

QUANTITATIVE LINGUISTICS

Vol. 31

**Zur linguistischen
Synergetik:
Struktur und Dynamik
der Lexik**

Reinhard Köhler



Studienverlag Dr. N. Brockmeyer
Bochum 1986

QUANTITATIVE LINGUISTICS

Editors

G. Altmann, Bochum
 R. Grotjahn, Bochum
 N. D. Andreev, Leningrad
 M. V. Arapov, Moscow
 J. Boy, Bochum
 B. Brainerd, Toronto
 H. Guiter, Montpellier
 D. Hérault, Paris
 E. Hopkins, Bochum
 W. Lehfeldt, Konstanz
 W. Matthäus, Bochum
 R. G. Piotrowski, Leningrad
 B. Rieger, Aachen/Amsterdam
 J. Sambor, Warsaw
 D. Wickmann, Aachen

Editorial Board

INHALT

1. EINLEITUNG	1
2. MESSUNG UND MODELLBILDUNG IN DER LINGUISTIK	8
2.1 Verteilung fundamentaler linguistischer Eigenschaften	9
2.2 Abhängigkeiten zwischen linguistischen Größen	10
2.3 Strukturgesetze	11
2.4 Gesetze der Sprachdynamik	13
3. ZIELSETZUNG UND METHODIK	16
4. DIE KONSTRUKTION DES MODELLS	20
4.1 Einführung: Die Elemente auf der Phonem-/Grammebene	22
4.2 Sprache - ein selbstregulierendes System	25
4.3 Linguistik und Synergetik	34
4.4 Zur Begriffsbestimmung	38
4.5 Systemtheoretische Hilfsmittel: Graphenalgebra und lineare Operatoren	43
4.5.1 Operatoren und Graphen	44
4.5.2 Struktur und Funktion	46
4.5.3 Vereinfachung von Graphen	48

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Köhler, Reinhard:

Zur linguistischen Synergetik: Struktur u. Dynamik d. Lexik / Reinhard Köhler. - Bochum: Studienverlag Brockmeyer, 1986. (Quantitative linguistics: Vol. 31)
 ISBN 3-88339-538-2

NE: GT

ISBN 3-88339-538-2

Alle Rechte vorbehalten

© 1986 by Studienverlag Dr. N. Brockmeyer

Querenburger Höhe 281, 4630 Bochum 1

Druck: Thiebes GmbH & Co. Kommanditgesellschaft Hagen

4.6	Elemente, Struktur und Funktion des lexikalischen Systems	50	
4.6.1	Das Phoneminventar	50	
4.6.2	Die Lexikongröße	52	
4.6.3	Die Länge	53	
4.6.4	Die Polylexie	57	
4.6.5	Die Polytextie	63	
4.6.6	Die Frequenz	66	
4.6.7	Die Länge (Fortsetzung)	69	
4.7	Die Funktionen der Teilstrukturen	75	
4.8	Die Bedürfnisse, Größen und Prozesse im Modell	75	
5.	DEDUKTION VON KONSEQUENZEN	79	
5.1	Direkte Zusammenhänge	79	
5.2	Indirekte Zusammenhänge	81	
5.3	Verteilungen der strukturellen Eigenschaften	83	
5.4	Dynamik	85	
6.	DIE EMPIRISCHE ÜBERPRÜFUNG DES MODELLS	87	
6.1	Operationalisierung	87	
6.2	Datenerhebung	94	
6.3	Das Testverfahren	97	
6.3.1	Die Polylexie	100	
6.3.2	Die Polytextie	103	
6.3.3	Die Frequenz	105	
6.3.4	Die Länge	110	
6.3.5	Überprüfung indirekter Abhängigkeiten	112	
6.3.6	Vergleich der Operationalisierungsmöglichkeiten	124	
6.3.7	Zusammenfassung der Testergebnisse	136	
7.	DIE OSZILLATION DER LEXIK	137	
8.	VOM BASISMODELL ZUR SPRACHTHEORIE	147	
8.1	Leistung des entwickelten Modells	147	
8.2	Skizze des weiteren Ausbaus	149	
8.2.1	Erweiterung des Gegenstandsbereichs und Strukturverfeinerung	149	
8.2.2	Interaktion zwischen System und Systemumgebung	152	
8.2.3	Modellierungstechnik	152	
8.2.4	Zeitverhalten des Systems	153	
9.	SCHLUSSBEMERKUNG	154	
	LITERATUR	156	
	DATENTABELLEN	169	

1. EINLEITUNG

Die quantitative Linguistik unterscheidet sich nicht hinsichtlich ihres Erkenntnisziels von den anderen Richtungen der Linguistik (sofern sie den Aufbau einer Sprachtheorie anstreben), sondern in erster Linie durch ihre Hilfsmittel und Methoden. Die Grundlegung der mathematischen Linguistik erfolgte durch die Erarbeitung der kategorialen (qualitativen) Begriffe und der Formulierung der Zusammenhänge zwischen diesen Begriffen. Ihrer Aufgabe entsprechend wurden vor allem die qualitativen Bereiche der Mathematik, besonders Logik, Mengenlehre, Gruppentheorie, Automaten-theorie, Semi-Thue-Systeme, lineare Algebra, Verbände, Graphentheorie, Topologie und Geometrie in die Sprachwissenschaft eingeführt, und damit wurden die Voraussetzungen für exakte und widerspruchsfreie sprachwissenschaftliche Aussagen geschaffen.

Wie das Beispiel anderer Wissenschaften zeigt, genügen solche Begriffe auf Nominalskalenniveau aber nicht für ein tieferes Eindringen in den Forschungsgegenstand: So verlangte die Entdeckung vieler neuer Stoffe in der Chemie nach immer neuen Kategorien und Benennungen, nach Regeln für die Zusammensetzung von Stoffen aus anderen Stoffen. Aber zu einer ersten Einsicht in die Struktur der Stoffe - zur Entdeckung des Moleküls - führte eine quantitative Analyse, bei der die festen Massenproportionen in solchen Verbindungen festgestellt wurden.

Auch in der Linguistik wurden ordinale und metrische, also quantitative, Begriffe und damit Modelle und Methoden erforderlich, die natürlich auf den qualitativen aufbauen. Die Notwendigkeit höher skaliertter Begriffe in der Linguistik läßt sich - über

die grundsätzlichen Überlegungen zur Begriffsverfeinerung hinaus (vgl. z. B. Stegmüller 1970 : 98-109) - anhand von vier Erwägungen verdeutlichen:

1. Mit Hilfe von deterministischen Systemen lassen sich lediglich "extrem regelmäßige" Spezialfälle unter den sprachlichen Erscheinungen erfassen, von denen es aber nur sehr wenige gibt. Zur Überwindung dieser Beschränkung bedarf es der stochastischen Modellierung der Sprache und damit quantitativer Hilfsmittel. Einer von vielen Belegen für diese Auffassung ist der Sprachwandel: Jedes Sprachsystem ist im Laufe der Zeit sehr vielen kleinen und seltenen großen, sich sprunghaft verändernden Einflüssen ausgesetzt. Einem synchronischen, deterministischen Regelsystem fehlt jedes Mittel zur Beschreibung der entsprechenden Abweichungen. Die Veränderungen in den Sprachsystemen sind aber nicht nur meßbar, sie unterliegen auch selbst bestimmten (stochastischen) Gesetzmäßigkeiten.

Ein weiteres, allgemein bekanntes Beispiel für die Unzulänglichkeit deterministischer Deskription ist die Vagheit der Bedeutung; in der modernen Semantik werden daher zunehmend quantitative Modelle (wie die unscharfen Mengen) verwendet.

2. Bei der Untersuchung der Sprache haben wir es mit einem besonders für sozialwissenschaftliche Bereiche typischen Informationsunterschuß zu tun: wir können stets nur einen kleinen Ausschnitt des Forschungsgegenstands erfassen. Entweder ist er nicht endlich (wie die Menge der Texte in deutscher Sprache), oder er verändert sich schneller, als er beschrieben werden kann (wie die Lexik einer Sprache). Diese Situation ist aber der klassische Bedarfsfall für die Methoden der Statistik, die eine kontrollierte Schlußfolgerung bei unvollständiger Information erlauben.

3. Die Theorie- und Hypothesenbildung ist ein deduktiver Vorgang: Neue Hypothesen werden aus bekannten Aussagen oder vermuteten Zusammenhängen abgeleitet, nicht aus den Daten. Die Deduktion ist ein "qualitativer" Vorgang, ob man nun eine Differentialgleichung aufstellt und löst, einen stochastischen Prozeß modelliert oder mit Hilfe der klassischen Logik Schlüsse zieht. Der Unter-

schied besteht im Resultat: Durch die Verwendung quantitativer Werte für die Parameter der resultierenden Form einer Differentialgleichung wird eine bessere Annäherung, eine größere Abbildungstreue gegenüber anderen Deduktionstypen erreicht, deren resultierende qualitative Form nicht weiter verfeinert wird und den Gegenstand nur vage abbildet. Dieser Nachteil qualitativer Endresultate von Schlußfolgerungen hat in der letzten Zeit zur Entwicklung der unscharfen Mengen, der unscharfen Logik und der Wahrscheinlichkeitslogik geführt.

4. Da wohl alle Prozesse und Relationen in der Sprache stochastisch sind (in der Sprache bedeutet nicht notwendigerweise auch im Sprecher), kann man die entsprechenden Modelle nur statistisch überprüfen. Für diesen induktiven Vorgang bedient sich die quantitative Linguistik der Hilfsmittel der Stichproben-, Schätz- und Testtheorie.

Die hauptsächlich verwendeten mathematischen Hilfsmittel - Differentialgleichungen, Differenzengleichungen, stochastische Prozesse, Wahrscheinlichkeitstheorie, Skalierung, Graphentheorie, Informationstheorie und Statistik - versetzen die quantitative Linguistik aufgrund ihrer differenzierteren Begriffe und des entsprechenden Forschungsinstrumentariums in die Lage, genauere deskriptive Approximationen an die sprachliche Wirklichkeit zu erreichen als die qualitativ orientierten Teildisziplinen.

Inzwischen ist in der Linguistik aber auch ein erstes Vortreten in die explanative Phase dieser Wissenschaft zu beobachten. Die Suche nach Erklärungen in der Linguistik, nach allgemeinen Gesetzen, geht - dem Stand der Begriffsbildung entsprechend - vornehmlich von Vertretern der quantitativen Linguistik aus. Den Vorläufern dieser Bemühungen wie den Junggrammatikern (allgemeine Lautgesetze) oder G. K. Zipf (principle of least effort) fehlten noch die zur Theoriebildung notwendigen Voraussetzungen. Erst mit dem Einsatz der Mathematik in der Linguistik wurde ein entscheidender Schritt in Richtung auf Erklärungen getan, wobei

unter einer Erklärung die Ableitung des Explanandums aus Anfangsbedingungen mit Hilfe von Gesetzen zu verstehen ist (vgl. Hempel 1965).

An dieser Stelle muß auf die ebenso weit verbreitete wie irriige Überzeugung eingegangen werden, die Linguistik habe ihr vortheoretisches Stadium längst überwunden. Möglicherweise beruht dieses Mißverständnis auf der Ähnlichkeit axiomatischer Systeme (wie sie aus den formalen Bereichen der Mathematik und der Logik bekannt sind) und ihrer Schlußregeln einerseits mit dem deduktiven Prinzip wissenschaftlicher Erklärung andererseits. Die Verwendung deduktiver Beschreibungsmodelle, z.B. formaler Grammatiken, verleitet zu Redeweisen wie "Der Satz S wird durch die Regel R erklärt" u.ä.; oft genug verwischen sich, durch solche Redeweisen begünstigt, die Unterschiede zwischen deskriptiven Regelsystemen (wie Grammatiken) und explanativen Gesetzessystemen (wie wissenschaftlichen Theorien). So wird häufig übersehen, daß es gerade die beobachteten und mit Hilfe von Regeln beschriebenen (mehr oder weniger starken) Regelmäßigkeiten sind, die der Erklärung bedürfen. Erklärungen sind aber nur mit Hilfe von Gesetzen möglich; Begriffssysteme, Axiomaten und Systematisierungen sind zwar notwendige Voraussetzungen, sie genügen jedoch nicht.

Wir können die allgemeinen Ziele der quantitativen Linguistik - soweit sie über die Ziele der qualitativen Linguistik hinausgehen - in folgenden Punkten zusammenfassen:

- Einführung von Begriffen auf Ordinal-, Intervall- und Ratiokalenniveau und Entwicklung bzw. Adaption entsprechender Methoden zur Verfeinerung der Deskription;
- Suche nach allgemeinen Gesetzen für Struktur und Dynamik von Sprachen und Texten;
- Bestimmung der sprachlichen und außersprachlichen Parameter und Randbedingungen für diese Gesetze;
- Verknüpfen dieser Gesetze zu einer Sprach- bzw. Texttheorie.

Da die zur Erklärung dienenden Gesetze selbst wieder erklärungsbedürftig sind, hat die Suche nach Erklärungen weder ein logisches noch ein empirisches Ende (vgl. Popper 1973: 216ff). Für die Sprachwissenschaft sind die Grenzen für die Erklärungsketten durch die Grenzen ihres Gegenstandsbereichs vorgegeben; die "letzten" Erklärungen innerhalb einer zukünftigen linguistischen Theorie liegen also in den Schnittstellen zu den Nachbarwissenschaften Biologie, Physik, Psychologie und Soziologie¹.

Eine im Zusammenhang mit der linguistischen Theoriebildung besonders wichtige Frage ist die nach den wissenschaftstheoretischen Normen und methodologischen Standards, die dabei gelten sollen. In der vorliegenden Arbeit wird die von den meisten Vertretern der quantitativen Linguistik geteilte Ansicht vertreten, daß der Kritische Rationalismus (vgl. Popper 1972, 1973; Bunge 1967) die für diese Wissenschaft angemessene Position darstellt. Unter einer Theorie wird also ein System von Sätzen (Gesetzen) verstanden, das Erklärungen von anderen Sätzen (Einzelaussagen) ermöglicht. Der Vorgang der Erklärung ist deduktiv-nomologisch und muß den Bedingungen des Hempel-Oppenheim-Schemas (Hempel, Oppenheim 1948; Stegmüller 1969) genügen. Außerdem ist mit diesem Standpunkt die Forderung nach empirischer Falsifizierbarkeit der Theorie und jeder Hypothese verbunden².

¹ Natürlich ist zu berücksichtigen, daß die Gegenstände der Wissenschaften und auch das Selbstverständnis der Disziplinen Änderungen unterworfen sind.

² vgl. jedoch die modernen Auffassungen des Theorienpluralismus, nach denen eine Theorie nur zugunsten einer besser bestätigten anderen Theorie verworfen wird (Lakatos 1974, 1975; Spinner 1974).

schen Befunden, die als bedeutendster Meilenstein in die Geschichte der quantitativen Linguistik eingingen.

2. MESSUNG UND MODELLBILDUNG IN DER LINGUISTIK

Sie besagten, daß das Produkt aus Worthäufigkeiten in Texten und ihrem Rang (in der Rangfolge ihrer Häufigkeiten) für alle Ränge in einem Text etwa konstant sind. G.K. Zipf, der die gleiche Form der Rang-Frequenz-Verteilung auch für eine große Zahl anderer Erscheinungen aus dem weiten Gebiet des menschlichen Verhaltens fand, versuchte, diesen quantitativen Zusammenhang durch Annahme eines allgemeingültigen Prinzips der geringsten Anstrengung (principle of least effort) zu erklären. Er war davon überzeugt, ein universelles Gesetz gefunden zu haben, nach dem alle menschlichen Handlungen - gesteuert durch Unifikations- und Diversifikationskräfte - einer Aufwandsökonomie unterliegen.

Unter Ökonomie ist in diesem Sinne keinesfalls eine bewußte Handlungsstrategie zu verstehen wie etwa die Kosten-Nutzen-Optimierung in wirtschaftlichen Zusammenhängen, sondern vielmehr ein von der Möglichkeit bewußter Steuerung unabhängiges menschliches Verhaltensprinzip in Form einer anthropologischen Konstante.

Zipfs Überlegungen wurden von B. Mandelbrot (1953) begrifflich präzisiert. Insbesondere führt er eine von der Länge der Wörter abgeleitete Operationalisierung des Aufwandsbegriffs ein, so daß er ein mathematisches Modell für die Minimierung des durchschnittlichen Aufwands durch Optimierung der Wortlängen aufstellen konnte. Dieses Modell, das außerdem eine bessere Vorhersage des numerischen Zusammenhangs zwischen Frequenz und Rang erlaubte als Zipfs Formel, wurde als das Zipf-Mandelbrotsche Gesetz bezeichnet.

Für den gleichen Sachverhalt entwarf Simon (1955; 1960) ein alternatives Modell, das die Tätigkeit der Textproduktion als stochastischen Prozeß abbildet. Trotz der überzeugenden Herleitung und Argumentation Simons hatte dieses Modell keinen Bestand, da es sich mit den empirischen Daten zur Rang-Frequenz-Verteilung nicht sehr gut vereinbaren ließ.

Mit dem strukturalistischen Paradigma ist es der Sprachwissenschaft gelungen, sich ein Fundament exakter wissenschaftlicher Begriffe und damit die Voraussetzung für die Mathematisierung der Disziplin zu schaffen. Als zentraler Begriff wurde der des Systems eingeführt; im Rahmen des neuen Paradigmas wurde die Sprache zum Zeichensystem, dessen Elemente, die einzelnen sprachlichen Zeichen, in ihrer Form auf Konvention zurückgeführt wurden.

Von vornherein war mit dem durch den Strukturalismus gewonnenen Fortschritt eine beträchtliche Einseitigkeit verbunden. Zum einen bedeutete die (damals notwendige) Betonung der Trennung zwischen synchronischer und diachronischer Betrachtung der Sprache, daß nur ein statisches Bild vom Sprachsystem gezeichnet werden konnte. Prozesse und Wechselwirkungen zwischen Systemeigenschaften konnten nicht integriert werden. Zum anderen wurde das linguistische Interesse auf die Relationen der Zeichen im System beschränkt, die als (paradigmatische und syntagmatische) Beziehungen ausschließlich kategorialer Natur gesehen wurden.

Daneben begannen spätestens seit Beginn dieses Jahrhunderts quantitative Begriffe und Beschreibungen eine zunehmende Rolle zu spielen. Angeregt durch Erfordernisse der praktischen Anwendung, bestanden die ersten linguistischen Messungen in der Häufigkeitzählung von Buchstabenverbindungen und von Wörtern. Der Bedarf an Frequenzlisten für Sprachelemente zum Zweck der Kurzschriftentwicklung (Estoup 1916), der Sprachlehre und vergleichender philologischer Studien (Condon 1928; Zipf 1949) führte dann zu empiri-

Die Zipf-Mandelbrotsche Hypothese aber hat sich bis heute bewährt. Es ist sogar gelungen, auf der Grundlage des Zipfschen Modells weitere interessante Eigenschaften der Häufigkeitsstruktur von Texten zu finden. Zu den wichtigsten gehört sicherlich die Ableitung der sogenannten Zipfschen Zahl Z, mit deren Hilfe Texte verschiedener Länge bezüglich ihrer Häufigkeitsstruktur vergleichbar geworden sind. Sie steht für die optimale Länge eines Textes (die natürlich je nach Randbedingungen und Zweck des Textes sehr verschieden sein kann) und sagt den Informationsfluß im Text voraus (s. Orlov, Boroda, Nadarejškili 1982). Weitere Entwicklungen und Verbesserungen des Modells, etwa von Haight und Jones, findet man in Lánksy/Radil-Weiss (1980, 53-65). Außer an sprachlichen Texten ist die Gültigkeit des Zipf-Mandelbrotschen Gesetzes inzwischen an musikalischen Texten (die dem Wort entsprechende musikalische Texteinheit wird F-Motiv genannt; vgl. Boroda 1982a, 1982b, 1982c) und für die Verteilung von Farbflächen in Bildern (vgl. Vološin, Orlov 1982) nachgewiesen. Entsprechende Rang-Frequenz-Verteilungen kennt man auch u.a. in den Dokumentationswissenschaften der Soziologie, den Geowissenschaften und der Mathematik.

Die überragende Bedeutung des Zipf-Mandelbrotschen Gesetzes liegt vielleicht nicht so sehr in dem behandelten Forschungsgegenstand wie darin, daß mit ihm zum ersten Mal in der Sprachwissenschaft eine Hypothese formuliert wurde, die

- auf metrisierten Begriffen beruhte,
- aus theoretischen Überlegungen abgeleitet,
- falsifizierbar,
- in ein System von Aussagen (über das menschliche Verhalten) eingebettet und
- empirisch gut bestätigt

ist und daher durchaus mit Berechtigung die Bezeichnung Gesetz trägt.

2.1 VERTEILUNGEN FUNDAMENTALER LINGUISTISCHER EIGENSCHAFTEN

Ein weiteres mathematisches Modell für die Worthäufigkeit stammt von J.B. Carroll, der sich insbesondere für den Zusammenhang zwischen Type- und Token-Frequenz interessierte. Er zeigte (1967, 1968, 1969, 1970/72), daß die Lognormale Verteilung empirisch weit bessere Resultate lieferte als die Zipf-Mandelbrotsche. Zudem läßt sich mit diesem Modell die Texthäufigkeit aus der Lexikonhäufigkeit ableiten. Es gibt allerdings keine bekannte Möglichkeit, die Lognormale Verteilung als Modell für die Worthäufigkeit theoretisch zu begründen.

Für andere fundamentale Größen³ wurden ebenfalls statistische Verteilungen gefunden, die das Vorkommen dieser linguistischen Eigenschaften im Inventar der Sprache oder auch ihr Auftreten in Texten beschreiben. Schon früh (s. Fuchs 1955) wurde die Länge sprachlicher Entitäten untersucht und als Indikator für typologische und philologische Zwecke (Autorenbestimmung und Textklassifikation; vgl. bes. Pieper 1979) nutzbar gemacht. Die Ableitung der negativen Binomialverteilung als Modell für die Verteilung der Wortlänge gelang Grotjahn (1982).

Eine fundamentale Eigenschaft von Wörtern ist auch die Anzahl ihrer verschiedenen Bedeutungen. Krylov (1982a, 1982b) hat sich mit der Frage befaßt, wie diese Größe im Lexikon verteilt ist. Er stellte ein kombinatorisches Modell auf und überprüfte es an Material aus der russischen Sprache.

³ Unter einer fundamentalen Größe versteht man eine solche, die nicht aus anderen Größen zusammengesetzt oder abgeleitet werden kann (vgl. Bunge 1967).

2.2 ABHÄNGIGKEITEN ZWISCHEN LINGUISTISCHEN GRÖßEN

Zipf war es auch, der als erster den Zusammenhang zwischen der Länge eines Wortes und seiner Frequenz beobachtete. Er stellte zu seiner Erklärung die Hypothese auf, daß die Wörter einer Sprache bei häufiger Benutzung einem Kürzungsprozeß unterliegen. Andere Autoren beschrieben ebenfalls diese beiden Größen in Abhängigkeit voneinander; in einigen romanischen Sprachen wurden quantitative Untersuchungen durchgeführt (vgl. etwa Guiter 1974).

Über die Form dieser Abhängigkeit herrscht allerdings keine Übereinstimmung - selbst ihre Richtung ist umstritten (s. a. a. O.). Eine Aufgabe der vorliegenden Arbeit wird der Versuch sein, im Rahmen des hier entwickelten Modells eine befriedigende Antwort auf diese Frage zu geben.

Eine Voraussage über die Abhängigkeit der Bedeutungszahl eines Wortes von seiner Länge leitete Altmann (1980) als Konsequenz aus dem Menzerathschen Gesetz (s. u.) ab. In (Altmann, Beöthy, Best 1982) wurde diese Annahme aufgegriffen und überprüft; inzwischen ist sie durch Untersuchungen an Texten in zehn typologisch verschiedenen Sprachen bestätigt worden (s. auch Fickermann, Marker-Jäger, Rothe 1984; Samor 1984). Die Bedeutungskomplexität, wie Altmann diese Größe nannte, wurde von ihm als die Anzahl der Definitionen im Wörterbuch operationalisiert⁴, die Länge als Zahl der Silben oder Grapheme. In den genannten Arbeiten zu diesem Thema wird der Zusammenhang zwischen diesen Größen unter das Menzerathsche Gesetz subsumiert, indem der Bedeutungskomplexität die Rolle der Komponente zugewiesen wird. Diese Auffassung wird hier u. a. wegen der etwas vagen Analogie zwischen Bedeutungszahl und Komponentengröße nicht vertreten, obwohl

4 Eine solche Operationalisierung ist natürlich nicht ganz unproblematisch (vgl. hierzu die Ausführungen zur Polylexie in den Abschnitten 4.6.4 und 6.1); sie läßt sich jedoch rechtfertigen und führt zu stabilen und plausiblen Resultaten.

Interpretationen des Menzerathschen Gesetzes möglich sind, die eine Integration der beiden Bereiche Längen-Längen-Relation und Bedeutungszahl-Längen-Relation in ein Modell nahelegen (s. Altmann, Schwibbe 1985). Hier soll jedoch nicht näher auf das Problem eingegangen werden, da die fraglichen Beziehungen unten auf eine andere Weise gewonnen werden.

Nach Zipf (1949) ist die Bedeutungszahl eines Wortes von seiner Häufigkeit abhängig. Die Form der Beziehung, die Zipf als Hypothese aufstellte, kann als Sonderfall des Menzerathschen Gesetzes angesehen werden. Zipf konnte den Zusammenhang der beiden Größen empirisch nachweisen, zu einer theoretischen Ableitung war er jedoch nicht in der Lage. Seine Formel hat daher lediglich deskriptiven Status.

Eine synoptische Studie der Relationen zwischen den drei fundamentalen Größen Länge, Häufigkeit und Bedeutungszahl wurde von Guiter (1974) durchgeführt. Auch an dieser Stelle zeigte sich wieder, daß (mangels theoretischer Modellvorstellungen) nicht einmal über die Richtungen der Abhängigkeiten Klarheit herrscht - und erst recht nicht darüber, welche mathematischen Funktionen den sprachlichen Verhältnissen adäquat sind.

2.3 STRUKTURGESETZE

Im Bereich der strukturellen Beschreibung von Sprachen und Texten bedingt die Verwendung qualitativer Modelle wie Algebra und Logik eine Beschränkung auf die extrem regelmäßigen Erscheinungen. Alle grammatischen und eine große Zahl stilistischer Phänomene lassen sich aber mit Mitteln der Kombinatorik und der Wahrscheinlichkeitstheorie erfassen, wenn man sie in termini der direkten oder mittelbaren Verkettbarkeit gegebener Einheiten oder der simultanen Realisierung ihrer

Komponenten beschreibt. Mit Wahrscheinlichkeitsmodellen kann das ganze Spektrum zwischen strengem Determinismus und losem Zusammenhang abgedeckt werden; die traditionellen philologischen (sofern sie präzise genug sind) und die formal-linguistischen Darstellungen erweisen sich als durchaus eingeschlossene Spezialfälle der differenzierteren quantitativen Repräsentationsmöglichkeiten⁵.

Obwohl die Sprachwissenschaftler, die sich mit Sprach- oder Textstruktur beschäftigen, sich fast ausnahmslos - bewußt oder unbewußt - auf die Strukturbeschreibung beschränken, gibt es keinen Grund, dieses Gebiet aus den Bemühungen zur Schaffung von Modellen mit erklärender Kraft auszuklammern. Im Gegenteil: Gerade die Konstruktionsprinzipien von Texten und die Aufbaugesetze der Sprache zu finden sollte zentrales Anliegen der Sprachwissenschaft sein, sofern es ihr um den Aufbau einer Theorie geht. Die grammatischen, semantischen und phonologischen Struktureigenschaften aller Sprachen und ihre Verwendungsregularitäten sind erklärungsbedürftig und nur aus einer solchen Theorie erklärbar.

Neben dem Zipf-Mandelbrotschen Gesetz, dessen Gegenstand die Häufigkeitsstruktur von Texten ist, muß in diesem Zusammenhang vor allem auf das Menzerathsche Gesetz eingegangen werden.

Ausgehend von den Beobachtungen Menzeraths (1954), der an deutschen Wörtern eine Korrelation zwischen der Wortlänge

5 Dem immer wieder vorgebrachten Einwand gegen quantitative Syntaxmodelle (der sich in erster Linie auf die Verwendung von Markov-Ketten bezieht), die wegen größerer Allgemeinheit zweckmäßige Annahme einer unendlichen Einbettungstiefe (vgl. Chowsky 1957; Osgood 1963; Levelt 1974) könne mit stochastischen Mitteln nicht wiedergegeben werden, kann leicht durch Einsatz rekursiver probabilistischer Grammatiken (s. z. B. Suppes 1970; vgl. auch Levelt 1974) begegnet werden.

Eine entsprechende, heute in der Computerlinguistik verbreitete Anwendung stochastischer Modelle findet sich z. B. in vielen Ausprägungen von maschinellen Analysesystemen (Parsern), wo die syntaktischen Alternativen stillschweigend mit Wahrscheinlichkeiten bewertet sind.

(gemessen in Silben) und der Silbenlänge (gemessen in Phonemen) feststellte, leitete Altmann (1980) aus theoretischen Überlegungen eine Differentialgleichung ab, deren Lösung allgemein die Abhängigkeit der Komponentenlänge (z. B. Silbe) von der Konstruktlänge (z. B. Wort) darstellt. Er nannte diesen Zusammenhang, der für alle Sprachen und alle linguistischen Beschreibungsebenen Gültigkeit beansprucht, das Menzerathsche Gesetz.

Sämtliche bisher durchgeführten empirischen Untersuchungen (vgl. Gerlach 1980; Heups 1983; Köhler 1982; Schwibbe 1984; Teupenhayn, Altmann 1984) sind gute Bestätigungen dieser Hypothese, die in ihrer allgemeinsten Form folgendermaßen wiedergegeben werden kann:

"Die Länge der Komponenten ist eine Funktion der Länge des aus ihnen gebildeten sprachlichen Konstrukts" (Altmann 1980).

Das Menzerathsche Gesetz läßt u. a. eine Reihe interessanter psycholinguistischer Interpretationen zu und besitzt einige Konsequenzen, die sich - vor allem im lexikalischen Bereich - mit dem Vorhaben der vorliegenden Arbeit berühren: Wie später noch auszuführen sein wird, hat die Differentialgleichung, die dem Gesetz zugrundeliegt, eine fundamentale Bedeutung auch für weitere Zusammenhänge zwischen linguistischen Größen.

2.4 GESETZE DER SPRACHDYNAMIK

Schon die Junggrammatiker versuchten, bestimmte Veränderungen der Sprachsysteme, nämlich Veränderungen von Eigenschaften ihrer Elemente, über die bloße Beschreibung hinaus mit Hilfe von

Gesetzen - den Lautgesetzen - zu erklären. Die Annahme einer ausnahmslosen, deterministischen Gesetzmäßigkeit für den Lautwandel ist die Hauptsache für das Scheitern ihres Programms; stochastische Modelle, die Tendenzen und allmähliche Zu- und Abnahme einer Erscheinung erfassen, standen ihnen noch nicht zur Verfügung.

Diese Beschränkung besteht heute nicht mehr, so daß der Sprachwandel im Sinne einer "exakten Sprachwandelforschung" (vgl. Best, Kohlhasse 1983) als von stochastischen Prozessen gesteuert modelliert werden kann. Empirische Grundlage für diese Forschung sind einerseits sich entsprechende Daten aus verschiedenen Sprachstufen, andererseits die gesetzmäßige Organisation und Verteilung der lexikalischen Einheiten (auf einer Sprachstufe betrachtet) im Hinblick auf ihr Alter und ihre Herkunft.

Piotrovskij, Bektaev und Piotrovskaja (1977) formulierten ein Exponentialgesetz für die Zunahme von Wörtern eines jeden semantischen, terminologischen oder anders bestimmten Bereichs der Lexik im Laufe der Zeit. Weiter läßt sich zeigen, daß die Aufnahme sprachfremder lexikalischer Elemente nach einem anderen 'Piotrovskij'-Gesetz (vgl. etwa Beöthy, Altmann 1982) verläuft. Dieses Gesetz wurde von Altmann aufgrund einiger theoretischer Überlegungen zur Ausbreitung sprachlicher Neuerungen beliebiger Art abgeleitet. Es wurde bereits für Entwicklungstendenzen im Russischen sowie für die Entlehnungsdynamik des Deutschen und des Ungarischen angewendet (s. Altmann, v. Buttlar, Rott, Strauß 1983; Best, Kohlhasse 1982).

Ein weiterer Ansatz, der im Prinzip auf die Glottochronologie zurückgeht, stammt von Arapov und Cherc (1983). Er beschreibt den Zerfall des Lexikons einer Sprache in Abhängigkeit von Frequenzklasse und Alter der Wörter.

In allen Bereichen ist im Verlauf der letzten dreißig Jahre ein beträchtlicher methodologischer Fortschritt zu verzeichnen. Sowohl im empirischen Bereich, in dem die Einführung

quantitativer Begriffe und immer anspruchsvollerer Verfahren zu zahlreichen wichtigen Erkenntnissen und Hypothesen geführt hat, als auch in der theoretischen quantitativen Linguistik, in der die deduktive Modellgewinnung eine zunehmende Rolle spielt, ist die Tendenz zu explanativen Fragestellungen und Zielsetzungen unverkennbar.

Obwohl in jüngster Zeit ein gutes Dutzend von Sprach- und Textgesetzen gefunden worden sind, ist es bisher nicht gelungen, die Randbedingungen oder die für ihre Parameter bestimmenden Faktoren zu determinieren. Auch besteht bisher kein Zusammenhang zwischen den verschiedenen Hypothesen und Gesetzen; es existiert vor allem noch kein allen gemeinsamer theoretischer Hintergrund.

die Parameter der Gesetze sowie ihre Randbedingungen zu determinieren,

- Erweiterungen in zweifacher Hinsicht vorzunehmen: zusätzliche Variablen aufzunehmen und zusätzliche Relationen zu berücksichtigen.

Schließlich soll aus dem Modell hervorgehen, auf welche Weise es selbst aus wiederum höheren Prinzipien abgeleitet und erklärt werden kann.

Die Modellierung erfolgt axiomatisch und auf einer Abstraktionsebene, die oberhalb der linguistischen Ebene liegt; die vorliegende Untersuchung setzt also bereits eine linguistische Begriffsbildung voraus.

An dieser Stelle muß betont werden, daß hier nicht das Ziel verfolgt wird, einen linguistischen Bereich mit dem zu entwickelnden Modell in jeder Hinsicht vollständig zu erfassen, zumal ein geeignetes Vollständigkeitskriterium für ein solches dynamisches Modellsystem zu finden nur unter Zuhilfenahme theoretischer Überlegungen mit Erklärungsgehalt gelingen könnte; diese aber sind ja gerade erst zu entwickeln. Ein wichtiger pragmatischer Grund für die Selbstbeschränkung liegt in der Gefahr der Komplexitätsexplosion. Hauptanliegen dieser Arbeit ist die Entwicklung eines Ansatzes bzw. eines Modellprinzips; diese soll zwar anhand eines nichttrivialen Ausschnitts des Gegenstandsbereichs vorgenommen werden, andererseits aber nicht in der Darstellung von Detailproblemen untergehen.

Das vorgestellte Modell enthält daher als Elemente einige wenige Größen, die auch bisher als besonders wichtig angesehen und bereits intensiv (wenn auch nur jede für sich isoliert) untersucht worden sind.

Es handelt sich um die Variablen

3. ZIELSETZUNG UND METHODIK

Ziel der vorliegenden Arbeit ist es, einen Schritt in Richtung auf die Integration der verschiedenen Ansätze in ein gemeinsames Modell zu tun; denn ein Fortschritt in diesem Punkt ist Voraussetzung für die Entwicklung einer linguistischen Theorie, wie sie in Altmann (1978) skizziert ist.

Der vordringlichste Schritt auf dem Weg zur Theoriebildung ist gegenwärtig die Verknüpfung bekannter und neuer Gesetze bzw. Gesetzeskandidaten untereinander, um so ein System⁶ von allgemeinen Aussagen zu erhalten. Hier soll nun versucht werden, ausgehend von einem Kern linguistischer Zusammenhänge, ein Basismodell zu erstellen, das zunächst aus wenigen, als zentral angenommenen Variablen besteht. Dieses Modell soll es aber bereits erlauben,

- wenigstens einige der wichtigsten, bisher theoretisch isolierten Gesetze der quantitativen Linguistik abzuleiten,

- weitere Gesetze zu deduzieren und bisher nicht bekannte Zusammenhänge zu prognostizieren,

⁶ Der Ausdruck "System" wird in dieser Arbeit in mehreren Bedeutungen verwendet:

1. System von Gesetzen (Theorie);
2. Sprach(sub)system (Forschungsgegenstand der Linguistik);
3. Modellsystem (Teil des linguistischen Modells, systemtheoretische Abbildung).

Nur wenn der Kontext Zweifel offen läßt, wird auf die richtige Lesart hingewiesen.

- a) Länge,
- b) Bedeutungspotenz (Polylexie) und
- c) Gebrauchshäufigkeit (Frequenz)

lexikalischer Einheiten. Aufgrund theoretischer Überlegungen hat sich gezeigt, daß eine weitere Größe, die

- d) Kontextgebundenheit (Polytextie),

eng mit diesen verbunden ist und in das Modell aufgenommen werden muß. Hinzu kommen ergänzend die globalen Größen

- e) Lexikonumfang und
- f) Phonemanzahl.

Dargestellt wird das Modellsystem mit Mitteln der Graphenalgebra und der linearen Operatoren in Form eines Regelkreises.

Das vorgeschlagene Modell soll es insbesondere gestatten, beobachtbare Zusammenhänge zwischen linguistischen Größen zu erklären. Dies wird erreicht, indem die Relationen zwischen den Elementen des Systems in Form von Differentialgleichungen theoretisch begründet werden, so daß Aussagen mit dem Status von Gesetzhypothesen entstehen.

Grundlage und Voraussetzung des gesamten Ansatzes ist die Annahme, daß die menschliche Sprache (neben anderen Zeichensystemen) ein selbstregulierendes System ist. Dadurch ist es möglich, auf die Funktionalanalyse als Erklärungsrahmen zurückzugreifen, ohne Gefahr zu laufen, teleologischen oder anderen Pseudoerklärungen aufzusitzen. Allerdings wird dazu die herkömmliche funktionale Erklärungslogik, wie sie aus der wissenschaftstheoretischen Literatur vor allem der Sozialwissenschaften bekannt ist, zu modifizieren und um einige Elemente zu erweitern sein.

Soll das Modell, wie angekündigt, zur Theoriebildung in der Linguistik beitragen, so ist die empirische Überprüfung der auf theoretischem Wege gewonnenen Hypothesen zwingend erforderlich, um Realitätsbezug und Gültigkeit nachzuweisen. Daher werden die postulierten Zusammenhänge und die Hauptkonsequenzen, die beim gegenwärtigen Stand der Modellbildung aus dem Modell ableitbar sind, anhand von Stichproben aus Textmaterial in deutscher Sprache mit Hilfe statistischer Testmethoden überprüft.

Abschließend soll die mögliche Bedeutung des vorgestellten Ansatzes für den Aufbau einer Sprachtheorie und für die Verankerung linguistischer Erklärungen in einer allgemeinen Theorie der selbstregulierenden Systeme diskutiert werden.

Betrachtet man die Eigenschaften von Phonemen oder Graphemen im Hinblick auf diese beiden Bedürfnisse, so stellt man leicht fest, daß - bei einem gegebenen Inventar von Phonemen / Graphemen - sowohl der Gedächtnis- wie auch der Produktionsaufwand für diese Elemente um so geringer ist, je ähnlicher sie sich sind. Das Gedächtnis wird um so weniger belastet, je weniger Merkmale behalten werden müssen; das Optimum⁸ ist erreicht, wenn sämtliche Elemente⁸ lediglich die Merkmale mit distinktiver Funktion tragen.

4.1 EINFÜHRUNG: DIE ELEMENTE AUF DER PHONEM- / GRAPHEMBEBENE

Zipf hatte zur Erklärung der Rang-Frequenz-Verteilungen von Wörtern das von ihm postulierte principle of least effort (Zipf 1949) herangezogen, das auf einem optimalen Ausgleich der einander entgegengesetzt wirkenden Unifikations- und Diversifikationskräfte beruht.

Ein ähnlicher Ansatz wird hier als Ausgangspunkt für die Modellierung der Mechanismen auf der Ebene der Phoneme oder Grapheme gewählt.

Postuliert wird zunächst die Existenz zweier Bedürfnisse der Sprachbenutzer, denen ein Sprachsystem mehr oder weniger gerecht werden kann:

- die Minimierung des erforderlichen Gedächtnisaufwands (im folgenden minG) und
- die Minimierung des erforderlichen Produktionsaufwands (im folgenden minP).

Diese Bedürfnisse sind als Anforderungen an das Sprachsystem zu interpretieren; sie rufen die "Unifikationskraft" hervor und gehören nicht selbst zum System.

Die Produktion ist weniger aufwendig, wenn sie aus einheitlichen Artikulations- bzw. Bewegungsvorgängen besteht: Beim Sprechen ist es einfacher, Artikulationsmerkmale (Zungenposition, Stimmhaftigkeit ...) beizubehalten, als sie zu wechseln (vgl. z.B. Lindblom 1983). Analoges gilt für das Schreiben (vgl. hierzu auch Watt 1983). Daher werden bei der Aufnahme von Fremdwörtern die neuen Phoneme, sofern sie überhaupt beibehalten werden, an die Merkmale des aufnehmenden Systems angepaßt; daher entwickeln sich Phonem- und Schriftsysteme weiter.

Der Zusammenhang zwischen den Bedürfnissen und der Ähnlichkeit der Elemente eines Inventars ist in Abbildung 1 schematisch dargestellt. Die Ähnlichkeit der Zeichen untereinander wird als globale Eigenschaft des Zeicheninventars aufgefaßt.

In dieser und den folgenden Abbildungen bedeutet das Pluszeichen auf der Kante eines Pfeils einen positiven Wirkungszusammenhang zwischen der Ausgangsgröße und der Größe, auf die der Pfeil weist. Ein Minuszeichen symbolisiert einen inversen Zusammenhang: Je stärker die Ausgangsgröße ausgeprägt ist, desto mehr wirkt sie auf die Verringerung der Werte der Zielgröße.

- 7 Für Inventare mit mehr als einem Element.
- 8 Ein Inventar aus einem einzigen Element mit einem einzigen Merkmal wäre das theoretische Optimum in Bezug auf minG.

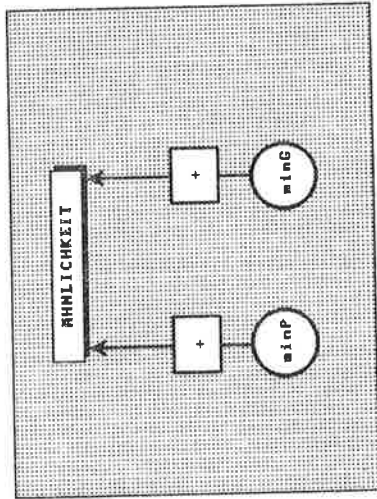


Abb. 1: Die Systemgröße ÄHNLICHKEIT ist eine Funktion der beiden Bedürfnisse minP und minG.

Nach dem bisher Gesagten müßte ein Zeichensystem von Phonemen oder Graphemen, das den genannten Bedürfnissen seiner Benutzer optimal gerecht würde, aus Zeichen maximaler Ähnlichkeit bestehen. Im Extremfall bedeutet maximale Ähnlichkeit den Zusammenfall aller Zeichen zu einem einzigen. Dies beträfe dann zusätzlich eine Systemgröße P-INVENTARUMFANG, die mit dem Bedürfnis minG im gleichen Verhältnis steht wie die Zeichenähnlichkeit: Je kleiner der P-INVENTARUMFANG, desto geringer der Gedächtnisaufwand.

Nehmen wir jedoch ein (der Zipfschen Diversifikation entsprechendes) Bedürfnis, nämlich

die Minimierung des Wahrnehmungs- und Dekodierungsaufwands (minW),

hinzu, bietet sich ein anderes Bild. Um die Verständlichkeit bzw. Lesbarkeit zu erleichtern, müssen die Zeichen möglichst unähnlich

sein (Merkmalsredundanz); die minW wirkt also den beiden anderen entgegen: verringern auf die ÄHNLICHKEIT und damit indirekt erhöhend auf den P-INVENTARUMFANG.

In der Linguistik wurde dieses Prinzip mit der Bezeichnung maximale Distinktivität von den Strukturalisten eingeführt und bei der Modellierung von Phonemsystemen angewandt. Eine Verwendung bei der quantitativen Betrachtung von Phonemsystemen findet man in Strauß (1980).

Eine Veränderung der Größe P-INVENTARUMFANG hat aber ihrerseits eine Konsequenz, die berücksichtigt werden muß. Je weniger Elemente auf der Ebene der zweiten Gliederung zur Verfügung stehen, desto länger werden die aus ihnen gebildeten Elemente der ersten Gliederung, die lexikalischen Einheiten, wenn deren Zahl konstant gehalten wird. So sind aus einem Phoneminventar, das aus dem einzigen Element /a/ besteht, nur die Wörter /a/, /aa/, /aaa/,... bildbar - also nur ein Wort jeder Länge n. Aus einem Inventar der Größe x lassen sich x Wörter der Länge 1, x² Wörter der Länge 2, allgemein xⁿ Wörter der Länge n bilden.

Die Bedürfnisse minP und minG verlangen aber eine möglichst kleine Ausprägung der Größe WORTLÄNGE. Ein Modell, das diese Beziehungen zwischen den drei Größen ÄHNLICHKEIT, P-INVENTARUMFANG und WORTLÄNGE sowie den drei Bedürfnissen minG, minP und minW berücksichtigt, ist in Abbildung 2 dargestellt.

In dieser Konstellation sind einander direkt und indirekt widersprechende Anforderungen enthalten, daher kann die optimale Ausprägung einer Größe nicht mehr durch Annehmen eines Extremwerts erreicht werden. Vielmehr verlangt die gleichzeitige Berücksichtigung aller Bedürfnisse nach der Herausbildung eines optimalen Gleichgewichts als Kompromiß.

9 Wie eine große ÄHNLICHKEIT zu einem geringen P-INVENTARUMFANG führt, so verursacht umgekehrt die Verringerung der ÄHNLICHKEIT durch minW einen erhöhten P-INVENTARUMFANG.

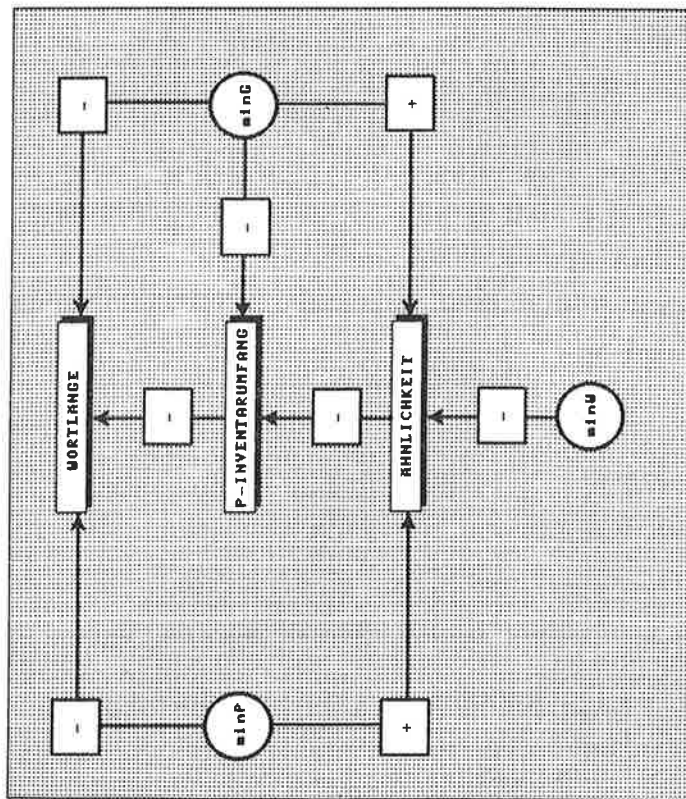


Abb. 2: Subsystem mit drei Größen und drei Systembedürfnissen

Dieses einfache Modell vernachlässigt jedoch u. a. den Einfluß von Tönen und Akzenten auf die Wortlänge: Tonsprachen kommen mit kürzeren Wörtern aus, da sie über ein zusätzliches Merkmal verfügen. In solchen Sprachen wirkt sich die Anzahl der Töne ver- ringernd auf die Wortlänge aus. Sprachen mit beweglichem Akzent besitzen ebenfalls ein zusätzliches Mittel zur Differenzierung, das man z. B. häufig in der Flexion findet.

Selbstverständlich wären in dieses System weitere wichtige Größen einzubeziehen wie die Phonen- bzw. Graphemdistribution,

die HÄUFIGKEIT der einzelnen Elemente, ihre KOMPLEXITÄT und die Parameter ihrer Verteilungen. Doch interessieren sie an dieser Stelle nicht direkt, da hier nicht primär ein Modell für diese Ebene der Phoneme oder Grapheme erstellt werden soll; es genügt für unsere Zwecke, den Zusammenhang zwischen der Bedürfnissen und den Größen P-INVENTARUMFANG und WORTLÄNGE als Schnittstelle eines solchen zu dem System auf der lexikalischen Ebene zu erfassen¹⁰.

4.2 SPRACHE - EIN SELBSTREGULIERENDES SYSTEM

In den Überlegungen des letzten Abschnitts dienen die Systembedürfnisse als Instanzen zur Erklärung von beobachteten oder angenommenen Phänomenen. So sagen wir vom Inventarumfang des Phonemsystems, er habe eine bestimmte Größe, weil er aufgrund des Bedürfnisses minG den geringsten Wert annimmt, der bei einem gegebenem Wert der Größe Phonomähnlichkeit möglich ist. Offenbar handelt es sich dabei nicht um eine Kausalerklärung. Nach Hempel und Oppenheim (z. B. Hempel 1965) folgt die deduktiv-nomologische Erklärung dem allgemeinen Schema:

$G_1, G_2, \dots, G_n$	}
$S_1, S_2, \dots, S_n$	}
	Explanans

E	Explanandum

¹⁰ Weitere Einflüsse auf die Wortlänge wie die Notwendigkeit von Redundanz und die Aufwandsökonomie betrachten wir als zu lexikalischen Ebene gehörend; sie werden daher in einem späteren Abschnitt behandelt.

G₁ ... G_n sind allgemeine Gesetze, während S₁ ... S_n Sätze darstellen, die bestimmte Umstände beschreiben. Gesetze gelten immer unter spezifizierten Randbedingungen; wenn die Umstände den Randbedingungen entsprechen, kann man von ihnen und den Gesetzen deduktiv auf das Explanandum schließen.

Für die Sprache sind jedoch Gesetze, die allgemeine Bedürfnisse der menschlichen Kommunikation (wie die Minimierung des Gedächtnisaufwands) mit einer Eigenschaft des sprachlichen Systems (wie die Größe des Phoneminventars) kausal verknüpfen, weder bekannt, noch ist es plausibel, die Existenz solcher Gesetze anzunehmen.

Eine Systemeigenschaft wird vielmehr damit erklärt, daß sie für die Funktion des Systems notwendig ist. So ist z.B. ein Phoneminventar von ausreichender Größe notwendig, damit genügend viele verschiedene Morpheme gebildet werden können. Eine solche funktionale Erklärung ist ein besonderer Typ der deduktiv-nomologischen Erklärung, der eine Reihe von Problemen aufwirft.

Ein wichtiges Problem ist das der funktionalen Äquivalente (Hempel 1965a; Easer, Klenovitz, Zehnpfennig 1977): Die Existenz eines Elements, welches eine bestimmte Funktion innerhalb des Systems hat, ist durch sie noch nicht hinlänglich erklärt, wenn auch andere Elemente die gleiche Funktion (evtl. auf andere Weise) ausüben könnten. In praktisch allen komplexen Systemen (z.B. sprachlichen, biologischen, soziologischen u.a.) besteht diese Möglichkeit. Sehr oft existieren in verschiedenen verwandten oder vergleichbaren Systemen unterschiedliche Elemente für die gleiche Funktion; sogar die gleichzeitige Existenz mehrerer funktionaler Äquivalente ist häufig zu finden. So genügt es nicht, eine notwendige Anzahl von Morphemen als Erklärung für die Größe eines Phoneminventars anzugeben, denn die Bildung längerer Morpheme aus weniger verschiedenen Phonemen und/oder die Verwendung von Tönen ist im Hinblick auf die Morphemzahl äquivalent.

Die gleiche Problematik ist aus anderen Wissenschaftsgebieten bekannt: Die Existenz eines Herzens im Säugetierorganismus läßt

sich nicht durch Hinweis auf seine Funktion als Blutkreislaufpumpe erklären, da andere Lösungen für die Versorgung des Organismus nicht nur denkbar sondern auch in der Natur verwirklicht sind (so z.B. durch ein offenes System bei Würmern). - Die Tätigkeiten von Menschengruppen lassen sich unter Anwendung des Prinzips der Führung und Delegation koordinieren, aber auch durch Diskussion und Absprache, durch vererbte Rollenzuweisungen mit Arbeitsteilung und vieles mehr.

Eine befriedigende Erklärung eines Äquivalents ist nur denkbar, wenn die Beziehungen zwischen allen funktionalen Äquivalenten einer Funktion angegeben werden können¹¹. Altmann fordert (in 1981) die Ableitung dieser Beziehung aus Axiomen, theoretischen Annahmen oder Gesetzen. In der vorliegenden Arbeit wird versucht, dieser Forderung nachzukommen: Die Beziehungen der Systemelemente untereinander folgen aus theoretischen Annahmen und werden explizit modelliert¹².

Die zweite fundamentale Frage in Zusammenhang mit der Erklärung in der Funktionalanalyse ist die nach der logischen Verbindung zwischen der Existenz eines Bedürfnisses und seiner Befriedigung durch ein Element bzw. eine Eigenschaft des Systems. Auf allgemeine Gesetze, aus denen bei Vorliegen von Bedürfnissen auf deren Bedienung kausal geschlossen werden dürfte, können wir nicht zurückgreifen. Viel einleuchtender ist es dagegen, Sprache als ein selbstregulierendes System aufzufassen (vgl. Altmann 1981). Als solches besitzt die Sprache Mechanismen, die den Systemzustand ständig optimal an den Sollzustand anzupassen versuchen; dabei kann der Sollzustand als Vektor der Systembedürfnisse interpretiert werden. Die Existenz einer Systemeigenschaft,

11 Ein grundsätzliches logisches Problem liegt darin, daß nicht alle funktionalen Äquivalente bekannt sind. Theoretisch können unendlich viele verschiedene Äquivalente für eine gegebene Funktion existieren.

12 Die Existenz von im Modell nicht berücksichtigten Größen wird dagegen nicht ausgeschlossen - im Gegenteil: Es handelt sich ausdrücklich um ein embryonales Teilmodell.

die ein Bedürfnis bedient, kann dann unter Berufung auf einen derartigen Anpassungsmechanismus erklärt werden.

Eine vollständige funktionalanalytische Erklärung einer Erscheinung E , müßte in etwa nachstehendem Schema folgen:

- a) Das System S ist selbstregulierend: Für jedes Bedürfnis besteht ein Mechanismus, der den Systemzustand so verändert, daß es bedient wird.
- b) An das System S sind die Bedürfnisse $B_1 \dots B_n$ gestellt.
- c) Das Bedürfnis B kann durch die funktionalen Äquivalente $E_1 \dots E_r \dots E_n$ bedient werden.
- d) Zwischen den funktionalen Äquivalenten besteht die Relation $R(E_1, E_2, \dots, E_n)$.
- e) Aufgrund der Systemstruktur besteht zwischen den Elementen $s_1 \dots s_n$ des Systems S die Relation $Q(s_1, \dots, s_n)$.

E_r ist Element des Systems mit der Ausprägung R_r

Diese Erklärung ist nur schlüssig, wenn aufgrund der Relationen R und Q alle anderen Möglichkeiten zur Bedienung der B_i ausgeschlossen oder ungünstiger als E_r sind.

Abgerundet werden müßte eine solche Funktionsanalyse durch Angabe der Funktionen $R_i(E_1, \dots, E_n)$, die für jedes Bedürfnis B_i alle Ausprägungen der funktionalen Äquivalente E_r so bestimmt, daß die B_i optimal bedient werden. Solche Funktionen können nicht empirisch gewonnen werden; sie bedürfen der theoretischen Ableitung.

Das oben entworfene Schema erlaubt nur ein statisches Bild vom Zusammenhang zwischen Bedürfnissen und Systemeigenschaften. Selbstregulierende Systeme sind aber notwendigerweise als dynamisch zu betrachten. Ein theoretisch wie praktisch wichtiger Aspekt ist der Zeitbedarf der Anpassungsmechanismen. Eine Veränderung in der Systemumgebung (z.B. den benutzten Kommunikationsmedien) veranlaßt das System zur allmählichen Veränderung einiger Eigenschaften. Wegen der Verknüpfung der Systemeigenschaften untereinander führt dies wieder zu anderen Veränderungen und so fort, bis sich ein neuer Systemzustand (ein Gleichgewicht) eingependelt hat. Durch Überkompensation kann dieser Prozess noch verzögert werden. Aufgrund der Komplexität des Systems ist ein solches Gleichgewicht wohl immer ein Kompromiß zwischen einer großen Zahl von divergierenden Anforderungen an die meisten Systemelemente, das zudem noch zeitlich hinter den Entwicklungen der Bedürfnisse hinterherhinkt.

Der durch die Systemumgebung definierte Sollzustand ist selbst keine konstante Größe. Die gesellschaftlichen Bedingungen, in die eine Sprache eingebettet ist, ändern sich recht schnell, so daß sogar vor Erreichen eines optimalen Kompromisses die Bedürfnisse bereits wieder anders ausgeprägt sein können. Wir sprechen daher auch von einem Fließgleichgewicht der Systemeigenschaften.

Aus all diesen Gründen befindet sich eine Sprache zu keinem Zeitpunkt in einem Zustand, der für alle Eigenschaften dem Sollzustand entspricht, sondern stets in Bewegung¹³ (die Dynamik

¹³ Hierzu und zum Problem der Ausdrucksadäquatheit vgl. Altmann (1981: 28).

einer einzelnen Eigenschaft kann durch die Abbildung 3 veranschaulicht werden).

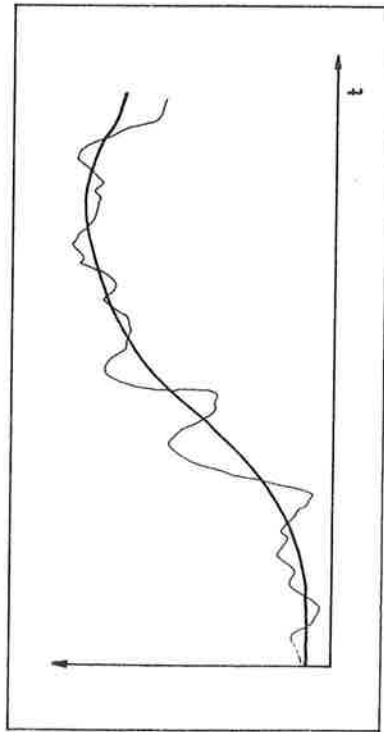


Abb. 3: Die Dynamik einer Spracheigenschaft. Die fette Linie repräsentiert die theoretische Funktion $R(E, \dots, E)$, die der optimalen Bedienung eines Bedürfnisses entspricht. Die dünne Linie gibt die empirische Ausprägung der Funktionswerte wieder.

Linguistische Daten lassen sich daher meist nur mit Hilfe statistischer Mittel erfassen. Die in Form von deterministischen Regeln darstellbaren Erscheinungen sind in der Minderzahl, und auch sie sind infolge der Anpassungsmechanismen dem Wandel unterworfen. Ein linguistischer Ansatz, der Sprache als selbstregulierendes, dynamisches System begreift, muß unter die Oberfläche dieser momentanen Phänomene schauen und die tieferliegenden Gesetze der Systemstruktur und -dynamik erforschen.

Wie sehen nun die Anpassungsmechanismen aus? Wie wirken sie auf die Systemeigenschaften ein, und wie werden Soll- und Ist-Werte miteinander verglichen? Die Annahme der Selbstregulation

bedeutet keineswegs den Versuch einer teleologischen Erklärung: Die in der Sprache stattfindenden Veränderungen werden nicht bewirkt, um ein vorgegebenes Ziel zu erreichen. Vielmehr ist die mit der Selbstregulation verbundene Vorstellung im großen und ganzen dem biologischen Modell für die Entwicklung der Arten, der Evolution, bzw. der Anpassung von Organismen¹⁴ vergleichbar. Analog zum Prinzip von Mutation und Selektion entstehen bei der Verwendung von Sprache ständig Abweichungen und Variationen, von denen sich einige durchsetzen können, andere nicht. Langfristig bleiben nur solche Varianten erhalten, die in irgendeiner Weise dazu beitragen, daß das System insgesamt den Bedürfnissen gerecht zu machen. Der eigentliche explanative Hintergrund beruht also auf der Behauptung, daß durch Anpassungsdruck diejenigen Systemeigenschaften entstehen bzw. überleben, die die "Lebensfähigkeit" des Systems in seiner Umwelt erhalten bzw. erhöhen.

Diese Aussage wird als das Axiom der sprachlichen Selbstregulation in das Modell eingeführt.

Zur Erläuterung der Mutations- und Selektionsprozesse in der Sprache kommen wir noch einmal auf das Beispiel der Bedürfnisse minP, minG und minW zurück. Jede individuelle Äußerung eines Ausdrucks unterscheidet sich von allen anderen in einer Vielzahl von Eigenschaften, jede Realisation eines Phons fällt anders aus - auch bei demselben Sprecher. Das ist eine stete Quelle von phonetischer Variation, die von Sprechern und Hörern die meiste Zeit praktisch unbemerkt bleibt.

14

Unter den Sprachwissenschaftlern des 19. Jahrhunderts (vgl. bes. Steinthal 1881) war die Auffassung von der Sprache als Organismus vorherrschend. Ihre Sichtweise war jedoch sehr idealistisch und einseitig; vor allem fehlten ihnen die methodologischen Voraussetzungen, um über ihren ersten, naiven Ansatz hinauszukommen.

Eine ähnliche Situation ergab sich beim Aufkommen des strukturalistischen Paradigmas mit seinem ebenfalls einseitigen und beschränkten Systembegriff, der erst seit der Ausarbeitung der Systemtheorie überwindbar geworden ist. Unten wird zu zeigen versucht, auf welche Weise die beiden Anschauungen in dem viel umfassenderen, dynamischen Systemkonzept der modernen Systemtheorie aufgehen.

In bezug auf einen einzelnen Sprecher bedeutet das Bedürfnis zur Minimierung des Produktionsaufwands nichts anderes als seine Tendenz, verstärkt solche phonetischen Variationen hervorzu- bringen, die eine Einsparung an artikulatorischer Anstrengung bewirken. Das Bedürfnis nach Minimierung des Gedächtnisaufwands repräsentiert die Tendenz, häufiger auf Unterscheidungen zu ver- zichten, distinktive Elemente nicht zu benutzen und sie im Extremfall ganz aus dem Inventar zu streichen. Beide Tendenzen sorgen für die Selektion vereinfachender Mutationen.

Da der gleiche Vorgang sich bei jedem Sprecher unabhängig von allen anderen abspielt, würde er, wenn es keine Rückkopplung gäbe, theoretisch zu einer großen Zahl verschiedener, äußerst simpler phonetisch-phonologischer Subsysteme führen. Diese loka- len Mikroprozesse interagieren jedoch z.B. mit den entgegenge- setzten Wirkungen der Hörer-Bedürfnisse (der eigenen und derer anderer Personen). Das Bedürfnis nach Minimierung des Wahr- nehmungs- und Dekodierungsaufwands verhindert ein solches Resul- tat gründlich. Mangelnde Distinktivität bei der Artikulation erschwert die Dekodierung durch den Hörer, so daß zusätzlicher Aufwand getrieben werden muß (größere Konzentration, evtl. mehr- fache Wiederholung der Äußerung). Wenn der auf diese Weise ent- stehende Mehraufwand die Aufwandsersparnis durch Vereinfachung der Artikulation übersteigt, werden die entsprechenden Varianten vermieden (die Mutationen sterben aus).

Abstrahiert man von den lokalen Mikroprozessen, die mit den menschlichen Individuen assoziiert sind, so stellt sich ihr sum- mativer Effekt als Anpassungsmechanismus dar, der ein Gleichge- wicht zwischen den Sprecher- und Hörer-orientierten Bedürfnissen herstellt.

Für jeden Laut einer Sprache kann man daher die Veränderungs- wahrscheinlichkeit als Funktion der Artikulationsanstrengung und des Dekodierungsaufwands ableiten. Dies haben Job und Altmann (1985) mit der Differentialgleichung

$$\frac{dW}{W} = \left(\frac{k}{1-A} - \frac{b}{A} \right) dA$$

getan, in der W für den Wandelrang, A für die Artikulationsanstrengung und $1-A$ für den Dekodierungsaufwand¹⁵ stehen. Die Lösung $W = c(1-A)^{-k}A^{-b}$ ist in der Abbildung 4 gra- phisch dargestellt. Es ist leicht zu sehen, daß die Wahr- scheinlichkeit¹⁶ für einen Lautwandel dann am geringsten ist, wenn zwischen den Wirkungen beider Bedürfnisse ein Ausgleich besteht.

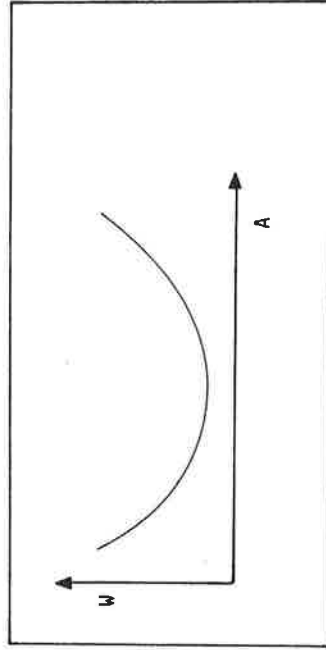


Abb. 4: Der Wandelrang (W) als Funktion der Artikulations- anstrengung (A)

15 Der Dekodierungsaufwand ist um so größer, je geringer die Artikulationsanstrengung ist, daher $1-A$.

16 Die Wahrscheinlichkeit läßt sich durch Normierung der Fläche unter der Kurve berechnen.

4.3 LINGUISTIK UND SYNERGETIK

Selbstregulationsphänomene werden in anderen Disziplinen bereits seit einiger Zeit systematisch erforscht; vor allem in der Biologie (Eigen 1971), der Chemie (Prigogine 1979: V) und der Physik (Haken 1978). Der Physiker Hermann Haken baute darauf - besonders aber auf seinen eigenen Forschungen zum Laser - eine umfassende Theorie der kooperativen Phänomene auf: die Synergetik. Ihre grundlegenden Prinzipien sind so allgemein, daß die sprachliche Selbstregulierung lückenlos in ihren Beschreibungs- und Erklärungsrahmen paßt.

Zur Illustration soll wieder die Anpassung der phonetischen phonologischen Systemeigenschaften dienen. Der effektive Phonemvorrat einer Sprache ist direkt abhängig von der Anzahl von Merkmalen, die von den Sprechern der Sprache zur Distinktion verwendet werden. Wenn sich nämlich aufgrund der Bedürfnisse $\min P$ und $\min G$ phonetische Varianten durchsetzen, die bestimmte Unterschiede verwischen, fallen wenigstens zwei Phoneme zusammen, und das Phoneminventar verringert sich entsprechend. Dies ist z.B. dann der Fall, wenn ein bislang distinktives Merkmal aus Gründen der Artikulationserleichterung mehr und mehr in komplementärer Distribution auftritt, wenn also zwei Phoneme zu bedingten Allophenen werden. Der umgekehrte Vorgang findet statt, wenn aufgrund des Bedürfnisses $\min W$ stärker distinktive Mutationen selektiert werden: Dies kann zur Einführung eines Phonems (z.B. durch Aufspaltung von bis dahin komplementär distribuierten Allophenen in eigenständige Phoneme) führen.

Obwohl Phoneme abstrakte, theoretische Konstrukte sind, stellen sie insofern also ein direktes Maß für den Gesamtaufwand an phonetischer Distinktion, die von einer Sprachgemeinschaft geleistet wird, dar. Was im Zusammenhang mit den Bedürfnissen $\min P$, $\min G$ und $\min W$ für phonetische Merkmale gesagt wurde, gilt ebenso

für das gesamte Phonemsystem: je weniger Phoneme, desto größer der Dekodierungsaufwand (entweder sind längere Morpheme und Wörter zu verarbeiten, oder es entsteht ein Mehraufwand bei der Disambiguierung von Homophonen), je mehr Phoneme, desto größer der Produktionsaufwand (erhöhte Distinktivität ist mit erschwerter Artikulation verbunden).

Für die Veränderungswahrscheinlichkeit eines jeden einzelnen Phonems gibt es, wie oben gezeigt, ein Minimum, das bei einem bestimmten Verhältnis zwischen den beiden Aufwandsarten A und 1-A liegt. In der Sprache der Synergetik gesprochen besteht an der Stelle dieses Minimums ein stabiles Gleichgewicht. Ein Phonem, dessen Produktionsaufwand an einer anderen Stelle liegt, verändert sich mit hoher Wahrscheinlichkeit so, daß sich seine Koordinaten schließlich im Minimum einpendeln - so wie eine Kugel die Wand einer Schüssel hinabrollt und nach einigem Auf und Ab an der tiefsten Stelle liegen bleibt (s. Abb. 5).

Dieses Gleichgewicht bleibt aber nur ceteris paribus stabil. Verändert sich ein anderer Parameter, eine mit dem Subsystem verbundene Größe, so verschiebt sich auch das Minimum. Unter Umständen kann nun eine besondere Situation eintreten.

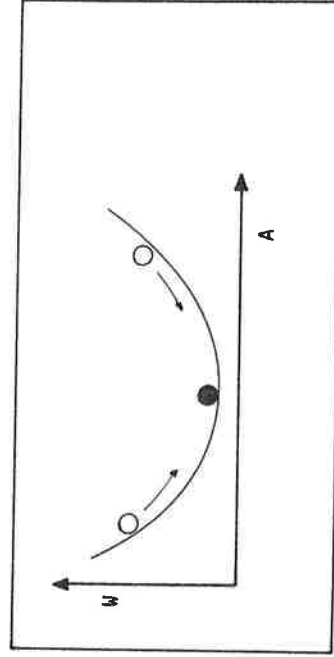


Abb. 5: Stabiles Gleichgewicht zwischen den Aufwandsarten A und 1-A für ein Phonem

Angenommen, durch erhöhte Frequenz von aus einer anderen Sprache übernommenen Ausdrücken spielt eine vorher nicht vorhandene oder redundante Distinktion eine immer größere Rolle, d.h. es wird ein zusätzliches Phonem importiert. Dann kommt es in jedem Fall (wegen der Maximierung der Distinktivität¹⁷) bei einem, meist mehreren der alten Phoneme zu einem Symmetriebruch (vgl. Abb. 6).

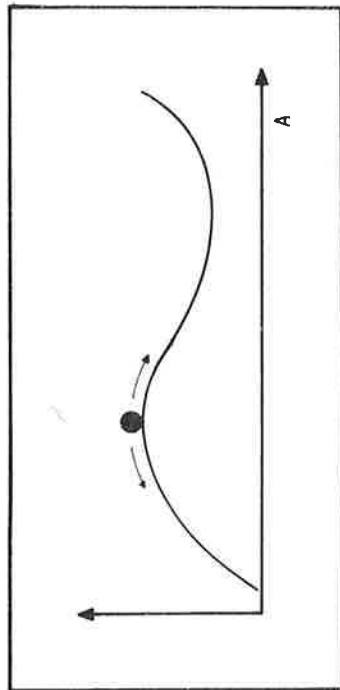


Abb. 6: Instabiles Gleichgewicht

Da die Koeffizienten in der Funktion für den Wandelrang eines Phonems Funktionen von weiteren Systemeigenschaften sind, können Veränderungen wie die beschriebene die Kurve so verformen¹⁸, daß aus dem stabilem Gleichgewicht ein instabiles wird. Der Zufall entscheidet nun darüber, in welche Richtung der Lautwandel geht; die kleinste phonetische Fluktuation reicht jetzt aus, um das Artikulations-Dekodierungs-Gleichgewicht zu zerstören.

17 vgl. Strauß 1980; Moulton 1961

18 Wollte man dies mathematisch ausführen, so müßte die Funktion noch entsprechend verallgemeinert werden.

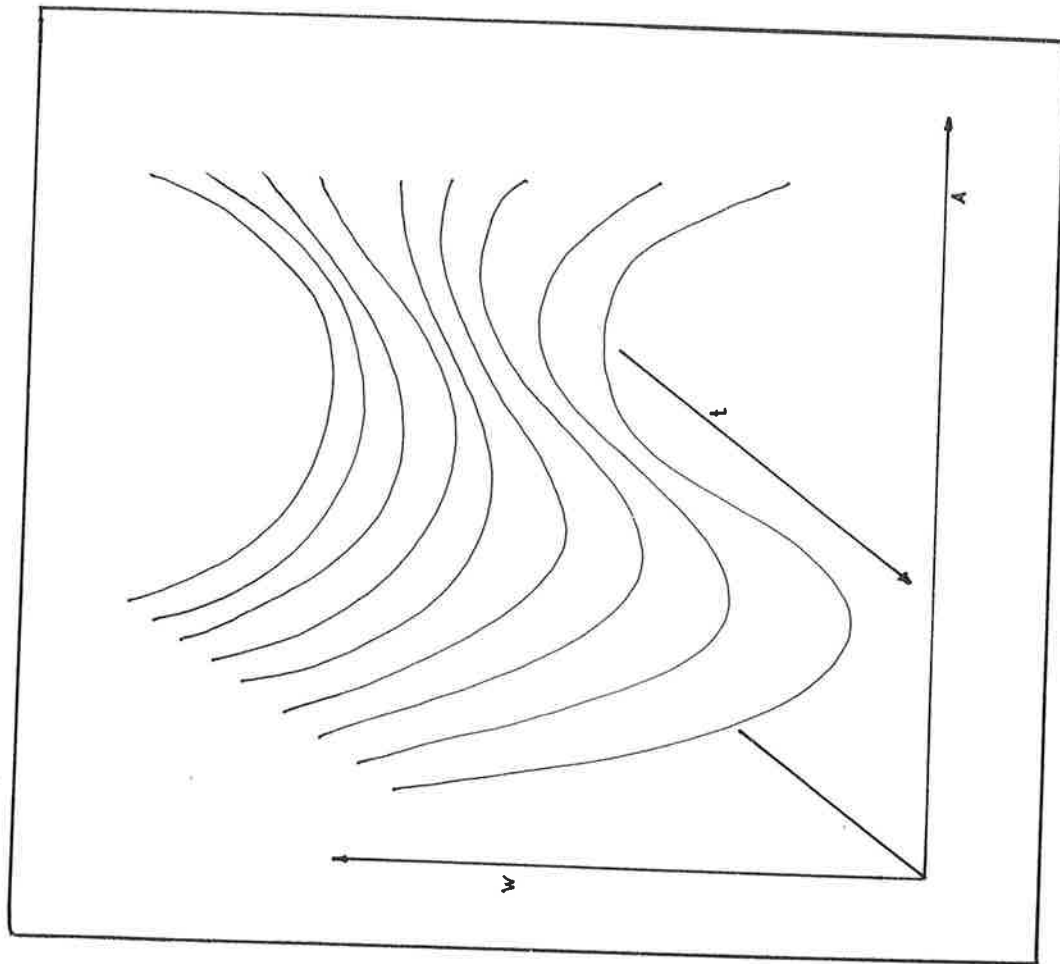


Abb. 6a: Übergang vom stabilen zum instabilen Gleichgewicht

Die lexikalische Einheit wird als undefinierter, primitiver Begriff eingeführt:

Unter einer lexikalischen Einheit verstehen wir ein in der Rede beobachtbares, bedeutungstragendes, ein- oder mehrteiliges Element des Lexikons einer natürlichen Sprache.

Diesen lexikalischen Einheiten kann - wie allen sprachlichen Einheiten - eine Vielzahl linguistischer Eigenschaften zugeschrieben werden. Damit ist jede Eigenschaft gemeint, die im Rahmen einer theoretischen linguistischen Betrachtung der jeweiligen Einheit aufgrund der Struktur des Modells für diese Einheit zukommt (vgl. hierzu und zur Klassifikation von linguistischen Eigenschaften Altmann, Lehfeldt 1980 : 37-39).

Nicht jede linguistische Eigenschaft interessiert bei der Konstruktion dieses Basismodells; diejenigen Eigenschaften, die als Elemente in der Struktur des Modells betrachtet werden - im gegenwärtigen Modell also die vier oben genannten -, sollen als strukturelle Eigenschaften bezeichnet werden. Die vorläufige Beschränkung auf diese Größen schließt keineswegs einen späteren Einschluß weiterer Eigenschaften aus; er wird sogar notwendig, wenn theoretische Überlegungen oder die empirische Überprüfung des Modells entsprechende Defizite aufzeigen. Sobald solche neu hinzugenommenen Eigenschaften der lexikalischen Einheiten ihren Platz im Modell zugewiesen bekommen haben, wenn also ihre Relationen zu den anderen Größen formuliert worden sind, erhalten sie ebenfalls den Status struktureller Eigenschaften.

Def. 1

Eine strukturelle Eigenschaft einer lexikalischen Einheit ist jede Eigenschaft, die sich bei einer linguistischen Analyse der Verwendung dieser Einheit in der Rede als manifestes oder latentes Merkmal beobachten läßt, und die Element des der Beobachtung zugrundeliegenden Modells ist.

Vergleicht man die Vorgänge bei der Entstehung und Selektion von Mutationen auf der phonetischen Ebene mit den Veränderungsprozessen auf der phonologischen, so stellt man fest, daß erstere sich um Größenordnungen schneller abspielen als letztere. Die Veränderung von Phonemsystemen benötigt Generationen, während die Prozesse spontan phonetischer Variation bei jeder Äußerung stattfinden, die ein Sprecher von sich gibt. Daher kann man sagen, daß die Bedingungen dafür, ob eine Mutation in die Richtung des Gleichgewichts führt oder nicht, ob sie sich also durchsetzen kann, durch den Rahmen der vorgegebenen phonologischen Situation zu jeder Zeit feststehen. Diesen Zusammenhang nennt man in der Synergetik die "Versklavung" des einen Prozesses durch den anderen. Das Phonemsystem ist ein "Ordnungsparameter" für den phonetischen Prozeß.

4.4 ZUR BEGRIFFSBESTIMMUNG

Zum besseren Verständnis der Struktur des Modells, seiner Elemente und ihrer Relationen zueinander ist es notwendig, einige Begriffe zu definieren, die dabei verwendet werden; dadurch wird auch der zu modellierende Gegenstand genauer abgegrenzt.

Die vier Größen Länge, Bedeutungskomplexität (Polylexie), Kontextgebundenheit (Polytextie) und Gebrauchshäufigkeit, die die Kernelemente des Basismodells bilden sollen, sind Eigenschaften von sprachlichen Einheiten wie Morphemen, Wörtern oder Syntagen. Man darf davon ausgehen, daß die in dem Basismodell postulierten Relationen und Prozesse für alle Typen von Ausdrücken gültig sind, die als lexikalisiert angesehen werden können. Solche Ausdrücke werden im folgenden als lexikalische Einheiten bezeichnet. Ohne Verlust der Allgemeinheit wird in der weiteren Darstellung auch einfach von Wörtern gesprochen.

Anhand der ihnen zugeschriebenen Eigenschaften lassen sich die lexikalischen Einheiten in lexikalische Klassen einteilen, z.B. in Frequenzklassen, Altersklassen, Wortarten und Flexionsklassen.

Def. 2

Eine Menge von lexikalischen Einheiten wird lexikalische Klasse genannt, wenn diese in einer oder mehreren Eigenschaften übereinstimmen.

Handelt es sich bei der gemeinsamen Eigenschaft um eine strukturelle Eigenschaft, soll von einer strukturellen Klasse gesprochen werden. Beispiel: Frequenz- oder Längenkategorie.

Def. 3

Strukturelle Klasse heißt eine lexikalische Klasse, deren Elemente eine oder mehrere strukturelle Eigenschaften teilen.

Wir fordern, daß jede lexikalische Einheit eine eindeutige Ausprägung jeder strukturellen Eigenschaft besitzt. Daher kann sie eindeutig als Punkt im lexikalischen Raum abgebildet werden, dessen Koordinaten durch den n -Tupel $\langle e_1, e_2, \dots, e_n \rangle$ der Ausprägungen ihrer strukturellen Eigenschaften angegeben wird.

Def. 4

Der Raum, der durch die Dimensionen der strukturellen Eigenschaften der lexikalischen Einheiten definiert ist, wird lexikalischer Raum genannt.

Die bisher angestellten Betrachtungen beschränken sich auf den Bereich der Synchronie; das zu konstruierende Modell soll aber den dynamischen Aspekt, die Diachronie, einschließen. Die Abbildung von Veränderungen der lexikalischen Einheiten ist möglich, wenn das Achsensystem der strukturellen Eigenschaften zusätzlich eine Zeitachse erhält. Eine synchrone Darstellung entspricht dann einer Momentaufnahme des lexikalischen Raums zu einem gegebenen

Zeitpunkt t , eine diachrone der Abbildung des Verlaufs einer strukturellen Eigenschaft über die Zeit. Entsprechend werden die dynamischen Eigenschaften definiert:

Def. 5

Die dynamischen Eigenschaften einer lexikalischen Einheit sind Dauer, Geschwindigkeit und Beschleunigung der Bewegung dieser Einheit im lexikalischen Raum.

Die lexikalischen Eigenschaften sind auf vielfältige Weise untereinander verknüpft, z.B. über paradigmatische Bedeutungsbeziehungen, über syntagmatische Kookurrenzwahrscheinlichkeiten, über Derivation oder Komposition u.v.m.

Es handelt sich bei der Menge der lexikalischen Eigenschaften also um einen Gegenstand, der eine interne Struktur besitzt. Nach außen hin ist er relativ abgeschlossen und autonom. Die Struktur, die wir in ihr sehen, ist eine Abbildung unseres theoretischen Wissens über die Beziehungen der Elemente zueinander. Kurz: Wir betrachten diese Menge als System.

Def. 6

Die Menge der lexikalischen Eigenschaften mit ihren Relationen untereinander bildet das lexikalische System. Alle Elemente sind so miteinander verknüpft, daß keine isolierte Teilmenge entsteht.

Im Lichte einer bestimmten Theorie sind natürlich nicht alle denkbaren Relationen im System relevant, ebenso, wie nicht in jedem Zusammenhang alle zuschreibbaren Eigenschaften einer lexikalischen Einheit von Interesse sind. Als Strukturelemente des Modellsystems kommen also nur diejenigen Relationen in Betracht, die mit den strukturellen Eigenschaften der lexikalischen Einheiten interdependent sind. Das Gesamtsystem ist somit zu jedem Zeitpunkt durch Angabe aller Ausprägungen aller Eigenschaften aller Elemente beschreibbar.

Def. 7

Der Vektor der strukturellen Eigenschaften der lexikalischen Einheiten des Systems zu einem Zeitpunkt wird Systemzustand zu t genannt.

Ein so modelliertes lexikalisches System sieht man zwar als relativ autonom an, aber nicht als hermetisch abgeschlossen. Dies bedeutet, daß es - wie alle Systeme - über eine oder mehrere Schnittstellen mit seiner Umwelt in Kontakt steht. Die Teile der Umwelt, mit der das System in Wechselwirkung steht, ist die Systemumgebung.

Def. 8

Die Menge von Elementen, die nicht als Teil des lexikalischen Systems betrachtet werden, aber Änderungen im Systemzustand hervorrufen oder an denen Änderungen durch das System hervorgerufen werden können, heißt Systemumgebung.

Die Elemente der Systemumgebung lassen sich nach zwei Kriterien in Gruppen einteilen: nach den Änderungen, die sie im System hervorrufen (bzw. nach den Systemelementen, auf die sie wirken), und nach ihrer Rolle in der Systemumgebung.

Die Elemente der unter dem zweiten Aspekt betrachtete Gruppe sollen Systembedürfnisse heißen.

Ein Systembedürfnis ist also ein Element (welches selbst ein System sein kann) der Systemumgebung, das unter dem Gesichtspunkt seiner Beziehungen zu anderen Elementen außerhalb und wenigstens einem Element innerhalb des betrachteten Systems betrachtet wird. Ein Beispiel ist die Minimierung des Wahrnehmungsaufwands; dieses theoretische Konstrukt steht in Beziehung zu Elementen von Bereichen wie physikalische Akustik, Physiologie der Perzeptionsorgane, Signalverarbeitung im Nervensystem und Psychologie der Symbolrepräsentation und -verarbeitung einerseits und Zeichenähnlichkeit andererseits.

Def. 9

Unter einem Systembedürfnis verstehen wir ein Element der Systemumgebung, welches in Relation zu anderen Elementen außerhalb des Systems steht und Änderungen im System hervorrufen kann.

Der (vor allem in der Soziologie) verbreitete Terminus "Systembedürfnis" ist recht ungeschickt gewählt und führt immer wieder zu Mißverständnissen. Es sei deshalb noch einmal betont, daß ein solches Bedürfnis nicht zum System gehört, sondern eine von außen an das System gestellte Anforderung ist.

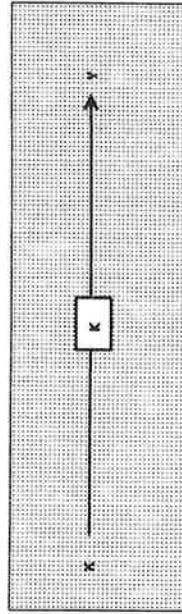
4.5 SYSTEMTHEORETISCHE HILFSMITTEL: GRAPHENALGEBRA UND LINEARE OPERATOREN

Die Darstellungsform, in der die oben betrachteten Zusammenhänge erörtert worden sind, ist noch nicht ausreichend präzise, um den an das Modell gestellten Anforderungen zu genügen. Insbesondere läßt sie Raum für Vagheiten hinsichtlich des Grades einer Abhängigkeit (bisher ist lediglich die Richtung der Abhängigkeit markiert) und hinsichtlich des Zusammenwirkens von Abhängigkeiten (was passiert, wenn zwei oder mehrere Pfeile auf dieselbe Größe weisen?). Hier kann die allgemeine Systemtheorie als Hilfsmittel dienen; sie basiert auf einer Menge von Regeln, mit denen Aussagen über Funktion aus Aussagen über Struktur deduziert werden können. Diese Regeln bilden eine spezielle Algebra: die der linearen Graphen und der linearen Operatoren. Soweit möglich, wird ihre Darstellung den Konventionen folgen, wie sie in Cortes, Przeworski, Sprague (1974) verwendet werden, ergänzt

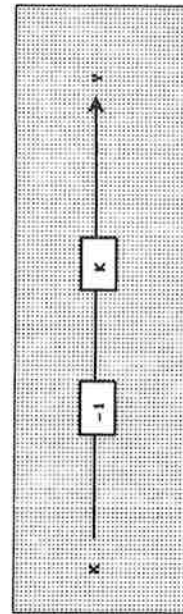
durch das oben eingeführte Konzept der Bedürfnisse¹⁹ in unserer Notation.

4.5.1 Operatoren und Graphen

Die Aussage, daß eine Größe X sich auf eine andere Größe Y auswirkt, nimmt in Graphendarstellung folgende Gestalt an:

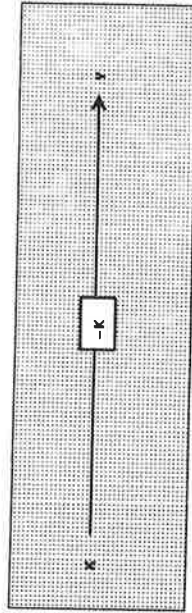


Das Kästchen symbolisiert den Proportionalitätsoperator, der angibt, in welchem Maße X auf Y wirkt. Handelt es sich um die Abbildung einer inversen Proportion, so kann dies entweder unter Zuhilfenahme des Umkehr-Operators

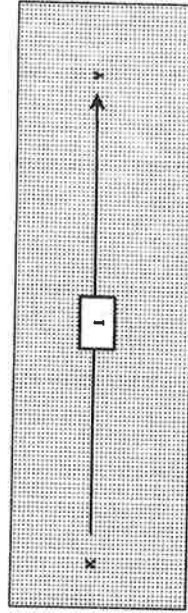


¹⁹ vgl. zu diesem Begriff auch Abschn. 4.2.

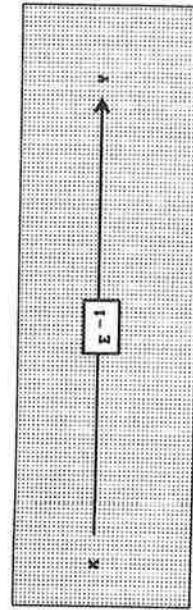
oder kürzer durch Verwendung eines negativen Vorzeichens für den Proportionalitätsfaktor



geschehen. Zur Straffung der Darstellung wird im folgenden, wo möglich, auf die explizite Verwendung des Identitätsoperators



verzichtet. Schließlich wird noch als einziger weiterer Operator der Verzögerungsoperator benötigt,



der besagt, daß X nach einer Verzögerung von einer Zeiteinheit auf Y wirkt.

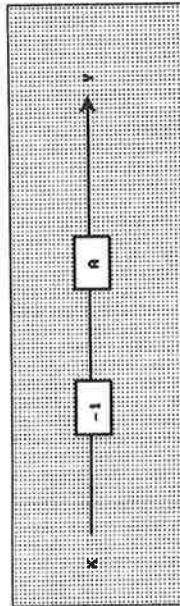
4.5.2 Struktur und Funktion

Mit Hilfe des Konzepts der Operatoren kann man nun folgende Definitionen formulieren:

1. Die Struktur eines Systems ist die Menge seiner untereinander verknüpften Operatoren;
2. Die Funktion eines Systems ist die Gesamtheit der Wirkungen aller Operatoren des Systems.

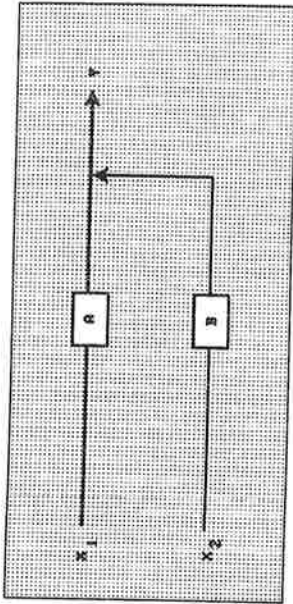
Zur Ermittlung der Funktion eines (Sub-)Systems aus seiner Struktur legen die Regeln der Graphenalgebra fest:

1. In Vorwärtsrichtung werden die Operatoren miteinander multipliziert. Aus der Struktur

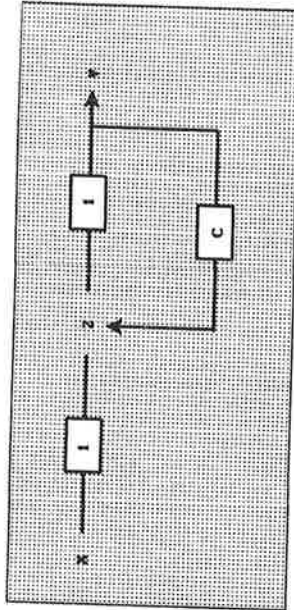


folgt die Funktion $y = -Ax$.

2. Aufeinander treffende Kanten sind additiv verknüpft. Die Struktur



hat die Funktion $y = Ax_1 + Bx_2$. Diese beiden Regeln genügen bereits, um auch Rückkopplungskanten zu berechnen. Der Struktur



entsprechen die Gleichungen

$$\begin{aligned} y &= 1z \\ z &= 1x + cy \end{aligned}$$

Daher ist

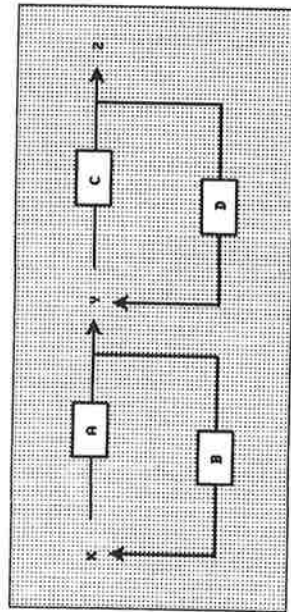
$$\begin{aligned} y &= x + cy \\ y - cy &= x \end{aligned}$$

$$y = \frac{x}{1 - c}$$

4.5.3 Vereinfachung von Graphen

Die Berechnung der Funktion komplexer Strukturen läßt sich vereinfachen, indem man zunächst die Funktionen von Teilstrukturen ermittelt und die Resultate auf die gleiche Weise miteinander verknüpft wie einfache Operatoren.

Gegeben sei die Struktur



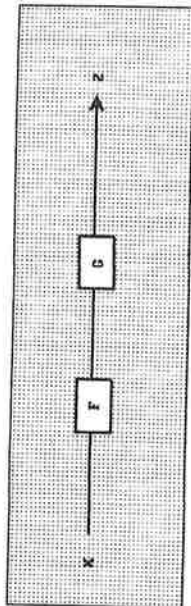
Die Funktion der Teilstruktur, die der linken Rückkopplungsschleife entspricht, ergibt sich nach den bereits bekannten Regeln der Graphenalgebra als

$$y = \frac{a}{1 - ab} x;$$

der rechten Teilstruktur kommt die Funktion

$$z = \frac{c}{1 - cd} y$$

zu. Bezeichnet man diese beiden Funktionen mit F und G, so läßt sich die Gesamtstruktur auch vereinfacht darstellen:



Die Funktion dieser Struktur ist nun einfach

$$z = f g x.$$

Setzt man die oben gefundenen Funktionen für f und g ein, so erhält man

$$z = \frac{ac}{(1 - ab)(1 - cd)} x.$$

Auf analoge Weise lassen sich auch komplizierte Systeme durch schrittweise Reduktion leicht berechnen, selbst wenn eine größere Zahl von Eingangs- und Ausgangsgrößen gleichzeitig berücksichtigt werden muß.

Diese Darstellung ist natürlich sehr stark verkürzt; sie ist jedoch für die hier verfolgten Zwecke vorläufig vollkommen ausreichend. Eine ausführliche Einführung in die Graphenalgebra und weitere Literatur findet man z.B. in (a.a.O.).

In den nachfolgenden Abbildungen werden Systemgrößen als Rechtecke, Operatoren als Quadrate und Systembedürfnisses als Kreise repräsentiert. Systembedürfnisse werden entweder als über einen Operator mit einer Systemgröße verknüpft dargestellt oder als Interpretation eines Operators. Im zweiten Fall wird das Bedürfnis (ohne Pfeil) mit dem betreffenden Operator verbunden.

4.6 ELEMENTE, STRUKTUR UND FUNKTION DES LEXIKALISCHEN SYSTEMS

4.6.1 Das Phoneminventar

Zur Entwicklung des Basismodells wird hier noch einmal bei der Schnittstelle zwischen dem Phonem-/Graphemsystem und dem lexikalischen System angesetzt. Dies geschieht in einer Weise, daß diese Schnittstelle mit dem Einfluß der Phonemzahl²⁰ auf die Wortlänge zusammenfällt. Die beteiligten Bedürfnisse werden hier ebenenspezifisch behandelt, wenn dies die Darstellung vereinfacht; auf der Phonemebene können alle Bedürfnisse in zwei einander entgegengesetzte zusammengefaßt werden:

- die Minimierung des Kodierungsaufwands (minK) und
- die Minimierung des Dekodierungsaufwands (mind).

In Abwandlung und Präzisierung des ersten Ansatzes oben wird nicht nur die Richtung, sondern auch der Grad des Einflusses dieser Bedürfnisse auf die Phonemzahl angegeben. Um Verwechslungen zwischen den Zeicheninventaren zu vermeiden, wird hier nicht mehr von P-INVENTARUMFANG gesprochen; diese Größe soll von jetzt an PHONEMZAHL heißen. Außerdem wird eine Transformation durchgeführt: Um das Modell linear darstellen zu können, wird nicht die PHONEMZAHL selbst verwendet, sondern ihr Logarithmus, der im folgenden als L-PHONEMZAHL bezeichnet werden soll.

Dem Bedürfnis minK entspricht auf dieser Ebene die Zipfsche Unifikationskraft, die sich reduzierend auf die L-PHONEMZAHL

²⁰ Hier und im folgenden ist die Möglichkeit, anstelle der Phoneme die Grapheme zu betrachten, immer mitgedacht, wenn von der Phonemzahl die Rede ist. Der Unterschied, der sich zwischen den Maßen bei gesprochenen bzw. geschriebener Sprache ergibt, läßt sich durch einen konstanten Faktor ausdrücken.

auswirkt. Diese Kraft wird daher als ein inverser Proportionalitätsoperator $-Y_2$ dargestellt (vgl. Abbildung 7). Demgegenüber entspricht dem Bedürfnis mind die Zipfsche Diversifikationskraft, die als Operator Y_1 wiedergegeben wird.

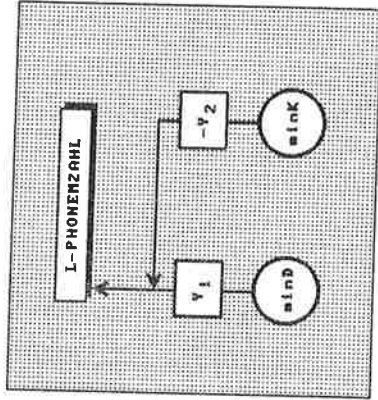


Abb. 7: Die L-PHONEMZAHL resultiert aus einem Kräftegleichgewicht zwischen Unifikation und Diversifikation.

Die Kanten treffen direkt aufeinander, so daß der auf die Größe L-PHONEMZAHL wirkende resultierende Einfluß durch die Addition der beiden entgegengesetzten Teilkräfte²¹ zu berechnen ist. Nach den Regeln der Graphenalgebra darf man also schreiben:

$$L\text{-PHONEMZAHL} = Y_1 \text{ mind} - Y_2 \text{ minK},$$

wobei es bei einer gegebenen Phonemzahl unendlich viele Werte für die beiden Teilausdrücke auf der rechten Seite der Gleichung gibt, die diese Bedingung erfüllen. Ohne weitere Information ist also hier keine numerische Bestimmung der Operatoren möglich.

²¹ Die gleiche lineare Verknüpfung setzt Altmann (1985) für entsprechende Teilkräfte an, um die semantische Diversifikation mit Hilfe eines stochastischen Prozesses zu modellieren.

4.6.2 Die Lexikongröße

Einen weiteren Parameter, der für die anderen Elemente des lexikalischen Systems bestimmend ist, stellt die LEXIKONGRÖßE dar. Sie ist direkt mit einem elementaren Systembedürfnis verbunden,

- dem Kodierungsbedürfnis (Kod).

Es repräsentiert die Notwendigkeit, mehrere verschiedene Bedeutungen ausdrücken zu können. Je mehr Bedeutungen kodiert werden müssen, desto größer muß die Anzahl der lexikalischen Einheiten einer Sprache sein.

Zwecks Linearisierung wird auch diese Größe transformiert; Im Modell wird ihr Logarithmus L-LEXIKONGRÖßE (s. Abbildung 8) verwendet.

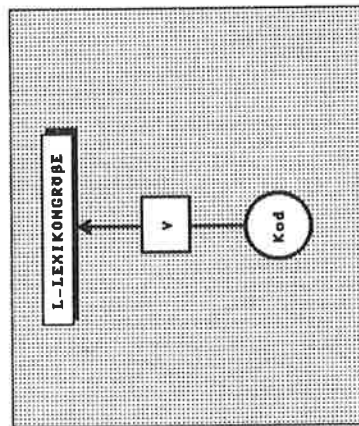


Abb. 8: Das Kodierungsbedürfnis wirkt über einen Proportionalitätsoperator V auf die L-LEXIKONGRÖßE.

Für die Wirkung des Kodierungsbedürfnisses auf die Lexikongröße wird nicht der Identitätsoperator, sondern ein Proportionalitätsoperator angesetzt, da das Verhältnis von zu kodierenden Bedeutungen und der Anzahl der lexikalischen Einheiten - außer in Konstruktssprachen - niemals 1 : 1 beträgt (s dazu 4.6.4).

4.6.3 Die Länge

Als erstes Systemelement, das einer strukturellen Eigenschaft der lexikalischen Einheiten entspricht, wird die fundamentale Größe LÄNGE eingeführt. Die Länge eines sprachlichen Ausdrucks kann auf vielerlei Weise gemessen werden - je nachdem, welche Bezugseinheit gewählt wird und welche Aspekte eines solchen Ausdrucks im Rahmen des der Untersuchung zugrundeliegenden Modells als relevant angesehen werden. Gewöhnlich versteht man in der Linguistik unter der Länge die Anzahl der konstituierenden Elemente: je nach Modell Phoneme, Grapheme, Silben²², Morpheme oder Wörter. In vielen Fällen wird die Länge aber auch als die zeitliche Dauer des akustischen Signals aufgefaßt²³.

Für das Basismodell wird die LÄNGE einer lexikalischen Einheit als die Anzahl der sie konstituierenden Phoneme festgelegt. Um aber die Relationen der Variablen LÄNGE zu anderen Elementen des Systems linear darstellen zu können, wird hier transformiert:

Def.10

Die L-LÄNGE einer lexikalischen Einheit ist der Logarithmus der Anzahl der sie konstituierenden Phoneme

22 Die Silbe ist ein Beobachtungsbegriff, der sich auf die lautliche Repräsentation von Ausdrücken bezieht. Wenn Eigenschaften von Silben an schriftlich vorliegenden Texten untersucht werden (wie es gang und gäbe ist), sollte das Abbildungsverhältnis zwischen Silben und Graphenkombinationen explizit gemacht werden.

23 vgl. z. B. Fonagy, Magdics (1960)

Sie ist im Intervall $\langle 0, \infty \rangle$ definiert, da die kleinste Länge eines sprachlichen Ausdrucks 1 Phonem beträgt ($\log 1 = 0$), während eine theoretische Obergrenze nicht angegeben werden kann.

Im Gegensatz zur L-PHONEMZAHL und zur L-LEXIKONGRÖßE ist die L-LÄNGE kein Attribut des gesamten lexikalischen Systems, sondern eine Eigenschaft, die jeder einzelnen lexikalischen Einheit zukommt. Es muß daher genauer bestimmt werden, was gemeint ist, wenn von der L-LÄNGE als Funktion der beiden anderen Größen die Rede ist.

Wie bereits festgestellt wurde, ist die Länge der Elemente der Lexik von der Anzahl der Elemente auf der Phonemebene abhängig. Nun trifft diese Aussage keinesfalls auf ein bestimmtes einzelnes lexikalisches Element zu: Ein individuelles Wort wird nicht, dadurch kürzer, daß die Sprache ein zusätzliches Phonem in ihr Inventar aufnimmt. Es handelt sich vielmehr um einen stochastischen Zusammenhang, der zwischen der Phonemzahl und der von ihr durch die Gesetze der Kombinatorik abhängigen durchschnittlichen Ausdruckslänge besteht.

Von gleicher Art ist die Beziehung zwischen der Lexikongröße und der Ausdruckslänge: Je mehr lexikalische Einheiten erforderlich sind, desto länger müssen sie (bei konstanter Phonemzahl) durchschnittlich sein²⁴.

Die Systemgrößen L-PHONEMZAHL und L-LEXIKONGRÖßE repräsentieren somit einen globalen stochastischen Beitrag zur L-LÄNGE einer lexikalischen Einheit, über deren individuelle Ausprägung - ihre Abweichung von Durchschnitt - bisher nichts ausgesagt wurde.

Wie wir wissen, wird von keiner natürlichen Sprache der kombinatorisch mögliche Vorrat an Formen zur Bildung von lexikalischen Einheiten ausgeschöpft. In der Regel ist der Ausnutzungsgrad so-

24 Zusätzlich zu dieser Beziehung sind gegebenenfalls die Töne bzw. Akzente zu berücksichtigen.

gar recht gering. Als Beispiel kann die nachfolgende Aufstellung der formbaren Phonemfolgen aus den drei Phonemen /t/, /n/ und /u:/ dienen:

/ttt/	/tu:t/	/tnt/
/ttu:/	/tu:u:/	/tnu:/
/ttn/	/tu:n/	/tnn/
/u:tt/	/u:u:t/	/u:nt/
/u:tu:/	/u:u:u:/	/u:nu:/
/u:tn/	/u:u:n/	/u:nn/
/ntt/	/nu:t/	/nnt/
/ntu:/	/nu:u:/	/nnu:/
/ntn/	/nu:n/	/nnn/

Nur die vier unterstrichenen Phonemfolgen werden im heutigen Hochdeutsch zur Bildung von Morphemen herangezogen.

Es lassen sich sogar sprachtypische phonotaktische Regeln formulieren, mit denen die Auswahlprinzipien einer Sprache beschrieben werden. Der Grund für diese Beschränkung ist offensichtlich und seit langem bekannt: Zur Erhöhung der Übertragungssicherheit bzw. zur Verringerung der Verwechslungsgefahr besitzt jedes natürliche Kommunikationssystem eine den Anforderungen entsprechende Redundanz. Daher soll hier das Systembedürfnis zur

- Sicherung der Übertragung (Red)

postuliert werden, das direkt proportional auf die L-LÄNGE wirkt (Abbildung 9).

Als Funktion des Subsystems, das aus den bisher eingeführten Elemente besteht, läßt sich nach den Regeln der Graphenalgebra die L-LÄNGE als

$$\log(l) = A \log(g) + Z \text{ Red} - p \log(p)$$

angeben, wo mit $\log(l)$ die L-LÄNGE, mit $\log(g)$ die L-LEXIKONGRÖßE und mit $\log(p)$ die L-PHONEMZAHL bezeichnet wurden. Bei Verwendung

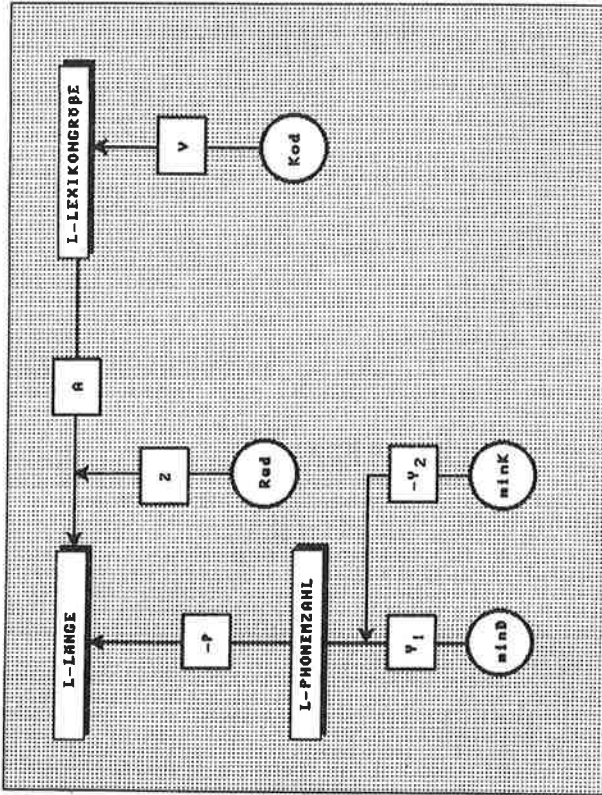


Abb. 9: Die durchschnittliche L-LÄNGE bestimmt sich aus einander entgegenwirkenden Systemgrößen: L-LEXIKON-GRÖßE und Redundanz einerseits und L-PHONENZAHL andererseits.

der nichtlogarithmischen Größen LÄNGE (l), LEXIKONGRÖßE (g) und PHONENZAHL (p) erhält man die Gleichung

$$l = \frac{A \cdot Z}{g \cdot \text{Red}} \cdot \frac{p}{p}$$

oder

$$l = \frac{A \cdot Z}{g \cdot \text{Red}} \cdot \frac{-p}{p}$$

Es sei zur Erinnerung noch einmal betont, daß die Länge einer lexikalischen Einheit damit noch nicht vollständig determiniert ist - bisher sind nur die globalen Einflüsse der beiden besprochenen Systemparameter berücksichtigt.

4.6.4 Die Polylexie

Ein gemeinsames Merkmal aller natürlichen Sprachen ist ihre Mehrdeutigkeit, die auf allen Analyseebenen zu finden ist: Es gibt phonologische, morphologische, syntaktische, semantische u.a. Mehrdeutigkeiten. So kommt auch den lexikalischen Einheiten meist nicht eine einzige, fest umrissene Bedeutung zu.

Ob man nun auf dem Standpunkt steht, daß die Bedeutung eines sprachlichen Zeichens nur bei seiner Verwendung in dem jeweiligen Kontext existent ist, oder ob man die Bedeutungen als Entitäten auffaßt, die auch losgelöst vom Gebrauch vorhanden sind - in jedem Fall ist eine Bedeutung eine Abstraktion von Inhalten einer großen Zahl situativer und kontextueller Verwendungsumgebungen.

Als Resultat erhält man entweder selbständige Konstrukte oder aber mehr oder weniger große Mengen von Situationen und Kontexten, in denen eine Bedeutung des betrachteten Ausdrucks realisierbar ist. Die Anzahl der verschiedenen Bedeutungen, die eine lexikalische Einheit zu einem gegebenen Zeitpunkt trägt (in Altmann, Beöthy, Best 1982 "Bedeutungskomplexität"), soll die POLYLEXIE dieser Einheit genannt werden. Dieser Begriff soll nicht zwischen semantischen und grammatischen Bedeutungen differenzieren, so daß auch den Funktionswörtern eine POLYLEXIE größer Null zukommt.

Der durchschnittliche POLYLEXIE-Wert einer lexikalischen Einheit liegt also im Intervall $\langle 1, \infty \rangle$, der seines Logarithmus - der L-POLYLEXIE - im Intervall $\langle 0, \infty \rangle$. Es ist offensichtlich, daß die durchschnittliche POLYLEXIE einen Einfluß auf die LEXIKONGRÖßE hat: Je mehr Bedeutungen ein Ausdruck im Mittel trägt, desto weniger Ausdrücke muß die Lexik (bei konstanter Zahl der zu kodierenden Bedeutungen) umfassen.

An dieser Stelle greift wieder das Bedürfnis nach Minimierung des Kodieraufwands; die Unifikation wirkt in Richtung auf verstärkte POLYLEXIE. Im Extremfall wäre eine Lexik denkbar, die aus einem einzigen Wort bestünde, das jeweils alle Bedeutungen trägt, die der jeweilige Expedient ausdrücken möchte. Aber auch hier besteht das entgegengesetzte Bedürfnis nach Minimierung des Dekodieraufwands: Ein Text ist um so leichter zu dekodieren, je eindeutiger die lexikalischen Einheiten sind. Die Diversifikation wirkt in Richtung auf Verringerung der POLYLEXIE. Dem entspricht ein hypothetischer Extremfall einer Lexik, die für jede Bedeutung einen separaten lexikalisierten Ausdruck besitzt.

Die empirisch vorfindbaren Ausprägungen der POLYLEXIE in den verschiedenen Sprachen stellen jeweils das Resultat eines Gleichgewichts zwischen den beteiligten Einflüssen dar.

Wieder wird angenommen, daß die (logarithmisch transformierten) Kräfte additiv verknüpft sind; für die Wirkung der L-POLYLEXIE auf die L-LEXIKONGRÖßE wird der Proportionalitätsoperator -L eingesetzt (vgl. Abbildung 10).

Wenn man jetzt den Einfluß des Kodierungsbedürfnisses als die (ebenfalls logarithmierte) Zahl der zu kodierenden Bedeutungen definiert, kann man die Proportionalitätsfaktoren L und V numerisch bestimmen: Wenn keine weiteren Einflüsse im Regelkreis zu berücksichtigen sind, haben beide den Wert 1.

Beispiel: Angenommen, eine Sprache unterscheidet zu einem gegebenen Zeitpunkt 100000 Bedeutungen. Wenn jede lexikalische

Einheit durchschnittlich 4 Bedeutungen trägt, ist ein Inventar von 25000 Wörtern erforderlich, da

$$\text{LEXIKONGRÖßE} = \text{Bedeutungszahl} / \text{POLYLEXIE}.$$

Mit logarithmierten Variablen ergibt sich, wenn man für L und V jeweils 1 einsetzt,

$$\text{L-LEXIKONGRÖßE} = \text{Kod} - \text{L-POLYLEXIE}.$$

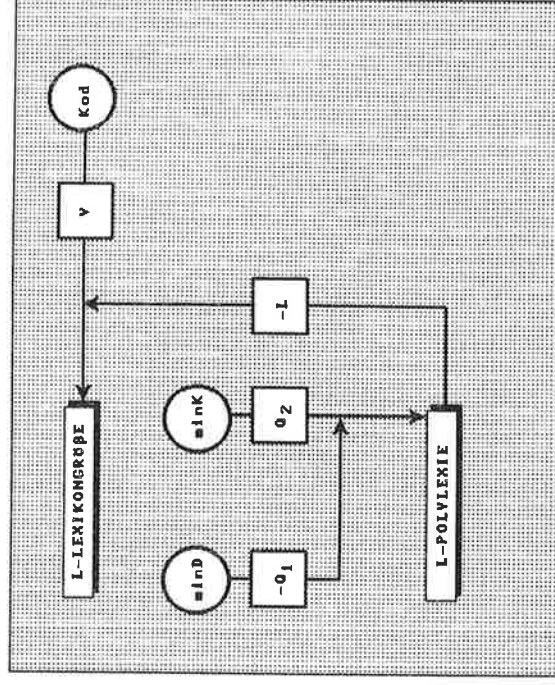


Abb. 10: Die durchschnittliche POLYLEXIE und das Kodierungsbedürfnis bestimmen gemeinsam die LEXIKONGRÖßE.

Ein Zusammenhang zwischen der Bedeutungszahl und anderen Größen wurde auch schon früher bemerkt: Zipf (1949) vermutete eine Abhängigkeit dieser Zahl von der Worthäufigkeit, und Guiter

(1974) veröffentlichte empirische Funktionen zwischen Frequenz und Bedeutungszahl bzw. Länge und Bedeutungszahl für mehrere romanische Sprachen und das Englische. Diese Autoren haben jedoch ihre Befunde nicht linguistisch interpretiert oder gar theoretische Ableitungen vorgelegt. Erst seit Altmann²⁵ besteht eine plausible Deduktionsmöglichkeit.

Nehmen wir für einen gegebenen Ausdruck einer Sprache einen beliebigen POLYLEXIE-Wert größer 1 an und eine Kommunikationssituation, in der die durch diese POLYLEXIE entstehende Mehrdeutigkeit unerwünscht ist: wir postulieren das

- Spezifikationsbedürfnis (Spz).

Die natürlichen Sprachen bieten zur Bedienung dieses Systembedürfnisses verschiedene Hilfsmittel (funktionale Äquivalente) an: Spezifikation kann - ebenso wie Modifikation - durch syntaktische Mittel (Beispiel: "syntaktisches Mittel" vs. "statistisches Mittel") oder durch morphologische wie Komposition und Affigierung (Beispiel: "Hilfsmittel" vs. "Mittelwert") erreicht werden. In jedem Fall ist die Länge des resultierenden Ausdrucks betroffen²⁶; bei der praktischen Messung wirkt sich jedoch - wegen der relativen Seltenheit lexikalisierten Syntagmen gegenüber Wörtern - fast nur die morphologische Lösung auf die Länge aus. Das bedeutet: Die Abhängigkeit der POLYLEXIE von der LÄNGE ist um so stärker, je mehr eine Sprache von morphologischen gegenüber syntaktischen Mitteln zur Bedeutungsspezifikation Gebrauch macht. Diese typologische Eigenschaft einer Sprache soll SYNTHEZITÄT heißen und mit dem Buchstaben T bezeichnet werden. Über den Zusammenhang der Größen machen wir folgende Annahmen:

25 vgl. Altmann 1980 und Altmann, Beßth, Best 1982

26 Die Wirkung auf die Länge kann durch weitere Zusammenhänge modifiziert oder gar aufgehoben sein, z.B. wenn die Sprache von Tönen oder Akzenten Gebrauch macht.

a) Die Veränderungsrate der POLYLEXIE ist umgekehrt proportional zur LÄNGE.

b) Der Proportionalitätsfaktor ist ausschließlich durch die Synthesizität der Sprache bestimmt.

Bezeichnet man die POLYLEXIE mit y , ihre erste Ableitung mit y' und die LÄNGE mit x , so kann man unter Einbeziehung der Synthesizität T eine Differentialgleichung aufstellen, die diesen Annahmen entspricht:

$$\frac{y'}{y} = \frac{-T}{x}$$

Diese Differentialgleichung hat die Lösung

$$y = Cx^{-T}$$

wobei C eine beliebige Konstante ist. Für die Modellvariablen wird die Gleichung logarithmiert, woraus sich

$$\ln \text{POLYLEXIE} = \ln C - T \ln \text{LÄNGE},$$

mit $c = \ln C$ ergibt. Aus den Überlegungen im letzten Abschnitt ist die Größe, die hier für c einzusetzen ist, bereits bekannt; es handelt sich um den globalen Einfluß des Gleichgewichts zwischen Unifikation und Diversifikation. Die vollständige Gleichung lautet damit

$$\ln \text{POLYLEXIE} = Q_2 \cdot \min K - Q_1 \cdot \min D - T \cdot \ln \text{LÄNGE}.$$

Die Abbildung 11 zeigt diese Relation in Zusammenhang mit den bisher behandelten.

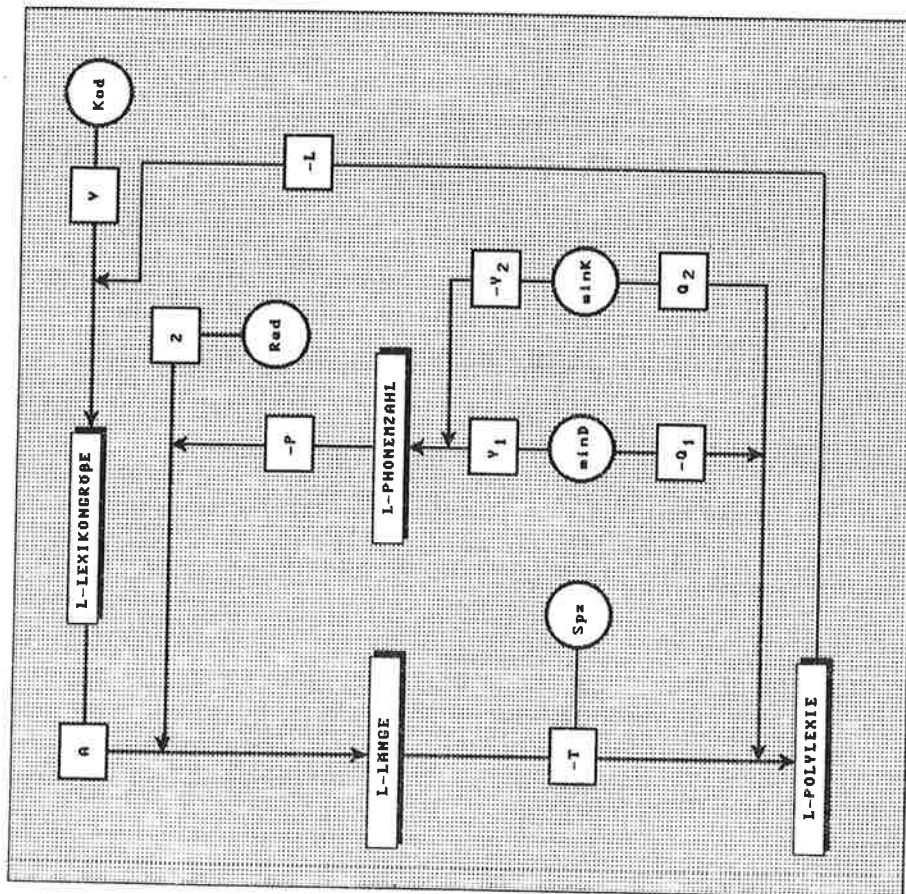


Abb. 11: In diesem Regelkreis ist die L-POLYLEXIE als Funktion der L-LÄNGE und des globalen Einflusses der Bedürfnisse mind und mink dargestellt.

4.6.5 Die Polytextie

Die Bedeutung einer lexikalischen Einheit abstrahiert von den individuellen Situationen des Gebrauchs in der Rede - jedenfalls soweit, daß sie weitgehend kontextunabhängig verwendet werden kann. Es wäre äußerst unökonomisch, wenn eine Sprache Bedeutungen wie "Haus,..." für ein bestimmtes Haus zu einer Zeit t an einem Ort o in einer Situation s lexikalisierte. Kontextspezifische Bedeutungskomponenten werden in den natürlichen Sprachen gewöhnlich entweder zusätzlich kodiert oder vom Rezipienten erschlossen. Auf diese Weise kann die Lexikongröße endlich gehalten werden. Es soll daher ein entsprechendes Bedürfnis postuliert werden:

- die Kontextökonomie (KÖ).

Die Eigenschaft der lexikalischen Einheiten, kontextunabhängig verwendbar zu sein, nennen wir POLYTEXTIE. Sie wird als metrischer Begriff definiert:

Def.11

Die POLYTEXTIE einer lexikalischen Einheit ist die Anzahl der Kontexte, in denen sie verwendet wird.

Die Kontextökonomie wirkt in Richtung auf eine hohe durchschnittliche POLYTEXTIE; den entsprechenden Prozeß nennen wir Globalisierung. Natürlich hat auch dieses Bedürfnis einen Widerpart. In vielen Fällen wird für eine optimale Kommunikation eine so weitgehende Eindeutigkeit der Ausdrücke erforderlich, daß

- a) eine lexikalische Einheit eine zusätzliche und unverwechselbare, kontextspezifische Bedeutung erhält (Bsp. "Gruppe", "Familie", "Ring", "Körper", ... in der Mathematik),

Nun ist die POLYTEXTIE nicht gleichmäßig über die lexikalischen Einheiten einer Sprache verteilt. So besitzt etwa der Ausdruck 'haben' eine sehr hohe, der Ausdruck 'hell' eine geringe und schließlich der Ausdruck 'Kontrapunkt' eine sehr geringe Ausprägung der POLYTEXTIE. Es wird also eine Relation zu integrieren sein, die für die Abweichung der POLYTEXTIE-Werte individueller lexikalischer Einheiten von der durchschnittlichen POLYTEXTIE verantwortlich ist.

Es scheint plausibel, anzunehmen, daß die Anzahl der Kontexte, in denen ein Ausdruck verwendet werden kann, vor allem von seiner Bedeutungsflexibilität abhängt. Ein Wort, das lediglich eine einzige, eng abgegrenzte Bedeutung trägt, wird nicht in so vielen Kommunikationssituationen Verwendung finden wie eines, das eine höhere Bedeutungspotenz besitzt. Man kann sich diesen Zusammenhang anhand eines hypothetischen Beispiels mit Hilfe nachstehender Tabelle verdeutlichen:

	Wort 1	Wort 2	Wort 3	Wort 4	Wort 5	...
Kontext 1	3	5	2	2	2	
Kontext 2	1	3	1	1		
Kontext 3	1	2		1		
Kontext 4	1	1				
Kontext 5		1				
Summe der Bed.	5	12	3	4	2	
Anz. d. Kontx.	4	5	2	3	1	

Diese Hypothese formulieren wir folgendermaßen:

Die POLYTEXTIE einer lexikalischen Einheit ändert sich proportional zu ihrer POLYLEXIE.

- b) ein neuer Terminus geschaffen wird, der nur in seinem Ursprungs-kontext eine Bedeutung hat (Bsp. "Laser", "Polylexie", ...),
- c) ein fremdsprachlicher Ausdruck oder ein ausgestorbenes bzw. selten gewordenes Wort der Gemeinsprache zu diesem Zweck importiert bzw. wiederbelebt wird (Bsp. "Tokamak", "Hort...).

Wir bezeichnen dieses Systembedürfnis als

- die Kontextspezifikation (KS).

Dieses Bedürfnis wirkt der KÜ entgegen auf die POLYTEXTIE ein, so daß sich als Resultat des Gleichgewichts zwischen der Globalisierung und diesem Zentralisierungsprozeß eine durchschnittliche, optimale POLYTEXTIE im System einstellt. Dies ist in Abbildung 12 als linearer Graph wiedergegeben, wobei auch hier von der Additivität der (logarithmierten) Wirkungen ausgegangen wird. Als Systemkomponenten verwenden wir wieder eine logarithmische Transformationsfunktion, die L-POLYTEXTIE.

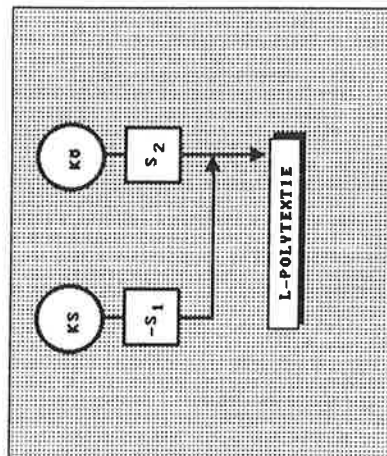


Abb. 12: Die durchschnittliche L-POLYTEXTIE als Funktion der Bedürfnisse KS und KÜ

Wenn wir die POLYTEXTIE mit y , die POLYLEXIE mit x und den Proportionalitätsfaktor mit G bezeichnen, können wir diese Hypothese in Form der Differentialgleichung

$$\begin{aligned} y' &= G \\ - &= - \\ y &= x \end{aligned}$$

mit der Lösung

$$y = Cx$$

aufstellen. Logarithmisch ergibt dies die Gleichung

$$L\text{-POLYTEXTIE} = c + G L\text{-POLYLEXIE},$$

mit $c = \ln C$. In dieser Gleichung ist die unbekannte Größe c unschwer als der Einfluß des KS-KÜ-Gleichgewichts zu interpretieren:

$$L\text{-POLYTEXTIE} = S_2 KÜ - S_1 KS + G L\text{-POLYLEXIE},$$

und dies entspricht exakt der Funktion, die sich nach den Regeln der Graphenalgebra aus der Struktur ergibt, wie sie in Abbildung 13 dargestellt ist.

4.6.6 Die Frequenz

Als Frequenz einer lexikalischen Einheit bezeichnet man die Häufigkeit ihrer Verwendung in der Rede. Sie ist umso größer, je öfter die jeweilige lexikalische Einheit von den Sprachbenutzern im Durchschnitt herangezogen wird, um eine beliebige Bedeutung auszudrücken. Dies gilt unabhängig davon, ob der Sprecher oder

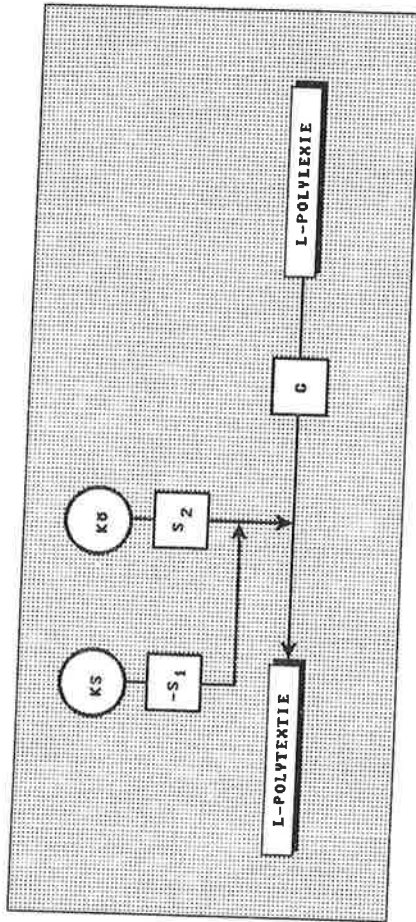


Abb. 13: Die L-POLYLEXIE beeinflusst die L-POLYTEXTIE im Maße G , so daß den lexikalischen Einheiten individuelle Ausprägungen der L-POLYTEXTIE zukommt.

Schreiber eine unter dem benutzten Ausdruck bereits lexikalisierte Bedeutung übermittelt, der lexikalischen Einheit in der aktuellen Kommunikationssituation eine neue, evtl. nur momentane Bedeutung verleiht oder einen ganz neuen Ausdruck prägt. Sie hängt also insbesondere davon ab, wie oft eine der Bedeutungen, die die Einheit tragen kann, kommunikationsrelevant wird.

Wir führen ein Systembedürfnis ein, das wir

- das Anwendungsbedürfnis (Anw)

nennen wollen. Dieses Bedürfnis wirkt sich auf die (transformierte) L-FREQUENZ einer lexikalischen Einheit in einem Maße aus, das der Proportionalitätsoperator R angibt.

Die Wahrscheinlichkeit dafür, daß ein bestimmter Ausdruck sich für eine gegebene Kommunikationssituation eignet, ist aber nicht konstant, sondern eben von der Situation bzw. allgemeiner vom Kontext abhängig - und damit um so größer, in je mehr verschiedenen Kontexten der Ausdruck verwendet werden kann. Daher liegt die

Vermutung nahe, daß die Änderungsrate der L-FREQUENZ von der L-POLYTEXTIE abhängt - und zwar proportional zu einem Faktor K.

Dieser Hypothese entspricht die Differentialgleichung

$$\begin{aligned} y' &= K \\ &= - \frac{K}{y} \end{aligned}$$

wo y für die FREQUENZ und x für die POLYTEXTIE steht. Ihre Lösung

$$y = Cx^K$$

führt zur logarithmischen Gleichung

$$L-FREQUENZ = c + K \cdot L-POLYTEXTIE.$$

Setzen wir für c = ln C den Einfluß des Anwendungsbedürfnisses ein, erhalten wir mit

$$L-FREQUENZ = R \cdot Anw + K \cdot L-POLYTEXTIE$$

die Funktion, die aus der Struktur in Abbildung 14 resultiert.

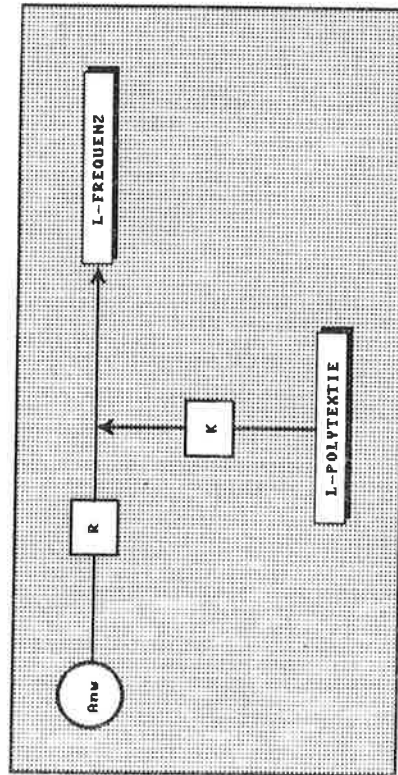


Abb. 14: Die L-FREQUENZ als Funktion des Anw-Bedürfnisses und der L-POLYTEXTIE

der Rang-Frequenz-Verteilung, die unter dem Namen Zipf-Mandelbrot'sches Gesetz bekannt ist, indem er die Wortlänge (interpretiert als "Kosten") im Durchschnitt minimierte.

Im Gegensatz dazu dient der Zusammenhang zwischen den beiden Größen in unserem Modell zur Erklärung der individuellen Ausprägung der LÄNGE.

Es ist offensichtlich, daß die Länge lexikalischer Einheiten von ihrer Frequenz etwa in der Weise abhängt, wie Zipf es darstellte. Als z.B. das Wort 'Automobil' in die Lexik mehrerer Sprachen aufgenommen wurde, war es noch sehr selten (wegen der Rarität des Automobils wurde das Designat des neuen Worts noch nicht oft kommunikationsrelevant). Mit zunehmender Häufigkeit (wegen des wachsenden Anwendungsbedürfnisses aufgrund der großen Verbreitung der Automobile) wurde der Ausdruck in allen Sprachen stark gekürzt: dt. 'Auto', skand. 'bil', indon. 'mobil' etc.

Aber es gibt nicht nur den Kürzungsprozeß; bei abnehmender Frequenz besteht die Tendenz, die kurzen Ausdrücke auch wieder zu verlängern (oder sie durch längere zu ersetzen). Lüttke (1980)

4.6.7 Die Länge (Fortsetzung)

Die Größen, die die durchschnittliche LÄNGE der lexikalischen Einheiten global bestimmen, haben wir oben bereits zusammengestellt: LEXIKONGRÖßE, PHONEMZAHL und Redundanz. In diesem Abschnitt werden wir auch die Relation in das Modellsystem aufnehmen, die für die Variabilität der LÄNGE unter den individuellen lexikalischen Einheiten verantwortlich ist.

Der Zusammenhang zwischen der Länge und der Häufigkeit von Wörtern ist wohl die bekannteste quantitative Beziehung zwischen zwei linguistischen Eigenschaften (s. Zipf 1949; Guiter 1974; Herdan 1966). Die Zipfsche Hypothese zu ihr besteht in der Vermutung, daß Wörter bei häufigem Gebrauch abgekürzt werden. Mandelbrot (1953) benutzte die Frequenz-Längen-Relation zur Herleitung

bringt, neben vielen anderen das Beispiel des heutigen französischen Ausdrucks für "heute": 'aujourd'hui', der an die Stelle des zu kurzen 'hui' (aus lat. 'hodie') getreten ist. Man müsste also eher folgendermaßen formulieren: Jede lexikalische Einheit hat eine Länge, die ihrer Häufigkeit entspricht. Diese Abhängigkeit betrachten wir als Funktion eines weiteren Systembedürfnisses und postulieren daher

die Minimierung des Produktionsaufwands (minP)

als Ursache dafür. Die Hypothese

Die Veränderungsrate der LÄNGE ist umgekehrt proportional zur FREQUENZ einer lexikalischen Einheit

fassen wir als Differentialgleichung

$$\frac{y'}{y} = \frac{-N}{x}$$

wo y für die LÄNGE, x für die FREQUENZ und -N für den Proportionalitätsfaktor des Bedürfnisses minP steht. Die Lösung

$$y = Cx^{-N}$$

ergibt die logarithmische Funktion

$$L-LÄNGE = c - N L-FREQUENZ$$

Die unbekannte Größe c = ln C ist leicht als der globale Einfluß auf die durchschnittliche LÄNGE zu interpretieren, von der die individuellen Längen der lexikalischen Einheiten gemäß ihren Frequenzen abweichen. Die vollständige Funktionsgleichung lautet also

$$L-LÄNGE = A L-LEXIKONGRÖßE + Z Red - P L-PHONEMZAHL - N L-FREQUENZ$$

oder

$$LÄNGE = \frac{A \quad Z}{LEXIKONGRÖßE \quad Red} + \frac{P}{PHONEMZAHL} - FREQUENZ$$

Dies ist auch die Funktion, die sich aus der Struktur in Abbildung 15 ergibt.

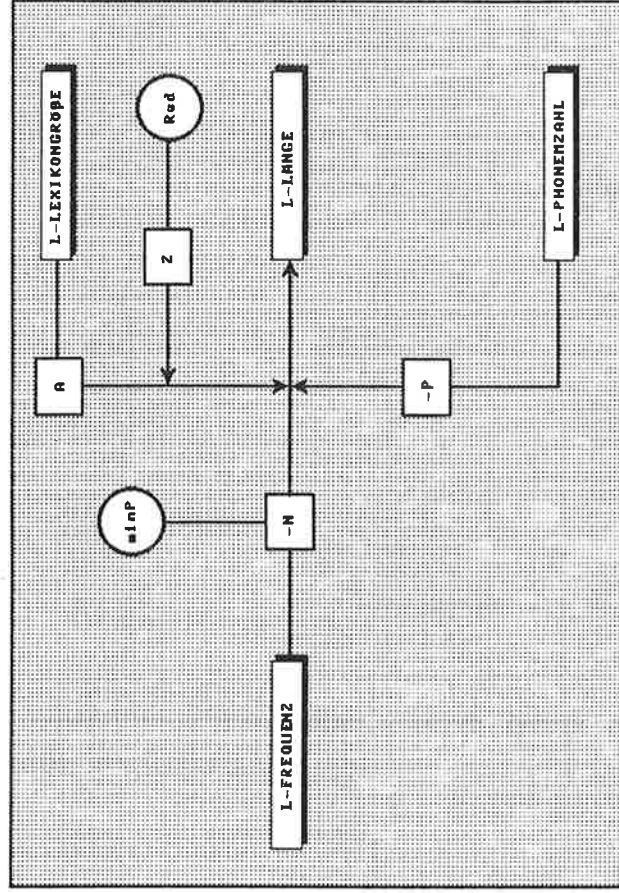


Abb. 15: Vier Systemgrößen und zwei Bedürfnisse bestimmen die LÄNGE der lexikalischen Einheiten durch Ausbildung eines Gleichgewichts

Zwischen Länge und Frequenz können weitere plausible Beziehungen angenommen werden, die einem komplizierteren Zusammenhang entsprechen. So kann man annehmen, daß die Kürzung eines Aus-

drucks auch von der augenblicklichen Länge des Ausdrucks abhängt. Ein Wort wie 'Auto' wird auch bei größerer Häufigkeitszunahme nicht mehr so stark verkürzt werden wie 'Automobil'. Dieser Annahme entspräche einer Struktur, wie sie die Abbildung 16 zeigt. Der Verzögerungsoperator E^{-1} drückt aus, daß auf die Kürzung diejenige Ausdruckslänge wirkt, die er "einen Augenblick" vor der Kürzung hatte.

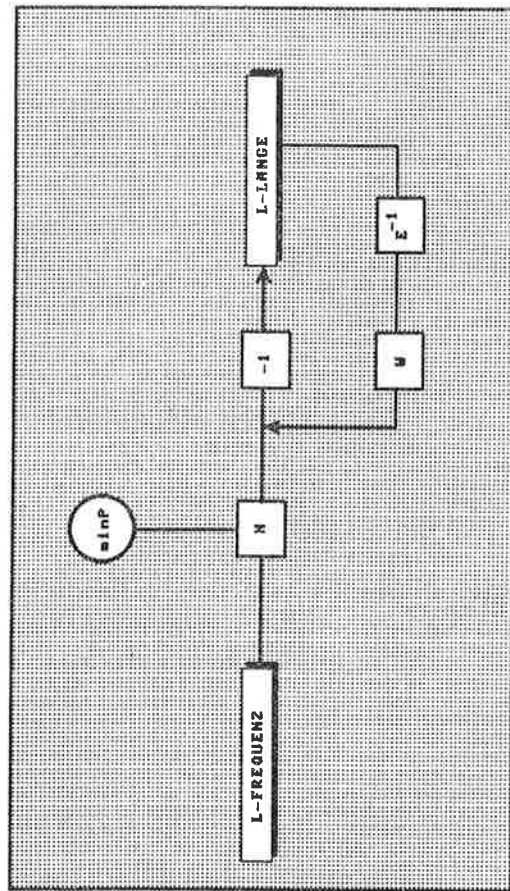


Abb. 16: Der Rückkopplungseinfluß der LÄNGE auf die Kürzung

Eine weitere Zusatzannahme könnte darin bestehen, einen Einfluß der LÄNGE auf die FREQUENZ einzubeziehen: Von zwei in Hinsicht auf die Bedeutung und den Kontext gleichermaßen geeigneten Ausdrücken wird häufiger der kürzere gewählt. Die entsprechende Struktur zeigt die Abbildung 17:

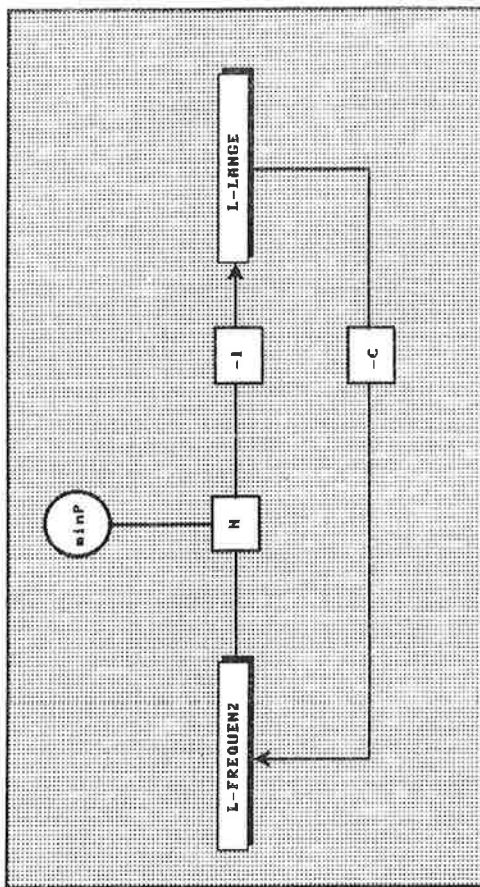


Abb. 17: Der Rückkopplungseinfluß der LÄNGE auf die FREQUENZ

Die Frage, ob diese oder weitere Verfeinerungen zu einem realistischeren Modell der sprachlichen Wirklichkeit beitragen, ist in diesem theoretischen Stadium der Untersuchung kaum beantwortbar. Wir werden daher auf alle denkbaren Erweiterungen des Modells verzichten, bis seine wesentlichen Strukturzüge einer empirischen Überprüfung (s. unten) unterzogen worden sind.

Die Abbildung 18 zeigt die Gesamtstruktur des Basismodells in der jetzigen Form.

4.7 DIE FUNKTIONEN DER TEILSTRUKTUREN

Aus den Teilstrukturen des Modells ergeben sich für alle vier strukturellen Eigenschaften der lexikalischen Einheiten Funktionen der Form

$$y^B = Ax^B,$$

die der Lösung der Differentialgleichung

$$\begin{aligned} B &= X \\ &= \\ Y &= Y \end{aligned}$$

entspricht. Für alle Abhängigkeiten zwischen den Systemgrößen, die für strukturelle Eigenschaften stehen, wurde hier auch diese Form der Differentialgleichung angesetzt²⁷. Diese vier Gleichungen stellen den Kern des Modells dar.

Die beiden Variablen LEXIKONGRÖßE und PHONEMZAHL ergeben sich aus kombinatorischen Überlegungen bzw. aus Größen, die außerhalb des Modells liegen und daher hier als empirische Parameter aufgefaßt werden können.

4.8 DIE BEDÜRFNISSE, GRÖßEN UND PROZESSE IM MODELL

Nachstehend sind in tabellarischer Form die Systembedürfnisse und ihre Wirkungen auf das System zusammengefaßt.

27 Je nach Vorzeichen der Konstanten B geben die Gleichungen direkte oder umgekehrte Proportionalitäten wieder.

Tabelle 1: Übersicht über die Systembedürfnisse

Bedürfnis	wirkt auf die Größe	in Richtung
Minimierung des Kodierungsaufwands (minK)	Phonemzahl Polylexie	- +
Minimierung des Dekodierungsaufwands (minD)	Phonemzahl Polylexie	+ -
Übertragungssicherung (Red)	Länge	+
Kodierung (Kod)	Lexikongröße	+
Spezifizierung (Spz)	Polylexie	-
Kontextökonomie (KÖ)	Polytextie	+
Kontextspezifizierung (KS)	Polytextie	-
Anwendung (Anw)	Frequenz	+
Minimierung des Produktionsaufwands (minP)	Länge	-
Minimierung der Inventargröße (minI)	Lexikongröße	-

In Tabelle 2 sind die Systemgrößen mit ihren Funktionen dargestellt, in Tabelle 3 die Prozesse im Modellsystem, die direkt auf Bedürfnisse zurückzuführen sind.

Tabelle 2: Systemgrößen und Funktionen

Systemgröße	Funktion
LEXIKONGRÖßE	LG = KOD V -L PL
PHONEMANZAHL	PH = Y1 -Y2 mind minK
LÄNGE	L = A Z -P -N LG Red PH F
POLYLEXIE	PL = Q2 -Q1 -T mind L
POLYTEXTIE	PT = KÖ KS -S1 G PL
FREQUENZ	F = R R Anw PT

Tabelle 3: Prozesse

Prozeß	bedientes Bedürfnis	Resultat
Phonologische Unifikation	minK	Verringerung d. PHONEMZAHL
Phonologische Diversifikation	minD	Vergrößerung d. PHONEMZAHL
Phonologische Restriktion	Red	Vergrößerung d. LÄNGE
Lexikalisisierung	Kod	Vergrößerung d. LEXIKONGRÖßE
Lexikalische Unifikation	minK	Vergrößerung d. POLYLEXIE
Lexikalische Diversifikation	minD	Verringerung d. POLYLEXIE
Lexikalische Reduktion	minI	Verringerung d. LEXIKONGRÖßE
Spezifizierung	Spz	Verringerung d. POLYLEXIE
Globalisierung	KÜ	Vergrößerung d. POLYTEXTIE
Zentralisierung	KS	Verringerung d. POLYTEXTIE
Anwendung	Anw	Vergrößerung d. FREQUENZ
Kürzung	minP	Verringerung d. LÄNGE

5. DEDUKTION VON KONSEQUENZEN

Aus der Struktur des Modells ist eine Reihe von Konsequenzen ableitbar, die dazu herangezogen werden können, das Modell empirisch zu überprüfen²⁸. Einigen von ihnen kommt zusätzlich eine eigenständige theoretische Bedeutung zu.

5.1 DIREKTE ZUSAMMENHÄNGE

Oben wurde die Form der Abhängigkeiten zwischen den Variablen aus Differentialgleichungsansätzen abgeleitet. Daraus folgt, daß das Basismodell den Zusammenhang zwischen den betrachteten Größen nicht nur beschreibt, sondern auch erklärt und vorhersagt.

Die Erklärung besteht in der deduktiven Ableitung der Funktionen²⁹ aus der Systemstruktur und damit aus den Gesetzhypothesen, die oben aufgestellt wurden; die Funktionen wiederum erklären die beobachteten Datenzusammenhänge. Da Gesetze (shypothesen) Allaussagen sind, folgt aus dem Modell die Vorhersage, daß die jeweiligen Variablenzusammenhänge in allen Sprachen und zu allen Zeiten den abgeleiteten Funktionen folgen, soweit die Randbedingungen erfüllt sind.

28 vgl. Kap. 6

29 s. Tabelle 2

Dies bietet bereits ausgezeichnete theoretische und empirische Überprüfungsmöglichkeiten für das Modell. Aus dem Ansatz lassen sich aber noch viel weitergehende Folgerungen für die zu erwartenden Funktionen ziehen: Die Parameter der Funktionen leiten sich ebenfalls aus der Systemstruktur her; sie schreiben damit eine bestimmte Interpretation empirisch gefundenen Parameterwerte vor; in einigen Fällen ist sogar heute schon eine Vorhersage über die Parameterausprägung möglich. Die Funktion

$$PL = Q L^{-T}$$

entspricht³⁰ der Vorhersage, daß in jeder Sprache mit variabler Ausdruckslänge

- a) die Polylexie eines Ausdrucks von seiner Länge abhängt;
- b) die Form dieser Abhängigkeit hyperbolisch ist;
- c) sich eine Längenveränderung um so mehr auf die Polylexie auswirkt, je synthetischer die Sprache ist.

Die erste Aussage impliziert eine Behauptung über die Richtung der Abhängigkeit. Es handelt sich also nicht um eine bloße Korrelation, sondern über eine quasi-kausale funktionale Beziehung. Der letzte Punkt ist bereits eine quantitative³¹ Prädiktion über den Parameter T, die das Modell sehr empfindlichen empirischen Falsifikationstests aussetzt.

30 Das Symbol Q drückt den Einfluß des Gleichgewichts zwischen den Auswirkungen der Bedürfnisse $\min K$ und $\min D$ aus.

31 Zunächst auf einer ordinalen Skala. Nach eingehenden typologischen Studien werden sich sogar numerische Vorhersagen treffen lassen: Durch Vergleich einer großen Zahl von Stichproben aus verschiedenen Sprachen und den typologischen Merkmalen dieser Sprachen ist es möglich, die für die Ausprägung des Parameters T verantwortlichen Faktoren zu finden. Hat man dann die genaue Beziehung zwischen ihnen und T bestimmt, so läßt sich die Form der Abhängigkeit $PL(L)$ daraus gesetzmäßig ableiten.

Sollte sich das Modell bewähren, bietet die Aussage c) zudem die Möglichkeit, nun umgekehrt den Parameter T als quantitatives Maß für den Synthetizitätsgrad einer Sprache zu verwenden. Analoge Überlegungen können für die weiteren Funktionsparameter angestellt werden.

5.2 INDIREKTE ZUSAMMENHÄNGE

Die systemtheoretischen Hilfsmittel erlauben es auch, aus der Modellstruktur die indirekten Variablenzusammenhänge abzuleiten. Man kann z.B. fragen, wie die Polylexie eines Ausdrucks von seiner Frequenz abhängt. Dazu werden zwei Funktionsgleichungen herangezogen: diejenige, in der die Polylexie als abhängige Variable, und diejenige, in der die Frequenz als unabhängige Variable vorkommt.

Zu jedem beliebigen Zeitpunkt sind die Einflüsse der Bedürfnisse konstant, das Gleiche gilt für die Lexikongröße und die Phonemanzahl sowie die Operatoren. Wir schreiben daher kurz

$$\begin{aligned} PL &= Q L^{-T} \text{ und} \\ L &= R F^{-N} \end{aligned}$$

Hier stehen die Konstanten Q für $\min K^2$ und R für den Ausdruck $LG^A \text{ Red}^T PH^{-r}$. Um die Frage nach der indirekten Abhängigkeit der Polylexie von der Frequenz zu beantworten, eliminieren wir die vermittelnde Variable LÄNGE, indem wir die rechte Seite der zweiten Gleichung in der ersten Gleichung für die unabhängige Variable L einsetzen:

$$\begin{aligned} PL &= Q (R F^{-N})^{-T} \\ &= Q R^{-T} F^{NT} \end{aligned}$$

Mit $S = QR^{-T}$ und $Z = NT$ ergibt sich

$$PL = SF^{TZ}$$

Während die Ausgangsgleichungen monoton fallende, hyperbolische Abhängigkeiten wiedergeben, resultiert für die indirekte Abhängigkeit der Polylexie von der Frequenz eine monoton steigende Potenzfunktion. Auch diese abgeleitete Aussage ist empirisch leicht überprüfbar.

Noch ergiebiger ist wieder, wie unter 5.1, die Betrachtung der Parameter: Die Potenzfunktion, die wir erhielten, ist durch die Ausgangsgleichungen vollständig determiniert; ihre Parameter bestimmen sich, wie gezeigt, ausschließlich aus ihnen. Dieses formale Ergebnis entspricht der Behauptung, daß für die Beschreibung des Zusammenhangs zwischen Polylexie und Frequenz keine weiteren als die in den beiden Gleichungen bereits enthaltenen Größen erforderlich sind.

Eine empirische Untersuchung dieses indirekten Zusammenhangs an einer beliebigen Sprache sollte also

- a) zu einer Potenzfunktion $PL = SF^{TZ}$ führen und
- b) Parameter ergeben, die sich numerisch als $S = QR^{-T}$ und $Z = NT$ vorhersagen lassen.

Wieder kann im Prinzip mit allen indirekten Zusammenhängen zwischen den Modellgrößen analog verfahren werden. Dabei können solche indirekten Abhängigkeiten auch über mehr als eine vermittelnde Variable gewonnen werden, indem mehrfach nacheinander eingesetzt wird. Auf diese Weise ist es möglich, sämtliche Beziehungen zwischen strukturellen Eigenschaften lexikalischer Einheiten aus dem Modellsystem zu deduzieren.

5.3 VERTEILUNGEN DER STRUKTURELLEN EIGENSCHAFTEN

Außer den Resultaten von Zipf und Mandelbrot liegt bereits eine ganze Reihe von Untersuchungen zu den Verteilungen von Eigenschaften linguistischer Einheiten vor (vgl. z.B. Carroll 1967, 1968, 1970/72; Fucks 1955; Grotjahn 1982; Herdan 1966:201-206; Krylov 1982). All diesen Arbeiten ist gemeinsam, daß sie die untersuchten Eigenschaften als isolierte Phänomene behandeln, deren empirisch beobachtbare Verteilungen es mit statistischen Mitteln möglichst gut anzunähern gilt. Eine Ausnahme bildet Krylov insofern, als er ein aus theoretischen Überlegungen abgeleitetes, also deduktiv gewonnenes Modell für die Verteilung der Bedeutungszahl von Wörtern im Lexikon vorschlägt³². Andere Autoren (Arapov und Cherc 1983; Beöthy und Altmann 1985) deduzieren überzeugend die Verteilung linguistischer Größen aus ihrer Interdependenz mit anderen Größen.

In dem hier vorgeschlagenen Modell ist die Ausprägung einer strukturellen Eigenschaft in jedem Fall von zwei Hauptkomponenten abhängig.

Die erste ist die Auswirkung eines Bedürfnisses, eines Gleichgewichts zwischen gegenläufigen Bedürfnissen, oder einer indirekt aus Bedürfnissen resultierende Größe³³. Gemeinsam ist den Varianten dieser Komponente, daß sie zu jedem Zeitpunkt Konstanten bilden, durch die den Variablen FREQUENZ, LÄNGE, POLYLEXIE und POLYTEXTIE jeweils momentan fixe Grundanteile zukommen³⁴.

32 Krylovs eigene empirische Tests seines Modells an russischen Wörterbüchern können jedoch nicht als Bestätigung angesehen werden, obwohl er selbst sie als Untermauerung wertet. Später (1982 b) hat er ein verbessertes Modell vorgeschlagen.

33 Hier die PHONEMANZAHL und die LEXIKONGRÖÖE.

34 Die Verknüpfung ist in der linearisierten Darstellung additiv.

Im Fall der POLYLEXIE haben wir diesen konstanten Anteil, den das Gleichgewicht zwischen den gegenläufigen Bedürfnissen mind und minK hervorruft, mit dem Symbol Q bezeichnet.

Die zweite Komponente besteht dagegen aus dem Einfluß der unabhängigen Variablen. Dieser ist nicht konstant, sondern variiert zu jedem Zeitpunkt über den Bereich der Werte, die die als unabhängig betrachtete strukturelle Größe annehmen kann, und bildet damit den Anteil der Variablenausprägung, der für die individuellen Abweichungen der Eigenschaften der lexikalischen Einheiten von dem fixen Grundanteil verantwortlich ist³⁵.

Für die POLYLEXIE ist die unabhängige Größe die LÄNGE der lexikalischen Einheiten; ihr Einfluß auf die POLYLEXIE wird, wie oben beschrieben, vom Spezifikationsbedürfnis über den Parameter -T modifiziert. Jede lexikalische Einheit besitzt nun einen individuellen POLYLEXIE-Wert, dessen Abweichung von dem Festwert, der aus mind und minK resultiert, eben von der LÄNGE dieser Einheit abhängt.

Aus diesem Mechanismus ergibt sich eine weitere Konsequenz. Da jede Ausprägung der unabhängigen Variablen zu einer bestimmten Ausprägung der abhängigen Variablen führt, bestimmt die Anzahl der lexikalischen Einheiten, denen ein gegebener Wert der unabhängigen Variablen zukommt, die Anzahl der Vorkommen des Funktionwertes (also der Ausprägung der abhängigen Variablen). Das Modell führt so zu der Folgerung, daß die Wahrscheinlichkeitsverteilung einer strukturellen Eigenschaft der lexikalischen Einheiten im Lexikon eine Funktion der Verteilung der sie bestimmenden strukturellen Eigenschaft ist. Wenn z. B. $n(x)$ die Anzahl der lexikalischen Einheiten mit der Länge x ist, so ist wegen $y = Qx^{-T}$ die Anzahl der lexikalischen Einheiten mit der Polylexie y im einfachsten Fall theoretisch $n(y) = n(x)$.

Dieser Zusammenhang ist empirisch nicht offensichtlich, weil die beteiligten Mechanismen nicht deterministisch wirken. So

35 Der variable Anteil entspricht der multiplikativen Verknüpfung.

stellen sich etwa die gekürzten Ausdruckslängen bei erhöhter Frequenz weder sofort noch für jeden Ausdruck zu gleicher Zeit ein: Die Anpassung der Systemgrößen unterliegt einer gewissen Verzögerung, wobei auch das Ausmaß der Trägheit uneinheitlich und von vielen, zumeist noch unbekannten Faktoren abhängig ist. Es ist zu erwarten, daß die meisten Variablenausprägungen stets zeitlich hinter ihrem Optimum zurückbleiben. Da nun also die Abhängigkeiten der Systemgrößen untereinander stochastischen Charakter besitzen, ergibt sich für die Abhängigkeit der Wahrscheinlichkeitsverteilungen voneinander die Notwendigkeit, sie mit Hilfe eines stochastischen Prozesses abzuleiten, der den Fluktuationen Rechnung trägt.

Die Lösung dieser Aufgabe wird jedoch einer gesonderten Untersuchung vorbehalten bleiben.

5.4 DYNAMIK

Weitere Konsequenzen, die sich aus dem Modell folgern lassen, betreffen die Systemdynamik.

Die Verfolgung der Ausprägung einer oder mehrerer Modellgrößen über die Zeit entspricht der diachronischen Betrachtung der modellierten linguistischen Eigenschaften. Nun läßt sich die Entwicklung einer Größe als Funktion der Entwicklung einer anderen formulieren, ähnlich wie es oben für die Verteilung von Modellgrößen skizziert wurde, so daß sämtliche strukturellen Eigenschaften in ihrem zeitlichen Verlauf vorhersagbar sind, wenn das dynamische Verhalten einer einzigen bekannt ist. Damit ergibt sich zum ersten Mal die Möglichkeit einer linguistischen Simulation.

Für die Veränderung der FREQUENZ linguistischer Einheiten haben Altmann et al (Altmann 1983; Altmann, v. Buttlar, Rott,

Strauß 1983) eine Differentialgleichung aus einem interaktionistischen Modell zur Verbreitung neuer Elemente in der Sprache abgeleitet; die Zeitfunktionen, die sich als Lösungen ergeben, sind empirisch gut bestätigt (vgl. Müller-Hasemann 1983; Imsiepen 1983; Best 1983; Kohlhasse 1983; Best, Kohlhasse 1983). Altmann gab diesem Gesetz des Sprachwandels den Namen Piotrowski-Gesetz.

Nach diesem Gesetz bestimmt sich der Anteil einer beliebigen Spracherscheinung an der Gesamtzahl der Verwendungen in Abhängigkeit von der Zeit t bei unvollständiger Veränderung mit

$$p = \frac{c}{1 + ae^{-bt}}$$

bei vollständiger Veränderung (wenn die neue Erscheinung sich durchsetzt und alle anderen Verwendungen verdrängt) mit

$$p = \frac{1}{1 + ae^{-bt}}$$

und bei reversibler Entwicklung (wenn die Erscheinung nach Erreichen eines Maximums wieder seltener wird) mit

$$p = \frac{1}{1 + ae^{-bt+ct^2}}$$

Mit diesem Gesetz sind die Voraussetzungen gegeben, das hier entwickelte Modell anhand diachroner Daten zu überprüfen. Ein solches Vorhaben ist jedoch mit einem nicht ohne weiteres leistbaren Aufwand verbunden: Für die meisten Sprachen stellt sich eine repräsentative Datengewinnung als nahezu aussichtslos dar.

An dieser Stelle soll daher der Hinweis auf die prinzipielle Möglichkeit genügen. Praktische Untersuchungen werden sich vermutlich auf einzelne Aspekte beschränken müssen.

Auch der Versuch einer breiter angelegten Simulation, deren Auswertung wertvolle Hinweise verspricht, kann hier nur in Aussicht gestellt werden.

6. DIE EMPIRISCHE ÜBERPRÜFUNG DES MODELLS

Theoretische Plausibilität und Konsistenz der Hypothesen sind die ersten und wichtigsten Anforderungen an ein Modell mit Erklärungsanspruch; darüber hinaus jedoch verlangt unser wissenschaftstheoretischer Standpunkt eine empirische Überprüfung der Aussagen³⁶.

Um das hier entwickelte Basismodell zu testen, werden einige der oben besprochenen Konsequenzen aus der Modellstruktur herangezogen.

6.1 OPERATIONALISIERUNG

Zur Gegenüberstellung der oben abgeleiteten Behauptungen mit empirischen Daten bedarf es noch einer Angabe darüber, welche linguistischen Beobachtungsbegriffe mit den theoretischen Konstrukten des Modells korrespondieren; es muß festgelegt werden, wie die entsprechenden Einheiten zu identifizieren und auf welche Weise ihre Eigenschaften zu messen sind.

³⁶ Daher sind auch nur solche Modelle für den Aufbau einer linguistischen Theorie von Interesse, die (wenigstens im Prinzip) solchen Tests zugänglich sind. Nichtempirische Modelle und andere gegen Überprüfung immunen Hypothesen können dazu nicht beitragen.

Wörter unterliegen die Ausprägungen ihrer Eigenschaften den im Modell abgebildeten Prozessen; verwendet wird ein Wort aber stets in einer morphosyntaktischen Umgebung, die seine Aktualisierung auf eine bestimmte Wortform festlegt.³⁷

Für die Wahl der Wortform als Beobachtungsgröße spricht auch die Tatsache, daß die verschiedenen Wortformen eines Paradigmas ja in bezug auf die Eigenschaften Frequenz und Länge differieren. Die strukturellen Größen Polylexie und Polytextie dagegen, deren Rolle im Modell ebenso zentral ist, verhalten sich gegenüber flexionsmorphologischen Unterschieden indifferent:

Bedeutungen und Verwendungskontexte sind Entitäten, die durch Abstraktion von der Aktualisierung der Wörter in einem grammatischen Gefüge zustande kommen. Dieses Argument spricht für die Verwendung von Lemmata anstelle von Wortformen.

Der Gedanke, je nach untersuchtem Zusammenhang einmal diese und einmal jene Definition zu wählen, muß aus zwei Gründen zurückgewiesen werden. Obwohl an sich für die Überprüfung der Abhängigkeit der Länge von der Frequenz die Verwendung von Wortformen und für die anderen Funktionen die Verwendung von Lemmata zweifellos unproblematisch ist, wären die resultierenden Meßwerte wegen der Definitionsinkonsistenz nicht vergleichbar.³⁸ Der zweite Einwand schließt ein solches Vorgehen vollkommen aus: Das Modell bildet die lexikalischen Einheiten als Punkte im lexikalischen Raum ab, dessen Dimensionen durch die strukturellen Eigenschaften gebildet werden. Damit ist festgelegt, daß empirische Messungen dieser Eigenschaften als Vektormessungen durchgeführt werden müssen, d.h. für jede untersuchte lexikalische Einheit sind alle strukturellen Eigenschaften gleichzeitig festzustellen³⁹, was eine einheitliche Operationalisierung voraussetzt.

37 Für jede syntaktische, lexikalische und morphologische Deutigkeit der Konstruktion

38 Die gemessenen Frequenzen und Längen bezögen sich dann auf jeweils auf unterschiedliche Entitäten.

39 Damit sind einfachere Verfahren ausgeschlossen, durch die z.B. die Frequenz von Wörtern bestimmter Länge aus einer

Die Angemessenheit einer Operationalisierung läßt sich allerdings niemals - jedenfalls nicht auf logisch haltbare Weise - nachweisen. Zwar soll die allgemeine Problematik der Konstruktion von Korrespondenzregeln zur Übersetzung von theoretischen Konstrukten in Beobachtungsbegriffe hier nicht näher diskutiert werden (einen Überblick über mögliche Vorgehensweisen findet man z.B. in Scheuch, Zehnpfennig 1974: 165-170), eine Bemerkung über die Konsequenz einer durchgeführten Operationalisierung ist jedoch angebracht:

Gegen die Definitionen und Meßvorschriften, die unten für die theoretischen Begriffe des Modells angegeben werden, lassen sich ohne Zweifel zahlreiche Einwände, die sich aus unterschiedlichen Sichtweisen des Gegenstands ergeben können, finden. Aber auch ein gut begründeter Einwand gegen eine Operationalisierung allein entkräftet das Ergebnis der Untersuchung nicht. Je nach Operationalisierung sind verschiedene Resultate einer Analyse zu erwarten; konkrete Ergebnisse haben lediglich relativ zu einer gegebenen Menge von Korrespondenzregeln Gültigkeitsanspruch und sind nur unter dieser Voraussetzung bewertbar (vgl. auch Galtung 1967). Unter dieser Voraussetzung werden in den folgenden Abschnitten Operationalisierungen und Meßvorschriften für die Konstrukte aus dem Modell gesucht.

Ein grundlegender Begriff ist der der lexikalischen Einheit. Obwohl das Modell so angelegt ist, daß als lexikalische Einheiten alle Arten von lexikalisierten Ausdrücken erfaßt werden können, soll hier aus praktischen Gründen die Anwendung zunächst auf Wörter beschränkt werden. Zur Operationalisierung des Begriffs Wort bieten sich mehrere Möglichkeiten an, von denen keine vollständig befriedigen kann. Abgesehen von den immer auftretenden Schwierigkeiten bei der Behandlung von Diskontinuitäten besteht ein grundsätzliches Dilemma:

Als strukturelle Größen spielen die Länge und die Frequenz von lexikalischen Einheiten eine zentrale Rolle. Diese Eigenschaften lassen sich am ehesten den Wortformen zuordnen, die in der Rede beobachtbar sind. Denn nur in der tatsächlichen Verwendung der

Andernfalls wären zwar die einzelnen aus dem Modell ableitbaren Gesetzhypothesen testbar, nicht aber das Modell insgesamt und direkt.

Das zu testende Modell gibt keiner der beiden Lösungsprinzipien einen theoretischen Vorzug, so daß hier beide Operationsalisierungen parallel⁴⁰ verwendet werden sollen.

Als nächstes sind die vier strukturellen Eigenschaften der lexikalischen Einheiten zu operationalisieren. Die Länge bietet hier die geringsten Schwierigkeiten. Wegen der Beschränkung auf Wörter kommen als linguistisch sinnvolle Längeneinheiten nur Morpheme, Silben, Phoneme, und Grapheme in Frage. Da im Modell Morpheme und Silben nicht explizit⁴¹ abgebildet sind, die Phoneme / Graphem - Ebene dagegen vertreten ist, liegt die Verwendung letzterer zur Längenmessung näher. Die Auswahl unter diesen beiden spielt keine Rolle; zwar gibt es keine eindeutige Zuordnung zwischen den Phonemen und Graphemen der meisten Sprachen, aber beide Zeichensysteme sind Abbildungen der sprachlichen Realität, die auf gleicher Weise durch Abstraktion entstehen und von dieser etwa gleich weit entfernt sind.

Da Testmaterial, wie Textkorpora und Wörterbücher, überwiegend in orthographischer Form vorliegt, soll als Maß für die Länge einer lexikalischen Einheit die Anzahl ihrer Grapheme festgesetzt werden. Im Fall von Wortformen in Texten gilt die von zwei Leerzeichen umschlossene Graphemfolge⁴², im Fall von Lemmata

Stichprobe mit der Polylexie von Wörtern der gleichen Länge aus einer anderen Stichprobe in Beziehung gesetzt werden könnte.

40 selbstverständlich voneinander unabhängig und vollständig getrennt

41 Implizit wird auf Morpheme bei der Modellierung der POLY-LEXIE - LÄNGE - Relation und des Einflusses des Spezifikationsbedürfnisses Bezug genommen.

42 wobei Diskontinuitäten vernachlässigt werden

die Graphemfolge der zugehörigen Grundform⁴³ als die zu messende lexikalische Einheit.

Für die Länge der Lemmata wären natürlich auch andere Lösungen denkbar. So kann man einwenden, daß ja nicht die Länge einer willkürlich gewählten - "Grundform" diejenige ist, die beim Gebrauch einer lexikalischen Einheit in Erscheinung tritt, sondern eben die Länge der jeweils verwendeten Wortform. Aus dieser Argumentation wäre als Maß für die Länge folgerichtig eine "wirksame Länge" abzuleiten, die aus den Längen der tatsächlich verwendeten Wortformen des Lemmas (gewichtet nach deren Einzelhäufigkeiten) zu mitteln wäre.

Diese Möglichkeit wurde vor allem wegen des hohen Meßaufwandes nicht gewählt, sie sollte aber zur Verbesserung der Messungen herangezogen werden, wenn die oben beschriebene, größere Meßvorschrift sich als unzureichend herausstellte.

Ein weiterer Einwand könnte entstehen, wenn man bedenkt, daß ein systematischer Fehler durch die unterschiedlichen Längen der Flexionsmorpheme der Lemmata verschiedener Wortarten hervorgerufen wird (vgl. die endungslose prädikative Form der Adjektiv-Grundform mit der Infinitivendung der Verben). Auch hier sollte eine entsprechende Korrektur der Meßvorschrift (Längenmessung ohne Flexionsmorpheme) vorgenommen werden, soweit die Notwendigkeit dazu aufgrund der Datenlage nicht ausgeschlossen werden kann.

Die Messung der Polylexie bietet erheblich größere Schwierigkeiten. Leider existiert zur Zeit kein Algorithmus zur Bestimmung der Bedeutungspotenz von Ausdrücken, so daß der einzige praktikable Weg derjenige zu sein scheint, den Altmann et al.⁴⁴

43 wie sie in der traditionellen Grammatik üblich sind:

Infinitiv, Nom. sing. mask. etc. Selbstverständlich wäre jede beliebige Flexionsform eines Paradigmas gleichermaßen geeignet.

44 s. z. B. Altmann, Beöthy, Best (1982)

gehen: Als Schätzung für die Anzahl der unterschiedlichen Bedeutungen einer lexikalischen Einheit, wird die Zahl der Wörterbucheinträge zu dieser Einheit verwendet. Obwohl, wie zu erwarten, beträchtliche Diskrepanzen zwischen verschiedenen Wörterbüchern bestehen, kann man davon ausgehen, daß sich die vom Modell vorhergesagten Gesetzmäßigkeiten zumindest im Großen und Ganzen wiederfinden lassen. Oder andersherum ausgedrückt: Wenn sogar eine solch grobe Meßmethode zu Resultaten führen sollte, die den Prädiktionen entsprechen, dann ist eine gute Bestätigung des Modells erreicht worden. Sollte der Test fehlschlagen, so müßte erst durch weitere Anstrengungen entschieden werden, ob ein besser geeignetes Maß nötig ist, oder ob das Modell in diesem Punkt als fehlerhaft zu gelten hat.

Die Polytextie ist als Maß für die Anzahl der Kontexte eingeführt worden, in denen die Verwendung eines bestimmten Ausdrucks möglich ist. Es ist vollkommen klar, daß für eine solche Größe nur sehr grobe Messungen vorgenommen werden können. Für die empirische Überprüfung des Modells legen wir fest: Die Polytextie einer lexikalischen Einheit wird durch die Anzahl der Texte bestimmt, in denen diese Einheit wenigstens einmal vorkommt. Da die Zahl aller Texte unendlich ist, kann nur eine Stichprobenschätzung anhand eines Textkorpus erfolgen.⁴⁵

Mehrere nichttriviale Probleme entstehen bei der Operationalisierung des Frequenz-Begriffs. In (1983 c) hat Orlov plausibel gezeigt, daß die Auftretenswahrscheinlichkeit von Wörtern in bezug auf eine Sprache insgesamt nicht ermittelbar ist⁴⁶, da jede

45 Keine Zusammenstellung von Texten zu einem Korpus kann für sich in Anspruch nehmen, repräsentativ für alle Texte einer Sprache in einem gegebenen Zeitintervall zu sein. Trotzdem sollte sich die zu überprüfende Abhängigkeit der Polytextie von der Polylexie selbst dann nachweisen lassen, wenn ein sehr einseitiges Korpus mit vielen gleichartigen Texten als Textgrundlage dient. Die vorhergesagte hyperbolische Funktion könnte dann zwar eine extreme Form annehmen, müßte aber dennoch den Daten entsprechen.

46 Dies gilt jedenfalls für den heutigen Stand der empirischen Forschung in der Sprachwissenschaft.

Stichprobe - also jedes Textkorpus - für eine valide Wahrscheinlichkeitsschätzung zu klein ist, selbst wenn sie mehrere Millionen laufende Wörter umfaßt. Dies ist zum einen eine Konsequenz der - für statistische Methoden äußerst ungünstigen - steil fallenden Frequenz-Belegungs-Verteilung der Wörter, die zwar seit Zipf bekannt ist, deren Auswirkungen auf die Stichprobenparameter aber bisher nicht beachtet wurden: Ein Großteil der Wörter eines Textes (etwa die Hälfte der Wortvorkommen) sind hapax legomena, während für eine verlässliche Schätzung Häufigkeiten von mindestens 10 Vorkommen nötig sind. Gleichzeitig sorgt aber gerade die Wiederholung von Wörtern in Texten dafür, daß die Anzahl der in einer Stichprobe erfaßten Wörter immer sehr klein gegenüber dem lexikalischen Inventar einer Sprache ist, so daß die Unterrepräsentation der meisten Wörter zu einer starken Häufigkeitsüberschätzung der vorgekommenen führt.

Orlovs Argumentation läßt keinen gangbaren Ausweg offen; eine Textstichprobe, deren Umfang hinreichend über dem Vokabularumfang liegt - also in der Größenordnung von einigen zehn Millionen laufenden Wörtern - ist heute nicht zu verwirklichen. Die Konsequenzen daraus sind jedoch für die Testbarkeit des Modells nicht so katastrophal, wie es auf den ersten Blick scheint. Einerseits besteht zwar keine Möglichkeit, eine unverzerrte Schätzung für die Frequenz-Verteilung zu erhalten, andererseits ist diese Verteilung selbst auch nicht der Untersuchungsgegenstand. Vielmehr geht es um den Zusammenhang zwischen Ausprägungen dieser Variablen und Ausprägungen der Länge bzw. der Polytextie. Selbstverständlich werden alle entsprechenden Berechnungen wegen der Orlovsschen Verzerrung ebenfalls verzerrt, aber alle auf genau die gleiche Weise. Die empirisch ermittelten Kurven werden proportional zur Frequenzüberschätzung zu steil gegenüber der nicht erfaßbaren Sprachwirklichkeit ausfallen, und gerade dadurch in bezug auf die Abhängigkeiten abbildungsstreu sein. Jedenfalls gibt es keinen theoretischen Grund, eine Abweichung des KurvenTyps aufgrund dieser Verzerrung zu erwarten.

Die einfache Vorkommenshäufigkeit einer lexikalischen Einheit im untersuchten Textkorpus soll daher als Entsprechung der theoretischen Größe Frequenz angenommen werden.

Je nachdem, auf welche Weise der Begriff der lexikalischen Einheit operationalisiert wurde, muß eine bestimmte Meßvorschrift eingehalten werden, damit die Messungen der strukturellen Eigenschaften aufeinander beziehbar sind:

Als Frequenz von Wortformen wird direkt ihre absolute Häufigkeit im Korpus angesetzt; die Frequenz eines Lemmas soll sich aus der Summe der absoluten Häufigkeiten der zum Lemma gehörenden Wortformen ergeben.

6.2 DATENERHEBUNG

Zur konkreten empirischen Überprüfung des Modells wurden aufgrund der oben dargestellten Überlegungen zur Operationalisierung Daten erhoben, die den theoretischen Begriffen entsprechen: die Größen Länge, Häufigkeit und Polytextie von Wörtern der deutschen Sprache aus einer Analyse eines Textkorpus, die Polylexie unter Verwendung eines Wörterbuchs.

Wegen des großen Aufwands, den eine solche Untersuchung verursacht, wurde die Durchführung der Tests zunächst auf eine einzige Sprache beschränkt; weitere sollen zu einem späteren Zeitpunkt folgen.

Als erste Testsprache wurde naheliegenderweise die deutsche gewählt. Als Testmaterial diente das LIMAS-Korpus⁴⁷, eine Zusammenstellung von 500 Texten (bzw. Fragmenten), deren Auswahl die

47 vergl. Schaefer (1976)

Autoren im Hinblick auf Textsortenrepräsentativität⁴⁸ betrieben haben. Das Korpus ist so konstruiert, daß jeder Einzelextrakt eine Länge von (nicht immer genau) 2000 Wörtern aufweist. Abgesehen von den übrigen Vor- und Nachteilen eines solchen Aufbaus eignet sich das LIMAS-Korpus daher sehr gut zur Gewinnung der erforderlichen Daten und besitzt auch einen vertretbaren Umfang (1 Mio. laufende Wörter).

Aus diesem Material wurden die Größen Frequenz, Polytextie und Länge von Wortformen direkt erhoben⁴⁹. Mit Hilfe des Lemmatisierungsprogrammes LEMMA2 von Willee⁵⁰ erfolgte eine halbautomatische Zuordnung der Wortformen zu Lemmata und Wortarten. Da der manuelle Anteil der Arbeiten beträchtlich ist (sämtliche Lemmatisierungsvorschläge des Programms müssen überprüft und korrigiert werden, dasselbe gilt für die Wortartenzuordnung), war die Beschränkung der Stichprobe auf etwa ein Fünftel des Vokabulars notwendig: Von den 125000 Wortformen, die im LIMAS-Korpus vorkommen, wurden nur diejenigen mit den Anfangsbuchstaben D, G, H, L oder R weiter betrachtet⁵¹.

Aus diesen entstand die Liste der 13000 Lemmata als Stichprobe aus der gleichen Grundgesamtheit wie oben, bei deren Erhebung aber eine andere Einheit als Korrelat zur "lexikalischen Einheit" diente.

48 Dieser Anspruch soll hier nicht diskutiert werden. Im Prinzip ergeben sich durch Nichterfüllung in diesem Zusammenhang sehr ähnliche Probleme wie bei der Behandlung der Frequenz (s.o.). Allerdings erscheint auch eine analoge Einschätzung der Konsequenzen plausibel, so daß mit dem Material gearbeitet werden konnte.

49 Die Gewinnung der Rohdaten und deren weitere rechnerische Verarbeitung wurden auf der Rechenanlage des Essener Hochschulinformatikzentrums durchgeführt.

50 vgl. Willee (1979)

51 Auf eine - methodologisch vorzuziehende - zufällige Stichprobe mußte aus technischen Gründen verzichtet werden.

Die Messung der strukturellen Eigenschaften der Lexeme erfolgte durch erneute Auszählung der Vorkommen der zugehörigen Wortformen im Korpus, so daß sich die Frequenz eines Lemmas aus der Summe der Frequenzen der zugehörigen Wortformen ergab. Die Polytextie eines Lemmas wurde als die Zahl der Texte ermittelt, in denen wenigstens eine Wortform des Lemmas vorkam. Als Länge wurde die tatsächliche Länge des Lemmas in Buchstabenzahl gemessen⁵².

Alle aus dem Modell ableitbaren Zusammenhänge wurden parallel anhand beider Stichproben überprüft.

Darüber hinaus wurden Gegenproben im Hinblick auf systematische Fehler durchgeführt, die durch offene oder verdeckte unterschiedliche Behandlung der Wortarten induziert sein könnten. Dazu wurden aus jeder der beiden Stichproben drei zusätzliche Stichproben erzeugt, die aus den offenen Lexikonklassen der Nomina, Verben und Adjektive bestanden. Auf diese Weise wurde jeder Test achtfach ausgeführt; die folgenden Tabelle zeigt das Verhältnis der Stichproben zueinander.

Tabelle 4: Stichproben aus der Wortliste zum LIMAS-Korpus

Wortformen		Lemmata			
gesamt		gesamt			
Nom	Vb	Adj	Nom	Vb	Adj

52 Obwohl die Umlaute und das 'ß' im Limas-Korpus zweigraphemig dargestellt sind, wurden sie bei der Zählung, wie sonst üblich, als ein einziges Zeichen angesehen.

Zur Überprüfung der Beziehungen, in denen die Polylexie eine Rolle spielt, wurden nun noch die entsprechenden Daten aus einem Wörterbuch benötigt. Zu diesem Zweck wurde das "Deutsche Wörterbuch" von Wahrig⁵³ herangezogen.

Da die Erhebung der POLYLEXIE-Werte rein manuell durchgeführt werden mußte, wurde eine weitere Reduktion des Stichprobenumfangs vorgenommen. Aus der Stichprobe der Lemmata wurde eine kleinere Stichprobe gezogen, deren Umfang 1325 Einheiten betrug. Für diese wurden der POLYLEXIE - Wert aus dem Wörterbuch von Wahrig bestimmt, indem die Anzahl der darin als unterschiedlich gekennzeichneten Bedeutungen festgehalten wurde.

Zu den acht oben erwähnten kommt also für Fragen, die mit der POLYLEXIE zusammenhängen, diese Stichprobe als neunte hinzu.

6.3 DAS TESTVERFAHREN

Der wichtigste Teil der empirischen Überprüfung des Modells besteht darin, die theoretisch behaupteten Zusammenhänge zwischen Paaren von Systemgrößen den Variablenabhängigkeiten gegenüberzustellen, die sich aus den Stichproben ergeben. Jede strukturelle Verknüpfung von zwei Modellgrößen entspricht einer Hypothese nach folgendem Hypothesenschema:

H_v : Die Ausprägung der Variablen y ist eine Funktion der Ausprägung der Variablen x mit der Funktionsgleichung

$$y = Ax^b$$

53 Wahrig (1981)

Aus diesem Schema werden alle benötigten Einzelhypothesen generiert, indem für die Symbole x und y die den Systemgrößen entsprechenden Variablen eingesetzt werden (alle strukturellen Eigenschaften sind ja nach dem Modell auf gleichartige Weise miteinander verknüpft). Im Einzelnen sind dies die Hypothesen⁵⁴:

- $H_L: L = A F^B$ (Länge als Funktion der Frequenz);
 $H_{PL}: PL = A L^B$ (Polylexie als Funktion der Länge);
 $H_{PT}: PT = A PL^B$ (Polytextie als Funktion der Polylexie);
 $H_F: F = A PT^B$ (Frequenz als Funktion der Polytextie).

Die Konstanten A fassen in den Funktionsgleichungen diejenigen Wirkungskomponenten zusammen, die keine Funktion der unabhängigen Größe sind. In H_L steht beispielsweise diese Konstante für das Produkt

$$A \quad Z \quad -P \\ LG \quad Red \quad PH,$$

also den gemeinsamen Einfluß von Lexikongröße, dem Bedürfnis nach Übertragungssicherheit und der Inventargröße des Phonemsystems.

Die so aufgestellten Hypothesen werden nun auf Verträglichkeit mit den Daten der Stichproben überprüft. Dazu werden aus allen Stichproben Mengen von Wertepaaren aus den gemessenen Ausprägungen der strukturellen Eigenschaften gebildet⁵⁵. Dann werden die

54 Die Konstanten-Symbole A und B stehen in jeder Hypothese für andere Konstanten! Die Bezeichnungen wurden wegen der größeren Einfachheit jedesmal gleich gewählt.

55 Da es sich bei den untersuchten Erscheinungen um stochastische Abhängigkeiten handelt, kommen bei der Messung für die x - Werte oft mehrere verschiedene y - Werte vor, aus welchen dann der Mittelwert gebildet wird. Aus diesem Grund ergeben sich in der Regel weit weniger Datenpunkte für die Anpassung, als Werte in der Stichprobe enthalten sind.

Funktionsgleichungen linearisiert und nach dem Algorithmus von Householder⁵⁶ (s. zB. Stoer 1979: 164-171) diejenigen Werte für die Parameter A und B ermittelt, die die geringsten Abweichungen der Daten von den theoretischen Werten ergeben (Kurvenganpassung). Schließlich wird mit Hilfe des F - Testes festgestellt, ob die Abweichungen der empirisch gemessenen Variablenausprägungen y_i von den theoretisch zu erwartenden Werten y_i gering genug sind, um als zufällig angesehen werden zu können.

Dieser Test setzt die Summe der erklärten Abweichungen der Datenpunkte von der Regressionsgeraden in Beziehung zu den unerklärten Abweichungen:

$$F_{1,1,f2} = \frac{SSR / (k-1)}{SSE / (n-k)}$$

$$SSR = (y - y_1)^2 \quad \text{(erklärte Varianz)}$$

$$SSE = (y_1 - y_1)^2 \quad \text{(unerklärte Varianz),}$$

wobei k die Anzahl der Parameter und n die Anzahl der Datenpunkte angibt. Die Anzahl der Freiheitsgrade bestimmt sich als

$$f1 = k-1; \quad f2 = n-k$$

(vgl. Walker, Lev 1953: 230ff). Der resultierende F - Wert ist ein Maß für die Güte der Anpassung. Berechnet man noch die Wahrscheinlichkeit dafür, einen ebensogroßen oder größeren F - Wert als den gefundenen durch Zufall zu erreichen, so verfügt man über die Möglichkeit, verschiedene Anpassungen miteinander zu vergleichen.

Alle Einzelhypothesen wurden diesem Testverfahren unterzogen, wobei festgesetzt wurde, daß eine Hypothese dann abzulehnen ist, wenn der resultierende F - Wert mit einer größeren Wahrscheinlichkeit

56 In erster Approximation wäre auch die Methode der kleinsten Fehlerquadrate verwendbar.

lichkeit als $P = 0.01$ auch zufällig erreichbar ist. Eine Hypothese wird also nur dann vorläufig⁵⁷ angenommen, wenn sichergestellt ist, daß die Übereinstimmung von theoretischen und empirischen Werten mit mindestens 99-prozentiger Konfidenz für die Richtigkeit der Hypothese spricht.

6.3.1 Die Polylexie

Aus der Struktur des Modells folgt die Hypothese, daß die Anzahl der Bedeutungen eines Ausdrucks eine Funktion der Länge des Ausdrucks ist. Die Konstante A in der Funktionsgleichung $PL = A L^B$ gibt den globalen Einfluß des Gleichgewichts zwischen den Wirkungen der Bedürfnisse $minD$ und $minK$ wieder, während die Konstante B ein Maß der Synthetizität der Sprache ist und auf das Bedürfnis "Spezifizierung" zurückzuführen ist.

Die Anpassung der Hyperbel an die Daten aus der Stichprobe ergab die Schätzung in Tabelle 5⁵⁸.

Einen optischen Eindruck vom Kurvenverlauf und von der Anpassung vermittelt Abbildung 19. Hier und in allen weiteren Graphiken sind die empirischen Datenpunkte durch das Symbol "x" dargestellt; die kontinuierliche Linie entspricht der theoretischen Funktion mit den aus der Stichprobe geschätzten Parameterwerten.

57 vorläufig deshalb, weil induktive Tests keine Verifikation von Hypothesen erlauben. Eine mißlungene Falsifikation berechtigt dazu, die Hypothese so lange als gültig anzusehen, bis sie möglicherweise doch noch widerlegt wird.

58 Die empirischen und theoretischen Werte der Stichprobe sind in Anhang A 1 tabelliert.

Tabelle 5: Anpassung der Funktion $PL = A L^B$

A	B	FG	F	F	P
12.5	-0.828	1,22	130	7.95	-10 10

A, B : Empirische Konstanten

FG : Anzahl der Freiheitsgrade

F : F - Variable. Die Hypothese wird auf dem Signifikanzniveau 0.01 angenommen, wenn dieser Wert größer als der Tabellenwert ist.

F_{0.01} : Theoretischer F-Wert auf dem 0.01-Niveau

P : Wahrscheinlichkeit, durch Zufall einen gleichgroßen oder größeren F - Wert zu erhalten

Die Ergebnisse entsprechen völlig den Erwartungen: der Parameter B (Synthetizität) hat ein negatives Vorzeichen⁵⁹, und die theoretische Kurve gibt die gemessenen Werte ausgezeichnet wieder. Aufgrund des F - Wertes, der den geforderten um Größenordnungen übersteigt, wird die Hypothese angenommen.

Für die Überprüfung des Zusammenhangs dieser beiden Größen kommt nur die eine Stichprobe in Frage⁶⁰.

59 vgl. Abschn. 5

60 vgl. Abschn. 6.1

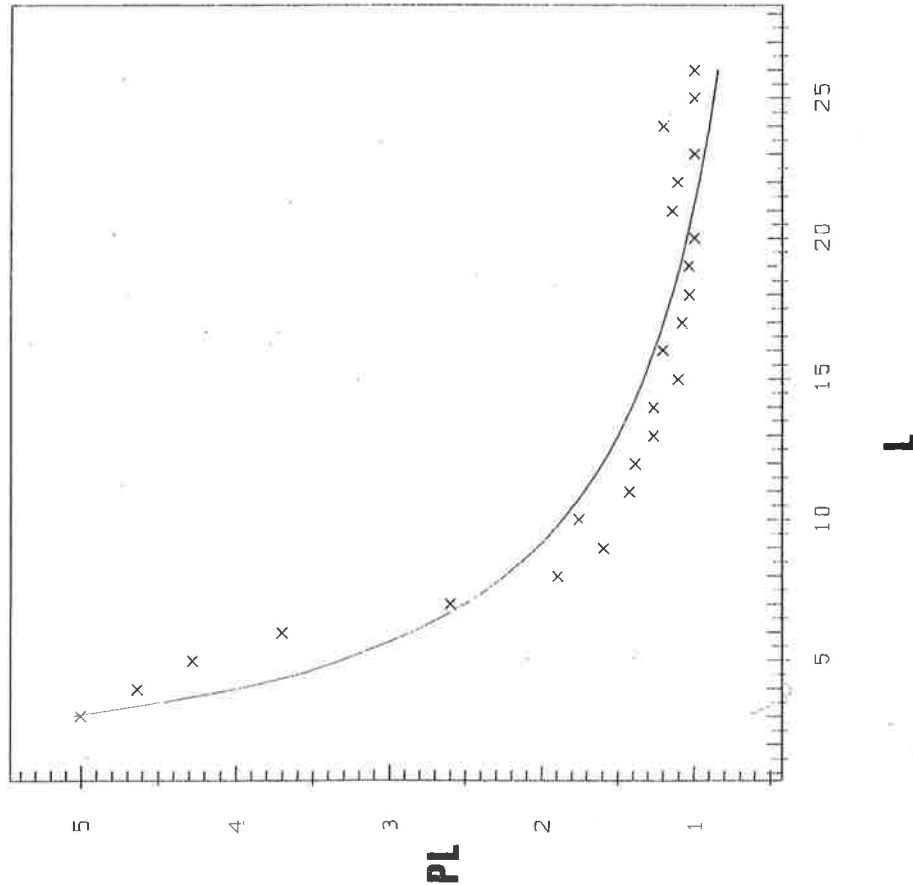


Abb. 19 : Die Polylexie (PL) als Funktion der Länge (L)

6.3.2 Die Polytextie

Die aus dem Modell abgeleitete Hypothese besagt, daß die Polytextie eines Ausdrucks hyperbolisch von seiner Polylexie abhängt. Die Funktionsparameter haben folgende Bedeutung: A steht für das Gleichgewicht aus den Wirkungen der Kontextökonomie und der Kontextspezifikation, B gibt den Einfluß der Polylexie wieder.

Die Ergebnisse der Anpassung dieser Funktion an die Daten der gleichen Stichprobe geben die Tabelle 6 und die Abbildung 20 wieder⁶¹.

Tabelle 6: Anpassung der Funktion $PT = PL^B$

A	B	FG	F	P
3.92	1.56	1,14	17.4	8.86 0.0009

Auch hier wird die Hypothese in vollem Umfang bestätigt. Der Parameter B ist positiv, wie es die Annahmen bei der Modellkonstruktion nahelegen: ein größerer Bedeutungsumfang führt zu einer größeren Menge von Kontexten (und damit Texten), in denen ein Ausdruck verwendet werden kann.

Der F - Wert ist etwa doppelt so hoch wie erforderlich, um die Hypothese vorläufig anzunehmen.

61 Die Datenpunkte und die entsprechenden theoretische Werte finden sich im Anhang A 2

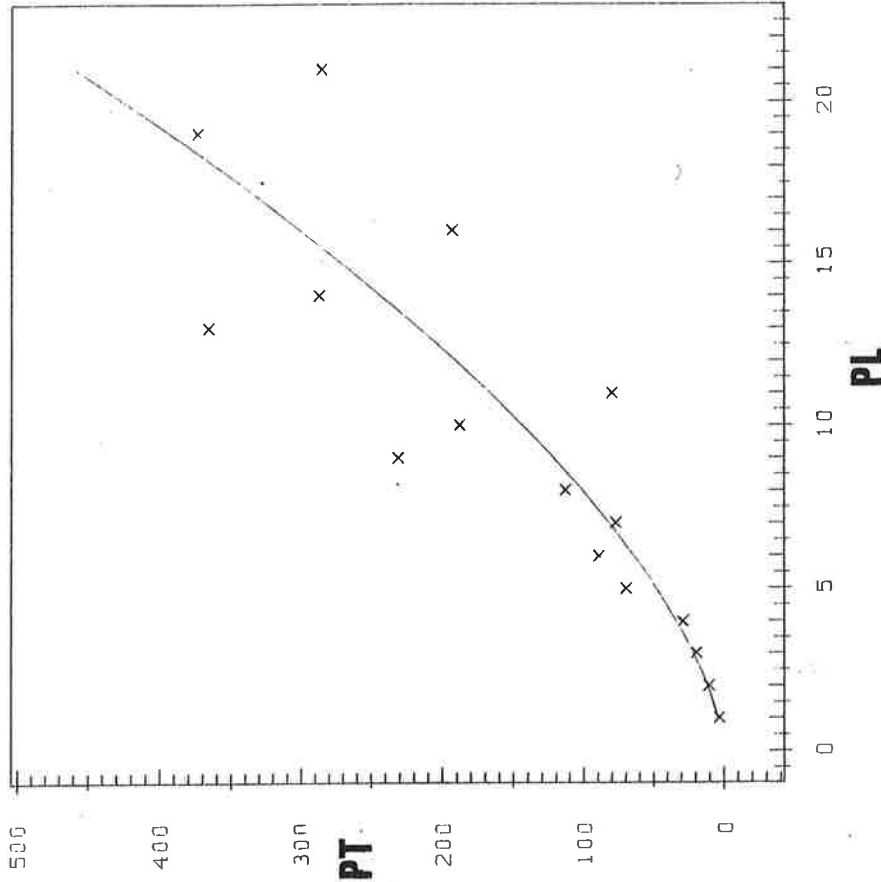


Abb. 20: Die Polytextie (PT) als Funktion der Polylexie (PL)

6.3.3 Die Frequenz

Die Frequenz einer lexikalischen Einheit bestimmt sich nach dem Modell aus der Wirkung des Anw - Bedürfnisses und aus der Polytextie dieser Einheit. In der Funktionsgleichung steht die Konstante A für den Einfluß von Anw, die Konstante B für die Auswirkung der Polytextie.

Zunächst sollen die Resultate der Anpassung an die Daten der gleichen Stichprobe dargestellt werden, die für die ersten beiden Zusammenhänge herangezogen wurde. Die Tabelle 7 und Abbildung 21 zeigen eine sehr gute Anpassung⁶². Der F - Wert beträgt fast das 50-fache des zur Annahme der Hypothese erforderlichen Wertes.

Tabelle 7: Anpassung der Funktion $F = A \cdot B^P$

A	B	FG	F	F'	P
1.12	1.10	1,122	332	6.86	$1.2 \cdot 10^{-11}$

Der Parameter B ist positiv und beschreibt die erwartete Tendenz: In je mehr Kontexten (Kotexten) ein Ausdruck verwendet werden kann, desto öfter wird er benutzt.

Gegenüber dem mathematisch überragenden Ergebnis ist der optische Eindruck der Anpassung etwas unbefriedigend. Die Graphik

62 Datentabelle s. Anhang A 3

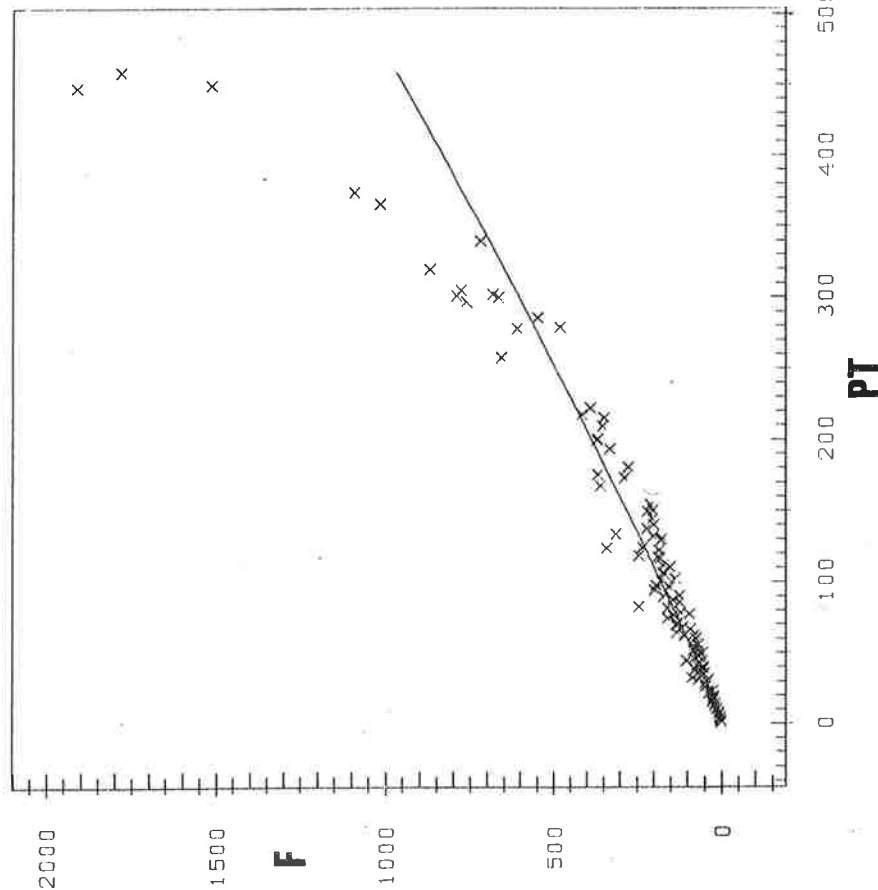


Abb. 21: Die Frequenz (F) als Funktion der Polytextie (PT)

läßt klar erkennen, daß eine Kurve mit größerer Steilheit die Daten noch weit besser beschreiben könnte. Eine solche geht aber aus dem Differentialgleichungsansatz der Modellkonstruktion nicht hervor.

Diese Sachlage läßt zwei Interpretationen zu. Entweder ist der Ansatz richtig; dann ist in den Daten des Limas - Korpus eine Verzerrung enthalten, die als untypisch gelten muß. Oder das Modell bedarf einer Ergänzung, weil ein Einfluß auf die Systemgröße Frequenz (z.B. das Alter der lexikalischen Einheiten) übersehen worden ist. Welche dieser beiden Interpretationen zutrifft, läßt sich erst entscheiden, wenn eine genügende Anzahl von Textkorpora in verschiedenen Sprachen untersucht worden ist.

Sollte die erste Vermutung zutreffen, so muß im Hinblick auf das Modell nichts weiter unternommen werden, da es ja dann als mit der Beobachtung verträglich gilt. Da sich aber auch die zweite Vermutung als richtig erweisen könnte, soll bereits hier eine mögliche Modellergänzung diskutiert werden.

Die Struktur des Modells beruht in diesem Punkt bisher auf der Annahme, daß die Veränderungsrate der Frequenz nur von der Polytextie abhängt. Diese Annahme wurde oben mit Hilfe der Differentialgleichung

$$\frac{y'}{y} = \frac{B}{x}$$

ausgedrückt. Die einfachste Zusatzannahme an dieser Stelle ist die eines konstanten Gliedes; d.h. daß die Frequenzänderung einer lexikalischen Einheit außer von der Polytextie noch von einer unbekannten konstanten Größe abhängt. Die Differentialgleichung lautet dann

$$\frac{y'}{y} = \frac{B}{x} + c,$$

und ihre Lösung führt zu der Funktion

$$y = A \cdot x^B \cdot e^{cx}$$

Die Anpassung dieser Funktion an die gleichen Daten wie oben brachte die Ergebnisse, die in Tabelle 8 und Abbildung 22 wiedergegeben sind³.

Tabelle 8: Anpassung der Funktion $F = A \cdot PT^B \cdot e^{cPT}$

A	B	c	FG	F	F	P
1.62	0.964	0.00193	2,121	824	4.79	$33 \cdot 10^{-71}$

Das Ergebnis ist sowohl mathematisch noch besser (im Vergleich zu der rein hyperbolischen Funktion) als auch optisch näher an den Daten.

Ob diese Modellerweiterung tatsächlich nötig ist, werden erst weitere empirische Studien zeigen. Entscheidend wird aber letztlich sein, ob sich eine plausible Interpretation für die Zusatzannahme finden läßt. Sie könnte in der Annahme einer Rückkopplung von der Länge auf die Häufigkeit bestehen: kürzere Ausdrücke werden vorgezogen, wenn möglich (vgl. Abschn. 4.6.2).

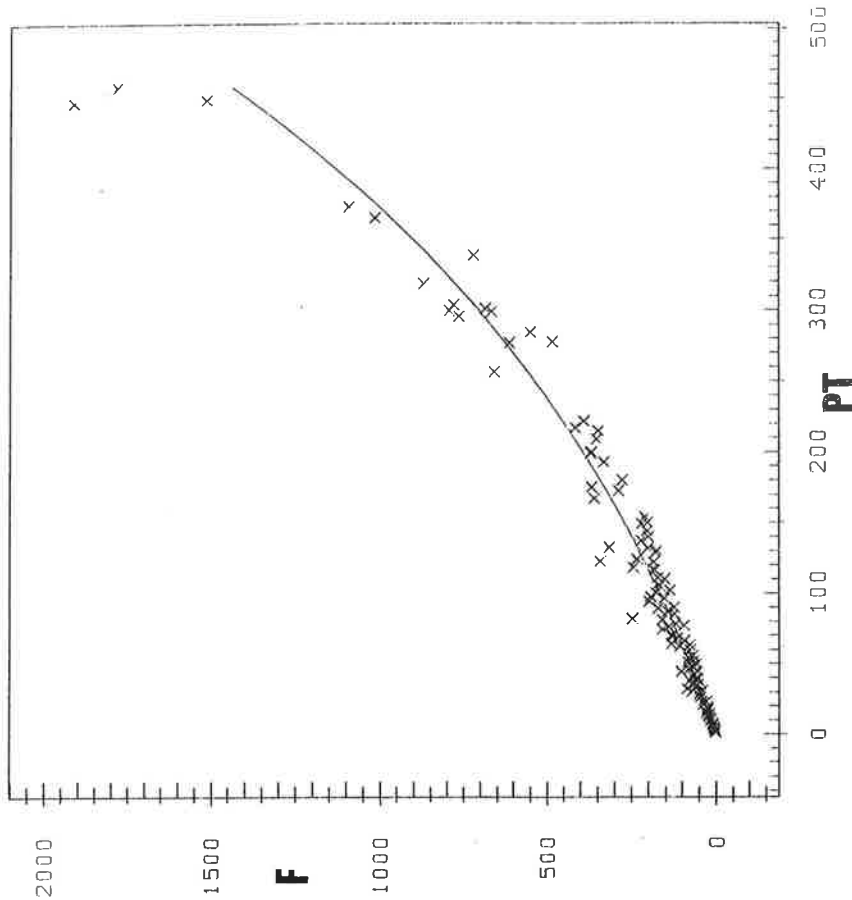


Abb. 22: Die Frequenz (F) als Funktion der Polytextie (PT); erweitertes Modell

6.3.4 Die Länge

Die Länge einer lexikalischen Einheit bestimmt sich aus zwei Faktoren - der globalen Komponente, die sich aus den Elementen Lexikongröße, Phonemzahl und dem Bedürfnis zur Übertragungsicherung Red zusammensetzt, und der individuellen Abweichung der betreffenden Einheit aufgrund ihrer Frequenz.

Die globale Komponente wird in der Funktionsgleichung durch die Konstante A repräsentiert. Der Parameter B, die Kürzungssteilheit, entspricht dem Einfluß der Frequenz auf die Länge.

Tabelle 9: Anpassung der Funktion $L = A F^B$

A	B	FG	F	F	P
10.9	-0.108	1,158	105	6.80	$2.2 \cdot 10^{-11}$

Die Anpassung an die Daten (der gleichen Stichprobe) ergibt wieder ein ausgezeichnetes Ergebnis: die F - Variable hat das Fünfzehnfache des erforderlichen Tabellenwertes (s. Tabelle 9). Auch der optische Eindruck der Abbildung 23 bestätigt, daß die theoretische Funktion die gemessene Tendenz gut erfäßt⁶⁴

Trotzdem auffällig ist allerdings die im Vergleich zu den anderen Anpassungen recht große Streuung der empirischen Daten-

64 Datentabelle s. Anhang A 5

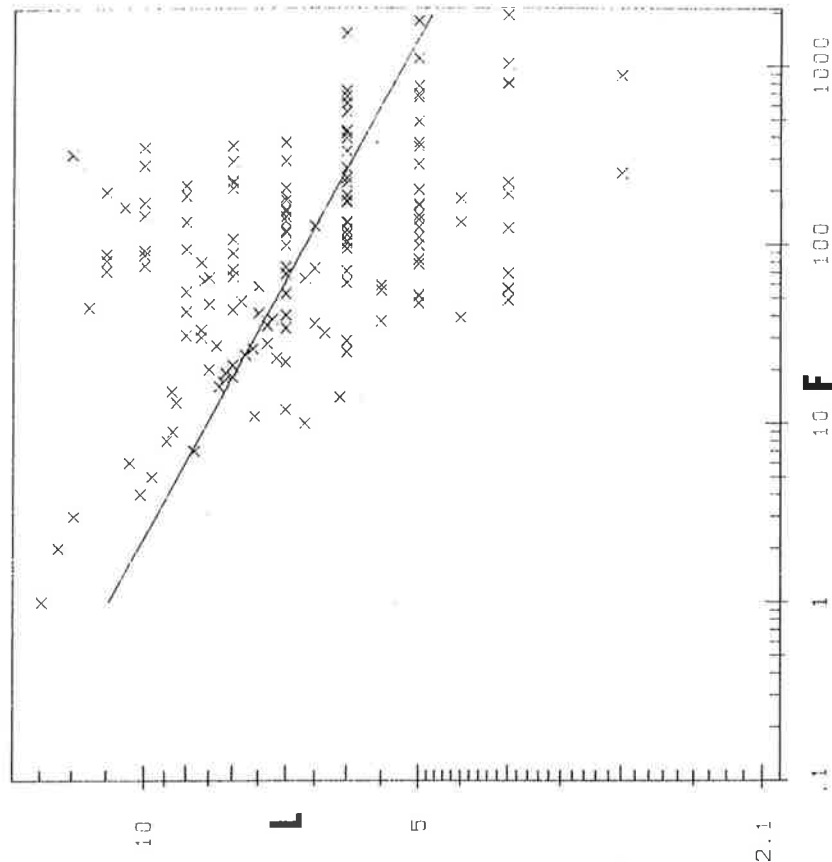


Abb. 23: Die Länge (L) als Funktion der Frequenz (F). Beide Achsen sind logarithmisch skaliert

punkte um den Graphen der Funktion. Dieser Eindruck ist zum Teil auf optische Verzerrung aufgrund der unterschiedlichen Skalen für F und L zurückzuführen. Zum anderen Teil aber spiegelt der Graph tatsächlich auch größere Abweichungen wieder. Darauf wird später noch zurückzukommen sein⁴⁵; vorläufig soll der Hinweis auf das doch außerordentlich gute Resultat des F - Tests genügen.

6.3.5 Überprüfung indirekter Abhängigkeiten

Einen noch wesentlich härteren Test für das Modell stellt die Einbeziehung mehrerer Variablen gleichzeitig durch Überprüfung von indirekten Abhängigkeiten dar. Wird nämlich nur die direkte Beziehung zweier Systemgrößen in Hypothesenform getestet, so läßt sich auch nur eine isolierte Aussage über den entsprechenden Strukturausschnitt des Modells treffen.

Daher wurden ergänzende Hypothesen aufgestellt, die Teilstrukturen mit den Verknüpfungen von je drei Systemgrößen betreffen, und schließlich eine solche, die den gesamten Modellkern, d.h. alle vier strukturellen Eigenschaften, in den Test einbeziehen. Zu deren Überprüfung wurden für die betroffenen Abhängigkeiten weitere empirische Daten aus den Stichproben erhoben.

Die zusätzlichen Hypothesen wurden, wie im Abschnitt 5 beschrieben, durch Verknüpfung von je zwei bzw. drei der oben behandelten primären Hypothesen gewonnen, indem die vermittelnde Größe durch Einsetzen der entsprechenden Funktion aus der Funktionsgleichung eliminiert wurde. Dabei wurden die Konstanten A und B aus denen der Ausgangsfunktionen berechnet (In Tabelle 10 sind diese Ausgangswerte zusammengefaßt).

65 vgl. Abschn. 7

Tabelle 10: Die empirischen Konstanten der vier direkten Abhängigkeiten

	L(F)	PL(L)	PT(PL)	F(PT)
A	10.9	12.5	3.92	1.12
B	-0.108	-0.828	1.56	1.1

Wenn auch davon auszugehen ist, daß die empirischen Datenpunkte aus der Messung solcher indirekter Abhängigkeiten eine beträchtlich größere Streuung aufweisen als die der direkten, sollte doch in jedem Fall eine ausreichend gute Anpassung der Funktion $y = A x^B$ möglich sein. Nun darf zwar nicht erwartet werden, daß die aus den Daten geschätzten Parameter den errechneten Werten sehr nahe kommen, wohl aber die Übereinstimmung von Vorzeichen und Größenordnung. Ein negatives Resultat in dieser Hinsicht würde dafür sprechen, daß an der betreffenden Stelle in der Modellstruktur ein weiterer, bisher nicht berücksichtigter Faktor auf eine der Größen einwirkt.

Nachstehend sollen kurz die Resultate dieses Versuches dargestellt werden.

1. Zur Aufstellung der theoretischen Funktionsgleichung für die Abhängigkeit der Polylexie von der Frequenz werden die ersten beiden direkten Abhängigkeiten miteinander verknüpft. Mit

$$L = 10.9 F^{-0.108} \quad \text{und}$$

$$PL = 12.5 L^{-0.828}$$

ergibt sich durch Einsetzen

$$PL = 12.5 (10.9 F^{-0.108} - 0.829)$$

und damit für diese Beziehung die Voraussage

$$A_v = 1.73 \text{ und } B_v = 0.09$$

Die Anpassung der Funktion (vgl. Tabelle 11 und Abbildung 24)⁶⁶ ergab die aus den Daten geschätzten Parameter

$$A_0 = 1.03 \text{ und } B_0 = 0.283,$$

bei wieder ausgezeichnete Anpassungsgüte.

Wie aus der Abbildung ersichtlich, scheint in den Daten eine Trendaufspaltung in drei separate Teilfunktionen enthalten zu sein, die auf einem unterschiedlichen Verhalten der Wortarten beruhen könnten. Eine getrennte Behandlung der Teilfunktionen würde vermutlich zu einem noch besseren Ergebnis führen (vgl. Abschnitt 6.3.6.)

Überraschend ist die Übereinstimmung der berechneten und der empirischen Konstantenwerte. Es liegt daher nahe, zu testen, ob sich die Parameter überhaupt signifikant unterscheiden. Wenn dies nicht der Fall ist, kann die Vorhersage als korrekt angesehen werden.

Interessant ist hier weniger der Parameter A, der lediglich eine Parallelverschiebung der Regressionsgeraden⁶⁷ anzeigt, sondern vor allem die Differenz von B_v und B₀. Die Nullhypothese

$$H_0 : B_v - B_0 = 0$$

66 Datentabelle s. Anhang A 6

67 Die Anpassung wurden nach Logarithmierung mit Hilfe der linearen Regression durchgeführt.

Tabelle 11: Anpassung der Funktion $PL = A F^B$

A	B	FG	F	P
1.03	0.283	1,158	121	6.80
				2.2*10 ⁻¹¹

steht für die Annahme einer korrekten Vorhersage. Sie muß abgelehnt werden, wenn der Wert von

$$t_{2n-4} = \frac{|B_v - B_0|}{S \sqrt{\frac{1}{\sum (x_{vi} - \bar{x}_v)^2} + \frac{1}{\sum (x_{0i} - \bar{x}_0)^2}}}$$

$$\text{mit } S^2 = \frac{(n-2) s_b^2 + (n-2) s_v^2}{2n - 4}$$

größer als der Wert der t-Verteilung mit 2n-4 Freiheitsgraden ist. s_b^2 und s_v^2 sind die Varianzen, x_{sv} und x_{0v} die Mittelwerte der mit B_v bzw. B₀ theoretisch berechneten Werte x_{1sv} bzw. x_{10v} . In diesem Fall ergab sich ein Wert von $t = 0.079$ bei 316 Freiheitsgraden gegenüber einem kritischen Wert von 2.35 auf dem Signifikanzniveau 0.01. Die Nullhypothese braucht also nicht abgelehnt zu werden, und die theoretische Vorhersage des Parameterwerts gilt als bestätigt.

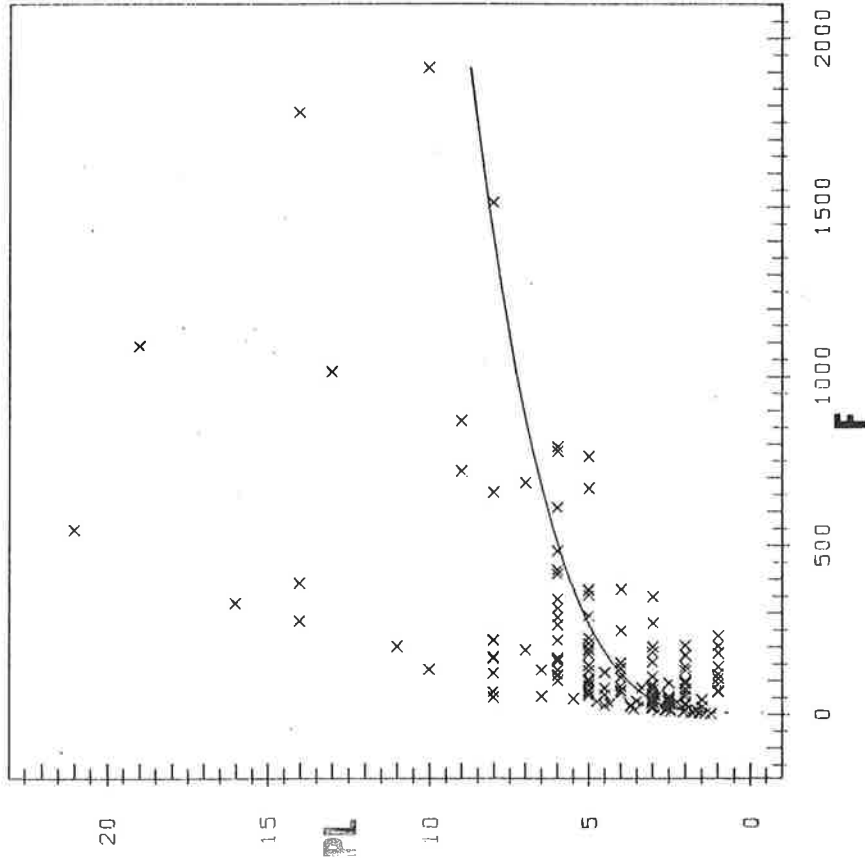


Abbildung 24: Die indirekte Abhängigkeit der Polylexie (PL) von der Frequenz (F)

2. Die indirekte Abhängigkeit der Polytextie von der Länge berechnet sich analog. Aus

$$PL = 12.5 L^{-0.828} \quad \text{und}$$

$$PT = 3.92 PL^{1.56}$$

erhält man durch Einsetzen die Funktion

$$PT = 3.92 (12.5 L^{-0.828})^{1.56}$$

und damit die theoretischen Werte $A_v = 201.6$ und $B_v = -1.29$. Die Anpassung der Hyperbel an die Stichprobendaten⁶⁸ erbrachte bei einem gutem F -Wert (vgl. Tabelle 12 und Abbildung 25) die empirische Schätzung $A_0 = 1710$ und $B_0 = -2.39$.

Tabelle 12: Anpassung der Funktion $PT = AL^B$

A	B	FG	F	P
1710	-2.39	1,22	17.8	7.95 0.00035

Während der Parameter B sehr genau erreicht wurde, ist der Wert für A_0 viel zu hoch, um mit dem errechneten A_v identifiziert zu werden. Dennoch ist die Abweichung nicht sonderlich beunruhig-

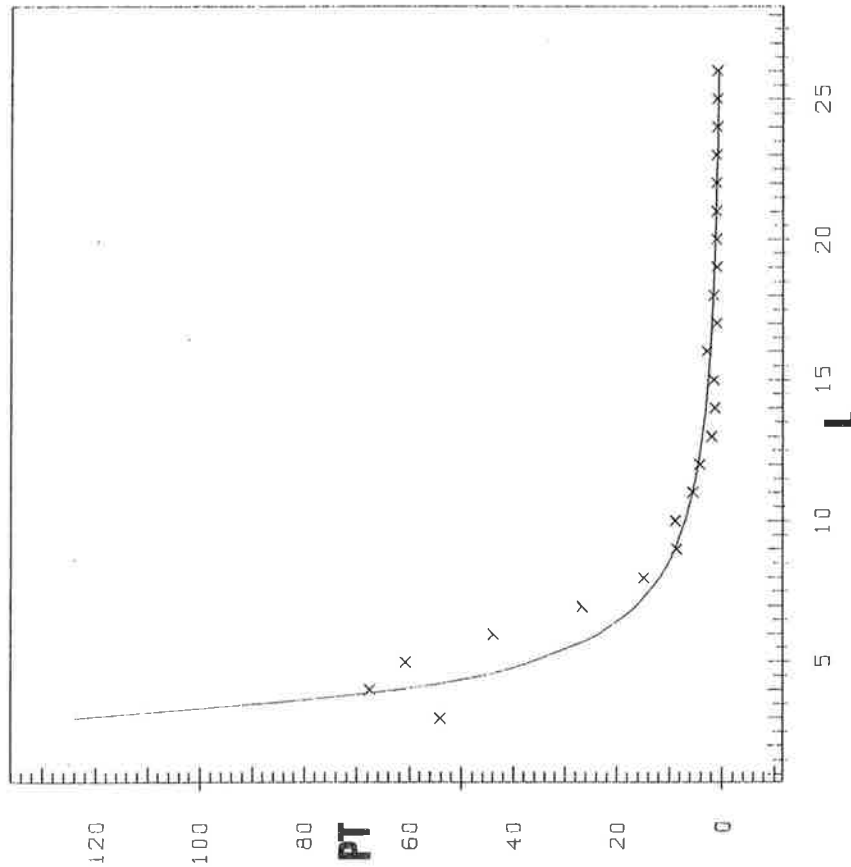


Abb. 25: Die indirekte Abhängigkeit der Polytextie (PT) von der Länge (L)

gend, wenn man die Bedeutung dieses Parameters berücksichtigt: Er gibt die Polytextie von Wörtern der Länge 1 an, d. h. er bestimmt den Kurvenverlauf vor allem im Bereich kleiner Längen. Es gibt aber nicht ein einziges Lemma der Länge 1, so daß hier mit besonders großen Meßfehlern gerechnet werden muß. Der Test der Differenz $B_v - B_0$ wurde mit dem t-Test wie oben vorgenommen und ergab einen Wert von $t = 0.6502$ bei 44 Freiheitsgraden gegenüber einem Tabellenwert von 2.41 auf dem 0.01-Niveau. Auch hier muß die Vorhersage des Parameterwerts als korrekt betrachtet werden.

3. Die dritte indirekte Beziehung besteht theoretisch zwischen der Frequenz und der Polylexie. Wie oben erhält man durch Einsetzen aus

$$PT = 3.92 PL^{1.56} \text{ und}$$

$$F = 1.12 PT^{1.1}$$

die Gleichung

$$F = 1.12 (3.92 PL^{1.56})^{1.1}$$

was zu den theoretischen Parametern $A_v = 5.033$ und $B_v = 1.716$ führt. In diesem Fall erbrachte die Anpassung (vgl. Tabelle 13 und Abbildung 26)⁴⁹ geschätzte Parameterwerte, die mit den vorhergesagten numerisch nahezu gleich sind. Der Test der Parameterdifferenz erbrachte auch einen entsprechend minimalen Wert von $t = 0.0311$ mit 28 Freiheitsgraden bei einem theoretischen $t_{0.01}$ von 2.47.

4. Schließlich bleibt der Zusammenhang zwischen Länge und Polytextie. Die Ausgangsgleichungen

69 Datentabelle s. Anhang A 8

Tabelle 13: Anpassung der Funktion $F = A \cdot PL^B$

A	B	FG	F	F	P
5.50	1.76	1,14	11.2	8.86	0.0048

$$F = 1.12 \cdot PL^{1.1} \quad \text{und}$$

$$L = 10.9 \cdot F^{-0.108}$$

ergeben

$$L = 10.9 \cdot (1.12 \cdot PL^{1.1})^{-0.108}$$

woraus $A_v = 10.77$ und $B_v = -0.12$ bestimmt werden. Wiederum ergibt sich aus der Anpassung (vgl. Tabelle 14 und Abbildung 27) eine außerordentliche numerische Übereinstimmung mit den geschätzten Werten. Der t-Wert ist noch geringer als oben, nämlich 0.0056 (996 Freiheitsgrade), während der kritische Wert auf dem 0.01-Niveau 2.33 beträgt.

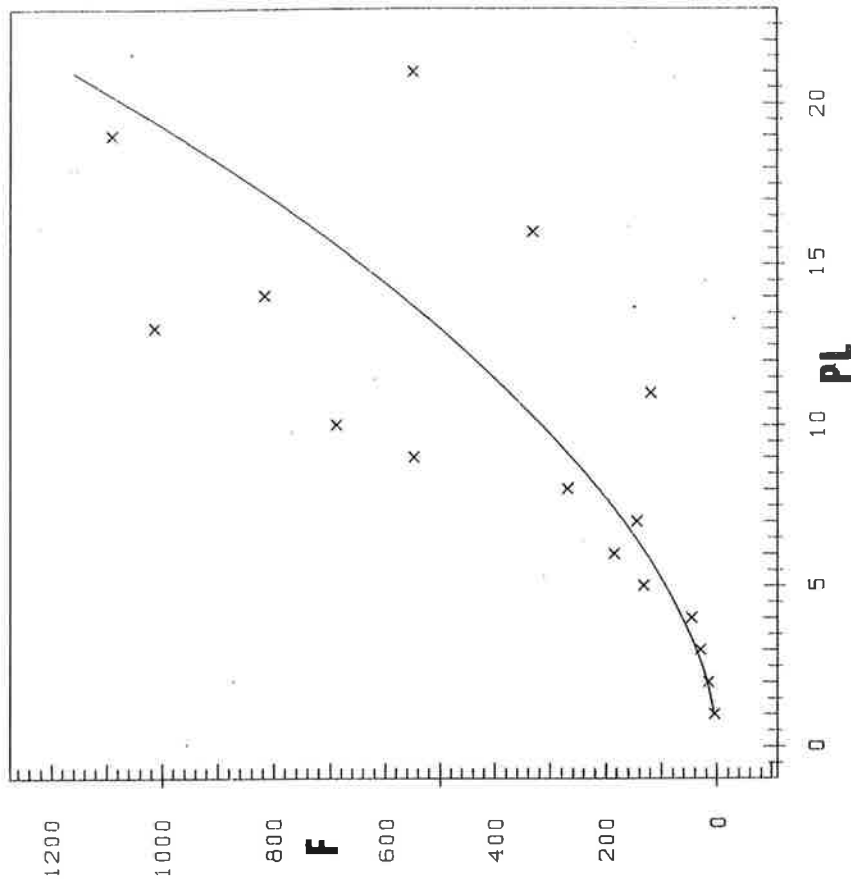


Abb. 26: Die indirekte Abhängigkeit der Frequenz (F) von der Polylexie (PL)

Tabelle 14: Anpassung der Funktion $L = A \cdot P^B$

A	B	FG	F	F	P
				0.01	
11.5	-0.112	1,180	110	6.77	2*10 ⁻¹¹

5. Aufgrund dieser positiven Resultate wurde schließlich noch der Versuch unternommen, die Struktur des gesamten Modellkerns mit einer indirekten Abhängigkeit über alle vier Modellvariablen zu testen.

Analog zu oben wurde durch Einsetzen mit

$$F = 1.12 \cdot P^{1.1}$$

$$L = 10.9 \cdot (1.12 \cdot P^{1.1})^{-0.108} \quad \text{und}$$

$$PL = 12.5 \cdot (10.767 \cdot L^{-0.1188})^{-0.828}$$

eine theoretische Vorhersage für die Funktionsparameter berechnet. Daraus ergab sich $A_v = 1.747$ und $B_v = 0.10$. Trotz des "weiten Weges" zwischen der unabhängigen und der abhängigen Variablen ist die Übereinstimmung dieser mit den aus der Anpassung (vgl. Tabelle 15 und Abbildung 28) stammenden Werten über-

71 Datentabelle s. Anhang A 10

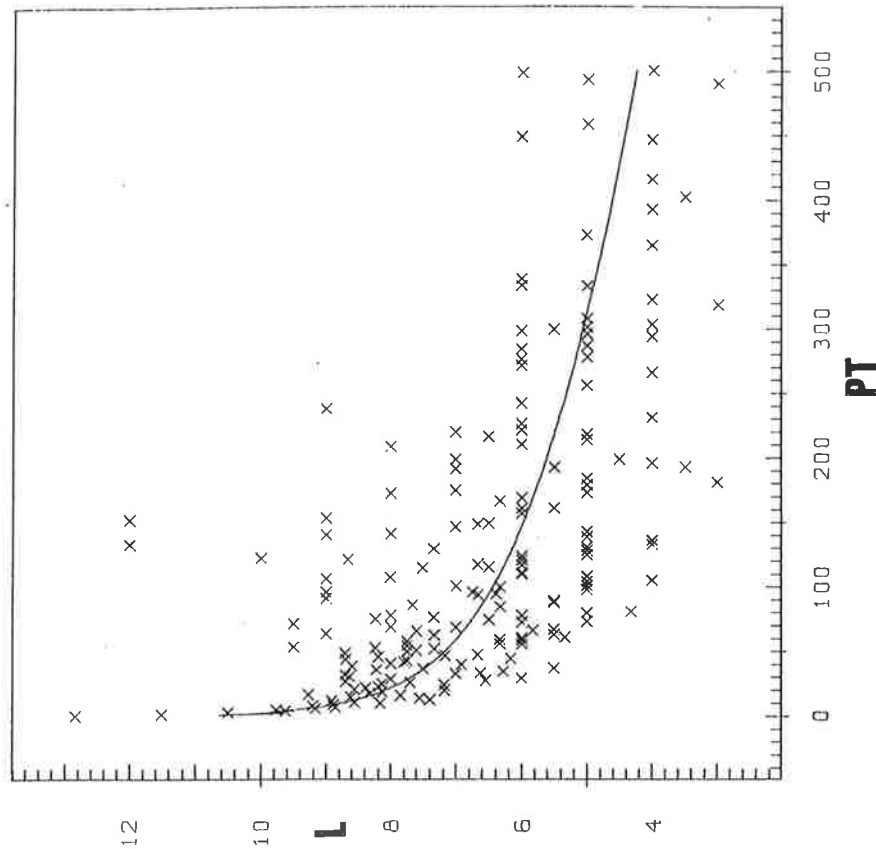


Abb. 27: Die indirekte Abhängigkeit der Länge (L) von der der Polytextile (PT)

raschend gut - mehr als eine Übereinstimmung in der Größenordnung war nicht zu erwarten, es ergab sich aber auch hier annähernd numerische Gleichheit. Tatsächlich muß auch aufgrund des Tests der Parameterdifferenz ($t = 0.0868$ mit 244 Freiheitsgraden gegenüber dem Wert der t-Verteilung von 2.35 auf dem Signifikanzniveau 0.01) von der Gleichheit des vorgesehenen und des empirischen Parameters ausgegangen werden.

Tabelle 15: Die Anpassung der Funktion $PL = A \cdot PT^B$

A	B	FG	F	P
			0.01	
1.11	0.31	1,122	88.8	6.86
				$2.1 \cdot 10^{-11}$

Insgesamt spricht auch diese zusätzliche Auswertung der indirekten Abhängigkeiten für die Richtigkeit der Modellkonstruktion; sie läßt sogar (vorläufig) darauf schließen, daß das modellierte Subsystem relative Autonomie besitzt.

6.3.6 Vergleich der Operationalisierungsmöglichkeiten

Alle bisher dargestellten Tests beruhten auf Stichprobendaten, die aus solchen Messungen stammen, bei denen als Beobachtungs-korrelat zum theoretischen Begriff der lexikalischen Einheit das Lemma der im Text gefundenen Wortform diente. Zur Kontrolle

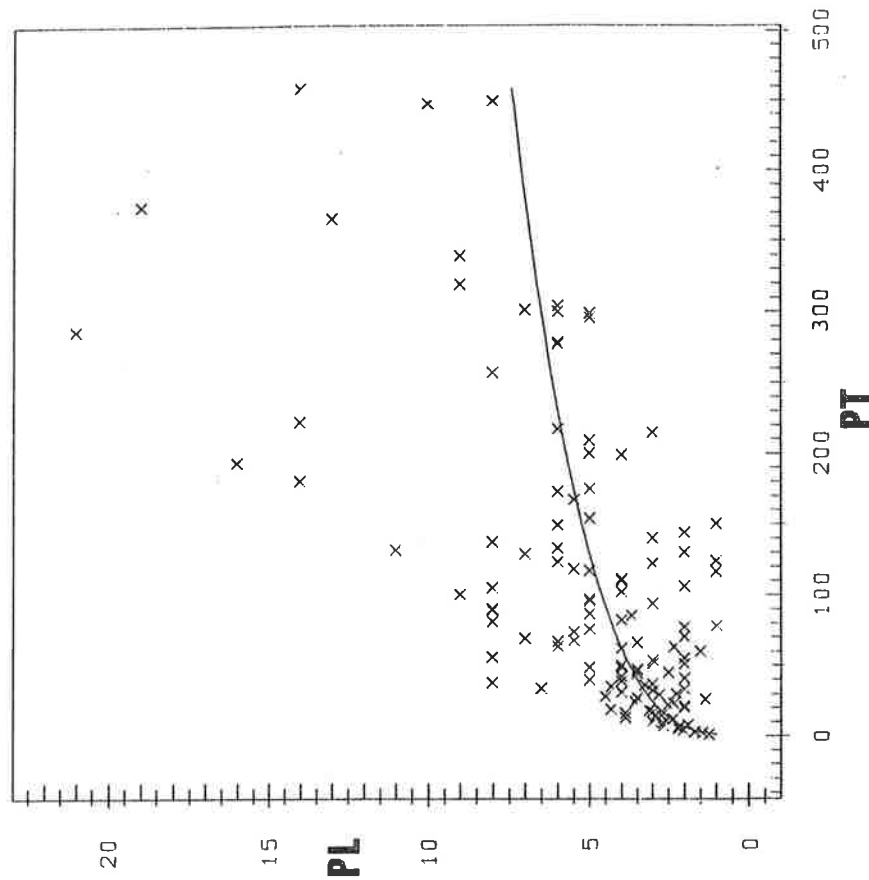


Abb. 28: Die doppelt indirekte Abhängigkeit der Polylexie (PL) von der Polytextie (PT)

wurden zusätzliche, unabhängige Tests mit Daten durchgeführt, die auf der Einheit Wortform beruhen. Möglich war dies nur in bezug auf diejenigen Variablenzusammenhänge, bei denen die Polylexie keine Rolle spielt, also bei der Verknüpfung der Länge mit der Frequenz und der Verknüpfung der Frequenz mit der Polytextie.

Da Verzerrungen, die durch eine ungeeignete Operationalisierung entstehen könnten, vermutlich wortartenspezifisch wären², wurde außerdem nach der Zugehörigkeit der Wörter zu einer der offenen Wortklassen unterschieden.

Tabelle 16 und Abbildung 29 zeigen die Resultate der Anpassungen der Funktion $y = A x^b e^c$ an acht Teilstichproben für die Abhängigkeit der Frequenz von der Polytextie³.

In keinem Fall durfte die Hypothese abgelehnt werden; vielmehr sind fast sämtliche Testergebnisse überragend gut. Darüber hinaus läßt sich feststellen, daß sogar die Parameterschätzungen sehr dicht beieinander liegen.

Offensichtlich unterscheiden sich die lexikalischen Einheiten (ob Lemma oder Wortart) in Abhängigkeit von der Wortklassenzugehörigkeit in ihrem Abhängigkeitsverhalten in bezug auf die Systemgrößen. Vor einer Einbeziehung dieser Unterscheidung in das Modell müßten die Differenzen der Parameter jedoch ebenfalls noch auf Signifikanz getestet werden.

72 Der quantitative Unterschied bei der Behandlung von Wortarten bzw. Lemmata hängt vor allem vom Formenreichtum der betreffenden lexikalischen Einheit ab.

73 Datentabellen S. Anhänge A 11 bis A 18

Tabelle 16: Anpassung der Funktion $F = A P^B C P^T$ an acht verschiedene Teilstichproben

	Verben	Nomina	Adjektive	Gesamt
Lemmata	A=1.25 B=0.96 C=0.00234 FG: 2,120 F=508 F =4.8 0.01 P=2.68*10 ⁻⁵⁹	A=1.53 B=1.03 C=0.000705 FG: 2,98 F=237 F =4.82 0.01 P=2.87*10 ⁻³⁸	A=1.47 B=0.96 C=0.00234 FG: 2,120 F=508 F =4.79 0.01 P=3.48*10 ⁻⁸¹	A=2.16 B=0.848 C=0.00391 FG: 2,180 F=66.1 F =4.72 0.01 P=2.99*10 ⁻²²
Wortformen	A=1.46 B=0.982 C=0.00394 FG: 2,94 F=48.3 F =4.85 0.01 P=3.38*10 ⁻¹⁵	A=1.35 B=1.12 C=0.00215 FG: 2,82 F=215 F =4.88 0.01 P=2.43*10 ⁻³³	A=1.52 B=1.02 C=0.000838 FG: 2,109 F=242 F =4.80 0.01 P=8.09*10 ⁻⁴¹	A=1.77 B=1.02 C=0.0015 FG: 2,157 F=56.1 F =4.74 0.01 P=4.14*10 ⁻¹¹

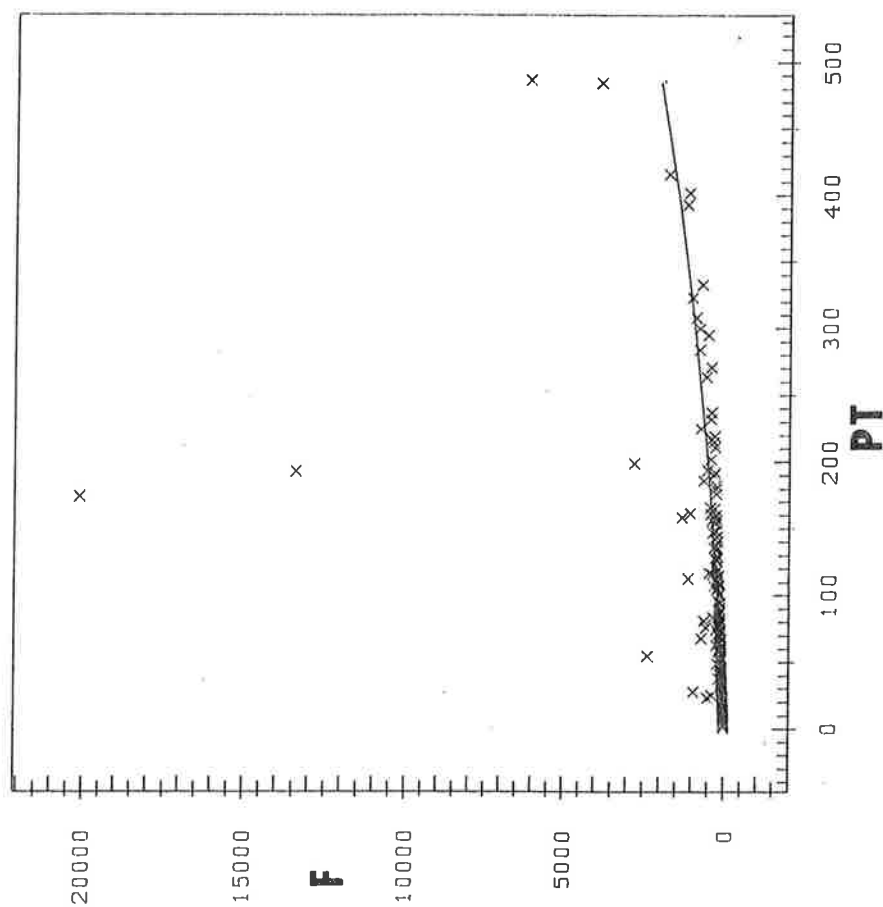


Abb. 29 a: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Wortformen gesamt

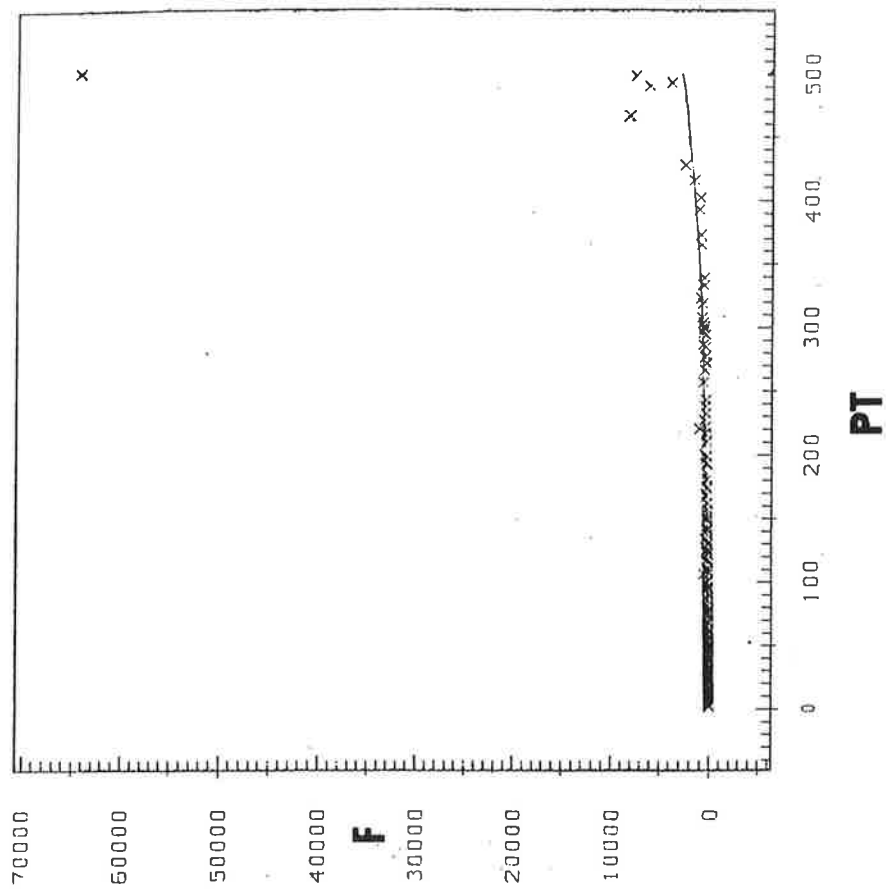


Abb. 29 b: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Lemmata gesamt

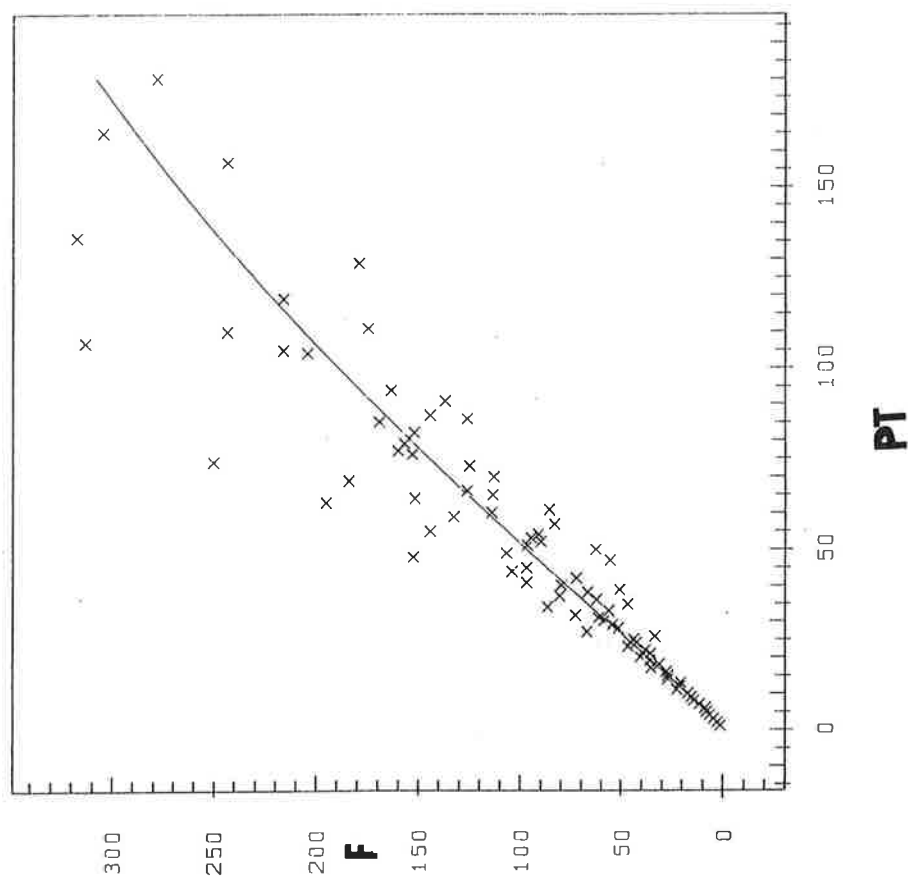


Abb. 29 c: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Wortformen Nomina

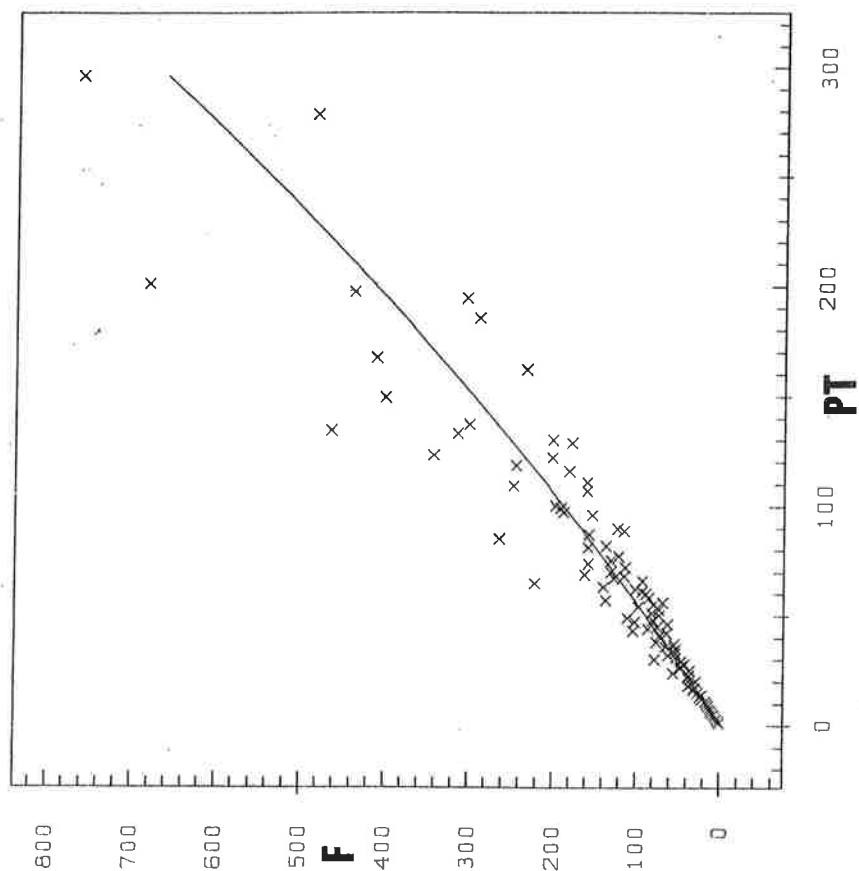


Abb. 29 d: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Lemmata Nomina

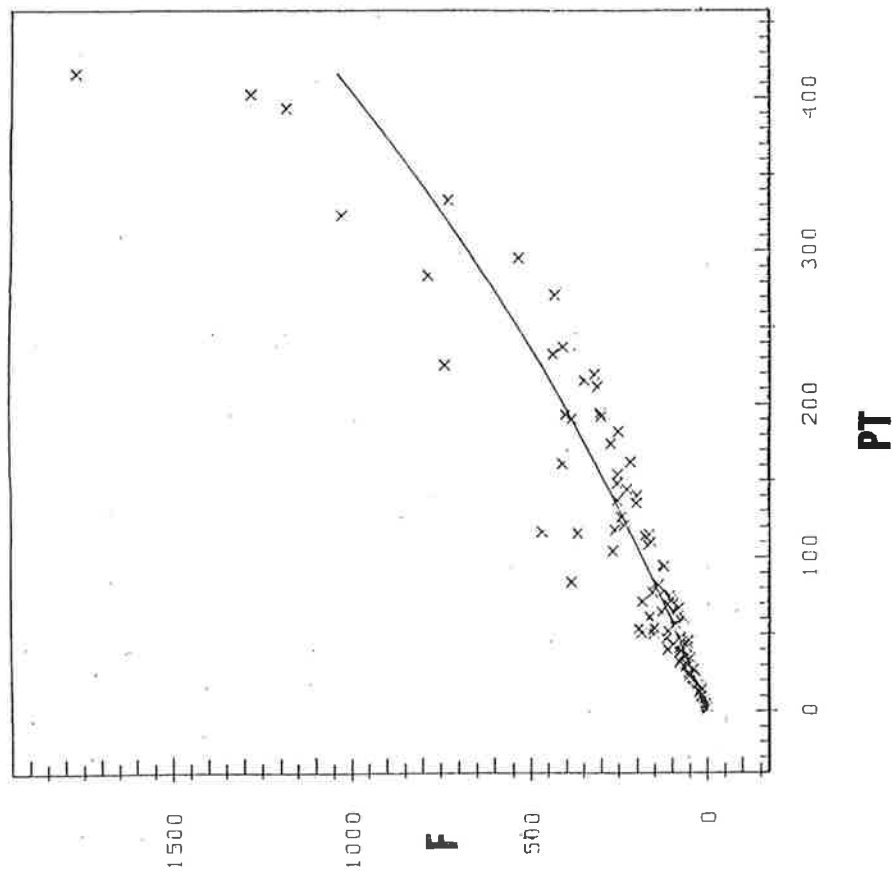


Abb. 29 e: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Wortformen Adjektive

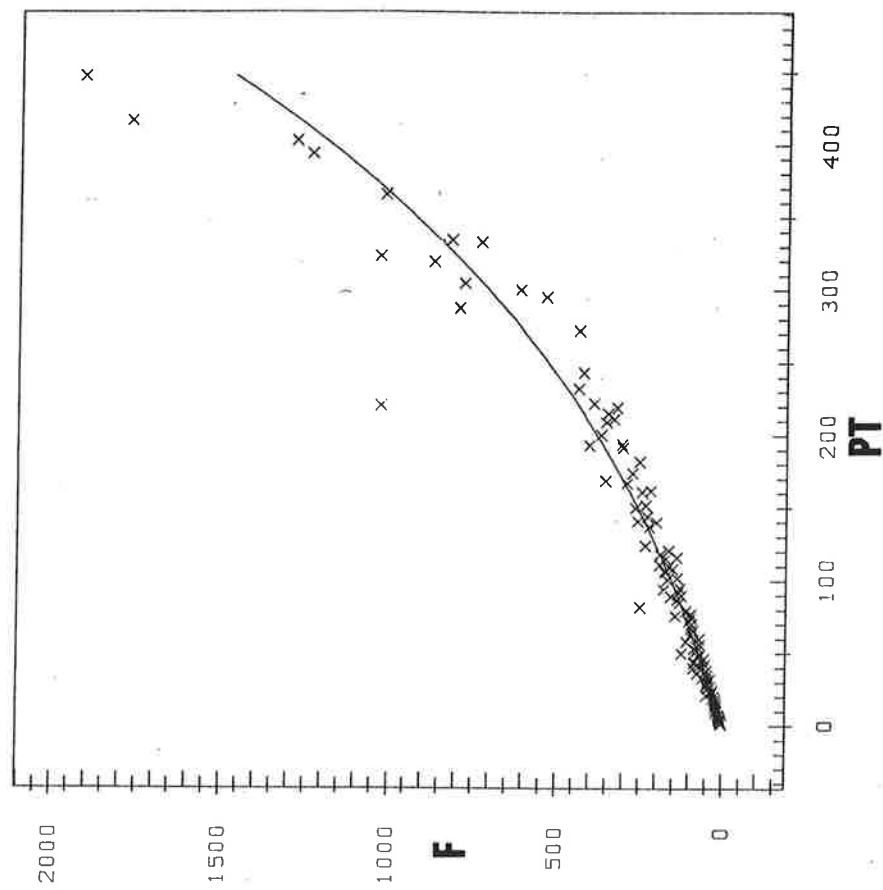


Abb. 29 f: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Lemmata Adjektive

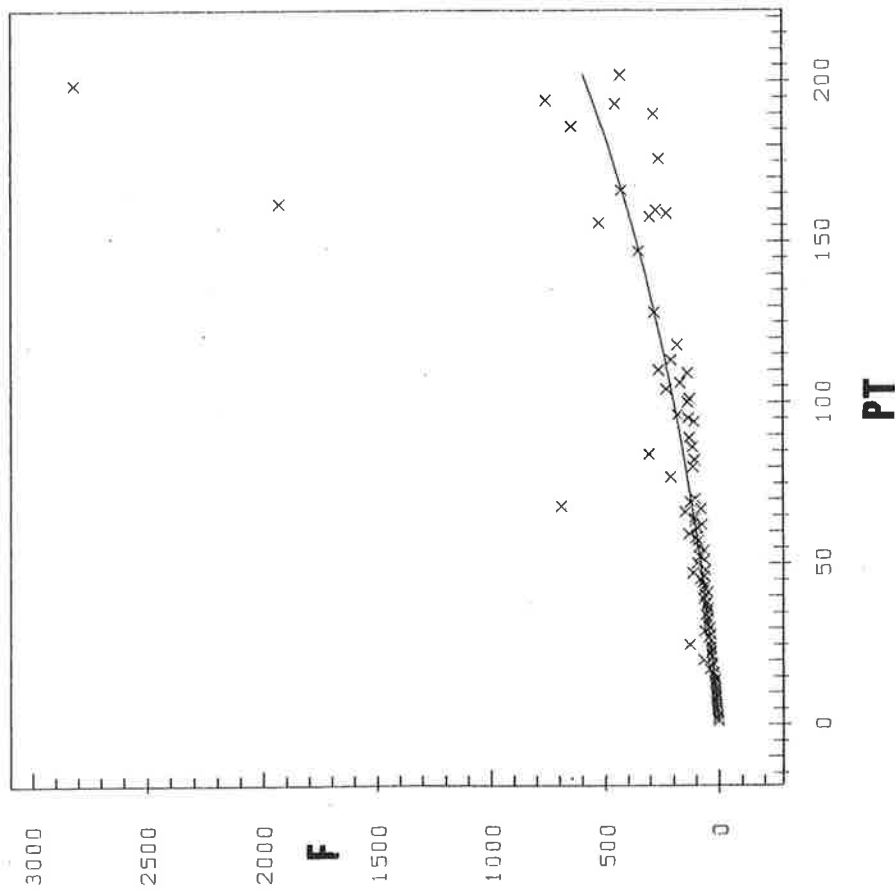


Abb. 29 g: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Wortformen Verben

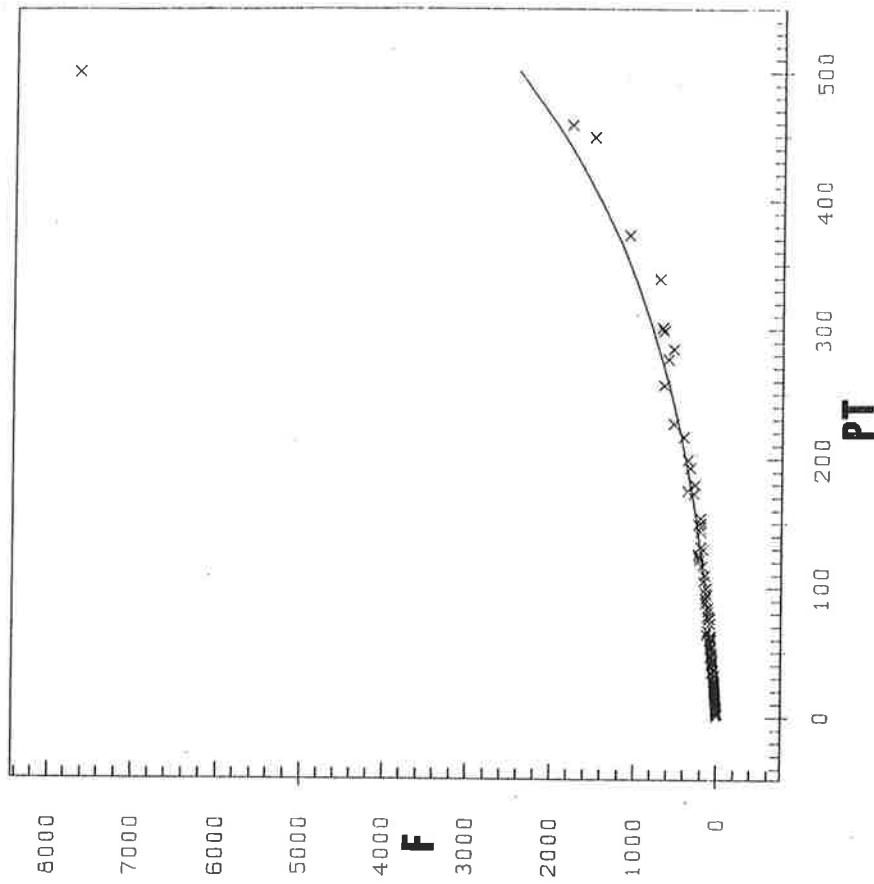


Abb. 29 h: Die Abhängigkeit der Frequenz (F) von der Polytextie (PT): Lemmata Verben

6.3.7 Zusammenfassung der Testergebnisse

Sämtliche Tests, einschließlich der Überprüfung von Zusatzhypothesen und der Verwendung von Kontrollstichproben, haben eindeutig die Verträglichkeit des Modells mit den untersuchten empirischen Daten erwiesen. Die Entscheidungen über die Annahme der Hypothesen erfolgten auf dem relativ hohen Signifikanzniveau von $P = 0.01$.

Die gefundenen Werte der Testvariablen überstiegen die Tabellenwerte in fast allen Fällen um mehrere Größenordnungen; die Irrtumswahrscheinlichkeit ist damit äußerst gering. In allen Fällen stimmten die theoretische Vorhersage der Parameterwerte mit den aus den Daten berechneten auch numerisch überein.

Ein besonderer Erfolg war die Tatsache, daß die aus der Modellstruktur deduzierten Variablenzusammenhänge sogar im Fall der doppelte indirekten Abhängigkeit mit vier beteiligten Systemgrößen eine ebenso gute Bestätigung erhielten.

Im Fall der Abhängigkeit der Frequenz von der Polytextie liegt ein Indiz vor, daß an dieser Stelle eine weitere, bisher unbekannte Einwirkung vorliegen könnte. Diese Frage kann aber erst aufgrund einer wesentlich größeren Datenbasis entschieden werden, wenn keine theoretischen Notwendigkeit dazu gefunden wird, eine solche neue Größe anzunehmen.

Die Überprüfung des Modells an dem erhobenen Datenmaterial ergibt eine eindeutige Bestätigung. Für größere Evidenz und eine Aussage über die Reichweite bedarf es noch einer größeren Zahl von verschiedenen Studien, insbesondere für andere Sprachen.

7. DIE OSZILLATION DER LEXIK

Bei der Anpassung der Funktion $y = A x^B$ an die empirischen Daten zur Abhängigkeit der Länge von der Frequenz ergab sich ein etwas unerwartetes Bild: die Güte der Anpassung steht nach dem F-Test außer Frage, und dennoch scheinen die Datenpunkte im Vergleich zu den anderen Anpassungen relativ stark um die theoretische Funktion zu streuen. Die Abbildung 30 zeigt die Daten noch einmal in nichtlogarithmischer Form. Die Tendenz ist unverkennbar, und das Resultat des statistischen Tests weist die erklärte Varianz gegenüber der unerklärten als so groß aus, daß die Hypothese nicht verworfen werden darf. Trotzdem wäre es möglich, daß sich hinter der scheinbar zufälligen Streuung ein systematischer methodischer Fehler oder ein bisher nicht berücksichtigter systematischer Einfluß verbirgt. Im ersten Fall müßte die gesamte Untersuchung nach der Korrektur des Fehlers wiederholt werden. Der zweite Fall wäre interessant, weil eine solche "Anomalie" entweder zu einer Verbesserung des Modells führen muß (falls eine entsprechende Erweiterung oder Modifikation die Anomalie erklären kann), oder aber die Unverträglichkeit des Modells mit den Beobachtungen nachweist.

Wenn die Abweichungen der Datenpunkte von der theoretischen Kurve zufällig sind, so sollte sich ihr Verlauf mit Hilfe der Methode der gleitenden Mittelwerte glätten lassen. Daher wurden aus der gleichen Stichprobe Daten generiert, deren Werte für die Länge sich aus den Mittelwerten eines gleitenden Intervalls der Länge 20 auf der Frequenz - Achse ergaben.

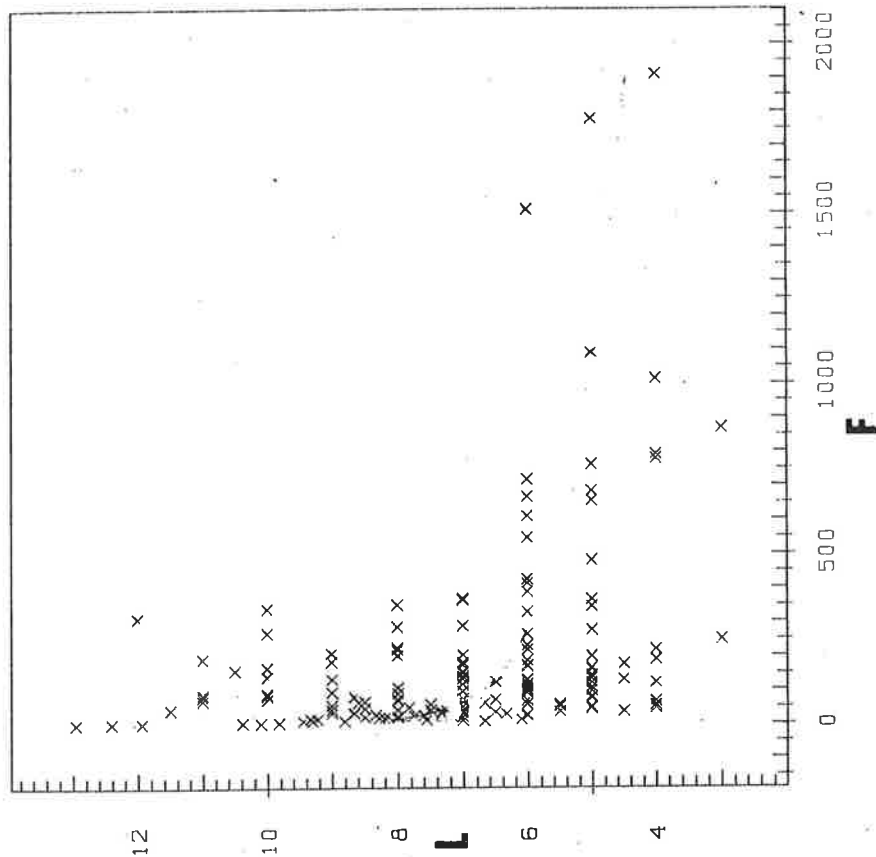


Abbildung 30: Die Länge (L) in Abhängigkeit von der Frequenz (F)

Anstelle der erwarteten schlichten Annäherung der Punkte an die Kurve zeigte sich ein überraschendes Bild (vgl. Abbildung 31): Der geglättete Verlauf der Datenpunkte ergab eine die theoretische hyperbolische Funktion schwingende Kurve; die Anpassung (s. Tabelle 17) der aus dem Modell abgeleiteten Funktion war wieder ausgezeichnet.

Tabelle 17: Die Anpassung der Funktion $L = A F^B$ bei gleitenden Mittelwerten (Intervall 20)

A	B	FG	F	P
9.36	-0.066	1,138	138	$6.83 \cdot 10^{-11}$

Zur Überprüfung der Stabilität dieses Phänomens wurde die Anpassung mehrfach unter Verwendung unterschiedlicher Intervallgrößen wiederholt (vgl. die Abbildung 32 und 33 sowie die Tabellen 18 und 19). Die beobachtete Oszillation um die vorhergesagte Kurve stellte sich jedesmal mit unterschiedlicher Auflösung, aber unverkennbarer Deutlichkeit ein. Ein zufälliges Zustandekommen dieser Erscheinung kann nahezu mit Sicherheit ausgeschlossen werden.

Das stellt uns vor die Aufgabe, eine sprachwissenschaftliche Erklärung für diese Beobachtung zu finden. Auf den ersten Blick scheint es keinen bekannten Zusammenhang zu geben, der einen Schwingungsvorgang in der Lexik einer Sprache verursachen könnte, zumal die Oszillation nicht in der Zeit-Dimension, sondern an der Frequenzachse festgestellt wurde.

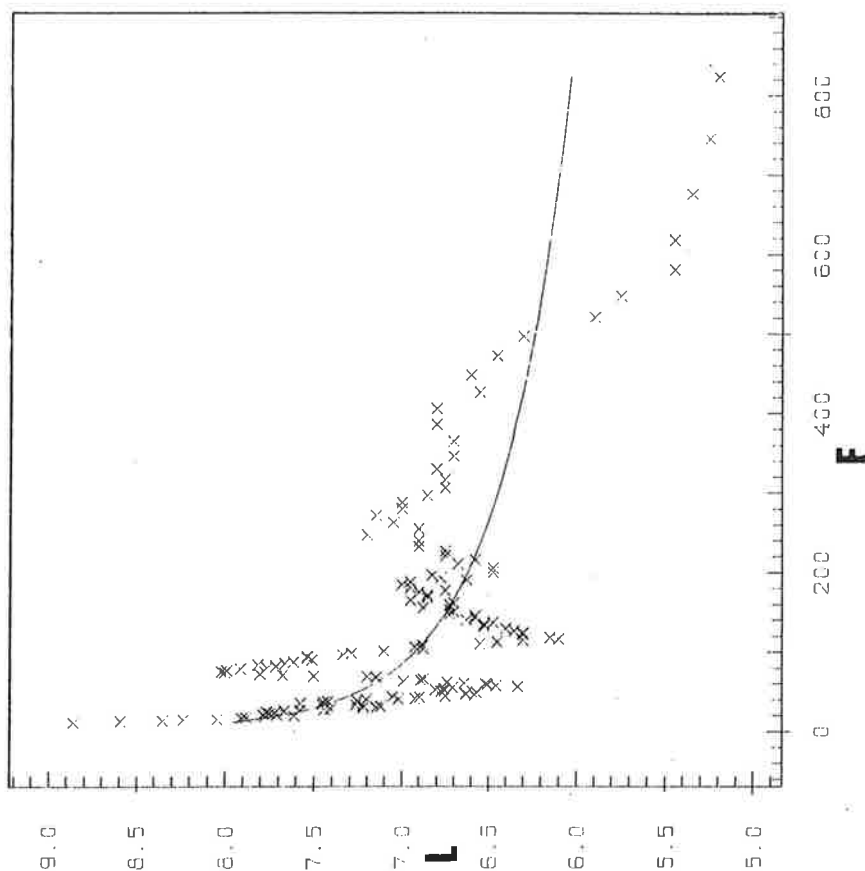


Abbildung 31: Die Länge (L) als Funktion der Frequenz (F).
Berechnet mit Hilfe gleitender Mittel über
ein Intervall von 20

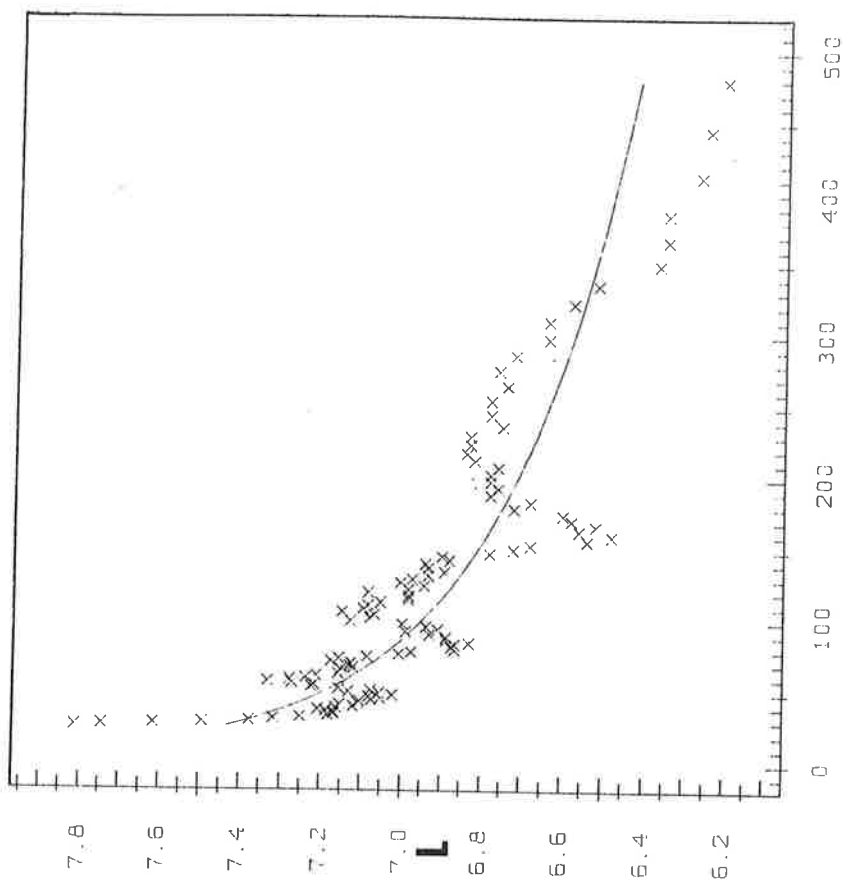


Abbildung 32: Die Länge (L) als Funktion der Frequenz (F).
Berechnet mit Hilfe gleitender Mittel über
ein Intervall von 50

Tabelle 18: Die Anpassung der Funktion $L = A F^B$ bei gleitenden Mittelwerten (Intervall 50)

A	B	FG	F	F	P
8.76	-0.05	1,108	238	6.75	$1.5 \cdot 10^{-11}$

Tabelle 19: Die Anpassung der Funktion $L = A F^B$ bei gleitenden Mittelwerten (Intervall 100)

A	B	FG	F	F	P
8.91	-0.052	1,58	276	6.75	$7 \cdot 10^{-12}$

An dieser Stelle kommen wir auf eine Überlegung zurück, die bei der Konstruktion des Modells anstellt, dann aber zunächst nicht weiter verfolgt wurde: Die Modellstruktur für den Zusammenhang zwischen Länge und Frequenz beruht auf der Annahme, daß die Veränderungsrate der Länge umgekehrt proportional zur Frequenz einer lexikalischen Einheit ist. Diese Annahme haben wir in Form der Differentialgleichung

$$\frac{y'}{y} = \frac{-N}{x}$$

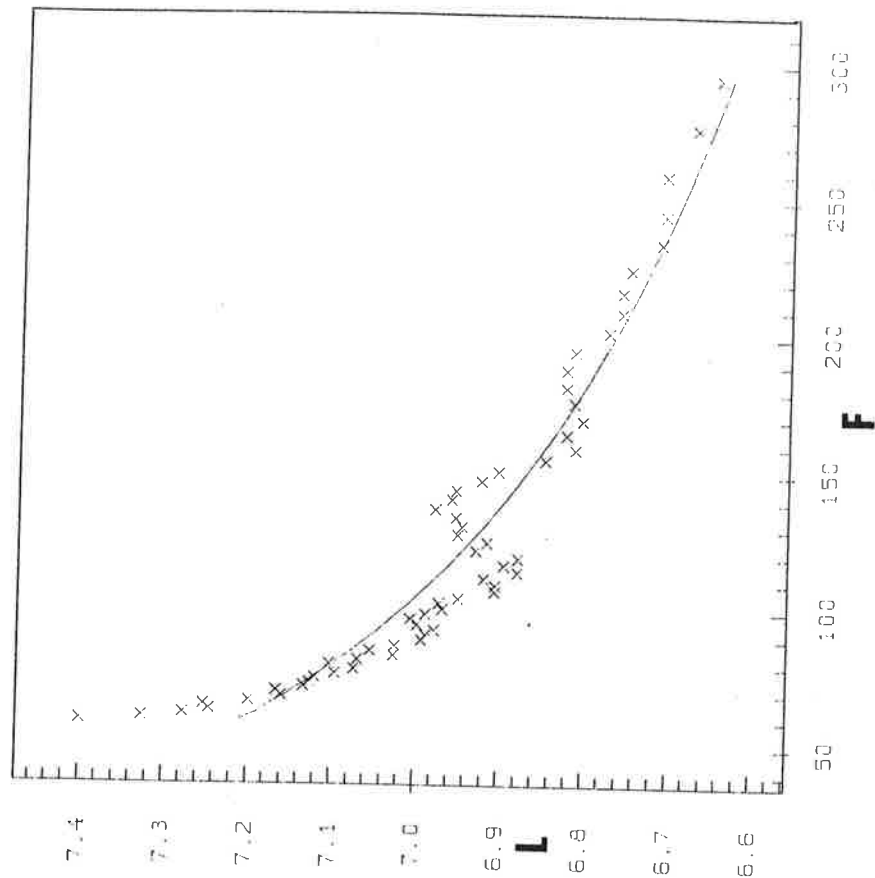


Abbildung 33: Die Länge (L) der Funktion der Frequenz (F). Berechnet mit Hilfe gleitender Mittel über ein Intervall von 100

ausgedrückt. Mit einer zusätzlichen Überlegung wollen wir diesen Ansatz verfeinern:

Eine lexikalische Einheit wird um so stärker gekürzt, je länger sie ist. Ein infolge häufiger Verwendung bereits kurzer Ausdruck wird auch bei weiterhin hoher oder sogar steigender Frequenz nicht mehr sehr viel kürzer. So ist nicht zu erwarten, daß deutsche Wörter wie Luft, Buch oder Fuß noch kürzer werden können. Dagegen ist es durchaus wahrscheinlich, daß die Ausdrücke Katalysator oder Umweltverschmutzung durch steigende Gebrauchs- häufigkeit in naher Zukunft eine solche Verkürzung erfahren oder aber durch kürzere ersetzt werden⁷⁵. Die Kürzung eines Wortes kann also je nach Länge beschleunigt oder gebremst werden. Wir formulieren dies als Hypothese:

Die Veränderung der Kürzungsrate einer lexikalischen Einheit ist eine Funktion der Länge und der Frequenz.

Für die exakte Darstellung dieser Annahme als Differentialgleichung gibt es mehrere naheliegende Möglichkeiten; für eine eindeutige Auswahl unter ihnen besitzen wir heute noch keine ausreichenden theoretischen Kenntnisse. In jedem Fall aber entspricht der Aussage eine Differentialgleichung zweiter Ordnung. Ein plausibler Ansatz ist der folgende:

$$L'' = -q L + k e^{-rF}$$

wo L für die Länge, F für die Frequenz steht, und q und r Konstanten sind. Diese Gleichung drückt einen linearen Zusammenhang aus, nach dem

- a) bei großer Länge eine kleine Veränderung der Kürzung (= geringe Bremsung der noch starken Kürzung),

75 Die Fälle Bahn für Eisenbahn, Cafete für Gafeteria und Mikro für Mikrocomputer sind Beispiele, die zur Zeit beobachtbar sind. Die Prozesse haben verschiedene Geschwindigkeiten, die aus der unterschiedlich schnellen Häufigkeitszunahme resultieren.

bei kleiner Länge eine starke Veränderung der Kürzung (= starke Bremsung der schon geringen Kürzung)

- b) eine exponentiell abnehmende positive Abhängigkeit von der Frequenz

angenommen wird. Für q , k und r ist damit ein positives Vorzeichen festgelegt.

Die Wurzeln der charakteristischen Gleichung sind $i q$ und $-i q$. Da es sich um komplexe Zahlen handelt, ist die Lösung der homogenen Gleichung (mit $Q = q$),

$$L_h = a \cos QF + b \sin QF,$$

eine oszillierende Funktion der Frequenz. Mit der partikulären Lösung $L_p = c e^{rF}$ erhält man

$$L = a \cos QF + b \sin QF + c e^{rF}.$$

Diese Funktion repräsentiert eine Exponentialfunktion mit aufmodulierter Schwingung. Auf den Versuch einer Anpassung an die Daten soll hier verzichtet werden.

An dieser Stelle sollte keine detaillierte Lösung des Problems ausgeführt werden (das wäre vom Umfang her bereits eine eigenständige Arbeit). Vielmehr soll hier eine grundsätzliche Einordnung der beobachteten Phänomene in den Modellzusammenhang erfolgen:

Die Oszillation der Länge in der Frequenz läßt sich aus der Annahme einer veränderlichen Kürzungsrate, die von der Länge abhängt, erklären. Eine schwingende Funktion, deren Gesamtamplitude abnimmt, ergibt sich durch den Ansatz einer linearen Differentialgleichung zweiter Ordnung, wenn bestimmte Bedingungen erfüllt sind:

- a) Die Veränderung der Kürzungsrate ist positiv proportional zur Länge

b) Die Veränderung der Kürzungsrate ist nicht abhängig von der Kürzung; allgemeiner: der Koeffizient des Gliedes erster Ordnung ist kleiner als ein Viertel des Koeffizienten nullter Ordnung. Wenn also die Differentialgleichung allgemein

$$L'' + aL' + bL = f(F)$$

lautet, so ergibt sich eine oszillierende Funktion als Lösung dann, wenn $4a < b$. Wenn $a = 0$, so ist die Schwingung ungedämpft.

c) Die Funktion $f(F)$ ist monoton fallend.

Die erste Bedingung ist wohl für natürliche Sprachen immer erfüllt, über die zweite können wir heute noch keine Aussage machen; aus den untersuchten Daten läßt sich nur ein Schluß für das Deutsche ziehen. Die dritte Bedingung folgt direkt aus dem Modell.

8. VOM BASISMODELL ZUR

SPRACHTHEORIE

8.1 LEISTUNG DES ENTWICKELTEN MODELLS

Hauptziel der vorliegenden Studie war es, ein Modell aufzustellen, das mehrere bisher isolierte, für sich stehende Sprachgesetze aus einem einheitlichen Ansatz erklärt und darüber hinaus komplex genug ist, um aus ihm bisher unbekannte Gesetze zu deduzieren.

Das dargestellte Modell umfaßt sechs linguistische Größen, zehn Außenbezüge in Form von Bedürfnissen und zwölf interpretierte Prozesse⁷⁴. Die aus der angesetzten Modellstruktur ableitbaren Funktionen befinden sich in Übereinstimmung mit den bisher bekannten Variablenzusammenhängen; sie sind sogar formal identisch mit dem "Menzerathschen Gesetz" von Altmann (1980). Darüber hinaus wurde eine Reihe bisher nicht bekannter Abhängigkeiten und Prozesse aus dem Aufbau des Modells gefolgert. Eine weitere Konsequenz war die Folgerung, daß die Verteilungen der Eigenschaften lexikalischer Einheiten Funktionen voneinander sind.

Sowohl die einzelnen theoretisch aus der Struktur abgeleiteten Konsequenzen als auch das Modell insgesamt sind anhand einer empirischen Überprüfung an einem Korpus

deutschsprachiger Texte bestätigt worden. Die aus den Daten hervorgehenden weiterführenden Indizien⁷⁷ sind vollständig mit den dem Modell zugrundeliegenden Annahmen verträglich. Direkte und sogar mehrfach indirekte Tests der vorhergesagten Variablenabhängigkeiten ergeben sehr hoch signifikante Resultate.

Die unerwartete Entdeckung einer Schwingung im lexikalischen Raum entlang der Frequenz - Achse war müheles durch Einbeziehung einer bereits vorher vermuteten Rückkopplung in das Modell erklärbar.

Der Grundgedanke von Zipfs Arbeit, das principle of least effort, und seine Unifikations- und Diversifikationskräfte erfahren durch das Modell eine Bestätigung und Neuinterpretation auf der Grundlage der Funktionalanalyse und der Synergetik. Ein wesentlicher Zug des Ansatzes ist das strukturelle Axiom der sprachlichen Selbstregulation, wie es von Altmann (1981) für die linguistische Funktionalanalyse gefordert wurde.

Das Modell ist so angelegt, daß es in ein umfassendes Kommunikationsmodell aufgrund soziologischer und psychophysischer Funktionalanalysen eingebettet werden kann. Als Anknüpfungspunkte dienen die Bedürfnisse und die Funktionsparameter, soweit sie nicht bereits modellintern determiniert sind. Wegen dieser Eigenschaft erlaubt es das Modell, die Systemstruktur selbst wieder aus höheren Prinzipien abzuleiten.

77 vgl. die Anpassung der Abhängigkeit der Frequenz von der Polytextie, deren Resultat auf die Existenz einer noch nicht berücksichtigten Größe hindeuten könnte.

8.2 SKIZZE DES WEITEREN AUSBAUS

Aus diesen Gründen darf das Modell als Ansatz zu einer zu entwickelnden Sprachtheorie charakterisiert werden. Der hier herausgearbeitete Kern ist gewiß nicht die einzige oder auch nur die beste Startmöglichkeit für ein solches Vorhaben; eine Vielzahl anderer Größen eignet sich sicher ebenso gut. Insofern ist nicht der Ausgangspunkt der Modellierung wesentlich, sondern die Erweiterbarkeit. Daher sollen noch kurz einige Aspekte des Modellausbaus angesprochen werden.

8.2.1 Erweiterung des Gegenstandsbereichs und Strukturverfeinerung

Die spezifisch linguistische Leistung beim Ausbau des Modells wird in der schrittweisen Ausdehnung auf weitere Phänomene bestehen. Hierbei lassen sich zwei einander ergänzende Vorgehensweisen unterscheiden: die theoretisch und die empirisch motivierte Erweiterung.

Im ersten Fall erfolgt die Aufnahme neuer Größen oder die Modellierung zusätzlicher Regelkreise auf der Grundlage von bereits etabliertem linguistischen Wissen oder auf plausiblen Annahmen. So ist z.B. die grobe Modellierung von phonologischen, morphologischen, syntaktischen und lexotaktischen Zusammenhängen sowie die Verknüpfung der resultierenden Subsysteme relativ leicht in Anlehnung an das Begriffssystem geeigneter Grammatikmodelle⁷⁸ zu verwirklichen. So findet die Konstituenz-Relation linguistischer Einheiten verschiedener Ebenen im Modellsystem ihren Niederschlag direkt in der Beziehung zwischen den Größen Inventarumfang (der Einheiten der unteren Ebene) und Länge (der Einheiten der übergeordneten Ebene). Die syntaktischen Eigenschaften einer Sprache gehen in den Synthesizitätsparameter (z.B.

78 Hier ist in erster Linie an die Stratifikationsgrammatik (Laab: 1966; s. auch Lockwood: 1972; Schreyer 1977) gedacht.

im Zusammenhang mit dem Bedürfnis nach Bedeutungsspezifisierung', in Wortklassenparameter und natürlich in die Charakteristika satzbezogener Regelkreise ein.

Auf ähnliche Weise lassen sich viele Grobstrukturen aufstellen; eine feinere Modellierung von Subsystemen und ihren Interrelationen bedarf aber sicherlich auch empirischer Forschung.

Ebenfalls aus theoretischen Überlegungen heraus werden Hierarchisierungen von Regelkreisen und Prozessen vorgenommen werden. Als Beispiel für eine sehr hohe Ebene der Selbstregulation soll hier noch die Regelung der Steuerungseffektivität' in das Modell aufgenommen werden. Unter diesem Terminus wollen wir ein Maß dafür verstehen, wie stark Impulse aus der Systemumgebung zu einer gegebenen Zeit auf das Modell einwirken können. Einerseits ist diese Einwirkung der Bedürfnisse (bzw. die Ausrichtung des Systems nach den Bedürfnissen) zur Anpassung an veränderliche Umstände erforderlich. Andererseits kann eine zu schnelle Reaktion der Prozesse auf solche Veränderungen zu einer Katastrophe führen. Sie beruht darauf, daß sich an verschiedenen Elementen des Systems Änderungen vollziehen, die nicht miteinander vereinbar sind - eine Situation die entstehen kann, wenn die Anpassung schneller ist als der systeminterne Signalfluß. Für den menschlichen Teilnehmer an der Sprachgemeinschaft würde sich eine solche Katastrophe darin äußern, daß sich die Sprache schneller verändert, als er es nachvollziehen kann, so daß sie die Funktionsfähigkeit als Kommunikationsmittel für ihn verliert. Dem Bedürfnis nach Anpassungsfähigkeit des Systems steht daher ein Bedürfnis nach Stabilität entgegen. Es wirkt bremsend und moderierend, um den Zerfall des Systems zu verhindern. Die Abbildung 34 stellt die Regelung der Steuerungseffektivität durch ein Gleichgewicht zwischen Anpassung und Stabilisierung schematisch und stark vereinfacht dar. Bei einer ausführlichen

79 vgl. Köhler, Altmann (1983)

Integration dieser Steuerungsebene müsste berücksichtigt werden, daß den Verbindungen zwischen jedem Bedürfnis und den von ihm betroffenen Systemgrößen individuelle Ausprägungen der Steuerungseffektivität zukommen.

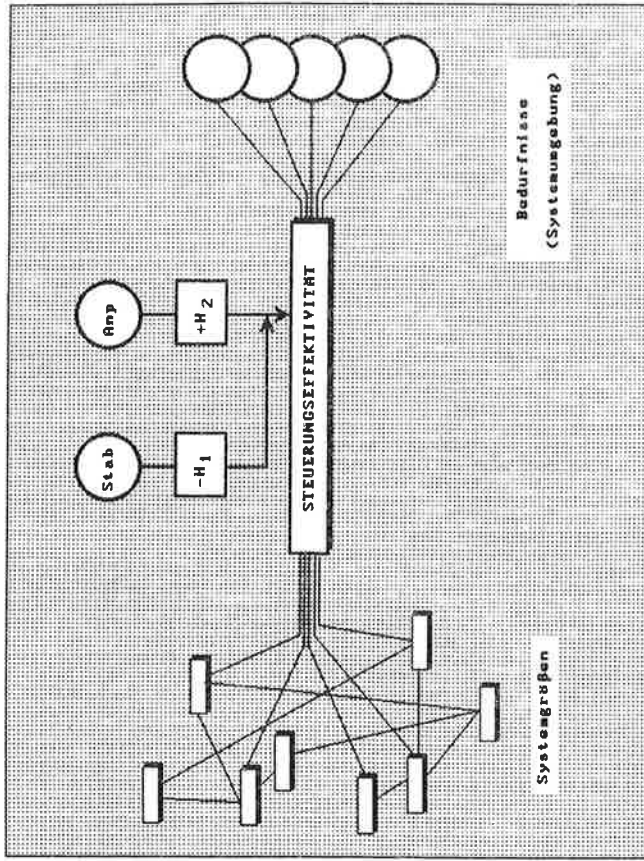


Abbildung 34: Die Regelung der Effektivität, mit der die Systemumgebung auf das System wirkt, durch die Bedürfnisse Stabilität und Anpassung

Andere Modellerweiterungen werden aus Gründen der empirischen Evidenz vorgenommen. So ist es z.B. möglich, für jeden Variablenzusammenhang zu messen, welcher Anteil der Variabilität aufgrund des Modells noch unerklärt ist und daher die Einführung einer weiteren Größe erfordert.

8.2.2 Interaktion zwischen System und Systemumgebung

Von diesem Bereich des weiteren Ausbaus ist zu erwarten, daß er vor allem von der Suche nach Bedürfnissen geprägt sein wird. Bislang sind die linguistischen Überlegungen mit funktionalanalytischer Zielsetzung noch weit von verwertbaren Ergebnissen entfernt; die Formulierungen von "Kommunikationsbedürfnissen" bleiben noch sehr unkonkret und eher programmatisch (vgl. Kanngießer 1977). Schwieriger als das Aufstellen der Bedürfnisse wird die Zuordnung der funktionalen Äquivalente sein, am schwersten die Angabe der Relationen der Äquivalente untereinander.

Um eines Tages zu einem ernstzunehmenden Sprachmodell zu kommen, das das Verhältnis von Sprache und Gesellschaft und die psychobiologischen Voraussetzungen von Sprache mitabbildet, müssen auch die Rückwirkungen des Systems auf seine Umgebung zum Gegenstand der Forschung gemacht werden. So wie die Bedürfnisse als Eingangsgrößen Teil des Modells (nicht des Systems) sind, so gehören auch die Ausgangsgrößen dazu. Sie bilden die Schnittstellen zu Untersuchungsbereichen wie 'sprachliche Relativität' (im Sinne v. Sapir/Whorf) bis 'Sprache als politisches Instrument'.

8.2.3 Modellierungstechnik

Sicher werden im Laufe der Zeit die verwendeten Methoden ausgearbeitet und verbessert. Schon jetzt ist absehbar, daß lineare Operatoren und Graphenalgebra nicht genügen, selbst wenn oben fast alle Zusammenhänge durch Logarithmieren linearisiert werden könnten.

Ein möglicher Schritt könnte auch darin bestehen, die Modellierung mit Hilfe von stochastischen Prozessen anstelle von

Differentialgleichungen vorzunehmen; dies würde auch die Ableitung der Verteilungen von linguistischen Größen erleichtern.

Eine wesentliche Erleichterung für die zu erarbeitende Methodologie wird mit Sicherheit die interdisziplinäre Zusammenarbeit mit Wissenschaftlern, die in anderen Bereichen mit selbstregulierenden Systemen zu tun haben, - mit Synergetikern - spielen, 'zumal die Erforschung kooperativer Phänomene in einigen Disziplinen schon sehr weit ausgearbeitet ist.

8.2.4 Zeitverhalten des Systems

Von einem ausreichend hohem Komplexitätsgrad an wird die Einbeziehung der Zeitachse interessant werden. Dieser Aspekt der Dynamik wurde hier bisher vor allem deshalb vernachlässigt, weil (gerade für das Deutsche) kaum verlässliche sprachhistorische Angaben verfügbar sind. In den nächsten Jahren wird es aber möglich sein, sowohl empirische Studien zur Überprüfung des gegenwärtigen Modellstands auch an Sprachen durchzuführen, für die bessere diachrone Daten vorhanden sind⁸⁰, als auch schließlich zu einem verbesserten, komplexen Modell zu kommen, welches nach ausgiebigen empirischen Tests zum Zweck der Simulation verwendet werden kann.

Die linguistische Simulation an einem solchen systemtheoretischen Modell versetzt die Sprachwissenschaft in die Lage, quasi-experimentelle Untersuchungen durchzuführen. Damit wäre es möglich, Vorgänge, die in der Realität einmalig und unwiederholbar sind, unter laborähnlichen Bedingungen zu studieren.

80 z.B. Russisch, Englisch oder Schwedisch

9. S C H L U S S B E M E R K U N G

Hier soll nicht der Anspruch erhoben werden, durch die oben skizzierten Schritte sei bereits der Weg zu einer langersehten Sprachtheorie vorgezeichnet. Noch bleiben zu viele Fragen offen, zu viele Gebiete nicht einmal berührt, als daß die Konturen dieser Theorie heute schon sichtbar wären.

So sind etwa die Bereiche der Syntax, der Pragmatik, vor allem der Semantik in dem vorgestellten Modell kaum angeschnitten worden, und nur Schnittstellen zu diesen linguistischen Aspekten werden geboten - nicht aber genauere Vorstellungen zu ihrer Integration.

Andererseits bietet der dargestellte Ansatz - insbesondere dank des fundamentalen Axioms der Selbstregulation - die Möglichkeit, sämtliche sprachwissenschaftlichen Forschungsaspekte in einem Modell zu vereinheitlichen. Darüber hinaus kann die Linguistik sich mit seiner Hilfe - gemeinsam mit einer Reihe von Nachbarwissenschaften und den modernen Naturwissenschaften - der interdisziplinär verbindenden Theorie der Synergetik bedienen und so an einen beträchtlichen methodologischen Fortschritt für sich nutzbar machen. Auf den eigenen Forschungsgegenstand bezogen bedeutet das für die Linguistik unter anderem, daß eine Reihe von Scheinfragen endlich aufgegeben bzw. aufgelöst werden kann.

Nicht zuletzt betreffen solche Scheinfragen die Identität des sprachwissenschaftlichen Forschungsgegenstands selbst. Kein Versuch, eine bestimmte Sprache territorial, ethnisch oder historisch abzugrenzen, kann linguistisch befriedigen. In jedem Fall ist das Resultat entweder noch immer zu heterogen, um einheitlich beschrieben werden zu können, oder zu begrenzt, um als Sprache bezeichnet zu werden: 'Die Sprache' besitzt eine unscharfe Identität.

Im Lichte des verwendeten Ansatzes wird deutlich, daß der klassifikatorische Begriff "Sprache" im Sinne von Nationalsprache ebenso unzulänglich ist wie die anderen Kategorialbegriffe. Er ist für bestimmte Zwecke nützlich, seine Überstapazierung aber muß in die Irre führen. Zur Verdeutlichung sei ein Bild erlaubt: Die Anwendung dieses klassifikatorischen Begriffs auf das Phänomen Sprache gleicht dem Versuch, im Atlantik eine einzelne Welle abzugrenzen und zu etikettieren. Wie diese ist die einzelne Sprache, wenn Sprache als selbstregulierendes System verstanden wird, eine durch die Aufmerksamkeit des Beobachters hervor gehobene, zeitlich begrenzte Makrostruktur, die durch kooperative Prozesse innerhalb des Systems und äußere Einwirkungen entsteht. Unterschiede in den äußeren Einwirkungen, also geographisch, kulturell, soziologisch und historisch differierende Bedürfnisse, führen zu unterschiedlichen Anpassungen durch die Selbstregulationsmechanismen. Wenn die Stabilität gegenüber der Anpassung nur lokal bewahrt bleiben kann, wenn der Anpassungsdruck bei unvereinbaren Bedürfnissen überwiegt, entstehen 'Sprachen', Dialekte, Soziolekte und Fachsprachen. Bei entsprechenden Umgebungsbedingungen, d.h. bei Konvergenz der Systembedürfnisse, können Sprachen und Sprachvarietäten auch "zusammenfließen".

Ebenso deutlich tritt bei systemtheoretischer Modellierung das Verhältnis von linguistischer Beschreibung zur linguistischen Erklärung hervor. Die mit Hilfe von Grammatiken, Wörterbüchern, Statistiken u.a. formulierten Beschreibungen sind Abbildungen von Ausschnitten aus Momentaufnahmen der dynamischen Entwicklung des Sprachsystems.

Von einer linguistischen Theorie wird man erst dann sprechen dürfen, wenn aus ihr die Eigenschaften der Lexik, die grammatischen Regularitäten (und Unregelmäßigkeiten) und andere Aspekte der linguistischen Deskription abgeleitet, d.h. erklärt und vorhergesagt werden können.

L I T E R A T U R

- Arapov, M. V.; Cherc, M. M.**
1983 Mathematische Methoden in der historischen Linguistik. Bochum
- Beőthy, E.; Altmann, G.**
1985 Semantic Diversification of Hungarian Verbal Prefixes III. "fől-", "el-", "be-". In: Rothe, U. [Hrsg.]: Glottometrika 7, Bochum, 45-56
- Best, K. -H.**
1983 Zum morphologischen Wandel einiger deutscher Verben. In: Best, Kohlhasse 1983a, 107-118
- Best, K. -H.; Kohlhasse, J. [Hrsg.]**
1983a Exakte Sprachwandelersuchung. Göttingen
- Best, K. -H.; Kohlhasse, J.**
1983b Der Wandel von ward zu wurde. In: Best, Kohlhasse 1983a, 91-102
- Boroda, M. G.**
1982 Die melodische Elementareinheit. In: Orlov, Ju. K.; Boroda, M. G.; Nadarejşvili, I. S.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 205-221
- 1982b Häufigkeitsstrukturen musikalischer Texte. In: Orlov, Ju. K.; Boroda, M. G.; Nadarejşvili, I. S.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 231-262
- 1982c Zur Bestimmung einer phrasenähnlichen melodischen Informationseinheit in der Musik. In: Orlov, Ju. K.; Boroda, M. G.; Nadarejşvili, I. S.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 222-232
- Bunge, M.**
1967 Scientific Research I, II. Berlin. Heidelberg. New York
- Altmann, G.**
1980 Prolegomena to Menzerath's Law. In: Grotjahn, R. [Hrsg.]: Glottometrika 2. Bochum, 1-10
- 1981 Zur Funktionalanalyse in der Linguistik. In: Esser, J.; Hübler, A. [Hrsg.]: Forms and Functions. Tübingen 25-32
- 1983 Das Piotrovski-Gesetz und seine Verallgemeinerungen. In: Best, K. -H.; Kohlhasse, J. [Hrsg.]: Exakte Sprachwandelersuchung. Göttingen, 59-90
- 1984a Vortrag auf der Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft in Bielefeld 1984
- 1984b Semantische Diversifikation. In: Folia Linguistica (erscheint)
- Altmann, G.; von Buttlar, H.; Rott, W.; Strause, U.**
1983 A Law of Language Change. In: Brainerd, B. [Hrsg.]: Historical Linguistics. Bochum, 104-115
- Altmann, G.; Beőthy, E.; Best, K. -H.**
1982 Die Bedeutungskomplexität der Wörter und das Menzerathsche Gesetz. In: Zs. für Phonetik, Sprachwissenschaft und Kommunikationsforschung 35, 537-543
- Altmann, G.; Lohfeldt, W.**
1980 Einführung in die Quantitative Phonologie. Bochum
- Altmann, G.; Schwibbe, M.**
1985 Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. (erscheint)

- Burghes, D.N.; Graham, A.
1980 Introduction to Control Theory, Including Optimal Control. Chichester
- Carroll, J.B.
1967 On Sampling from a Lognormal Model of Word-Frequency Distribution. In: Kucera, H.; Francis, W.N.: Computational Analysis of Present-Day American English. Providence, Rhode Island, 406-424
- 1968 Word-Frequency Studies and the Lognormal Distribution. In: Zale, E.M. [Hrsg.]: Proceedings of the Conference on Language and Language Behavior. New York, 213-235
- 1969 A rationale for an asymptotic lognormal form of word-frequency distribution. Princeton
- 1970/72 An Alternative to Juillard's Usage Coefficient for Lexical Frequencies, and a Proposal for a Standard Frequency Index (SFI). In: Computer Studies in the Humanities and Verbal Behavior 3, 61-65
- Chomsky, N.
1957 Syntactic Structures. The Hague
- Condon, E.V.
1923 Statistics of vocabulary. In: Science 67, 300
- Cortés, F.; Przeworski, A.; Sprague, J.
1974 Systems Analysis for Social Scientists. New York. London. Sydney. Toronto
- Dym, C.L.; Ivey, E.S.
1980 Principles of mathematical modelling. London
- Eigen, M.
1971 Selforganization of Matter and the Evolution of Biological Macromolecules. In: Die Naturwissenschaften 58, 465-523

- Esser, H.; Klenovits, K.; Zehnppennig, H.
1977 Wissenschaftstheorie 2, Stuttgart
- Estoup, J.B.
1916 Gammes Stenographiques. Paris
- Fickermann, I.; Markner-Jäger, B.; Rothe, U.
1984 Wortlänge und Bedeutungskomplexität. In: Boy, J.; Köhler, R. [Hrsg.]: Glottometrika 6. Bochum, 115-126
- Fónagy, I.; Magdics, K.
1960 Speed of utterance in Phrases of different lengths. In: Language and Speech 3, 179-192
- Fucks, W.
1955a Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen. Köln. Opladen
1955b Theorie der Wortbildung. In: Mathematisch-physikalische Semesterberichte 4, 195-212
- Galtung, J.
1967 Theory and Methods of Social Research. Oslo
- Gerlach, R.
1982 Zur Überprüfung des Menzerathschen Gesetzes im Bereich der Morphologie. In: Lehfeldt, W., Strauß, U. [Hrsg.]: Glottometrika 4. Bochum, 95-102
- Geršić, S., Altmann, G.
1980 Laut - Silbe - Wort und das Menzerathsche Gesetz. In: Forum Phonetikum 21. Frankfurt/M., 115-123
- Gigerenzer, G.
1981 Messung und Modellbildung in der Psychologie. München. Basel

Grotjahn, R.

- 1982 Ein statistisches Modell für die Verteilung der Wortlänge. In: Zeitschrift für Sprachwissenschaft 1, 44-75

Guiter, H.

- 1974 Les relations fréquence - longueur - sens des mots (langues romanes et anglaises). XIV Congresso Internazionale di linguistica e filologia romanza, Napoli 1974, 15-20

Haight, F. A.

- 1969 Two probability distributions connected with Zipf's rank-size conjecture. In: Applicationes Mathematicae case 10, 225-228

Haken, H.

- 1978 Synergetics. Berlin. Heidelberg. New York

Haken, H. [Hrsg.]

- 1981 Chaos and order in nature. Berlin

Haken, H.; Wagner, M. [Hrsg.]

- 1973 Cooperative phenomena. Berlin

Hempel, C. G.

- 1965a The logic of functional analysis. In: Hempel 1965b, 297-330

Hempel, C. G. [Hrsg.]

- 1965b Aspects of scientific explanation. New York

Hempel, C. G.; Oppenheim, P.

- 1948 Studies in the Logic of Explanation. In: Philosophy of Science 15, 135-175

Herdan, G.

- 1958 The Relation between the Dictionary Distribution and the Occurrence Distribution of Word length and its importance for the study of Quantitative Linguistics. In: Biometrika 45, 222-228

1966

- The advanced theory of language as choice and chance. Berlin. Heidelberg. New York

Heupa, G.

- 1983 Untersuchungen zum Verhältnis von Satzlänge zu Clauselänge am Beispiel deutscher Texte verschiedener Textklassen. In: Köhler, R., Boy, J. [Hrsg.]: Glottometrika 5. Bochum, 113-133

Hirsch, M. W.; Smale, S.

- 1974 Differential equations, dynamical systems, and linear algebra. London

Imsiepen, U.

- 1983 Die e-Epithese bei starken Verben im Deutschen. In: Best, Kohlhasse 1983a, 119-142

Job, U.; Altmann, G.

- 1985 Ein Modell für anstrengungsbedingte Lautveränderung. In: Folia Linguistica Historica, 1985 (erscheint)

Kanngießer, S.

- 1977 Skizze des linguistischen Funktionalismus. In: Osnabrücker Beiträge zur Sprachtheorie 3, 132-187

Kohlhasse, J.

- 1983 Die Entwicklung von ward zu wurde beim Nürnberger Chronisten Heinrich Deichsler. In: Best, Kohlhasse 1983a, 103-106

Köhler, R.

- 1982 Das Menzerathsche Gesetz auf Satzebene. In: Lehfeldt, W., Strauß, U. [Hrsg.]: Glottometrika 4. Bochum, 103-113

Köhler, R.; Altmann, G.

- 1983 Systemtheorie und Semiotik. In: Zeitschrift für Semiotik 5/4, 424-431

Krylov, Ju. K.

- 1982a Eine Untersuchung statistischer Gesetzmäßigkeiten auf der paradigmatischen Ebene der Lexik natürlicher Sprachen. In: Guiter, H.; Arapov, M. V. [Hrsg.]: Studies on Zipf's Law. Bochum, 234-262

- 1982b Ob odnož paradigme lingvostatističeskich rasprede- lenij. In: Ja. Soontak [Hrsg.]: Lingvostatistika i vyčislitel'naja lingvistika, Tartu, 80-102

Kuhn, T. S.

- 1967 Die Struktur wissenschaftlicher Revolutionen. Frankfurt/M.

Kulla, B.

- 1979 Angewandte Systemwissenschaft. Würzburg, Wien

Lakatos, I.

- 1974 Falsifikation und Methodologie wissenschaftlicher Forschungsprogramme. In: Lakatos, I.; Musgrave, A. [Hrsg.]: Kritik und Erkenntnisfortschritt. Braunschweig, 88-189

1975

- Kritischer Rationalismus und die Methodologie wissenschaftlicher Forschungsprogramme. In: Weingart [Hrsg.]: Wissenschaftsforschung. Frankfurt, 91-132

Lamb, S. M.

- 1966 Outline of Stratificational Grammar. Washington

Lánaky, P., Radil-Weiss, T.

- 1980 A Generalization of the Yule-Simon Model, With Special Reference to Word Association Tests and Neural Cell Assembly Formation. In: Journal of Mathematical Psychology 21, 53-65

Lehfeldt, W.

- 1980 Zur numerischen Erfassung der Schwierigkeit des Sprechbewegungsablaufs. In: Grotjahn, R. [Hrsg.]: Glottometrika 2. Bochum, 44-61

Levelt, W. J. M.

- 1974 Formal Grammars in Linguistics and Psycholinguistics I-III. The Hague, Paris

Lindblom, B.

- 1983 Economy of Speech Gestures. In: MacNeilage, P. [Hrsg.]: Speech Production. New York

Lockwood, D. G.

- 1972 Introduction to Stratificational Linguistics. New York

Lüdtke, H.

- 1980 Auf dem Weg zu einer Theorie des Sprachwandels. In: Lüdtke, H. [Hrsg.]: Kommunikationstheoretische Grundlagen des Sprachwandels. Berlin, New York, 182-252

Mandelbrot, B.

- 1953 An informational theory of statistical structure of language. In: Jackson, W. [Hrsg.]: Communication Theory. London, 486-502

1961

- On the Theory of Word Frequencies and on Related Markovian Models of Discourse. In: Proceedings of Symposia in Applied Mathematics 12, 190-219

1965

- Information Theory and Psycholinguistics. In: Wolman, B. B., Nagel, E. [Hrsg.]: Scientific Psychology. New York, London, 550-562

Menzerath, P.

- 1954 Die Architektonik des deutschen Wortschatzes. Bonn

Moulton, W. G.

- 1961 Lautwandel durch innere Kausalität: die ost-schweizerische Vokalspaltung. In: Zeitschrift für Mundartforschung 28, 227-251

Müller-Hasemann, W.

- 1983 Das Eindringen englischer Wörter ins Deutsche ab 1945. In: Best, Kohlhasse 1983a, 143-160

Orth, B.

- 1974 Einführung in die Theorie des Messens. Stuttgart. Berlin. Köln. Mainz

Orlov, Ju. K.

- 1970a Obosženie zakona Cipfa-Mandelbrota. In: Soobsženija akademii nauk Gruzinskoi SSR 57/1, 37-40
- 1970b O statističeskoj strukture soobsženij, optimal'nych dlja čelovečeskogo vosprijatija. In: Naučno - Tehničeska Informacija, Serija 2, 8, 11-15
- 1982a Dynamik der Häufigkeitsstrukturen. In: Orlov, Ju. K.; Boroda, M. G.; Nadarejšvili, I. Š.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 82-117
- 1982b Ein Modell der Häufigkeitsstruktur des Vokabulars. In: Orlov, Ju. K.; Boroda, M. G.; Nadarejšvili, I. Š.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 118-192
- 1982c Linguostatistik: Aufstellung von Sprachnormen oder Analyse des Redeprozesses? (Die Antinomie "Sprache-Rede" in der statistischen Linguistik). In: Orlov, Ju. K.; Boroda, M. G.; Nadarejšvili, I. Š.: Sprache, Text, Kunst. Bochum, 1-55

Orlov, Ju. K.; Boroda, M. G.; Nadarejšvili, I. Š.

- 1982 Sprache, Text, Kunst. Quantitative Analysen. Bochum

Osgood, C. E.

- 1963 On Understanding and Creating Sentences. In: Amer. Psychol. 18, 735-751

Paulre, E. (Hrsg.)

- 1980 System Dynamics and the Analysis of Change. Proceedings of the 6th International Conference on System Dynamics, University of Paris-Dauphine, November 1980. Amsterdam. New York. Oxford

Pieper, U.

- 1979 Über die Aussagekraft statistischer Methoden für die linguistische Stilanalyse. Tübingen

Piotrovski, J. R. G.; Bektaev, K. B.; Piotrovskaja, A. A.

- 1977 Matematičeskaja Lingvistika. Moskva

Popper, K. R.

- 1972 Conjectures and Refutations. The Growth of Scientific Knowledge. London
- 1973 Objektive Erkenntnis. Hamburg

Prigogine, I.

- 1979 Vom Sein zum Werden. München

Rapoport, A.

- 1957 The Stochastic and the "Teleological" Rationales of Certain Distributions and the So-Called Principle of Least Effort. In: Behavioural Science, 147-161

- 1982 Zipf's Law Re-visited. In: Guitier, H.; Arapov, M. V. [Hrsg.]: Studies on Zipf's Law. Bochum, 1-28

Schaefer, B.

- 1976 Maschinenlesbare Textkorpora des Deutschen und des Englischen. In: Deutsche Sprache 4, 356-369

Scheuch, E.K., Zehnpfennig, H.

- 1974(3) Skalierungsverfahren in der Sozialforschung. In: Koenig, R. (Hrsg.): Handbuch der empirischen Sozialforschung, Bd. 3a, 97-186. Stuttgart

Schreyer, R.

- 1977 Stratifikationsgrammatik. Tübingen

Schwibbe, M.

- 1984 Text- und wortstatistische Untersuchungen zur Validität der Menzerath'schen Regel (MR). In: Boy, J., Köhler, R. (Hrsg.): Glottometrika 6. Bochum, 152-176

Simon, H.A.

- 1955 On a Class of Skew Distribution Functions. In: Biometrika 42, 425-440
1960 Some Further Notes on a Class of Skew Distribution Functions. In: Information and Control 3, 80-88

Spinner, H.F.

- 1974 Pluralismus als Erkenntnismodell. Frankfurt/M. 1974

Stegmüller, W.

- 1969 Wissenschaftliche Erklärung und Begründung. Berlin
1970 Theorie und Erfahrung. Berlin. Heidelberg. New York

Steinthal, H.

- 1881 Abriss der Sprachwissenschaft. Berlin

Stoer, J.

- 1979 Einführung in die Numerische Mathematik I. Berlin. Heidelberg. New York

Strauß, U.

- 1980 Struktur und Leistung der Vokalsysteme. Bochum

Suppes, P.

- 1970 Probabilistic grammars for natural languages. In: Synthese 22, 95-116

Teupenhayn, R., Altmann, G.

- 1984 Clause length and Menzerath's law. In: Boy, J., Köhler, R. (Hrsg.): Glottometrika 6. Bochum, 127-138

Vološin, B.A.; Orlov, Ju.K.

- 1982 Das verallgemeinerte Zipf-Mandelbrotsche Gesetz und die Verteilung der Anteile von Farbflächen in der Malerei. In: Orlov, Ju.K.; Boroda, M.G.; Naderjškvič I.Š.: Sprache, Text, Kunst. Bochum, 263-270

Wahrig, G.

- 1981 Deutsches Wörterbuch. Gütersloh

Walker, H.M., Lav, J.

- 1953 Statistical Inference. New York

Watt, W.C.

- 1983 Grade der Systemhaftigkeit. Zur Homogenität der Alphabetschrift. In: Zs. für Semiotik 5/4, 371-399

Weber, H.J.

- 1974 Mehrdeutige Wortformen im heutigen Deutsch. Tübingen

Willee, G.

- 1979 Lemma - ein Programmsystem zur automatischen Lemmatisierung deutscher Wortformen. In: Sprache und Datenverarbeitung 1-2, 45-60

Yule, G.U.

- 1924 A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. In: Philosophical Transactions, B, 213, 21

Zipf, G. K.

1949 Human Behaviour and the Principle of Least Effort. Reading, Mass.

1968(2) The psycho-biology of language. An introduction to dynamic philology. Cambridge, Mass.

Tabelle A 1: Die Polylexe in Abhängigkeit von der Länge

i	x_i	y_i y_e	y_t
1	3	5.0000	5.0485
2	4	4.6316	3.9779
3	5	4.2740	3.5066
4	6	3.6981	2.8430
5	7	2.6000	2.5022
6	8	1.8938	2.2402
7	9	1.5943	2.0319
8	10	1.7537	1.8521
9	11	1.4215	1.7207
10	12	1.3853	1.6010
11	13	1.2657	1.4983
12	14	1.2658	1.4091
13	15	1.1071	1.3308
14	16	1.2037	1.2615
15	17	1.5789	1.1998
16	18	1.0333	1.1443
17	19	1.0357	1.0941
18	20	1.0006	1.0486
19	21	1.1429	1.0071
20	22	1.1111	0.9690
21	23	1.0000	0.9340
22	24	1.2000	0.9016
23	25	1.0000	0.8716
24	26	1.0000	0.8438

* y_e - empirische Werte y_t - theoretische Werte

Tabelle A 2: Die Polytextile in Abhängigkeit von der Polytextile

i	x_i	y_{i_e}	y_{i_t}
1	1.0	3.1792	3.9158
2	2.0	10.5417	11.5634
3	3.0	19.3306	21.8456
4	4.0	28.2745	34.2652
5	5.0	69.6389	48.5833
6	6.0	89.3235	64.6220
7	7.0	77.3000	82.2491
8	8.0	112.8462	101.3609
9	9.0	230.6667	121.8732
10	10.0	187.6667	143.7157
11	11.0	80.0000	166.8287
12	13.0	365.0000	216.6655
13	14.0	286.6667	243.3037
14	16.0	193.0000	299.8389
15	19.0	373.0000	392.3432
16	21.0	285.0000	458.8559

Tabelle A 3: Die Frequenz in Abhängigkeit von der Polytextile

i	x_i	y_{i_e}	y_{i_t}	f	x_i	y_{i_e}	y_{i_t}
1	1.0	1.3299	1.1158	51	54.0	81.5000	91.2732
2	2.0	2.7801	2.3986	52	55.0	70.0000	93.1412
3	3.0	4.4861	3.7530	53	56.0	69.0000	95.0128
4	4.0	5.5128	5.1561	54	60.0	76.5000	102.5333
5	5.0	6.4474	6.5956	55	62.0	109.0000	106.3134
6	6.0	8.6970	8.0677	56	63.0	77.3333	108.2083
7	7.0	10.6667	9.5646	57	64.0	133.0000	110.1033
8	8.0	10.6875	11.0840	58	66.0	93.0000	113.9115
9	9.0	17.0000	12.6233	59	67.0	118.6667	115.6186
10	10.0	14.9091	14.1606	60	68.0	115.0000	117.7287
11	11.0	15.4286	15.7542	61	69.0	132.0000	119.6417
12	12.0	19.6429	17.3428	62	70.0	131.0000	121.5576
13	13.0	23.0000	18.9455	63	74.0	159.0000	129.2494
14	14.0	20.9000	20.5607	64	76.0	141.5000	133.1117
15	15.0	28.2857	22.1581	65	77.0	94.0000	135.0469
16	16.0	27.4286	23.8269	66	78.0	120.0000	136.5646
17	17.0	22.8333	25.4754	67	81.0	159.0000	142.6133
18	18.0	24.6667	27.1360	68	82.0	248.0000	144.7613
19	19.0	26.0000	28.8052	69	85.0	124.0000	150.6199
20	20.0	28.2500	30.4837	70	87.0	144.0000	154.5376
21	21.0	38.7500	32.1705	71	89.0	171.0000	158.4648
22	22.0	35.2500	33.8664	72	90.0	135.0000	160.4318
23	23.0	27.0000	35.5701	73	94.0	197.0000	168.3224
24	24.0	35.0000	37.2814	74	96.0	154.0000	172.2610
25	26.0	48.0000	40.7262	75	97.0	192.0000	174.2635
26	27.0	43.5000	42.4591	76	100.0	176.5000	180.2238
27	28.0	50.2500	44.1987	77	102.0	137.0000	184.2076
28	29.0	43.6000	45.9448	78	105.0	167.0000	190.1987
29	30.0	71.2500	47.6972	79	106.0	173.0000	192.1997
30	31.0	40.5000	49.4556	80	110.0	152.0000	200.2232
31	32.0	86.0000	51.2200	81	111.0	172.5000	202.2339
32	33.0	63.0000	52.9901	82	116.0	182.0000	212.3153
33	34.0	61.5000	54.7659	83	117.0	187.0000	214.3370
34	35.0	52.6667	56.5470	84	118.0	245.0000	216.3606
35	36.0	62.2500	58.3335	85	122.0	184.0000	224.4726
36	37.0	48.0000	60.1251	86	123.0	342.0000	226.5050
37	38.0	77.0000	61.9218	87	124.0	234.0000	228.5391
38	39.0	55.0000	63.7235	88	129.0	179.0000	238.7350
39	40.0	56.0000	65.5259	89	130.0	175.0000	240.7791
40	41.0	72.3333	67.3411	90	132.0	201.0000	244.8724
41	42.0	62.3333	69.1568	91	133.0	314.0000	246.9214
42	44.0	102.0000	72.8018	92	137.0	221.0000	255.1335
43	45.0	57.0000	74.6308	93	140.0	200.0000	261.3091
44	46.0	75.0000	76.4640	94	144.0	205.0000	269.5645
45	47.0	67.5000	78.3014	95	149.0	220.0000	279.9174
46	48.0	71.0000	80.1429	96	150.0	203.0000	281.8924
47	49.0	61.0000	81.9884	97	154.0	210.0000	290.3066
48	50.0	58.0000	83.8378	98	167.0	358.5000	317.4605
49	51.0	82.2500	85.6910	99	173.0	287.0000	330.0979
50	52.0	68.5000	87.5480	100	175.0	367.0000	334.3139

Table A 3 (Forts.)

i	x_i	y_{i_e}	y_{i_t}
101	180.0	277.000	344.8758
102	193.0	331.000	372.4781
103	199.0	369.000	385.2339
104	200.0	368.000	387.4221
105	209.0	353.000	406.7158
106	215.0	348.000	419.6266
107	217.0	416.000	423.9386
108	222.0	390.000	434.7367
109	257.0	657.000	511.0062
110	277.0	611.000	555.0876
111	278.0	483.000	557.3006
112	285.0	548.000	572.8145
113	296.0	762.000	597.2735
114	299.0	667.000	603.9608
115	300.0	791.000	606.1914
116	301.0	685.000	608.4226
117	304.0	778.000	615.1216
118	319.0	868.000	648.7180
119	339.0	719.000	653.7685
120	365.0	1014.000	752.7473
121	373.0	1090.000	770.9843
122	447.0	1912.000	941.5159
123	449.0	1513.000	946.1682
124	458.0	1761.000	967.1301

Table A 4: Die Frequenz in Abhängigkeit von der Polytextie (3 Parameter)

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
1	1.0	1.3299	1.6193	51	54.0	81.5000	83.9928
2	2.0	2.7801	3.1656	52	55.0	70.0000	85.6570
3	3.0	4.4861	4.6892	53	56.0	69.0000	87.3263
4	4.0	5.5128	5.2003	54	60.0	76.5000	94.0556
5	5.0	6.4474	7.7038	55	62.0	109.0000	97.4517
6	6.0	8.6976	9.2023	56	63.0	77.3333	99.1578
7	7.0	10.6667	10.6978	57	64.0	133.0000	100.8692
8	8.0	10.6875	12.1914	58	66.0	93.0000	104.3081
9	9.0	17.0000	13.6841	59	67.0	118.6667	106.0358
10	10.0	14.9091	15.1767	60	68.0	115.0000	107.7688
11	11.0	15.4286	16.6698	61	69.0	132.0000	109.5074
12	12.0	19.6429	18.1638	62	70.0	131.0000	111.2515
13	13.0	23.0000	19.6592	63	74.0	159.0000	118.2833
14	14.0	20.9000	21.1563	64	76.0	141.5000	121.8330
15	15.0	28.2857	22.6553	65	77.0	94.0000	123.6163
16	16.0	27.4286	24.1565	66	78.0	120.0000	125.4053
17	17.0	22.8333	25.6602	67	81.0	159.0000	130.8068
18	18.0	24.6667	27.1666	68	82.0	248.0000	132.6189
19	19.0	26.0000	28.6758	69	83.0	124.0000	138.0901
20	20.0	28.2500	30.1879	70	87.0	144.0000	141.7670
21	21.0	38.7500	31.7032	71	89.0	171.0000	145.4676
22	22.0	35.2500	33.2217	72	90.0	123.0000	147.3268
23	23.0	27.0000	34.7436	73	94.0	197.0000	154.8241
24	24.0	35.0000	36.2689	74	96.0	154.0000	158.6091
25	26.0	48.0000	39.3305	75	97.0	192.0000	160.5106
26	27.0	43.5000	40.8669	76	100.0	176.5000	166.2530
27	28.0	50.2500	42.4071	77	102.0	137.0000	170.1122
28	29.0	43.6000	43.9512	78	105.0	167.0000	175.9479
29	30.0	71.2500	45.4994	79	106.0	173.0000	177.9058
30	31.0	40.5000	47.0516	80	110.0	152.0000	185.8010
31	32.0	86.0000	48.6079	81	111.0	172.5000	187.7909
32	33.0	63.0000	50.1684	82	116.0	182.0000	197.8373
33	34.0	61.5000	51.7332	83	117.0	187.0000	199.8661
34	35.0	52.6667	53.3022	84	118.0	245.0000	201.8016
35	36.0	62.2500	54.8756	85	122.0	184.0000	210.1097
36	37.0	48.0000	56.4535	86	123.0	342.0000	212.1784
37	38.0	77.0000	58.0357	87	124.0	234.0000	214.2538
38	39.0	55.0000	59.6225	88	129.0	179.0000	224.7323
39	40.0	56.0000	61.2135	89	130.0	175.0000	226.8484
40	41.0	72.3333	62.8098	90	133.0	201.0000	231.1012
41	42.0	62.3333	64.4104	91	133.0	314.0000	233.2380
42	44.0	102.0000	67.6256	92	137.0	221.0000	241.8545
43	45.0	57.0000	69.2403	93	140.0	200.0000	246.3904
44	46.0	79.0000	70.8558	94	144.0	205.0000	257.2040
45	47.0	67.5000	72.4842	95	149.0	220.0000	268.3823
46	48.0	71.0000	74.1134	96	150.0	203.0000	270.6397
47	49.0	61.0000	75.7475	97	154.0	210.0000	279.7424
48	50.0	58.0000	77.3865	98	167.0	358.5000	310.1485
49	51.0	82.2500	79.0306	99	173.0	287.0000	324.6160
50	52.0	68.5000	80.6796	100	175.0	367.0000	329.5007

Tabelle A 5: Die Länge in Abhängigkeit von der Frequenz

i	x_i	y_i	y_i	i	x_i	y_i	y_i
		e	t			e	t
1	1.0	12.5497	10.9358	51	53.0	7.0000	7.1277
2	2.0	12.4107	10.1484	52	54.0	9.0000	7.1134
3	3.0	11.9375	9.7143	53	55.0	5.5000	7.0993
4	4.0	10.0893	9.4176	54	56.0	4.0000	7.0855
5	5.0	9.8140	9.1938	55	57.0	4.0000	7.0720
6	6.0	10.3750	9.0146	56	58.0	7.5000	7.0588
7	7.0	8.8095	8.8562	57	59.0	5.5000	7.0452
8	8.0	9.4444	8.7395	58	60.0	6.0000	7.0305
9	9.0	9.3077	8.6292	59	61.0	8.5000	7.0062
10	10.0	6.6667	8.5317	60	62.0	6.6667	6.9842
11	11.0	7.5625	8.4445	61	63.0	8.5000	6.9726
12	12.0	7.0000	8.3657	62	64.0	8.0000	6.9611
13	13.0	9.2222	8.2938	63	65.0	7.0000	6.9367
14	14.0	6.1000	8.2272	64	66.0	4.0000	6.9278
15	15.0	9.3333	8.1668	65	67.0	11.0000	6.9171
16	16.0	8.2857	8.1102	66	68.0	6.0000	6.9065
17	17.0	8.2000	8.0573	67	69.0	8.0000	6.8961
18	18.0	8.0000	7.9078	68	70.0	6.5000	6.8859
19	19.0	8.1429	7.8613	69	71.0	7.0000	6.8758
20	20.0	8.5000	7.9174	70	72.0	10.0000	6.8658
21	21.0	8.0000	7.8758	71	73.0	5.0000	6.8464
22	22.0	7.0000	7.8364	72	74.0	8.6667	6.8275
23	23.0	7.1667	7.7950	73	75.0	11.0000	6.8182
24	24.0	7.7500	7.7533	74	76.0	5.0000	6.8001
25	25.0	6.0000	7.7292	75	77.0	10.0000	6.7853
26	26.0	7.6000	7.6966	76	78.0	11.0000	6.7766
27	27.0	8.3333	7.6653	77	79.0	6.0000	6.7403
28	28.0	7.3333	7.6353	78	80.0	10.0000	6.7242
29	29.0	6.0000	7.6065	79	81.0	9.0000	6.7084
30	30.0	8.6667	7.5787	80	82.0	6.0000	6.7007
31	31.0	9.0000	7.5520	81	83.0	7.0000	6.6707
32	32.0	6.3333	7.5262	82	84.0	5.0000	6.6634
33	33.0	8.6667	7.5013	83	85.0	6.0000	6.6490
34	34.0	7.0000	7.4772	84	86.0	6.0000	6.6281
35	35.0	7.3333	7.4538	85	87.0	8.0000	6.6145
36	36.0	6.5000	7.4312	86	88.0	5.0000	6.5946
37	37.0	5.5000	7.4092	87	89.0	6.0000	6.5690
38	38.0	7.2500	7.3880	88	90.0	7.0000	6.5566
39	39.0	4.5000	7.3674	89	91.0	7.0000	6.5354
40	40.0	7.0000	7.3473	90	92.0	6.0000	6.5266
41	41.0	7.5000	7.3277	91	93.0	4.0000	6.5092
42	42.0	9.0000	7.3087	92	94.0	5.0000	6.4979
43	43.0	8.0000	7.2902	93	95.0	6.5000	6.4705
44	44.0	11.5000	7.2722	94	96.0	6.0000	6.4552
45	45.0	8.5000	7.2374	95	97.0	9.0000	6.4395
46	46.0	5.0000	7.2206	96	98.0	4.5000	6.4246
47	47.0	7.8333	7.2043	97	99.0	6.0000	6.4094
48	48.0	4.0000	7.1863	98	100.0	5.0000	6.3940
49	49.0	5.0000	7.1573	99	101.0	7.0000	6.3741
50	50.0	5.0000	7.1424	100	102.0	7.0000	6.3541

Tabelle A 4 (Forts.)

i	x_i	y_i	y_i
		e	t
101	180.0	277.0000	341.8500
102	193.0	331.5000	374.8964
103	199.0	369.0000	390.6162
104	200.0	368.0000	393.2654
105	209.0	353.0000	417.4902
106	215.0	348.0000	434.0277
107	217.0	416.0000	439.6102
108	222.0	350.0000	453.7215
109	257.0	657.0000	558.9462
110	277.0	611.0000	624.4253
111	278.0	483.0000	627.8067
112	285.0	548.0000	651.7702
113	296.0	762.0000	690.4835
114	299.0	567.0000	701.2701
115	300.0	791.0000	704.6878
116	301.0	665.0000	708.5163
117	304.0	778.0000	719.4689
118	319.0	668.0000	775.7586
119	339.0	719.0000	854.9061
120	365.0	1014.0000	965.1817
121	373.0	1090.0000	1008.8734
122	447.0	1512.0000	1374.2194
123	449.0	1513.0000	1385.4734
124	458.0	1781.0000	1436.9315

Tabelle A 5 (Forts.)

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
101	142.0	10.0000	6.4092	151	719.0	6.0000	5.3809
102	144.0	5.0000	6.3995	152	762.0	5.0000	5.3473
103	152.0	7.0000	6.3623	153	778.0	4.0000	5.3353
104	155.0	7.0000	6.3490	154	791.0	4.0000	5.3258
105	159.0	10.0000	6.3315	155	868.0	3.0000	5.2727
106	163.0	5.0000	6.3146	156	1014.0	4.0000	5.1851
107	167.0	5.0000	6.2981	157	1090.0	5.0000	5.1446
108	168.0	10.0000	6.2941	158	1513.0	6.0000	4.9661
109	171.0	6.0000	6.2821	159	1781.0	5.0000	4.8796
110	173.0	6.0000	6.2742	160	1912.0	4.0000	4.8424
111	175.0	7.0000	6.2664				
112	179.0	4.5000	6.2512				
113	182.0	7.0000	6.2400				
114	184.0	7.0000	6.2325				
115	187.0	6.0000	6.2216				
116	190.0	4.0000	6.2111				
117	192.0	11.0000	6.2041				
118	201.0	5.0000	6.1735				
119	203.0	8.0000	6.1669				
120	205.0	7.0000	6.1604				
121	208.0	9.0000	6.1444				
122	219.0	8.0000	6.1167				
123	220.0	6.0000	6.1137				
124	221.0	4.0000	6.1107				
125	225.0	8.0000	6.0989				
126	234.0	6.0000	6.0732				
127	246.0	3.0000	6.0352				
128	265.0	6.0000	5.9923				
129	270.0	10.0000	5.9802				
130	277.0	5.0000	5.9637				
131	287.0	8.0000	5.9410				
132	290.0	7.0000	5.9343				
133	314.0	12.0000	5.8836				
134	321.0	6.0000	5.8503				
135	342.0	10.0000	5.8297				
136	348.0	5.0000	5.8188				
137	353.0	8.0000	5.8098				
138	367.0	5.0000	5.7855				
139	369.0	7.0000	5.7832				
140	390.0	6.0000	5.7477				
141	416.0	6.0000	5.7079				
142	427.0	6.0000	5.6918				
143	483.0	5.0000	5.6167				
144	548.0	6.0000	5.5408				
145	557.0	5.0000	5.4762				
146	567.0	6.0000	5.4335				
147	567.0	6.0000	5.4246				
148	685.0	5.0000	5.4091				

Tabelle A 6: Die Polylexte in Abhängigkeit von der Frequenz

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
1	1.0	1.2114	1.0269	51	53.0	2.5000	3.1587
2	2.0	1.3586	1.2495	52	54.0	2.0000	3.1754
3	3.0	1.5250	1.4014	53	55.0	6.5000	3.1920
4	4.0	1.5893	1.5203	54	56.0	3.0000	3.2083
5	5.0	1.8140	1.6194	55	57.0	3.0000	3.2244
6	6.0	1.7168	1.7051	56	58.0	4.5000	3.2403
7	7.0	2.0476	1.7811	57	59.0	3.0000	3.2560
8	8.0	2.0556	1.8497	58	61.0	4.0000	3.2869
9	9.0	2.4615	1.9124	59	62.0	2.8000	3.3020
10	10.0	2.8333	1.9703	60	64.0	2.0000	3.3318
11	11.0	2.6250	2.0242	61	65.0	5.0000	3.3465
12	12.0	2.4286	2.0746	62	66.0	1.0000	3.3610
13	13.0	2.0000	2.1222	63	68.0	4.0000	3.3895
14	14.0	2.8000	2.1671	64	69.0	8.0000	3.4036
15	15.0	1.6667	2.2099	65	70.0	1.0000	3.4174
16	16.0	3.0000	2.2506	66	71.0	5.0000	3.4312
17	17.0	3.6000	2.2896	67	72.0	4.0000	3.4448
18	18.0	2.5000	2.3265	68	73.0	3.0000	3.4582
19	19.0	2.0000	2.3626	69	74.0	2.0000	3.4716
20	20.0	2.5000	2.3973	70	75.0	2.0000	3.4848
21	21.0	3.0000	2.4306	71	77.0	4.5000	3.5108
22	22.0	1.5000	2.4629	72	79.0	3.3333	3.5364
23	23.0	3.5667	2.4940	73	80.0	3.0000	3.5490
24	24.0	3.0000	2.5243	74	82.0	4.0000	3.5735
25	25.0	4.5000	2.5536	75	86.0	3.0000	3.6224
26	26.0	2.4000	2.5821	76	87.0	5.0000	3.6343
27	27.0	3.6667	2.6098	77	89.0	1.0000	3.6577
28	28.0	2.6000	2.6368	78	91.0	2.0000	3.6808
29	29.0	2.5000	2.6631	79	93.0	2.5000	3.7035
30	30.0	2.3333	2.6868	80	94.0	2.0000	3.7147
31	31.0	3.3333	2.7139	81	98.0	5.0000	3.7586
32	32.0	4.3333	2.7384	82	99.0	2.0000	3.7656
33	33.0	2.0000	2.7623	83	101.0	6.0000	3.7910
34	34.0	3.3000	2.7856	84	104.0	5.0000	3.8226
35	35.0	3.0000	2.8087	85	106.0	1.0000	3.8432
36	36.0	2.2500	2.8312	86	109.0	4.0000	3.8737
37	37.0	3.0000	2.8532	87	113.0	3.0000	3.9134
38	38.0	4.7500	2.8748	88	115.0	6.0000	3.9329
39	39.0	2.5000	2.8960	89	116.0	6.0000	3.9617
40	40.0	2.5000	2.9169	90	120.0	1.0000	3.9805
41	41.0	2.5000	2.9373	91	123.0	8.0000	4.0085
42	42.0	3.0000	2.9574	92	124.0	4.5000	4.0177
43	43.0	1.5000	2.9772	93	125.0	4.5000	4.0268
44	44.0	1.5000	2.9966	94	126.0	5.0000	4.0717
45	45.0	2.5000	3.0346	95	131.0	2.0000	4.0805
46	46.0	5.5000	3.0531	96	132.0	6.5000	4.0894
47	47.0	2.8333	3.0713	97	133.0	6.0000	4.0981
48	48.0	3.0000	3.0893	98	134.0	10.0000	4.1068
49	49.0	8.0000	3.1245	99	137.0	4.0000	4.1326
50	52.0	2.5000	3.1417	100	141.0	2.0000	4.1664

Tabelle A 6 (Forts.)

i	x _i	y _i _e	y _i _t	i	x _i	y _i _e	y _i _t
101	142.0	1.0000	4.3748	151	719.0	9.0000	6.6068
102	144.0	5.0000	4.1913	152	762.0	5.0000	6.7153
103	152.0	4.0000	4.2559	153	778.0	6.0000	6.7559
104	155.0	3.0000	4.2796	154	791.0	6.0000	6.7876
105	159.0	6.0000	4.3105	155	868.0	9.0000	6.9684
106	163.0	8.0000	4.3409	156	1014.0	13.0000	7.2819
107	167.0	6.0000	4.3708	157	1050.0	19.0000	7.4323
108	168.0	6.0000	4.3788	158	1513.0	8.0000	8.1551
109	171.0	6.0000	4.4002	159	1761.0	14.0000	8.5403
110	173.0	2.0000	4.4147	160	1912.0	10.0000	8.7135
111	175.0	2.0000	4.4291				
112	179.0	5.0000	4.4575				
113	182.0	1.0000	4.4785				
114	184.0	3.0000	4.4924				
115	187.0	5.0000	4.5130				
116	190.0	7.0000	4.5334				
117	192.0	5.0000	4.5468				
118	200.0	5.0000	4.5997				
119	201.0	11.0000	4.6052				
120	203.0	1.0000	4.6191				
121	205.0	2.0000	4.6319				
122	210.0	5.0000	4.6636				
123	219.0	8.0000	4.7193				
124	220.0	6.0000	4.7254				
125	221.0	8.0000	4.7315				
126	225.0	5.0000	4.7536				
127	234.0	1.0000	4.8086				
128	245.0	4.0000	4.8654				
129	255.0	6.0000	4.9810				
130	270.0	3.0000	5.0074				
131	277.0	14.0000	5.0438				
132	287.0	6.0000	5.0547				
133	290.0	5.0000	5.1097				
134	314.0	6.0000	5.2260				
135	331.0	16.0000	5.3045				
136	342.0	6.0000	5.3338				
137	348.0	3.0000	5.3602				
138	353.0	5.0000	5.4020				
139	367.0	5.0000	5.4616				
140	368.0	5.0000	5.4660				
141	369.0	4.0000	5.4702				
142	390.0	14.0000	5.5566				
143	416.0	6.0000	5.6590				
144	427.0	6.0000	5.7009				
145	463.0	6.0000	5.9033				
146	468.0	21.0000	6.1180				
147	611.0	6.0000	6.3093				
148	657.0	8.0000	6.4403				
149	667.0	5.0000	6.4675				
150	685.0	7.0000	6.5168				

Tabelle A 7: Die Polylexie in Abhängigkeit von der Länge

i	x _i	y _i _e	y _i _t
1	3.0	54.2222	123.7918
2	4.0	67.5789	62.2276
3	5.0	60.7534	35.4995
4	6.0	43.9057	23.6035
5	7.0	26.8125	16.3274
6	8.0	14.9646	11.8650
7	9.0	8.6415	8.9530
8	10.0	9.0075	6.9594
9	11.0	5.6364	5.5413
10	12.0	4.3303	4.5005
11	13.0	2.8989	3.7166
12	14.0	1.5190	3.1132
13	15.0	1.6786	2.6398
14	16.0	2.9444	2.2623
15	17.0	1.1316	1.9571
16	18.0	1.7333	1.7071
17	19.0	1.1071	1.5001
18	20.0	1.1667	1.3270
19	21.0	1.2143	1.1809
20	22.0	1.2222	1.0566
21	23.0	1.1667	0.9500
22	24.0	1.0000	0.8581
23	25.0	1.0000	0.7783
24	26.0	1.0000	0.7087

Tabelle A 8: Die Frequenz in Abhängigkeit von der Polylexie

i	x_i	y_{i_e}	y_{i_t}
1	1.0	4.7881	5.5048
2	2.0	15.5045	18.6104
3	3.0	30.6048	37.9499
4	4.0	46.3333	62.9175
5	5.0	132.0000	93.1272
6	6.0	184.5294	128.2997
7	7.0	145.5000	168.2189
8	8.0	268.7692	212.7095
9	9.0	547.3333	261.6254
10	10.0	689.3333	314.8415
11	11.0	119.5000	372.2489
12	13.0	1014.0000	499.2642
13	14.0	816.0000	566.7091
14	16.0	331.0000	719.1217
15	19.0	1090.0000	972.6573
16	21.0	548.0000	1159.6965

Tabelle A 9: Die Länge in Abhängigkeit von der Polylexie

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
1	1.0	12.8321	11.4462	51	51.0	7.1579	6.000
2	2.0	11.5065	10.5861	52	52.0	7.1382	1.3333
3	3.0	10.5072	10.1107	53	53.0	7.1188	8.2500
4	4.0	9.6329	9.7848	54	54.0	7.0997	9.5000
5	5.0	9.7745	9.5379	55	55.0	7.0809	7.7500
6	6.0	9.1667	9.3399	56	56.0	7.0625	6.0000
7	7.0	8.8561	9.1748	57	57.0	7.0443	6.3333
8	8.0	9.2165	9.0334	58	58.0	7.0265	7.7500
9	9.0	8.8913	8.9099	59	59.0	7.0089	6.0000
10	10.0	8.1688	8.8003	60	60.0	6.9916	6.3333
11	11.0	8.5593	8.7018	61	61.0	6.9745	6.0000
12	12.0	8.9111	8.6124	62	62.0	6.9577	5.3333
13	13.0	7.4000	8.5306	63	63.0	6.9411	7.3333
14	14.0	7.5556	8.4552	64	64.0	6.9248	5.5000
15	15.0	8.6216	8.3852	65	65.0	6.9087	9.0000
16	16.0	7.8519	8.3199	66	66.0	6.8928	7.6000
17	17.0	9.2800	8.2588	67	67.0	6.8772	5.8333
18	18.0	8.3333	8.2013	68	68.0	6.8617	5.5000
19	19.0	8.1364	8.1470	69	69.0	6.8465	7.0000
20	20.0	7.1765	8.0956	70	70.0	6.8314	8.0000
21	21.0	8.5625	8.0468	71	71.0	6.8167	9.5000
22	22.0	8.3846	8.0003	72	72.0	6.8018	9.5000
23	23.0	8.2000	7.9559	73	73.0	6.7873	6.5000
24	24.0	7.1875	7.9134	74	74.0	6.7730	5.0000
25	25.0	8.1333	7.8727	75	75.0	6.7588	6.5000
26	26.0	7.7000	7.8336	76	76.0	6.7448	8.2500
27	27.0	8.7059	7.7959	77	77.0	6.7309	7.3333
28	28.0	6.5455	7.7597	78	78.0	6.7172	6.0000
29	29.0	8.0000	7.7247	79	79.0	6.7037	8.0000
30	30.0	6.0000	7.6909	80	80.0	6.6893	5.0000
31	31.0	8.6364	7.6582	81	81.0	6.6754	4.3333
32	32.0	8.7143	7.6265	82	82.0	6.6639	6.3333
33	33.0	7.0000	7.5958	83	83.0	6.6516	5.5000
34	34.0	6.6250	7.5660	84	84.0	6.6398	5.5000
35	35.0	6.2857	7.5370	85	85.0	6.6287	7.6667
36	36.0	8.2308	7.5089	86	86.0	6.6163	6.6667
37	37.0	7.5000	7.4814	87	87.0	6.6048	6.4000
38	38.0	5.5000	7.4547	88	88.0	6.5933	6.7500
39	39.0	8.6000	7.4287	89	89.0	6.5819	9.0000
40	40.0	6.9167	7.4033	90	90.0	6.5695	5.0000
41	41.0	8.0000	7.3785	91	91.0	6.5584	6.3333
42	42.0	7.8000	7.3543	92	92.0	6.5474	7.0000
43	43.0	7.7500	7.3306	93	93.0	6.5367	5.0000
44	44.0	8.7000	7.3074	94	94.0	6.5263	5.0000
45	45.0	6.1667	7.2848	95	95.0	6.5163	4.0000
46	46.0	8.2000	7.2626	96	96.0	6.5067	4.0000
47	47.0	7.1667	7.2408	97	97.0	6.4974	9.0000
48	48.0	6.6667	7.2195	98	98.0	6.4884	8.0000
49	49.0	8.7143	7.1986	99	99.0	6.4795	5.0000
50	50.0	7.7500	7.1781	100	100.0	6.4707	6.3321

Tabelle A 9 (Forts.)

i	x _i	y _i _e	i	x _i	y _i _e	y _i _t
101	115.0	7.5000	151	238.0	9.0000	5.3921
102	116.0	6.5000	152	243.0	6.0000	5.3636
103	117.0	6.0000	153	257.0	5.0000	5.2860
104	118.0	6.6667	154	267.0	4.0000	5.2324
105	121.0	6.0000	155	272.0	6.0000	5.2060
106	122.0	8.6667	156	277.0	6.0000	5.1801
107	123.0	10.0000	157	278.0	5.0000	5.1749
108	124.0	6.0000	158	285.0	6.0000	5.1391
109	126.0	5.0000	159	287.0	5.0000	5.1290
110	120.0	7.3333	160	295.0	4.0000	5.0890
111	130.0	5.0000	161	296.0	5.0000	5.0841
112	132.0	5.0000	162	299.0	6.0000	5.0693
113	133.0	12.0000	163	300.0	5.5000	5.0644
114	134.0	4.0000	164	301.0	5.0000	5.0595
115	137.0	4.0000	165	304.0	4.0000	5.0449
116	140.0	5.0000	166	308.0	5.0000	5.0256
117	141.0	9.0000	167	319.0	3.0000	4.9734
118	142.0	8.0000	168	323.0	4.0000	4.9547
119	144.0	5.0000	169	333.0	5.0000	4.9086
120	147.0	7.0000	170	334.0	6.0000	4.9040
121	149.0	6.6667	171	336.0	6.0000	4.8814
122	150.0	6.5000	172	365.0	4.0000	4.7670
123	152.0	12.0000	173	373.0	5.0000	4.7330
124	154.0	9.0000	174	393.0	4.0000	4.6499
125	157.0	6.0000	175	402.0	3.5000	4.6134
126	161.0	6.0000	176	416.0	4.0000	4.5577
127	162.0	5.5000	177	447.0	4.0000	4.4388
128	167.0	6.3333	178	449.0	6.0000	4.4713
129	169.0	6.0000	179	458.0	5.0000	4.3980
130	173.0	8.0000	180	490.0	3.0000	4.2827
131	174.0	5.0000	181	493.0	5.0000	4.2722
132	175.0	7.0000	182	498.0	6.0000	4.2547
133	180.0	5.0000	183	500.0	4.0000	4.2478
134	182.0	3.0000				
135	185.0	5.0000				
136	192.0	7.0000				
137	193.0	5.5000				
138	194.0	3.5000				
139	197.0	4.0000				
140	199.0	7.0000				
141	200.0	4.5000				
142	209.0	8.0000				
143	211.0	6.0000				
144	215.0	5.0000				
145	217.0	6.5000				
146	219.0	7.0000				
147	220.0	7.0000				
148	222.0	6.0000				
149	227.0	6.0000				
150	232.0	4.0000				

Tabelle A 10: Die Polylexie in Abhängigkeit von der Polytextie

i	x _i	y _i _e	y _i _t	i	x _i	y _i _e	y _i _t
1	1.0	1.2180	1.1133	51	54.0	3.0000	3.6314
2	2.0	1.4326	1.3807	52	55.0	2.0000	3.8523
3	3.0	1.6667	1.5654	53	56.0	8.0000	3.8748
4	4.0	2.0256	1.7112	54	60.0	1.5000	3.9585
5	5.0	2.1842	1.8337	55	62.0	4.5000	3.9998
6	6.0	2.1818	1.9402	56	63.0	2.3333	4.0188
7	7.0	2.6667	2.0351	57	64.0	6.0000	4.0334
8	8.0	1.8750	2.1210	58	66.0	3.5000	4.0771
9	9.0	2.6667	2.1998	59	67.0	6.0000	4.0961
10	10.0	5.0000	2.2727	60	68.0	5.5000	4.1149
11	11.0	2.5714	2.3408	61	69.0	7.0000	4.1336
12	12.0	2.3571	2.4047	62	70.0	2.0000	4.1521
13	13.0	3.8571	2.4651	63	74.0	5.5000	4.2241
14	14.0	2.5223	2.5223	64	76.0	5.0000	4.2592
15	15.0	2.7143	2.5768	65	77.0	2.0000	4.2764
16	16.0	3.8571	2.6289	66	78.0	1.0000	4.2936
17	17.0	3.0000	2.6786	67	81.0	8.0000	4.3440
18	18.0	3.1111	2.7265	68	82.0	4.0000	4.3606
19	19.0	4.3333	2.7735	69	85.0	3.6667	4.4094
20	20.0	2.0000	2.8159	70	87.0	5.0000	4.4412
21	21.0	2.5000	2.8598	71	89.0	8.0000	4.4726
22	22.0	2.5000	2.9013	72	90.0	8.0000	4.4881
23	23.0	2.3333	2.9415	73	94.0	3.0000	4.5490
24	24.0	5.5556	2.9805	74	96.0	5.0000	4.5787
25	25.0	1.3333	3.0583	75	97.0	5.0000	4.5934
26	27.0	3.5000	3.0913	76	100.0	9.0000	4.6370
27	26.0	4.5000	3.1263	77	102.0	4.0000	4.6655
28	29.0	2.8000	3.1604	78	105.0	8.0000	4.7076
29	30.0	2.2500	3.1938	79	106.0	2.0000	4.7214
30	31.0	4.0000	3.2264	80	110.0	4.0000	4.7759
31	32.0	3.0000	3.2583	81	111.0	4.0000	4.7893
32	33.0	2.0000	3.2895	82	116.0	1.0000	4.8351
33	34.0	6.5000	3.3200	83	117.0	5.0000	4.8680
34	35.0	4.3333	3.3459	84	118.0	5.5000	4.8808
35	36.0	3.2500	3.3753	85	122.0	3.0000	4.9315
36	37.0	3.0000	3.4081	86	123.0	6.0000	4.9440
37	38.0	8.0000	3.4364	87	124.0	1.0000	4.9564
38	39.0	4.0000	3.4641	88	125.0	7.0000	5.0174
39	40.0	5.0000	3.4914	89	130.0	2.0000	5.0295
40	41.0	2.0000	3.5182	90	132.0	11.0000	5.0533
41	42.0	4.0000	3.5445	91	133.0	6.0000	5.0651
42	44.0	3.5000	3.5980	92	137.0	8.0000	5.1118
43	45.0	2.5000	3.6211	93	140.0	3.0000	5.1462
44	46.0	3.5000	3.6458	94	144.0	2.0000	5.1913
45	47.0	3.5000	3.6702	95	149.0	6.0000	5.2465
46	48.0	4.0000	3.6942	96	150.0	1.0000	5.2574
47	49.0	5.0000	3.7178	97	154.0	5.0000	5.3004
48	50.0	4.0000	3.7412	98	167.0	5.5000	5.3351
49	51.0	2.0000	3.7642	99	173.0	6.0000	5.4948
50	52.0	3.0000	3.7865	100	175.0	5.0000	5.5144

Tabelle A 10 (Forts.)

i	x _i	y _i e	y _i t
101	190.0	14.0000	5.5627
102	193.0	16.0000	5.6842
103	199.0	4.0000	5.7383
104	200.0	5.0000	5.7472
105	209.0	5.0000	5.8261
106	215.0	3.0000	5.8774
107	217.0	6.0000	5.8543
108	222.0	14.0000	5.9360
109	257.0	8.0000	6.2113
110	277.0	6.0000	6.3572
111	278.0	6.0000	6.3643
112	265.0	21.0000	6.4135
113	296.0	5.0000	6.4891
114	299.0	5.0000	6.5094
115	300.0	6.0000	6.5162
116	301.0	7.0000	6.5229
117	304.0	6.0000	6.5429
118	319.0	9.0000	6.6413
119	339.0	9.0000	6.7675
120	355.0	13.0000	6.9242
121	373.0	19.0000	6.9708
122	447.0	10.0000	7.3727
123	449.0	8.0000	7.3829
124	458.0	14.0000	7.4284

Tabelle A 11: Die Frequenz in Abhängigkeit von der Polytextile:
Wortformen insgesamt

i	x _i	y _i e	y _i t	i	x _i	y _i e	y _i t
1	1.0	1.3965	1.7724	51	51.0	79.6250	106.4431
2	2.0	2.7702	3.6059	52	52.0	93.1250	108.7406
3	3.0	4.4235	5.4656	53	53.0	94.6667	111.0457
4	4.0	6.0548	7.3472	54	54.0	98.4000	113.3584
5	5.0	8.0157	9.2440	55	55.0	2372.6000	115.6787
6	6.0	8.6530	11.1552	56	56.0	99.4286	120.3420
7	7.0	10.9950	13.0792	57	57.0	122.0000	122.6851
8	8.0	12.8392	15.0151	58	58.0	124.8753	125.0357
9	9.0	14.5639	16.9622	59	59.0	25.8000	127.3940
10	10.0	15.6514	18.9200	60	60.0	109.0000	129.7599
11	11.0	18.7262	20.8880	61	61.0	76.0000	132.1334
12	12.0	23.0149	22.8658	62	62.0	244.0000	134.5146
13	13.0	22.9383	24.8532	63	63.0	126.2500	136.9033
14	14.0	20.9333	26.8499	64	64.0	108.1667	139.2997
15	15.0	23.3043	28.8557	65	65.0	110.8000	141.7038
16	16.0	26.1522	30.8704	66	66.0	81.0000	144.1155
17	17.0	35.1316	32.8938	67	67.0	692.5000	146.5348
18	18.0	30.0222	34.9259	68	68.0	133.0000	148.9619
19	19.0	32.9048	36.9665	69	69.0	102.7500	151.3966
20	20.0	48.2143	39.0155	70	70.0	185.5000	153.8390
21	21.0	35.0000	41.0729	71	71.0	103.0000	156.2891
22	22.0	39.7222	43.1384	72	72.0	116.5000	158.7469
23	23.0	498.9048	45.2121	73	73.0	250.0000	161.2124
24	24.0	41.7846	47.2940	74	74.0	106.0000	163.6856
25	25.0	372.4091	49.3838	75	75.0	550.0000	166.1666
26	26.0	35.2222	51.4817	76	76.0	185.5000	168.6553
27	27.0	48.0250	53.5875	77	77.0	125.0000	171.1517
28	28.0	930.2857	55.7012	78	78.0	157.0000	173.6559
29	29.0	60.7500	57.8227	79	79.0	114.0000	176.1679
30	30.0	40.6500	59.9521	80	80.0	609.0000	178.6877
31	31.0	55.5000	62.0893	81	81.0	133.0000	181.2153
32	32.0	69.7692	64.2342	82	82.0	146.0000	183.7506
33	33.0	52.6667	66.3869	83	83.0	346.5000	186.2938
34	34.0	74.6250	68.5473	84	84.0	169.2000	188.8449
35	35.0	47.7143	70.7155	85	85.0	121.0000	191.4037
36	36.0	56.8235	72.8913	86	86.0	144.0000	193.9704
37	37.0	64.5556	75.0747	87	87.0	130.0000	199.1275
38	38.0	162.6000	77.2659	88	88.0	137.0000	204.3161
39	39.0	68.8000	79.4646	89	89.0	134.2000	212.1584
40	40.0	73.0000	81.6710	90	90.0	131.0000	214.7884
41	41.0	69.6000	83.8850	91	91.0	182.0000	217.4263
42	42.0	61.6667	86.1067	92	92.0	137.0000	228.0580
43	43.0	75.3333	88.3359	93	93.0	127.0000	230.7360
44	44.0	78.6000	90.5727	94	94.0	233.6667	238.8183
45	45.0	82.0000	92.8172	95	95.0	176.0000	241.5285
46	46.0	59.0000	95.0692	96	96.0	313.0000	244.2469
47	47.0	153.6000	97.3238	97	97.0	168.5000	249.7080
48	48.0	117.2500	99.5960	98	98.0	137.0000	252.4567
49	49.0	208.8333	101.8708	99	99.0	231.4000	255.2016
50	50.0	93.5000	104.1531	100	100.0		

Tabelle A 11 (Forts.)

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
101	111.0	174.5000	257.9407	151	295.0	535.0000	923.5563
102	113.0	1720.5000	263.5035	152	300.0	791.0000	946.6380
103	114.0	178.0000	266.2872	153	308.0	905.0000	984.1966
104	115.0	166.0000	269.0791	154	323.0	1028.0000	1056.7434
105	116.0	367.5000	271.8793	155	333.0	729.0000	1106.6833
106	117.0	407.0000	274.8797	156	393.0	1183.0000	1424.4169
107	118.0	222.5000	277.5045	157	402.0	1137.0000	1487.9656
108	119.0	216.0000	280.3295	158	416.0	1772.0000	1573.6730
109	121.0	237.0000	286.0045	159	485.0	3863.0000	2047.7857
110	126.0	243.0000	300.3383	160	487.0	6091.0000	2056.5555
111	128.0	282.0000	306.1307				
112	129.0	179.0000	309.0396				
113	135.0	203.0000	326.6712				
114	136.0	317.0000	329.6397				
115	137.0	255.0000	332.6168				
116	140.0	200.0000	341.5997				
117	144.0	228.0000	353.6982				
118	147.0	352.0000	362.8636				
119	148.0	255.0000	365.9363				
120	154.0	253.0000	384.5575				
121	156.0	385.0000	390.8355				
122	157.0	243.0000	393.9879				
123	158.0	303.0000	397.1493				
124	159.0	1328.0000	400.3197				
125	160.0	276.0000	403.4990				
126	161.0	412.0000	406.6873				
127	162.0	1073.0000	409.8847				
128	165.0	304.0000	419.5311				
129	166.0	428.0000	422.7647				
130	174.0	2079.5000	448.9636				
131	176.0	262.0000	455.6056				
132	180.0	278.0000	469.0014				
133	182.0	252.0000	475.7555				
134	186.0	645.0000	489.3768				
135	190.0	336.0000	503.1502				
136	192.0	302.0000	510.0942				
137	193.0	1375.3333	513.5806				
138	194.0	530.0000	517.0767				
139	195.0	2817.0000	534.7023				
140	202.0	431.0000	545.3945				
141	211.0	311.0000	578.0036				
142	215.0	348.0000	592.7556				
143	217.0	411.0000	600.1421				
144	219.0	320.0000	607.6690				
145	225.0	741.0000	630.3447				
146	232.0	437.0000	657.2685				
147	237.0	423.5000	676.8129				
148	264.0	585.0000	786.9852				
149	271.0	434.0000	816.8599				
150	284.0	786.0000	873.8217				

Tabelle A 12: Die Frequenz in Abhängigkeit von der Polytexitle: Lemmata insgesamt

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
1	1.0	1.3829	2.1685	51	51.0	77.0000	73.8930
2	2.0	2.7762	3.9178	52	52.0	75.0000	75.4139
3	3.0	4.2178	5.5464	53	53.0	80.7500	76.9420
4	4.0	5.9233	7.1058	54	54.0	86.7500	78.4773
5	5.0	7.3218	8.6191	55	55.0	75.1250	80.0200
6	6.0	8.8889	10.0991	56	56.0	72.3333	81.5702
7	7.0	10.4892	11.5540	57	57.0	115.3333	83.1279
8	8.0	11.3402	12.9894	58	58.0	88.0000	84.6933
9	9.0	14.5870	14.4095	59	59.0	99.0000	86.2664
10	10.0	14.6494	15.8174	60	60.0	72.8333	87.8473
11	11.0	15.8814	17.2156	61	61.0	86.5000	89.4361
12	12.0	20.1333	18.6061	62	62.0	92.6667	91.0329
13	13.0	21.3429	19.9904	63	63.0	114.0000	92.6378
14	14.0	19.9333	21.3699	64	64.0	113.0000	94.2509
15	15.0	23.6216	22.7458	65	65.0	222.0000	95.8721
16	16.0	26.8148	24.1191	66	66.0	90.8000	97.5018
17	17.0	24.2400	25.4905	67	67.0	111.8333	99.1398
18	18.0	28.1667	26.8609	68	68.0	115.0000	100.7863
19	19.0	31.1818	28.2309	69	69.0	163.0000	102.4413
20	20.0	27.5294	29.6010	70	70.0	111.0000	104.1050
21	21.0	35.0000	30.9719	71	72.0	106.5000	107.4567
22	22.0	34.5769	32.3440	72	73.0	88.0000	109.1488
23	23.0	34.2000	33.7176	73	74.0	159.0000	110.8478
24	24.0	40.2500	35.0933	74	75.0	112.5000	112.5558
25	25.0	34.2667	36.4714	75	76.0	116.5000	114.2730
26	26.0	40.7000	37.8521	76	77.0	105.0000	115.9997
27	27.0	38.2941	39.2359	77	78.0	120.0000	117.7349
28	28.0	42.8182	40.6229	78	79.0	110.0000	119.4797
29	29.0	41.0000	42.0135	79	81.0	159.0000	122.9977
30	30.0	60.3333	43.4080	80	82.0	166.0000	124.7710
31	31.0	44.3636	44.8064	81	85.0	168.6667	130.1488
32	32.0	47.4286	46.2091	82	87.0	151.3333	133.7829
33	33.0	51.9000	47.6165	83	89.0	139.7500	137.4570
34	34.0	52.6250	49.0281	84	90.0	123.5000	139.3091
35	35.0	52.5714	50.4447	85	92.0	138.0000	143.5440
36	36.0	48.8750	53.2932	86	94.0	176.3333	146.8202
37	37.0	68.7500	54.7254	87	95.0	124.0000	148.7240
38	38.0	68.7500	56.1631	88	96.0	147.7500	150.6383
39	39.0	60.4000	56.1631	89	97.0	188.5000	152.5632
40	40.0	72.5000	57.6055	90	99.0	191.0000	156.4451
41	41.0	68.5000	59.0556	91	100.0	177.0000	158.4023
42	42.0	74.1000	60.5106	92	101.0	167.0000	160.3733
43	43.0	87.2500	61.9718	93	102.0	137.0000	162.3492
44	44.0	76.8000	63.4390	94	105.0	167.0000	168.3528
45	45.0	63.3333	64.9126	95	106.0	551.0000	170.3763
46	46.0	68.4000	66.3926	96	107.0	167.0000	172.4111
47	47.0	76.0000	67.8792	97	108.0	151.0000	174.4574
48	48.0	77.3333	69.3724	98	109.0	248.0000	176.5151
49	49.0	71.4286	70.8724	99	110.0	157.0000	178.5843
50	50.0	77.7500	72.3792	100	111.0	168.6667	180.6651

Tabellle A 12 (Forts.)

i	x_i	y_i	y_i	x_i	y_i	y_i
101	115.0	173.0000	189.1059	238.0	439.000	566.8889
102	116.0	160.5000	191.2457	243.0	422.000	588.3880
103	117.0	187.0000	193.3977	257.0	657.000	651.7220
104	118.0	226.0000	195.5617	267.0	593.000	700.0186
105	121.0	164.0000	202.1271	272.0	435.000	725.1652
106	122.0	207.0000	204.3403	277.0	611.000	751.0000
107	123.0	342.0000	205.5659	278.0	483.000	756.2511
108	124.0	220.6667	208.8041	285.0	548.000	793.8108
109	126.0	236.0000	213.3184	287.0	792.000	804.8043
110	129.0	179.0000	220.1857	295.0	535.000	849.9768
111	130.0	192.6667	222.5007	296.0	762.000	855.7606
112	132.0	201.0000	227.1699	299.0	667.000	873.2990
113	133.0	314.0000	229.5243	300.0	701.000	879.2079
114	134.0	483.0000	231.8920	301.0	685.000	885.1485
115	137.0	260.5000	239.0755	304.0	778.000	903.1515
116	140.0	200.0000	246.3814	308.0	905.000	927.6315
117	141.0	257.0000	248.8442	319.0	868.000	997.6720
118	142.0	206.0000	251.3209	323.0	1028.000	1024.1722
119	144.0	216.5000	256.3164	333.0	729.000	1292.9250
120	147.0	204.0000	263.9159	334.0	818.000	1100.0065
121	149.0	291.3333	269.0539	339.0	719.000	1135.9590
122	150.0	233.5000	271.6447	365.0	1014.000	1338.3132
123	152.0	234.0000	276.8702	373.0	1090.000	1407.1148
124	154.0	210.0000	282.1548	393.0	1235.000	1590.5472
125	157.0	248.0000	290.1940	402.0	1137.000	1679.4866
126	161.0	245.0000	301.1258	416.0	1172.000	1826.2982
127	162.0	225.5000	303.8974	447.0	1912.000	2191.3589
128	167.0	370.6667	317.9909	449.0	1513.000	2216.9498
129	169.0	355.0000	323.7399	458.0	1781.000	2335.3767
130	173.0	287.0000	335.4330	490.0	6098.000	2902.8608
131	174.0	273.0000	338.3974	493.0	3884.000	2850.6695
132	175.0	367.0000	341.3785	498.0	7584.000	2931.0657
133	180.0	277.0000	356.5367	500.0	64275.000	2965.0584
134	182.0	252.0000	362.7195			
135	185.0	280.0000	372.1243			
136	192.0	302.5000	394.6905			
137	193.0	367.0000	397.9866			
138	194.0	304.5000	401.3012			
139	197.0	437.0000	411.3562			
140	199.0	369.0000	418.1533			
141	200.0	524.5000	421.5803			
142	209.0	353.0000	453.2936			
143	211.0	330.5000	460.5583			
144	215.0	348.0000	475.3308			
145	217.0	414.0000	482.8404			
146	219.0	320.0000	490.4332			
147	220.0	1024.0000	494.2612			
148	222.0	390.0000	501.9806			
149	227.0	542.0000	521.6553			
150	232.0	437.0000	541.8780			

Tabellle A 13: Die Frequenz in Abhängigkeit von der Polytextile Wortform Nomina

i	x_i	y_i	y_i	x_i	y_i	y_i
1	1.0	1.3496	1.3496	51	53.0	94.6667
2	2.0	2.9684	2.9261	52	54.0	91.3333
3	3.0	4.5869	4.5974	53	55.0	144.0000
4	4.0	6.6525	6.3308	54	57.0	83.5000
5	5.0	8.1444	8.1101	55	59.0	132.6667
6	6.0	8.8435	9.9253	56	60.0	114.0000
7	7.0	11.6495	11.7697	57	61.0	86.0000
8	8.0	14.5088	13.6383	58	63.0	195.0000
9	9.0	15.5932	15.5274	59	64.0	151.0000
10	10.0	17.2381	17.4340	60	65.0	113.5000
11	11.0	22.5313	19.3557	61	66.0	126.0000
12	12.0	21.4167	21.2904	62	69.0	184.0000
13	13.0	20.9167	23.2365	63	70.0	113.0000
14	14.0	27.2273	25.1926	64	73.0	125.0000
15	15.0	27.1176	27.1574	65	76.0	250.0000
16	16.0	27.9412	29.1297	66	76.0	153.0000
17	17.0	35.5556	31.1087	67	77.0	160.0000
18	18.0	31.2857	33.0935	68	79.0	157.0000
19	19.0	35.8667	35.0832	69	82.0	152.0000
20	20.0	40.6667	37.0773	70	85.0	169.2000
21	21.0	35.6364	39.0750	71	86.0	126.0000
22	22.0	38.2308	41.0759	72	87.0	144.0000
23	23.0	46.7143	43.0793	73	91.0	137.0000
24	24.0	42.5333	45.0848	74	94.0	163.5000
25	25.0	44.0909	47.0920	75	104.0	204.0000
26	26.0	33.6667	49.1005	76	105.0	216.0000
27	27.0	67.2500	51.1098	77	107.0	313.0000
28	28.0	51.6667	53.1197	78	110.0	243.3333
29	29.0	54.5000	55.1298	79	111.0	174.5000
30	30.0	58.8000	57.1397	80	119.0	216.0000
31	31.0	61.5000	59.1493	81	129.0	179.0000
32	32.0	73.0000	61.1582	82	136.0	317.0000
33	33.0	56.3333	63.1663	83	157.0	243.0000
34	34.0	86.7500	65.1731	84	165.0	304.0000
35	35.0	47.0000	67.1826	85	180.0	278.0000
36	36.0	62.3333	69.1826			
37	37.0	60.6667	71.1848			
38	38.0	67.0000	73.1850			
39	39.0	51.0000	75.1831			
40	40.0	80.0000	77.1789			
41	41.0	97.0000	79.1722			
42	42.0	72.5000	81.1630			
43	44.0	104.5000	85.1360			
44	45.0	97.0000	87.1181			
45	47.0	56.0000	91.0727			
46	48.0	152.0000	93.0450			
47	49.0	107.0000	95.0139			
48	50.0	63.0000	96.9791			
49	51.0	96.5000	98.9407			
50	52.0	90.0000	100.8985			

Tabelle A 14: Die Frequenz in Abhängigkeit von der Polytextie:
Lemmata Nomina

i	x _i	y _i _e	y _i _t	i	x _i	y _i _e	y _i _t
1	1.0	1.4227	1.5313	51	53.0	73.0000	94.8273
2	2.0	2.9231	3.1291	52	54.0	98.0000	96.7388
3	3.0	4.5571	4.7543	53	55.0	79.6667	98.6540
4	4.0	6.4224	6.3985	54	56.0	69.0000	100.5728
5	5.0	8.1412	8.0575	55	57.0	137.0000	162.4953
6	6.0	9.7217	9.7228	56	58.0	85.5000	104.4215
7	7.0	11.8442	11.4109	57	60.0	89.0000	108.2846
8	8.0	12.1957	13.1026	58	61.0	94.0000	110.2216
9	9.0	14.8056	14.8029	59	62.0	107.0000	112.1622
10	10.0	15.8108	16.5114	60	63.0	140.2333	114.1063
11	11.0	17.8333	18.2273	61	65.0	222.0000	118.0053
12	12.0	22.1071	19.9503	62	66.0	93.0000	119.9602
13	13.0	24.7443	21.6801	63	67.0	127.5000	121.9186
14	14.0	22.1538	23.4162	64	68.0	115.0000	123.8805
15	15.0	27.0000	25.1584	65	69.0	163.0000	125.8460
16	16.0	31.7059	26.9056	66	70.0	131.0000	127.8150
17	17.0	37.6364	28.6604	67	72.0	114.0000	131.7635
18	18.0	34.2727	30.4198	68	74.0	159.0000	135.7261
19	19.0	29.5823	32.1846	69	75.0	131.0000	137.7127
20	20.0	35.7143	33.9545	70	77.0	122.0000	141.6962
21	21.0	38.5000	37.5097	71	81.0	150.0000	149.7051
22	22.0	37.5000	39.2947	72	82.0	137.0000	151.7159
23	23.0	56.4286	39.2947	73	85.0	264.0000	157.7693
24	24.0	56.4286	41.0544	74	87.0	158.0000	161.8221
25	25.0	36.5000	42.6785	75	89.0	116.0000	165.8887
26	26.0	47.4000	44.6780	76	90.0	124.0000	167.9272
27	27.0	47.0000	46.4816	77	96.0	154.0000	180.2299
28	28.0	42.2500	48.2898	78	97.0	188.5000	182.2923
29	29.0	47.0000	50.1024	79	99.0	191.0000	186.4274
30	30.0	78.3333	51.9193	80	100.0	198.5000	188.5001
31	31.0	52.6000	53.7406	81	107.0	161.0000	203.1044
32	32.0	62.3333	55.5661	82	109.0	248.0000	207.3077
33	33.0	54.0000	57.3959	83	111.0	161.0000	211.5246
34	34.0	55.6000	59.2298	84	116.0	192.0000	222.1264
35	35.0	53.0000	61.0679	85	118.0	245.0000	226.3910
36	36.0	65.5000	62.9101	86	122.0	202.5000	234.0609
37	37.0	55.0000	64.7564	87	123.0	342.0000	237.1119
38	38.0	76.3333	66.6067	88	129.0	179.0000	250.0891
39	40.0	70.1667	70.3192	89	130.0	201.5000	252.2639
40	41.0	72.0000	72.1814	90	133.0	314.0000	258.8087
41	42.0	63.8000	74.0476	91	134.0	463.0000	263.9971
42	43.0	104.0000	75.9176	92	137.0	300.0000	267.5828
43	44.0	86.7500	77.7915	93	149.0	400.0000	294.2327
44	45.0	75.0000	79.6692	94	162.0	233.0000	323.6597
45	46.0	63.0000	81.5508	95	167.0	411.0000	335.1326
46	47.0	102.5000	83.4362	96	185.0	289.0000	377.1511
47	48.0	60.5000	85.3253	97	194.0	304.0000	398.5834
48	49.0	111.0000	87.2183	98	197.0	437.0000	405.7905
49	50.0	73.0000	89.1149	99	200.0	681.0000	413.0294
50	51.0	82.2500	91.0153	100	278.0	483.0000	612.5805
				101	296.0	762.0000	661.8166

Tabelle A 15: Die Frequenz in Abhängigkeit von der Polytextie:
Wortformen Adjektive

i	x _i	y _i _e	y _i _t
101	210.0	320.000	456.9829
102	225.0	741.000	472.1854
103	232.0	437.000	490.1116
104	237.0	410.000	503.0421
105	271.0	434.000	593.8014
106	284.0	786.000	629.8335
107	295.0	535.000	660.9092
108	323.0	1228.000	742.4873
109	333.0	729.000	772.5001
110	393.0	1183.000	962.6512
111	402.0	1282.000	992.7095
112	416.0	1772.000	1040.2888

Tabelle A 15 (Forts.)

i	x _i	y _i _e	y _i _t	i	x _i	y _i _e	y _i _t
1	1.0	1.2235	1.5219	51	54.0	152.0000	94.7935
2	2.0	2.6479	3.0988	52	57.0	91.0000	100.4456
3	3.0	4.3000	4.6989	53	58.0	261.0000	102.3374
4	4.0	5.2612	6.3151	54	59.0	75.0000	104.2330
5	5.0	7.3614	7.9440	55	60.0	67.0000	106.1326
6	6.0	8.7283	9.5838	56	61.0	164.0000	108.0360
7	7.0	11.1702	11.2331	57	63.0	99.5000	111.8543
8	8.0	10.9412	12.8905	58	64.0	129.6667	113.7692
9	9.0	16.3103	14.5566	59	65.0	97.5000	115.6880
10	10.0	15.1875	16.2297	60	66.0	87.5000	117.6106
11	11.0	18.2778	17.9097	61	67.0	81.0000	119.5370
12	12.0	23.4444	19.5962	62	69.0	113.0000	123.4014
13	13.0	22.1111	21.2890	63	70.0	96.5000	125.3392
14	14.0	17.1765	22.9878	64	71.0	185.5000	127.2809
15	15.0	19.7500	24.6924	65	72.0	103.0000	129.2264
16	16.0	24.3636	26.4026	66	73.0	106.0000	135.0857
17	17.0	27.4615	28.1183	67	75.0	106.0000	135.0857
18	18.0	30.1176	29.8393	68	77.0	155.5000	139.0109
19	19.0	34.4545	31.5655	69	78.0	125.0000	140.9792
20	20.0	39.2500	33.2967	70	82.0	140.0000	148.8903
21	21.0	41.3636	35.0330	71	83.0	146.0000	150.8775
22	22.0	44.3571	36.7742	72	84.0	386.0000	152.8685
23	23.0	52.8333	38.5202	73	94.0	124.0000	173.9864
24	24.0	48.1111	40.2709	74	95.0	128.0000	175.0190
25	25.0	53.7500	42.0263	75	104.0	267.0000	193.4822
26	26.0	55.5000	43.7863	76	108.0	168.5000	201.7565
27	27.0	41.0000	45.5509	77	110.0	162.0000	205.9613
28	28.0	41.0000	47.3200	78	114.0	178.0000	214.3566
29	29.0	62.5714	49.0935	79	115.0	166.0000	218.5770
30	30.0	55.5714	50.8715	80	116.0	367.5000	220.6929
31	31.0	60.6250	52.6539	81	117.0	470.0000	220.6929
32	32.0	78.1667	54.4405	82	118.0	262.0000	222.8126
33	33.0	55.2500	56.2315	83	121.0	237.0000	229.1945
34	34.0	74.0000	58.0268	84	126.0	243.0000	239.9074
35	35.0	47.2333	59.8263	85	135.0	203.0000	259.4315
36	36.0	59.5000	61.6300	86	140.0	255.0000	263.8124
37	37.0	64.5000	63.4379	87	140.0	200.0000	270.4126
38	38.0	79.0000	65.2500	88	144.0	228.0000	279.2669
39	39.0	77.5000	67.0662	89	148.0	255.0000	288.1632
40	40.0	113.0000	68.8865	90	154.0	253.0000	301.6740
41	41.0	79.6667	70.7109	91	161.0	412.0000	317.5908
42	42.0	56.2500	72.5394	92	162.0	218.0000	319.8803
43	43.0	74.3720	74.3720	93	174.0	273.0000	347.6613
44	44.0	56.0000	76.2086	94	182.0	252.0000	366.4990
45	45.0	70.6667	78.0492	95	190.0	386.0000	385.5926
46	46.0	53.0000	79.8938	96	192.0	302.0000	390.4062
47	47.0	117.0000	81.7424	97	193.0	403.0000	392.8190
48	48.0	76.0000	83.5950	98	194.0	305.0000	395.2359
49	49.0	153.0000	87.3121	99	211.0	311.0000	436.9450
50	50.0	111.5000	91.0449	100	215.0	348.0000	446.9309

Tabelle A 16: Die Frequenz in Abhängigkeit von der Polytextite:
Lemmata Adjektive

i	x _i	y _i _e	y _i _t	i	x _i	y _i _e	y _i _t
1	1.0	1.2856	1.4763	51	52.0	75.0000	73.7335
2	2.0	2.5472	2.8776	52	53.0	83.3333	75.2698
3	3.0	3.7739	4.2562	53	55.0	76.0000	78.3597
4	4.0	5.1625	5.6224	54	56.0	87.0000	79.9133
5	5.0	6.1695	6.9812	55	58.0	109.0000	83.0383
6	6.0	7.8421	8.3354	56	60.0	69.2500	86.1870
7	7.0	9.6765	9.6869	57	63.0	92.0000	90.9552
8	8.0	11.3333	11.0369	58	66.0	87.5000	95.7784
9	9.0	18.6316	12.3865	59	67.0	92.5000	97.3984
10	10.0	15.1429	13.7364	60	70.0	91.0000	102.2962
11	11.0	14.3820	15.0873	61	72.0	99.0000	103.5929
12	12.0	19.2500	16.4395	62	75.0	94.0000	110.5859
13	13.0	17.3333	17.7935	63	76.0	141.5000	112.2632
14	14.0	16.9286	19.1497	64	77.0	94.0000	113.9470
15	15.0	21.2727	20.5084	65	79.0	110.0000	117.3341
16	16.0	18.3333	21.8698	66	82.0	248.0000	122.4644
17	17.0	22.2000	23.2341	67	85.0	129.5000	127.6549
18	18.0	19.7778	24.6016	68	87.0	138.0000	131.1490
19	19.0	27.1429	25.9724	69	89.0	155.0000	134.6704
20	20.0	22.0000	27.3468	70	90.0	123.0000	136.4414
21	21.0	45.0000	28.7247	71	92.0	133.0000	140.0043
22	22.0	30.7778	30.1065	72	94.0	176.3333	143.5950
23	23.0	30.7500	31.4921	73	95.0	128.0000	145.4008
24	24.0	27.5000	32.8818	74	101.0	167.0000	156.3853
25	25.0	32.0000	34.2756	75	102.0	137.0000	158.7242
26	26.0	43.0000	35.6736	76	106.0	173.0000	165.7376
27	27.0	33.0000	37.0760	77	107.0	173.0000	167.6300
28	28.0	45.2500	38.4827	78	108.0	151.0000	169.5299
29	29.0	39.5000	39.8940	79	110.0	162.0000	173.3519
30	30.0	42.3333	41.3098	80	111.0	190.0000	175.2741
31	31.0	38.0000	42.7302	81	115.0	173.0000	183.0384
32	32.0	35.6667	44.1554	82	116.0	139.0000	184.9985
33	33.0	57.7500	45.5853	83	118.0	188.0000	188.9416
34	34.0	42.0000	47.0201	84	121.0	164.0000	194.9142
35	35.0	45.0000	48.4597	85	124.0	234.0000	200.9570
36	36.0	74.1250	49.9043	86	137.0	221.0000	227.9727
37	37.0	50.5000	51.3540	87	140.0	200.0000	234.4032
38	38.0	46.0000	52.8087	88	141.0	257.0000	236.5633
39	39.0	63.0000	54.2685	89	144.0	228.0000	243.0042
40	40.0	85.0000	55.7335	90	150.0	264.0000	256.3855
41	41.0	68.6667	57.2037	91	152.0	234.0000	260.8848
42	42.0	52.3333	58.6792	92	161.0	245.0000	281.5668
43	43.0	83.0000	60.1600	93	162.0	218.0000	283.9093
44	44.0	83.0000	61.6462	94	167.0	290.0000	295.7579
45	45.0	59.0000	63.1377	95	169.0	355.0000	300.5614
46	46.0	55.0000	64.6348	96	174.0	273.0000	312.7325
47	47.0	70.0000	66.1373	97	182.0	252.0000	332.6966
48	48.0	71.0000	67.6453	98	192.0	302.5000	358.5232
49	49.0	65.5000	69.1589	99	193.0	403.0000	361.1603
50	50.0	122.0000	70.6781	100	194.0	305.0000	363.8074

Tabelle A 16 (Forts.)

i	x_i	y_{i_e}	y_{i_t}
101	200.0	368.000	379.9023
102	209.0	353.000	404.7397
103	211.0	330.500	410.3745
104	215.0	348.000	421.7722
105	219.0	320.000	433.3425
106	220.0	1024.000	436.2623
107	222.0	390.000	442.1348
108	232.0	437.000	472.1657
109	243.0	422.000	506.5188
110	272.0	435.000	604.9820
111	287.0	792.000	658.7702
112	295.0	535.000	689.1823
113	300.0	611.000	708.6429
114	304.0	778.000	724.4670
115	319.0	868.000	785.8826
116	323.0	1028.000	802.8272
117	333.0	729.000	846.2660
118	334.0	918.000	850.6959
119	365.0	1014.000	996.1290
120	393.0	1235.000	1141.8716
121	402.0	1282.000	1191.8287
122	416.0	1772.000	1272.7115
123	447.0	1912.000	1466.3431

Tabelle A 17: Die Frequenz in Abhängigkeit von der Polytextie:
Wortformen Verben

i	x_i	y_{i_e}	y_{i_t}	i	x_i	y_{i_e}	y_{i_t}
1	1.0	1.1604	1.4657	51	52.0	35.5000	86.7983
2	2.0	2.4731	2.9067	52	54.0	66.5000	90.7878
3	3.0	3.7283	4.3456	53	55.0	90.0000	92.8033
4	4.0	5.6890	5.7870	54	57.0	94.3333	96.8763
5	5.0	8.3419	7.2333	55	58.0	103.0000	98.9339
6	6.0	8.3871	8.6858	56	59.0	131.5000	101.0057
7	7.0	9.1667	10.1454	57	61.0	93.0000	105.1022
8	8.0	12.2083	11.6127	58	62.0	76.5000	107.3072
9	9.0	12.0465	13.0881	59	63.0	105.5000	109.4367
10	10.0	14.1818	14.5722	60	64.0	106.0000	111.5808
11	11.0	15.3939	16.0652	61	66.0	147.0000	115.9134
12	12.0	17.9000	17.5674	62	67.0	81.0000	118.1020
13	13.0	20.0690	19.0790	63	68.0	692.5000	120.3056
14	14.0	17.2000	20.6003	64	69.0	125.2500	122.5243
15	15.0	22.2000	22.1315	65	70.0	105.0000	124.7582
16	16.0	24.1250	23.6727	66	77.0	214.0000	140.8278
17	17.0	42.4000	25.2240	67	80.0	114.0000	147.9516
18	18.0	28.6420	26.7857	68	82.0	107.3000	152.7815
19	19.0	29.0625	28.3578	69	84.0	307.0000	157.6770
20	20.0	68.5000	29.9406	70	86.0	116.0000	162.0387
21	21.0	29.3846	31.5340	71	89.0	130.0000	170.2372
22	22.0	35.5000	33.1383	72	94.0	110.0000	183.1636
23	23.0	32.1429	34.7536	73	95.0	134.0000	185.8071
24	24.0	34.8000	36.3799	74	96.0	182.0000	188.4683
25	25.0	129.8333	38.0174	75	100.0	137.0000	199.2914
26	26.0	36.2500	39.6662	76	101.0	127.0000	202.0422
27	27.0	45.6000	41.3264	77	104.0	230.0000	210.4048
28	28.0	35.0000	42.9980	78	106.0	171.0000	216.0725
29	29.0	63.2000	44.6812	79	109.0	137.0000	224.7151
30	30.0	38.7500	46.3761	80	110.0	265.0000	227.6340
31	31.0	48.0000	48.0827	81	113.0	211.0000	236.5064
32	32.0	48.6667	49.8012	82	118.0	183.0000	251.6856
33	33.0	58.6667	51.5317	83	128.0	282.0000	283.5639
34	34.0	48.3333	53.2742	84	147.0	352.0000	350.0796
35	35.0	48.3333	55.0288	85	156.0	525.0000	384.5028
36	36.0	48.4000	56.7957	86	158.0	303.0000	392.4211
37	37.0	52.5000	58.5749	87	159.0	229.0000	396.4177
38	38.0	53.5000	60.3664	88	160.0	276.0000	400.4395
39	39.0	69.0000	62.1705	89	162.0	1928.0000	408.5590
40	40.0	61.2500	63.9871	90	166.0	428.0000	425.1058
41	41.0	48.2500	65.8163	91	176.0	262.0000	468.3215
42	42.0	63.5000	67.6584	92	186.0	645.0000	514.2945
43	43.0	65.2500	69.5132	93	190.0	286.0000	533.4897
44	44.0	71.0000	71.3810	94	193.0	455.0000	548.1976
45	45.0	83.3333	73.2617	95	194.0	755.0000	553.1604
46	46.0	66.0000	75.1556	96	199.0	2817.0000	578.4332
47	47.0	114.0000	77.0626	97	202.0	431.0000	593.9697
48	48.0	60.0000	78.9829				
49	50.0	95.0000	82.8636				
50	51.0	65.0000	84.8241				

Tabelle A 18: Die Frequenz in Abhängigkeit von der Polytexhtie:
Lenmata Verben

i	x_i	y_i	y_i	x_i	y_i	y_i
1	1.0	1.1186	1.2572	56.0	81.000	73.6006
2	2.0	2.2067	2.4683	58.0	72.000	76.6180
3	3.0	3.2527	3.6672	59.0	99.000	78.1386
4	4.0	4.3627	4.8609	60.0	71.000	79.6672
5	5.0	5.6364	6.0526	61.0	79.000	81.2038
6	6.0	6.8077	7.2443	62.0	74.000	82.7485
7	7.0	7.7692	8.4371	63.0	79.000	84.3014
8	8.0	9.2353	9.6321	64.0	133.000	85.8624
9	9.0	10.9091	10.8297	65.0	115.000	90.5950
10	10.0	11.9444	12.0307	66.0	88.000	100.2870
11	11.0	12.8000	13.2353	67.0	91.500	105.2487
12	12.0	14.7500	14.4439	68.0	99.000	106.9201
13	13.0	14.4000	15.6569	69.0	120.000	108.6002
14	14.0	16.8000	16.8744	70.0	113.000	115.4096
15	15.0	19.8750	18.0966	71.0	112.500	120.6111
16	16.0	18.7500	19.3239	72.0	144.000	127.6746
17	17.0	19.6667	20.5563	73.0	143.000	133.0701
18	18.0	20.5667	21.7939	74.0	123.000	138.5507
19	19.0	29.7500	23.0370	75.0	141.500	140.3966
20	20.0	23.5000	24.2857	76.0	134.000	147.8774
21	21.0	29.1667	25.5400	77.0	167.000	157.4498
22	22.0	27.6667	26.8001	78.0	152.000	167.2735
23	23.0	30.0000	28.0660	79.0	187.000	181.4601
24	24.0	27.8000	29.3380	80.0	216.000	191.9111
25	25.0	31.7500	30.6160	81.0	231.000	196.1672
26	26.0	34.5000	33.1905	82.0	236.000	203.4672
27	27.0	38.0000	34.4872	83.0	175.000	209.2002
28	28.0	36.0000	35.7902	84.0	201.000	213.6341
29	29.0	35.5000	38.4157	85.0	205.000	241.2065
30	30.0	38.0000	39.7383	86.0	204.000	248.3657
31	31.0	36.0000	41.0676	87.0	237.000	253.1992
32	32.0	59.0000	42.4035	88.0	203.000	255.2342
33	33.0	63.5000	43.7462	89.0	210.000	265.4982
34	34.0	47.0000	45.0957	90.0	287.000	315.1491
35	35.0	41.5667	46.4521	91.0	367.000	320.6544
36	36.0	56.5000	49.1858	92.0	277.000	334.6585
37	37.0	64.6667	50.5632	93.0	331.000	372.7289
38	38.0	61.0000	51.9477	94.0	369.000	391.1366
39	39.0	58.0000	54.7382	95.0	416.000	449.7022
40	40.0	57.3333	56.1443	96.0	542.000	484.5123
41	41.0	58.0000	57.5577	97.0	657.000	599.1181
42	42.0	57.0000	58.9784	98.0	611.000	685.7750
43	43.0	80.5000	60.4065	99.0	548.000	722.6038
44	44.0	55.5000	61.8358	100.0	667.000	790.4429
45	45.0	60.6667	63.2852		685.000	800.4980
46	46.0	58.0000	64.7358		800.000	1010.1690
47	47.0	75.0000	67.6599		719.000	1230.9772
48	48.0	83.0000	70.6147		1090.000	1863.3650
49	49.0	67.0000	72.1038		1513.000	1953.1241
50	50.0	67.0000			1781.000	2421.1783

Tabelle A 19: Die Länge in Abhängigkeit von der Frequenz
(gleitende Mittel über 20)

i	x_i	y_i	y_i	x_i	y_i	y_i
1	11.5000	8.6601	7.9756	65.3000	5.0683	7.1204
2	12.5000	8.5695	7.9323	66.5500	6.8717	7.1116
3	13.5000	8.3510	7.8925	67.8000	7.1467	7.1029
4	14.5000	8.2340	7.8558	69.1000	7.1567	7.0941
5	15.5000	8.0433	7.8216	70.3500	7.4967	7.0845
6	16.5000	7.8045	7.7898	72.0000	7.6717	7.0751
7	17.5000	7.6808	7.7599	73.5000	7.7967	7.0656
8	18.5000	7.7752	7.7318	75.0000	7.9967	7.0562
9	19.5000	7.6098	7.7052	76.5500	8.0167	7.0468
10	20.5000	7.7098	7.6801	78.0500	7.9833	7.0379
11	21.5000	7.7817	7.6562	79.7000	7.9083	7.0283
12	22.5000	7.7484	7.6335	81.3500	7.7583	7.0189
13	23.5000	7.7206	7.6119	83.0000	7.7023	7.0097
14	24.5000	7.7656	7.5912	84.7500	7.8083	7.0001
15	25.5000	7.6556	7.5714	86.5500	7.6583	6.9905
16	26.5000	7.5763	7.5524	88.4500	7.6083	6.9806
17	27.5000	7.4413	7.5341	90.3000	7.5083	6.9702
18	28.5000	7.4038	7.5166	92.6000	7.5333	6.9598
19	29.5000	7.2217	7.4997	94.8000	7.5333	6.9491
20	30.5000	7.1467	7.4834	97.0500	7.3333	6.9385
21	31.5000	7.1217	7.4676	99.3500	7.2833	6.9278
22	32.5000	7.2217	7.4524	101.5999	7.1000	6.9177
23	33.5000	7.2633	7.4376	103.8499	6.8750	6.9078
24	34.5000	7.4508	7.4234	106.2500	6.9250	6.8975
25	35.5000	7.5758	7.4089	108.5000	6.8750	6.8881
26	36.5000	7.4456	7.3948	110.7500	6.5500	6.8789
27	37.5000	7.4208	7.3811	112.9499	6.4500	6.8700
28	38.7000	7.2542	7.3679	115.1000	6.3000	6.8616
29	39.8000	7.2042	7.3544	117.2999	6.1000	6.8531
30	40.9000	7.0208	7.3413	119.6499	6.1500	6.8442
31	42.0000	6.9208	7.3286	121.8499	6.3000	6.8361
32	43.1000	7.0542	7.3162	124.0999	6.3000	6.8279
33	44.2000	6.8958	7.3042	126.6499	6.3500	6.8189
34	45.3000	6.7458	7.2925	129.1999	6.4000	6.8100
35	46.4000	6.5792	7.2811	131.6499	6.5250	6.8010
36	47.5000	6.6292	7.2699	134.5499	6.5250	6.7920
37	48.6000	6.6292	7.2591	137.2499	6.4750	6.7832
38	49.7500	6.5667	7.2480	139.8999	6.6250	6.7747
39	50.9000	6.7717	7.2372	142.5499	6.5750	6.7664
40	52.1000	6.7550	7.2262	145.1999	6.5750	6.7583
41	53.3000	6.8050	7.2154	147.7999	6.7250	6.7504
42	54.5000	6.7550	7.2050	150.5499	6.7900	6.7423
43	55.7500	6.7050	7.1943	153.3999	6.7250	6.7340
44	57.0000	6.3300	7.1839	156.0999	6.8750	6.7264
45	58.2000	6.4550	7.1741	158.8999	6.7250	6.7186
46	59.4000	6.5050	7.1645	161.7999	6.7000	6.7106
47	60.6000	6.5133	7.1552	164.7500	6.9500	6.7027
48	61.8000	6.6383	7.1460	166.0499	5.9500	6.6940
49	62.9500	6.7383	7.1374	171.2500	5.8500	6.6858
50	64.1000	6.9883	7.1290	174.3499	6.9000	6.6780

Tabelle A 19 (Forts.)

i	x _i	y _i _e	y _i _t
101	177.5000	6.7500	6.6702
102	180.7999	6.9500	6.6621
103	184.1500	7.0000	6.6542
104	187.3999	6.9500	6.6466
105	190.4999	6.6250	6.6394
106	193.5999	6.7750	6.6324
107	196.9499	6.8250	6.6250
108	200.9500	6.4750	6.6163
109	205.6499	6.4750	6.6063
110	210.5000	6.6750	6.5963
111	215.5999	6.5750	6.5860
112	221.0000	6.7500	6.5754
113	226.3999	6.7500	6.5650
114	232.6999	6.9000	6.5529
115	240.0999	6.9000	6.5398
116	247.6999	7.2000	6.5265
117	255.4999	6.9000	6.5133
118	263.1499	7.0500	6.5008
119	271.4499	7.1500	6.4876
120	279.6999	7.0000	6.4750
121	287.8999	7.0000	6.4628
122	296.8999	6.8500	6.4498
123	306.7500	6.7500	6.4360
124	317.0999	6.7500	6.4221
125	326.1999	6.6000	6.4051
126	336.3499	6.7000	6.3852
127	345.1999	6.6000	6.3631
128	355.6999	6.8000	6.3405
129	405.7499	6.8000	6.3195
130	426.5000	6.5500	6.2990
131	446.5000	6.6000	6.2782
132	472.3499	6.4500	6.2571
133	496.7499	6.3000	6.2365
134	520.5999	5.9000	6.2175
135	547.4499	5.7500	6.1971
136	581.0499	5.4500	6.1730
137	618.1999	5.4500	6.1481
138	676.1499	5.3500	6.1122
139	746.8500	5.2500	6.0726
140	824.0499	5.2000	6.0337

Tabelle A 20: Die Länge in Abhängigkeit von der Frequenz (gleitende Mittel über 50)

i	x _i	y _i _e	y _i _t	f	x _i	y _i _e	y _i _t
1	26.70	7.8094	7.4280	51	92.4400	6.9887	6.5784
2	27.74	7.7411	7.4138	52	94.2400	6.9087	6.5717
3	28.78	7.6124	7.4001	53	96.1800	6.9387	6.5645
4	29.82	7.4906	7.3865	54	98.1600	6.9987	6.5574
5	30.86	7.3743	7.3742	55	100.1999	7.1287	6.5502
6	31.90	7.3188	7.3619	56	102.2999	7.0767	6.5430
7	32.94	7.2506	7.3500	57	104.4600	7.0687	6.5357
8	34.00	7.1817	7.3363	58	106.5999	7.1487	6.5286
9	35.06	7.1676	7.3270	59	108.7800	7.0967	6.5216
10	36.14	7.1676	7.3158	60	110.9599	7.0833	6.5147
11	37.22	7.1863	7.3050	61	113.1599	7.0533	6.5078
12	38.30	7.2063	7.2945	62	115.4200	6.9833	6.5010
13	39.40	7.1619	7.2841	63	117.7000	6.9833	6.4942
14	40.50	7.1195	7.2741	64	120.0000	7.0633	6.4875
15	41.60	7.1532	7.2643	65	122.3399	6.9833	6.4806
16	42.70	7.1075	7.2547	66	124.7199	6.9433	6.4741
17	43.80	7.1035	7.2455	67	127.1199	7.0633	6.4675
18	44.90	7.0735	7.2364	68	129.6599	6.9733	6.4607
19	46.00	7.0507	7.2276	69	132.1999	6.9333	6.4540
20	47.10	7.0607	7.2191	70	134.7599	6.8933	6.4474
21	48.22	7.0207	7.2105	71	137.3199	6.9333	6.4409
22	49.36	7.0540	7.2021	72	139.9399	6.9400	6.4344
23	50.50	7.1307	7.1938	73	142.7199	6.8600	6.4277
24	51.66	7.0757	7.1856	74	145.4799	6.9000	6.4211
25	52.88	7.1557	7.1772	75	148.1799	6.7500	6.4148
26	54.10	7.2237	7.1689	76	150.9399	6.7200	6.4085
27	55.34	7.2170	7.1608	77	153.6399	6.8000	6.4020
28	56.60	7.2703	7.1527	78	156.9799	6.5400	6.3951
29	57.88	7.3303	7.1446	79	160.4199	6.4800	6.3877
30	59.16	7.2770	7.1368	80	163.9399	6.5600	6.3803
31	60.50	7.2370	7.1288	81	167.5199	6.5200	6.3729
32	61.84	7.2103	7.1209	82	171.2800	6.5800	6.3654
33	63.20	7.1570	7.1131	83	175.0599	6.6000	6.3579
34	64.60	7.1370	7.1053	84	179.2599	6.7200	6.3499
35	66.02	7.1503	7.0975	85	183.7599	6.6800	6.3415
36	67.48	7.1203	7.0897	86	188.4199	6.7800	6.3330
37	69.00	7.1303	7.0818	87	193.1200	6.7600	6.3247
38	70.54	7.1253	7.0739	88	197.8799	6.7600	6.3164
39	72.12	7.1753	7.0661	89	202.8599	6.7800	6.3080
40	73.72	7.1553	7.0583	90	207.8199	6.7600	6.3000
41	75.36	7.0853	7.0505	91	212.7399	6.8200	6.2920
42	77.00	7.0053	7.0428	92	218.0600	6.8400	6.2837
43	78.64	6.9753	7.0354	93	223.8799	6.8300	6.2749
44	80.36	6.8653	7.0277	94	229.8199	6.8300	6.2661
45	82.06	6.8753	7.0203	95	236.8599	6.7500	6.2560
46	83.76	6.8653	7.0131	96	245.1799	6.7800	6.2444
47	85.46	6.8287	7.0060	97	254.7399	6.7800	6.2317
48	87.16	6.8887	6.9991	98	265.1999	6.7400	6.2183
49	88.88	6.8887	6.9922	99	275.7999	6.7600	6.2053
50	90.66	6.9287	6.9853	100	286.6799	6.7200	6.1924

Tabelle A 20 (Forts.)

i	x _i	y _i _e	y _i _t
101	296.2199	6.6400	6.5793
102	310.5799	6.6400	6.5659
103	323.0999	6.5800	6.5529
104	335.8199	6.5200	6.5402
105	350.0000	6.3700	6.5266
106	367.0199	6.3500	6.5110
107	385.4799	6.3500	6.4930
108	412.3799	6.2700	6.4730
109	444.5799	6.2500	6.4486
110	479.3600	6.2100	6.4242

Tabelle A 21: Die Länge in Abhängigkeit von der Frequenz (gleitende Mittel über 100)

i	x _i	y _i _e	y _i _t	f	y _i _e	y _i _t
1	59.5700	7.3990	7.2061	51	195.3299	6.8143
2	60.9900	7.3249	7.1992	52	202.4099	6.7743
3	62.4800	7.2755	7.1902	53	209.6299	6.7593
4	63.9900	7.2446	7.1813	54	216.9899	6.7593
5	65.5300	7.2515	7.1724	55	225.0999	6.7493
6	67.1000	7.1977	7.1636	56	234.6599	6.7143
7	68.7000	7.1596	7.1545	57	244.9699	6.7093
8	70.3000	7.1652	7.1463	58	259.4900	6.7093
9	71.9200	7.1321	7.1378	59	276.6799	6.6733
10	73.5500	7.1255	7.1295	60	295.1599	6.6467
11	75.1900	7.1198	7.1214			
12	76.8600	7.0948	7.1132			
13	78.5500	7.0726	7.1052			
14	80.2500	7.1016	7.0973			
15	81.9700	7.0683	7.0895			
16	83.7100	7.0254	7.0818			
17	85.4600	7.0534	7.0742			
18	87.2800	7.0234	7.0664			
19	89.1000	6.9920	7.0588			
20	90.9300	6.9870	7.0514			
21	92.7700	6.9770	7.0441			
22	94.6500	6.9570	7.0367			
23	96.6100	7.0053	7.0292			
24	98.5700	6.9878	7.0219			
25	100.5300	6.9678	7.0147			
26	102.5199	6.9716	7.0076			
27	104.5899	6.9485	7.0003			
28	106.7899	6.9052	6.9927			
29	109.1500	6.9052	6.9848			
30	111.5499	6.9185	6.9769			
31	114.0100	6.8785	6.9690			
32	116.5600	6.8952	6.9610			
33	119.1299	6.8785	6.9531			
34	121.9300	6.9285	6.9447			
35	124.8899	6.9152	6.9361			
36	127.9500	6.9502	6.9274			
37	131.0600	6.9452	6.9167			
38	134.2099	6.9527	6.9102			
39	137.4899	6.9777	6.9015			
40	140.7699	6.9577	6.8931			
41	144.0499	6.9527	6.8848			
42	147.5300	6.9227	6.8763			
43	151.2599	6.9027	6.8674			
44	155.0899	6.8477	6.8585			
45	159.4599	6.8127	6.8486			
46	164.4699	6.8227	6.8376			
47	170.0999	6.8043	6.8256			
48	176.1799	6.8143	6.8132			
49	182.3400	6.8243	6.8010			
50	188.6659	6.8243	6.7890			

TOTAL INFORMATION from Language and Language Behavior Abstracts

Lengthy, informative English abstracts—regardless of source language—which include authors' mailing addresses.

Complete indices—author name, book review, subject, and periodical sources at your fingertips.

Numerous advertisements for books and journals of interest to language practitioners.

NOW over 1200 periodicals searched from 40 countries—in 32 languages—from 25 disciplines.

Complete copy service for most articles.

ACCESS TO THE WORLD'S STUDIES ON LANGUAGE—IN ONE CONVENIENT PLACE!

What's the alternative?

Time consuming manual search through dusty, incomplete archives.

Limited access to foreign and specialized sources.

Need for professional translations to remain informed.

Make sure YOU have access to

LANGUAGE AND LANGUAGE BEHAVIOR ABSTRACTS when you need it . . .

For complete information about current and back volumes, write to: P.O. Box 22206, San Diego, CA. 92122, USA.