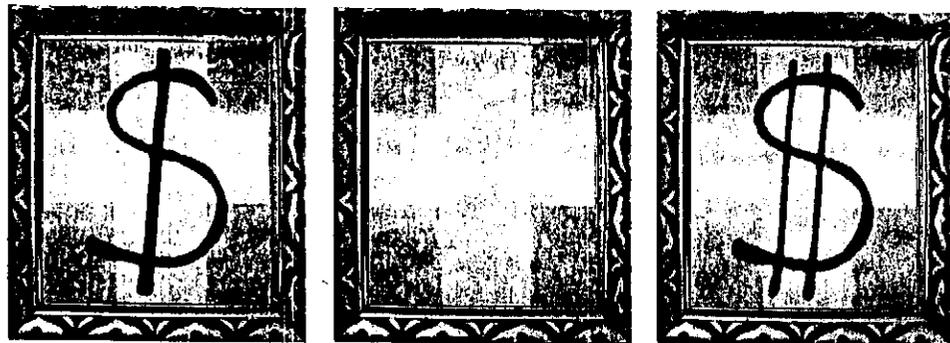


KULTUR. SPRACHE. ÖKONOMIE.

Beiträge zur gleichnamigen Tagung an
der Wirtschaftsuniversität Wien
3. – 5. Dezember 1999



WIENER SLAWISTISCHER ALMANACH
SONDERBAND 54
WIEN 2001

Wiener Slawistischer Almanach, Sonderband 54 (2001) 485–509

Peter Grzybek

KULTUR-ÖKONOMIE Zur Häufigkeit text-konstitutiver Elemente

1. Einleitung

Die vorliegenden Überlegungen zum Zusammenhang von *Sprache – Kultur – Ökonomie* sind in gewissem Sinne eher programmatischen denn analytischen Charakters, insofern sie weniger fertige Ergebnisse präsentieren als vielmehr aussichtsreiche Forschungsperspektiven aufweisen. Andererseits aber sollen sie gerade auch auf analytische Perspektiven verweisen, die – durchaus von der Sprache ausgehend – auf spezifische Art und Weise wesentliche Aspekte einer Kultur-Ökonomie transparent machen sollen.

Bei dieser Art von Fragestellung ist von vornherein davon auszugehen, dass sich der Zusammenhang von *Sprache – Kultur – Ökonomie* auf verschiedene Art und Weise verstehen und untersuchen lässt; diese Heterogenität möglicher Herangehensweisen wird in erster Linie durch die Verstehensweise des Begriffs der 'Ökonomie' selbst begünstigt, der sich in unterschiedlicher Form zur 'Sprache' und zur 'Kultur' in Beziehung setzen lässt. Die gängigste Praxis innerhalb der geisteswissenschaftlichen Tradition dürfte die *thematische* und/oder *metaphorische Aneignung* des Ökonomie-Begriffes sein. Solche „Aneignungen“ finden sich nicht zuletzt auch und gerade in der Literaturwissenschaft zuhauf, etwa wenn – um nur ein provokant-prototypisches Beispiel zu geben – das 'Chaos' im Hause der Familie Oblonskij am Anfang von Tolstojs *Anna Karenina* auf einer inhaltlich-bildlichen Ebene unverblümt zur Chaostheorie in Bezug gesetzt wird o. ä. Dabei ist diese Art von „thematischer Aneignung“ noch relativ harmlos im Vergleich zu terminologischen Usurpationen aus anderen (vor allem naturwissenschaftlichen) Bereichen und Diskursen, mit deren (vor allem) postmoderner Umgangsweise unlängst Sokal/Bricmont in ihrer Abhandlung *Eleganter Unsinn* (1999) ebenso radikal wie arrogant abgerechnet haben.

Im Sinne der „thematischen Aneignung“ würde es also – um wiederum nur einfachste, doch in gewissem Sinne charakteristische Beispiele zu geben – bei der Untersuchung des Zusammenhangs von Sprache – Kultur – Ökonomie etwa um sprachliche (lexikalische, phraseologische, sprachspezifische oder sprach-

übergreifende u. a.) Konzeptualisierungen ökonomischer Konzepte wie 'Armut', 'Reichtum', 'Geld' o. a. gehen; oder um die historische Dimension der Semantisierung dieser Begriffe; oder um die Verarbeitung ökonomisch relevanter Themen und Konzepte in literarischen Texten (unter Berücksichtigung autoren-, gattungs-, oder epochenbedingter Spezifika), u. v. a. m. – all dies mit dem Ziel, die kulturhistorischen Implikationen der Relevanz ökonomischer Konzepte im sozialen und kulturellen Geflecht zu (re-)konstruieren.

Der Wert solcher diskursanalytischer und kulturhistorischer Rekonstruktionen soll hier mitnichten in Frage gestellt werden. Dies sei umso mehr betont, als den folgenden Ausführungen ein anderes Verständnis des Ökonomie-Begriffs hinsichtlich seiner Relevanz für den Zusammenhang von 'Sprache' und 'Kultur' zugrunde liegt: 'Ökonomie' wird im Folgenden nämlich nicht als Thema, sondern als (strukturelles) *Prinzip* verstanden. Auch diese Verstehensweise ist natürlich nicht neu. Gerade im Kontext der Beschäftigung mit slawi(sti)scher Sprach- und Kulturtheorie ruft der Begriff der 'Ökonomie' unmittelbar frühe Arbeiten des Russischen Formalismus in Erinnerung: So stand das 'Gesetz der Ökonomie der schöpferischen Kräfte' [закон экономики творческих сил] als „Gesetz und Ziel des Schaffens“ ganz am Anfang der Überlegungen, die Viktor Šklovskij (1916) in seinem oft als Manifest des Russischen Formalismus bezeichneten Aufsatz „Kunst als Verfahren“ entwickelte. Unter Bezugnahme auf H. Spencer,¹ W. James,² R. Avenarius³ u. a. akzeptierte Šklovskij im Prinzip die von den Genannten entwickelten Grundannahmen zur Ökonomie psychischer Kräfte, schränkte die Relevanz der einschlägigen Annahmen aber auf den Bereich der (als 'Sonderfall der Sprache' verstandenen) 'praktischen Sprache' ein – im Hinblick auf die 'dichterische Sprache' hingegen dürfe man nicht aufgrund einer Analogie zur prosaischen Sprache von Gesetzen der Verschwendung und Ökonomie sprechen, sondern nur aufgrund ihrer eigenen Gesetze: Zwar bestehe die größte Ökonomie der Wahrnehmungskräfte in der (Ver-)Automatisierung,

¹ Spencer, Herbert (1820–1903), englischer Sozialtheoretiker, der in seinen Schriften stark auf die (Lamarck'sche) Evolutionstheorie Bezug nahm.

² James, William (1842–1910), amerikanischer Philosoph und Psychologe, u. a. bekannt dafür, dass er sich recht früh mit dem semiotischen Werk von Charles S. Peirce und dessen Pragmatizismus auseinandersetzte.

³ Avenarius, Richard (1843–1896), neben Ernst Mach führender Vertreter des sog. Empirio-kritizismus, dem es um eine biologische Begründung von Logik und Erkenntnistheorie ging. Ihm zufolge arbeitet die Psyche wie alles organische Leben nach dem Prinzip des kleinsten Kraftmaßes: Ein Wesen passt sich seiner Umwelt umso besser an, je schneller und einfacher es die zu seiner Erhaltung notwendigen Leistungen erbringen kann. (*Philosophie als Denken der Welt gemäß dem Prinzip des kleinsten Kraftmaßes*, 1876). Mach nannte dieses Modell einer möglichst rationalen Erkenntnisleistung *Denkökonomie*, die besagt, dass wissenschaftliche Sachverhalte auf einfachste Weise und mit dem geringst möglichen Aufwand an Denkvorgängen erfasst werden sollten – eine Idee, die Avenarius übernahm.

doch sei es Merkmal des Künstlerischen, dass es absichtlich für eine vom Automatismus befreite Wahrnehmung geschaffen sei.

Die hier aufgeworfene Frage nach der Relevanz des Gesetzes der Ökonomie der schöpferischen Kräfte und ihrer weiteren Ausarbeitung in der Literatur- und Kunsttheorie kann an dieser Stelle natürlich nicht im Detail systematisch verfolgt werden. Festzuhalten wären allerdings zwei wichtige Fragen:

1. Lässt sich die postulierte „Ökonomie der Kräfte“ in systematischen (z. B. textuellen) Gegebenheiten nachweisen?
2. Wenn ja, unterscheiden sich diesbezüglich künstlerische und nicht-künstlerische Texte?

Wie gesagt, diese Fragen lassen sich im gegebenen Zusammenhang nicht im Detail verfolgen, geschweige denn befriedigend beantworten. Es soll lediglich eingangs auch auf die historische Dimension dieser Fragestellung hingewiesen werden, die sich in gewissem Sinne sogar weiter bis in die antike Rhetorik zurückverfolgen, wenn auch nicht direkt aus ihr ableiten lässt. Bevor es im folgenden also um die konkrete Frage gehen soll, inwiefern ein als Strukturprinzip verstandener Ökonomiebegriff sich für kulturelle und kulturwissenschaftliche Fragestellungen als von Relevanz erweisen kann, soll deshalb die Tragweite dieser historischen Dimension der Fragestellung zumindest ansatzweise skizziert sein.

2. Die historische Dimension des Ökonomie-Begriffs

Wenn soeben gesagt wurde, dass sich der für uns relevante Ökonomie-Begriff bis in die antike Rhetorik zurückverfolgen lässt, so gilt dazu einschränkend zu sagen, dass er im Zusammenhang mit den fünf Haupttätigkeiten (den 'officia' bzw. 'εργα') des Redners zunächst überhaupt nicht in Erscheinung trat. Noch im ersten nachchristlichen Jahrhundert stufte Quintilian in seinem *Lehrbuch der Rhetorik* (3,3,9) 'oeconomia' als griechisches (Fremd-)Wort ein, für das es im Lateinischen keine unmittelbare Entsprechung gebe; es beziehe sich eigentlich auf die Sorge um die Führung eines Haushalts und stehe im Kontext der Rhetorik in uneigentlicher Bedeutung. Damit ist jedoch indirekt ein Zusammenhang zwischen Rhetorik und dem Ökonomie-Begriff hergestellt, und in der Tat kann es als gesichert angesehen werden, dass etwa ab dem ersten Jahrhundert v. u. Z. der Begriff 'oikonomia' im Griechischen dem zuvor schon verwendeten Begriff der 'taxis' zunehmend gleichgestellt wurde. Gemeint war in jedem Fall die „Gliederung des Stoffes“, die im Lateinischen mit 'dispositio' bezeichnet und in der Folge dann auch mit 'ordo' wiedergegeben wurde. Somit ergab sich das „klassische“ fünfgliedrige Schema der rednerischen Tätigkeit:

1	Auffindung des Materials; Stoffsamml.	inventio		heuresis
2	Anordnung und Gliederung des Stoffes	dispositio		taksis
				↓
		ordo	←	oikonomia
3	sprachl. Formulierung; stilist. Gestaltung	elocutio		lexis
4	Auswendiglernen, Memorieren der Rede	memoria		mneme
5	Vortrag	pronuntiatio, actio		hypokrisis

Formuliert man diese Bedeutungsübertragung in moderne Terminologie um, so wurde hier, in der klassischen Rhetorik, der Ökonomiebegriff vom Bereich der „ordentlichen Haushaltsführung“ auf den Bereich des *Textes* und dessen formale Strukturierung übertragen. Ein anderer, jedoch ähnlicher Transfer fand später, in anderem Zusammenhang, auch auf den Bereich der Lexik (und damit auf den Bereich des Sprachsystems) statt. Das zeigt sich besonders deutlich am Beispiel des Englischen: Wurde hier der Begriff 'yconomie' um 1530 im Sinne von Haushaltsführung („Howsolde keepyng“) eingeführt, so gab ihm schon 1658 Edward Philips, ein Neffe und Schüler von John Milton, in einem der frühesten englischen Lexika, *The New World of English Words*, eine ganz andere Bedeutung: Seine Untersuchung über Herkunft, Sinn und Wortbildung der Fremdwörter resümiert Philips nämlich mit der Feststellung, er habe die gesamte Ökonomie der Fremdwörter untersucht: „Thus I have, in as brief a Method as I could device, run through the whole O e c o n o m y of our foreign words, and I have arranged them all into their several orders and distinctions.“

Man könnte meinen, 'Ökonomie' sei hier in Fortsetzung der auf den Haushalt bezogenen Verwendungsweise im Sinne eines (lexikalischen) Inventars zu verstehen; allerdings hat nach Sperk (1985: 31) der Wortgebrauch von 'Ökonomie' hier eigentlich nichts mehr mit Haushalt oder Haushaltsführung zu tun, sondern bezeichnet „ein komplexes, aber doch erklärbares lexikalisches System“, weshalb das Wort sinngemäß auch mit lateinisch *ordo*, englisch *order* oder deutsch *Ordnung* wiedergegeben werden könnte. In diesem Sinne verstanden, wurde der Ökonomiebegriff, nachdem er in der antiken Rhetorik auf den Bereich des *Textes* bezogen worden war, nun – ebenfalls in moderner Terminologie ausgedrückt – auch auf den Bereich des sprachlichen *Codes* bzw. der sprachlichen *Struktur* übertragen.

Solche Übertragungen des Ökonomiebegriffs auf den Bereich der Sprache lassen sich in den folgenden Jahrhunderten immer wieder finden, und sie sollten dabei auch auf dem einen oder anderen Weg zum Strukturalismus führen. In eine dieser Linien lässt sich z. B. Jacob Grimm einordnen, demzufolge die Sprache sich „überall haushälterisch“ zeigt: „[...] sie weist die kleinsten, unscheinbarsten Mittel auf und reicht damit doch zu großen Dingen hin“; in Fortführung dieser Gedanken (und in erweiternder Übertragung auf der Meta-Ebene des Sprachlichen) sollte später Chomsky seine Transformationsgrammatik nicht zu-

letzt deshalb für besonders ökonomisch halten, weil diese sich zur Erzeugung einer unendlichen Anzahl von Äußerungen mit endlichen Mitteln eigne. Eine andere, unmittelbar strukturalistische Linie nimmt ihren Ausgang bei Saussure, der seine system-orientierte Sprachwissenschaft in zahlreichen Analogien zu wirtschaftswissenschaftlichen Konzepten entwickelte und dabei u. a. die Sprache als ein „System reiner Werte“ [„un système de pures valeurs“] bezeichnete. Dabei hatte Saussure zwar der Sprachdiachronie noch systematischen Charakter abgesprochen; doch Zweifel, wie sie schon Ende der 20er Jahre von Jakobson, Tynjanov, oder Karcevskij geäußert wurden, legten recht bald die Annahme nahe, dass der Wandel eines Sprachsystems ebenfalls systematischen Charakters ist und die Ökonomie des Systems bewahrt. Der Nachweis dieser Auffassung wurde später häufig mit Martinet verbunden, der in den 50er und 60er Jahren von einer „ständigen Antinomie“ und einem daraus hervorgehenden „Gleichgewicht“ zwischen den Kommunikationsbedürfnissen des Menschen einerseits und seiner Tendenz, seine geistige und körperliche Tätigkeit auf ein Minimum zu beschränken, sprach. Dieser Sichtweise zufolge versteht sich also die Ökonomie des Systems nicht (primär) als Selbstzweck, sondern als Ergebnis des sprachlichen Verarbeitungs- oder auch Entwicklungsprozesses.

Allerdings ist diese Betrachtungsweise im wahrsten Sinne des Wortes einseitig: Betrachtet wird nämlich die sprachliche Ökonomie (und in der Folge die Ökonomie des Sprachsystems) als Ergebnis einer *produktionsseitigen* 'Ökonomie' – die damit letztendlich im Sinne von energetischer „Kosten“-Minimierung auf Produktionsseite zu verstehen ist. Was in dieser Sichtweise somit unberücksichtigt bleibt, ist die Komponente der Rezeption und damit auch der systematische Gesamtzusammenhang der Fragestellung. Und dieser Zusammenhang ist keineswegs trivial oder eindeutig, insofern maximale 'Ökonomie' (eben im Sinne von minimalem Aufwand) für den Produzenten keinesfalls einer erhöhten Ökonomie für den Rezipienten gleichkommen muss, sondern, ganz im Gegenteil, dessen Energien umso mehr erfordert. In diesem Sinne setzt ein umfassenderes Verständnis von Ökonomie eine gesamtheitliche Sicht auf den Kommunikationsprozess voraus.

Es ist das Verdienst von George Kingsley Zipf (1902–1950), bereits in den 30er und 40er Jahren die Komplexität dieses Zusammenhangs systematisch untersucht zu haben. Auf seinen Ansatz gilt es folglich, ein wenig detaillierter einzugehen.

3. Von Zipf zu Zipf-Mandelbrot zur Theorie der Verteilungen

Zipf hat nicht nur recht früh die Implikationen und Konsequenzen eines umfassenden Konzepts von System-Ökonomie erkannt – die gerade im jüngster Zeit vor dem Hintergrund synergetischer Konzeptionen neu beleuchtet werden (vgl.

Prün 1999) –, sondern darüber hinaus auch erste Hypothesen im Hinblick auf eine mathematische Formalisierung und damit auch Operationalisierbarkeit formuliert. Der mathematische Ansatz erweist sich dabei nicht als Selbstzweck, sondern als eine Möglichkeit, Analogien und Konvergenzen heterogener Systeme beschreibbar zu machen. Dies ist für die vorliegenden Überlegungen insofern von Bedeutung, als sich der Zipf'sche Ansatz so im Hinblick auf die Kultur in einem semiotischen Zugang für die Beschreibung heterogener Zeichensysteme nutzbar machen lässt.

So war auch Zipf selbst zunächst von der Beobachtung nur sprachlicher Phänomene ausgegangen, hatte in der Folge aber immer wieder parallele Untersuchungen zu nicht-sprachlichen Fragestellungen durchgeführt (wie z. B. zur Anzahl und Größe von Städten und deren Entfernung zueinander u. v. a. m.). Mittlerweile sind seine grundsätzlichen Ideen eigentlich seltener auf kultur-, sondern vielmehr naturwissenschaftliche Problemstellungen (wie z. B. in der DNA-Analyse, der Erdbebenforschung, u. v. a. m.) übertragen worden.

Gerade auch Parallelen zwischen kulturellen (Zeichen-)Systemen und dem naturwissenschaftlichen Bereich machen den Vergleich zur Sprache und zur Kultur besonders spannend und relevant; doch kann auf diese Querbeziehungen im hier gegebenen Zusammenhang nicht eingegangen werden. Statt dessen soll der Zipf'sche Ansatz als Ausgangspunkt genommen werden, um der These nachzugehen, dass die (im weiten semiotischen Sinne verstandenen) 'Texte' einer Kultur einer formalisierbaren ökonomischen Systematik bzw. einer systematischen Ökonomie unterliegen. Der Umstand, dass Zipfs ursprüngliche Vorschläge zur Mathematisierung seiner Überlegungen dabei verschiedentlich modifiziert, erweitert oder spezifiziert worden sind, ist dabei von zweitrangiger Bedeutung: Der Zipf'sche Ansatz soll vielmehr grundsätzlich darauf hin befragt werden, inwiefern sich mit ihm zeichensystem-übergreifende Konvergenzen und/oder Analogien metasprachlich erfassen bzw. beschreiben lassen. Wenn das gelingt, würde das wesentlichen Einblick in das Funktionieren von Zeichensystemen vor dem Hintergrund einer komplexen Kultur-Ökonomie bieten – zumindest würde es erlauben, entsprechende Hypothesen zu begründen, deren Überprüfung freilich an anderer Stelle im Detail nachzugehen wäre.

3.1. Zipf

In seinem 1935 publizierten Buch *The Psycho-Biology of Language* mit dem bezeichnenden Untertitel *An Introduction to Dynamic Philology* hatte Zipf u. a. Argumente dafür geliefert, dass die Vorkommenshäufigkeit von Wörtern in Texten nicht zufällig, sondern gesetzmäßigen Charakters ist. Genauer gesagt, hatte er einen Zusammenhang zwischen der Vorkommenshäufigkeit eines Wortes und der Anzahl von Wörtern, die diese Häufigkeit aufweisen, beobachtet:

Demnach sind es in einem Text relativ wenige Wörter, die häufig vorkommen, und es sind relative viele Wörter, die selten vorkommen. Diese Beobachtung über die Abnahme der Variabilität bei zunehmender Frequenz hatte Zipf mit der Annahme verbunden, dass dieses Wechselverhältnis gesetzmäßigen Charakters ist, und er hatte versucht, es mit einer relativ einfachen Gleichung mathematisch zu formulieren:

$$(1) a \cdot b^2 = k$$

In dieser Formel entspricht a der Anzahl von Wörtern mit einer bestimmten Vorkommenshäufigkeit (die üblicherweise eher mit n bezeichnet würde), b entspricht der konkreten Anzahl der Vorkommnisse (f_n), und k ist eine (für den gegebenen Text charakteristische) Konstante. Die beiden Komponenten a und b stellen somit Parameter dar, die von Text zu Text variieren können, auch wenn die Gleichung insgesamt sich nicht verändert (und auf mehr als einen Text zutreffen kann). Der genannten Formel zufolge würde sich also das Produkt der Vorkommenshäufigkeit eines bestimmten Wortes und der Summe seiner Vorkommenshäufigkeiten als eine Konstante darstellen. Seine Berechnungen hat Zipf u. a. an den Daten einer Untersuchung von Eldridge (1911) zur Worthäufigkeit im amerikanischen Zeitungsendlich veranschaulicht, die insgesamt ca. 44000 Wörter (ca. 6000 verschiedene Wörter) umfasste. In diesem Korpus kamen 2976 Wörter nur einmal, 1079 Wörter zweimal, 516 Wörter dreimal, usw. vor; ein Wort wie der Artikel 'the' hingegen kam 4290 mal vor. Aus Gründen der Anschaulichkeit hat Zipf bei der Überführung der Daten in eine Graphik die (beobachteten und theoretischen) Werte logarithmiert, so dass sich die postulierte Gesetzmäßigkeit in Form einer Geraden darstellt; Abb. 1 stellt das recht überzeugende Ergebnis anschaulich dar.

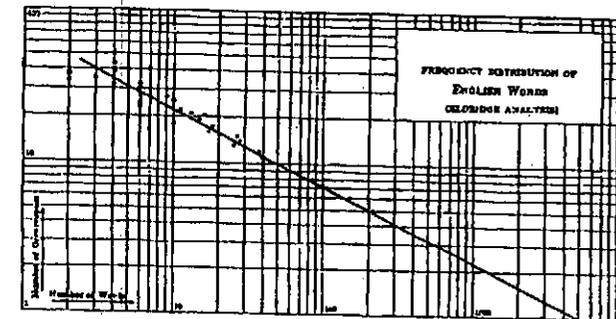


Abb. 1:
Rang-Häufigkeitsverteilung englischer Wörter
(Eldridge-Analyse) nach Zipf (1935: 46)

Bei der Verallgemeinerung seiner Beobachtungen aus dem Bereich der Phonetik sprach Zipf (1935: 129) zunächst noch recht allgemein von der Tendenz eines Systems, „die Bedingung eines Gleichgewichts aufrechtzuerhalten“, was die „Existenz eines zugrundeliegenden Gesetzes der Ökonomie der Kräfte“ [‘law of economy of effort’ bzw. ‘law of economy of energy’] nahe lege.

In seinem späteren Buch *Human Behavior and the Principle of Least Effort* von 1949 mit dem Untertitel *An Introduction to Human Ecology* arbeitete Zipf diese zunächst allgemein gehaltenen Überlegungen aus. Nun ordnete er seine vorherigen Beobachtungen in einen allgemeineren Rahmen ein, innerhalb dessen er zwischen der ‘Ökonomie des Sprechers’ einerseits und der ‘Ökonomie des Hörers’ andererseits differenzierte und beide zueinander in Beziehung setzte. Er postulierte dabei zwei grundsätzlich wirkende, antagonistische Kräfte, die er mit den Termini ‘Unifikation’ bzw. ‘Diversifikation’ bezeichnete. Zwischen diesen beiden widersprüchlichen „Interessen“, die Zipf (1949: 21) als „theoretical forces“ bezeichnete, müsse nun, so Zipf, eine Balance gefunden werden.

Zipf erläuterte seine Konzeption u. a. am Beispiel der Struktur des Vokabulars; hier lautete seine Argumentation etwa wie folgt: Während es für den Sprecher am „ökonomischsten“ sei, genau ein Wort für m verschiedene Bedeutungen zu haben, sei es für den Hörer am ökonomischsten, m verschiedene Wörter mit jeweils genau einer Bedeutung zu haben. Diese unterschiedlichen „Interessen“ lassen sich wie folgt formalisieren:

Ökonomie des Sprechers

$$\sum_{i=1}^m w = 1$$

vs.

Ökonomie des Hörers

$$\sum_{i=1}^m w = m$$

An dieser Stelle kommen wir genau zu dem oben angesprochenen Punkt zurück, dem zufolge maximale ‘Ökonomisierung’ auf der Produktionsseite womöglich eine rezeptionsseitige Reduktion der Ökonomie nach sich zieht: sinnvollerweise lässt sich nur dann von einer System-Ökonomie sprechen, wenn sich eine bestimmte Art von „Balance“ zwischen beiden ergibt. In diesem Zusammenhang darf eine wichtige Einschränkung allerdings nicht übersehen werden: Wie zu sehen ist, beschäftigte sich Zipf selbst im Rahmen seines Ansatzes in erster Linie mit Häufigkeiten von Elementen innerhalb eines Systems und mit Faktoren, welche die Häufigkeitsverteilung beeinflussen. Mit diesem Ansatz ist natürlich nur ein Teilaspekt der Ökonomie von Systemen zu erfassen: Das hängt insbesondere damit zusammen, dass wir uns, beim Beispiel sprachlicher Texte verbleibend, auf der paradigmatischen Ebene des Inventars, nicht auf der syntagmatischen (linearisierten) Ebene des realisierten Textes befinden. Außer Be-

tracht bleibt dabei also vorerst, dass natürlich auch in der spezifischen Kombination der Einheiten des Inventars „ökonomie-förderliche“ und „ökonomie-hinderliche“ Prinzipien zum Wirken kommen. Das wiederum bezieht sich nicht nur auf die Frage nach der Kombination von Elementen und deren Häufigkeiten, sondern auch nach der Frage der syntagmatischen Streuung bzw. möglicher lokaler Konzentrationen der Elemente.

Mit der Wahl des Zipf’schen Ansatz geht also zunächst eine bewusst in Kauf genommene methodologische Begrenzung einher, die hier nicht stillschweigend übergangen werden darf: Denn klarerweise haben wir es in kulturellen Systemen nicht einfach nur mit Häufigkeiten einzelner (isolierter) Elemente zu tun, sondern mit Elementen verschiedener Ebenen einschließlich der Kombination einzelner Elemente zu größeren Einheiten (auf einer „höheren“ Ebene) und deren Häufigkeiten. Als weitere Fragestellungen kommen sodann Wechselbeziehungen zwischen den Einheiten verschiedener Ebenen sowie die konkrete Verteilung der Einheiten in der Linearität des Textes hinzu, so dass die ursprüngliche Frage nach der Häufigkeit bestimmter Einheiten in einem Text in der Tat nur einen Teilaspekt innerhalb der gesamten Komplexität behandeln kann.

Genau dieser Teilaspekt soll die folgenden Überlegungen leiten, um die Relevanz der Zipf’schen Konzeption zu veranschaulichen – weiterführende Fragen wären an anderer Stelle aufzugreifen. Veranschaulichen wir aus diesem Grunde kurz – bevor wir uns der Mathematisierung durch Zipf detaillierter zuwenden – an einem trivialen Beispiel, an dem zu sehen ist, inwiefern die „simple“ Untersuchung von Vorkommenshäufigkeiten der Elemente eines Systems als Voraussetzung für die Untersuchung von dessen Ökonomie von Relevanz ist. Gehen wir beispielsweise davon aus, dass es in einem (z. B. sprachlichen oder musikalischen) Text mit insgesamt n Einheiten (z. B. mit 1000 Lauten oder 1000 Musiknoten) 25 verschiedene Einheiten gibt. Wenn diese 25 Einheiten im Hinblick auf die Häufigkeit ihres Vorkommens maximal variiert werden sollen, ergibt sich für jede Einheit ein genau 40-maliges Vorkommen. Eine derartige, in einer „Gleichverteilung“ resultierende maximale Variation der Vorkommenshäufigkeit der Einheiten ist jedoch in der konkreten (Text-)Praxis kaum jemals zu beobachten – vielmehr kommen üblicherweise die einen Elemente häufiger, andere hingegen seltener vor (weswegen auch ihr Informationsgehalt unterschiedlich ist). Und genau dies ist die Frage, die Zipf am Beispiel lexikalischer Häufigkeiten detailliert analysiert und vor dem Hintergrund antagonistischer „System-Interessen“ zu formalisieren versucht hat – die Frage nämlich, ob sich das Verhältnis der Häufigkeiten, in dem die Einheiten x_1, x_2, \dots, x_n zueinander stehen, gesetzmäßig im Sinne einer System-Ökonomie erfassen lässt.

Die Tragweite dieser Argumentation, so trivial sie zu sein scheint, ist nicht zu unterschätzen: Es wird nämlich nicht mehr und nicht weniger gesagt, als dass sich die Ökonomie eines Systems als Ergebnis externer antagonistischer Kräfte

– im gegebenen Fall als Ergebnis der „ökonomischen“ Interessen der am Kommunikationsprozess Beteiligten – herausstellt. Was auch in der Systemtheorie à la Luhmann später als wesentlich herausgearbeitet wurde, ist hier also bereits immanenter Bestandteil der Konzeption: Jedwede Beschäftigung mit der „Ökonomie des Systems“ muss davon ausgehen, dass die Bedingungen, unter denen diese zustande kommt, nicht (allein) in den Spezifika des Systems selbst begründet sein müssen. Insofern liegt es nahe, die ursprünglichen Zipf'schen Kräfte aus gegenwärtiger Sicht umzuinterpretieren – insbesondere aus systemtheoretischer bzw. synergetischer Perspektive (vgl. Altmann/Köhler 1995; Prün 1999).

Beim Versuch der Formalisierung seiner Beobachtungen bezog Zipf (1949) im Gegensatz zu seiner früheren Herangehensweise zwar nach wie vor die absolute Vorkommenshäufigkeit der Elemente in seine Formel ein, in Ergänzung dazu aber nun den Rang (r), den ein Wort mit einer bestimmten Häufigkeit einnimmt. Dies führte zu der (ebenfalls noch recht einfachen) Formel

$$(2) r \cdot f = k.$$

Diesem Ansatz zufolge erweist sich nunmehr das Produkt der absoluten Vorkommenshäufigkeit eines Wortes und seines Ranges als eine Konstante. Veranschaulichen wir auch das Ergebnis ebenfalls an einem Beispiel, und zwar an den oben bereits genannten und dargestellten Daten aus dem Eldridge-Korpus. Die Berechnung nach Formel (2) ergibt nach bilogarithmischer Transformation (s. o.) die Abb. 2.

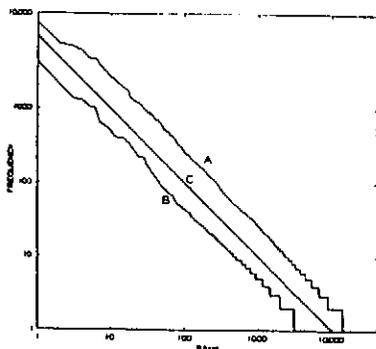


Abb. 2: Rang-Häufigkeitsverteilung von Wörtern nach Zipf (1949: 25).
(A) James Joyce; (B) Eldridge; (C) „Idealkurve“

Wie zu sehen ist, führen die Berechnungen sowohl nach (1) als auch nach (2) zu recht passablen Anpassungen; dies gilt allerdings, wie auch schon Zipf (1935) einräumte, nur in eingeschränktem Umfang für die besonders häufig (und, wie

im Anschluss an Zipf später von anderen Forschern bemängelt wurde) für die besonders selten vorkommenden Wörter. Aus diesem Grunde ist der Zipf'sche Ansatz verschiedentlich modifiziert, erweitert und in allgemeinere Verteilungsmodelle überführt worden.

3.2. Von Zipf zu allgemeinen Verteilungen

Die Vorschläge von Zipf sind bis auf den heutigen Tag in der quantitativen Linguistik immer wieder diskutiert und mitunter derart spezifisch ausgearbeitet worden, dass sie den Eindruck einer eigenen Wissenschaft erwecken, zu deren Verständnis ein mathematisches Grundstudium notwendig ist (vgl. Guiter/Arapov 1982, u. a.). Auf Sprache und sprachliche Texte bezogen, dürfte eine der wichtigsten und bekanntesten Ergänzungen wohl diejenige von Benoît Mandelbrot (1953, 1954) gewesen sein, der an die (auch von Zipf konzidierte) Beobachtung anknüpfte, dass die von ihm beschriebene Gesetzmäßigkeit wohl für den „mittleren“ Bereich des Vokabulars zutrifft, nicht aber für die extremen (d. h. die besonders häufigen und die besonders seltenen) Werte – und genau das zeigen ja auch die obigen Graphiken. Mandelbrot ging zunächst von der einfachen Zipf'schen Formel (2) $r \cdot f = C$ aus, die sich in geringfügiger Umformulierung als (2a) $f = C/r$ lesen lässt. Wenn man nun für die absolute Frequenz f die Wahrscheinlichkeit P sowie für den Rang r die absolute Vorkommenshäufigkeit n in die Formel einführt, ergibt sich die geringfügige Modifikation

$$(2b) P_i = \frac{C}{n}$$

Diese Gleichung erweiterte Mandelbrot um die beiden Parameter a und b , so dass sich im Endergebnis die auch als „Zipf-Mandelbrot'sches Gesetz“ (im Folgenden: ZMG) bekannte Formel ergibt:⁴

$$(3) P_i = \frac{C}{(b+i)^a} \quad i = 1, 2, 3, \dots, n$$

Bei genauerem Hinsehen zeigt sich, dass die ursprüngliche Zipf'sche Formulierung (2b) sich als Spezialfall der ZM-Formel (3) darstellt, für den Fall nämlich, dass $b = 0$ und $a = 1$.

⁴ Der Vollständigkeit halber sei erwähnt, dass sich die Konstante C hier aus den vorhandenen Parametern bestimmen lässt, nämlich als

$$(3') C = \sum_{i=1}^n (b+i)^a$$

Wir haben es hier nach wie vor mit einer relativ einfachen Formel zu tun; da wir im weiteren Verlauf unserer Überlegungen noch auf komplexere Formeln stoßen werden, sei das Prinzip, um das es geht, an dieser Stelle kurz etwas detaillierter erläutert. Ausgegangen waren wir mit Zipf von der einfachen Gleichung (2) $r \cdot f = C$. Diese lässt sich als eine Funktion verstehen: Wenn man hier z. B. die Ränge 1, 2, 3, ... n als Variable x definiert, ergibt sich nämlich die Funktion:

$$(2c) f(x) = \frac{C}{f}$$

In der oben angeführten ZM-Formel (3) ist diese Funktion bereits in den Status einer Wahrscheinlichkeitsfunktion überführt; Voraussetzung dafür ist, dass sich die Summe aller Wahrscheinlichkeiten auf 1 beläuft. In diesem Sinne lässt sich die ZM-Formel auch wie folgt schreiben:

$$(3a) P(x) = \frac{C}{(b+x)^a} \quad x = 1, 2, 3, \dots, n$$

Auch hier gilt, dass die Parameter a und b bei einzelnen konkreten Texten variieren können; dadurch bleibt das Modell (wenn es denn passt) insgesamt unverändert, auch wenn sich der Wert für $P(x)$ je nach Variierung der Parameter jeweils ändern kann. C erweist sich innerhalb der Wahrscheinlichkeitsfunktion als eine Normierungskonstante, die so zu bestimmen ist, dass sich die Summe aller Wahrscheinlichkeiten auf 1 beläuft.

Es ergibt sich damit insgesamt ein Verteilungsmodell, in dem sich für jedes $P(x)$ die theoretische Häufigkeit berechnen lässt. So lassen sich Textstrukturen gegebenenfalls einem gemeinsamen Typ oder Modell zuordnen (auf ein gemeinsames Gesetz zurückführen). Freilich gibt es sehr viele verschiedene Verteilungsmodelle, und es kommt darauf an herauszufinden, welches dieser Verteilungsmodelle sich am besten an die beobachteten Daten anpassen lässt. Damit ist gemeint, dass für jedes der Verteilungsmodelle die Parameter so variiert werden, dass sie jeweils am besten passen, und dass man daran anschließend prüft, welches der Verteilungsmodelle insgesamt am besten passt.⁵

Verschiedene dieser Verteilungen lassen sich aufgrund mathematischer Ähnlichkeiten in so genannten „Familien“ zusammenfassen. So haben Zörnig und Altmann (1995) zum Beispiel dafür argumentiert, dass sich sowohl das ur-

⁵ Gegenwärtig sind ca. 1000 verschiedene Verteilungsmodelle bekannt (natürlich nicht nur im Hinblick auf Sprache, sondern aus den verschiedensten Bereichen); etwa 700 von ihnen sind in einem Thesaurus erfasst (Wimmer/Altmann 1999); für etwa 200 von ihnen gibt es Spezialsoftware, mit der sich die Anpassungen iterativ automatisch berechnen lassen (Altmann-Fitter, 1995).

sprüngliche Zipf'sche Gesetz als auch das ZMG beide als Spezialfälle eines noch allgemeineren Gesetzes angesehen (und einer gemeinsamen „Familie“ zugeordnet) werden können. Diese Verteilungsfamilie haben sie als „Lerch-Verteilung“ bezeichnet; sie lässt sich wie folgt formalisieren:

$$(4) P(x) = \frac{P^x}{(b+x)^a}$$

In der Tat ergeben sich aus dieser allgemeinen Formel verschiedene Verteilungsfunktionen, wenn die Parameter (a bzw. b) bestimmte Grenzwerte einnehmen; die Tatsache, dass sich gegebenenfalls mehrere Verteilungen einer gemeinsamen „Familie“ zuordnen lassen, kann man also so interpretieren, dass die betreffenden Verteilungsmodelle bestimmte Eigenschaften teilen, und dass sie unter bestimmten Rand- oder Extrembedingungen (z. B. wenn ein Parameter gegen 0, oder 1, oder auch gegen ∞ tendiert) mit einem anderen Verteilungsmodell (das dann weniger Parameter aufweist) konvergieren.

Es seien der Anschaulichkeit halber zwei dieser Verteilungen angeführt. So ergibt sich für den Fall, dass $p = 1$, $b > 0$, $a > 1$, die oben dargestellte Zipf-Mandelbrot-Verteilung (ZMV):

$$(3a') P(x) = \frac{1}{(b+x)^a}$$

Und für den Fall, dass $p = 1$, $b > 0$, $a = 1$, ergibt sich die so genannte Good-Verteilung; diese sei deswegen hier erwähnt, weil wir ihr im nächsten Abschnitt noch begegnen werden:

$$(5) P(x) = \frac{1}{(b+x)}$$

Wie zu sehen ist, lässt sich die zuletzt genannte Good-Verteilung (4) somit entweder als Spezialfall der ZMV verstehen (für den Fall, dass $a = 1$) oder auch, nach Zörnig/Altmann, als Spezialfall der noch allgemeineren Lerch-Verteilung (für den Fall, dass $p = 1$, $b > 0$, $a = 1$).

Lassen wir damit die mathematischen Überlegungen auf sich beruhen und fassen statt dessen die bisherigen Überlegungen zusammen: Es kann festgehalten werden, dass sich mit der Untersuchung der Häufigkeiten (bzw. der Häufigkeitsverteilung) von Elementen einerseits individuelle Texte ein und desselben Zeichensystems (sprachliche Texte, Texte einer gegebenen Sprache, Texte eines bestimmten Autors, usw.) auf Gemeinsamkeiten hin untersuchen lassen; andererseits lassen sich strukturelle Analogien zwischen verschiedenen (künstlerischen und nicht-künstlerischen) Zeichensystemen untersuchen. Das wiederum eröffnet die Möglichkeit, Zeichensystem-übergreifende Prinzipien der Informa-

tionsorganisation im Hinblick auf eine ihnen zugrundeliegende System-Ökonomie zu studieren.

Dabei ist natürlich die Frage der zu untersuchenden Ebene der Zeichensysteme von besonderer Relevanz: In den obigen Überlegungen ging es im Zusammenhang mit dem Zipf'schen Ansatz ja ausschließlich um Sprache, und dabei sogar nur um die Organisation lexikalischer Häufigkeiten. Aber natürlich ist innerhalb des Systems eines sprachlichen Textes die lexikalische Ebene nur eine von mehreren, neben den Ebenen der Graphemik/Phonetik; der Morphologie/Syllabik, u. a. m. Und man kann mittlerweile davon ausgehen, dass die Einheiten jeder dieser Ebenen nicht demselben Verteilungsmodell wie der für die Lexeme gültigen *ZMV* folgen.

Im Folgenden sollen unter Berücksichtigung dieses Umstands sowie vor dem allgemeinen Hintergrund der Zipf'schen Überlegungen zur Unifikation und Diversifikation die Relevanz von Häufigkeitsverteilungen in Zeichensystemkonstituenten aus den Bereichen der Sprache (4), der Musik (5) und der Malerei (6) analysiert werden. Die zur Analyse notwendigen Daten sind – insbesondere im Hinblick auf Musik und Malerei – in den entsprechenden Studien leider nur unvollständig publiziert, so dass echte Analysen bzw. Re-Analysen kaum möglich sind. Genau dies ist einer der Gründe, warum die folgenden Überlegungen eher programmatischen Charakters bleiben müssen.

4. System-ökonomische Gesetzmäßigkeiten sprachlicher Texte (am Beispiel von Puškins *Царь Салтан*)

Im Folgenden sollen zwei ausgewählte Gesetzmäßigkeiten der Häufigkeitsverteilung von Elementen verschiedener Ebenen eines sprachlichen Textes am Beispiel von Puškins Versmärchen *Царь Салтан* demonstriert werden. In beiden Fällen handelt es sich um Ranghäufigkeiten (Wort- und Graphemranghäufigkeiten), anhand derer nur einige ausgewählte Fragen behandelt werden können, so dass andere Häufigkeitsregularitäten in Texten, vor allem auch sprach-, autoren-, gattungsspezifische o. ä. Aspekte unberücksichtigt bleiben müssen (vgl. hierzu Grzybek 2001a, b). Auch detaillierte mathematische Erläuterungen der einzelnen Verteilungsmodelle sollen nicht ausführlich besprochen werden, da es hier in erster Linie um den Nachweis geht, dass solche Häufigkeiten sich überhaupt (auch in künstlerischen Texten) manifestieren, und dass sich auf den verschiedenen Ebenen jeweils eigene Regularitäten nachweisen lassen.

4.1. Wortranghäufigkeit

Beginnen wir mit der Untersuchung der Wortranghäufigkeit. Von den insgesamt 3989 Wörtern des Textes kommt das häufigste 153mal, das zweithäufigste

133mal, das dritthäufigste 75mal, usw. vor; die 3989 tokens verteilen sich auf 1302 types (TTR = 0.33). Tab. 1 enthält in den Spalten 2 und 3 für die ersten 30 Ränge (X[i]) – diese entsprechen ca. 90% aller vorkommenden Werte – die absoluten (F[i]) und prozentualen (F[i]%) Häufigkeiten; in den Spalten 4 und 5 die nach *ZM* (3) berechneten theoretischen Werte, ebenfalls absolut (NP[i]) und prozentual (NP[i]%) ; Abb. 3 enthält die entsprechende graphische Darstellung für die ersten 130 Ränge.

Tab. 1: Wortranghäufigkeit in Puškins „Царь Салтан“

X[i]	F[i]	F[i]%	NP[i]	NP[i]%	X[i]	F[i]	F[i]%	NP[i]	NP[i]%
1	153	11,75	142,01	10,91	21	23	1,77	22,19	1,70
2	133	10,22	105,70	8,12	22	22	1,69	21,41	1,64
3	75	5,76	85,13	6,54	23	21	1,61	20,69	1,59
4	70	5,38	71,76	5,51	24	21	1,61	20,02	1,54
5	60	4,61	62,31	4,79	25	20	1,54	19,40	1,49
6	57	4,38	55,24	4,24	26	20	1,54	18,81	1,44
7	50	3,84	49,74	3,82	27	18	1,38	18,27	1,40
8	41	3,15	45,33	3,48	28	16	1,23	17,76	1,36
9	36	2,76	41,70	3,20	29	14	1,08	17,28	1,33
10	33	2,53	38,66	2,97	30	14	1,08	16,83	1,29
11	30	2,30	36,07	2,77			a 0,8162	χ^2	92,41
12	28	2,15	33,84	2,60			b 1,2938	FG	1143
13	27	2,07	31,90	2,45			n 1302	P(χ^2)	1,00
14	25	1,92	30,18	2,32				C	0,0232
15	25	1,92	28,66	2,20					
16	25	1,92	27,30	2,10					
17	24	1,84	26,08	2,00					
18	24	1,84	24,97	1,92					
19	24	1,84	23,96	1,84					
20	23	1,77	23,04	1,77					

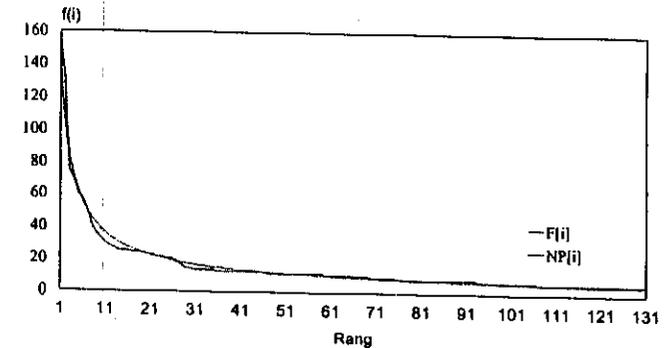


Abb. 3: Wortranghäufigkeit in Puškins „Царь Салтан“

4.2. Graphemranghäufigkeit

Ähnlich wie bei Wortranghäufigkeiten geht es auch bei Graphemranghäufigkeiten nicht um die Analyse der Vorkommenshäufigkeit spezifischer Grapheme, sondern um die gesetzmäßige Relation zwischen dem häufigsten, zweithäufigsten, dritthäufigsten usw. Graphem innerhalb eines Textes. Diese Relation lässt sich in der allgemeinsten Form als Proportionalitätsrelation der Art $P_x \sim P_{x-1}$ bzw. spezifisch als Funktion $P_x = g(x)P_{x-1}$ verstehen. Dieser Ansatz führte nach Formel (3) zur ZMV; für die Modellierung der Graphemranghäufigkeit in Puškins *Царь Салтан* (wie in vielen anderen Texten auch) hingegen erweist sich eine andere, die so genannte *negative hypergeometrische Verteilung*, als geeignet. Diese steht in keiner unmittelbaren Beziehung zur Zipf'schen oder Lerch'schen Verteilung; das heißt freilich nicht, dass die Zipf'schen Kräfte der Unifikation und Diversifikation hier nicht als Interpretationsmaßstab herangezogen werden könnten (zu einer entsprechenden Interpretation der Parameter s. Wimmer/Wimmerová); das heißt lediglich, dass die genannten Verteilungsmodelle in keiner direkten mathematischen Beziehung miteinander stehen. So ergibt sich aufgrund der Differenzgleichung

$$(6) P_x = \frac{b+x}{x} \frac{c-x}{a-x} P_{x-1}$$

die negative hypergeometrische Verteilung:

$$(7) P_x = \frac{(M+x-1)}{x} \frac{(n+1-x)}{(K-M+n-x)} P_{x-1}$$

Die Anpassungsergebnisse der negativen hypergeometrischen Verteilung an die in eine Rangreihenfolge gebrachten Graphemhäufigkeiten des *Царь Салтан* sind der Tab. 2 bzw. Abb. 4 zu entnehmen.

Tab. 2:
Graphemranghäufigkeit in Puškins „Царь Салтан“

X[i]	F[i]	F[i]%	NP[i]	NP[i]%	X[i]	F[i]	F[i]%	NP[i]	NP[i]%
1	1876	10,21	1937,65	10,54	21	306	1,66	284,215	1,66
2	1519	8,26	1559,26	8,48	22	274	1,49	251,013	1,49
3	1475	8,02	1372,32	7,47	23	219	1,19	219,353	1,19
4	1345	7,32	1240,94	6,75	24	204	1,11	189,231	1,11
5	1174	6,39	1136,12	6,18	25	187	1,02	160,656	1,02
6	949	5,16	1047,03	5,70	26	168	0,91	133,658	0,91
7	942	5,12	968,452	5,27	27	151	0,82	108,288	0,82
8	904	4,92	897,492	4,88	28	96	0,52	84,6265	0,52
9	881	4,79	832,373	4,53	29	37	0,20	62,7921	0,20
10	867	4,72	771,93	4,20	30	13	0,07	42,9662	0,07
11	678	3,69	715,355	3,89	31	13	0,07	25,438	0,07
12	585	3,18	662,072	3,60	32	3	0,02	10,7227	0,02
13	570	3,10	611,65	3,33			K 3,2001	χ^2 150,47	
14	532	2,89	563,766	3,07			M 0,8400	FG 28	
15	459	2,50	518,171	2,82			n 31	$P(\chi^2) < 0,001$	
16	439	2,39	474,668	2,58			C 0,0082		
17	438	2,38	433,105	2,36					
18	397	2,16	393,362	2,14					
19	369	2,01	355,344	1,93					
20	313	1,70	318,98	1,74					

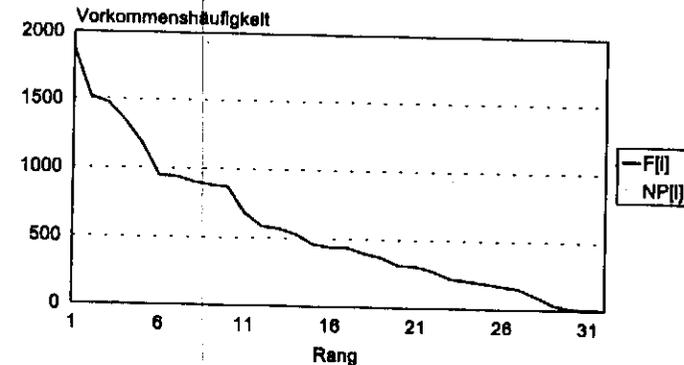


Abb. 4: Graphemranghäufigkeit in Puškins „Царь Салтан“

5. Musik

Um die Untersuchungen aus dem Bereich der Sprache mit solchen aus dem Bereich der Musik vergleichbar machen zu können, gilt es als Erstes, die Frage der 'Wort-Äquivalenz' in der Musik zu klären. Boroda (1982) hat sich explizit mit der Frage von Elementwiederholungen in musikalischen Texten auf der melodischen Ebene auseinander gesetzt. Bei der Wahl einer geeigneten Elementareinheit ergibt sich ihm zufolge folgendes Problem: Wählt man als Elementareinheit eine Einheit mit fixierter Länge (ein melodisches Intervall, eine Aufeinanderfolge von Intervallen, o. ä.), dann führt das zu einer „unnatürlichen“ Segmentierung des Textes, vergleichbar etwa mit der Zerstückelung eines Textes in Segmente von jeweils n Buchstaben Länge. Andererseits sind Einheiten mit variabler Länge (wie z. B. ein Motiv oder Submotiv) in der Musikwissenschaft nicht eindeutig definiert.

Aus diesem Grund hat Boroda (1982) als Elementareinheit das so genannte „formale Motiv“ (*F*-Motiv) vorgeschlagen; als ein solches versteht er ein Segment der melodischen Linie, das innerhalb einer der vier folgenden metrisch-rhythmischen Gruppierungen einzugrenzen ist (vgl. Abb. 5a, b):

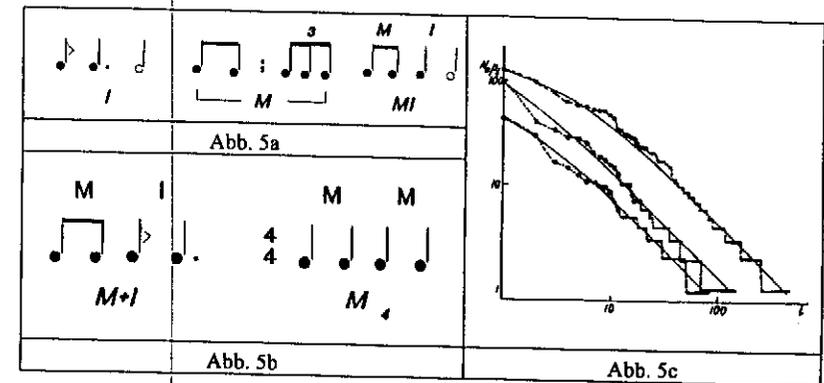
1. eine Abfolge (*I*) dar, in der jeder Folgeton länger ist als der bzw. die vorausgehende(n);
2. eine Takt-Struktur (*M*) von zwei bzw. drei Tönen, von denen der erste Ton metrisch stärker als der bzw. die folgende(n) ist;
3. eine „synthetische“ Struktur (*MI*), in der eine Abfolge *I* mit einem Endton von *M* beginnt.

Auf dieser Grundlage hat Boroda verschiedenen (homophonen und polyphonen) Musiktexten aus dem Stil-Spektrum des 18. bis 20. Jahrhunderts folgende Charakteristika zugewiesen:

- a. die Summe aller vorkommenden *F*-Motive (in Entsprechung zur 'Textlänge');
- b. die Anzahl aller nach einem bestimmten Kriterium distinktiven *F*-Motive (entsprechend einem 'Motivinventar');
- c. die Vorkommenshäufigkeiten eines jeden distinktiven *F*-Motivs (entsprechend der Häufigkeitsstruktur des Textes)

Wie oben bereits gesagt wurde, gibt es mittlerweile für die Güte der Anpassung verschiedener Verteilungsmodelle an Daten Spezialsoftware; Boroda freilich ist einen anderen Weg gegangen, nämlich den der argumentativen Analogie zur *ZMV*, wie sie für das lexikalische Inventar sprachlicher Texte charakteristisch ist (s. o.). Dabei hat er den Parameter a der Gleichung (3a) – der Mandelbrot zufolge zwar > 1 ist, stets aber gegen 1 tendiert – gleich 1 gesetzt. Dies führt, wie oben bereits dargestellt wurde, zur sog. Good-Verteilung (5) als Spezialfall

der *ZMV*. Anpassungsergebnisse für Werke von Chopin, Bach und Beethoven hat Boroda (1982: 244) graphisch wie in Abb. 5c dargestellt.



Borodas (1982: 247) eigene Schlussfolgerungen lauten dahingehend, dass in den von ihm untersuchten musikalischen Texten die rekursive Organisation von *F*-Motiven allgemein gültigen Abhängigkeitsverhältnissen unterliegt, die mit dem *ZMG* beschrieben werden können.

Die graphische Darstellung (5c) ist auf den ersten (subjektiven) Blick ohne Frage überzeugend; dennoch stellen sich aufgrund des Vorgehens von Boroda eine Reihe wichtiger Probleme. Das Hauptproblem liegt darin, dass in den Arbeiten keine Rohdaten angegeben werden, so dass sich die Ergebnisse nicht re-analysieren lassen; es lässt sich de facto nicht einmal etwas über die Güte der Anpassungen an die *ZMV* bzw. an die Good'sche Verteilung sagen bzw. darüber, ob (eine) andere als diese Verteilungen an die Daten zu anderen (besseren?, noch besseren?) Ergebnissen führen würde(n). Eine Lösung dieser Probleme wäre insbesondere für zeichensystem-übergreifende Schlussfolgerungen absolute Voraussetzung; sollten sich die Beobachtungen von Boroda dabei bestätigen, wäre dies ein wichtiger Hinweis auf eine Analogie zwischen der lexikalischen Ebene der Sprache und der *F*-Motiv-Struktur musikalischer Texte. So aber lassen die Untersuchungen von Boroda de facto nicht mehr (und nicht weniger) als die Formulierung interessanter Hypothesen zu.

Während Boroda, wie zu sehen war, den linguistisch geebneten Weg der „Argumentation per analogiam“ gegangen ist, haben Wimmer/Wimmerová (Ms.) einen anderen Weg gewählt, den sie als „Weg der Konstruktion“ bezeichnen: Als Datenmaterial dienten hierbei Werke von Bach, Beethoven, Liszt und Chopin, von denen die Häufigkeiten, mit denen Töne einer gegebenen Tonhöhe vorkommen, berechnet wurden. An die in eine Rangreihenfolge gebrachten Häufigkeitsdaten wurden iterativ zahlreiche Verteilungsmodelle angepasst und die jeweilige Anpassungsgüte getestet; im Ergebnis stellte sich heraus, dass sich

bei diesen Daten insgesamt die oben bereits bei den Graphemranghäufigkeiten dargestellte *negative hypergeometrische Verteilung* als geeignetes Modell erwies (vgl. auch Wimmer/Altmann 2001). Dieser Befund ist insofern von besonderem Interesse, weil auch Köhler/Martináková-Rendeková (1998) bei der Untersuchung der Häufigkeiten von Tonhöhe und Tonstärke einer Chopin-Étude zu demselben Ergebnis gekommen waren. Hier wurde die Étude op. 25, no. 11 analysiert; Abb. 6 stellt die empirischen und theoretischen Werte in anschaulicher Form dar ($\chi^2_{DF=70} = 25.55$; $P(\chi^2) = 0.99$; $C = 0.008$).

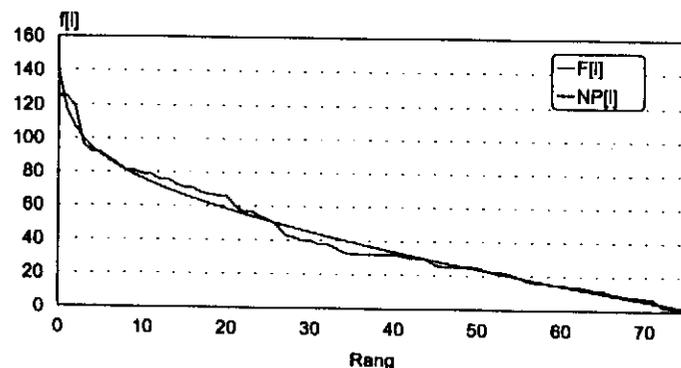


Abb. 6: Tonranghäufigkeit in Chopin (Étude 25, op. 25, no. 11); nach Köhler/Martináková-Rendeková (1998)

Die Divergenz der Ergebnisse von Wimmer/Wimmerová und Köhler/Martináková-Rendeková (1998) einerseits und der Beobachtungen von Boroda andererseits wirft die entscheidende Frage danach auf, wie die Unterschiedlichkeit der Ergebnisse zu erklären sein könnte; es bieten sich als Interpretation zwei verschiedene Optionen an, die sich nicht zwangsläufig einander ausschließen, über deren Relevanz aufgrund der (publizierten) Datenlage derzeit aber auch nicht entschieden werden kann:

1. die Divergenz der Ergebnisse kann in der sprach-analogie-bedingten Argumentation von Boroda begründet sein – das würde bedeuten, dass sich gegebenenfalls eine andere als die *ZMV* bzw. an die Good'sche Verteilung, deren Güte von Boroda nicht getestet wurde, zur Modellierung der entsprechenden Daten eignet;
2. in den Untersuchungen von Wimmer/Wimmerová und Köhler/Martináková-Rendeková (1998) wurde schlichtweg eine andere Ebene der musikalischen Textorganisation untersucht als in den Texten von Boroda.

Sollten sich jedenfalls die bislang vorliegenden Ergebnisse und Hypothesen bestätigen, ergäbe sich eine interessante Analogie: Wenn sich nämlich in (bestimmten?) sprachlichen Texten die Häufigkeitsverteilung von Graphemen mit der negativen hypergeometrischen und die Wortranghäufigkeit mit der *ZMV* modellieren lässt, und wenn sich in (bestimmten?) musikalischen Texten die Häufigkeitsverteilung von Tonhöhen bzw. die Häufigkeitsverteilung von *F*-Motiven mit eben diesen beiden Verteilungsmodellen theoretisch erfassen lassen, so ergäbe sich in der Tat eine Analogie zwischen Graphemen ~ Tonhöhen einerseits und Lexemen ~ *F*-Motiven andererseits.

Die Feststellung einer solchen Analogie wäre natürlich für Zeichensystem-übergreifende Untersuchungen der Informationsorganisation semiotischer Systeme von außerordentlicher Relevanz. Doch wie gesagt: Solange (für den Bereich der Musik) keine solideren Daten vorliegen (bzw. veröffentlicht werden), lässt sich eine solche Vermutung bestenfalls im Sinne von Spekulationen formulieren. Genau dies gilt auch für Texte aus der Malerei, auf die abschließend eingegangen werden soll – hier stellt sich die (Daten)-Lage de facto freilich als noch schlechter dar.

6. Malerei

Einen Versuch, den auf die oben beschriebene Art erweiterte *ZMV* (Good'sche Verteilung) ebenfalls in Analogie zur Sprache auf die Malerei anzuwenden, findet sich bei Vološin/Orlov (1982). Die Autoren haben die nach abnehmender Häufigkeit geordnete Reihenfolge der Flächen eines Bildes von I. Levitan, „Ewige Ruhe“ [Над вечным покоем] (1893/94), untersucht. Die 'Fläche', die von einer bestimmten Farbe eingenommen wird, entspricht dabei der Summe aller Teilflächen eben dieser Farbe.

Auch in dieser Studie werden keine Rohdaten angeführt, was dieselben Schwierigkeiten mit sich bringt bzw. nach sich zieht wie die Studien von Boroda zur Musik. Die Autoren postulieren aufgrund ihrer Untersuchung jedenfalls, dass die nach abnehmender Häufigkeit geordnete Reihenfolge der Farbflächenanteile (auch 'Farbreihe' genannt) eben der Formel folgt, die sich aus dem verallgemeinerten *ZMG* ergibt (womit sie die Good'sche Verteilung meinen). Einschränkend müssen die Autoren ihren Beobachtungen allerdings hinzufügen, dass dies „nur für den Anfangsteil“ der Farbreihe gilt. Das stellt die Güte des (Anpassungs-)Ergebnisses insgesamt natürlich vollkommen in Frage – und da keinerlei Rohdaten angeführt werden, ist eine Re-Analyse nicht möglich.



Abb. 7a: „Ewige Ruhe“ von I. Levitan

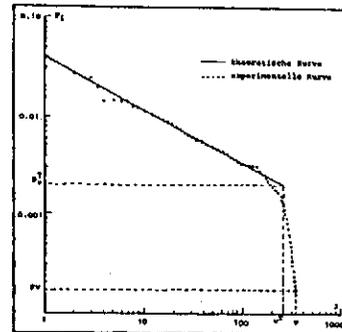
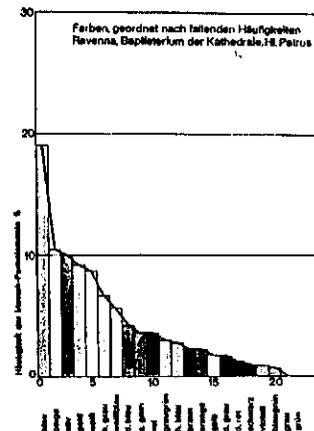


Abb. 7b: Anpassung an ZMV

Ebenfalls nicht möglich ist die Re-Analyse einer Untersuchung, die Fucks bereits in den 60er Jahren durchgeführt hat. Er hat einen Mosaikausschnitt des Mosaiks vom Apostel Petrus aus dem Baptisterium der Kathedrale in Ravenna analysiert (vgl. Abb. 8a). Für die „insgesamt rund 5000 Mosaik-elemente“ (Fucks 1968: 30) hat er 20 verschiedene Farbtöne differenziert und deren Vorkommenshäufigkeit ausgezählt; so kommt seinen Angaben nach der am stärksten vertretene Blauton exakt 955 (19%) mal vor, die absoluten Häufigkeiten der übrigen Farbtöne werden nicht angeführt, sondern nur graphisch in Form der Abb. 8b angeführt:

Abb. 8a:
Mosaik des Apostel Petrus (5. Jhd.)Abb. 8b:
„Rangordnung“ der Farben
nach Fucks (1968)

Wie gesagt, ist aufgrund der fehlenden Rohdaten eine Re-Analyse nicht möglich; wenn man ungeachtet dessen – in vollem Bewusstsein der gefährlichen Naivität dieses Vorgehens – aufgrund der Abb. 8a die absoluten Vorkommenshäufigkeiten zu rekonstruieren und an diese ein theoretisches Modell anzupassen versucht, stellt sich abermals die negative hypergeometrische Verteilung als geeignetes Modell dar ($\chi^2_{DF=16} = 84.97$; $P(\chi^2) < 0.001$; $C = 0.017$). Abb. 9 veranschaulicht die empirischen und theoretischen Werte:

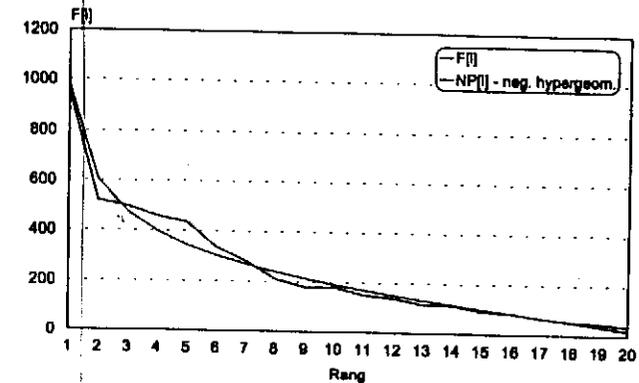


Abb. 9: Rekonstruierte Ranghäufigkeitsverteilung der Farbsteine nach Fucks (1968)

Sollten sich auch diese Befunde – deren Validität hiermit nochmals nachdrücklich in Frage gestellt sei – bestätigen, hätten wir es nicht nur im Bereich von Sprache und Musik, sondern auch bei der Malerei mit jeweils zwei identischen Verteilungsmodellen zu tun, die unterschiedliche (womöglich miteinander vergleichbare) Ebenen der Organisation (hier: Farbflächen und Anzahl der jeweils x -farbigen Mosaiksteine unabhängig von ihrer Größe) betreffen.

7. Zusammenfassung

Wie aus den oben Überlegungen und Darstellung deutlich hervorgeht, steht die Untersuchung der System-Ökonomie kultureller Texte ganz am Anfang der Erforschung – und das, obwohl sich grundlegende theoretische Annahmen wie diejenigen von Zipf über die Kräfte der Unifikation und Diversifikation bereits in den 30er und 40er Jahren finden, und obwohl diese Annahmen sich problemlos in modernere (systemtheoretische bzw. synergetische) Konzeptionen einbinden ließen.

In den vorangegangenen Überlegungen ging es allgemein um Fragen der ökonomischen Organisation von (künstlerischen und nicht-künstlerischen) Tex-

ten, die vor dem Hintergrund der Zipf'schen Kräfte der Diversifikation und Unifikation betrachtet werden sollte. Die Wirksamkeit dieser Kräfte sollte an einer spezifischen Fragestellung, der Häufigkeitsverteilung bestimmter Elemente verschiedener Zeichensysteme aus dem Bereich der Sprache, der Musik und der Malerei nachgewiesen werden. Aufgrund fehlender Daten insbesondere zu den nicht-sprachlichen Texten sind solide Rückschlüsse schlicht und einfach nicht zulässig; ungeachtet dessen haben sich über die verschiedenen Zeichensysteme hinweg auf verschiedenen Ebenen ihrer inneren Organisation zwei Verteilungsmodelle als relevant erweisen: die sog. Zipf-Mandelbrot'sche (z. T. in Form der sog. Good'schen) und die negative hypergeometrische Verteilung. Inwiefern diese beiden Verteilungsmodelle in der Tat bei der Organisation der Zeichensysteme eine besondere Rolle spielen und ob sich jeweils entsprechende Ebenen dieser Zeichensysteme mit ihnen erfassen lassen, ist eine Frage, die nur in der Zukunft seriös zu beantworten sein wird.

Ungeachtet scheinbar widersprüchlicher Ergebnisse in Detailfragen lassen die obigen Ausführungen jedenfalls eine Reihe von (zugegebenermaßen teils spekulativen) Hypothesen zu:

1. die Häufigkeitsverteilung von Elementen in sprachlichen Texten sowie in Texten der Musik und Malerei ist nicht zufällig, sondern gesetzmäßig organisiert;
2. die Gesetzmäßigkeit der Organisation ist am besten in sprachlichen Texten dokumentiert;
3. zwischen künstlerischen und nicht-künstlerischen Texten sind keine prinzipiellen Unterschiede zu beobachten, diese scheinen sich vielmehr auf andere Art und Weise zu äußern;
4. in Abhängigkeit von der gewählten Größe (Ebene) der Elemente unterscheiden sich die Häufigkeitsmodelle.

Literatur

- Altmann, G.; Koch, W.A. (eds.) 1998. *Systems. New Paradigms for the Human Sciences*. Berlin/New York.
- Boretzky, N.; Enninger, W.; Stolz, Th. (Hrsg.) 1990. *Spielarten der Natürlichkeit – Spielarten der Ökonomie*. Bochum.
- Boroda, M.G. 1977. „Häufigkeitsstrukturen musikalischer Texte“, in: Orlov et al. (Hrsg.) (1982), 231–262.
- 1998. „Systemic organization and the development of the European musical language“, in: Altmann/Koch (eds.) (1998), 479–513.
- Cardauns, B. 1985. „Zum Begriff der 'oeconomia' in der lateinischen Rhetorik und Dichtungskritik“, in: Stemmler (Hrsg.) (1985), 9–18.

- Eldridge, R.C. 1911. *Six Thousand Common English Words*. Buffalo.
- Fenk-Oczlon, G. 1990. „Ökonomieprinzipien in Kognition und Kommunikation“, in: Boretzky et al. (Hrsg.) (1990), 37–51.
- Fucks, W. 1968. *Nach allen Regeln der Kunst. Diagnosen über Literatur, Musik, bildende Kunst – die Werke, ihre Autoren und Schöpfer*. Stuttgart.
- Grzybek, P. 2001a. „Grenzen der Individualität. Untersuchungen zu Puškins Vermärchen Царь Салтан“ [Vortrag beim Deutschen Slavistentag in Potsdam].
- 2001b. „Zur Entwicklung der Prosa Karel Čapeks – Quantitative Analysen“ [in Vorb.]
- Güter, H.; Arapov, M.V. (Hrsg.) 1982. *Studies on Zipf's Law*. Bochum [= Quantitative Linguistics; 16].
- Jackson, W. (ed.) 1953. *Communication Theory*. London.
- Köhler, R.; Martináková-Rendeková, Z. 1998. „A systems theoretical approach to language and music“, in: Altmann/Koch (eds.) (1998), 514–546.
- Madonia, G. 1969. „Ökonomie“, in: Martinet (ed.) (1969), 55–59.
- Mandelbrot, B. 1953. „An informational theory of the statistical structure of language“, in: Jackson (ed.) (1953), 486–502.
- 1954. „Structure formelle des textes et communication“, in: *Word* (10), 1–27.
- Martinet, A. 1960. *Grundzüge der Allgemeinen Sprachwissenschaft*. Stuttgart u. a.
- (ed.) 1969. *Linguistik. Ein Handbuch*. Stuttgart, 1973.
- Orlov, Ju.K.; Boroda, M.G.; Naderejšvili, I.Š. 1982. *Sprache, Text, Kunst. Quantitative Analysen*. Bochum [= Quantitative Linguistics; 15].
- Sperk, K. 1985. „Von Iconomique zu Reagonomics: Ein wortgeschichtlicher Beitrag“, in: Stemmler (Hrsg.) (1985), 29ff.
- Stemmler, Th. (Hrsg.) 1985. *Ökonomie. Sprachliche und literarische Aspekte eines 2000 Jahre alten Begriffs*. Tübingen.
- Uhlřířová, L.; Wimmer, G.; Altmann, G.; Köhler, R. (eds.) 2001. *Text as a Linguistic Paradigm: Levels, Constituents, Constructs*. Festschrift in Honour of Luděk Hřebřířek. Trier.
- Vološin, B.A.; Orlov, Ju.K. 1972. „Das verallgemeinerte Zipf-Mandelbrotsche Gesetz und die Verteilung der Anteile von Farbflächen in der Malerei“, in: Orlov et al. (Hrsg.) (1982), 263–270.
- Weinrich, H. (1991): „Ökonomie und Ökologie in der Sprache“, in: *französisch heute*, 4; 310–320.
- Wimmer, G.; Altmann, G. 2001. „Models of Rank Frequency Distributions in Language and Music“, in: Uhlřířová et al. (eds.) (2001), 283–294.
- Wimmer, G.; Wimmerová, S. „Ein musikalisches Rangordnungsgesetz“ [Manuskript]
- Zipf, G.K. 1935. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Cambridge, Mass., 1968.
- 1949. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge, Mass.