

# Das synergetische Basismodell der Lexik und die chinesische Schrift

*Cornelia Menzel*

## 1. Einleitung

Das chinesische Schriftzeichensystem scheint einige Zusammenhänge aufzuweisen, die denen entsprechen, die bereits am Wortschatz vor allem westlicher Sprachen überprüft wurden. So gibt es beispielsweise Häufigkeitsunterschiede unter den Schriftzeichen, genauso wie es mehr oder weniger häufige Wörter gibt. Und wie häufigere Wörter im Durchschnitt kürzer sind als seltenere, scheinen häufigere Schriftzeichen im Durchschnitt auch graphisch einfacher zu sein als seltenere. Diese Beobachtungen und die Beschäftigung mit dem Basismodell der Lexik, das R. Köhler entwickelt und 1986 veröffentlicht hat<sup>1</sup>, regte mich dazu an, die darin erfaßten Zusammenhänge auch für die chinesische Schrift zu untersuchen. Ich gelangte zu der Annahme, daß die Zusammenhänge zwischen Häufigkeit, Länge und Polylexie sowie Polytextie in ähnlicher Weise auch hier gelten müßten, obgleich es sich bei den Schriftzeichen nicht um dieselbe Analyseebene handelt wie beim Wortschatz. Das führte zu einer entsprechenden Übertragung des Modells und den Untersuchungen, deren Ergebnisse ich hier vorlege.

In meinen Ausführungen werde ich übrigens gelegentlich einfach „Zeichen“ schreiben, wenn ich „chinesisches Schriftzeichen“ meine. Ist doch das *principle of least effort* vermutlich ein universelles...

## 2. Die chinesische Schrift

Die moderne chinesische Schrift, d.h. die Schriftzeichen, wie sie zur Aufzeichnung moderner chinesischer Sprache verwendet werden, ist als morpho-sylla-

---

<sup>1</sup> Köhler (1986).

bisches Schriftzeichensystem zu betrachten.<sup>2</sup> Knapp 90% der Schriftzeichen im modernen Schriftzeicheninventar repräsentieren Morpheme, und zwar jeweils ca. zur Hälfte freie und gebundene Morpheme. Bei den übrigen gut 11% handelt es sich um unikale Morpheme oder um Bestandteile von Morphemen, die mit zwei oder mehr Schriftzeichen dargestellt werden. Ein Schriftzeichen (als Type) kann zur Darstellung mehrerer verschiedener Morpheme dienen und mehrere Aussprachen haben, was jedoch nicht bedeutet, daß es für jedes Morphem eine eigene Aussprache aufweist. Ein Schriftzeichen in einem Text wird als eine Silbe gelesen, mit vernachlässigbar wenigen Ausnahmen. Die überwältigende Mehrheit der Wörter im modernen Wortschatz ist zweisilbig und wird entsprechend mit zwei Schriftzeichen niedergeschrieben. Chinesische Wörter, das sei hier angemerkt, kennen keine Flexion, da die chinesische Sprache grammatikalische Beziehungen hauptsächlich durch lexikalische Mittel und die Satzstellung ausdrückt.

Es ist natürlich problematisch, von *dem* modernen Schriftzeicheninventar zu reden. Verschiedene Frequenzuntersuchungen kamen betreffs der Anzahl der heute gebräuchlichen Zeichen zu unterschiedlichen Ergebnissen, die von etwa 4.500 bis etwas über 7.500 Schriftzeichen-Types reichten. Auf letztere Zahl kam jedoch nur eine Untersuchung, die ein Korpus von knapp 12 Mio. Schriftzeichen zum Gegenstand hatte. Zwei weitere Untersuchungen, deren Korpusgrößen gut 21 Mio. bzw. gut 40 Mio. Schriftzeichen betragen, kamen auf 5.991 bzw. 6.001 Schriftzeichen-Types. Das *Xinhua zidian*, das verbreitetste Schriftzeichenbüchlein in der VR China, enthält rund 11.100 Schriftzeicheneinträge und ohne Berücksichtigung von Mehrfacheinträgen, z.B. wegen verschiedener Aussprachen, angeblich immer noch gut 8.000.<sup>3</sup> Wenn man 1.500 Schriftzeichen beherrscht, gilt man in der VR China als Halbanalphabet, mit 3.000 Schriftzeichen als alphabetisiert. Es gibt eine offizielle Liste von 2.500 häufigen und 1.000 „nächsthäufigen“ Zeichen. Wer diese insgesamt 3.500 Zeichen beherrscht, soll 99,48% aller Schriftzeichen in allgemeinen Texten erkennen können, wozu die „nächsthäufigen“ Zeichen nur 1,51% beitragen. Nähme man weitere 1.000 Schriftzeichen hinzu, so würde sich dieser „Abdeckungsgrad“ nur kaum spürbar erhöhen.

In der VR China wurde in den fünfziger Jahren eine Schriftreform durchgeführt, die das Ziel hatte, das Lesen und Schreiben für die breite Masse des Volkes zu erleichtern. Im Zuge dieser Reform wurden 2.264 „Langzeichen“ durch 2.236 „Kurzzeichen“ ersetzt. Dies geschah zum großen Teil dadurch, daß Schriftzeichenelemente, die in vielen Schriftzeichen vorkamen, vereinfacht wurden, so daß eine Vereinfachung viele Schriftzeichen betraf. In den meisten Fällen war die gewählte neue Gestalt des Elements eine in der Handschrift bereits weit

---

<sup>2</sup> DeFrancis (1984). Dies ist vermutlich die beste linguistische Einführung in die chinesische Sprache und Schrift in einer westlichen Sprache.

<sup>3</sup> Die Ausgabe von 1971 enthält 8.075 Einträge ohne Berücksichtigung von Mehrfacheinträgen, vgl. Zhou (1978:142).

verbreitete Vulgärform, die nunmehr zum Standard erhoben wurde. Weitere Methoden der Kürzung waren u.a. die Rückführung auf archaische Formen, die Ersetzung komplexer Komponenten durch einfache Symbole und die Schaffung völlig neuer einfacher Zeichen. Die Differenz von 28 Zeichen kommt dadurch zustande, daß in einigen Fällen Morpheme, die früher durch „eigene“ Zeichen geschrieben wurden, nunmehr mit demselben Zeichen geschrieben werden wie ein anderes Morphem.

In China beruht bis heute fast jeder gedruckte Text auf einer handgeschriebenen Vorlage, so daß das Bedürfnis nach Aufwandsreduzierung beim Schreiben durchaus zu Unterschieden in der Strichzahl von handgeschriebenen und gedruckten, dem Orthographiestandard entsprechenden Schriftzeichen führen könnte. In einer daraufhin untersuchten Stichprobe von 152 verschiedenen Kurzzeichen, die in handgeschriebener und gedruckter Form vorlagen, hatten knapp ein Drittel in handgeschriebener Form genauso viele Striche wie in der gedruckten Form. 43% dieser Schriftzeichen waren in der handschriftlichen Form nur einen Strich „kürzer“ als in Druckschrift. Insgesamt waren also knapp drei Viertel der Schriftzeichen in handschriftlicher Form gar nicht oder nur geringfügig „kürzer“ als in der gedruckten Form. Die Strichzahl der gedruckten Form der Kurzzeichen kann also als Näherung an den tatsächlichen Schreibaufwand angenommen werden.

Untersuchungen zur Häufigkeit von Schriftzeichen in Texten werden in China seit 1927 durchgeführt. Sie sind meist durch schriftreformerische oder pädagogische Ziele motiviert. Die Größe der analysierten Korpora hat mit der Entwicklung moderner Möglichkeiten der Datenverarbeitung stark zugenommen. Dennoch ist die quantitative Linguistik in China noch sehr entwicklungsbedürftig, eine Situation, die sich mit der stärkeren Beteiligung Chinas auch am internationalen wissenschaftlichen Austausch sicherlich bald ändern wird.

### **3. Das Korpus**

Das untersuchte Korpus besteht aus Texten, die in Kurzzeichen gedruckt waren. Es umfaßt insgesamt 1.808.114 Schriftzeichen Text und weist ein Inventar von 4.574 Schriftzeichen auf. Es war unter didaktischen Gesichtspunkten zusammengestellt worden und sollte an Inhalt und Textsorten widerspiegeln, was ein durchschnittlich gebildeter Chinese in seinem Leben liest. Es umfaßt daher Sach- und Fachtexte (zus. knapp 40%), Dramen, fiktionale Prosa sowie Essays und Märchen. Bei der Auswertung wurden nur chinesische Schriftzeichen berücksichtigt, während Satzzeichen, arabische Zahlzeichen, lateinische Buchstaben u.ä. übergangen wurden. Die ermittelten Daten wurden in einem Häufigkeitswörterbuch<sup>4</sup> veröffentlicht. Dieses enthält neben Wortlisten auch eine Liste, die für jedes Zei-

---

<sup>4</sup> *Xiandai Hanyu pinlü cidian* (1986).

chen die absolute und relative Häufigkeit sowie den Rang (genauer: Ordnungsnummer) verzeichnet und darüber hinaus Angaben dazu enthält, an der Darstellung wie vieler verschiedener Wörter ein Schriftzeichen beteiligt ist, ob es bei mehrsilbigen Wörtern am Anfang, Ende oder in der „Mitte“ steht bzw. ob es nur zur Darstellung einsilbiger Wörter dient. Zusammengefaßt stellt sich diese Verteilung wie folgt dar:

217 (= 4,7%) Schriftzeichen stellen ausschließlich einsilbige Wörter dar.

1620 (= 35,5%) Schriftzeichen sind ausschließlich an der Darstellung mehrsilbiger Wörter beteiligt, davon kommen 519 ausschließlich am Wortanfang, 39 ausschließlich in der (nicht näher spezifizierten) Wortmitte, 433 ausschließlich am Wortende und 168 in allen drei möglichen Positionen vor.

2737 (= 59,8%) Schriftzeichen können sowohl einsilbige Wörter darstellen als auch an der Darstellung mehrsilbiger Wörter beteiligt sein.

Die Daten aus diesem Korpus wurden als Untersuchungsgrundlage gewählt, weil das Korpus ausreichend groß erscheint und das Frequenzwörterbuch mehr als nur Angaben zur Zeichenhäufigkeit enthält. Im Lichte der weiter oben referierten Zahlen zu Korpus- bzw. Inventargrößen und Listen häufiger Zeichen scheint ein Inventar von 4.574 Schriftzeichen-Types ausreichend aussagekräftig zu sein.

#### **4. Das Basismodell und die chinesische Schrift**

Die Grundannahme der folgenden Übertragung des Basismodells von Köhler auf das chinesische Schriftzeichensystem ist, daß dieses System in bezug auf die hier untersuchten Eigenschaften und Prozesse eine Struktur hat, die der der Lexik entspricht, weshalb für die Zusammenhänge zwischen den Größen des Systems auch ein ähnliches Verhalten erwartet wird. Was die funktionalen Abhängigkeiten der Systemgrößen von einander betrifft, wird daher dieselbe Differentialgleichung als mathematisches Modell zugrunde gelegt, wie Köhler dies für sein Basismodell getan hat. Die Lösung der Differentialgleichung und ihre linearisierte Form werden natürlich ebenso übernommen.

Die untersuchte „Sprache“, das soll noch einmal ganz deutlich herausgestellt werden, ist im folgenden das chinesische Schriftzeichensystem und *nicht* die chinesische Sprache oder die Lexik der chinesischen Sprache. Rückschlüsse aus den weiter unten referierten Ergebnissen auf die chinesische Sprache oder ihre Lexik sind daher nicht zulässig.

*Der Inventarumfang* entspricht der Lexikongröße im Basismodell. Das Kodierungsbedürfnis (Kod) ist zu beschreiben als das Bedürfnis, Silben der chinesischen Sprache in Zeichen zu kodieren, die sich graphisch für verschiedene Morpheme unterscheiden. Je mehr Silben und Morpheme schriftlich dargestellt werden müssen, desto größer muß der Umfang des Zeicheninventars sein. Das ent-

gegenwirkende Bedürfnis nach Minimierung des Inventarumfangs (minI) wird dadurch bedient, daß viele Schriftzeichen mehrdeutig sind und für verschiedene Morpheme und Silben verwendet werden können.

Der Inventarumfang wird für die Überprüfung des Modells als die Anzahl der verschiedenen Schriftzeichen operationalisiert, die sich in dem untersuchten Korpus finden.

*Die Graphemzahl* im übertragenen Modell entspricht der Phonemzahl des Lexikmodells. Sie ist die Anzahl der bedeutungsunterscheidenden Zeichenelemente, die sich durch Minimalpaaranalyse als Grundbausteine der Schriftzeichen isolieren lassen.

Der Umfang des Grapheminventars wird durch die Bedürfnisse nach Minimierung des Kodieraufwands (minK) und Minimierung des Dekodieraufwands (minD) beeinflusst. ‚minK‘ verlangt, daß das Inventar möglichst klein und seine Elemente einander ähnlich und wenig komplex sind, damit man es sich leicht merken und seine Elemente schnell ausführen kann, ohne viele verschiedene Schreibbewegungen machen zu müssen. ‚minD‘ verlangt dagegen, daß es möglichst viele, optisch gut unterscheidbare Elemente gibt, die die Identifizierung des Schriftzeichens, zu dem sie sich zusammensetzen, erleichtern.

Zur Überprüfung der Hypothesen wurde das von Bohn durch eine modifizierte Minimalpaaranalyse ermittelte Grapheminventar zugrunde gelegt (Bohn, 1998:12-14). Die Graphemzahl ist die Anzahl der Bohnschen Grapheme, die in den Schriftzeichen (Types) des untersuchten Korpus vorkommen.

*Die graphische Komplexität* entspricht der Länge lexikalischer Einheiten im Basismodell. Wie bereits angedeutet unterscheiden sich auch chinesische Schriftzeichen hinsichtlich des Aufwands, den ihre schriftliche Artikulation kostet. Schriftzeichen mit mehr Strichen sind aufwendiger als solche mit weniger Strichen. Schriftzeichen mit mehr Graphemen sind aufwendiger als solche mit wenigen, selbst wenn sie insgesamt weniger Striche haben als jene. Wie Bohn (1998:20-24) gezeigt hat, gilt für das Verhältnis zwischen der mittleren Graphemzahl von Schriftzeichen und der durchschnittlichen Strichzahl ihrer Grapheme das Menzerathsche Gesetz. Das heißt, daß Schriftzeichen, die viele Grapheme haben, im Durchschnitt aus Graphemen mit weniger Strichen bestehen als Schriftzeichen, die sich aus wenigen Graphemen zusammensetzen. Jedoch ist die räumliche Anordnung einer größeren Anzahl von Graphemen in einem Schriftzeichen schwieriger als die Anordnung nur eines oder zweier Grapheme. Daher sollte die graphische Komplexität eines Schriftzeichens sowohl in Strichen als auch in Graphemen gemessen werden können.

Natürlich sind noch feinere Messungen für den Schreibaufwand denkbar, jedoch vergleichsweise aufwendig. Durchführbar ist dagegen die Berücksichtigung der verschiedenen Arten von Strichen, die unterschiedlich aufwendig sind. So

wurde die Komplexität der Schriftzeichen zur Überprüfung der Hypothesen auch als Summe der Aufwandsmaße ihrer Striche nach der Unterteilung von Bohn (1998:15) gemessen.

Der Inventarumfang und die Graphemzahl wirken in gleicher Weise auf die Zeichenkomplexität ein wie die Lexikongröße und die Phonemzahl auf die Länge lexikalischer Einheiten im Basismodell. Das Redundanzbedürfnis (Red), verhindert die Bildung allzu ähnlicher Zeichen und beeinflusst daher die Komplexität.

*Die Funktionskomplexität:* Die Eigenschaft chinesischer Schriftzeichen, zur Darstellung verschiedener Morpheme und Wörter verwendet werden zu können, soll als ihre Funktionskomplexität bezeichnet werden.

Es gibt einen Zusammenhang zwischen graphischer Komplexität und Funktionskomplexität, der durch das Spezifikationsbedürfnis (Spz) beeinflusst wird. ‚Spz‘ bewirkte, diachronisch betrachtet, daß Schriftzeichen, die zur schriftlichen Darstellung verschiedener Wörter oder Morpheme gebraucht wurden, durch Hinzufügung von Graphemen komplexer wurden. Dieser Prozeß führte z.B. dazu, daß das Schriftzeichen für *lái* ‚Weizen‘, urspr. 𪛗,<sup>5</sup> das aufgrund seiner Aussprache eine zeitlang zur Darstellung des gleichlautenden Wortes ‚kommen‘ entlehnt wurde, durch Hinzufügung eines Graphems für seine ursprüngliche Bedeutung komplexer wurde: 𪛗<sup>3</sup>. Das erweiterte Zeichen war in seiner Darstellungsfunktion auf die Wiedergabe der ursprünglichen Zeichenbedeutung, nämlich ‚Weizen‘ beschränkt. Ein phonetisch entlehntes Zeichen konnte aber auch für die neue Bedeutung, zu deren Darstellung es eine zeitlang verwendet worden war, durch Hinzufügung eines Graphems spezifischer werden. ‚Spz‘ wirkt also u.U. vergrößernd auf den Inventarumfang, was in der vorliegenden Modellierung in dem Einfluß des Kodierungsbedürfnisses auf den Inventarumfang enthalten sei.

Das Wissen um diesen historischen zeichenbildenden Prozeß läßt vermuten, daß im gegenwärtigen Schriftzeichensystem Zeichen mit mehr Graphemen durchschnittlich eine geringere Funktionskomplexität haben als Zeichen mit weniger Graphemen.

Auf die Funktionskomplexität wirken außerdem die Bedürfnisse ‚minK‘ und ‚minD‘ ein, die in einem Ausgleichsverhältnis zu einander stehen. ‚minD‘ strebt nach einer geringen Funktionskomplexität, da der Leser gern schnell und problemlos erkennen können möchte, welches Morphem durch ein Schriftzeichen jeweils repräsentiert werden soll. ‚minK‘ dagegen verlangt nach einer höheren Funktionskomplexität, da dies beim Schreiben die bequeme Verwendung weniger unterschiedlicher Schriftzeichen erlauben würde.

Die Funktionskomplexität selbst wirkt, indem sie das Bedürfnis nach Minimierung des Inventarumfangs bedient, verringern auf die Inventargröße zurück, da das Inventar kleiner sein kann, wenn die durchschnittliche Funktionskomplexi-

---

<sup>5</sup> Ich vereinfache hier etwas. Das archaische Zeichen sah noch etwas anders aus.

tät seiner Elemente größer ist, d.h., wenn viele Schriftzeichen zur schriftlichen Darstellung vieler verschiedener Morpheme verwendet werden können.

Ein Vergleich des übertragenen Modells mit Köhlers Basismodell zeigt, daß die Systemgröße Funktionskomplexität im übertragenen Modell eine Zusammenfassung der Teilstruktur des Basismodells ist, die den Zusammenhang zwischen den Größen Polylexie und Polytextie samt der auf letztere einwirkenden Systembedürfnisse nach Kontextspezifizierung und -ökonomie enthält. Derartige Vereinfachungen sind möglich und erlauben die Berechnung auch komplizierter Systeme (Köhler, 1986:48-49). Im Basismodell ist die Polytextie eine Funktion der Polylexie und die Frequenz eine Funktion der Polytextie. Ferner konnte für das Basismodell gezeigt werden, daß die Frequenz indirekt eine Funktion der Polylexie ist. Die Zusammenfassung dieser Teilstruktur zu einer Größe beschädigt mithin nicht das Modell. Der praktische Grund für diese Vereinfachung liegt darin, daß für die Schriftzeichen des untersuchten Inventars nur bekannt ist, in wieviel verschiedenen Wort-Types sie im Korpus vorkommen, nicht aber in wieviel verschiedenen Texten.

Der Zusammenhang zwischen Zeichenkomplexität und Funktionskomplexität im linearen Modell läßt sich ausdrücken durch die Gleichung<sup>6</sup>

$$L\text{-Funktionskomplexität} = Q_2 * \min K - Q_1 * \min D - T * L\text{-Komplexität.}$$

Zur Hypothesenüberprüfung wurde die Funktionskomplexität eines Schriftzeichens operationalisiert als die Anzahl der verschiedenen ein- und mehrsilbigen Wörter (Types) im Wortschatz des Korpus, an deren schriftlicher Darstellung das Schriftzeichen beteiligt ist.

*Die Frequenz:* Es ist anzunehmen, daß Schriftzeichen mit höherer Funktionskomplexität auch häufiger in Texten erscheinen. Die Frequenz von Schriftzeichen wäre dann eine Funktion ihrer Funktionskomplexität, wobei das Anwendungsbedürfnis zusätzlich auf die Frequenz wirkt. Dieser Zusammenhang wird im linearisierten Modell als direkt-proportionale Abhängigkeit dargestellt:

$$L\text{-Frequenz} = R * \text{Anw} + K * L\text{-Funktionskomplexität.}$$

Für jedes Schriftzeichen des untersuchten Korpus ist bekannt, wie oft es im Korpus vorkommt.

Noch einmal *zur graphischen Komplexität:* Für die chinesische Schrift ist ein Zusammenhang zwischen der Häufigkeit von Zeichen und ihrer graphischen Komplexität bekannt, ähnlich wie er für Wörter und ihre Länge gilt. Das Bedürfnis nach Minimierung des Produktionsaufwands (minP) führt dazu, daß Schrift-

---

<sup>6</sup> ‚L-‘ bedeutet, daß eine Größe in ihrer logarithmierten Form betrachtet wird.

zeichen, die besonders häufig gebraucht werden, auch besonders anfällig für graphische Vereinfachungen sind, sofern sie nicht schon sehr einfach sind. Man kann sogar sagen, daß ‚minP‘ die Entwicklung der chinesischen Schrift von komplexeren archaischen Formen zur auch heute noch verwendeten *Kai-shu*-Form sowie zu schnell schreibbaren handschriftlichen Varianten für den privaten oder halböffentlichen Gebrauch vorangetrieben hat. In dieser Entwicklung hat sich das Wirken von ‚minP‘ in der Zeit manifestiert. Indem die Schriftreformer in der Handschrift gebräuchliche Verkürzungen zum Standard erklärten, erkannten sie das Ergebnis dieser „natürlichen“ Entwicklung an. Auch bei synchronischer Betrachtung läßt sich erkennen, daß häufige Schriftzeichen durchschnittlich „kürzer“ bzw. weniger komplex sind als seltenere.

Die graphische Komplexität chinesischer Schriftzeichen wird modelliert in der Gleichung

$$L\text{-Komplexität} = A * L\text{-Inventarumfang} + Z * \text{Red} - P * L\text{-Graphemzahl} - N * L\text{-Frequenz}.$$

Das ganze übertragene Modell ist in Abbildung 1 wiedergegeben.

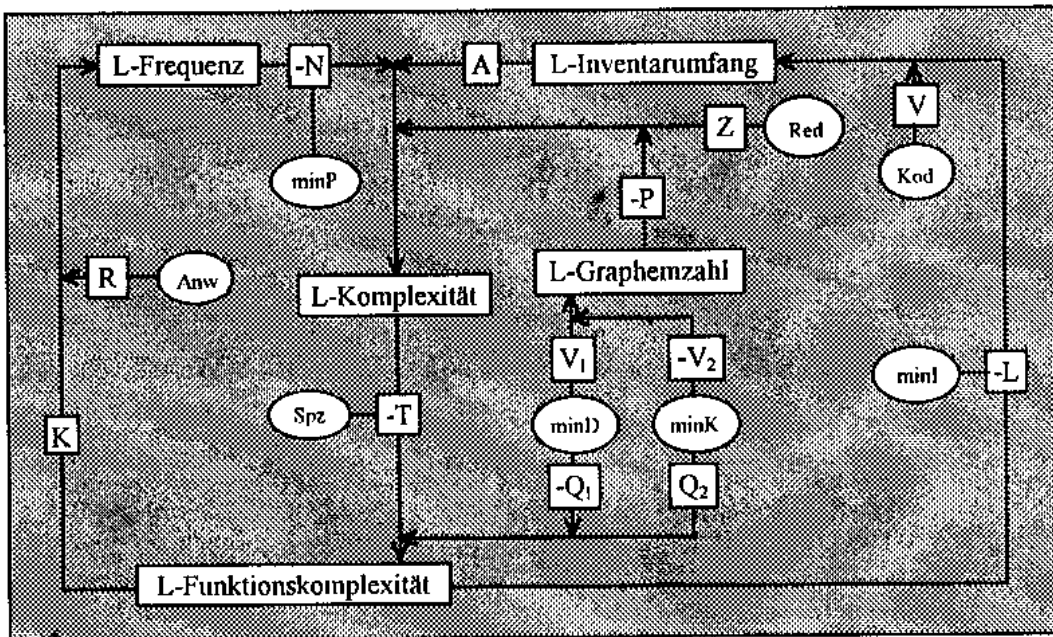


Abb. 1: Modell des chinesischen Schriftzeichensystems



## 4.1 Die Hypothesen

Aus dem übertragenen Modell lassen sich Hypothesen über drei direkte und drei indirekte Zusammenhänge ableiten. Bei den direkten funktionalen Abhängigkeiten handelt es sich um folgende:

$$H\ 1: \text{Funktionskomplexität} = A_1 * \text{Komplexität}^{B1}$$

$$H\ 2: \text{Frequenz} = A_2 * \text{Funktionskomplexität}^{B2}$$

$$H\ 3: \text{Komplexität} = A_3 * \text{Frequenz}^{B3}$$

Durch Einsetzen erhält man drei indirekte Abhängigkeiten:

$$H\ 4: \text{Komplexität} = A_4 * \text{Funktionskomplexität}^{B4}$$

$$H\ 5: \text{Funktionskomplexität} = A_5 * \text{Frequenz}^{B5}$$

$$H\ 6: \text{Frequenz} = A_6 * \text{Komplexität}^{B6}$$

## 5. Die Überprüfung der Hypothesen

Die sechs Hypothesen wurden unter Verwendung des linearisierten Modells überprüft. Dazu wurde das Statistikpaket SPSS<sup>7</sup> benutzt. Die multiple lineare Regression erfolgte mit der Methode der kleinsten Fehlerquadrate. Bei allen untersuchten Zusammenhängen gab es zu den Werten der unabhängigen Variable stets Mehrfachantworten (*replicated responses*). Es wurde daher jeweils der Mittelwert der Mehrfachantworten genommen und mit der Anzahl der beteiligten Werte gewichtet. Datenpunkte, deren Gewicht kleiner oder gleich 5 war, wurden in der Regel von der Regression ausgeschlossen. Wenn kein Ausschluß vorgenommen wurde, ist dies extra vermerkt. Als Maß für die Güte der Anpassung wurde der Determinationskoeffizient  $R^2$  - im folgenden einfach D - gewählt. Eine Anpassung wurde als gut betrachtet, wenn D mindestens den Wert 0,9 erreichte.<sup>8</sup>

Dezimalzahlen gebe ich oft auf zwei oder drei Stellen hinter dem Komma gerundet an. Allen Berechnungen liegen jedoch ungerundete Dezimalzahlen zugrunde.

### 5.1 Die direkten Abhängigkeiten

H 1: *Die Funktionskomplexität ist eine Funktion der graphischen Komplexität der Schriftzeichen. Sie ist geringer, wenn die graphische Komplexität höher ist. Im linearen Modell lautet die Gleichung für H 1*

---

<sup>7</sup> SPSS für Windows, Version 6.0.1.

<sup>8</sup> So geht auch Hammerl (1991:160) vor. Daß diese Vorgehensweise nicht unproblematisch ist, zeigt Grotjahn (1992).

$$L\text{-Funktionskomplexität} = \ln A + B * L\text{-Komplexität},$$

wobei für B ein negativer Wert erwartet wird.

Die graphische Komplexität wurde auf drei verschiedene Weisen gemessen, und zwar a) als Anzahl der Striche, b) als Anzahl der Grapheme und c) als die Summe der Aufwandsgewichtungen der Striche. Die Regression wurde für alle 3 Meßweisen durchgeführt, und zwar unter Einschluß aller Datensätze.

a) Strichzahl:	D = 0,956	A = e <sup>5,59</sup> = 268,12	B = -1,373
b) Graphemzahl:	D = 0,953	A = e <sup>3,666</sup> = 39,09	B = -1,133
c) Aufwandsmaß:	D = 0,95	A = e <sup>6,086</sup> = 439,72	B = -1,44

B ist wie erwartet negativ und hat überdies in allen drei Messungen sehr ähnliche Werte. Die Werte von A unterscheiden sich, weil die Meßwerte der drei verschiedenen Meßweisen in absoluten Zahlen verschieden hoch ausfallen. Die Abbildungen 2a - 2c zeigen die Datenpunkte sowie die Funktionskurve in nicht-logarithmierter Form.

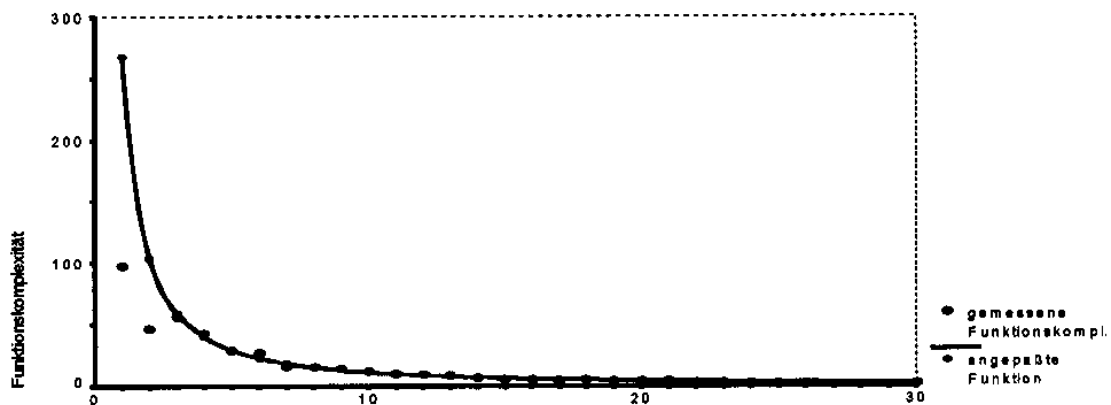


Abb. 2a

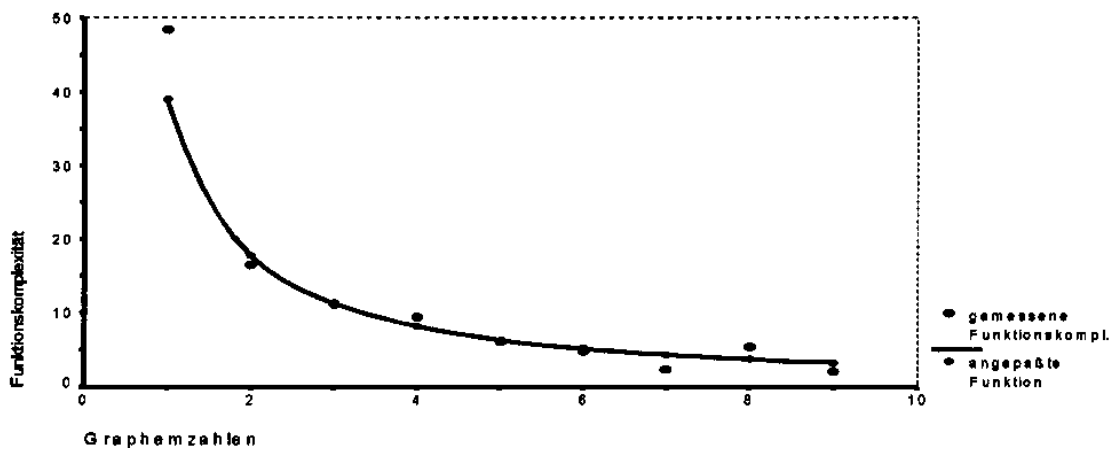


Abb. 2b

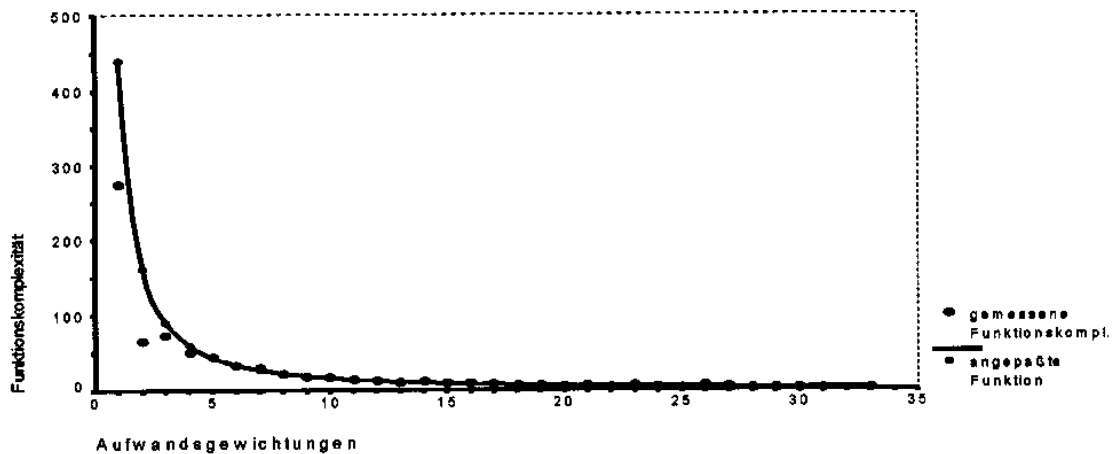


Abb. 2c

Sowohl die Güte der Anpassung als auch die augenscheinliche Übereinstimmung von Datenpunkten und Funktionskurven veranlassen dazu, die Hypothese als bestätigt zu betrachten.

H 2: Die Frequenz ist eine Funktion der Funktionskomplexität der Schriftzeichen. Sie ist höher für Schriftzeichen mit höherer Funktionskomplexität. Linearisiert wird H 2 ausgedrückt durch

$$L\text{-Frequenz} = \ln A + B * L\text{-Funktionskomplexität.}$$

Für B wird ein positiver Wert erwartet.

Für die Regression wurden alle Datensätze herangezogen. Die Anpassung war mit  $D = 0,958$  sehr gut. Abbildung 3 zeigt in nicht-logarithmierter Form die gemessenen Datenpunkte und die Funktionskurve mit aus den Daten geschätzten Parametern. Es werden nur die Datenpunkte mit einem Gewicht  $> 5$  dargestellt. Mit zunehmender Funktionskomplexität ist eine vermehrte Streuung der Datenpunkte um die Funktionskurve zu beobachten. Dennoch scheint die Funktionskurve die Tendenz der Abhängigkeit korrekt widerzuspiegeln. Daher und aufgrund der Güte der Regression soll auch diese Hypothese als bestätigt betrachtet werden.

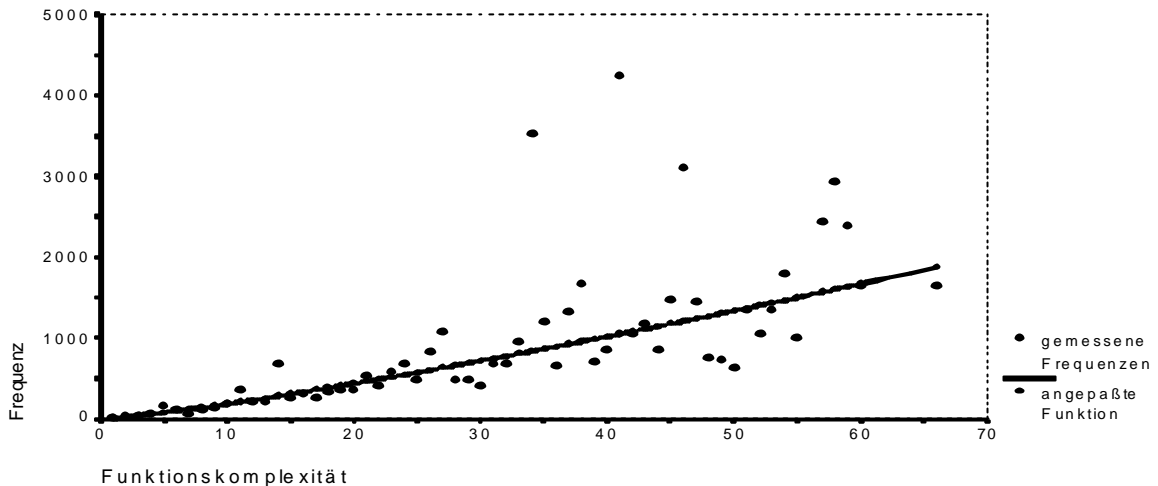


Abb. 3

H 3: Die graphische Komplexität der Schriftzeichen ist eine Funktion ihrer Frequenz, und zwar ist die graphische Komplexität im Durchschnitt geringer, wenn die Frequenz höher ist. In linearisierter Form lautet die Gleichung:

$$L\text{-Komplexität} = \ln A + B * L\text{-Frequenz},$$

wobei für B wieder ein negativer Wert erwartet wird.

Die Frequenz ist die absolute Häufigkeit jedes Schriftzeichens im Korpus. Die möglichen Frequenzwerte sind sehr disparat. Gerade unter den häufigen Schriftzeichen kommt der Fall, daß zwei Zeichen die gleiche Frequenz haben, praktisch nicht vor. Um die Frequenz sinnvoll als unabhängige Variable verarbeiten zu können, wurden die Daten daher in Frequenzklassen zusammengefaßt. Die Klassen wurden zusätzlich nach der Anzahl der Datenpunkte, die in ihnen zusammengefaßt waren, gewichtet. Es wurden die Klassenbreiten 50 und 100 gewählt. Frequenzklassen, die fünf Datenpunkte oder weniger zusammenfaßten, wurden von der Regression ausgeschlossen. Als Wert der unabhängigen Variable wurde jeweils die Klassenmitte gewählt. Die Regression wurde für jede der drei Meßweisen der graphischen Komplexität mit beiden Klassenbreiten durchgeführt. Die folgende Tabelle zeigt die Ergebnisse.

<u>Meßweise</u>	<u>Klassenbreite 100</u>	<u>Klassenbreite 50</u>
a) Strichzahl	D = 0,94	D = 0,93
	A = e <sup>2,846</sup> = 17,22	A = e <sup>2,72</sup> = 15,18
	B = -0,114	B = -0,094
b) Graphemzahl	D = 0,95	D = 0,897
	A = e <sup>1,51</sup> = 4,53	A = e <sup>1,4</sup> = 4,066
	B = -0,958	B = -0,78

c) Aufwandsmaß	$D = 0,946$	$D = 0,92$
	$A = e^{3,057} = 21,28$	$A = e^{2,94} = 18,88$
	$B = -0,11$	$B = -0,09$

B hat jeweils das erwartete negative Vorzeichen und liegt in fast jedem Fall nahe bei -0,1, wobei die Unterschiede für die Zusammenfassung der Daten in Frequenzklassen der Breite 50 etwas größer ausfallen als bei der Klassenbreite 100. Was die Werte von A betrifft, so unterscheiden sie sich natürlich erheblich zwischen den Ergebnissen für die verschiedenen Meßweisen, jedoch nur wenig zwischen den Regressionsergebnissen mit verschieden breiten Frequenzklassen bei gleicher Meßweise.

Die Abbildungen 4a - 4c beziehen sich auf die Daten mit der Klassenbreite 100, die Abbildungen 5a - 5c auf die mit der Klassenbreite 50. Datenpunkte und Funktionskurve werden jeweils in nicht-logarithmierter Form dargeboten. Es ist zu erkennen, daß in jedem Fall die Datenpunkte mit zunehmender Frequenz stärker um die Funktionskurve streuen.

Da die Funktionsanpassung stets gut bis sehr gut war und die Funktionskurven auch optisch einen recht überzeugenden Eindruck machen, wird diese Hypothese ebenfalls als bestätigt betrachtet. Das Ausschlußkriterium, das Frequenzklassen mit fünf Datenpunkten oder weniger von der Regression ausnimmt, führt jedoch zu einer erheblichen Reduktion der berücksichtigten Datenpunkte. Insbesondere werden Schriftzeichen mit sehr hohen Frequenzen ausgeschlossen, da hier die Frequenzklassen besonders dünn besetzt sind. Die Hypothese gilt also hauptsächlich für Schriftzeichen mit mittlerer und niedriger Häufigkeit, die - und das ist andererseits auch zu bedenken - den Großteil des Inventars ausmachen.

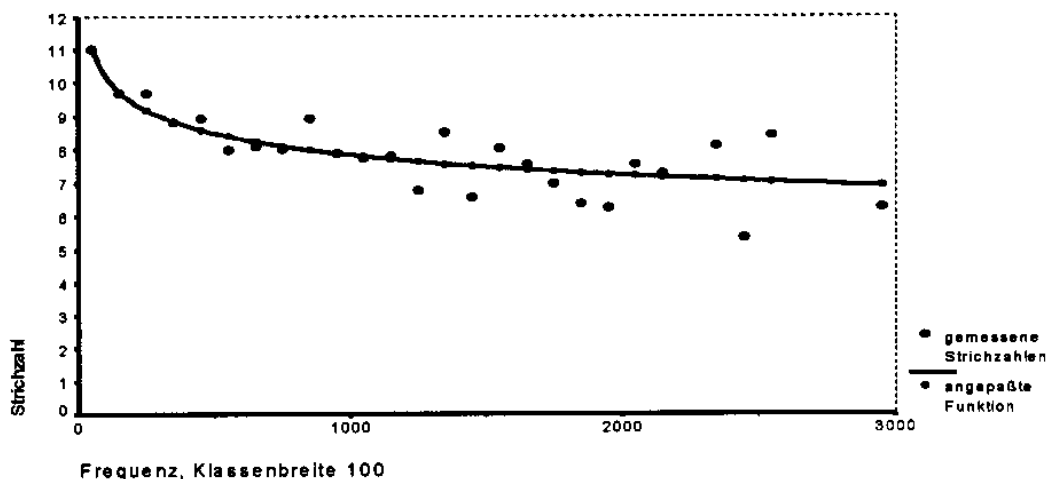


Abb. 4a

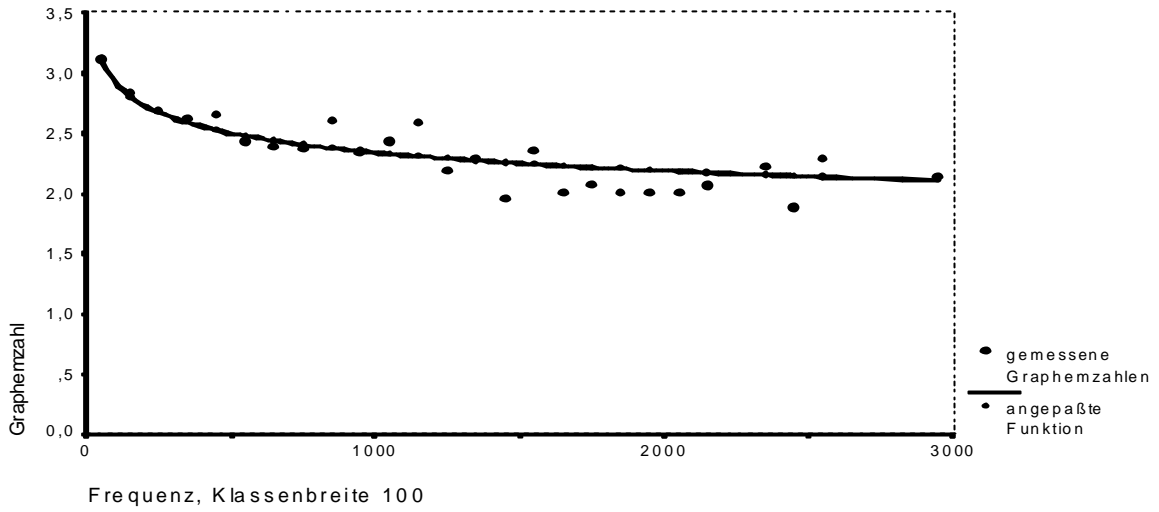


Abb. 4b

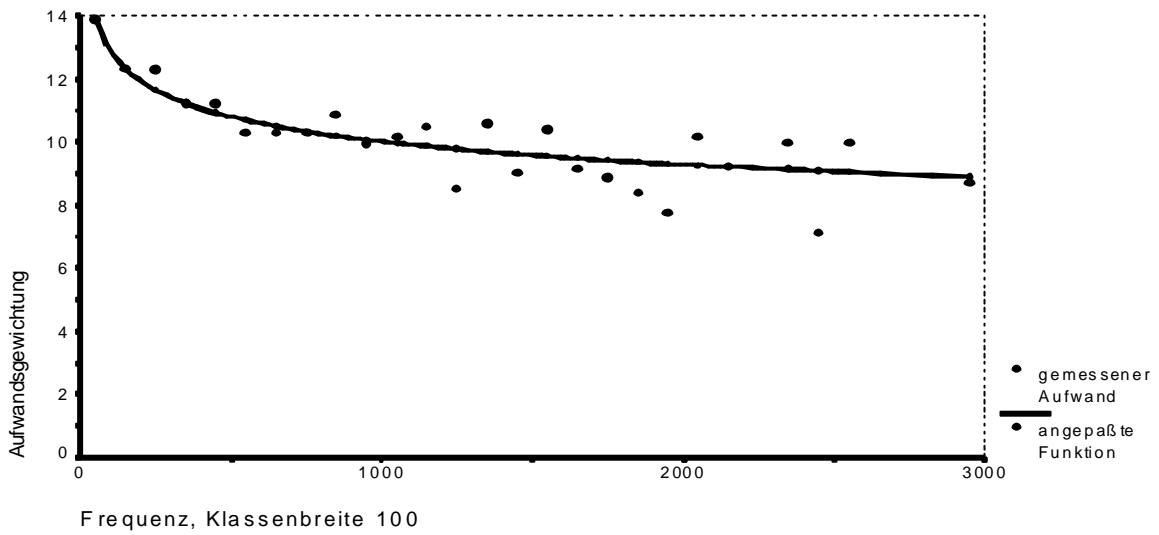


Abb. 4c

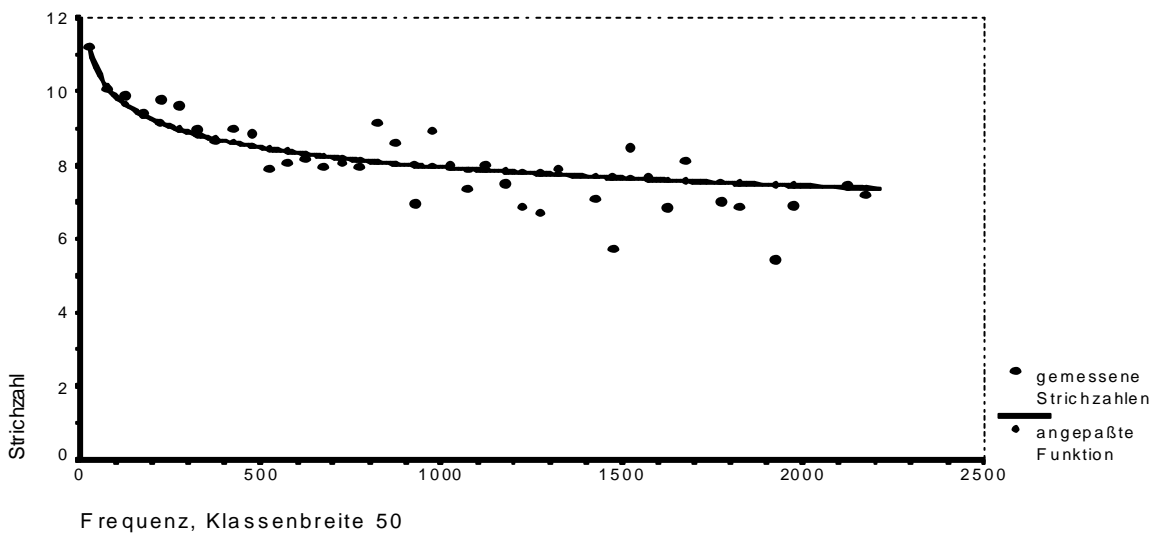


Abb. 5a

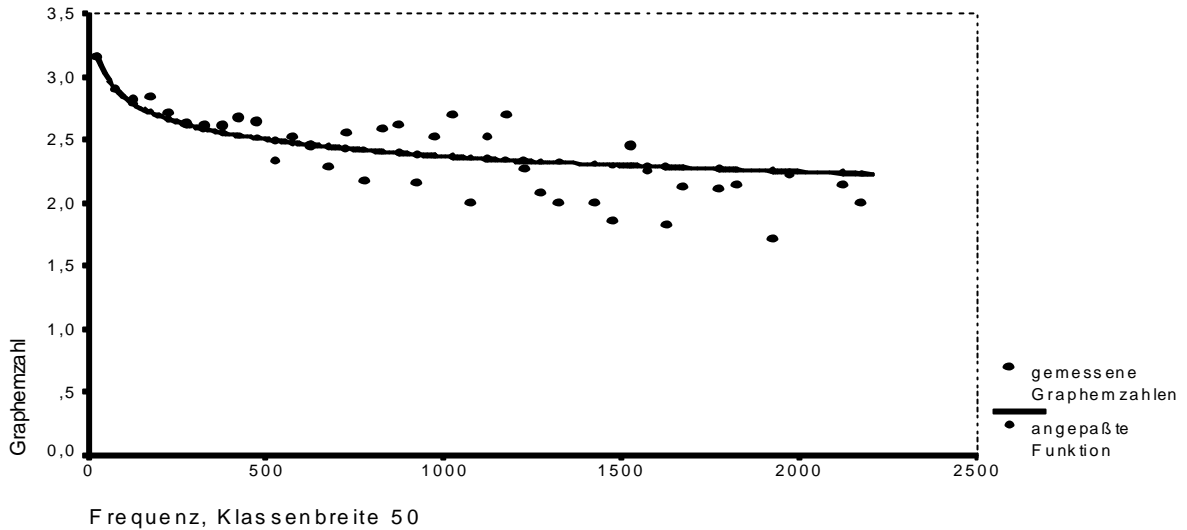


Abb. 5b

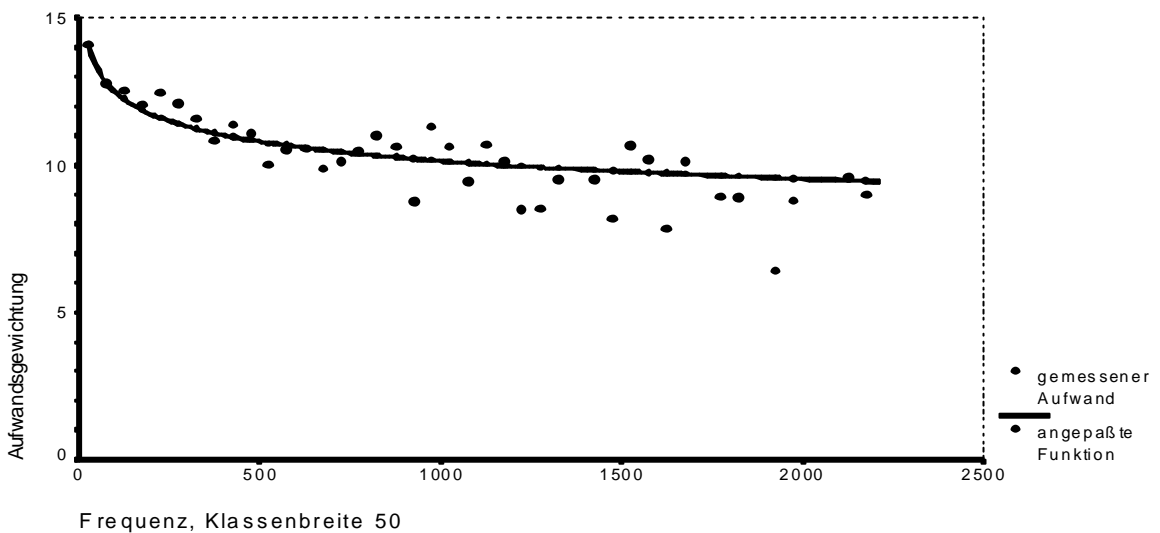


Abb. 5c

Die Regression wurde auch für die ungefilterten Daten ausgeführt, also ohne Ausschluß von Datenpunkten. Sie fiel mit  $D = 0,76$  bis  $0,85$  schlechter aus als für die gefilterten Daten. Die hierbei aus den Daten geschätzten Parameterwerte unterscheiden sich jedoch nur geringfügig von denen, die aus den gefilterten Daten geschätzt wurden.

Die Hypothesen über direkte Abhängigkeiten H1, H2 und H3 können vorläufig angenommen werden, da alle vorgenommenen Überprüfungen gute bzw. sehr gute Ergebnisse lieferten. Darüber hinaus ist auch der optische Eindruck, den die angepaßten Funktionskurven im Verhältnis zu den Datenpunkten machen, ziemlich überzeugend. Auf dieser Grundlage werden nun die indirekten Abhängigkeiten überprüft.

## 5.2 Die indirekten Abhängigkeiten

Für die direkten Abhängigkeiten sind die aus den Daten geschätzten Funktionsparameter nun bereits bekannt. Durch Einsetzung können jetzt die theoretischen Funktionsparameter für die Funktionen, die indirekte Abhängigkeiten modellieren, errechnet werden. Sie werden anschließend mit den empirischen Funktionsparametern, die für die indirekten Zusammenhänge ebenfalls aus den Daten geschätzt werden, verglichen. Ein statistischer Test soll zeigen, ob zwischen den theoretischen und empirischen Werten der Funktionsparameter ein signifikanter Unterschied besteht. Wenn kein signifikanter Unterschied besteht, so kann man sagen, daß das Modell den indirekten Zusammenhang korrekt voraussagt. Sollte der Unterschied jedoch signifikant sein, so muß die Vorhersage durch das Modell abgelehnt werden.

H 4: *Die graphische Komplexität von Schriftzeichen ist indirekt eine Funktion ihrer Funktionskomplexität, wobei die Frequenz diese indirekte Abhängigkeit vermittelt. In linearisierter Form lautet die Gleichung*

$$L\text{-Komplexität} = \ln X + Y * L\text{-Funktionskomplexität.}$$

Unter Berücksichtigung der verschiedenen Meßweisen für die graphische Komplexität und der verschiedenen Klassenbreiten der Frequenzdaten ergeben sich folgende sechs theoretische Voraussagen:<sup>9</sup>

a) graphische Komplexität gemessen in Strichzahlen:

$$\begin{aligned} \mathbf{L\text{-Komp}_{a1}} &= 2,72 - 0,094 * (2,444 + 1,215 * L\text{-Fk}) \\ &= \mathbf{2,49 - 0,114 * L\text{-Fk}} && \text{bzw.} \\ \mathbf{L\text{-Komp}_{a2}} &= 2,85 - 0,114 * (2,444 + 1,215 * L\text{-Fk}) \\ &= \mathbf{2,57 - 0,138 * L\text{-Fk}} \end{aligned}$$

b) graphische Komplexität gemessen in Graphemzahlen:

$$\begin{aligned} \mathbf{L\text{-Komp}_{b1}} &= 1,4 - 0,078 * (2,444 + 1,215 * L\text{-Fk}) \\ &= \mathbf{1,2 - 0,095 * L\text{-Fk}} && \text{bzw.} \\ \mathbf{L\text{-Komp}_{b2}} &= 1,51 - 0,96 * (2,444 + 1,215 * L\text{-Fk}) \\ &= \mathbf{1,277 - 0,116 * L\text{-Fk}} \end{aligned}$$

---

<sup>9</sup> ‚Komp‘ steht für die graphische Komplexität, ‚Fk‘ für die Funktionskomplexität. Der tiefgestellte Index 1 weist darauf hin, daß die Parameter des direkten Zusammenhangs zwischen Frequenz und graphischer Komplexität aus Daten mit der Frequenzklassenbreite 100 gewonnen wurden. Der Index 2 steht, wenn diese Parameter aus Daten mit der Frequenzklassenbreite 50 ermittelt wurden.



c) graphische Komplexität gemessen in Strichaufwandsgewichtungen:

$$\begin{aligned} \mathbf{L\text{-Komp}_{c1}} &= 2,94 - 0,09 * (2,444 + 1,215 * \mathbf{L\text{-Fk}}) \\ &= \mathbf{2,72 - 0,109 * \mathbf{L\text{-Fk}} \quad \text{bzw.}} \\ \mathbf{L\text{-Komp}_{c2}} &= 3,06 - 0,109 * (2,444 + 1,215 * \mathbf{L\text{-Fk}}) \\ &= \mathbf{2,79 - 0,13 * \mathbf{L\text{-Fk}} \end{aligned}$$

Die Regression hatte folgende Ergebnisse:

- a) Strichzahl:     D = 0,73     A = e<sup>2,55</sup> = 12,82     B = -0,116  
 b) Graphemzahl:   D = 0,60     A = e<sup>1,25</sup> = 3,49     B = -0,092  
 c) Aufwandsmaß:   D = 0,75     A = e<sup>2,78</sup> = 16,19     B = -0,114

Die Abbildungen 6a - 6c zeigen jeweils die empirischen Datenpunkte und die Funktionskurve mit aus den Daten ermittelten Parametern.

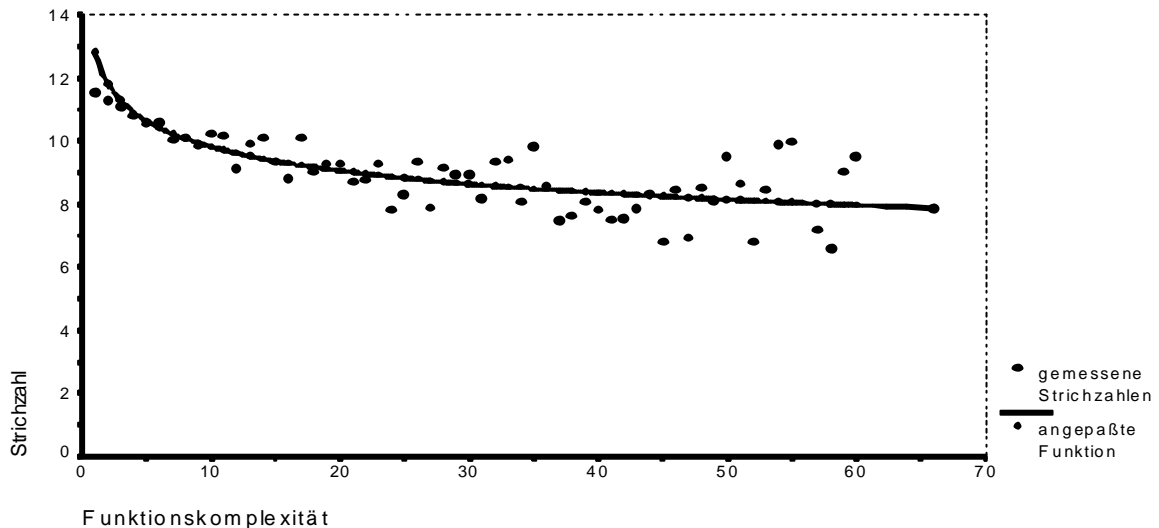


Abb. 6a

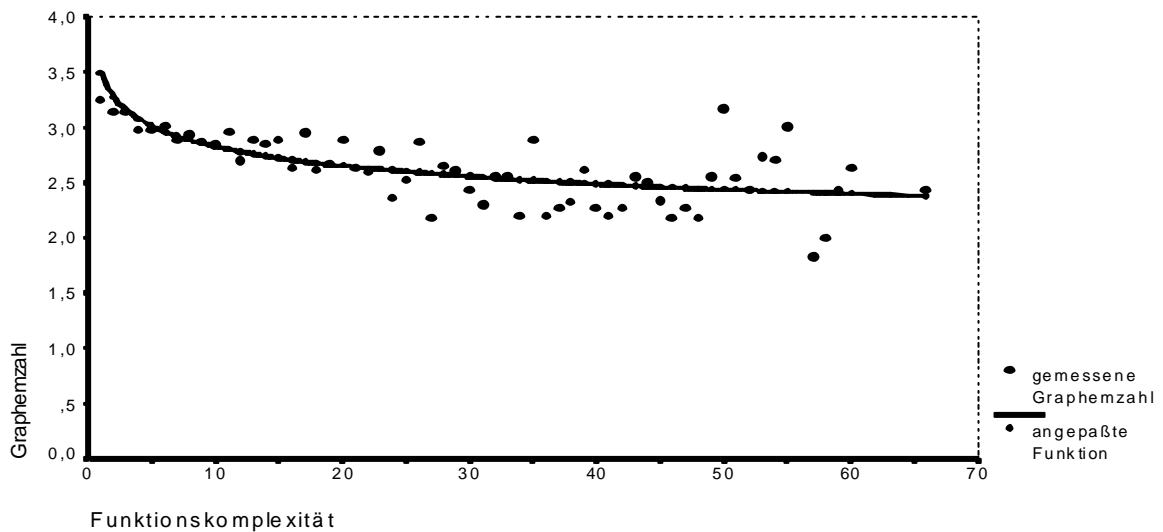


Abb. 6b

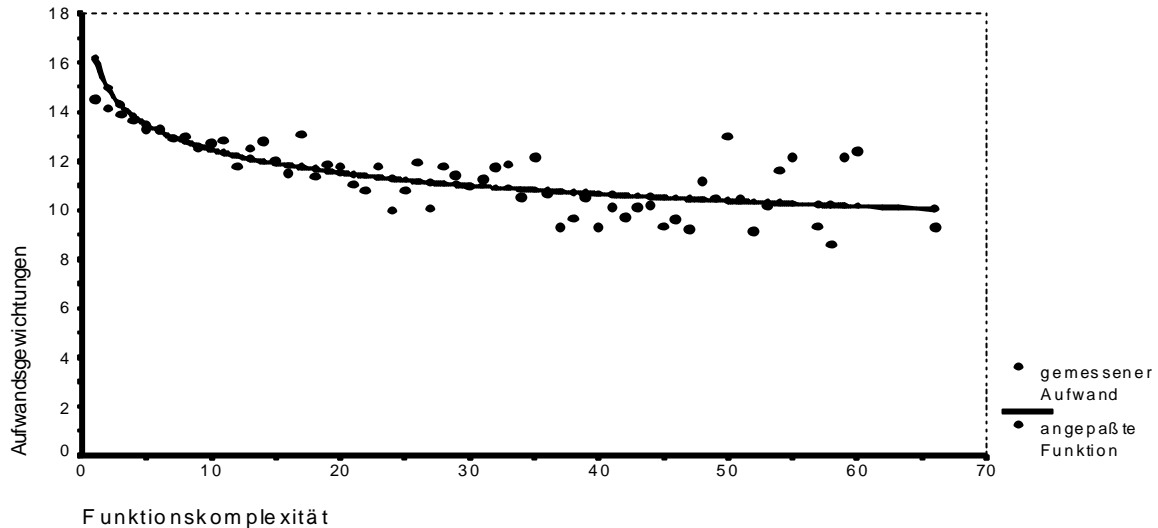


Abb. 6c

D ist nicht sehr gut ausgefallen. Die gefundenen Parameterwerte scheinen auf den ersten Blick jedoch gut zu den theoretisch vorausgesagten Werten zu passen. Die folgende Tabelle gibt eine Übersicht über die Funktionen mit den theoretisch erwarteten und den empirisch ermittelten Parameterwerten.

<u>Meßweise</u>	<u>theoretisch</u>	<u>empirisch</u>
a) Strichzahl	$Komp_{a1} = 12,06 * Fk^{-0,114}$ $Komp_{a2} = 13,04 * Fk^{-0,138}$	$Komp_{ae} = 12,82 * Fk^{-0,116}$
b) Graphemzahl	$Komp_{b1} = 3,36 * Fk^{-0,095}$ $Komp_{b2} = 3,59 * Fk^{-0,116}$	$Komp_{be} = 3,49 * Fk^{-0,092}$
c) Aufwandsmaß	$Komp_{c1} = 15,16 * Fk^{-0,109}$ $Komp_{c2} = 16,3 * Fk^{-0,13}$	$Komp_{ce} = 16,19 * Fk^{-0,114}$

Dem oberflächlichen Eindruck nach ist die Übereinstimmung zwischen der jeweils ersten theoretischen Funktion und der Funktion mit den aus den Daten geschätzten Parametern größer als zwischen der zweiten theoretischen Funktion und der angepaßten Funktion. Im Fall a) ist aus Abbildung 7a zu ersehen, daß sich die Kurven der beiden theoretischen Funktionen etwa bei  $Fk = 24$  überschneiden. Insgesamt bleibt die Funktionskurve für  $Komp_{a1}$  in etwa gleichem Abstand von der an die Daten angepaßten Funktionskurve, während sich die Kurve für  $Komp_{a2}$  mit wachsender Funktionskomplexität weiter entfernt. Vergleichbares gilt für den Fall b). In Abbildung 7c ist zu beobachten, daß sich die Kurve für  $Komp_{c2}$  in erheblich größerer Entfernung von der an die Daten angepaßten Funktionskurve befindet als die für  $Komp_{c1}$ .

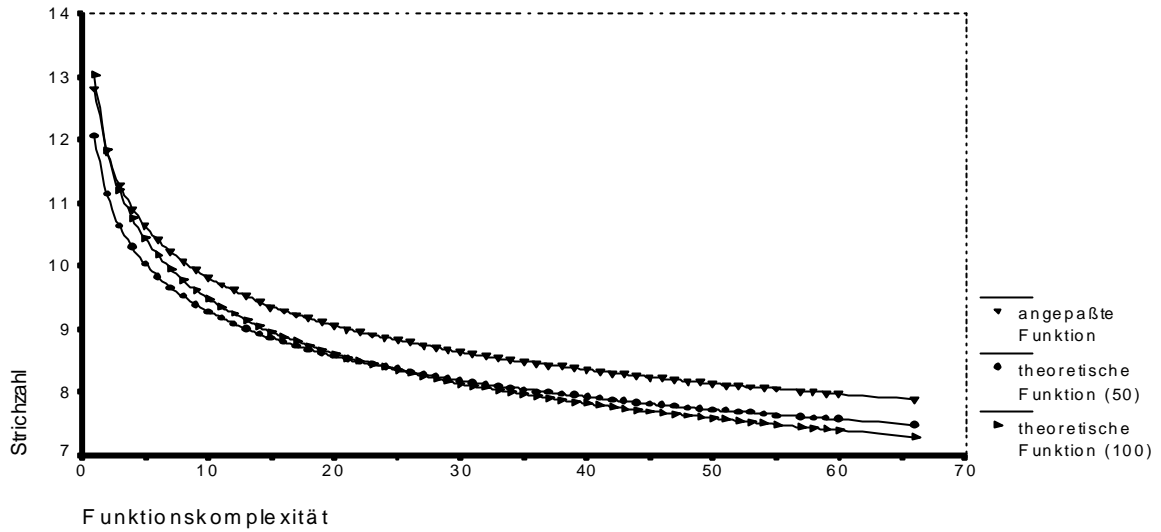


Abb. 7a

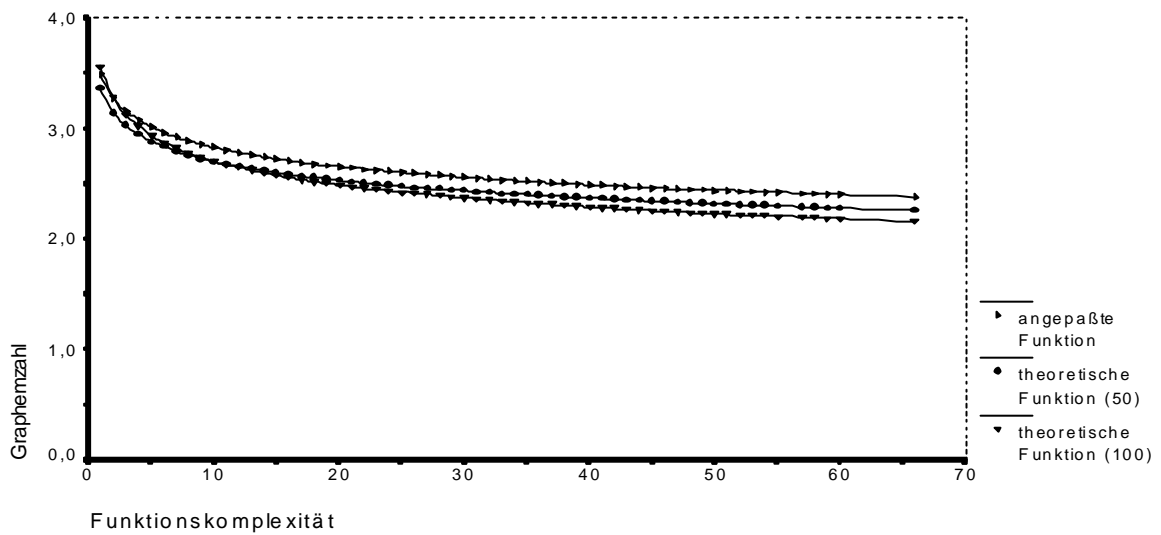


Abb. 7b

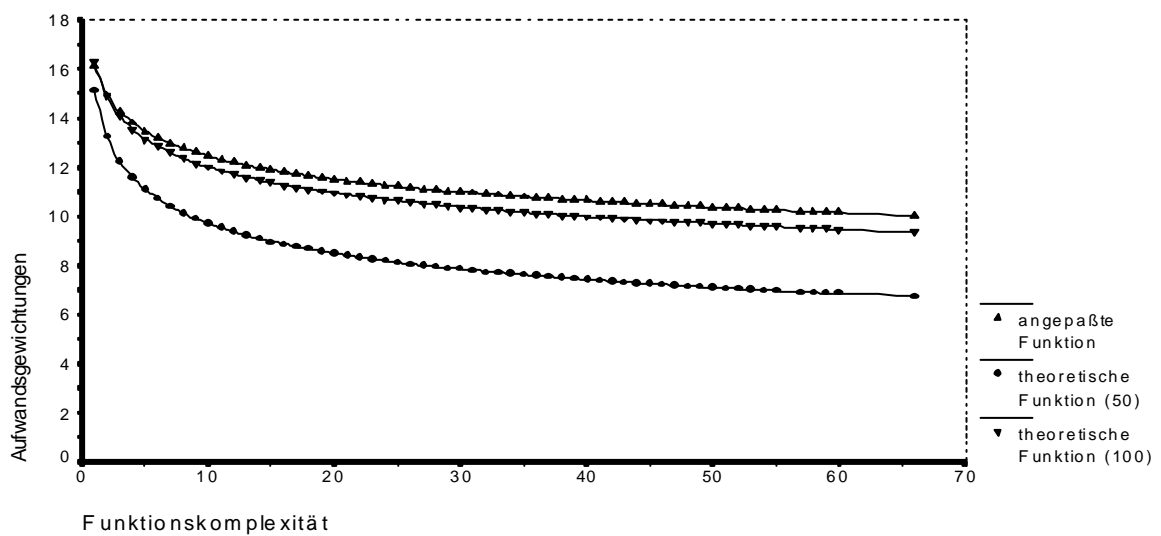


Abb. 7c

Köhler (1986) hat zur Bewertung der Unterschiede zwischen den theoretisch vorausgesagten Funktionswerten und den Werten der an die empirischen Daten angepaßten Funktion den *t*-Test genommen. Dies ist ein Test zum Vergleich von Mittelwerten. Ich bin ebenso vorgegangen.

Der *t*-Test, dessen Ergebnisse ich nicht im einzelnen referiere, ergab für alle sechs möglichen Vergleiche signifikante Mittelwertsunterschiede. Aufgrund dieses Ergebnisses muß die Hypothese zunächst abgelehnt werden. Da die Anpassung an die empirischen Daten auch nicht befriedigend gewesen ist, wird nach einer Verbesserung dieses Teils des Modells zu suchen sein.

In der linearisierten Form der Grafiken aufgrund logarithmierter Daten (hier nicht wiedergegeben) war in allen drei Fällen zu erkennen, daß die Gerade für die jeweils zweite vorgeschlagene theoretische Funktion (tiefgestellter Index 2) annähernd parallel zur Geraden für die angepaßte Funktion verlief, während dies bei der Geraden für die erste vorgeschlagene theoretische Funktion nicht der Fall war. Dies scheint mir ein Hinweis darauf zu sein, daß die Zusammenfassung in Frequenzklassen der Breite 50 zu besseren Ergebnissen führt als die Zusammenfassung in Klassen der Breite 100. Darüber hinaus könnte die augenscheinliche Parallelität der Geraden ein Indiz dafür sein, daß ein im Modell bisher unberücksichtigter Faktor für die Diskrepanz zwischen den theoretisch erwarteten und den empirisch gewonnenen Parametern verantwortlich ist. In der vorliegenden Form muß die Hypothese zwar abgelehnt werden, aber eine Verbesserung, die den bisher unberücksichtigten Einfluß mit einbezieht, wird die untersuchte Hypothese als Bestandteil enthalten, so daß man sie als einen Schritt in die richtige Richtung bezeichnen kann.

H 5: *Die Funktionskomplexität ist indirekt eine Funktion der Frequenz, wobei die graphische Komplexität diese indirekte Abhängigkeit vermittelt. Linearisiert lautet die Gleichung*

$$L\text{-Funktionskomplexität} = \ln X + Y * L\text{-Frequenz.}$$

Da es für die graphische Komplexität drei Meßweisen gibt und die Frequenz zum Zwecke der Regression in Klassen mit zwei verschiedenen Breiten zusammengefaßt wurden, sind insgesamt wiederum sechs verschiedene theoretische Funktionen möglich.<sup>10</sup>

$$\begin{aligned} \text{a) } \quad \mathbf{L-Fk_{a1}} &= 5,59 - 1,373 * (2,85 - 0,114 * L\text{-Freq}) \\ &= \mathbf{1,68 + 0,156 * L-Freq} \qquad \text{bzw.} \end{aligned}$$

---

<sup>10</sup> Die Namen der Systemgrößen werden wieder abgekürzt, so daß ‚Freq‘ für Frequenz steht. Die Zahlen im tiefgestellten Index haben die gleiche Bedeutung wie oben.

$$\begin{aligned}
 \mathbf{L-Fk_{a2}} &= 5,59 - 1,373 * (2,72 - 0,094 * \mathbf{L-Freq}) \\
 &= \mathbf{1,85 + 0,13 * \mathbf{L-Freq}} \\
 \text{b) } \mathbf{L-Fk_{b1}} &= 3,666 - 1,133 * (1,51 - 0,096 * \mathbf{L-Freq}) \\
 &= \mathbf{1,95 + 0,108 * \mathbf{L-Freq}} && \text{bzw.} \\
 \mathbf{L-Fk_{b2}} &= 3,666 - 1,133 * (1,4 - 0,08 * \mathbf{L-Freq}) \\
 &= \mathbf{2,076 + 0,088 * \mathbf{L-Freq}} \\
 \text{c) } \mathbf{L-Fk_{c1}} &= 6,086 - 1,441 * (3,06 - 0,109 * \mathbf{L-Freq}) \\
 &= \mathbf{1,68 + 0,157 * \mathbf{L-Freq}} && \text{bzw.} \\
 \mathbf{L-Fk_{c2}} &= 6,086 - 1,441 * (2,94 - 0,09 * \mathbf{L-Freq}) \\
 &= \mathbf{1,85 + 0,13 * \mathbf{L-Freq}}
 \end{aligned}$$

Da die unabhängige Variable die Frequenz war, wurden die Daten wiederum in Frequenzklassen der Breiten 50 und 100 zusammengefaßt und für die Regression von der Klassenmitte ausgegangen. Die Anpassungsergebnisse lauteten:

Klassenbreite 100:	D = 0,969	A = e <sup>-1,649</sup> = 0,192	B = 0,804
Klassenbreite 50:	D = 0,97	A = e <sup>-1,173</sup> = 0,31	B = 0,74

Die angepaßten Funktionskurven sind in den Abbildungen 8.1 und 8.2 zu sehen. Es fällt auf, daß die Datenpunkte ab einer Frequenz von ca. 800 unterhalb der Funktionskurve streuen.

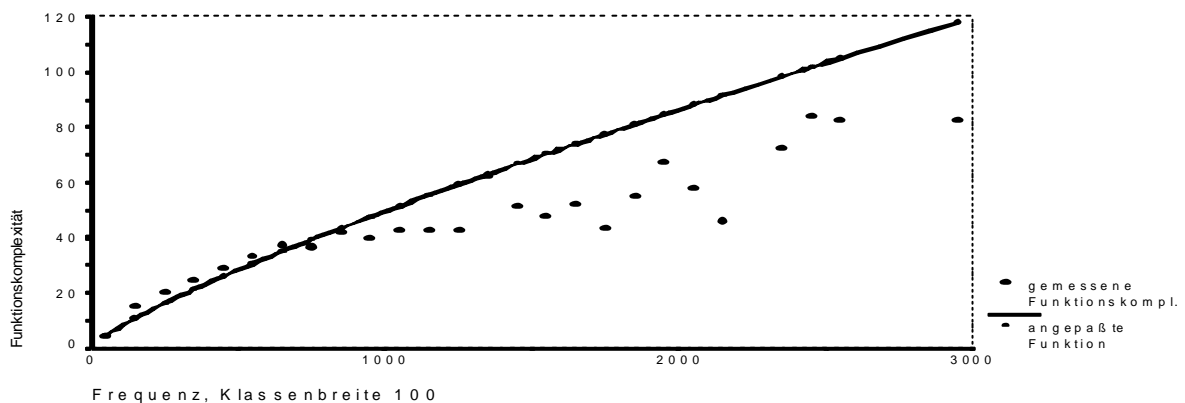


Abb. 8.1

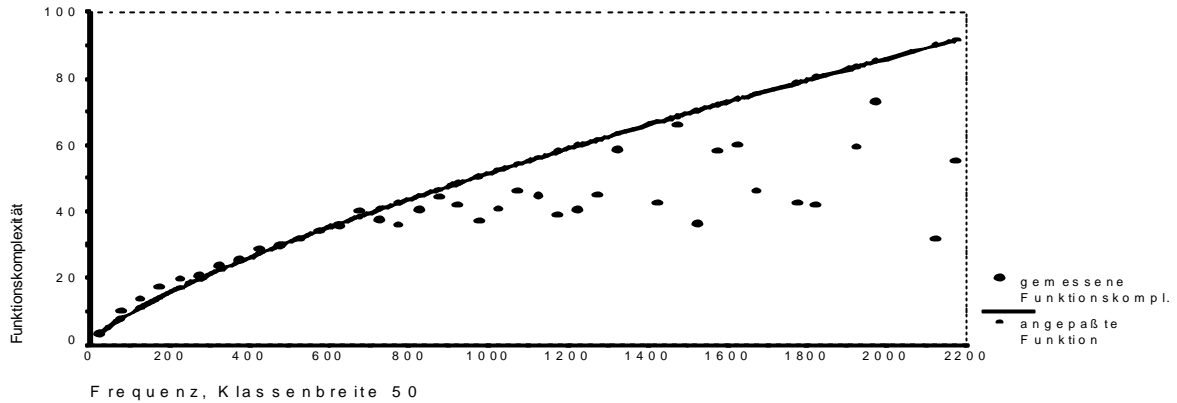


Abb. 8.2

Um den Vergleich zwischen den theoretisch erwarteten und den empirisch geschätzten Parametern zu erleichtern, werden die vergleichbaren Potenzfunktionen in der folgenden Übersicht zusammengestellt.

	<u>theoretisch</u>	<u>empirisch</u>
100:	$Fk_{a1} = 1,68 * Freq^{0,156}$ $Fk_{b1} = 7,05 * Freq^{0,108}$ $Fk_{c1} = 5,36 * Freq^{0,157}$	$Fk_{e1} = 0,192 * Freq^{0,804}$
50:	$Fk_{a2} = 1,85 * Freq^{0,13}$ $Fk_{b2} = 7,98 * Freq^{0,088}$ $Fk_{c2} = 6,37 * Freq^{0,13}$	$Fk_{e2} = 0,31 * Freq^{0,74}$

Während die Parameterwerte der theoretischen Funktionen sich in ihren Größenordnungen ähneln, ist in allen Fällen die Diskrepanz zwischen den Voraussetzungen und den empirisch gewonnenen Funktionsparametern augenfällig.

Die Abbildungen 9.1 und 9.2 zeigen jeweils die Kurve der angepaßten Funktion sowie die drei Kurven der entsprechenden theoretischen Funktionen. Es ist deutlich zu sehen, daß die drei theoretischen Kurven jeweils eng beieinander liegen, während die an die Daten angepaßte Kurve viel steiler verläuft.

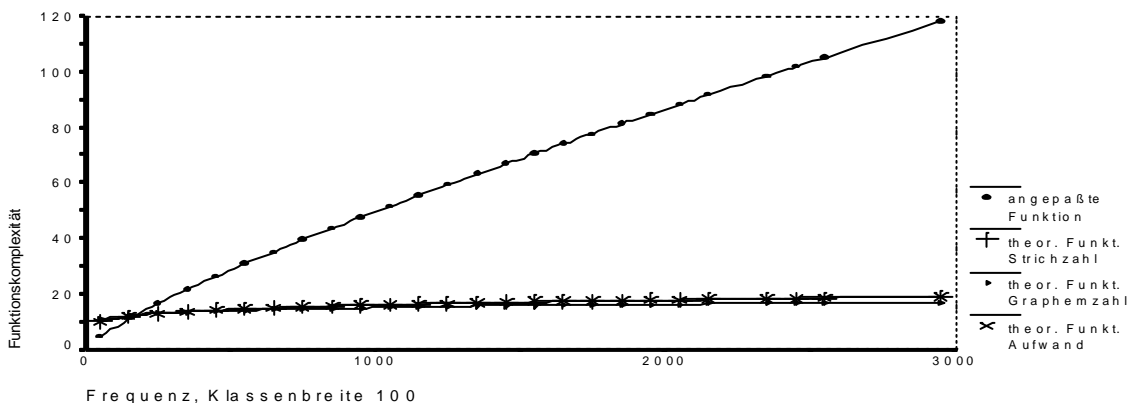


Abb. 9.1

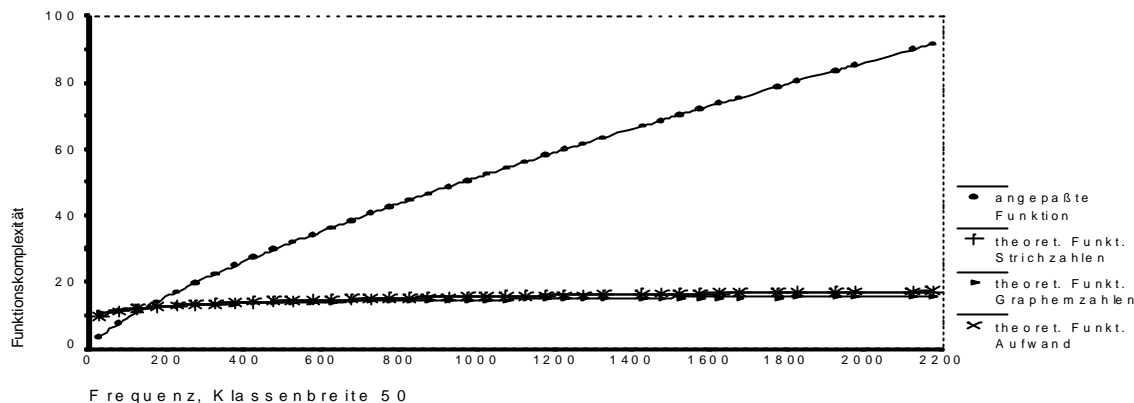


Abb. 9.2

Der *t*-Test zeigte auch hier, daß signifikante Unterschiede zwischen den Mittelwerten der verschiedenen Funktionen bestehen. Die Anpassung an die Daten war sehr gut, so daß ein Zusammenhang zu bestehen scheint. Die indirekte Hypothese in der vorliegenden Form muß jedoch abgelehnt werden, da sie den zu beobachtenden Zusammenhang nicht korrekt vorhersagt.

Die Filterung der Daten nach ihrem Gewicht, die bei der Überprüfung der hier vermittelnden Abhängigkeit der graphischen Komplexität von der Frequenz (H 3) vorgenommen worden war, kann für die erheblichen Abweichungen zwischen vorhergesagten und aus den Daten geschätzten Parametern nicht verantwortlich gemacht werden, denn wie weiter oben referiert, hatte die Überprüfung von H 3 unter Einbeziehung aller Datenpunkte zwar eine schlechtere Anpassungsgüte zur Folge, die geschätzten Parameter unterschieden sich jedoch nur äußerst geringfügig von den aus den gefilterten Daten geschätzten.

H 6: *Die Frequenz von Schriftzeichen ist indirekt eine Funktion ihrer graphischen Komplexität, wobei die Funktionskomplexität diese indirekte Abhängigkeit vermittelt:*

$$L\text{-Frequenz} = \ln X + Y * L\text{-Komplexität.}$$

Da die graphische Komplexität auf drei verschiedene Weisen gemessen werden kann, sind hier drei theoretische Vorhersagen möglich:

a) Strichzahl  $L\text{-Freq}_a = 2,444 + 1,215 * (5,59 - 1,373 * L\text{-Komp})$   
 $= 9,24 - 1,67 * L\text{-Komp}$

b) Graphemzahl  $L\text{-Freq}_b = 2,444 + 1,215 * (3,666 - 1,133 * L\text{-Komp})$   
 $= 6,9 - 1,377 * L\text{-Komp}$

c) Aufwandsmaß  $L\text{-Freq}_c = 2,444 + 1,215 * (6,086 - 1,441 * L\text{-Komp})$   
 $= 9,84 - 1,75 * L\text{-Komp}$

Die Regression hatte folgende Ergebnisse:

- a)  $D = 0,93$      $A = e^{11,077} = 64.690,26$      $B = -2,466$   
 b)  $D = 0,955$      $A = e^{7,63} = 2.058,5$      $B = -1,98$   
 c)  $D = 0,88$      $A = e^{11,675} = 117.557,75$      $B = -2,47$

Im Fall c) war die Anpassung nicht ganz befriedigend.

Die Abbildungen 10a - 10c zeigen jeweils die Datenpunkte und die angepaßte Funktionskurve. Die Abweichungen im Bereich niedriger Strichzahlen, Graphemzahlen und Aufwandsgewichtungen sind durch Datenpunkte mit niedrigen Gewichten zu erklären.

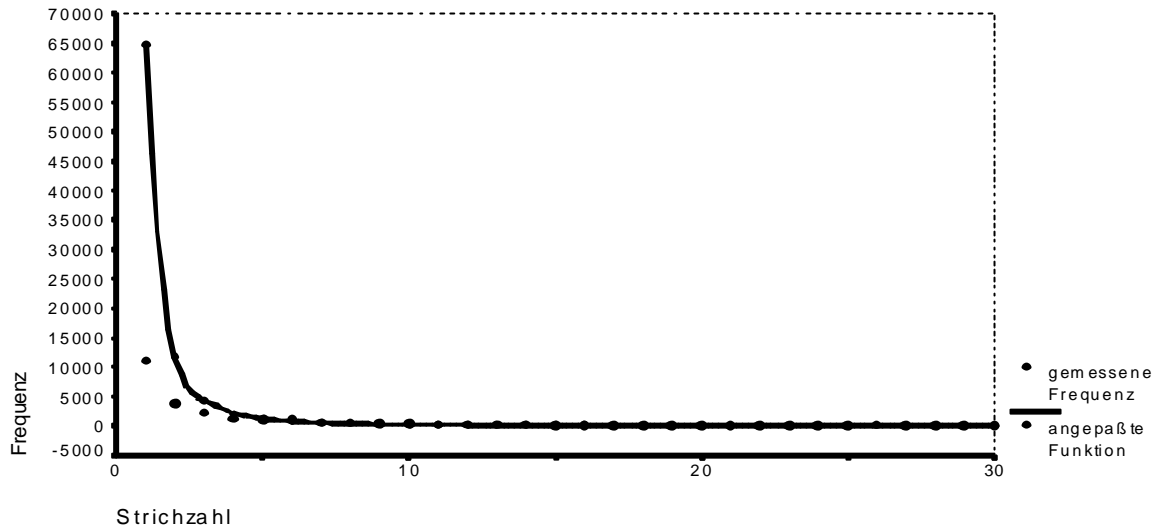


Abb. 10a

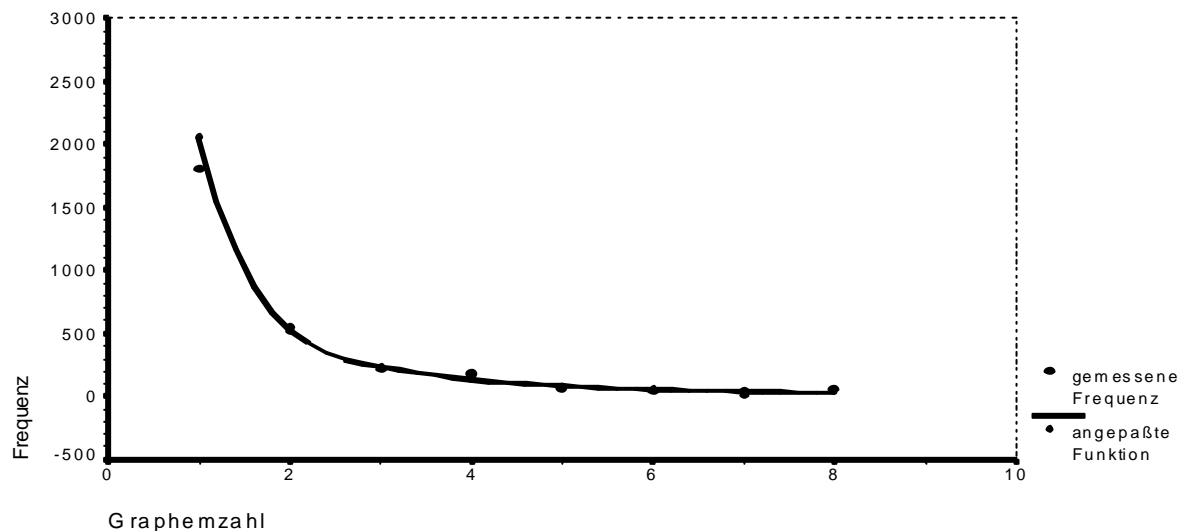


Abb. 10b



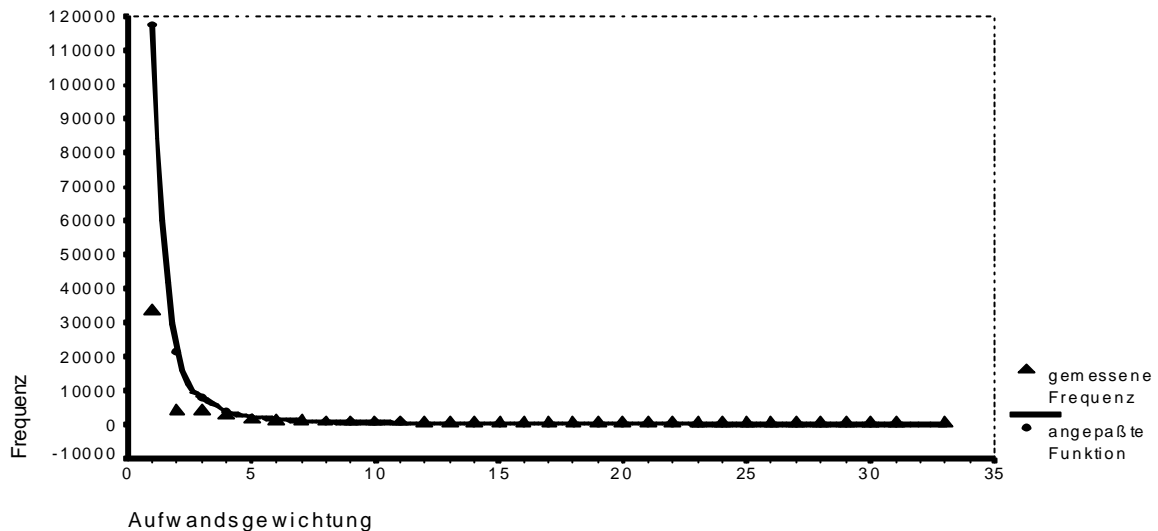


Abb. 10c

Die folgende Übersicht zeigt die Potenzfunktionen mit den theoretisch vorhergesagten und aus den empirischen Daten geschätzten Parameterwerten im Vergleich:

<u>theoretisch</u>	<u>empirisch</u>
a) $\text{Freq}_a = 10.287,14 * \text{Komp}^{-1,67}$	$\text{Freq}_{ea} = 64.690,26 * \text{Komp}^{-2,466}$
b) $\text{Freq}_b = 992,27 * \text{Komp}^{-1,377}$	$\text{Freq}_{eb} = 2.058,5 * \text{Komp}^{-1,98}$
c) $\text{Freq}_c = 18.797,89 * \text{Komp}^{-1,75}$	$\text{Freq}_{ec} = 117.557,75 * \text{Komp}^{-2,47}$

Die theoretischen und empirischen Werte der Exponenten ähneln einander jeweils in gewissem Maße, jedoch sind die empirischen Werte höher als die theoretischen, und zwar jeweils etwa um 0,6 bis 0,7.

Besonders auffällig sind die Größenunterschiede unter den Werten des Multiplikators. Die Unterschiede der theoretisch vorhergesagten Multiplikatorwerte untereinander sind viel geringer als die der empirisch gewonnenen untereinander. Letztere unterscheiden sich von ihrem theoretisch vorhergesagten Wert um Größenordnungen. Dabei ist zu beobachten, daß die Abweichungen zwischen theoretischem und empirischem Wert der beiden Parameter am geringsten sind, wenn die graphische Komplexität der Schriftzeichen in der Anzahl ihrer Grapheme gemessen wird (b).

Die Güte der Anpassungen besonders für die Messungen der graphischen Komplexität in Strichzahl (a) und Graphemzahl (b) legt nahe, daß tatsächlich eine Abhängigkeit der Frequenz von der graphischen Komplexität vorliegt. Es ist jedoch möglich, daß das Modell einer Verbesserung bedarf und diese Abhängigkeit nicht als durch die Systemgröße „Funktionskomplexität“ vermittelt zu modellieren ist oder ein weiterer Einfluß berücksichtigt werden muß, der gegenwärtig nicht im Modell enthalten ist.

Die Abbildungen 11a - 11c zeigen die Funktionskurven. Die theoretische und die an die Daten angepaßte Kurve fallen schon relativ früh zusammen. Für die Fälle a) und c) zeigt ein Vergleich mit den Abbildungen 10a und 10c, daß die theoretischen Kurven den Daten besser gerecht werden als die Kurven der angepaßten Funktion, was durch die niedrigen Gewichte der empirischen Datenpunkte mit geringer graphischer Komplexität zu erklären ist.

Auch in diesem Fall, der sehr positiv aussieht, soll mit Hilfe des *t*-Tests geklärt werden, wie die Unterschiede zwischen den theoretisch erwarteten Parametern und den empirisch geschätzten zu bewerten sind.

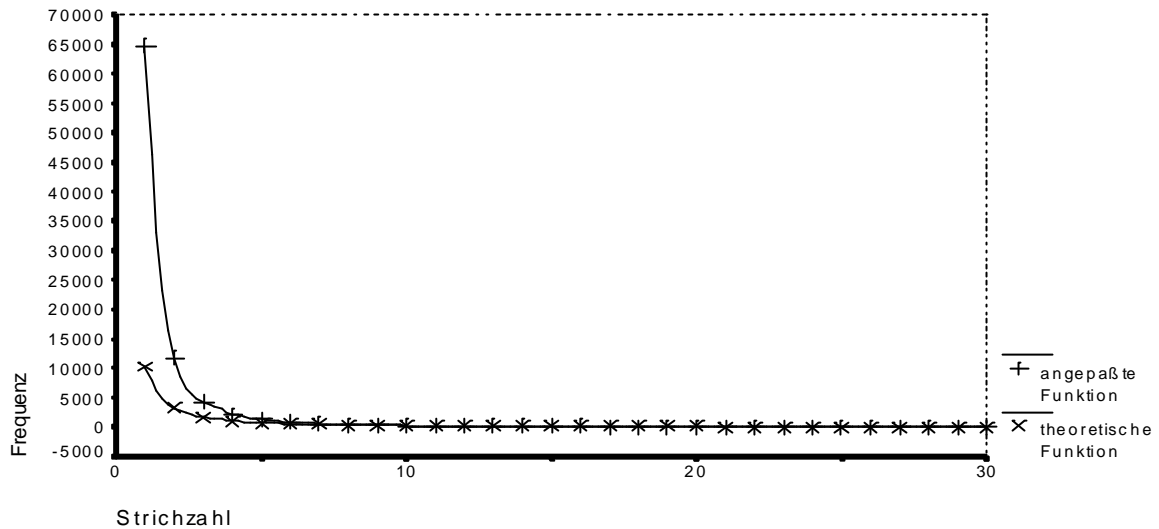


Abb. 11a

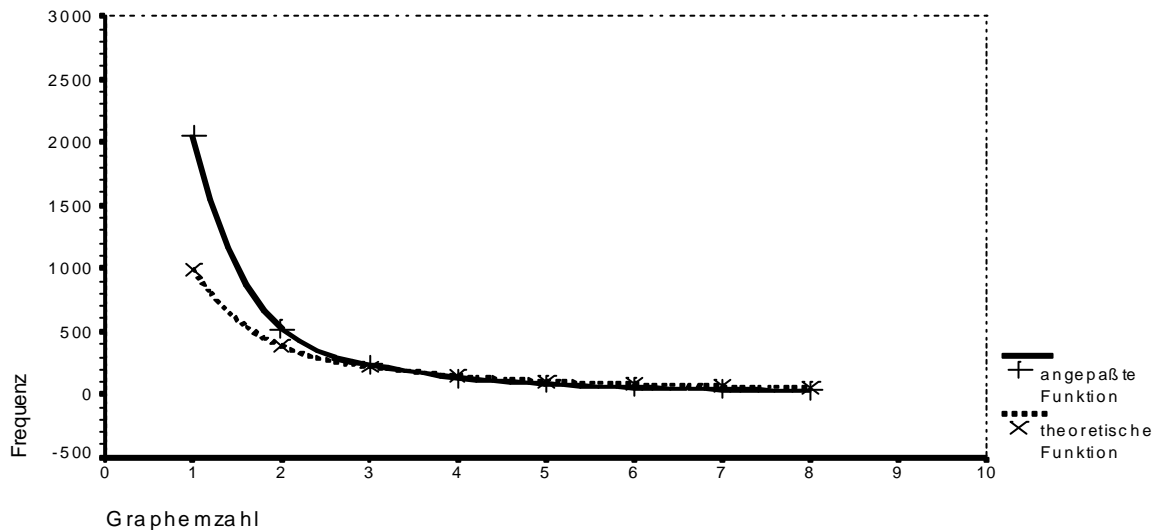


Abb. 11b

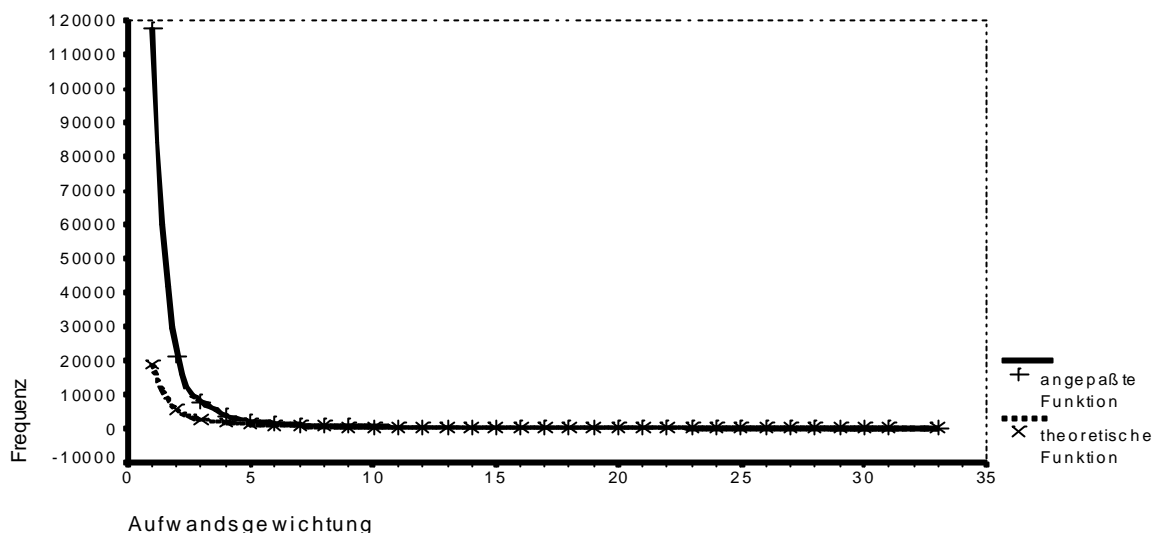


Abb. 11c

Der *t*-Test ergab für alle drei Vergleiche, die für diese Hypothese zu machen waren, keine signifikanten Mittelwertunterschiede.<sup>11</sup> Die Hypothese kann damit angenommen werden, was durch den optischen Eindruck der Funktionskurven und die hohen Werte des Determinationskoeffizienten bei der Anpassung an die empirischen Daten ebenfalls nahegelegt wird.

Im Gegensatz zu den Hypothesen über direkte Abhängigkeiten müssen zwei der drei Hypothesen über indirekte Abhängigkeiten abgelehnt werden. Einzig H 6 hielt der Überprüfung stand.

Die Hypothese H 4 mußte zwar auch abgelehnt werden, jedoch gibt es Anzeichen, daß eine Verbesserung des Modells und damit zusammenhängend eine Verfeinerung der Hypothese gute Ergebnisse bringen könnte. Daher ist das Modell hier als ein Schritt in die richtige Richtung zu werten.

Für die Hypothese H5, die die Abhängigkeit der Funktionskomplexität von der Frequenz betraf, wichen die theoretischen, aus dem Modell abgeleiteten Vorhersagen und das Anpassungsergebnis an die empirischen Daten sehr stark voneinander ab. Hier ist eine Verbesserung des Modells auf alle Fälle nötig, die sich u.U. nicht auf die Berücksichtigung weiterer Faktoren beschränken wird.

---

<sup>11</sup> Die Ergebnisse der drei durchgeführten zweiseitigen *t*-Tests waren:

- a) *t*-Wert = -0,31 auf einem Wahrscheinlichkeitsniveau von 0,755 bei 58 Freiheitsgraden
- b) *t*-Wert = 0,23 auf einem Wahrscheinlichkeitsniveau von 0,82 bei 16 Freiheitsgraden
- c) *t*-Wert = 0,00 auf einem Wahrscheinlichkeitsniveau von 1,0 bei 62 Freiheitsgraden

In allen drei Fällen wurde vor dem *t*-Test ein *F*-Test (nach Levene) auf Varianzhomogenität durchgeführt, der ebenfalls signifikant ausfiel (Fall a):  $F = 3,391$ ,  $P = 0,071$ ; Fall b):  $F = 1,06$ ,  $P = 0,319$ , Fall c):  $F = 2,827$ ,  $P = 0,98$ ).

## 6. Schlußbemerkungen

Die Übertragung von Köhlers Basismodell der Lexik auf das chinesische Schriftzeichensystem hat sich als lohnendes Experiment erwiesen. Die Überprüfung der Hypothesen über direkte Abhängigkeiten hat gezeigt, daß diese Abhängigkeiten auch im chinesischen Schriftzeichensystem bestehen. Für sie erwies sich das Modell als korrekt.

Für zwei von drei Hypothesen über indirekte Abhängigkeiten machte das Modell keine ausreichend korrekten Vorhersagen. Ich denke, daß die Operationalisierung der Funktionskomplexität eine Fehlerquelle sein könnte, da u.U. die Anzahl der Wort-Types, in denen ein Schriftzeichen im Korpus verwendet wird, nur ein ungenaues Maß dafür ist, zur Darstellung wie vieler verschiedener Morpheme ein Schriftzeichen tatsächlich dient. Die Ergebnisse der Überprüfung von Hypothese H 5 („die Funktionskomplexität ist indirekt eine Funktion der Frequenz“) wichen jedoch wahrscheinlich nicht allein aus diesem Grunde so stark von den Vorhersagen ab. Ich vermute, daß das Modell hier noch nicht vollständig ist.

Als eine spezifische Manifestation menschlichen Sprachvermögens und Sprechhandelns weist das chinesische Schriftzeichensystem Zusammenhänge auf, die auch in anderen Subsystemen von Sprache, etwa der Lexik, zu finden sind. Dies zu zeigen war ein Ziel dieser Arbeit. Darüber hinaus hoffe ich, durch die Untersuchung eines eher unzugänglichen Aspekts einer eher „fernliegenden“ Sprache gezeigt zu haben, daß das Basismodell auch auf andere Analyseebenen sinnvoll angewandt werden kann und dadurch neue Erkenntnisse über das Modell selbst wie auch über den Untersuchungsgegenstand gewonnen werden können.

## 7. Literatur

- Bohn, H.** (1998). *Quantitative Untersuchungen der modernen chinesischen Sprache und Schrift*. Hamburg: Verlag Dr. Kovac.
- DeFrancis, J.** (1984). *The Chinese language. Fact and fantasy*. Honolulu: University of Hawaii Press.
- Grotjahn, R.** (1992). Evaluating the adequacy of regression models: Some potential pitfalls. In B. Rieger (Hg.), *Glottometrika 13* (S. 121-172), Bochum: Brockmeyer.
- Hammerl, R.** (1991). *Untersuchungen zur Struktur der Lexik: Aufbau eines lexikalischen Basismodells*. Trier: Wissenschaftlicher Verlag Trier (Zugl.: Warschau, Univ., Habilitationsschrift, 1991)
- Köhler, R.** (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.
- Xiandai Hanyu pinlü cidian*** [Häufigkeitswörterbuch der modernen chinesischen Sprache] (1986). Beijing: Yuyan Xueyuan chubanshe.
- Zhou, Youguang** (1978). Xiandai Hanyu zhong shengpang de biaoyin gongneng wenti. [Das Problem der Phonetizität der Phonetika in den modernen chinesischen Schriftzeichen.] In ders. (1986). *Zhongguo yuwen de xiandaihua* (S. 141-151), [Die Modernisierung der Sprache und Schrift Chinas.] Shanghai: Jiaoyu chubanshe. (Zuerst erschienen in: *Zhongguo yuwen* 3/1978, 172-177.)