

Die linguistischen Hypothesen von G.K. Zipf

Claudia Prün

‘Das Werk von G.K. Zipf [hat] bis heute nichts von seiner Aktualität verloren; seine glänzende Funktionalanalyse des sprachlichen Verhaltens wird trotzdem immer wieder ignoriert. Deshalb sei an dieser Stelle noch der Hinweis gestattet, daß die Auswertung der Beobachtungen von Zipf und anderen bisher bereits zu gut einem Dutzend Gesetzhypothesen geführt hat. Die Weiterführung solcher funktionalanalytischen Untersuchungen ist ein vielversprechender Weg in Richtung auf die Entwicklung einer allgemeinen Zeichentheorie.’ (Köhler & Altmann, 1983:431).

1. Einleitung

G.K. Zipf war es wichtig, das menschliche Verhalten mit der selben Objektivität und den gleichen Methoden zu beobachten, wie ein Biologe das Sozialverhalten der Bienen oder den Nestbau von Vögeln beobachten würde (*Human Behavior*: Vorwort). Auch Zipf wollte *erklären* und nicht bei spekulativen Begründungen oder beschreibenden Regelwerken stehenbleiben. Seine Suche nach universalen Gesetzen sah er in Analogie zu anderen Wissenschaften wie der Physik, später der Soziologie, ohne die wissenschaftstheoretischen Voraussetzungen der funktionalen Erklärung (Altmann, 1981; Köhler, 1986) zu haben, die oben beschrieben wurden. Für ihn war die wissenschaftliche Methode ganz klar eine quantitative Methode und die Suche nach Gesetzen ein Kreislauf aus theoretischer Ableitung und empirischer Untersuchung. Obwohl er teilweise zu sehr überzeugt von seinen eigenen Vermutungen war („Monomanie“, Birkhan, 1979:50), betonte er stets, nicht die endgültigen Lösungen gefunden zu haben. Sein Ansatz hat in der systemtheoretischen Linguistik die Fortführung gefunden.

Die vorliegende Arbeit basiert auf einer gründlichen Exploration der Veröffentlichungen Zipfs, das sind die Monographien *Relative Frequency*¹, *Selected Studies* (erschienen 1932), *Psycho-Biology* (ersch. 1935), Zipf (1938e), *National Unity* (ersch. 1941) und *Human Behavior* (ersch. 1949), sowie 33 Aufsätze aus den Jahren 1937 bis 1950². Die gefundenen Hypothesen werden im Zusammenhang des Zipfschen Gedankengebäudes dargestellt. Kommentare beschränken sich auf das Nötigste.

Nach Lundberg und Dodd (1950), Miller (1968) und Birkhan (1979) ergibt sich die folgende tabellarische Biografie G.K. Zipfs:

- 07. Jan. 1902 geboren in Freeport, Illinois
- 1924 grad. Harvard, *summa cum laude*
- 1925 Studium in Bonn (F. Sommer) und Berlin (W. Schulze)
- 1929 Erscheinen der Dissertation *Relative Frequency*
- 1930 Ph.D. Harvard (*comparative philology*)
 - Lehrtätigkeit: deutsche Sprache³
- 1936 *Assistant Professor of German* (Harvard)
- 1939 *University Lecturer* (Harvard)
 - Geplante Veröffentlichung der deutschen, erweiterten Ausgabe von *Psycho-Biology* bei C. Winter, Heidelberg⁴

Birkhan (1979:50) berichtet: „Noch in den vierziger Jahren wurde Zipf, der dafür bekannt war, daß man in seinen Seminaren unablässig zu zählen hatte, nicht ‘wirklich Ernst genommen’ (mündliche Mitteilung eines ehemaligen Zipf-Schülers)“. Zipf erwähnte in der Regel die Namen derjenigen, die die mühsamen Zählungen durchgeführt hatten, in seinen Veröffentlichungen.

1950 *Guggenheim Fellowship* für das folgende Jahr

25. Sep. 1950 Tod.

Zipfs Nachruf (Lundberg & Dodd, 1950) erscheint nicht in einer linguistischen oder psychologischen, sondern in einer soziologischen Zeitschrift.

Im Lauf dieser Darstellung muß immer wieder auf einige Konzepte Zipfs zurückgegriffen werden, die am besten hier, sozusagen auf einer Metaebene, schon vorgestellt werden. Im Hauptteil der Arbeit beschränkt sich die Darstellung dann auf die jeweilige Instantiierung der beteiligten Konzepte und auf die Erklärung der Auswirkungen, die Zipf ableitet.

¹ Zipf (1929). Auf die wichtigeren Monographien wird in der vorliegenden Arbeit mit dem abgekürzten Titel bezug genommen (vgl. Literaturverzeichnis), um den Text leichter lesbar zu machen.

² Davon drei mit weiteren Autoren.

³ vgl. Deckblatt zu *Selected Studies*

⁴ vgl. Zipf (1938a; 1939:112)

Zwei Prinzipien: 'least effort' und 'relative frequency'

Das berühmteste Zipfsche Prinzip ist das der geringsten Anstrengung (*principle of least effort*, vgl. *Human Behavior*). Es besagt, daß ein handelndes Subjekt oder eine Gruppe von Subjekten sich seine Umgebung so organisiert, daß die zum Handeln erforderliche Energie „im Ganzen“ minimiert wird. Dabei ist das Minimum immer ein Optimum bezogen auf den zu erreichenden Effekt, die zeitliche Perspektive und andere Randbedingungen.

Weniger bekannt, aber in Zipfs Arbeiten früher als das erstgenannte Prinzip ist das *principle of relative frequency*. Dieses Prinzip ist für die Linguistik konzipiert und besagt, daß, je häufiger eine linguistische Einheit verwendet wird, sie umso weniger deutlich geäußert wird (*Relative Frequency* etc.).

Zwei Kräfte: Unifikation und Diversifikation

Unifikation und Diversifikation sind zwei „Kräfte“ oder Tendenzen, die einander in ihren Wirkungen entgegenstehen, aber gleichzeitig die Umwelt des „sich verhaltenden“ Subjekts bestimmen:

Unifikation bezeichnet die „egoistische“ Tendenz, sämtliche Handlungen nur auf die Bedürfnisse der eigenen Person auszurichten. Der ökonomische Vorteil von Unifikation nach dem Prinzip des geringsten Aufwands liegt darin, daß die Anzahl verschiedener Einheiten eines Typs (z.B. Siedlungen, Wörter) gering wird - idealerweise 1. Sämtliche Funktionen konzentrieren sich auf dieses Element, seine relative Verwendungshäufigkeit (z.B. Einwohnerzahl, Wortfrequenz) steigt auf 100%.

Diversifikation bezeichnet die Tendenz, Abläufe effektiver zu gestalten, indem verschiedene Funktionen auf verschiedene Elemente aufgeteilt werden. Maximale Diversifikation erreicht die 1:1-Zuordnung. Es gibt so viele Elemente, wie Funktionen ausgeübt werden müssen. Der ökonomische Vorteil liegt in der besseren Ausnutzung von gegebenen Potentialen (z.B. räumlich-geographisch, informationstheoretisch).

Zwei Analogien: Werkstatt und Glocken

Zipf bemüht zwei Analogien, um die Zusammenhänge von Funktionalität, Verwendungshäufigkeit und Aufwandsminimierung deutlich zu machen. Über ihren theoretischen Wert, die Art der Isomorphie von Modell und Wirklichkeit läßt sich streiten (vgl. Rapoport, 1982). Die Bilder müssen jedoch vorgestellt werden, um Zipfs Gedankenwelt besser zu verstehen.

Das Werkstattmodell (s. z.B. *Human Behavior*: 57-73; 158f.) beschreibt einen Handwerker, der idealerweise durch einen Punkt repräsentiert wird. Er hat verschiedene Werkstücke fertigzustellen, wobei einige Arbeitsabläufe (Aufgaben für die Werkzeuge) immer wieder vorkommen, andere speziell für ein Werkstück benötigt werden. Das Problem des Handwerkers ist die ökonomische Anordnung der Werkzeuge und ihre aufwandsparende Gestaltung. Dazu stehen ihm verschiedene Strategien zur Verfügung (vgl. Abschnitt 7).

Das Glockenläute-Modell (*Human Behavior*: 48-54) beschreibt eine Anordnung

von Glocken entlang einer geraden Linie. Der Glöckner, der sich an einem Ende der Anordnung befindet, hat in jedem Zeittakt genau eine Glocke zu läuten, egal, ob sie weit entfernt oder nahe an seinem Ausgangspunkt steht. Nach dem Läuten muß er noch im gleichen Zeittakt zum Ausgangspunkt zurückkommen. Jede Glocke muß mit einer gewissen relativen Häufigkeit betätigt werden, wobei die Häufigkeit mit steigender Entfernung abnimmt. Die Entscheidung des Glöckners, zu welcher Glocke er greift, beinhaltet nicht nur die vordergründige Aufwandsminimierung, zunächst alle nahe stehenden Glocken zu läuten, sondern erfordert auch, daß die weiten Wege gleichmäßig auf den vorgegebenen Zeitraum verteilt werden (vgl. besonders Abschnitt 11).

2. Phonologische Dynamik

In seiner Dissertation und ersten eruierten Veröffentlichung *Relative Frequency* entwickelt Zipf eine phonologische Hypothese. Sie lautet (als Zuspitzung des Prinzips der relativen Häufigkeit): „Der Deutlichkeitsgrad eines Lautes ist invers proportional zur relativen Häufigkeit dieses Lautes bezogen auf die Menge aller Laute in der Rede. Je häufiger ein Laut gebraucht wird, desto leichter ist er auszusprechen, und umgekehrt“.

Die Hypothese wird in *Selected Studies* und *Psycho-Biology* wieder aufgegriffen und Zipf liefert weitere Belege; sie wird jedoch nicht weiter entwickelt. Die einzige Modifikation ist die, daß Zipf später (*Psycho-Biology*: 49) die Beziehung als „nicht notwendigerweise proportional“ bezeichnet. Außerdem spricht er in späteren Veröffentlichungen von „Phonemen“ anstatt von „Lauten“ (*sounds*). In *Selected Studies* (S. 2, Fußnote) nimmt er begriffliche Gleichheit an, was aber undifferenziert und der Hypothese eher abträglich ist.

Zur Begründung dieser Hypothese, die eigentlich eine Instanz des Prinzips der relativen Häufigkeit darstellt, greift Zipf auf im Grunde genommen informationstheoretische Überlegungen zurück. Zur besseren Referenz zitiere ich hier nochmals das Prinzip der relativen Häufigkeit in voller Länge:

„Principle of Frequency. The accent, or degree of conspicuousness, of any word, syllable, or sound, is inversely proportionate to the relative frequency of that word, syllable, or sound, among its fellow words, syllables, or sounds, in the stream of spoken language. As usage becomes more frequent, form becomes less accented, or more easily pronounceable, and vice versa.“ (Relative Frequency: 4)

Zipf argumentiert, daß seltenere Spracheinheiten wichtiger sind, da sie mehr zum Inhalt beitragen, bzw. größeren Informationswert haben. Je häufiger ein sprachliches Element ist, desto stärker wird es erwartet und trägt damit nicht so viel zur Information bei. Neues, unerwartetes hat eine höhere „emotionale Intensität“ - es stört sozusagen den gleichmäßigen Reizstrom. Es muß wegen seiner Informationshaltigkeit unbedingt vom Hörer wahrgenommen werden, und das ist bei „größeren“

(wie auch immer „intensiveren“) Elementen eher gewährleistet.

Zur Illustration führt Zipf seinen Lesern den Extremzustand vor Augen, indem er nachvollzieht, was in einer Sprache geschehen muß, bei der jedes Wort mit demselben Laut beginnt (*Relative Frequency*: 38f.). Die Hörer können aus dem ersten Laut keine Menge erwarteter Wörter eingrenzen und legen darum keinen Wert darauf, diesen Laut zu hören. Der Sprecher kann deswegen vollkommen auf diesen Laut (als Anlaut) verzichten: Der häufigste Laut wird so zur Nichtexistenz abgeschwächt.

Zipf (*Relative Frequency*: 89) formuliert seine Hypothese mathematisch, indem er die Deutlichkeit bzw. die „Intensität“ eines Lautes als eine Funktion seiner Häufigkeit auffaßt (mit y = Intensität⁵ eines Lautes und x = Häufigkeit des Lautes; vgl. Abbildung 2.1):

$$(2.1) \quad Y = \frac{n}{x}$$

dabei ist n ein Proportionalitätsfaktor, Zipf interpretiert ihn in *Relative Frequency* als Faktor des Sprechtempos.

Das bedeutet, daß die oben genannte Hypothese überprüfbar wird, sobald die Begriffe geklärt und operationalisierbar sind.

Da die Hypothese in einem Bereich wirkt, der nicht unabhängig von anderen Prozessen, Wechselwirkungen und Zusammenhängen existiert, macht Zipf drei wichtige Zusatzannahmen.

Die erste bezieht sich auf die Randbedingungen phonetischer Veränderung: So beeinflußt z.B. die lautliche Umgebung eines Phonems dessen Disposition, sich zu verändern (z.B. assimilativ). Häufigkeitsveränderungen lösen dann lautliche Veränderungen bevorzugt an Stellen höherer Disposition aus. Die Umgebung kann eine Schiefe in der Phonemrealisation (*Psycho-Biology*: 97ff.) bewirken, die eventuell eine Vorstufe zur Phonemverschiebung darstellt. Ebenso können kontextbezogene Lautverschiebungen vorkommen, z.B. wurde im Deutschen /s/ vor /p/ in Initialstellung zu / / . Zipf geht davon aus, daß unter gleichen Bedingungen/Umgebungen gleiche Veränderungen stattfinden. Außerdem besteht die Möglichkeit analoger Veränderungen (*Relative Frequency*: 75): Wenn z.B. ein stimmhafter labialer Plosivlaut überhäufig vorkommt und sich verändert (z.B. stimmlos wird), dann ist es gut möglich, daß auch die anderen stimmhaften Plosivlaute diese Veränderung mitmachen, auch wenn ihre Häufigkeit noch „im Rahmen“ ist. Dies ist ein Hinweis darauf, daß die Hypothese unter Umständen von Lauten auch auf Merkmale und ihre Häufigkeit übertragen werden kann.

Die zweite Zusatzannahme ist die, daß die Häufigkeiten der Laute sich zwischen Schwellenwerten bewegen können, ohne sofort eine Wirkung im Lautsystem

⁵ In Anlehnung an das Weber-Fechnersche Gesetz der logarithmisch steigenden Reizintensität, nicht als Intensivierung eines Artikulationsparameters. Später bezeichnet Zipf diese Größe mit „Komplexität“.

zu zeitigen. Jeder Laut hat als Merkmalsansammlung eine ihm eigene Schwierigkeit, die mit der Deutlichkeit und damit mit der relativen Frequenz korreliert ist. Steigt die relative Frequenz des Lautes über einen bestimmten Wert, so wird die Aussprache abgeschwächt und einfacher. Sinkt die relative Häufigkeit unter einen bestimmten Schwellenwert, so wird die Artikulation verstärkt, schwieriger, aber damit deutlicher. Zipf macht diese Annahme, weil eine gewisse Stabilität des Systems gewährleistet sein muß. Lautverschiebungen bahnen sich graduell an, aber die Frequenzänderung muß so gravierend sein, daß eine systemische Veränderung spürbar wird (je höher die relative Häufigkeit, desto gravierender muß die Änderung sein, um zu wirken). Dazu kommt die Trägheit des Systems, das unausgewogene Zustände erst nach einer Zeitspanne ausgleicht.

Zuletzt braucht Zipf die Annahme der prinzipiellen Gleichheit der Sprechorgane aller Menschen, um seine Folgehypothesen aufzustellen.

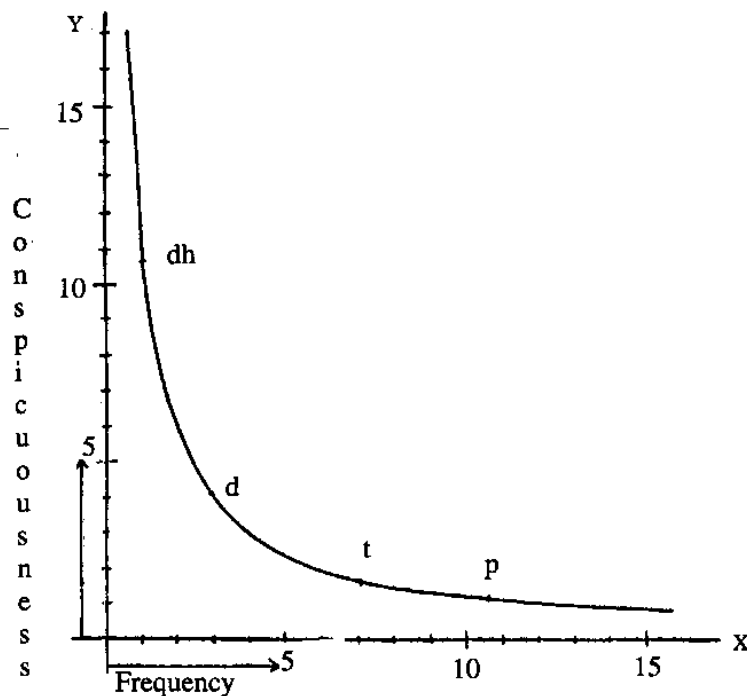


Abb. 2.1. Die hyperbolische Beziehung von Phonemhäufigkeit und -deutlichkeit nach Zipf (Relative Frequency: 89).

Aus dem Prinzip der relativen Häufigkeit und den oben genannten Zusatzannahmen ergibt sich, daß jeder Laut eine ihm eigene Schwierigkeit hat. So sind z.B. stimmhafte Laute intensiver (komplexer) als ihr stimmloses Pendant. Für Zipf folgen daraus drei universelle Hypothesen bezüglich der Struktur des Phoneminventars jeder Sprache (vgl. hauptsächlich *Relative Frequency, Selected Studies*, Zipf, 1949b sowie Zipf & Rogers, 1939):

- a) Die Häufigkeitsrangfolge oder sogar die Häufigkeitsverhältnisse vergleichbarer Laute oder Lautklassen in verschiedenen Sprachen sind stabil.

b) Die relativen Häufigkeiten vergleichbarer Laute sind in verschiedenen Sprachen nur leicht unterschiedlich (prinzipiell stabil/gleich)

a) und b) sollen aus Tabellen wie der folgenden (Tabelle 2.2) erkennbar sein:

Tabelle 2.2.

Relative Häufigkeiten [%] stimmhafter und stimmloser Verschlusslaute. Nach Zipf (1949b:396). Mit * markierte Werte beziehen sich auf „variphones“⁶, alle anderen auf Phoneme. Kursive Werte widersprechen der Hypothese p (stimmlos) $>$ p (stimmhaft).

Nr.	Sprache	<i>t</i>	<i>d</i>	<i>p</i>	<i>b</i>	<i>k</i>	<i>g</i>
1	Tschechisch	5.60	3.73	3.52	1.86	3.93	0.15
2	Niederländisch	7.83	4.67	1.99	1.20	3.21*	0.09*
3	Englisch	7.13	4.31	2.04	1.81	2.71	0.74
4	Ungarisch	7.18	3.30	<i>1.04</i>	<i>1.71</i>	5.72	2.45
5	Litauisch	5.76	2.61	3.71	1.35	4.61	1.36
6	Nord-Russisch	7.97	1.52	3.36	1.01	3.36	0.67
7	Süd-Russisch	7.05	2.46	2.79	1.51	3.97	1.66
8	Wendisch	6.26	3.02	2.55	1.56	3.29	2.41
9	Ost-Ukrainisch	3.83	3.24	2.82	2.11	4.11	-
10	Bulgarisch	7.54	3.55	2.82	1.32	2.98	1.46
11	Griechisch	7.58	2.87	3.38	0.49	4.07	1.74
12	Sanskrit	6.65	2.85	2.46	0.46	1.99	0.82
13	Latein	8.66	3.12	2.54	1.32	4.34	0.76
14	Italienisch	4.72	3.64	2.14	0.52	3.38	0.48
15	Spanisch	4.46	1.56*	2.92	0.46*	3.84	1.02
16	Portugiesisch	5.06	2.44*	2.68	0.30*	3.44	0.92
17	Französisch	4.90	4.54	3.96	1.82	3.30	0.36
Mittelwert		6.36	3.14	2.75	1.22	3.66	1.07
Standardabweichung		1.37	0.28	0.22	0.18	0.25	0.70

Die Annahme a) stellt sozusagen die horizontale Richtung des Vergleichs dar, die Annahme b) vergleicht die relativen Häufigkeiten der Laute in der Vertikalen und die Quotienten von Häufigkeiten der gleichen Artikulationsstelle, z.B. $p(t)/p(d)$. Die

⁶ Eine phonetisch begründete Abweichung von herkömmlichen Phonemklassifikationen, Anzeichen für ein in Veränderung befindliches Oppositionssystem. Für die genaue Bewertung siehe Zipf & Rogers (1939).

Abweichungen⁷ stellen für Zipf noch nicht automatisch einen Gegenbeweis dar. Vielmehr kann er dort, wo Ausnahmen in den Häufigkeitsverhältnissen (in einer Sprache) auftreten, diese durchaus begründen; z.B. wird das spanische „d“ nicht plosiv artikuliert (und damit deutlich, aufwendig), sondern interdental frikativ, d.h. schwächer als „t“, da es ohne Spannung artikuliert wird.

Damit geht Zipf einen wichtigen Schritt zur „Rettung“ seiner Hypothese (vgl. Altmann, erscheint), nämlich den der Überprüfung, ob Eigenschaften korrekt zugeschrieben wurden. Das spanische „d“ ist demzufolge nicht mit dem stimmhaften dentalen Okklusiven anderer Sprachen vergleichbar und fällt aus der Hypothese heraus. Es liefert sogar eine Bestätigung der Haupthypothese: der als „d“ notierte Laut wurde zu häufig, so daß er zu einem interdentalen Frikativ abgeschwächt wurde. Auf ähnliche Weise kann Zipf (*Relative Frequency, Selected Studies, Psycho-Biology*; ebenso Zipf & Rogers, 1939) auch die anderen Abweichungen erklären. In Zipf (1949b), also zu einer Zeit, als Zipf schon das gesamte Gebäude seiner Ökonomieprinzipien in Human Behavior dargelegt hatte, kann er seine Annahmen auch besser begründen: Phoneme werden aus der möglichen Menge der Artikulationen nach Prinzipien der Produktions- und der Wahrnehmungsökonomie ausgewählt. Eines dieser Prinzipien nennen wir heute maximale Distinktivität. Dies hat Phonemkorrespondenzen zwischen Sprachen zur Folge:

„We all feel from our experience with the phenomenon that the incidence of some phoneme-types such as those of the common vowels, spirants, liquids and nasals is far greater in the known phonetic systems than would be the case if sheer random chance were alone operative“ (Zipf, 1949b:2)

Da die Schwierigkeit (korreliert mit der Deutlichkeit) von Merkmalskombinationen universal ist (prinzipielle Gleichheit des Sprechapparats) entsteht die Balance, die unter a) und b) aufgeführt ist. Bei gleichem Schwierigkeitsgrad bilden sich Optionen, Auswahlmöglichkeiten.

Aus den Forderungen der Kombinationsökonomie (Zipf sagt „Permutationsökonomie“) und der maximalen Distinktivität ergibt sich die dritte Folgehypothese:

c) Das Phoneminventar jeder Sprache ist auf rund 20-60 Elemente begrenzt.

Diese Begrenzung der potentiell unendlichen Zahl von Einstellungskombinationen der Sprechwerkzeuge auf eine kleine Menge distinktiver Merkmalskombinationen ist für Zipf ein Indiz für das Gleichgewicht aus Phonemhäufigkeit und Phonemkomplexität (Zipf & Rogers, 1939). Es entsteht aus dem Zusammenwirken der konkurrierenden „Kräfte“ Kombinations-Ökonomie, Inventarsbegrenzung und Distinktivität der Einheiten.

In *Relative Frequency* behandelt Zipf „Laute“, ihre Häufigkeit und ihre Deut-

⁷ Hier z.B. Ungarisch: p/b. In ähnlichen Tabellen früherer Veröffentlichungen (z.B. *Relative Frequency*: 61, *Psycho-Biology*:75) treten mehr Ausnahmen auf, die hier durch das Konzept der *variphones* vermieden wurden.

lichkeit (Intensität). Später geht er dazu über, von Phonemen zu sprechen. Er richtet die Forderung an die Phonetik, die Artikulationsenergie für jedes Phonem festzustellen. Bis dies geleistet werden kann, tendiert er eher zu einer Art Markiertheitsauffassung (ein Merkmal mehr = schwierigere/deutlichere Artikulation) und vermischt so die Ebenen an manchen Stellen.

In (1938d) ordnet Zipf die Disziplinen Phonetik (als Phonometrie wie von Eberhard und Kurt Zwirner vertreten), Phonologie und seine eigene dynamische Philologie einander folgendermaßen zu:

Die Phonologie beschreibt die Normwerte, deren Erreichen von der Phonetik untersucht wird. D.h. die Phonetik liefert die Signifikanz der phonologischen bedeutungskonstituierenden Unterscheidungen. Die dynamische Philologie erklärt sprachliche Zusammenhänge durch die Formulierung von Gesetzen (z.B. „Zwirners Gesetz“: Die Bemühung um das phonetische Erreichen einer phonologischen Norm ist Gauß-normalverteilt (1938d) - was sie wahrscheinlich nicht ist).

Die Zusammenarbeit der drei Fächer besteht also in der

- Klassifikation durch die Phonologie
- Messung durch die Phonetik (Phonometrie)
- Erklärung durch die dynamische Philologie.

Wird z.B. eine Schiefe in der Lautverteilung festgestellt, so kann man unter Umständen eine Lautverschiebung vorhersagen. Durch eine zirkuläre (eher: spiralförmige) Vorgehensweise der Definition von Einheiten - Untersuchung der Eigenschaften - evtl. korrigierte Definition von Einheiten - Untersuchung der Eigenschaften... kann die dynamische Philologie als exakte Wissenschaft z.B. Klassifikationsfragen lösen. Im Beispiel (Zipf, 1938d) wendet er so die Annahme der stabilen relativen Lauthäufigkeit an, um zu entscheiden, ob eine /t/ + /s/-Kombination ein (verstärktes, seltenes) oder zwei (häufigere) Phoneme sind.

So können fragliche Klassifikationen oder nicht klassifizierbare Restmengen neu interpretiert werden, z.B. als beginnende Phonemverschiebungen. Auch komplementär distribuierte Laute (Variphone) haben z.T. möglicherweise die Grenze zum Phonemstatus schon überschritten (Zipf & Rogers, 1939). Die Technik der Oppositionsbildung reicht seiner Meinung nach noch nicht aus, da die Oppositionsfunktion nicht die einzige Funktion ist, die Form und Verhalten der Laute bestimmt. Zipf und Rogers (1939) plädieren daher für eine Ergänzung der operationellen Phonemdefinition aus den postulierten Erkenntnissen der dynamischen Philologie heraus. Ebenso können durch die dynamische Betrachtungsweise die „Gesetze“ der historischen Linguistik neu interpretiert werden. Es wird möglich, die Vorgänge bei Lautverschiebungen zu erklären (die damals z.T. noch üblichen Begründungen mit Klimaveränderungen, Einfluß der Geomorphologie auf Temperament und Sprache usw. werden hinfällig).

Zipf benutzt die Begriffe Deutlichkeit, Größe, Umfang, Intensität, Komplexität und Schwierigkeit von Phonemen fast austauschbar. Für ihn sind es Aspekte einer Eigenschaft der Sprachlaute. Was Deutlichkeit (leichte Wahrnehmbarkeit, Hörersei-

te) und Schwierigkeit (der Artikulation, Sprecher-seite) angeht, so nimmt er an, sie seien stark miteinander korreliert. Komplexität bezeichnet dasselbe wie Größe oder Umfang: Zipf wählt diese Begriffe, um deutlich zu machen, daß auf jeder linguistischen Analyseebene so etwas wie „Verkürzung“ oder „Verkleinerung“ möglich ist. Die Größe eines Lautes macht sich am ehesten an der Anzahl, aber auch an der Art seiner artikulatorischen Merkmale fest (ohne Unterscheidung relevanter und optionaler Merkmale). Allenfalls die Intensität drückt eine andere Nuance aus (aber nicht immer), nämlich die der „Intensivierung“ eines Lautes, z.B. bei gewollt deutlicher Aussprache durch die besondere artikulatorische Verstärkung eines oder mehrerer Merkmale. In den früheren Veröffentlichungen ist auch die Intensität eines Lautes gleichzusetzen mit seiner Größe. Je größer ein Laut, desto schwieriger ist er zu artikulieren (und umso seltener kommt er vor).

Die Beurteilung der Deutlichkeit ergibt sich aus der Kenntnis des Phonemsystems der untersuchten Sprache und dem paarweisen Vergleich von Lauten, die sich in einem Merkmal unterscheiden (z.B. /t/ + Stimme = /d/, d.h. /d/ ist deutlicher). Aus einer Reihe solcher Vergleiche folgt eine Ordinalskala bzw. eine geordnete Reihe der Laute (meistens bezogen auf eine Artikulationsstelle) nach ansteigender Deutlichkeit (z.B. / / - /t/ - /d/ - /dh/, vgl. *Relative Frequency*: 89 und Abbildung 2.1, oben). Natürlich ist auch „gleiche“ Deutlichkeit möglich. Später verwendet Zipf die Häufigkeit zur Abschätzung der Deutlichkeit, wegen der oben angegebenen mathematischen Beziehung, die er ja annimmt.

Die Schwierigkeit müßte Zipfs Ansicht nach von der Phonetik gemessen werden können (was allerdings bis heute nicht zufriedenstellend gelungen ist). Auch hier nimmt er einen direkten Zusammenhang mit der Anzahl der Merkmale an, wie bei der Deutlichkeit.

Eine weitere Möglichkeit, die Zipf (1946f) in Erwägung zieht, ist die Schätzung der Schwierigkeit durch die Irrtumsrate bei Taubstummen.

Den Zusammenhang der Artikulation mit der Lautumgebung läßt Zipf fast ganz außer acht. Der Gedanke findet nur Eingang in seine Überlegungen zur Schiefe phonetischer Realisation.

Aus dem Gesagten ergeben sich unter anderem folgende Methoden zur Veränderung der Phonemgröße:

- Vergrößerung durch Zufügung von Hauch, Spannung, Länge, Stimme usw., durch Diphthongierung usw. oder durch Intensivierung vorhandener Merkmale (stärkere Spannung, deutlicherer Hauch bis zur frikativen Artikulation).
- Verkleinerung durch entgegengesetztes Vorgehen (Ersetzung durch Elemente, die eventuell noch eine Teilmenge der Merkmale des ursprünglichen Lautes enthalten), Assimilation bis zur Löschung oder Übergang eines Merkmalsrestes auf benachbarte Laute, sowie die einfache Abschwächung „kontinuierlicher“ Merkmale wie Spannung oder Länge oder Kürzung.

Die Dynamik des Phonemsystems ergibt sich aus diesen Mechanismen. Veränderte Phonemhäufigkeiten (z.B. durch morphologische Prozesse) resultieren in Veränderungen der lautlichen Gestalt der Phoneme, wenn bestimmte Schwellenwerte er-

reicht wurden. Da die Veränderungen in unterschiedlichen Umgebungen möglicherweise verschieden ablaufen, kann die Veränderung eines Lautes wiederum die Häufigkeit eines anderen Lautes erhöhen (wobei die „eigene“ Häufigkeit auf ein vertretbares Maß sinkt) und diesen damit erneut zur lautlichen Veränderung zwingen. Es kommt zu Kettenreaktionen („Konsonantenrochade“ - Birkhan, 1979).

Das Phoneminventar, das sich in seiner Struktur der Oppositionen wie auch in seinem Umfang verändern kann, ist von solchen Verschiebungen mit betroffen. Änderungen des Inventarumfangs verändern wiederum die durchschnittlichen Längen der aus Phonemen zusammengesetzten Elemente. Da Phoneme zu Morphemen bzw. Wörtern kombiniert werden, hängen lexikalisches Inventar, Morpheminventar sowie Wort- und Morphemlängen eng mit dem Phoneminventar zusammen (Zipf, 1949b).

3. Wortlänge und Verwendungshäufigkeit

Die Hypothese über die Wortlänge stellt bei Zipf nur einen Zwischenschritt zur n - f -Verteilung dar (*Selected Studies*). Er präsentiert zwar Längendaten, die in Abbildung 3.1 graphisch wiedergegeben sind, stellt aber nur die Frequenzklassenbelegung als „Gesetz“ heraus. Zipfs Hypothese zur Wortlänge lautet (*Selected Studies*: 22): „Je länger ein Wort, desto weniger wird es benutzt. Je häufiger ein Wort benutzt wird, desto kürzer ist es.“

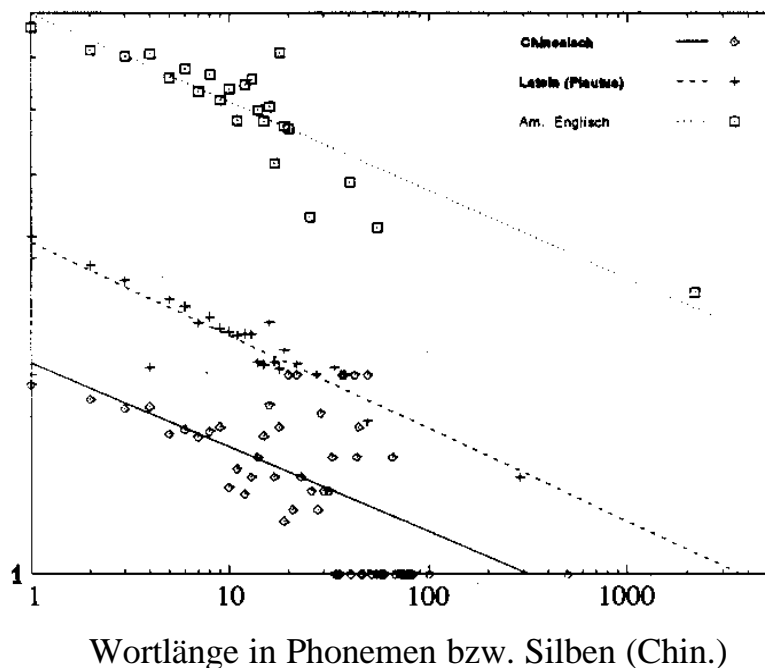


Abb. 3.1. Graphische Darstellung einiger Daten Zipfs (*Selected Studies*: 23; *Psycho-Biology*: 26-28) mit theoretischen Funktionsgraphen (Anpassungen der Gleichung Länge = Häufigkeit^{-a}). Logarithmische Achsen. Bei kleineren Häufigkeiten mittlere Längen.

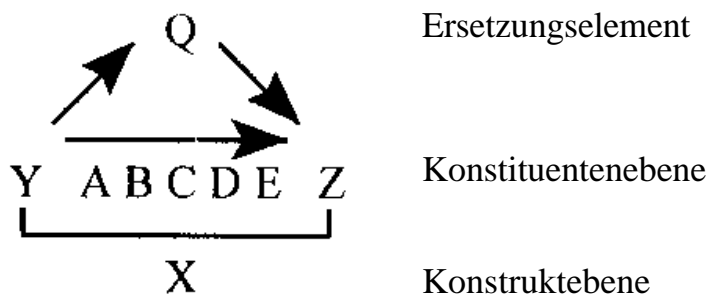
Der Zusammenhang ist invers, aber nicht notwendigerweise proportional (*Psychobiology*: 25). Zipf faßt die Hypothese nicht mathematisch, sondern sieht sie durch die Gültigkeit der Funktion $nf^2 = c$ (Frequenzklassenbelegungsgesetz, s. Abschnitt 10) als gezeigt an.

Es gibt mehrere Prozesse oder Prinzipien, die die Wortlängenhypothese stützen. Zunächst das Prinzip der relativen Häufigkeit: Häufigkeit und Deutlichkeit (conspicuousness) stehen in einem inversen Zusammenhang. Deutlichkeit hängt auf Wortebene für Zipf mit der Wortlänge zusammen. Das Prinzip und seine informationstheoretische Argumentation wurden im Abschnitt 2 ausführlicher dargestellt.

Ein weiterer Grund für die relative Kürze häufiger Wörter liegt in den syntagmatischen Zusammenhängen, die in Abschnitt 8 detailliert behandelt werden. Zipf geht davon aus, daß syntagmatische Kombinationen kleinerer Einheiten (Phoneme, Silben, Morpheme) sich mit häufigerem Gebrauch stärker festigen und dann auf charakteristische Elemente reduziert werden können. Festigkeit der Fügung liefert also die Kürzungs„disposition“.

Daß solche Wörter mit erhöhter Disposition zur Abkürzung dann auch tatsächlich verkürzt werden, liegt an der Ökonomieforderung, mit möglichst wenig Eingriff in das Sprachsystem Aufwand zu sparen. Dies ist durch Kürzungen an den häufigsten Elementen am ehesten gewährleistet (*Human Behavior*: 61). Die Ersparnis multipliziert sich mit der Häufigkeit des Elements.

In *Selected Studies* (S. 19ff.) will Zipf ein „Zeitdreieck“ als linguistische Maßeinheit einführen - es bleibt unklar, wie dies in die Praxis umgesetzt werden soll. Seiner Meinung nach wirken Abkürzungsprozesse auf allen linguistischen Ebenen, indem die Elemente ABCDE eines Ausdrucks X zu einer geringeren Anzahl Elemente, z.B. AQE zusammenschmolzen werden (bei gegebener Kürzungsdisposition, sprich vergleichsweise hoher Verwendungshäufigkeit und Kristallisation des Ausdrucks X). Das Abkürzungsdreieck sieht dann so aus:



Seine Manifestation findet das Zeitdreieck nicht etwa in Längenuntersuchungen, sondern in der n - f -Verteilung (*number-frequency: number of words with the same frequency of usage*). Die Verringerung der Anzahl von Wörtern gleicher Häufigkeit mit zunehmender Frequenz ist eine Folge der Wortlängenhypothese.

„There is not only a pull toward abbreviation, but the pull toward abbreviation varies in an orderly fashion as the frequency increases. That is, regardless of whether the language is highly inflected or

not (and this is very important to show the functional interrelationship of morphology and syntax), the number of words decreases with a marked similarity upon increase of usage.“ (Selected Studies: 24)

Dies ist die schon genannte *n-f*-Verteilung. Sie ist ein Ergebnis von Abkürzungsprozessen, indem eine relativ geringe Anzahl (kurzer) Wörter sehr häufig gebraucht wird - und zwar logarithmisch häufiger: Die „Anziehungskraft“ nimmt nach oben hin zu - und eine relativ große Anzahl (durchschnittlich längerer) Wörter sich die unteren Plätze teilt. Da die Länge außer in den Tabellen (s. oben) nie mehr explizit auftaucht, fällt es etwas schwer, Zipfs Argumentation nachzuvollziehen. Er scheint derart von den doppelt-logarithmischen Geraden der *n-f*-Verteilung begeistert zu sein, daß ihn die nicht sehr ordentlichen, stark variierenden Längendaten nicht mehr interessieren. Abkürzung und häufige Verwendung hält Zipf für so eindeutig und eng verknüpft, daß er meint, das eine durch das andere zeigen zu können.

Die Längen innerhalb einer Frequenzklasse hält Zipf für normalverteilt: „lengths of words (...) occurring once, twice, thrice, etc. are apparently distributed along the normal curve with the mean ever more to the left as the frequency becomes greater“ (1938b:266). Zur Verkürzung von Wörtern stehen verschiedene sprachliche Mittel zur Verfügung (vgl. *Psycho-Biology*: 29ff.). Sie lassen sich alle unter dem Begriff „Ersetzen im Zeitdreieck“ fassen:

- Ersetzen durch eine Teilmenge der schon vorhandenen Elemente des abzukürzenden Ausdrucks, vorzugsweise durch ein oder mehrere „charakteristische Elemente“. Dies ist die sogenannte „echte“ Abkürzung, *truncation*.
- Ersetzen durch einen anderen, kürzeren Ausdruck
- lokales Ersetzen von Wörtern, Satzteilen und noch komplexeren Ausdrücken z.B. durch Pronomen.

Abkürzung findet auch innerhalb eines bestimmten (z.B. sozialen, textlichen) Kontexts, wegen der kontextbedingt erhöhten Häufigkeit einer Bedeutung statt, d.h. sie muß nicht „global“ in einer Sprache durchgeführt werden, sondern kann sich auch nur in Subsystemen auswirken. Die genannten Mittel und die Wortlänge an sich stehen in Regelungs- bzw. Gleichgewichtsbeziehungen zu anderen sprachlichen Größen. Zipf erwähnt die folgenden:

Wortlänge und Bedeutungsspezifität stehen in einer inversen Gleichgewichtsbeziehung zueinander. Der Grund dafür liegt in dem Spezifikationsvorgang⁸ durch Kombination von Bezeichnern von Erfahrungsklassen (z.B. Wörtern) zu Syntagmen oder, bei stärkerer Kristallisation solcher Syntagmen, zu neuen Wörtern. In *Human Behavior* (S. 189) interpretiert Zipf die Abkürzung dahingehend, daß „müßige“ Einheiten, die nichts zur Spezifikation beitragen, gelöscht werden können.

Die Ersetzung von Wörtern durch kürzere Wörter hat zur Folge, daß diese Wörter (falls sie keine echten Neubildungen sind) mit einer zusätzlichen Bedeutung beladen werden.

⁸ Spezifikation bedeutet Bezeichnen einer selteneren Erfahrungsklasse.

Bei der Ersetzung durch ein neues Wort spielt auch die relative Länge dieses Wortes eine Rolle, damit der Abkürzungseffekt zustandekommt. So bestimmt die Länge dieses Wortes, im Verhältnis zu seiner Frequenz und zur Länge des zu ersetzenden Wortes, seine Verwendungshäufigkeit mit. Hier schaukeln sich die Prozesse gegenseitig hoch: schon häufig benutzte, kurze Wörter ersetzen andere, bekommen deren Bedeutungen und deren Häufigkeiten dazu; die Folge sind sehr steile Kurven, logarithmische Zusammenhänge. Wörter mit freien Frequenz-„Kapazitäten“ werden wieder in eine Position gesetzt, die ihrer Länge eher entspricht.

4. Die Ebene der Morpheme

Die Hypothese zur Morphemlänge wird analog zum Wortlängenzusammenhang begründet, der in Abschnitt 2 ausführlich dargestellt wurde. Sie lautet: Morphemhäufigkeit und Morphemlänge stehen in einem inversen Zusammenhang (*Psycho-Biology*: 172-176).

Zipf hat diese Annahme nicht mathematisch formuliert. Seine Beispiele aus *Psycho-Biology* sind in der Tabelle 4.1 wiedergegeben.

Tabelle 4.1.

Mittlere Länge (Phonemzahl) der dreißig bzw. 400 häufigsten autochthonen deutschen *Verbalpräfixe*, *Verbstämme* und *Verbalsuffixe* nach Kaeding, rangiert und in Klassen zu je fünf bzw. 50 benachbarten Werten. Nach Zipf, *Psycho-Biology*: 174-175. Alle Zählungen nach Kaeding, *Häufigkeitwörterbuch der Deutschen Sprache* (Steglitz, 1897-1898).

30 häufigste deutsche Präfixe		400 häufigste deutsche Verbstämme		30 häufigste deutsche Suffixe	
gruppierte Ränge	mittlere Phonemzahl	gruppierte Ränge	mittlere Phonemzahl	gruppierte Ränge	mittlere Phonemzahl
1 - 5	2.2	1 - 50	2.72	1 - 5	1.8
6 - 10	2.4	51 - 100	3.12	6 - 10	2.4
11 - 15	2.4	101 - 150	3.08	11 - 15	2.0
16 - 20	3.2	151 - 200	3.16	16 - 20	2.4
21 - 25	3.2	201 - 250	3.34	21 - 25	2.4
25 - 30	4.0	251 - 300	3.42	25 - 30	2.8
		301 - 350	3.42		
		351 - 400	3.52		

In Abbildung 4.1 sind die Werte der Tabelle 4.1 graphisch dargestellt. Die Daten weisen einen deutlichen Trend zur „Verlängerung“ der Morpheme mit abnehmen-

der Verwendungsfrequenz auf.

Die systemischen Zusammenhänge der Morphemlänge mit Morphemfrequenz, Bedeutungszahl, Kristallisationsgrad usw. gleichen denen der lexikalischen Ebene und werden hier nicht mehr gesondert aufgeführt. Interessant ist jedoch, daß die n - f -Verteilung der Morpheme in der Regel steiler ist als die der Wörter. Zipf (*Human Behavior*: 188) hält sie für „*rectilinear on arithmetic-logarithmic co-ordinates* (i.e. $\log f + \text{rank} = C$)“⁹. Ihre Diversifikation ist nicht so stark, ihr Inventar ist kleiner und ihre Bedeutung ist „spezifischer“ als die von Wörtern. Diese Zusammenhänge werden im Abschnitt über syntagmatische Beziehungen (8) ausführlicher besprochen.

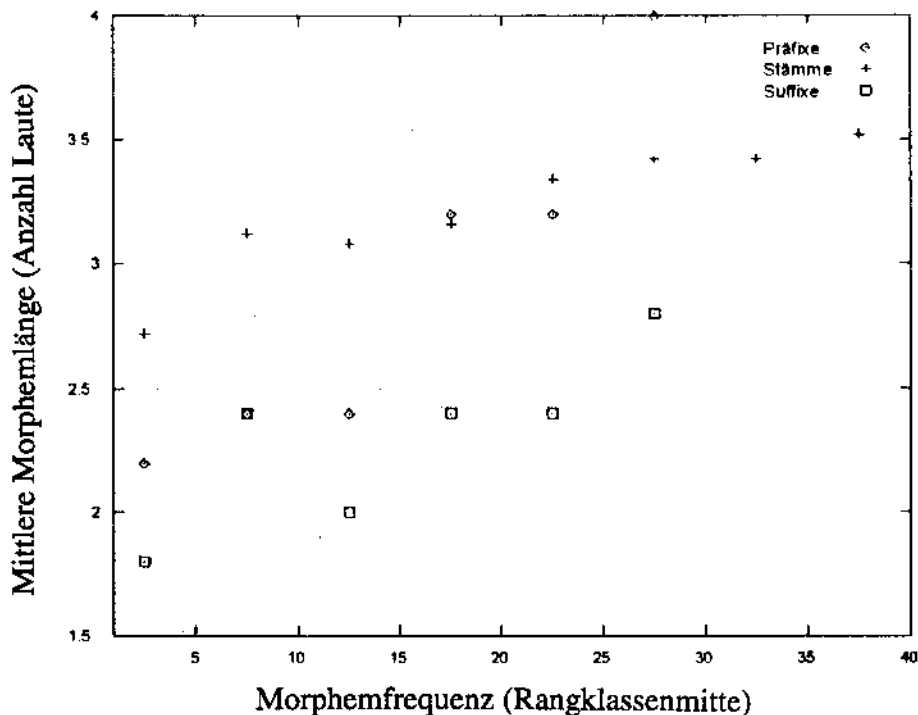


Abb. 4.1. Mittlere Morphemlängen der häufigsten deutschen Verbalpräfixe, -stämme und -suffixe (nach Kaeding). Für die Stämme wurden die Rangklassenmittlen mit $\frac{1}{10}$ multipliziert (bessere Darstellbarkeit).

5. Prosodie: Die Betonung von Morphemen und Wörtern

Die Betonungshypothese gehört zu den ersten Hypothesen Zipfs, sie ist so alt wie die Überlegungen zur Phonologie. Später tritt am Rande noch eine Annahme zum Satzaccent dazu. Die Hypothesen lauten: Die Betonung eines Morphems und seine Häufigkeit stehen in einem inversen Zusammenhang (*Relative Frequency*: 4). Die

⁹ Dieser Vermutung geht er an keiner anderen Stelle nach.

Position des Satzakkzents und die Frequenz des betonten Wortes stehen in einem inversen Zusammenhang (*Psycho-Biology*: 132).

Sie werden (a.a.O.) mit dem Prinzip der relativen Häufigkeit erklärt: Die Betonung stellt eine Intensivierung des betreffenden sprachlichen Elements dar und steigert somit seine Deutlichkeit (*conspicuousness*). Sprecherinteressen - nämlich die Minimierung des Artikulationsaufwandes - und Hörerinteressen - ausreichende Verständlichkeit besonders der informationsreichen, seltenen Elemente - stehen in einer solchen Beziehung zueinander, daß seltenere Elemente eher betont werden und häufigere eher unbetont bleiben.

Eine mathematische Funktion gibt Zipf nicht an, er zeigt seine Annahme nur anhand geeignet gewählter Beispiele. In Tabelle 5.1 sind Häufigkeitswerte für die maskuline Sanskrit-Deklination dargestellt. Die Häufigkeiten wurden in der *Rigveda* (früheste Sanskrit-Literatur) ermittelt.

Tabelle 5.1.

Häufigkeiten männlicher Kasusendungen (Substantive) im Sanskrit (*Rigveda*) nach Zipf (*Psycho-Biology*: 139). Fälle in Kapitälchen sind stammbetont, die anderen endungsbetont.

<i>Singular</i>		<i>Plural</i>	
NOMINATIV	24 286	NOMINATIV	10 981
AKKUSATIV	17 551	Akkusativ	5 353
Instrumental	4 234	Instrumental	3 360
Dativ	4 092	Dativ	363
Ablativ	923	Ablativ	124
Genitiv	5 274	Genitiv	1 595
Lokativ	3 789	Lokativ	1 546

Die drei häufigsten Klassen sind stammbetont, die seltenen Klassen sind endungsbetont. Weitere Beispiele in *Relative Frequency, Selected Studies* und *Psycho-Biology* sollen die Annahme belegen. Die Betonungshypothese hat Auswirkungen im System: Erhöht sich die Verwendungshäufigkeit eines Morphems, so kann dies den Verlust der Betonung bewirken, weil das Morphem nicht mehr so stark gegen andere abgegrenzt werden muß. Zipf führt als Beispiel die präfigierten Verben im Lateinischen an (*Relative Frequency*: 25ff.): Zunächst wurde das Präfix betont, um es gegen die anderen präfigierten Formen des gleichen Verbs abzugrenzen (Abgrenzung im Präfix-Paradigma). Später verschob sich die Betonung auf den Stamm, um das Verb gegen andere Verben mit dem gleichen Präfix abzugrenzen (Abgrenzung im Verbal-Paradigma)¹⁰.

¹⁰ Zipf gebraucht in *Psycho-Biology* (S. 166-172) diese Argumentation, um das feste Akzentmuster des Lateinischen (3-Silben-Regel) zu erklären. Besonders häufige Verben sollen mittels

Je fester ein Morphem in eine Verbindung eingefügt ist, desto weniger wahrscheinlich wird es betont. Zipfs Beispiel (*Relative Frequency*: 33) sind die deutschen präfigierten Verben: ist das Präfix abtrennbar, so ist es in der Regel betont. Ist das Präfix fest mit dem Stamm verbunden, so ist es meist unbetont. Die Hypothese zur Satzbetonung ergibt sich für Zipf so selbstverständlich, daß er ihr nicht weiter nachgeht. Systemeffekte gibt er zu diesem Problem ebenfalls nicht an. Mandelbrot (1957) kann den Zusammenhang mathematisch begründen.

6. Historische Hypothesen

In 1946b, 1947c und *Human Behavior* (S. 109-120) entfaltet Zipf einige Beobachtungen und Hypothesen zur Verteilung von verschiedenen Herkunfts- und Altersklassen im Lexikon. Er setzt sie in Beziehung zur Verwendungshäufigkeit, indem er Frequenzränge und Klassen von Frequenzrängen auf das Alter der in ihnen zusammengefassten Wörter hin untersucht.

Häufigkeit und Alter sind demnach positiv miteinander korreliert. Die Korrelation ist möglicherweise systematisch nicht-linear (bei der Operationalisierung von Häufigkeit als „Rangzahl“ ist die Beziehung natürlich negativ), bzw. folgt einer arithmetisch-logarithmischen Kurve (Zipf, 1947c:527).

Alter und Wortlänge (Silbenzahl) sind innerhalb einer Frequenzklasse negativ korreliert.

Im Wortschatz spiegeln sich Kulturkontakte durch kulturelle Stratifikation, d.h. Bildung von Schichten gleicher Herkunft. Deren Verteilung über die Häufigkeitsränge läßt Schlüsse über die Geschichte des Volkes zu, das die Sprache benutzt.

Zipf konnte diese Hypothesen nicht in konkrete mathematische Funktionen fassen. Er vermutete aber einen mathematischen Zusammenhang und zeigte ihn indirekt durch die Rang-Frequenz- und Frequenzklassen-Belegungs-Graphen. Ein Beispiel ist in Abbildung 6.1 gegeben.

Die *r-f*-Kurven A (für Pennsylvanisch insgesamt) und B (nur amerikanische Lehnwörter) zeigen deutlich, daß im Bereich höherer Frequenzen der Anteil der Lehnwörter immer stärker abnimmt. Die dazugehörigen *n-f*-Kurven C und D illustrieren denselben Sachverhalt (Frequenzen zwischen 10 und 1); Je höher die Frequenz, desto kleiner ist der Anteil amerikanischer Lehnwörter (D, die Anzahl Wörter mit gleicher Frequenz ist auf der Ordinate abgetragen). Die Bildung von etymologischen Schichten über Klassen von Frequenzrängen ist für das Englische in Abbildung 6.2 (A und C) und für Gotisch in Abbildung 6.3 illustriert.

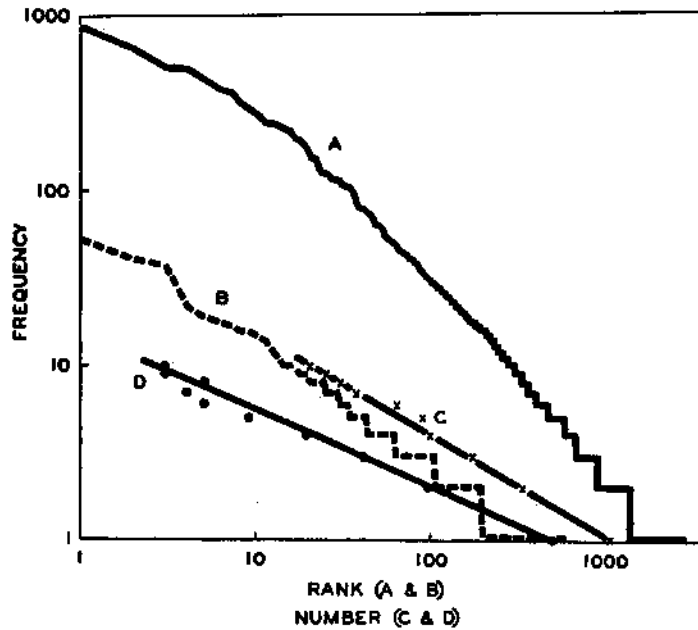


Abb. 6.1. R - f - (A, B) und n - f -Kurven (C, D) für Pennsylvanisch. A, C: Pennsylvanisch insgesamt; B, D: amerikanische Lehnwörter. Nach Zipf (1946b:354).

Der Zusammenhang von Alter und Länge ist aus der ersten Hypothese abgeleitet, denn den negativen Zusammenhang von Frequenz und Länge nimmt Zipf als gegeben an. Die Abbildungen 6.2 B und D zeigen Grafiken Zipfs, die die Herkunft der Wörter für die verschiedenen Längen (in Silben) angeben. Die Abbildungen beziehen sich nur auf Wörter der Frequenz 1.

Für Zipf (1946b) hängt das Phänomen damit zusammen, daß die psychologische Trägheit, wie er es nennt, mit einer Potenz der relativen Häufigkeit steigt. Gewohnheit stabilisiert die Überlebenswahrscheinlichkeit von sprachlichen Ausdrücken. Dazu kommt, daß die ökonomische Verwendung des lexikalischen Inventars die Anzahl der Bedeutungen bei den häufigeren Wörtern im Mittel erhöht (siehe Abschnitt 7). Der Verlust dieser Wörter wäre wesentlich, darum sind sie über lange Zeit stabil.

Zu den oben genannten Hypothesen gehören noch einige weitere Vorstellungen. Zipf prägt die Begriffe *senescent* und *nascent* für „alte“ und „neue“ Wörter (*Human Behavior*). Die alten Wörter, die langsam aussterben (*senescent words*), finden sich neben neuen Wörtern, die langsam eingeführt werden. Sie sind zu kurz für ihre Seltenheit, falls sie im Laufe ihres Lebens einmal häufiger waren und gekürzt wurden. Ihr Gegenstück sind relativ häufig gebrauchte *nascent words*, die (noch) zu lang für ihre Häufigkeit sind.

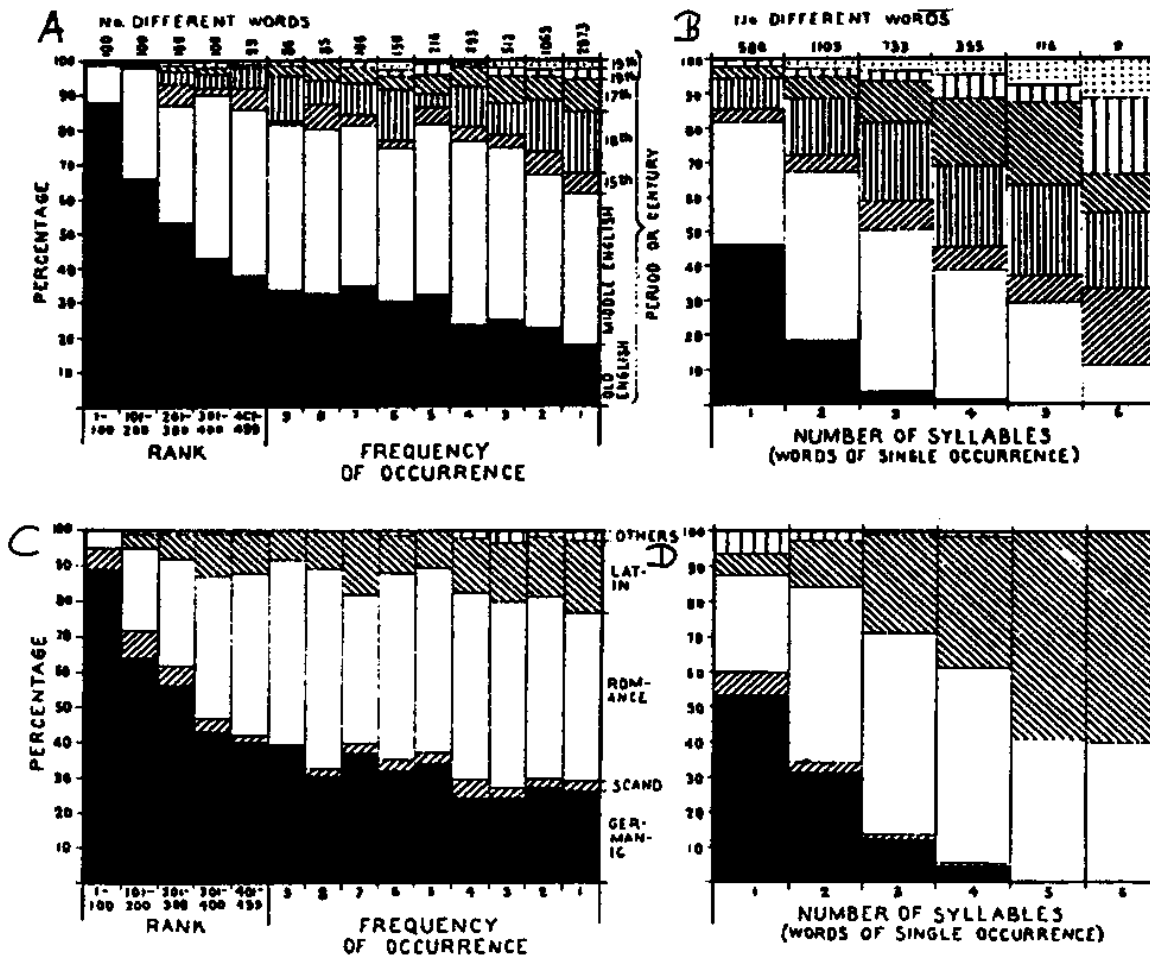


Abb. 6.2. A: Historische Schichten der engl. Lexik über dem Frequenzrang. B: Historische Schichten der engl. Lexik über der Wortlänge [Silben] für die Frequenz $f = 1$. C: Herkunftsschichten der engl. Lexik über dem Frequenzrang. D: Herkunftsschichten der engl. Lexik über der Wortlänge [Silben] für die Frequenz $f = 1$. Zählungen nach Eldridge. Abbildung nach Zipf (*Human Behavior*: 111).

Analog zur kulturellen Stratifikation einer Sprache ist im Wortschatz einer Person eine *Persönlichkeitsstratifikation* zu finden. Schichten der Reihenfolge, Art, Qualität etc. von Erfahrungen spiegeln die Entwicklung der Persönlichkeit in ihrer Erfahrungswelt. Dies entspricht der Rekapitulationshypothese der Biologie auf psychologischer Ebene (Zipf, 1946b; *Human Behavior*: 246).

Länge, Häufigkeit und Alter von Wörtern hängen nach Zipf aufgrund ökonomischer Ausnutzung des (lexikalischen) Inventars miteinander zusammen. Darum bestehen im System Gleichgewichtsbeziehungen nicht nur zwischen Länge und Häufigkeit (invers, vgl. Abschnitt 3), Länge und Alter (invers) sowie Häufigkeit und Alter (direkt), sondern auch zwischen Häufigkeit und Bedeutungszahl, Länge und Spezifität und - indirekt - Alter und Bedeutungszahl bzw. Spezifität. Die Beziehungen zu diesen „semantischen“ Größen werden im folgenden Abschnitt (7) ausführlich besprochen. Hier werden sie nur festgestellt.

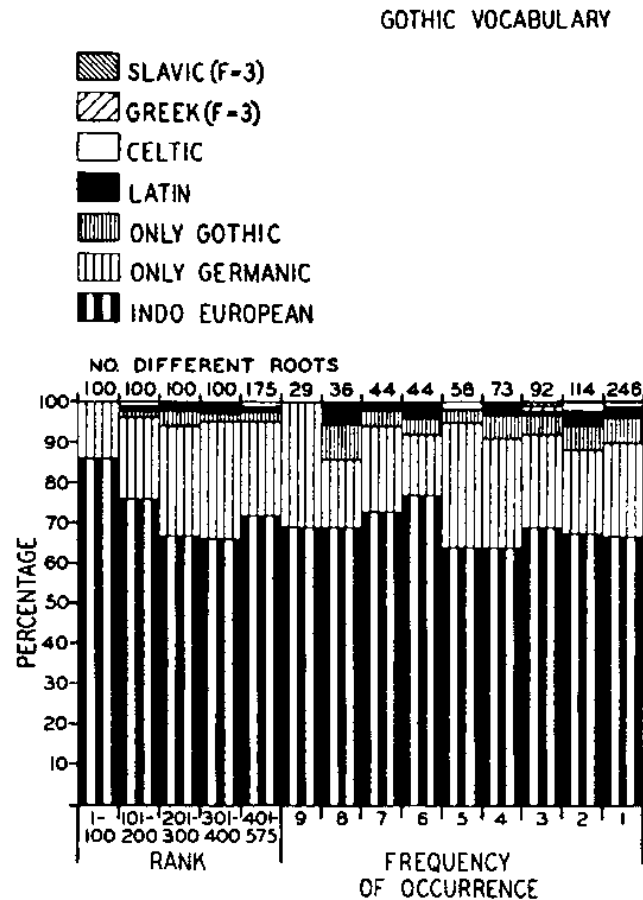


Abb. 6.3. Alters- und Herkunftsschichten der gotischen Lexik im rangierten Häufigkeitswörterbuch. Nach Zipf (Zipf, 1946b:352; *Human Behavior*: 114).

Zipfs Hypothesen wurden von Arapov und Cherc (1983) mathematisch formuliert und überprüft. Die Zusammenhänge von Alter und Frequenzrang bzw. Länge konnten so bestätigt werden.

7. Die Bedeutung sprachlicher Einheiten

Was den Bereich der sprachlichen Bedeutungen angeht, so zerfällt dieser bei G. K. Zipf in zwei Teile: in eine sehr explizite Hypothese zur Bedeutungszahl von Wörtern und daneben in Überlegungen zur Bedeutungskonstitution.

7.1. Zipfs Hypothese zur Bedeutungszahl von Wörtern

Häufigkeit und Bedeutungszahl eines Wortes hängen funktional zusammen: je häufiger ein Wort ist, desto mehr Bedeutungen trägt es (*Human Behavior*: 27ff.; 67ff.;

75ff.). Seine Ökonomieüberlegungen überzeugen Zipf davon, daß erst die vergleichsweise hohe Frequenz eines Wortes dazu führt, daß dieses Wort noch mehr Bedeutungen annehmen kann.

Der Handwerker aus Zipfs Werkstattmodell „spart“ sich Arbeit, indem er Werkzeuge, die sowieso häufig gebraucht werden (und sich damit in seiner Griffnähe befinden, sowie relativ klein und leicht gestaltet sind) für andere Zwecke und Aufgaben mitbenutzt (vgl. v.a. *Human Behavior*: 67). D.h. ein Sprecher kann, wenn der Hörer ihn zu verstehen in der Lage ist, ein Wort verwenden, das in einem anderen Kontext eine ganz andere Bedeutung hat¹¹. Motivation dafür ist die ökonomische Inventar-Ausnutzung (vgl. Abschnitt 9).

Ein anderer Prozeß, der zu einer erhöhten Bedeutungslast von Wörtern führt, ist die Abkürzung durch Ersetzen (*Psycho-Biology*: 274ff.). Der Sprecher hat Interesse daran, ein kurzes (und damit erwartungsgemäß häufiges) Wort zu verwenden. Diese Wörter bekommen deshalb eher noch eine Bedeutung zusätzlich.

Insgesamt dient also die Erhöhung der Bedeutungslast der Inventarsminimierung. Der Extremfall ist die Unifikation: alle Bedeutungen auf einem Ausdruck. Da das Hörerinteresse (in der Regel Diversifikation: jeder Bedeutung ein Ausdruck) mit der Häufigkeit des Ausdrucks abnimmt, wirkt die Unifikation stärker im Bereich hoher Frequenzen.

Zur mathematischen Ableitung des Zusammenhangs braucht Zipf eine ergänzende Annahme, und zwar die, daß sich aus dem Wirken von Unifikations- und Diversifikationskräften eine Verteilung $P(m)$ der Einzelbedeutungen mit den Einzelhäufigkeiten f_{mi} derart ergibt, daß der Mittelwert \bar{f}_m etwa gleich der Anzahl der Wortbedeutungen m ist (vgl. *Human Behavior*: 28)¹².

Die eigentliche Hypothese zur Bedeutungszahl wird unten als (7.7) bzw. (7.8) dargestellt. Zipf leitet aber eigentlich die Bedeutungszahl als Funktion des Frequenzranges ab (vgl. Zipf, 1942b, 1945a, *Human Behavior*). Die Bezeichner heißen:

F (frequency):	Häufigkeit eines Wortes
f_{mi} (frequency):	Häufigkeit der i -ten Bedeutung eines Wortes
\bar{f}_m (frequency):	durchschnittliche Häufigkeit einer Bedeutung eines Wortes
m (meaning):	Anzahl Bedeutungen des Wortes ($0 < i \# m$)
r (rank):	Häufigkeitsrang

¹¹ Damit eine Verwendungserweiterung mit Zunahme der Bedeutungszahl möglich wird (Hörerinteresse: Verständlichkeit), müssen sich zunächst bestimmte Aspekte (Bedeutungsgene, s. unten) zweier Bedeutungen entsprechen. Homonymie (gleichlautende Wörter mit z.B. unterschiedlicher Etymologie) und Polysemie unterscheidet Zipf nicht.

¹² Kritik siehe Rapoport (1982). Altmann und Schwibbe (1989:77) geben Zipfs Argumentation wörtlich wieder.

Es gilt (Wortfrequenz ist Summe der Bedeutungsfrequenzen):

$$(7.1) \quad F = \sum_{i=1}^m f_{m_i}$$

bzw. Wortfrequenz als Produkt aus Bedeutungszahl und mittlerer Bedeutungsfrequenz:

$$(7.2) \quad F = m \cdot \bar{f}_m.$$

Laut der ergänzenden Annahme oben ist:

$$(7.3) \quad m \approx \bar{f}_m$$

das ergibt sich aus (7.2)

$$(7.4) \quad f \approx m^2$$

bzw.

$$(7.5) \quad F \approx \bar{f}_m^2$$

und mit der Rang-Frequenz-Verteilung (ihre Gültigkeit ist notwendig, darum gilt die Hypothese (7.7) bzw. (7.12) unten nur auf der lexikalischen Ebene, Zipf, 1946f.)

$$(7.6) \quad r \cdot F = \text{const.}$$

gilt genauso auch

$$(7.7) \quad r \cdot m^2 = \text{const.}, \text{ die Rang-Bedeutungszahl-Verteilung.}$$

Variante: an anderer Stelle (1949b) benutzt Zipf eine Umformung von (7.4) zur Herleitung, nämlich

$$(7.8) \quad m \cdot F^2 = \text{const.},$$

was bedeutet, daß die Anzahl Bedeutungen m eines Wortes mit Frequenz F proportional zur Quadratwurzel der Frequenz ist. Daraus läßt sich (7.7) ableiten, denn aus

$$(7.9) \quad m = \alpha F^{\frac{1}{2}} \text{ und der Rang-Frequenz-Verteilung}$$

$$(7.10) \quad F = \frac{\beta}{r_F}$$

ergibt sich durch Einsetzen von F in (7.9) ($\alpha \cdot \beta^{\frac{1}{2}} = c^{\frac{1}{2}}$)

$$(7.11) \quad m_F = \left(\frac{c}{r_F} \right)^{\frac{1}{2}}$$

bzw.

$$(7.12) \quad r_F m_F^2 = c$$

mit $c = \text{const.}$ wie oben.

Der Index F soll verdeutlichen, daß es sich um die Ränge r und Bedeutungs-
zahlen m handelt, die zum Wort mit der Frequenz F gehören. Die Herleitung ist aus
den genannten Veröffentlichungen zusammengefaßt.

In Abbildung 7.1 ist Zipfs Beispiel einer Rang-Bedeutungszahl-Verteilung dar-
gestellt.

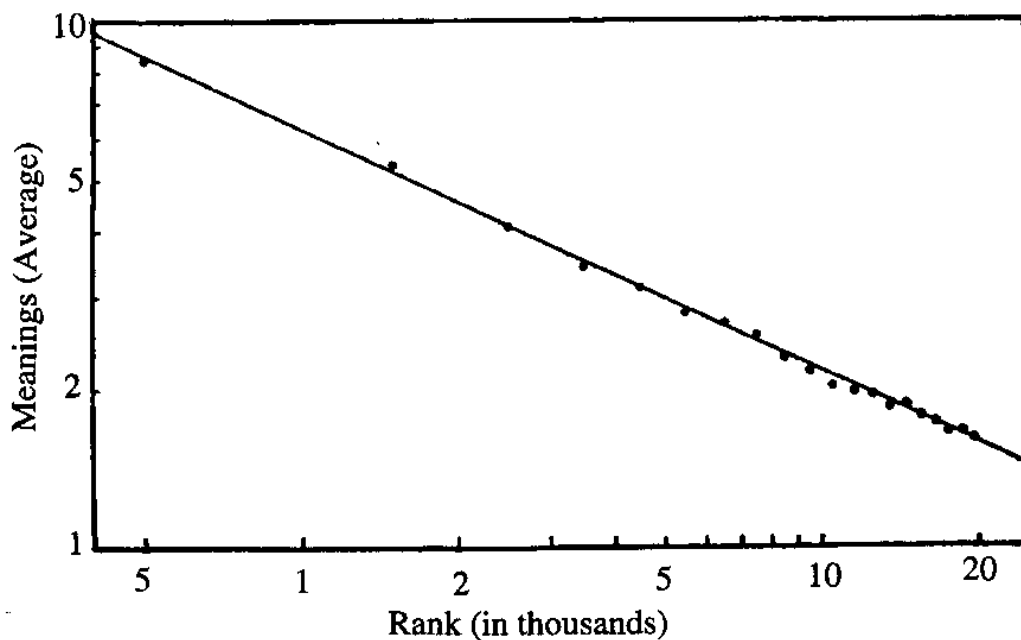


Abb. 7.1. Beziehung von Bedeutungszahl m und Frequenzrang r (Sprache: Eng-
lisch, Zählungen nach Thorndike). Theoretische Gerade $m_r = r^{0.4605}$.
Abbildung nach Zipf (*Human Behavior*: 30).

Aus den beiden funktionalen Beziehungen Frequenz-Länge und Frequenz-
Bedeutungszahl leitet sich ab, daß sich Länge und Bedeutungszahl invers zueinan-
der verhalten (vgl. *Human Behavior*: 61). Dieser Zusammenhang kann ebenfalls
funktional interpretiert werden, wenn man sich die Werkstatt-Analogie vor Augen

hält: es spart Mühe, ein kleines Werkzeug zu benutzen; zur Kodierung von zusätzlichen Bedeutungen werden daher die kurzen Elemente den längeren vorgezogen (von der Sprecherseite her), vor allem wenn die zu kodierende Bedeutung relativ häufig ist (sinkende Hörerinteressen).

Die Bedeutungszahl hängt auch mit dem Kristallisationsgrad eines Ausdrucks zusammen. Dies aber nur mittelbar, da die Kristallisation zur Verkürzung führt. Relative Häufigkeit führt wiederum zur Kristallisation, auch auf diesem Wege regelt sich also das Gleichgewicht von Frequenz und Bedeutungszahl (*Human Behavior*: 76ff.).

Syntaktische und syntagmatische Beziehungen (Juxtaposition, Reihenfolge) zwischen Wörtern ergeben neue Bedeutungen, die in den Wörtern selbst nicht notwendigerweise vorhanden waren (*Psycho-Biology*: 278ff.).

Zwischen Bedeutungen und Ausdrücken (z.B. Wörtern) besteht eine $m:n$ -Beziehung: n Ausdrücke kodieren m Bedeutungen. Ein Ausdruck kann mehrere Bedeutungen haben und eine Bedeutung kann durch mehrere Ausdrücke kodiert werden. Dies entspricht dem Konzept der funktionalen Äquivalente auf Bedeutungsebene (vgl. *Human Behavior*: 264ff., 270ff.).

Als *Primärbedeutung* eines Ausdrucks gilt die Bedeutung, in der der Ausdruck am häufigsten verwendet wird (*Psycho-Biology*: 288). Da sich die Verwendung mit der Zeit verschieben kann, ist natürlich die Primärbedeutung variabel.

Der Zusammenhang von Bedeutungszahl und Alter wurde schon genannt: je mehr Bedeutungen ein Wort hat, desto höher ist seine Chance, zu „überleben“. Außerdem steigt die Kristallisation mit dem Alter, wenn ein Ausdruck häufig verwendet wird (*Human Behavior*: 109ff.).

7.2. Zipfs Semantik: Bedeutungen als Wahrnehmungsklassen¹³

Der menschlichen Wahrnehmung ist direkt ein mentaler Klassifikationsprozeß nachgeschaltet. Wahrnehmungsreize evozieren Klassifikation als Reaktion der menschlichen Informationsverarbeitung (*mental reaction; mentation*).

Klassifikatorische Minimalleistungen, die nicht weiter aufgeteilt werden, bezeichnet Zipf als „Bedeutungsgene“.

Ein sprachliches Symbol steht für eine Menge von Genen: die sprachlichen Reize werden (im Spracherwerb und auch weiterhin) mit Erfahrungswerten korreliert, wodurch sie „Bedeutung“ erlangen. Die Erfahrungswerte der Wahrnehmung liegen außerhalb und natürlich auch innerhalb der Sprache. Durch diesen dauernden Abgleich (*correlation*) von Werten, ihren Klassen und deren Bezeichnern wird Verständigung möglich. Es entstehen konventionalisierte und ständig neu abgegliche (dynamische) Mengen von Erfahrungsklassen. Sowohl die minimalen Einteilungen, die Bedeutungsgene, wie auch Kombinationen von Genen - Klassen von z.T. äü-

¹³ Der Abschnitt bezieht sich in der Hauptsache zusammenfassend auf *Psycho-Biology* (S. 287ff.) und *Human Behavior* (v. a. 5. Kapitel, S. 106ff.).

ßerster Komplexität - sind konventionell. Nicht die Minimalklassen, sondern der Prozeß der Wahrnehmungsklassifikation ist universal¹⁴.

Bedeutungsgene entstehen als Abstraktionsleistungen bei der Bildung von Wahrnehmungsklassen: „Kann die Wahrnehmung XY hier eingeordnet werden oder nicht? Warum nicht?“ Die Klassifikation ist relevant für das Überleben des Menschen in seiner Umwelt. Die Funktionalität, die jedes Element der wahrgenommenen Welt für den Menschen hat, ist Grundlage seiner Einordnung. Die Klassen sind funktionale Klassen, die Bedeutungsgene funktionale „Eigenschaften“ von Elementen. Dadurch können Elemente, die dem einen heterogen erscheinen, für jemand anderen völlig „gleich“ sein - bezogen auf die funktionale Relevanz dieser Dinge für die betreffende Person.

Nichtkonventionelle Klassenbildung wird als pathologisch bezeichnet: der kranke (z.B. schizophrene) Mensch denkt zu sehr auf sein *ego* bezogen und verkennt die Ökonomie der sozial sanktionierten Klassenbildung. Jedoch ist seine eigene, krankheitsbedingt „verzerrte“ Einteilung der Welt durchaus in sich ökonomisch strukturiert (vgl. Whitehorn & Zipf, 1943; *Human Behavior*: 284ff.).

Sprachliche Ausdrücke sind Bezeichner solcher Erfahrungsklassen. Je mehr Bedeutungsgene eine Klasse umfaßt, desto *spezifischer* ist der Ausdruck, der sie bezeichnet. Das Gegenteil sind *generische* Klassen und dazugehörige Ausdrücke, die mit sehr wenigen (möglichst gegen 1) Bedeutungsgenen korreliert sind.

Die Größe der Klassen, ihre Anzahl und ihre Spezifität sind ökonomisch organisiert, je nach den Bedürfnissen der Gruppe oder der Person. Unifikation und Diversifikation sind auch hier die formenden Kräfte. Sie bewirken ein charakteristisches Verhältnis spezifischer und generischer Ausdrücke (Klassensymbole) auf einer Sprachebene. Syntagmatische (und syntaktische, d.h. funktional gesteuerte) Kombination von Ausdrücken einer niedrigeren Ebene ermöglicht die Kodierung von mehr und spezifischeren Bedeutungen, ohne daß auf dieser Ebene ein Inventarszuwachs entsteht. Die Bedeutungen der Einzelausdrücke werden - konventionsgesteuert - zu neuen Bedeutungen kombiniert.

Bedeutungsüberschneidungen, d.h. Bedeutungsgene, die zwei Ausdrücke gemeinsam haben, ermöglichen die Ersetzung des einen durch den anderen Ausdruck (vgl. oben). Dies kann verschiedene Funktionen haben: Abkürzung, Emphase, Spezifikation (z.B. durch metaphorischen Gebrauch). Metaphern, ja, spontane Ersetzungen überhaupt, die noch nicht konventionalisiert sind, werden nur verstanden, wenn eine Schnittmenge von Bedeutungsgenen existiert, die funktional äquivalent sind.

Hat sich eine Ersetzung allerdings in ihrem neuen Gebrauch (d.i. in der neuen Bedeutung) etabliert, entwickelt sie eine Eigendynamik, die unabhängig vom alten Ausdruck abläuft. Die Bedeutung kann sich verschieben, Synonyme entstehen (und unterscheiden sich trotz ihrer „Bedeutungsgleichheit“), Primär- und Sekundärbedeutungen kommen zustande und vertauschen ihre Plätze. Bedeutungsklassen können sich spalten und je ein Ausdruck bezeichnet die eine und die andere neue Klas-

¹⁴ Daß dabei universal scheinende Klassen auftauchen, ist zu erwarten, aber nicht zwingend.

se.

Zipf (*Human Behavior*: 308ff.) unterscheidet drei Möglichkeiten der Beschreibungssprache für sein semantisches System:

- semantische Algebra (z.B. Boolesche Algebra)
- semantische Geometrie (z.B. ein semantischer Raum, in dem die Korrelation funktionaler Abstraktionen (= Bedeutungsgene und ihre Kombinationen) mit Wahrnehmungsklassen stattfindet)
- semantische Dynamik (*dynamic symbolism*): erklärend durch das Prinzip des geringsten Aufwands; Bildung empirischer Gesetze.

Sprachliche Bedeutung wird also aus der Benutzung von Ausdrücken durch Abgleich (*correlation*) von Wahrnehmungen und Erfahrungsklassen inferiert.

Die Verwendung von Ausdrücken geschieht „ökonomisch“ (Zipf, 1945b). Nach dem Werkstattmodell (Wörter etc. als möglichst ökonomisch einzusetzende Werkzeuge für „Aufgaben“) gibt es folgende Mittel zur ökonomischen Organisation der Werkzeuge in der Werkstatt. Grundlage ist die zu minimierende Gleichung „Energieaufwand = Masse mal Frequenz mal Weg“.

Zipf überträgt das Bild nicht minutiös auf die Sprache, die Übertragung der Vorstellungen ist von mir in Klammern dazugesetzt worden. Manche Mechanismen wurden im Abschnitt 7.1 schon genannt.

- Gleichverteilung des Gewichts auf alle Werkzeuge
- Verkleinerung der am häufigsten gebrauchten Werkzeuge bei Erhalt der Funktionalität
- Zusammenfassung von Funktionen in neuen „Universal“-werkzeugen (= Ersetzung durch Neubildungen)
- Übertragung von Funktionen auf schon vorhandene Werkzeuge: am ökonomischsten auf kleine, häufig benutzte Werkzeuge (= Ausdrücke; Erhöhung der Bedeutungszahl sowie Generalisierung)
- Suche nach Aufgaben, die mit den vorhandenen Werkzeugen gut gelöst werden können (= ökonomische Reklassifikation; Aktivierung alter (*senescent*) Ausdrücke)
- Anpassende Veränderung der Werkzeuge (Spezialisierung eines Ausdrucks)
- Wichtig ist dabei, nicht das momentan einfachste zu tun, sondern die „Zeitperspektive“ im Auge zu behalten: kurzfristig erhöhter Aufwand (z.B. Neubildung, Spezialisierung) kann sich hinterher vielfach bezahlt machen. Das Prinzip des geringsten Aufwands bezieht immer die Zeit mit ein.

Über Zusammenhänge der Spezifität eines Ausdrucks mit anderen Größen z.B. des lexikalischen Systems macht Zipf nur wenige Aussagen. Er nimmt zwar an, daß mit steigendem Kristallisationsgrad ein Ausdruck immer weniger „distinkt“ wird (*Psycho-Biology*: 155). Ob damit aber die steigende Bedeutungszahl oder die sinkende Spezifität gemeint ist, bleibt unklar.

Zum Zusammenhang mit der Häufigkeit läßt sich sagen, daß generische Ausdrücke häufiger gebraucht werden, denn jedesmal, wenn ein seltenes Ereignis bezeichnet werden muß, das keinen eigenen bezeichnenden Ausdruck auf der entsprechenden linguistischen Ebene hat, bilden eher generische Kategorien durch Kombination ihrer Bezeichner einen Ausdruck auf einer höheren Ebene für die seltenere und spezifischere Kategorie.

Viel mehr Augenmerk legt Zipf im Zusammenhang mit Spezifität und Generalität auf die verschiedenen linguistischen Analyseebenen und ihr Verhältnis zueinander: je kleiner die Elemente sind, desto unspezifischer sind sie, was sich in einem kleineren Inventar und steileren *n-f*- und *r-f*-Kurven bemerkbar macht. Wie später gezeigt wird, sind steilere Kurven Hinweise auf geringere Diversifikation.

Das heißt, Komposita sind spezifischer als Simplizia (immer „durchschnittlich“), Wortformen sind spezifischer als Lexeme, und diese sind wiederum spezifischer als Morpheme ... ; Phoneme haben keine sehr spezifischen Bedeutungen¹⁵. Die Beispiele und Abbildungen dazu folgen in Abschnitt 8.

8. Syntagmatische Beziehungen: Die „Konfiguration“ sprachlicher Elemente bei G.K. Zipf¹⁶

Die Elemente jeder sprachlichen Ausdrucksebene konfigurieren sich zu Elementen einer höheren Ebene. Es gibt keine unabhängige höchste Ebene. Aus den wahrgenommenen Konfigurationen inferieren die Benutzer einer Sprache die Muster, nach denen sie wiederum Konfigurationen bilden.

Die Konfigurationen haben unterschiedliche relative Häufigkeit, emotionale Intensität, semantische Spezifität und einen jeweiligen Kristallisationsgrad (*Psycho-Biology*: 185ff.).

Emotionale Intensität bezieht sich auf den „Überraschungseffekt“ seltener Zusammenstellungen von Elementen (Zipf: *permutations*). Eng damit verbunden ist ein hoher Informationswert dieser Konfiguration.

Semantische Spezifität ergibt sich wie im vorigen Abschnitt als Eigenschaft der Bedeutung des sprachlichen Zeichens. Je seltener ein Ereignis wahrgenommen und in seiner funktionalen Relevanz klassifiziert werden muß, desto spezifischer ist die Klasse, in die es eingeordnet wird. Es ist als Kombination häufigerer, generischerer Bedeutungen beschreibbar.

Der Kristallisationsgrad beschreibt die Festigkeit der Konfiguration, wie z.B. auf der Morphemebene: ist das Präfix eines deutschen Verbs abtrennbar oder nicht? Die Distribution von Phonemen und die Freiheit der Wortstellung in der Syntax können den Kristallisationsgrad charakterisieren.

Die Unabhängigkeit eines Elements oder einer Klasse sprachlicher Elemente

¹⁵ Natürlich haben sie „Bedeutungen“: lautliche Reize werden bezüglich ihrer distinktiven Funktion klassifiziert.

¹⁶ Die folgende Darstellung bezieht sich hauptsächlich auf Zipfs *Psycho-Biology* (S. 185 ff.), und *Human Behavior* (S. 73ff.).

bezeichnet, wie sehr oder wie wenig es im Mittel auf Konfiguration angewiesen ist, um Bedeutung haben zu können.

Zipfs Grundannahme ist nun, daß die Häufigkeit einer Konfiguration ihre syntaktischen, psychologischen und semantischen Eigenschaften, nämlich Kristallisationsgrad, Intensität und Spezifität bestimmt. Diese drei Größen lassen sich ineinander umrechnen (*Psycho-Biology*: 201), da sie in Gleichgewichtsbeziehungen zueinander stehen.

Im einzelnen sind die Hypothesen zu den Gleichgewichtsbeziehungen:

- a) Je häufiger eine Konfiguration vorkommt, desto fester (kristalliner) wird sie (*Psycho-Biology*: 157).
- b) Je seltener eine Konfiguration ist, desto intensiver, spezifischer und unabhängiger ist sie (*Psycho-Biology*: 191).

Damit gilt auch:

- c) Je fester (kristalliner) eine Konfiguration ist, desto weniger emotional intensiv (desto erwarteter) ist sie (*Psycho-Biology*: 208).
- d) Je fester (kristalliner) eine Konfiguration ist, desto weniger spezifisch ist sie (*Psycho-Biology*: 157).
- e) Je häufiger ein Element ist, desto stärker ist seine Beteiligung an unterschiedlichen Konfigurationen (*Human Behavior*: 71).
- f) Je häufiger ein Element ist, desto länger sind die Konfigurationen, an denen es beteiligt ist (*Human Behavior*: 71),
- g) und desto häufiger¹⁷ sind die Konfigurationen, an denen es beteiligt ist (*Human Behavior*: 71).

Die dynamischen Faktoren Unifikation und Diversifikation wirken gleichzeitig in der Weise, daß sich die o.g. Zusammenhänge ergeben müssen.

Zu den Unifikationskräften gehören die Gewohnheit und die Trägheit (*Psycho-Biology*: 187f.). Sie sind im Allgemeinen den Sprecherinteressen zugeordnet.

Die Gewohnheit und die Trägheit wirken umso stärker, je häufiger eine Konfiguration ist (der Einfluß der Hörerinteressen läßt nach). Dadurch werden Konkurrenten einer Konfiguration umso leichter verdrängt, je häufiger diese Konfiguration ist. Außerdem festigt sich die Konfiguration psychologisch durch häufigeren Gebrauch.

Die Diversifikationskräfte sind im vorliegenden Problembereich Spezifikation und Intensivierung (*Psycho-Biology*: 209f.). Sie gehören zu den Hörerinteressen.

Spezifikation bewirkt, daß Konfigurationen länger, aber bckerer und seltener werden. Intensivierung bewirkt neue Kombinationen mit geringer Erwartung. Diese sind dadurch (noch) nicht kristallisiert.

Außerdem stellt die Bildung von Zusammensetzungen eine ökonomische Aus-

¹⁷ (*Sic!* unter (B) am a. O.) Die Vermutungen f) und g) widersprechen der Hypothese von Länge und Frequenz.

nutzung des Inventars dar (*Human Behavior*: 87). Zusätzliche Bedeutungen können ohne Zuwachs am Inventar ausgedrückt werden. Die Häufigkeit, mit der eine Funktion erfüllt werden muß, bestimmt die Größenordnung der sprachlichen Ebene, auf der sie erfüllt wird.

Durch die Kombination einer geringen Anzahl sehr häufiger Elemente mit einer größeren Anzahl seltener Elemente wird unter anderem auf Wortbildungsebene morphologische Transparenz gewährleistet (*Human Behavior*: 88).

Der Nachweis der o.g. Gleichgewichtszusammenhänge geschieht einerseits durch die Rang-Frequenz-Verteilung in der Form

$$(8.1) \quad r \cdot f^a = \text{const.}$$

wobei r die Rangzahl des Elements mit der Frequenz f ist. Der Parameter a im Exponenten wird als Maß für die emotionale Intensität interpretiert, denn die Steilheit der Kurve zeigt das Verhältnis von Diversifikations- und Unifikationskräften an. Ist die Kurve steil, also die Diversifikation in $n = r_{max}$ Elemente gering im Vergleich zur Häufigkeit f_{max} an der Stelle $r = 1$, so sind die Ausdrücke auch weniger intensiv und weniger spezifisch. Dies zeigt sich im Vergleich der Kurven für die unterschiedlichen Analyseebenen der Sprache. Ein Beispiel zeigt Abbildung 8.1.

Auch bei der Frequenzklassenbelegung sind die Kurven für die „kleineren“ Einheiten steiler. Abbildung 8.2. zeigt ein Beispiel Zipfs aus *Psycho-Biology*.

Vergleicht man die n - f -Kurven von Sprachen mit fester Wortstellung und von flektierenden Sprachen miteinander, so stellt man fest, daß die Kurven für Wörter und Morpheme sich bei fester Wortstellung fast entsprechen, bei flektierenden Sprachen aber deutlich voneinander abweichen. Da

$$(8.2) \quad n \cdot f^2 = \text{const.}$$

als ideale Kurve für die Lexik gilt, kann die Differenz zwischen dem empirischen Exponenten der Morphemkurve und der „2“ als Maß für die Flektiertheit der Sprache interpretiert werden.

Für Französisch ergibt sich die Abweichung 0.61 und für Chinesisch 0.22 (*Psycho-Biology*: 258). Die Kurven sind oben (vgl. Abbildungen 8.2 und 8.3) dargestellt.

Zipf (*Psycho-Biology*: 259f.) schildert den großen Kreislauf der Sprachentwicklung, von fester Wortstellung über die Kristallisation bestimmter Syntagmen hin zu Wörtern mit Affixen und relativ freier Wortstellung im Satz; nach weiterer Abkürzung der flektierten Wörter, und wenn die Affigierung funktional nicht mehr ausreicht, festigt sich die Wortstellung erneut und der Kreislauf schließt sich.

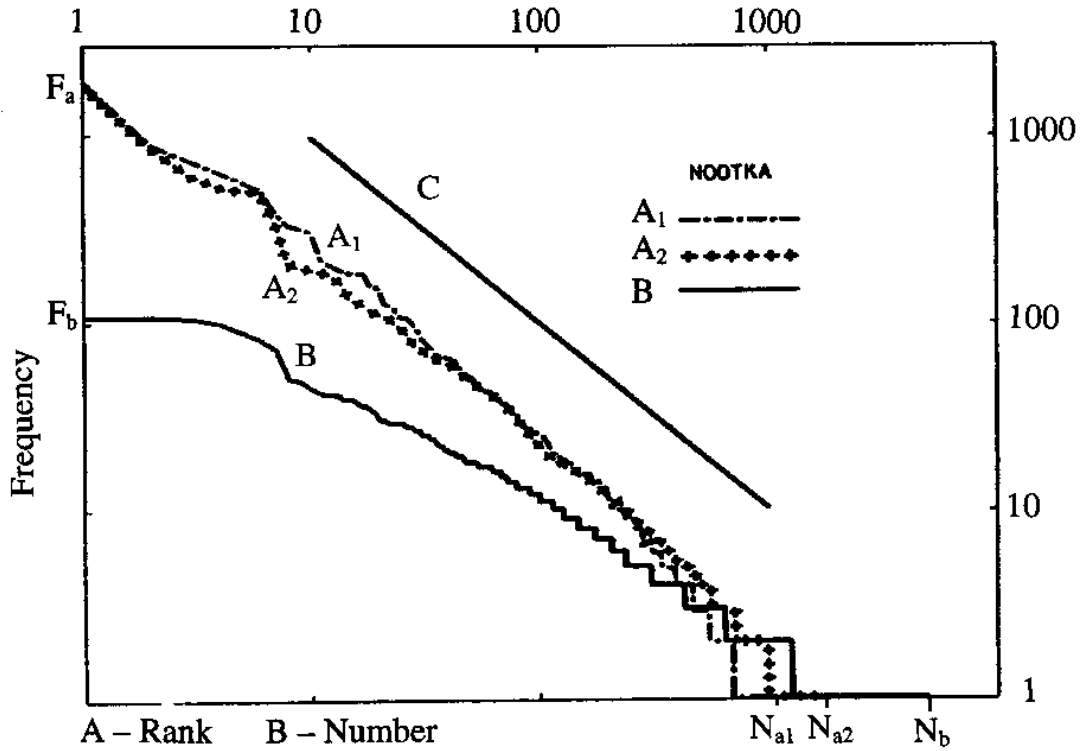


Abb. 8.1. Rang-Frequenzverteilungen für verschiedene Einheiten in Nootka. A₁: varimorphs¹⁸, A₂: Morpheme, B: holophrases¹⁹, C: theoretische Kurve für Wörter. Nach Zipf (Human Behavior: 80).

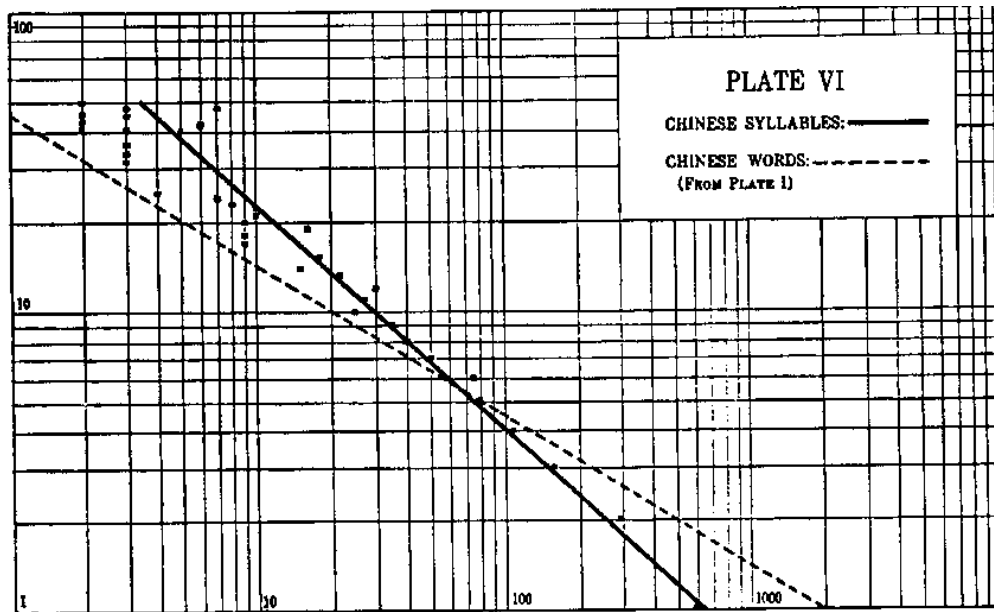


Abb. 8.2. n-f-Verteilung für chinesische Silben (empirische Werte) und theoretische Kurven für Silben und Wörter. Nach Zipf (Psycho-Biology).

¹⁸ Entspricht *variphones* auf der Morphemebene.

¹⁹ Kombination von Wörtern, zwischen Komposita und festen Syntagmen anzusiedeln.

Für die Syntax, die Konfiguration von Wörtern zu Sätzen, ergibt sich eine Reihe von Folgerungen. So entstehen Wortarten (*Psycho-Biology*: 224ff.) als funktionale Klassifikationen, die aus den Gebrauchshäufigkeiten und dem Verhalten der Wörter in der Struktur gebildet werden. Feste Wortstellung kann allerdings dasselbe leisten wie die Wortartenunterscheidung in der flektierenden Sprache: „Though parts of speech are not essential to sentence-structure, their existence reflects peculiarities of sentence-structure in which they have their origin“ (*Psycho-Biology*: 233).

Nach der Frequenz können grob zwei Wortklassen unterschieden werden: „Konzeptwörter“, die seltener sind, aber deren Klasse viele Elemente umfaßt, und „artikulatorische Wörter“, d.h. häufige Wörter, die die Relationen der Konzeptwörter zueinander festlegen. Die Bedeutung der Konzeptwörter liegt größtenteils in der außersprachlichen Erfahrungswelt, während die Bedeutung der artikulatorischen Wörter sich erst aus der Texterfahrung ergibt. Zipf (1938c) macht die Beobachtung, daß Schüler einer Fremdsprache zu übermäßigem Gebrauch von Konzeptwörtern tendieren, bevor sie Sicherheit im Gebrauch der artikulatorischen Wörter erlangen. In Tabelle 8.1 ist Zipfs Beispiel für die Wortartenverteilung im Englischen wiedergegeben.

Zipf nimmt außerdem an, daß die Wortart mit dem „Charakter“ dessen, was sie bezeichnet, zusammenhängt: Dinge, die sich verändern und Eigenschaften haben. Dabei weist er aber relativierend auf den „Aspekt“ hin, unter dem ein Wort in dem jeweiligen Zusammenhang klassifiziert wird.

Subjekt und Verb fügen sich zur Konfiguration „Satzkern“ (*Psycho-Biology*: 233 ff.). Fehlt ein Element, so wird es impliziert. Subjekt und Verb bilden wiederum Kerne, um die sich weitere Elemente gruppieren können. Der Satz ist also eine Konfiguration, die durch größere Unabhängigkeit, geringere Kristallisation, höhere Spezifität, Seltenheit und größeren Umfang charakterisiert wird - natürlich immer „im Durchschnitt“ und im Verhältnis zu den Elementen, aus denen er zusammengesetzt wird. Im Einzelfall kann auch ein Wort ein beliebig komplexes Konzept ausdrücken.

Die Wortstellung (*Psycho-Biology*: 239ff.) hängt mit dem emotionalen Eindruck einer Phrase so zusammen, daß das abhängigste, unwichtigste Element an erster Stelle steht, gefolgt von den unabhängigeren, wichtigeren (selteneren) Elementen. Zipf meint, das mit Beispielen aus dem Deutschen und Englischen belegen zu können, z.B. steht der Artikel in diesen Sprachen vor dem Substantiv/Nomen; Sätze der Art „*I shall go*“/„Ich werde gehen“ sollen ebenfalls nach diesem Prinzip gebaut sein.

Allgemein hat jede Konfiguration charakteristische Elemente (*Psycho-Biology*: 280ff.) Das sind die Elemente, deren Häufigkeit innerhalb einer Konfiguration deutlich höher ist, als insgesamt. Hat z.B. die Konfiguration XY eine Frequenz von 50, X insgesamt eine Frequenz von 52, Y eine Frequenz von 1034, so stellt X das charakteristische Element der Konfiguration XY dar. Bei Abkürzungen bleiben die charakteristischen Elemente eher erhalten, sofern sie die Information der gesamten Konfiguration allein übernehmen können.

Tabelle 8.1.

Wortartenverteilungen für amerikanische Telefongespräche. Nach Zipf (*Psychobiology*: 227).

Wortart	Häufigkeiten		Quotient
	insgesamt (tokens)	verschiedene (types)	$\frac{\text{tokens}}{\text{types}}$
<i>Substantives</i>	11 660	1 029	11.3
<i>Adjectives and Adverbs</i>	9 880	634	15.6
<i>Main Verbs</i>	12 550	456	27.5
<i>Auxiliary Verbs</i>	9 450	37	255.
<i>Pronouns</i>	17 900	45	398.
<i>Prepositions and Conjunctions</i>	12 400	36	344.
<i>Articles</i>	5 550	3	1 850.
Insgesamt	$E_1=79\ 390$	$E_2=2\ 240$	$E_1/E_2 = 35.4$

9. Inventar-Ökonomie und sprachliches Gleichgewicht

Die ökonomische Gestaltung eines Inventars sprachlicher Einheiten ist auf allen Ebenen eine wichtige Triebfeder für dynamische Prozesse. Zur Inventarökonomie gehören für Zipf die Optimierung der Inventargröße im Spannungsfeld funktionaler Unifikation und Diversifikation der einzelnen Elemente und die Strukturierung des Inventars in Frequenzklassen (nach Rang und Klassengröße), die im folgenden Abschnitt beschrieben wird. Im Abschnitt 7.2 (Hypothese des semantischen Gleichgewichts) wurden Mittel zur Optimierung der Verwendbarkeit von Elementen aufgezeigt, die zu einer aufwandsparenden Strukturierung des Inventars und der Verwendung der einzelnen Elemente führt. Die Kombination von Elementen ist ein weiteres wichtiges Mittel zur Optimierung der Inventargröße.

Der Einfluß der Inventarminimierung auf verschiedene Eigenschaften der Elemente im Inventar kann folgendermaßen zusammengefaßt werden (vgl. *Human Behavior*: 67ff.):

Inventarminimierung bewirkt durchschnittlich

- höhere Beteiligung der Elemente an syntagmatischen Kombinationen (auch indirekt über geringere Spezifität der Bedeutungen)
- niedrigere Spezifität
- höhere Bedeutungszahl
- kleineres Inventar
- indirekt: kleinere Elemente.

Abbildung 9.1 zeigt zusammenfassend das System der linguistischen Hypothesen G.K. Zipfs, wie sie bisher vorgestellt wurden. Die Darstellung ist an die linearen Graphen Köhlers (1986) angelehnt, wird aber nicht in überprüfbare mathematische Hypothesen übersetzt. Das Hypothesengebäude Zipfs ist bis auf wenige Stellen (mit „?“ markiert) konsistent, die Vorzeichen der indirekten und direkten Zusammenhänge stimmen überein, wenn man die verschiedenen Pfade zwischen zwei Systemgrößen (z.T. auch entgegen der Pfeilrichtung) betrachtet.

10. Die Zipfschen Gesetze

Als Zipfsche Gesetze werden außer den hier vorgestellten Zusammenhängen auch das Prinzip der relativen Häufigkeit (Birkhan, 1979) oder die Zusammenhänge von Bedeutungszahl und Länge bzw. Frequenz bezeichnet. Die vorliegende Arbeit verzichtet auf eine Numerierung, vielmehr dienen die beteiligten Größen (Anzahl n , Frequenz f , Rang r , Größe s etc.) als Bezeichner und Kürzel für die jeweiligen Zusammenhänge. Sie werden in der Reihenfolge ihrer Veröffentlichung behandelt. Eine andere übliche Bezeichnung ist *frequency distribution* für die Frequenzklassenbelegung und *frequency structure* für die Rang-Frequenz-Verteilung (Zörnig & Boroda, 1992).

10.1. n - f -Verteilung (Frequenzklassenbelegungsgesetz, Häufigkeitsverteilung)

Was Zipf *number-frequency relationship* nennt, soll hier mit dem Kürzel „ n - f -Verteilung“ bezeichnet werden. Die Bezeichnung nennt die beteiligten Größen n und f in der Reihenfolge, in der Zipf sie auf der x - bzw. y -Achse abträgt. Dieser Grundsatz wird auch in den Abschnitten 10.2 und 10.3 beibehalten.

Die n - f -Verteilung dient zunächst nur der Verdeutlichung einiger Annahmen Zipfs. Durch seine Linearität auf doppelt logarithmischen Achsen soll der Datenverlauf die Leser von der Überzufälligkeit der Daten und von der Gesetzmäßigkeit der zugrundeliegenden Annahmen überzeugen. Zipf wählt diese Darstellungsweise zum ersten Mal in *Selected Studies*, als es um die Gebrauchshäufigkeiten von Wörtern verschiedener Längen geht (vgl. Tabelle 3.1, dritte Spalten). Abbildung 10.1 zeigt zwei Graphiken Zipfs.

Es verwirrt am Anfang, daß Zipf die Zufallsvariable f auf der y -Achse und die Häufigkeit n des Ereignisses f („Wort mit der Frequenz f “) in der Stichprobe auf der x -Achse abträgt. Die Erklärung, es handle sich möglicherweise um einen funktionalen Zusammenhang („Die Frequenz eines Wortes hängt von der Anzahl Wörter ab, die diese Frequenz haben“) ist nicht Zipfs Aussage. Ihm geht es einfach um die „Korrelation“. Da ein funktionaler Zusammenhang mit f als abhängiger Variablen für die n - s -Verteilungen nicht haltbar ist (vgl. Abschnitt 10.3), ist er vermutlich auch im linguistischen Zusammenhang zu verwerfen.

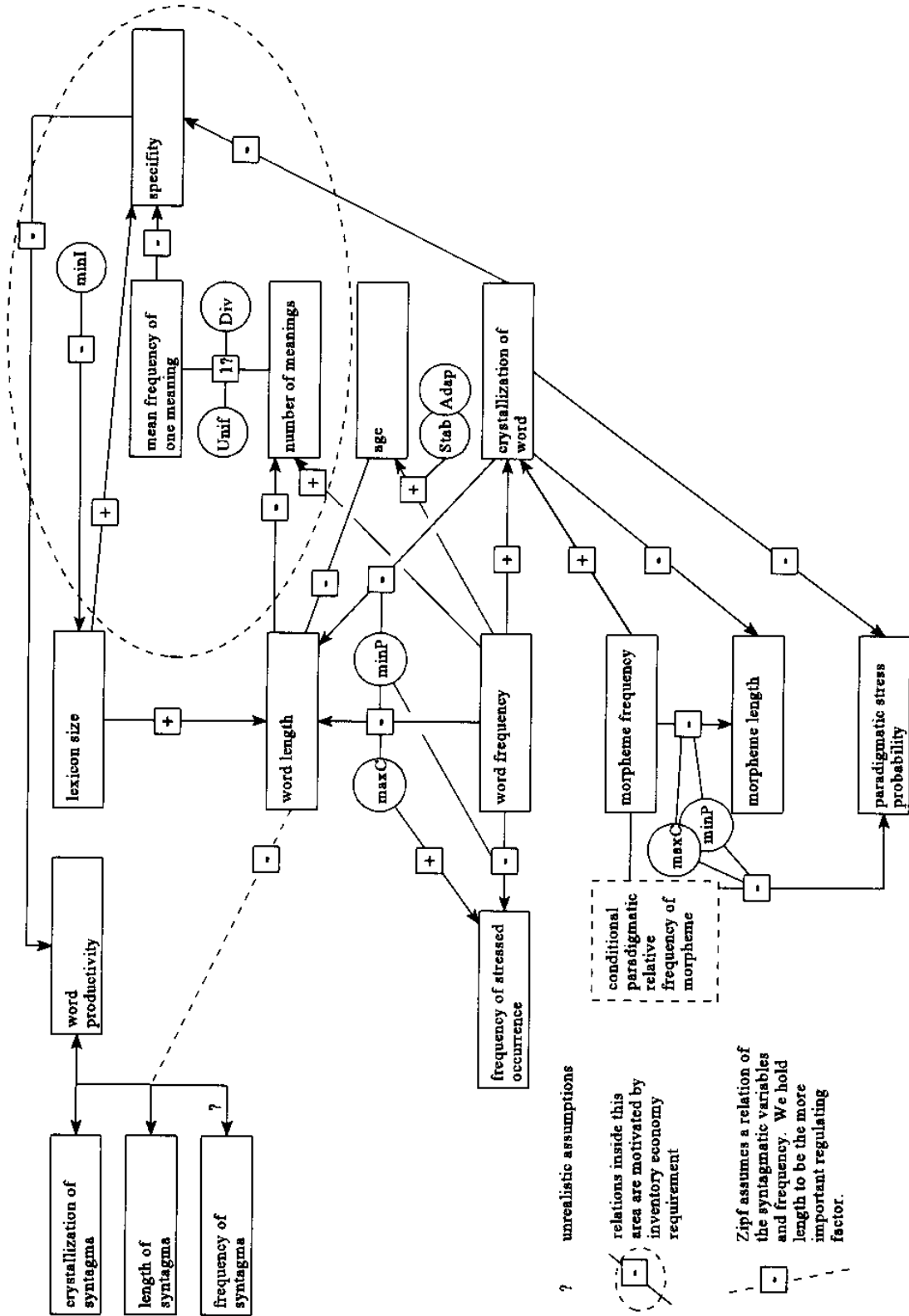


Abb. 9.1. Zusammenschau der linguistischen Hypothesen G.K. Zipfs.

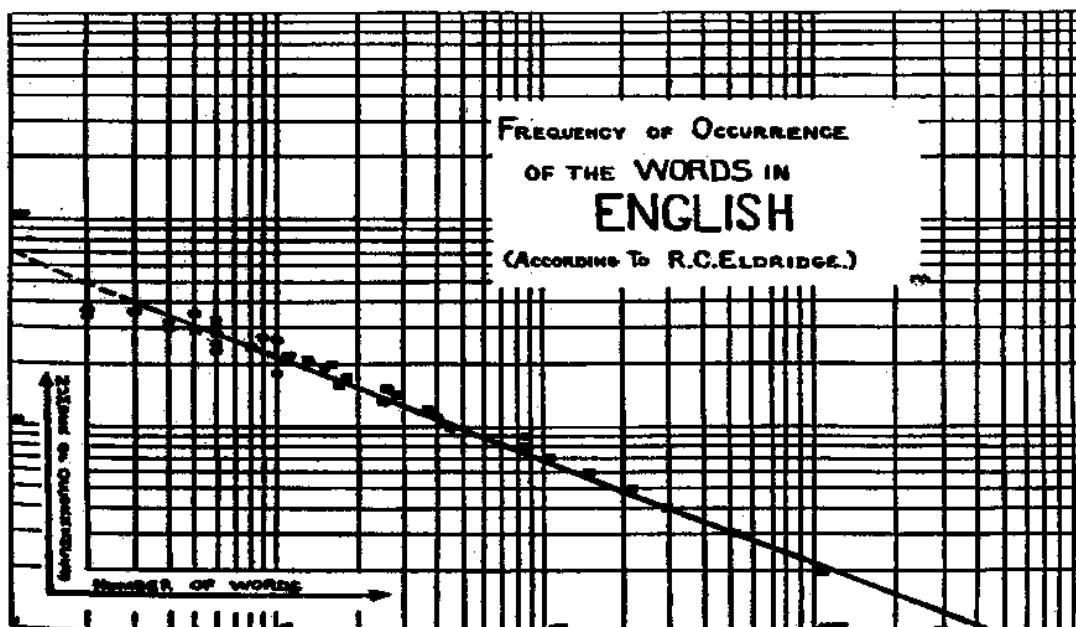
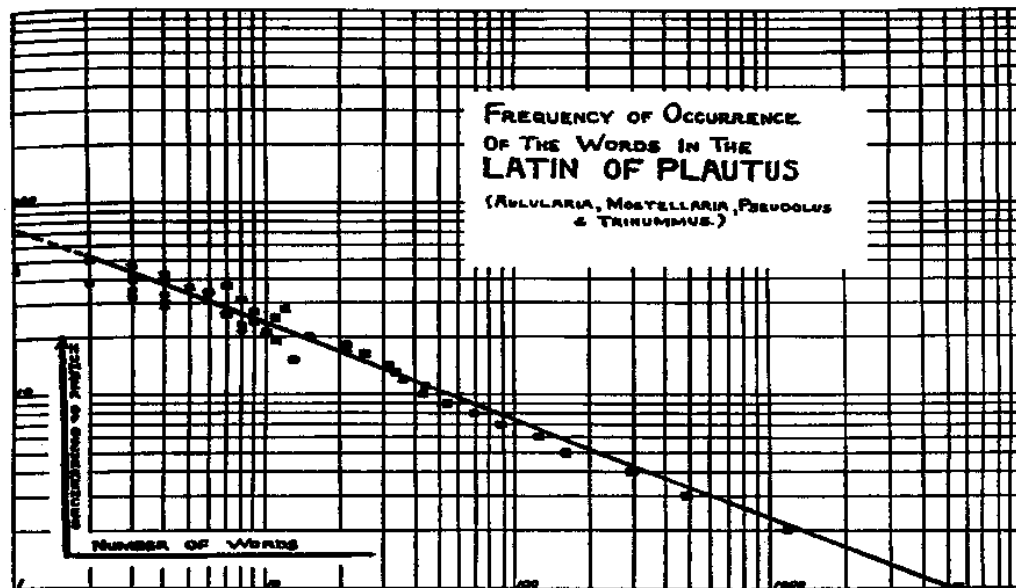


Abb. 10.1. n - f -Verteilungen der Lexik für Latein (Plautus) und Englisch. Logarithmische Achsen. Nach Zipf (*Selected Studies*: 25).

Die n - f -Verteilung gilt für den größten Teil der Daten, nicht jedoch im Bereich kleiner n (großer f) (*Psycho-Biology*: 43). Ihre Formel lautet (*Selected Studies* etc.):

$$(10.1) \quad n \cdot f^2 = \text{const.}$$

mit n als Anzahl der Elemente, die die Häufigkeit f haben. Eine allgemeinere Form lautet (*Psycho-Biology*: 258):

$$(10.2) \quad n \cdot f^q = \text{const.}$$

Eine andere Fassung der Formel (Zipf, 1938b) versucht, den Stufen der empirischen Verteilung Rechnung zu tragen:

$$(10.3) \quad n_f \left(f^2 - \frac{1}{4} \right) = c.$$

Diese Verfeinerung ist nur im Bereich höherer Klassengrößen n überhaupt wirksam; die Abweichung von 3 ist sonst verschwindend gering.

Zipf leitet die n - f -Verteilung unter anderem aus dem Werkstattmodell ab. Der Handwerker ordnet diejenigen Werkzeuge, die er am häufigsten braucht, in seiner direkten Nähe an. Die ökonomischste Anordnung der Werkzeuge, bei der dann auch das Verhältnis von Verwendungsfrequenz und Distanz konstant ist, ist die auf konzentrischen Kugelflächen um den Handwerker, wobei der ideale Handwerker ein Punkt ist. Wachsen die Distanzradien um einen bestimmten Betrag, so wachsen die Kugelflächen um das Quadrat dieses Betrages. Die Anzahl Werkzeuge auf einer Kugelfläche ist proportional zur Fläche selbst. Daraus ergibt sich, daß die Anzahl N der Werkzeuge mit der gleichen Distanz proportional zum Quadrat der Verwendungshäufigkeit ist (Zipf, 1940b).

Der Betrag des Exponenten sinkt mit der Stichprobengröße und die Kurve wird steiler (*Psycho-Biology*: 43). Die Feststellung wird aber nicht weiter untersucht.

In (1937b) rechtfertigt Zipf seine Denkweise und Methode gegenüber Joos (1936). Dabei zitiert er auf Seite 64 M.H. Stone, der den Zusammenhang des Wertes für den Exponenten mit dem Type-Token-Quotient einer Stichprobe herstellt. Indem nämlich die Anzahl Types v als Summe aller Elemente in den Frequenzklassen (k : Konstante der n - f -Verteilung; Frequenz f , $f_{max} = m$)

$$(10.4) \quad v = k \sum_{f=1}^{f=m} f^{-q}$$

und die Anzahl Tokens l als²⁰

$$(10.5) \quad l = k \sum_{f=1}^{f=m} f^{1-q}$$

berechnet werden können, darf für eine sinnvolle Stichprobe q nicht viel größer sein als 2, da der Quotient $\frac{v}{l}$ klein sein muß.

Eine andere Interpretation wurde schon vorgestellt. Sie besagt, daß der Exponent sich mit der Analyseebene verändert, d.h. bei kleineren Einheiten sind die n - f -Kurven steiler, bei größeren flacher, und daß die Flektiertheit einer Sprache Einfluß auf den Exponenten hat (vgl. Abschnitt 8.1.6).

Auch das geistige Alter beeinflusst den empirischen Wert für q , weshalb Zipf

²⁰ Die Gleichung für l leitet sich aus der r - f -Verteilung ab.

(1937a) ihn als Intelligenzmaß vorschlägt.

10.2. *r-f*-Verteilung (Rang-Frequenz-Verteilung, Häufigkeitsstruktur, das Zipfsche Gesetz)

Genau wie die *n-f*-Verteilung ist die *r-f*-Verteilung zunächst als Argumentationshilfe gedacht. Sie läuft der *n-f*-Verteilung bald den Rang ab, in *Psycho-Biology* (S. 44) wird sie als „Standardkurve“ bezeichnet. Die Linearität der Daten auf doppelt logarithmischen Achsen soll den Leser von ihrer Überzufälligkeit überzeugen. Die Annahmen sind aus dem Prinzip der relativen Häufigkeit und der *n-f*-Verteilung abgeleitet und sollen diese unterstützen. Abbildung 10.2 zeigt zwei repräsentative *r-f*-Verteilungen.

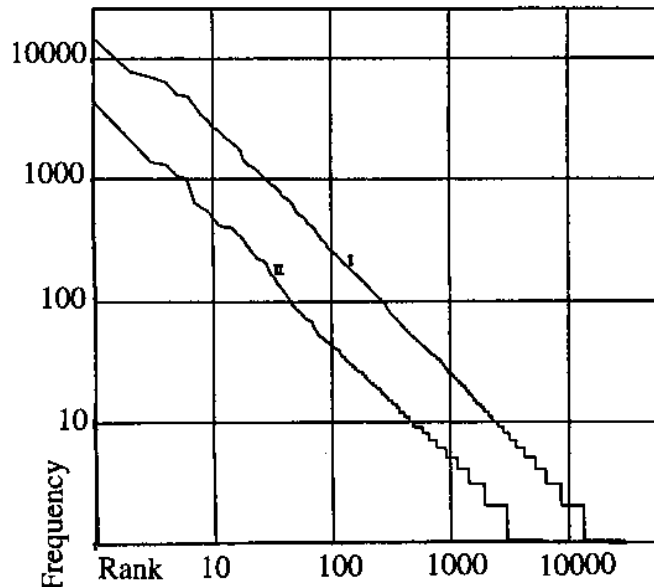


Abb. 10.2. Rang-Frequenz-Verteilungen in *Ulysses* von James Joyce (I) und für Stichproben amerikanischer Zeitungstexte nach R.C. Eldridge: Six thousand common english words, Buffalo, 1911; 43.989 Wörter (II). Logarithmische Achsen. Nach Zipf (1946f.:336).

Die erste Formulierung der Rang-Frequenz-Beziehung ist der „Wellenlängen-zusammenhang“ (*Psycho-Biology*: 45; vgl. Abschnitt 11)

$$(10.6) \quad \frac{N}{f_w} = 10r_w .$$

Die Wellenlänge N / f_w ist der reziproke Wert der relativen Häufigkeit (Stichprobenumfang N , Häufigkeit f_w eines Wortes w mit dem Frequenzrang r_w) und gibt an, wieviele Wörter im Mittel zwischen den einzelnen Tokens von w vorkommen. Daraus ergibt sich der direkt proportionale Zusammenhang als Verallgemeinerung

(*Human Behavior*: 24).

$$(10.7) \quad r \cdot f = \text{const.}$$

wobei die Konstante in (10.6) $N / 10$ beträgt.

Gleichbedeutend ist die Formulierung als harmonische Reihe der relativen Häufigkeiten, die für Zipf (z.B. 1937a) noch aussagekräftiger ist:

$$(10.8) \quad 1 = \frac{1}{10} \left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right)$$

Die Nenner der Brüche in der Klammer sind die Ränge $r = 1, 2, 3, \dots, n$; n ist die Anzahl unterschiedlicher Wörter (*types*) in der Stichprobe. Steht Sn („*sum n*“) für die Summe der harmonischen Reihe $\sum_{r=1}^n \frac{1}{r}$, dann ist die Konstante allgemein aus N / Sn zu schätzen. Sie verändert sich mit der Wiederholungsrate einer Stichprobe, wird umso kleiner, je stärker die einzelnen Wörter wiederholt werden (d.h. je kleiner das Type/Token-Verhältnis wird). Zipf hielt die Konstante $1/10$ zunächst für wesentlich für die Harmonie einer Sprache. Die „ausmultiplizierten“ Brüche $1/(10r)$ geben im Nenner die erwartete Wellenlänge des Wortes mit dem Rang r an. Später beschäftigt der Wert der Konstante Zipf kaum noch.

Zipf (1940a) wendet eine verallgemeinerte Form der harmonischen Reihe an, da der Stichprobenumfang durch die Konvergenz der Reihe begrenzt ist:

$$(10.9) \quad F \cdot Sn = \frac{F}{1^p} + \frac{F}{2^p} + \frac{F}{3^p} + \dots + \frac{F}{n^p}$$

Dabei ist F die Frequenz des häufigsten Wortes ($r_F = 1$). Der Parameter p für die ideale Verteilung des Wortschatzes („Standardkurve“) ist gleich 1, denn in diesem Fall ist $r_{\max} = f_{\max}$, d.h. der höchste Rang n ist gleich der höchsten Frequenz F und $F / n^p = 1$ ist die Häufigkeit des seltensten Wortes.

Die Gleichung (10.7) ist eine Tautologie. Besser formuliert wäre (N als Stichprobengröße):

$$(10.10) \quad N = F \sum_{r=1}^n \frac{1}{r^p}$$

Der Parameter p aus der Summen-(Reihen-)schreibweise findet sich in der r - f -Verteilung in folgender Form ($c = \text{const.}$):

$$(10.11) \quad r \cdot \frac{1}{f^p} = c.$$

Unter den Prinzipien und Analogien, aus denen Zipf die r - f -Verteilung herleitet, sind für die synergetische Linguistik Unifikation und Diversifikation die wichtigsten.

Weil beide Kräfte als Interessen von Sprecher und Hörer gleichzeitig wirken, ergibt sich eine bestimmte Anzahl n unterschiedlicher Elemente und eine „ideale“ höchste Frequenz an der Stelle $r = 1$, sowie die genannte Verteilung zwischen $r = 1$ und $r_{max} = n$.

Der Exponent der Verteilung wird dann als Quotient aus der Wirkung von Diversifikation und Unifikationskraft interpretiert (*Human Behavior*: 131):

$$(10.12) \quad p = \frac{Divs.}{Unif.}$$

Zipf sieht die Verteilung als Nachweis einer semantischen Balance, also des Gleichgewichts zwischen Gebrauchshäufigkeit und Inventarumfang sprachlicher Ausdrücke an, wobei die Größe der Elemente mit beiden Systemgrößen ebenfalls eng zusammenhängt. Daraus abgeleitet wird eine typologische oder Ebenenabhängige Interpretation des Parameters über den Type-Token-Quotient möglich, ebenso wie bei der n - f -Verteilung (vgl. Abschnitt 8).

Auch hier beeinflußt das (geistige) Alter den Parameter p (die Neigung der log-log-Geraden) (Zipf, 1942a). Bei neurotischen Patienten finden sich typische pathologische Formen extremer Unifikation (steile Kurven, kleines p ; Whitehorn & Zipf, 1943 und *Human Behavior*: 288ff.).

Die harmonische Reihe erlaubt einige Folgerungen über die Struktur der Population, für die sie gilt. So kann eine Menge von Elementen, die nach der harmonischen Reihe organisiert („homogen“) ist, nicht in Teilmengen aufgeteilt werden, die bezüglich derselben Kriterien ebenfalls „harmonisch“ organisiert sind. Ist das Ganze homogen, so sind es seine Teile nicht, bzw. zwei homogene Mengen oder Stichproben können zusammengefaßt nicht eine neue homogene Menge ergeben (Zipf, 1938b; das in *National Unity*: 36ff. gesagte gilt sinngemäß auch für die Sprache).

Außerdem bildet die Konvergenz der Summenreihe S_n möglicherweise eine Grenze der menschlichen Kapazität zur Informationsverarbeitung ab (Zipf, 1937b). Die Sättigung der Reihe bildet ferner eine Wachstumsgrenze: Durch die (Token-) Anzahl N der Elemente ist die maximale Diversifikation in n Types (Klassen) festgelegt, wenn der Exponent p fest ist.

10.3. n - s - und r - s -Verteilungen

Für manche Untersuchungsgegenstände muß anstatt der Frequenz f die Größe (*size*) s als Zufallsvariable für die Rangierung betrachtet werden. Dies ist z.B. für Intervalle der Wortwiederholung (s. Abschnitt 11) oder für manche soziologischen oder siedlungsgeographischen Fragestellungen der Fall (es gibt keine Frequenz von Ortschaften, wie es die Wortfrequenz gibt). Die Form der n - s - und r - s -Verteilungen ist ähnlich wie die der n - f - und r - f -Verteilungen; die Formeln sind dieselben allgemeinen Ausdrücke. Die Größenordnung der Parameter (Geradenneigung) wird im gegebenen Fall abgeleitet. Abbildung 10.3 zeigt eine typische Siedlungsgrößenverteilung, Abbildung 10.4 eine Einkommensverteilung.

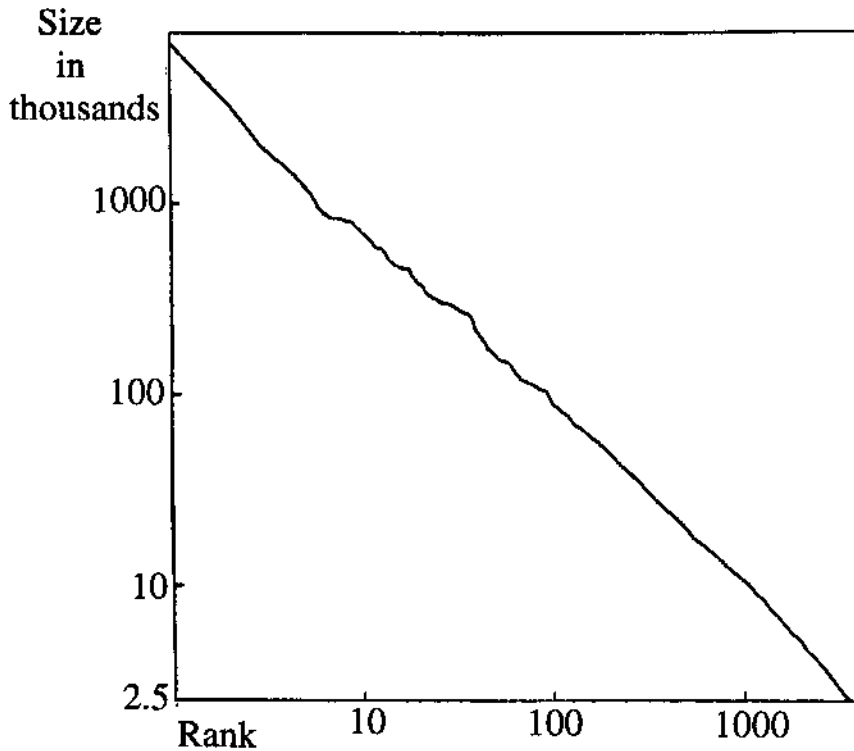


Abb. 10.3. r - s -Verteilung von Ortschaften mit mehr als 2500 Ew., nach Einwohnerzahl rangiert (U.S.A. 1930). Logarithmische Achsen. Nach Zipf (National Unity: 12).

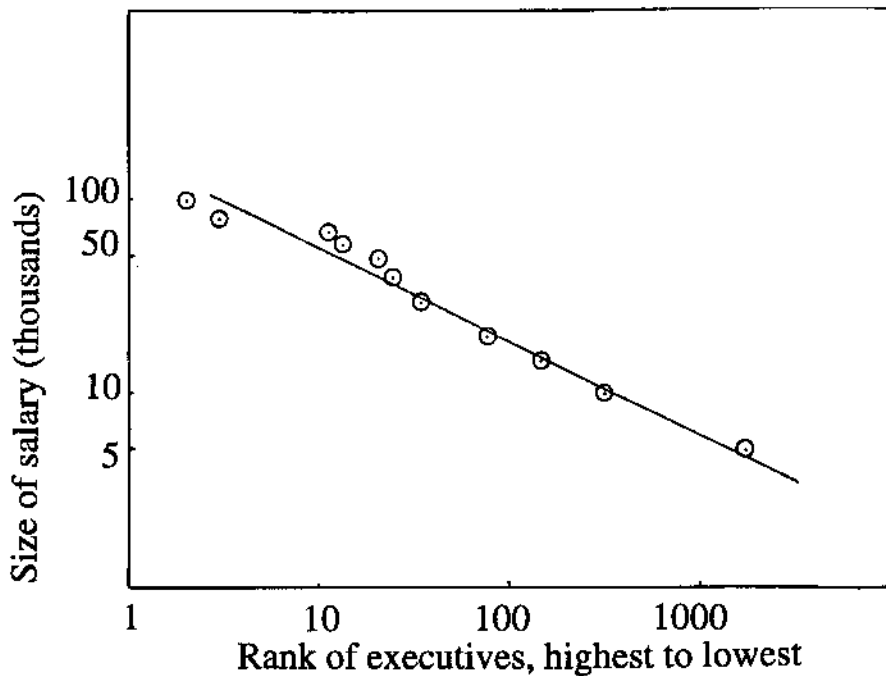


Abb. 10.4. Einkommensverteilung bei Führungskräften (General Motors, 1928). Nach Zipf und Rucker (1949:5).

Die Rangverteilung von Einkommen (Zipf, 1950c; Zipf & Rucker, 1949 etc.) ist in der Soziologie seit Pareto bekannt. Allerdings weichen die r - s -Kurven erheblich von

der hypothetischen log-Geraden ab. Noch deutlicher wird diese Abweichung in den n - s -Kurven, wo zum Teil sogar eine rückläufige Tendenz der Daten zu beobachten ist. Daß die Darstellung gekippt ist, wurde schon angemerkt. Neu ist allerdings, daß das monotone Modell $n \cdot s^q = c$ die Daten nicht beschreibt. Offensichtlich ist diese Form der n - s -Verteilung ein Spezialfall. Um festzustellen, ob dieser Spezialfall auch den Idealfall darstellt, wie Zipf es annimmt, müßte eine ausführliche Analyse der Regelungsmechanismen, die die Verteilung von Löhnen und Gehältern auf eine Belegschaft regeln, durchgeführt werden.

Orlov (1982) hat die Abgeschlossenheit eines Textes als Randbedingung für die Gültigkeit des Zipf-Mandelbrot-Gesetzes formuliert und so gezeigt, daß dieser Zusammenhang ebenfalls als Spezialfall einer allgemeineren Formel aufgefaßt werden muß.

11. Wortwiederholung und die Verteilung von *token*-Abständen im Text

Die mittlere Wellenlänge w_f eines Wortes mit der Frequenz f ist der Umkehrwert der relativen Häufigkeit (*Psycho-Biology*: 45), nämlich die durchschnittliche Anzahl von Wörtern zwischen zwei Tokens eines Types (N ist der Stichprobenumfang):

$$(11.1) \quad w_f = \frac{N}{f}$$

Die Wellenlängenproportion x ist der Wert, mit dem der Rang r multipliziert werden muß, um die Wellenlänge für den betreffenden Rang zu erhalten (*Psycho-Biology*; Zipf, 1937a).

Da sich für das Englische (Eldridge) $x = 10$ ergab, nahm Zipf an, eine universale, charakteristische Konstante gefunden zu haben. Sn (vgl. Abschnitt 10) bezeichnet die gleiche Größe wie x .

G.K. Zipf interessierten außer den Worthäufigkeiten auch die Häufigkeitsdynamik der Wörter innerhalb eines Textes. Ein globales Maß für die Wiederholungsrate in einer Stichprobe stellt der Exponent der n - f -Verteilung dar (1937a). Je steiler diese Kurve wird, desto stärker wiederholen sich die häufigsten Wörter, der Quotient aus laufenden Wörtern und verschiedenen Wörtern (Tokens : Types) wird immer größer. Das ist die typische Auswirkung jedes Unifikationsprozesses.

Über die Abstände zwischen den einzelnen Wortvorkommen macht Zipf folgende Annahmen:

1. Die n - s -Verteilung der Intervallgrößenklassen gilt innerhalb einer Wortfrequenzklasse mit $n^p \cdot s = c$ (n : Anzahl Intervalle mit gleiche Größe s) (1945b).

Um dies zu zeigen, zählt Zipf die Intervalle in Seiten aus Hanleys Index²¹ zu *Ulysses* (durch Subtraktion der Seitennummern im Index). Für die Frequenz $f = 5$ sind die Intervallgrößen auf der Abszisse und die Anzahl Intervalle gleicher Größe auf der Ordinate in Abbildung 11.1 dargestellt. Weitere Zählungen finden sich in *Human Behavior* (S.53).

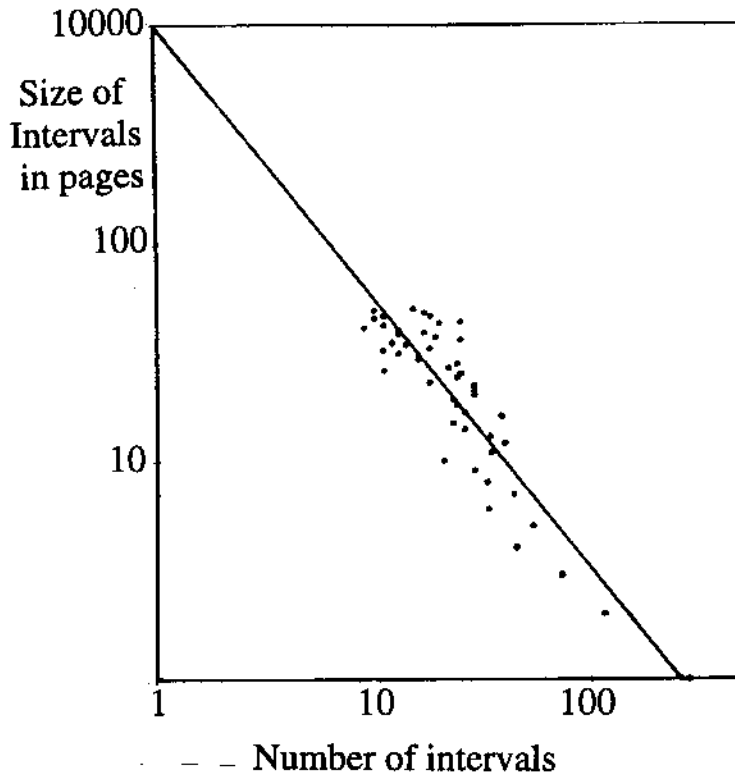


Abb. 11.1. N - s -Verteilung der Intervallgrößen für Wörter mit der Frequenz 5. Logarithmische Achsen. Zählung aus *Ulysses*. Nach Zipf (1945b:2).

2. Innerhalb einer Wortfrequenzklasse sind die verschiedenen Intervallgrößen über alle möglichen Positionen gleichmäßig verteilt.

Zipf illustriert die Annahme in einem Schaubild (Abbildung 11.2), ohne die Urlisten anzugeben oder einen mathematischen Nachweis zu bringen.

²¹ Siehe Literaturverzeichnis, Joos (1965).

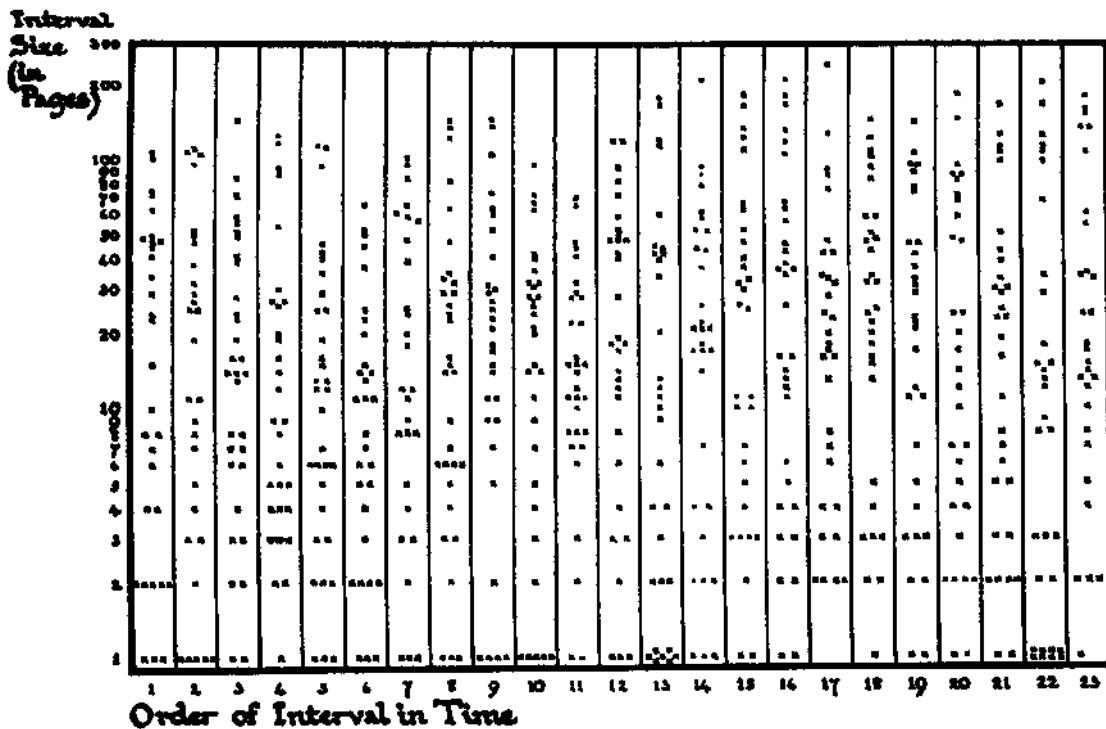


Abb. 11.2. Relative Intervallposition und Intervallgröße (Anzahl Seiten) zwischen den Wiederholungen der Wörter mit Frequenz $f = 24$ (Joyce: Ulysses). Aus Zipf (1946f:337). Die Zelle $I = 1$, Spalte 23 müßte drei 'x' haben, vgl. Human Behavior 46 und Tabelle 11.1.

In *Human Behavior* (S. 45, vgl. Tabelle 11.1) betrachtet Zipf die Verteilung von Intervallen der Länge 1 aus der gleichen Erhebung. Die Zufallsvariable X bezeichnet das Ereignis „die Länge des x -ten Intervalls bei einem Wort der Frequenz f ist 1“. Dabei ist $1 \leq x \leq (f - 1)$. Die kürzesten Intervalle sind dieser Tabelle nach gleichmäßig über die verschiedenen relativen Textpositionen verteilt, es gibt keine Tendenz, ein Wort langsam (mit größeren Abständen) einzuführen, oder daß direkt bei der Einführung bzw. in der „Mitte“ eine Klumpung mit vorzugsweise kurzen Abständen auftritt. Dabei werden Verteilungen in den Reihen der Tabelle betrachtet. Über die absolute Position eines Wortes im Text sagt diese Zählung nichts.

χ^2 -Tests auf Gleichverteilung führen bei $\alpha = 0.05$ nicht dazu, daß Zipfs Hypothese verworfen werden müßte.

Der Text stellt ein „Projekt“ dar, für das sich der Sprecher gewisse inhaltliche und strukturelle Vorgaben (Länge, Dichte) macht und darum das, was er sagen will, gleichmäßig über den Text verteilt. Um die „Arbeit“ gleichmäßig auf den projektierten Zeitraum aufzuteilen, müssen leichte und schwierige (das sind häufige und seltene) Wörter gleichmäßig im Text angeordnet werden. Damit verteilen sich in einer Frequenzklasse auch die Größen ihrer Abstände gleichmäßig (horizontal in Abbildung 11.2) bzw. nach den bekannten Organisationsprinzipien der Verwendungsökonomie (vertikal), der n - s -Verteilung. Cluster von Wörtern mit kleiner Wellenlänge wechseln sich mit Wörtern ab, deren Wellenlänge größer ist. Illustriert wird

diese Vorstellung durch die Glockenläute-Analogie. Die „Zeitperspektive“ Zipfs kann man in den Untersuchungen Orlovs (1982) wiederfinden, die zeigen, daß die Häufigkeitsstruktur künstlerischer Texte häufig sehr nahe an den „Zipfschen Umfang“ herankommt.

Tabelle 11.1.

Die Dispersion der $(f - 1)$ kleinsten Intervalle (1 Seite) zwischen den f Vorkommen der Wörter aus der Frequenzklasse f . Nach Zipf (*Human Behavior*:45), aus *Ulysses* (Hanley-Index). Die unterste Reihe ($f = 24$) entspricht der untersten Reihe (Intervallgröße 1) in Abbildung 11.2.

f	$f-1$	i -tes Intervall																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
6	5	62	55	62	58	52																		
12	11	7	19	15	16	9	12	18	16	12	15	14												
16	15	6	10	10	13	18	11	16	11	11	9	11	9	6	8	12								
17	16	4	3	5	6	4	8	5	10	11	9	14	5	8	6	7	8							
18	17	9	11	6	5	6	7	7	6	9	6	2	6	5	6	6	5	4						
19	18	3	8	5	11	5	6	13	9	6	5	6	8	2	7	10	5	7	4					
21	20	3	4	10	5	8	9	3	10	8	11	7	7	6	6	2	1	7	8	4	2			
22	21	7	5	8	12	5	9	5	9	6	7	5	8	6	6	7	10	7	10	9	5	2		
23	22	3	5	6	4	8	4	3	2	7	3	4	4	5	7	3	6	2	7	2	3	1	3	
24	23	3	5	2	1	3	3	3	3	4	5	2	3	7	3	2	2	0	1	2	2	2	8	3

Zipf leitet aus diesen Überlegungen auch eine „Geschlossenheit“ von sprachlichen Stichproben ab (*Human Behavior*: 39), was bedeutet, daß jedes vorkommende Wort mindestens einmal benutzt worden sein muß²². Der Gedanke wird aber nicht mathematisch ausgeführt.

Das gleiche Ökonomieprinzip entdeckt Zipf auch in der Gestaltung von Konzertprogrammen (1946d). Zählungen von Werken gleicher Komponisten in einem Konzertzyklus über mehrere Jahre ergeben für die Intervalle zwischen den Aufführungen annähernd eine n - s -Kurve. Auch r - f - und n - f -Verteilungen für die Aufführungshäufigkeit der Werke verschiedener Komponisten entsprechen einigermaßen den erwarteten Kurven.

12. Wortähnlichkeit und Redundanz

In Zipf (1949b) sowie in *Human Behavior* (S. 108) beschreibt Zipf ein Experiment zur phonemischen Redundanz. In einer Wortliste werden zunächst Homophone eliminiert bzw. zusammengefaßt. Danach werden durch sukzessive Aufhebung von

²² Da nur benutzte Wörter vorkommen, eine Tautologie. Zipf meint den Umfang, für den die r - f -Verteilung gilt, und bezeichnet ihn als closure.

Oppositionen neue Homophone gebildet bzw. zusammengefaßt. Aus dieser fortschreitenden Unifikation ergeben sich die Veränderungen für die Anzahl unterschiedlicher Wörter in einer Wortliste von 2544 Wörtern (Sprache: Englisch, nach Eldridge; vgl. *Human Behavior*: 108) wie in Tabelle 12.1.

Tabelle 12.1.

Fortschreitende phonologische Unifikation und Homophonzahlen nach Zipf (*Human Behavior*: 108).

Klasse	Anzahl	% von 2544	Rest (kumulativ)
Homophone	23	0.90	2521 Wörter
l = r; m = n	40	1.57	2481 Wörter
kurze Vokale: a=e=i=o=u ²³	21	0.83	2460 Wörter
alle Vokale gleich	136	5.35	2264 Wörter

In einer Bemerkung (*Human Behavior*: 549 Nr.37) weist Zipf darauf hin, daß nach vorläufigen Studien die *r-f*-Kurven dieser Daten log-Geraden bleiben, aber steiler werden.

Zipf tut gut daran, dieses sehr unsystematisch durchgeführte Experiment nur vorsichtig zu interpretieren. Die Aufhebung der Oppositionen bzw. ihre Reihenfolge ist völlig willkürlich und zum Teil fehlerhaft durchgeführt. Daß die *r-f*-Verteilungen steiler werden, ist eine Auswirkung der Unifikation, und inwieweit wirklich die theoretische log-Gerade solche Daten noch beschreibt, bleibt zu prüfen.

Im Rahmen von Untersuchungen zur Ausnutzung des kombinatorischen Potentials auf Phonemebene bzw. zur funktionalen Leistung von Oppositionen sind solche Versuche sicherlich interessant. Zipf war von der Redundanzleistung der „restlichen“ Phoneme und ihrer Anordnung überzeugt:

„we must not overlook the basic problem of the selection of particular phonetic entities for particular permutations in order to save work for the auditor by making the succession of words of the stream of speech as different als possible“ (*Human Behavior*: 108).

Zipf sah das Problem weniger im Zusammenhang von Bedeutungsverwechslung, da der Kontext stark disambiguierend wirkt, sondern als syntagmatisches Problem: „a too close juxtaposition of homophones, or even "otherwise homophonous" words might confuse the auditor“ (ebd.).

Diese Annahme berücksichtigt natürlich überhaupt nicht die Redundanzleistungen anderer sprachlicher Erscheinungen wie Wortarten, Morphosyntax und Syntax,

²³ „... removal of all differences between short vowels (so that bit, bet, bat, but, bought would fall together)“ (*Human Behavior*: 108). „bought“?!

Prosodie sowie distributionelle Varianten auf Phonemebene etc.

Vom synergetisch-linguistischen Standpunkt aus gesehen ist die Redundanz *Red* eine sprachformende Systemanforderung (Köhler, 1990b), die durch eine ganze Reihe funktional äquivalenter Mittel erfüllt werden kann. Zipfs Versuch bezieht sich auf die Redundanz auf lexikalischer Ebene und auf zwei Mittel, die phonemische Opposition und die syntagmatische Silbenstruktur, denn darauf läuft die Elimination aller Oppositionen außer der K-V-Opposition hinaus. Im Versuch müßten diese beiden Bereiche zunächst getrennt, dann aber auch im Zusammenhang mit anderen Mitteln zur Redundanzbildung untersucht werden, z.B. „Haben diese beiden Homophone die gleiche Wortart, gleiches Genus usw.“ - abhängig davon, wie stark die betreffende Sprache von den jeweiligen Mitteln Gebrauch macht.

Es ist zu überlegen, ob der Einfluß der Redundanz nicht in vielen Fällen von der Häufigkeit des entsprechenden Elements abhängt, da das Bedürfnis nach Übertragungssicherung - nach Zipfs *principle of relative frequency* - mit der Informationslast eines Elements steigt.

13. Geographie

Neben der *r-s*-Verteilung der Siedlungsgrößen stellt Zipf (1946a,c; *Human Behavior*: 386ff.) eine weitere geographische Hypothese auf:

Die Distanz D zweier Siedlungen ist proportional zum Produkt ihrer Einwohnerzahlen P :

$$(13.1) \quad \frac{P_1 P_2}{D} = \text{const.}$$

Die $\frac{P_1 P_2}{D}$ -Hypothese begründet Zipf (z.B. 1946a) damit, daß der Verkehr (Waren- und Personenumsatz, Informationsaustausch) zwischen zwei Ortschaften von den jeweiligen Bevölkerungsgrößen abhängt. Außerdem sinkt der Umsatz wegen des höheren Aufwands mit der Distanz. Da die Siedlungsgrößen *r-s*- und *n-s*-verteilt sind, ist die dazugehörige günstigste Distanz eben jeweils D .

Zum Nachweis führt Zipf (1946a) eine Untersuchung zum Transport von Paketen per Eisenbahn zwischen 78 Städtepaaren durch. Er erhält die Menge der Güter C (in 1000 Pfund) als Funktion des Quotienten $\frac{P_1 P_2}{D}$, und zwar logarithmiert:

$$(13.2) \quad \log C = 0.2157 + 0.8472 \log \left(\frac{P_1 P_2}{D} \right)$$

Abb. 13.1 zeigt die entsprechende Graphik Zipfs.

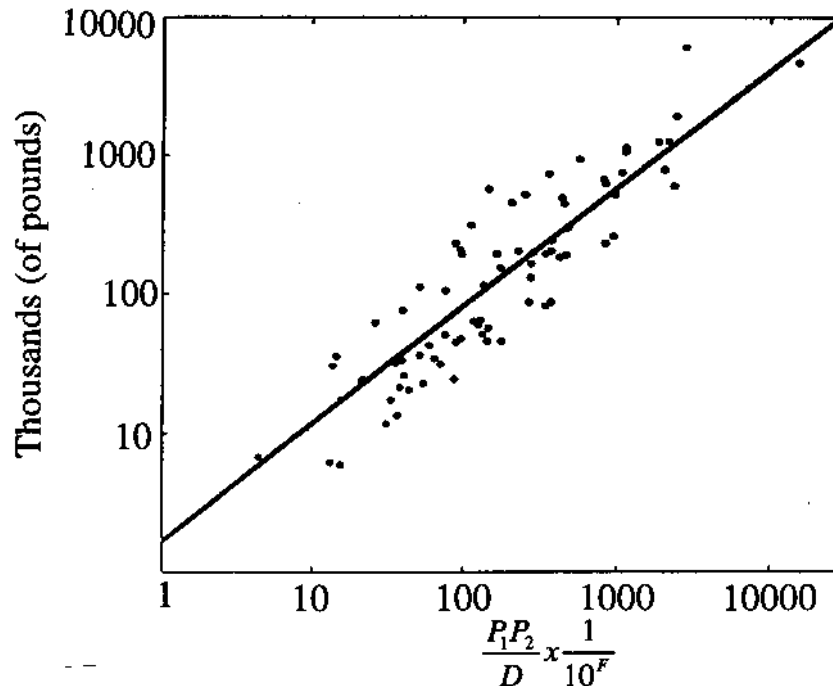


Abb. 13.1. Die Bewegung von Gütern per Bahn (railway express) zwischen 13 Städten in Tausend Pfund über den Quotienten $\frac{P_1 P_2}{D}$. Parameter der theoretischen Kurve s. (13.3); logarithmische Achsen. Nach Zipf (*Human Behavior*: 393).

Dieselben Überlegungen stellt Zipf (1946c) bezüglich des Personentransports an, und zwar für Bus, Bahn- und Flugverkehr. Hier ist die Anzahl der Fahrgäste bzw. die Menge der bezahlten Fahrpreise multipliziert mit der Distanz D proportional zu $P_1 \cdot P_2$. Ebenso ist die Anzahl von Ereignissen proportional zur Bevölkerung (Zipf, 1947a), darum ist auch die Zirkulation von Information (Anzahl und Umfang von Meldungen, Zirkulation von Zeitungen, Telefongespräche, Telegramme) proportional zu $\frac{P_1 P_2}{D}$ (1946e, 1948, vgl. Abbildung 13.4). Eine große Menge interessanter siedlungs- und stadtgeographischer, wirtschaftsgeographischer und soziologischer Zusammenhänge kann hier nicht mehr referiert werden (vgl. z.B. Zipf, 1947d). In *Human Behavior* sind im Prinzip alle diese Veröffentlichungen zusammengefaßt.

Zipf leitet aus seinen Ergebnissen ab, daß sich abhängig von der Bevölkerungszahl eines Ortes ein Einflußbereich (*domain*) um diesen Ort bildet (1946e). Er bedauert (Zipf, 1946f: 341), daß die Demographie in der Dialektforschung nicht beachtet wird. Die Siedlungsgröße beeinflusst die Lage und die Form von Kontaktgrenzen (Isoglossen; *Human Behavior*: 527f.): Sind P_1 und P_2 gleich, so sind die Kontaktgrenzen in der Mitte und „gerade“. Je größer einer der beiden Orte ist, desto größer ist sein Einflußbereich. Hier bietet sich eventuell auch eine Möglichkeit, ältere Sprachkontakte nachzuvollziehen.

Literatur

Titelabkürzungen im Text bedeuten:

- Relative Frequency*: Zipf (1929)
Selected Studies: Zipf (1931)
Psycho-Biology: Zipf (1935)
National Unity: Zipf (1941)
Human Behavior: Zipf (1949a)

Buchstaben (a, b ...) nach den Jahreszahlen dienen der Differenzierung und geben keine Erscheinungsreihenfolge an.

- Altmann, G.** (1981). Zur Funktionalanalyse in der Linguistik. In J. Esser & A. Hübler (Hg.), *Forms and functions*, (S. 25-32), Tübingen: Narr.
- Altmann, G.** (erscheint). Phonemzählung. In G. Altmann & R. Köhler, *Einführung in die quantitative Linguistik*.
- Altmann, G., Schwibbe, M.H., et al.** (1989). *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Hildesheim: Olms.
- Arapov, M.V., & Cherc, M.M.** (1983). *Mathematische Methoden in der historischen Linguistik* (= QL 17). Bochum: Brockmeyer.
- Birkhan, H.** (1979). Das „Zipfsche Gesetz“, das schwache Präteritum und die germanische Lautverschiebung. Sitzungsberichte der Österreichischen Akademie der Wissenschaften, philosophisch-historische Klasse 348.
- Joos, M.** (1936). Review of G.K. Zipf. The psycho-biology of language. *Language*, 12, 196-210.
- Joos, M.** (1965). Remarks on *Ulysses* word-frequencies. In M.L. Hanley et al. (Hg.), *Word index to James Joyce's Ulysses*. Madison/Wisc.
- Köhler, R.** (1986). *Zur linguistischen Synergetik. Struktur und Dynamik der Lexik* (=QL 31). Bochum: Brockmeyer.
- Köhler, R.** (1990). Elemente der synergetischen Linguistik. In R. Hammerl (Hg.), *Glottometrika 12* (S. 179-187), Bochum: Brockmeyer.
- Köhler, R., & Altmann, G.** (1983). Systemtheorie und Semiotik. *Zeitschrift für Semiotik*, 5, 424-431.
- Lundberg, G.A., & Dodd, S.C.** (1950). Obituary. *American sociological review*, 15, 104.
- Mandelbrot, B.** (1957). A note on a law of Berry and on insistence stress. *Information and control*, 1, 76-81.
- Miller, G.A.** (1968). Introduction. In G.K. Zipf (Hg.), *The psycho-biology of language. An introduction to dynamic philology*. Cambridge, Mass.: M.I.T. Press. 2nd printing, iii-x.
- Orlov, J.K.** (1982). Dynamik der Häufigkeitsstrukturen. In H. Guiter & M.V. Arapov (Hg.), *Studies on Zipf's law* (S. 116-153), Bochum: Brockmeyer.

- Rapoport, A.** (1982). Zipf's law re-visited. In H. Guiter & M.V. Arapov (Hg.), *Studies on Zipf's law* (S. 1-28), Bochum: Brockmeyer.
- Whitehorn, J.C., & Zipf, G.K.** (1943). Schizophrenic language. *Archive of neurology and psychiatry*, 49, 831-851.
- Zipf, G.K.** (1929). Relative frequency as a determinant of phonetic change. *Harvard studies in classical philology*, 40.
- Zipf, G.K.** (1932). *Selected studies of the principle of relative frequency in language*. Cambridge, Mass.: Harvard Univ.Press.
- Zipf, G.K.** (1935). *The psycho-biology of language. An introduction to dynamic philology*. Cambridge, Mass.: M.I.T. Press, 2nd ed. 1968 [Erstausgabe: Boston, Houghton-Mifflin, 1935].
- Zipf, G.K.** (1937a). Observations on the possible effect of mental age upon the frequency-distribution of words from the viewpoint of dynamic philology. *Journal of psychology*, 4, 239-244.
- Zipf, G.K.** (1937b). Statistical methods in dynamic philology (Reply to M. Joos). *Language*, 13, 60-70.
- Zipf, G.K.** (1938a). Erwiderung. *Indogermanische Forschungen*, 56, 75-77. [zu: Jost (1937)]
- Zipf, G.K.** (1938b). Homogeneity and heterogeneity in language. In answer to Edward L. Thorndike. *Psychological record*, 2, 347-367.
- Zipf, G.K.** (1938c). On the problem of grammatical rules and the study of 'General Language'. *Modern language journal*, 22/4, 243-249.
- Zipf, G.K.** (1938d). Phonometry, phonology, and dynamic philology. An attempted synthesis. *American speech*, 13, 275-285.
- Zipf, G.K.** (1938e). New facts in the early life of George Meredith. *Harvard studies and notes in philology and literature*, 20.
- Zipf, G.K.** (1940a). The generalized harmonic series as a fundamental principle of social organization. *Psychological record*, 4, 43.
- Zipf, G.K.** (1940b). On the economical arrangement of tools, the harmonic series and the properties of space. *Psychological record*, 4, 147-159.
- Zipf, G.K.** (1941). *National unity and disunity; The nation as a bio-social organism*. Bloomington, Ind.: Princeton Press.
- Zipf, G.K.** (1942a). Children's speech. *Science*, 96, 344-345.
- Zipf, G.K.** (1942b). The unity of nature, least-action, and natural social science. *Sociometry*, 5, 48-62.
- Zipf, G.K.** (1945a). The meaning-frequency relationship of words. *Journal of general psychology*, 33, 251-256.
- Zipf, G.K.** (1945b). The repetition of words, time-perspective and semantic balance. *Journal of general psychology*, 32, 127-148.
- Zipf, G.K.** (1945c). Some psychological determinants of the structure of publications. *American journal of psychology*, 58, 425-442.
- Zipf, G.K.** (1946a). The $\frac{P_1 P_2}{D}$ hypothesis: The case of railway express. *Journal of psychology*, 22, 3-8.

- Zipf, G.K.** (1946b). Cultural-chronological strata in speech. *Journal of abnormal and social psychology*, 41/3, 351-355.
- Zipf, G.K.** (1946c). The $\frac{P_1 P_2}{D}$ hypothesis: On the intercity movement of persons. *American sociological review*, 11, 677-686.
- Zipf, G.K.** (1946d). On the dynamic structure of concert programs. *Journal of abnormal and social psychology*, 41, 25-36.
- Zipf, G.K.** (1946e). Some determinants of the circulation of information. *American journal of psychology*, 59, 401-421.
- Zipf, G.K.** (1946f) The psychology of language. In P.H. Hariman (Hg.), *Encyclopedia of psychology*, 332-341.
- Zipf, G.K.** (1947a). On Dr. Miller's contribution to the $\frac{P_1 P_2}{D}$ hypothesis. *American journal of psychology*, 60, 284-287.
- Zipf, G.K.** (1947b). The frequency and diversity of business establishments and personal occupations. A study of social stereotypes and cultural roles. *Journal of psychology*, 24, 139-148.
- Zipf, G.K.** (1947c). Prehistoric 'cultural strata' in the evolution of Germanic: The case of Gothic. *Modern language notes*, 62, 522-530.
- Zipf, G.K.** (1947d). The hypothesis of the 'Minimum Equation' as a unifying social principle. With attempted synthesis. *American sociological review*, 12, 627-650.
- Zipf, G.K.** (1948). On the number, circulation-sizes and the probable purchasers of newspapers. *American journal of psychology*, 61, 79-89.
- Zipf, G.K.** (1949a). *Human behavior and the principle of least effort. An introduction to human ecology*. New York: Hafner reprint, 1972. [1. Aufl.: Cambridge/Mass., Addison-Wesley, 1949.]
- Zipf, G.K.** (1949b). Relative frequency and dynamic equilibrium in phonology and morphology. In *Proceedings of the 6th international congress of linguists*. Paris. S. 391-408.
- Zipf, G.K.** (1950a). Brand names and related social phenomena. *American journal of psychology*, 63, 342-366.
- Zipf, G.K.** (1950b). Empiric regularities in the frequency-distribution of chemical manufacturers and chemical distributors by product-diversity in the U.S.A. *American journal of psychology*, 30, 195-198.
- Zipf, G.K.** (1950c). The frequency-distribution of wages and the problem of labor unrest. *American journal of psychology*, 29, 315-324.
- Zipf, G.K.** (1950d). Empiric regularities in the frequency-distribution of directorships in american corporations. *American psychologist*, 5, 245 [= Ankündigung eines Vortrags für den 7.9.1950].
- Zipf, G.K.** (1950e). Wage distribution and the problem of labor harmony. In P.A. Sorokin (Hg.), *Explorations in altruistic love and behavior* (S. 333-346), Boston: .
- [**Zipf, G.K.** (1950f). *Quantitative analysis of a Sears Roebuck catalog*. New - York. - Nicht greifbare Veröffentlichung, nachgewiesen in G. Billmeier & D.

Krallmann, *Bibliographie zur statistischen Linguistik. IPK Forschungsbericht 69/3* (S. 144), Hamburg: Buske (1969).]

Zipf, G.K., & Rogers, F.M. (1939). Phonemes and variphones in four present-day Romance languages and Classical Latin from the viewpoint of dynamic philology. *Archives néerlandaises de phonétique expérimentale*, 15, 111-147.

Zipf, G.K., & Rucker, A.W. (1949). How to set salary brackets that spur achievement. *Modern management*, 9, 4-7.

Zörnig, P., & Boroda, M. (1992). The Zipf-Mandelbrot law and the interdependencies between frequency structure and frequency distribution in coherent texts. In B. Rieger (Hg.), *Glottometrika 13* (S. 205-218), Bochum: Brockmeyer.