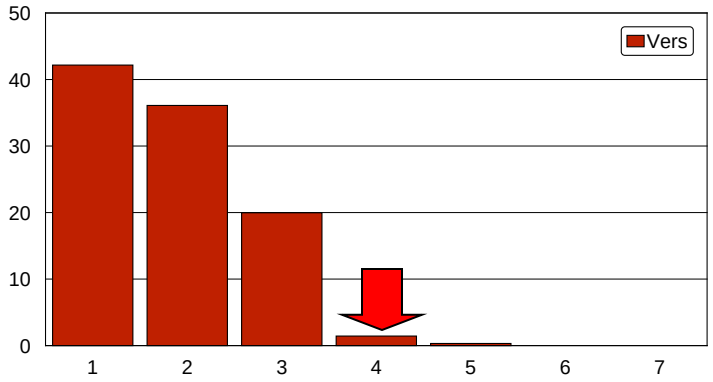
(Erfahrungsbericht aus einem Forschungsprojekt 2002-2005)

Alltag	Wissenschaft	Administration	Journalistik	Kunst		
				Prosa	Poesie	Dramatik
1	*2*	*3*	*4*	*5*	*6*	*7*
Kochrezept	Abstract	Anleitung	Agenturmeldung	Autobiographie	Elegie	**Drama**
Privatbrief	Aufsatz	Geschäftsbrief	Auslandsbericht	Biographie	Epos	Komödie
Tagebucheintrag	Autorreferat	Gesetzestext	Fachartikel	**Briefroman**	**Gedicht**	Tragödie
Witz	Diplomarbeit	Gutachten	Feuilleton	Epilog	Ode	Versdrama
	Dissertation	**Offener Brief**	Glosse	Erinnerungen	Sonett	
	Referat	Parteitagsbeschluss	Kolumne	Erzählung	Verserzählung	
	Rezension	Predigt	**Kommentar**	Fabel	Versroman	
	Tagungsbericht	Schreiben	Kritik	Gleichnis		
		Vertrag	**Leserbrief**	Kunstmärchen		
		Vortrag	Meldung	**Kurzroman**		
			Sportbericht	Legende		
			Wetterbericht	Mythos		
			Zeitschriftenaufsatz	Novelle		
			Zeitungsartikel	Roman		
				Sage		
				Schwank		
				Tagebuchroman		
				Volksmärchen		

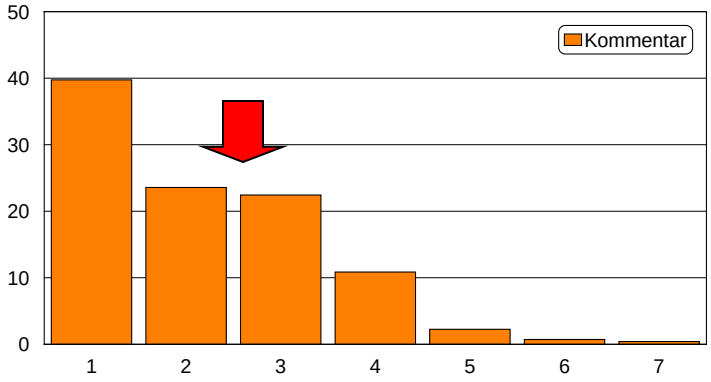
Wordlängenhäufigkeiten in unterschiedlichen Texten (in %)



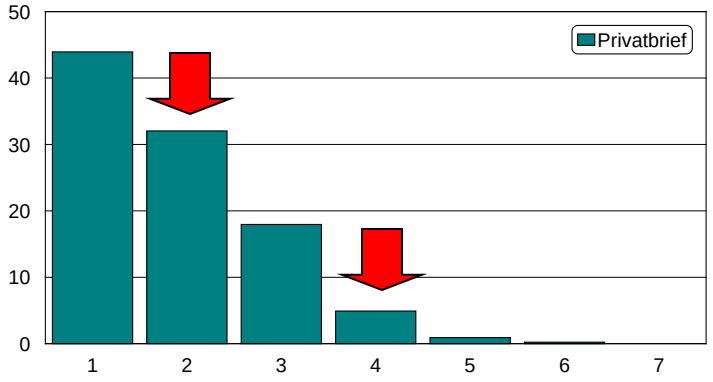
Literary Prose Text (#256)



Versified Poetic Text (#359)



Journalistic Comment (#324)

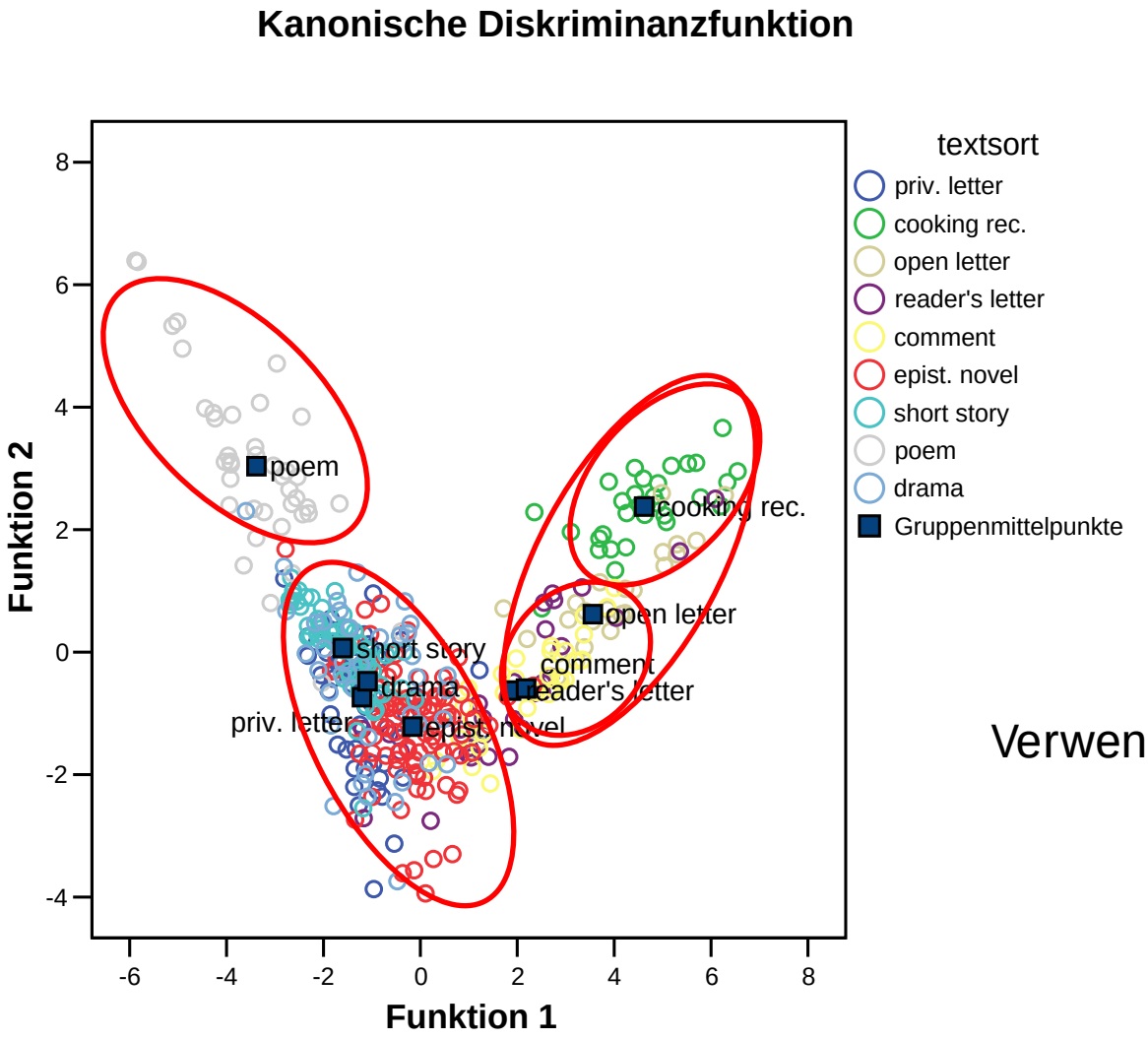


Private Letter (#1)

Text-Classification (429 Slovenian Texts)

FUNCTIONAL STYLE	AUTHOR(S)	TEXT TYPE(S)	NUMBER
EVERYDAY LANGUAGE	Cankar, Jurčič	Private **Letters**	61
PUBLIC STYLE	div. anon.	Open **Letters**	29
JOURNALISM	div. anon.	Readers' **Letters**, Comments	65
ARTISTIC STYLE *Prose*	Cankar Švigelj-Mérat / Kolšek	Individual Chapters from Short Novels („povest") **Letters** from an Epistolary Novel	68 93
Poetry	Gregorčič	Versified Poems	40
Everyday language	n.n.	Cooking recepies	32
Drama	Jančar	Individual Acts from Dramas	42

Wortlänge (WL)



429 Slowenische Texte

(9 Textsorten)

Verwendete Variablen:

m_1, s_1, v, p_1

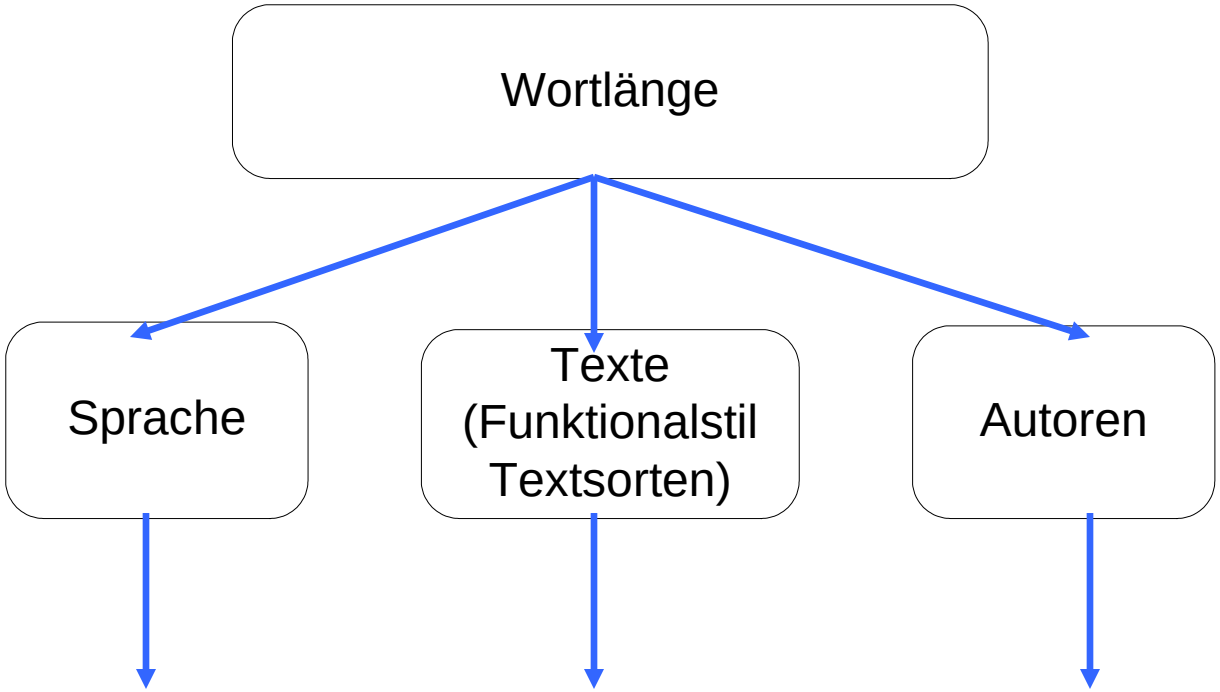
(59.70%)

Bestimmung der relevanten Variablen

Variable	Bezeichnung
$m_2 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \overline{x} \right)^2$	Varianz (2. Zentralmoment)
$o_i = \frac{m_2}{m_1}$	Ord'sches Kriterium *i*
$m_4 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \overline{x} \right)^4$	4. Zentralmoment
$p_4 = \frac{f_4}{N}$	rel. Häufigkeit 4-silb. Wörter
$v = \frac{\sqrt{m_2}}{m_1}$	Variationskoeffizient
$d = \frac{m_2}{m_1 - 1}$	Dispersionsquotient

→ Texte werden als 6-dimensionaler Vektor dargestellt

Ist die Wortlänge autorenspezifisch?



Russisch

Alltag→ Briefe

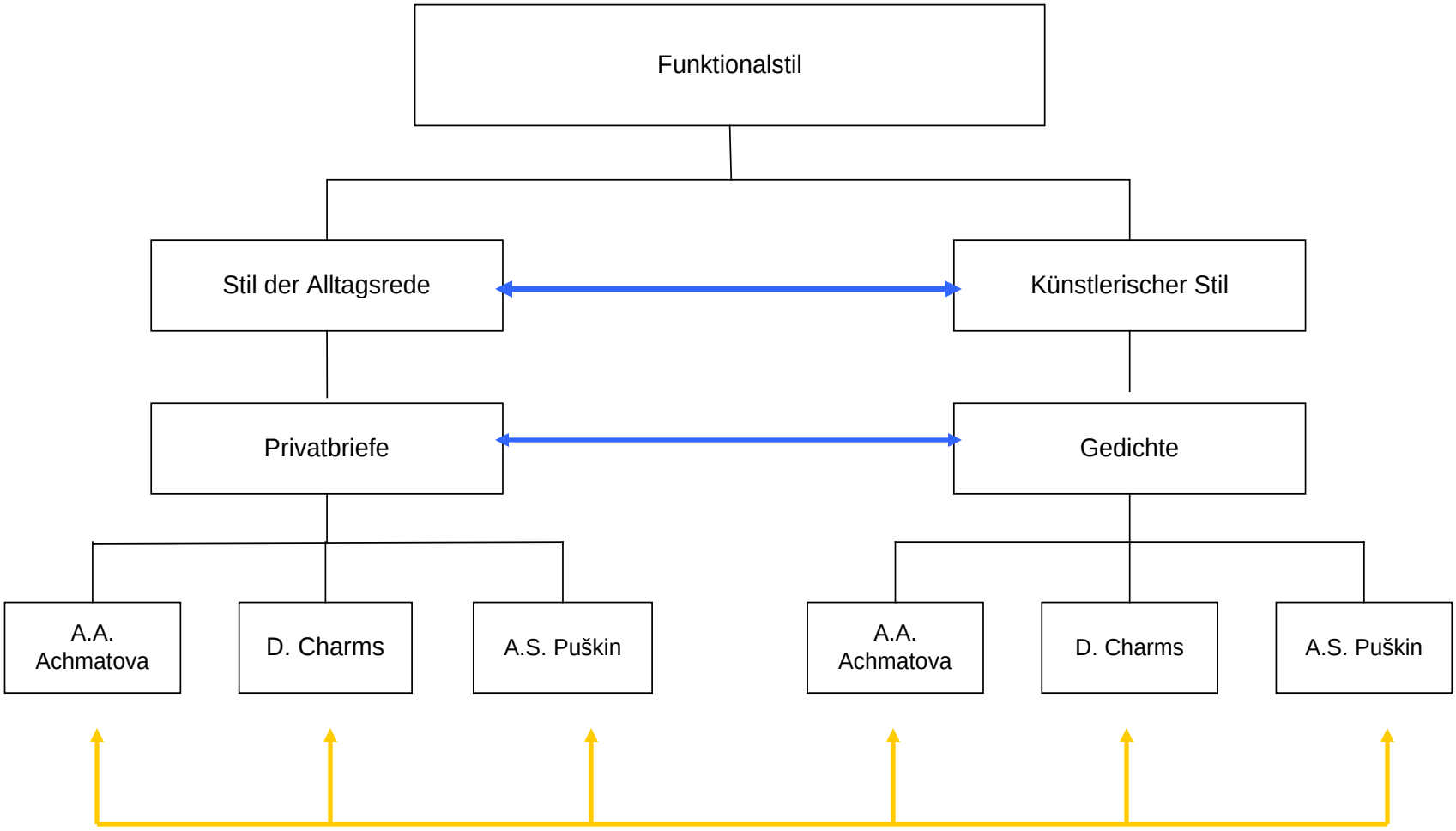
Kunst→ Gedichte

A. Achmatova

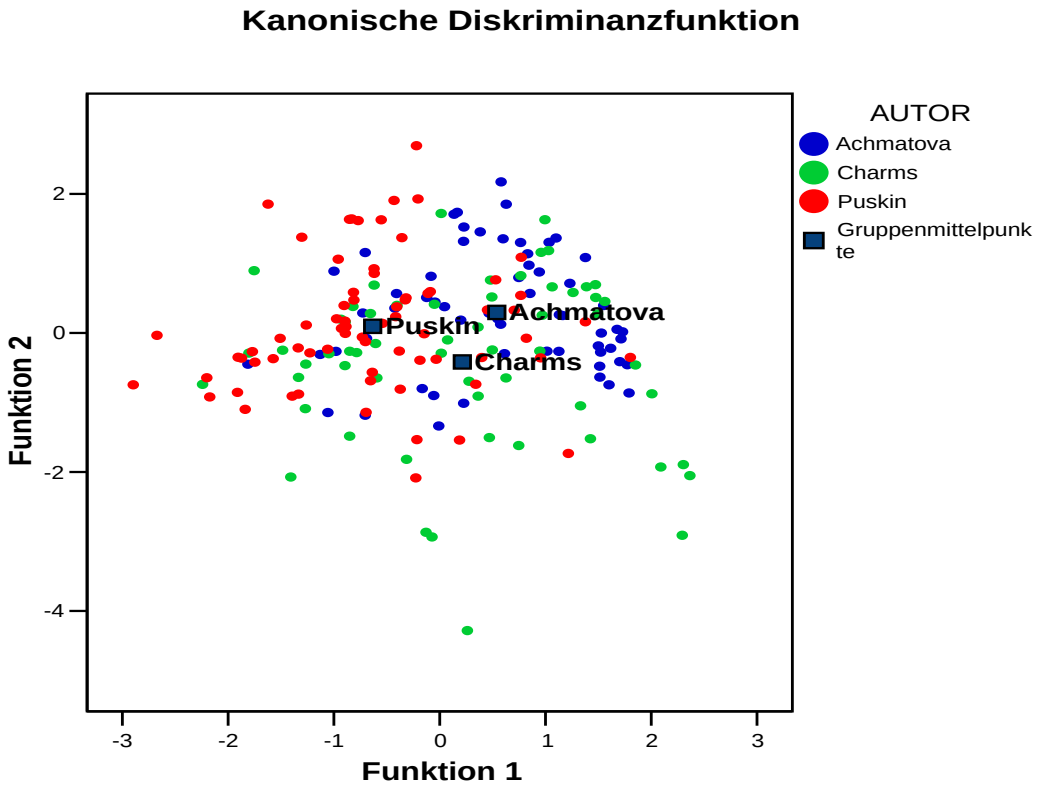
D. Charms

A. Puškin

Textklassifizierung nach Funktionalstilen/Textsorten und Autoren



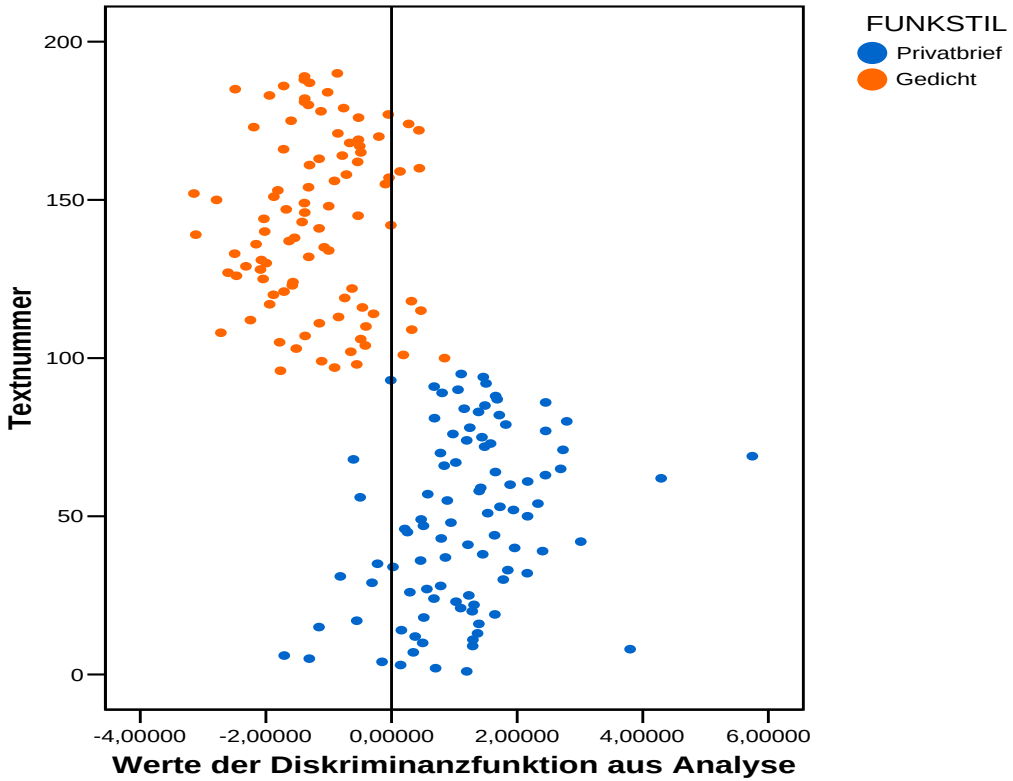
Klassifizierung der Autoren nach der Wortlänge?



→ 57,4% korrekt zugeordnete Texte

→ Nach Autoren ist keine sinnvolle Trennung von Texten möglich!

Trennung der Texttypen (nach p_4, d)



→ korrekte Klassifizierung:

88,4% der Briefe

90,5% der Gedichte

Klassifizierungsergebnisse			
Texttyp	Briefe	Gedichte	Gesamt
Briefe	84	11	95
Gedichte	9	86	95

Ausgewählte Ergebnisse zur Relevanz der Wortlänge

- WL keine Autorenspezifität
- Text (individuell) ist überlagert von Textsorte/Funktionalstil
- offene Fragen und Probleme:
- WL in unterschiedlichen Einheiten messen (!)
- WL: Test auf statistische Unterschiede: Problem: (i.d.R.) keine Normalverteilung von linguistischen Daten, Anwendung: nicht parametrische Testverfahren

Perspektiven und Grenzen der quantitativen Text- und Sprachanalyse

- Auswahl von linguistischen Eigenschaften/Merkmalen
- linguistische Interpretation/Einbettung vor der Operationalsierung / der Messung
- Bedeutung der Hypothesen-Bildung
- Übersetzung von der „Sprache der Linguistik“ in die „Sprache der Statistik“
- keine „blinde“ Anwendung statistischen Verfahren auf linguistische Daten
- statistisches Lehrbuchwissen in vielen Fällen für sprachliche Probleme nicht geeignet
- zentrale Probleme: Homogenität, fehlende Normalverteilung, extreme Schiefe der linguistischen Häufigkeitsverteilungen, Signifikanztests können z.T. nicht standardmäßig angewandt werden … …
- interdisziplinäre Zusammenarbeit gefordert

ABER: jede empirische Untersuchung = Betreten von Neuland = kleiner Mosaikstein

Ausgewählte Literatur:

Altmann, Gabriel (1980): Prolegomena to Menzerath's law. In: Rüdiger Grotjahn (ed.): *Glottometrika 2*. Bochum: Brockmeyer (Quantitative Linguistics, 3), S. 1–10.

Bußmann, Hadumod (Hg.) (2008): Lexikon der Sprachwissenschaft. Vierte, durchgesehene und bibliographisch ergänzte Auflage. Stuttgart: Kröner.

Köhler, Reinhard (2005): Gegenstand und Arbeitsweise der Quantitativen Linguistik. In: R. Köhler, G. Altmann und R.G. Piotrowski (Hg.): *Quantitative Linguistik. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook.* Berlin, New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 27), S. 1–16.

Zipf, George K. (1935): *The Psycho-Biology of Language. An Introduction to Dynamic Philology.* Cambridge: M.I.T. Press.

Zipf, George K. (1949): *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology.* Cambridge, MA: Addison-Wesley.