QUANTITATIVE LINGUISTICS

Vol. 14

GLOTTOMETRIKA 4

edited

by

W. Lehfeldt

U. Strauss

Studienverlag Dr. N. Brockmeyer

Bochum 1982

QUANTITATIVE LINGUISTICS

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Glottometrika. – Bochum: Studienverlag Brockmeyer
4. Ed. by W. Lehfeldt; U. Strauss. – 1982.
(Quantitative linguistics; Vol. 14)
ISBN 3-88339-250-2

NE: Lehfeldt, Werner [Hrsg.]; GT

ISBN 3-88339-250-2

Querenburger Höhe 281, 4630 Bochum 1
Druck Thiebes GmbH & Co Kommanditgesellschaft Hagen

We would like to express our gratitude

to the

STIFTUNG VOLKSWAGENWERK

a generous grant from which made possible the publication of the German edition of this book.

INHALT

Quantitative Analyse phonetischer Transkripte (Umgangssprache)

Peter Winkler
Universität Konstanz

1.1 Einleitung

Statistische Analysen von Lauthäufigkeiten und Lautkombinationen sind in der Linguistik und Phonetik schon verschiedentlich durchgeführt worden. SEGAL (1972, 22ff.) zitiert eine Reihe von Autoren, die bereits vor der Jahrhundertwende statistische Auswertungen von Sprachmaterial publiziert haben, das entweder schrifttextlich oder in phonetischen Transkriptionen vorlag. Später hatte die Phonologie die Anwendung quantitativer Methoden in der Linguistik besonders gefördert und angeregt; die Mehrzahl statistischer Sprachlautanalysen sind phonematisch orientiert. Mit phonetischer Zielsetzung sind noch weitere Applikationen quantitativ-statistischer Methoden möglich, von denen einige im Überblick von KARLGREN (1974[3], 129ff.) aufgeführt sind. Auch hier ließe sich eine Tradition bis vor 1900 zurückverfolgen, in der in jüngerer Zeit die Phonometrie von ZWIRNER das prominenteste Beispiel ist.

Quantitative Auswertungen phonetischer Transkripte sind im Wesentlichen für drei verschiedene Bereiche eingesetzt worden:

(1) Methodologie: Untersuchung der phänomenalen und perzeptiven Leistung von Transkribenten (z.B. OLLER und EILERS 1975 über phonetische Erwartbarkeit); Leistungsspektren von Transkriptionssystemen, Validitäts- und Reliabilitätsstudien (vgl. RICHTER und RICHTER 1981).

(2) Ausspracheuntersuchungen: Standard-, Umgangs- und Dialektaussprache, alltägliche, künstlerische, rhetorische Artikulationsstile; Basisuntersuchungen für die Kodifizierung von Lautungsformstufen oder für die Herstellung von Übungsmaterial (z.B. ORTMANN 1975), Sprechbewegungsabläufe (LINDNER 1975), soziolinguistische Erhebungen zur Aussprachevariation u.a.

(3) Paralinguistik: Beschreibung sprechereigener Artikulationsbesonderheiten, situations- und kontextgebundener Unterschiede, der Anpassung verschiedener Sprecher untereinander im Artikulationsniveau usw.

In der nachfolgenden quantitativen Analyse wird je ein Einzelproblem aus den o.g. Bereichen bearbeitet:

(1) Die auditive Festlegung von Lautlängen (Segmentdauer) bei ohrenphonetischen Segmentationen.

(2) Lauthäufigkeit und -kombinatorik umgangssprachlich beeinflußter Aussprache.

(3) Individuelle Artikulationseigenheiten zweier Sprecher während eines dyadischen Gesprächs im colloquial style.

1.2 MATERIAL UND METHODEN

Es werden die Transkripte der Tonaufnahmen zu TAKE D des Konstanzer Forschungsprojektes "Analyse unmittelbarer Kommunikation und Interaktion als Zugang zum Problem der Entstehung sozialwissenschaftlicher Daten" (finanziert von der Fritz-Thyssen-Stiftung; Leitung: Prof. Th. Luckmann/ Prof. Dr. P. Gross) zugrundegelegt.

(1) Sprachmaterial: Die Aufnahmen entstammen einer realen Gesprächssituation, in der untrainierte Sprecher (ein weiblicher und ein männlicher Gesprächsteilnehmer) frei ein zehnminütiges Gespräch gestalteten. Das Rahmenthema dieser Dyade wurde nur global vorgegeben. Die Aufnahmen enthalten "small-Talk"-Sequenzen, die in umgangssprachlich beeinflußten Sprechstil ausgeführt wurden. Die Sprecherin läßt in der Artikulation, teilweise auch in der Wortwahl schwäbisch-alemannische Dialektanklänge erkennen; der Sprecher zeigt Einflüsse der berliner Aussprache. Typisch für die Umgangsaussprache ist auch in dieser Aufnahme, daß die Sprecher frei zwischen hochgelauteten und dialektal eingefärbten Realisierungen oszillieren. Der schwäbisch-alemannische Einfluß zeigte sich u.a. im Ersatz von -[st] durch -[ʃt], in der Reduktion von [a] im Wort "das" zu [ə] und in Vokalrundungen bzw. im Ersatz von [a] durch [ɔ]. Als Merkmale der berliner Aussprache traten z.B. übermäßige Öffnungen der vokalisch aufgelösten R-Laute der Kombination -er und Ersatzrealisationen von [ts] durch [s] auf. Bei beiden Sprechern waren die üblichen Verschleifungen, Elisionen und Neubildungen typisch für den colloquial style (z.B. wurde der Satzanfang "Na, weil ich es..." als [navaeks] ausgesprochen).

(2) Aufnahme: Normalhalliger Raum (Filmstudio); Tonaufnahme mit für jeden Sprecher separaten Mikrofonen und Aufnahmegeräten (1 Pärchen Beyer Dynamic Mikrofone mit Richtwirkung; 2 NAGRA-Tonbandmaschinen mit 19 cm/sec; Abstand der Mikrofone vom Sprecher ca. 1m).

(3) Segmentation und Transkription: Die Tonaufnahmen wurden einzeln digitalisiert und jeweils gesondert computerunterstützt segmentiert sowie in broad transcription nach dem IPA-System transkribiert (unter Benutzung der PDP-11/50 Rechenanlage und der Software des Instituts für Phonetik und sprachliche Kommunikation der Universität München; Leitung: Prof. Dr. H.G. Tillmann; Beschreibung des Segmentations- und Transkriptionssystems vgl. WINKLER 1981, 28 ff.).

2.1 AUDITIVE FESTLEGUNG VON LAUTEXTENSIONEN (LÄNGE VON SEGMENTEN)

TIFFANY und CARRELL (1977[2], 45 ff.) bezeichnen die Transkriptionsmethode als eine perzeptuelle Analyse der "sound structure of language and articulatory possibilities", wobei sie vorsichtig einschränken, daß "...phonetic transcription is a record of the listener's impression of what he or she hears." (60). Das entspricht einer weitverbreiteten methodologischen Auffassung über die Transkriptionsmethode, in deren Rahmen bestimmte "Mängel" der Methode im Wesentlichen auf subjektive und individuelle Perzeptionen, Abweichungen von der interindividuellen Norm usw. zurückgeführt werden. Demgegenüber sieht die Phänomenologie in den immer wieder auftretenden Diskrepanzen zwischen den Notaten verschiedener Transkribenten oder zwischen Schallstruktur und Transkriptionen keine subjektiven Einzelfälle, sondern generelle Strukturen der menschlichen Perzeptions- und Apperzeptionsvorgänge. Verkürzt ausgedrückt beschreibt ein Transkript nicht die Schallstruktur von Lauten oder die Artikulationsmöglichkeiten der Sprechwerkzeuge, sondern die apperzeptiven Möglichkeiten zum Erkennen von Artikulationsbewegungen oder Sprachzeichen bzw. deren Varianten. In der Konsequenz der phänomenologischen Position liegt der Verzicht auf den Anspruch einer sogenannten "Abbildungstreue" von Notaten und weiterhin auch auf die Möglichkeit, mit ohrenphonetischen Methoden den Sprachschall oder die Sprechphysiologie so untersuchen zu können, wie es die signal-

und artikulationsphonetischen Methoden tun. Die phonetische Transkription kann kein Ersatz für Schall- und Lautanalysen sein, sondern sie dokumentiert den phänomenal-apperzeptiven Bereich gesprochener Sprache.

Analysen phonetischer Transkripte sind, immer der phänomenologischen Position zufolge, Untersuchungen zur Struktur der intersubjektiven Konstitution von Sprachzeichen aus Prototypen von Schallkonfigurationen (Phonemsystem), zur Technik des analytischen und funktionellen Hörens, zum apperzeptiven Aufbau von phonetischen Idealtypisierungen (Lauten), zur natürlichen Fähigkeit des Menschen, aus Schallprodukten auf die sprechphysiologischen Bewegungen schließen zu können, und zur fachspezifischen Wissensstruktur, den spezifischen Wissensvorräten und Elementen des Vorwissens. Die Phonologie kann zurecht Transkriptionen für die Untersuchung des Phonemsystems heranziehen, ebenso wie die Phonetik Notationen zur Beschreibung der diskriminierbaren Laute, Lautvarianten, Suprasegmentale und der erschließbaren Organbewegungen verwendet, nicht aber zur Analyse von Vorgängen, die während der normalen Kommunikation transphänomenal bleiben (z.B. die akustische Schallgestalt oder die tatsächlichen Sprechbewegungen).

Schon die Aufgliederung des Schallkontinuums in Laute (Segmentierung) ist eine rein phänomenal begründete Prozedur, bei der diskrete Muster von Lauttypen und Untertypen auf den Schallstrom "aufgeprägt" werden. Die akustische Struktur des Schalls ist als ein "Signal" aufzufassen, das durch weitreichende Apperzeptionsprozesse zur "Nachricht" wird (so beschreibt es das Signal-Nachricht-Modell der modernen Phonetik). Bei der computerunterstützten Segmentation, bei der der Schall gleichzeitig gesehen und gehört werden kann, werden die Unterschiede zwischen Signalstruktur und auditivem Korrelat unmittelbar evident.

Zum Beispiel kann sich während der Segmentationsarbeit der Transkribent entweder mehr nach dem visualisierten Schall (Oszillogramm) richten oder mehr nach dem auditiven Eindruck. Die Grenzen der Segmente würden jeweils anders festgelegt werden.

Als Beispiel sind in Abb. 1 das segmentierte Oszillogramm und fünf weitere Parameter des Anfangs von TAKE D, Sprecherin, angeführt. Die Diskrepanz von Schallverlauf und Segmenteintei-

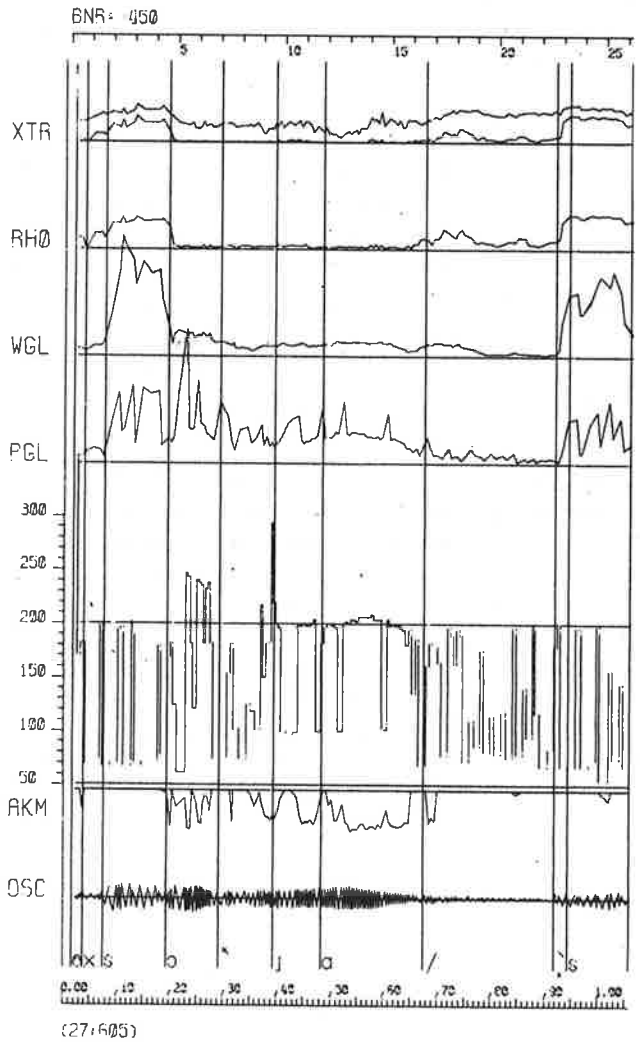


Abb.1: Segmentiertes Oszillogramm und Parameter-Darstellung des Anfangs von TAKE D, Sprecherin

lung aufgrund auditiver Kriterien ist deutlich bei der Kombination von [s] und [o] zu sehen. Die akustische Einteilung würde den Laut [o] später anfangen lassen; etwa mit dem Einsetzen des Stimmtones (Kurve über "AKM" in treppenförmiger Abstufung bei ca. 240 Hz) und dem gleichzeitigen Pegelanstieg (Kurve PGL).

Die auditive Bestimmung von Lautextensionen ist zwar nicht völlig unabhängig vom Schallverlauf, sie folgt aber zugleich einer eigenen, phänomenalen Regelhaftigkeit. Sie hat so gut wie keinen Bezug zur Umschrift der phonematischen Länge oder Kürze eines Lautes: Als "lang" transkribierte Vokale können so kurz sein wie "kurze" Laute und umgekehrt. (In der Phonetik wird diese Erscheinung damit erklärt, daß die Vokalqualität in der deutschen Sprache die -quantität dominiert: Offene Vokale werden zugleich "kurz" gehört, geschlossene "lang".) Die Transkription der (phonematischen) Lautdauer ist tatsächlich keine perzeptuelle Analyse der Lautextension, sondern die Kennzeichnung von bedeutungsunterscheidender Dauer eines Sprachzeichens. Die Konstitution von bedeutungsdifferenzierenden Elementen ist nur mittelbar von der akustischen Substanz abhängig.

Dazu zwei Beispiele aus dem untersuchten Material: Die Zeitstrecke zwischen Segmentanfang und -ende (Extension des Lautes) wurde bei einer Samplingfrequenz von 20KHz mit der Genauigkeit von 1/20 000 sec. digital bestimmt und in sec. umgerechnet. Die auditiv festgesetzten Zeitspannen für alle Realisationen ein- und desselben Phonems wurden maschinell sortiert und die statistischen Parameter $\bar{x}$, s und s^2 berechnet.

Tabelle 1: Auditive Längen der Vokale [ø], [œ], [o] und [ɔ]
TAKE D, Sprecherin, ca. 1o Min. Gespräch

	[ø]	[œ]	[o]	[ɔ]
sec	.22645	.23o4	.2177	.2325
	.2o38	.23o4	.2o845	.2291
	.16875	.14o8	.2o845	.2176
	.1612	.14o8	.2o665	.21o5
	.1o48	.12295	.2o5o5	.2o38
	.1o36	.12125	.2o44	.192
	.1oo2	.1o24	.1792	.1916
	.o512	.o9575	.1792	.18835
		.o768	.1696	.1786
		.o767	.1664	.16875
		.o7615	...	...
		.o7175	...	...
		.o64	.o384	.o384
		.o595	.o384	.o3655
		.o5785	.o377	.o362
		.o512	.o336	.o356
		.o434	.o3335	.o3385
		.o4o3	.o26	.o3o3
			.oo64	.o256
			.oo64	.o256
			.oo64	.o256
			.oo64	.o256
n	8	19	86	88
$\bar{x}$	1.1199	1.872	8.1288	8.3971
s^2	.13999	.o9891	.o9452	.o9542
s	.oo355	.oo3o6	.oo272	.oo277
	.o5959	.o536o	.o5216	.o5266

Die Absolutzeitwerte sind bei [o] niedriger als bei [ɔ] (Extremzeitwerte), wobei im mittleren Bereich gleiche Zeitwerte auftreten (z.B. .0384 sec.). Beim geschlossenen und offenen OE-Laut sind die Extremwerte ähnlich; es treten wiederum identische Längen auf (.0512 sec.).

Die theoretisch zu erwartenden Minimallängen - das Minimum, das beim Segmentieren nicht unterschritten werden kann, weil der Laut dann in eine andere Lautklasse eingeordnet wird - kann annähernd durch lineare Regression (x = 0) geschätzt werden.

	Minimum	empirisch	r (Anpassungsgüte)
[ø]	.033	.0512	.97
[œ]	.015	.0403	.87
[o]	.0003	.0064	.98
[ɔ]	.0213	.0256	.95

Die durchschnittliche Länge ist bei [ɔ] etwas größer als bei [o]; [ø] ist im Durchschnitt länger als [œ]. Zum Vergleich der Standardabweichungen wird zunächst der Variabilitätskoeffizient berechnet: V = 42.57[ø], =54.19[œ], = 55.28[o], = 55.19[ɔ]. Mit Ausnahme von [ø] liegen die Variationsbreiten in ähnlichen Größenordnungen.

Die Beispiele zeigten, in welchen Zeitdimensionen Segmente festgelegt werden. In der Tabelle 2 sind die Maximal- und Minimalextensionen aller Segmente des Materials aufgelistet. Die zwischen den Extremwerten liegenden Quantelungen lassen sich theoretisch berechnen: Die auditiven Zeitabschnitte steigen vom kürzesten bis längsten Segment eines Lauttyps logarithmisch (ln) oder wie 3. Wurzel aus einem Polynom dritten Grades an. Beim Segmentieren wird demnach ein "längerer" Laut nicht doppelt, dreifach usw. länger gehört als ein "kürzerer", sondern entsprechend der Funktion $y = ae^{bx}$, die bereits für biologische und physiologische Prozesse bekannt ist.

In Tabelle 2 sind selbstverständlich die empirischen Extremwerte angeführt, die theoretische Funktion ist aus allen aufgetretenen Extensionswerten berechnet (an einem Beispiel werden in Tabelle 4 empirische und theoretische Werte vollständig aufgeführt).

Tabelle 2: Mittel- und Extremwerte sowie Anpassungsgrad (r) aller Längenangaben (TAKE D, beide Sprecher)

1. Zahl: Sprecherin, 2. Zahl: Sprecher; n' = Zahl der Zeit-/Extensionsklassen, $\bar{x}'$ basiert auf n'

Symbol	n'	y_{min}	$\bar{x}'$	y_{max}	theoretische Funktion linear	ln	$\sqrt[3]{}$	$\sqrt{}$
i	49	.0298	.1079	.2945	.8986	.9725	.9621	.9512
	40	.0327	.1171	.3073	.9382	.9882	.9824	.9749
ɪ	79	.007	.0826	.307	.9362	.9786	.9884	.9831
	56	.0143	.0808	.3578	.8584	.9819	.9642	.9457
e	79	.02185	.1222	.3055	.9553	.9653	.9785	.9782
	65	.032	.132	.5625	.8012	.9593	.9273	.9026
ɛ	58	.0256	.1095	.35145	.9036	.9763	.9701	.9596
	37	.0346	.1309	.4993	.8399	.9746	.9496	.9288
a	79	.0217	.1070	.4352	.8385	.9764	.9515	.9298
	67	.02455	.1091	.2688	.9530	.9925	.9893	.9834
ɑ	65	.02545	.1467	.4352	.9267	.9725	.9777	.9719
	40	.0525	.1489	.4352	.9155	.9724	.9702	.9622
ä	79	.01445	.1156	.362	.9124	.9389	.9580	.9567
	79	.0064	.0971	.3072	.9247	.9265	.9611	.9627
ø	8	.0512	.1399	.22645	.9772	.9507	.9657	.9708
	7	.1024	.1635	.3072	.8573	.9270	.9058	.8943
œ	16	.0403	.08945	.2304	.8728	.9678	.9451	.9300
	3	.0768	.1707	.35605	.9966	.9733	.9831	.9873
o	63	.0064	.09755	.2177	.9679	.9343	.9749	.9810
	43	.022	.1446	.4048	.9356	.9771	.9820	.9766
ɔ	70	.0256	.09949	.2325	.9523	.9904	.9861	.9447
	36	.0346	.1100	.36115	.8461	.9650	.9422	.9237
ʏ	7	.02535	.09703	.2055	.9442	.9565	.9682	.9676
	10	.0609	.1810	.5249	.8567	.9733	.9454	.9266
Y	10	.0298	.0963	.2325	.9430	.9708	.9663	.9624
	4	.0435	.0496	.05625	.9977	.9992	.9988	.9986
u	26	.0256	.0981	.2816	.8945	.9708	.9617	.9504
	16	.0128	.0920	.21825	.9321	.9312	.9550	.9560

(Tabelle 2 Fortsetzung 1)

Symbol	n'	y_{min}	$\bar{x}'$	y_{max}	theoretische Funktion linear	ln	$\sqrt[3]{}$	$\sqrt{}$
o	46	.o128	.o959	.2561	.95o6	.9877	.9891	.9836
	13	.o165	.o586	.13915	.9186	.96oo	.9622	.9568
p	6	.oo56	.o471	.11335	.9589	.9674	.9858	.9866
	1o	.o1o7	.o656	.12355	.9841	.9569	.976o	.9815
t	79	.oo56	.o674	.3327	.79o2	.9649	.9457	.92o6
	45	.o145	.o726	.23o4	.91oo	.9844	.9729	.9616
k	42	.o128	.o6825	.23125	.8322	.9393	.93o8	.9153
	11	.o261	.o6665	.192	.8544	.9651	.9397	.9222
ʔ	79	.o129	.141o	.8o1	.79o4	.9685	.9442	.9152
	54	.o122	.1229	.4252	.9431	.9871	.9949	.9894
b	46	.oo275	.o6873	.3o275	.8686	.9613	.972o	.9562
	33	.oo915	.o8493	.2815	.9212	.99o8	.9897	.9795
d	79	.oo2	.o7o27	.7371	.6717	.9667	.9314	.8846
	79	.oo32	.o6672	.3o73	.8361	.9671	.9495	.9271
g	64	.o128	.o5317	.21o55	.8622	.9852	.9591	.9398
	29	.o16	.o5485	.1663	.8685	.9689	.9421	.9256
v	79	.o128	.o61o9	.25545	.8972	.99o9	.9778	.9643
	4o	.oo435	.o7955	.5249	.6933	.9238	.9o3o	.864o
z	35	.o256	.o74o7	.1792	.93o5	.9697	.9677	.9624
	38	.o2o45	.o7842	.32o1	.7643	.9348	.8932	.8651
ʒ	-	-	-	-	-	-	-	-
	1	.o426	-	-	-	-	-	-
j	4o	.o1o35	.1133	.4352	.8867	.9934	.9765	.9595
	19	.o1615	.o929	.4352	.752o	.9623	.9188	.8839
f	67	.o256	.1298	.768	.69o9	.9234	.8753	.8376
	36	.o2585	.1373	.5376	.7788	.93o9	.9o16	.8767
s	79	.o512	.1638	.46o3	.8999	.9777	.963o	.95o9
	79	.o256	.12o1	.4352	.826o	.9466	.9266	.9o78
ʃ	5o	.o512	.12o3	.3548	.8114	.9318	.9o49	.886o
	16	.o512	.o9725	.1568	.98o1	.9846	.9863	.9859
ç	74	.o23o4	.1183	.4224	.8598	.9415	.94o9	.9299
	46	.o3245	.o9119	.1791	.98o6	.9845	.99oo	.99o1
χ	32	.o2275	.o968	.19845	.9864	.9656	.98o3	.9848
	23	.o257	.1o49	.4252	.7721	.9462	.9o7o	.8787

(Tabelle 2 Fortsetzung 2)

Symbol	n'	y_{min}	$\bar{x}'$	y_{max}	theoretische Funktion linear	ln	$\sqrt[3]{}$	$\sqrt{}$
h	68	.o128	.3265	2.1632	.8138	.9748	.9769	955o
	57	.o149	.21o2	.789	.9431	.9749	.99o9	.9883
l	79	.oo88	.o835	.2711	.9542	.9754	.9912	.9897
	48	.o257	.1129	.32o1	.943o	.9766	.9822	.9785
ʁ	-	-	-	-	-	-	-	-
	4	.o384	.o521	.o6415	.9778	.9661	.97o8	.9728
ɓ	44	.oo9o5	.o7185	.22535	.9o38	.9662	.9694	.96o3
	16	.o128	.o66	.2o98	.9o11	.9898	.9775	.9636
ɒ	44	.o5o6	.14356	.3o435	.9717	.9932	.9936	.99o8
	58	.o19	.14597	.5249	.9o63	.9653	.971o	.9637
x	2	.o1656	.11255	.o6455	1.	1.	1.	1.
	-	-	-	-	-	-	-	-
m	79	.o128	.1452	.8o1	.7887	.9677	.94o7	.9116
	65	.o256	.1899	2.2656	.5529	.9556	.8741	.8o42
n	79	.ooo5	.1277	.7714	.7643	.81o3	.9137	.9o43
	79	.o135	.1644	1.6869	.5818	.9338	.8732	.8136
ŋ	25	.o385	.1261	.2561	.9611	.9799	.9824	.9799
	13	.o384	.1335	.27995	.9449	.9711	.9729	.9693
/	79	.oo81	.8178	1.5345	.2458	.9284	.6717	.5oo3
	79	.o255	.3148	1.1637	.9168	.9979	.9878	.975o
Pause	78	.o23o5	3.3916	2o.1728	.6969	.9343	.9o68	.8616
	76	.1464	4.6775	3o.9514	.8324	.9787	.98oo	.9583

Die Extensionsanalyse hat einen methodologischen Zweck: Dieser Längenanstieg könnte zwanglos sprechphysiologisch oder akustisch erklärt werden. Der Mehraufwand an Bewegungsaktivität beim Aussprechen eines längeren Lautes vergrößert sich gemäß der biologischen Zunahmefunktion; die zugehörigen akustischen Manifestationen steigen demnach in logarithmischen Zeitabschnitten.

Die akustischen Regelmäßigkeiten wie Vollständigkeit der Schwingungen etc. sind bei der Segmentation so gut wie nicht berücksichtigt worden. Das Segmentierungsprogramm sieht zwar eine Routine zum automatischen Aufsuchen der Nullstellen vor, die der Transkribent zum endgültigen Festlegen der Grenzen benutzen kann; diese wurde für das vorliegende Material sehr selten verwendet.

Weil sich die Quantelungen durch die auditive Festsetzung ergeben haben, ist es einleuchtender, phänomenale Zeitstrukturen und kategoriale Wahrnehmungsprozesse in erster Linie als Erklärung heranzuziehen und erst in zweiter Linie die sprechphysiologischen Abläufe. Von Lauttypen oder Sprechern hängt die Funktion nicht ab; es gibt kaum systematische Unterschiede innerhalb der Lautgruppen bzw. zwischen ihnen (vgl. Tab. 2, rechter Teil).

Verifizierung: Model fitting durch lineare Regression mit semilogarithmischer, semiquadratischer usw. Anpassung. In der Tab. 4 sind als ein Beispiel alle empirischen und theoretischen Werte für [k] und "Standard"-Extensionen angegeben. Die Standardlängen sind aus allen Zeitangaben aller Symbole als häufig wiederkehrende, z.T. mehrfach bei einem Symbol vertretene Extensionen zusammengestellt worden.

Tabelle 3: Anpassung theoretischer Funktionen an die empirische Anstiegsfunktion der Segmentlängen (r)

Funktion	Beispiel [k]	Standardextensionen
$y=ax+b$	.8322	.9497
$y=ae^{xb}$	.9393	.98oo
$y=\sqrt[3]{ax^3+bx^2+cx+d}$	.93o8	.9927
$y=\sqrt{ax^2+bx+c}$	.9153	.98oo
$y=ax^2+bx+c$	.6198	.855o
$y=ax^3+bx^2+cx+d$	.4796	.7845
$y=1+ax+b$	-.8144	-.75o2
$y=1/ax^2+bx+c$	-.41oo	-.5332

Tabelle 4: Empirische und theoretische Werte der Längenangaben von [k] (TAKE D, Sprecherin) und Standardextensionen

[k] empirisch	$y=ax+b$	$y=ae^{bx}$	$y=\sqrt[3]{ax^3+bx^2+cx+d}$	Standardextensionen empirisch	$y=ax+b$	$y=ae^{bx}$	$y=\sqrt[3]{ax^3+bx^2+cx+}$
.o128	.o1o4	.o239	.o2o8	.o128	.o216	.o25o	.o184
.o1795	.o132	.o25o	.o232	.o256	.o432	.o294	.o243
.o2295	.o16o	.o261	.o246	.o384	.o648	.o345	.o314
.o256	.o189	.o273	.o26o	.o512	.o864	.o4o4	.o398
.o2785	.o217	.o285	.o275	.o548	.1o8o	.o475	.o495
.o3365	.o245	.o298	.o291	.o64o	.1279	.o557	.o6o7
.o345	.o273	.o311	.o3o8	.o768	.1513	.o654	.o735
.o353	.o3o2	.o325	.o324	.o896	.1729	.o768	.o879
.o384	.o33o	.o339	.o342	.1o24	.1945	.o9o1	.1o41
.o4o15	.o358	.o354	.o36o	.1152	.2161	.1o58	.1222
.o416	.o386	.o37o	.o379	.128o	.2377	.1242	.1423
.o418	.o414	.o387	.o398	.14o8	.2593	.1458	.1644
.o425	.o443	.o4o4	.o418	.1536	.28o9	.1712	.1887
.o4455	.o471	.o422	.o439	.2o45	.3o25	.2oo9	.2154
.o462	.o499	.o441	.o46o	.2312	.3241	.2358	.2444
.o548	.o527	.o46o	.o482	.2816	.3457	.2768	.2759
.o5855	.o556	.o481	.o498	.3328	.3674	.325o	.31oo
.o5985	.o584	.o5o2	.o529	.384o	.389o	.3815	.3469
.o635	.o612	.o525	.o553	.4o96	.41o6	.4478	.3865
.o64	.o64o	.o548	.o578	.4352	.4322	.5256	.429o
.o6625	.o668	.o573	.o6o3				
.o6795	.o697	.o598	.o63o	r =	.9497	.98oo	.9927
.o69	.o725	.o625	.o657				
.o7o65	.o753	.o653	.o685				
.o71o5	.o781	.o682	.o714				
.o7185	.o81o	.o712	.o743				
.o7255	.o838	.o744	.o774				
.o7275	.o866	.o777	.o8o5				
.o7725	.o894	.o812	.o837				
.o793	.o922	.o848	.o869				
.o794	.o951	.o886	.o9o3				
.o815	.o979	.o926	.o938				
.o8375	.1oo7	.o967	.o973				
.o846	.1o35	.1o1o	.1oo9				
.o85o5	.1o63	.1o55	.1o46				
.o856	.1o92	.11o2	.1o85				
.o896	.112o	.1151	.1123				
.1o24	.1148	.12o3	.1163				
.1o365	.1176	.1256	.12o4				
.1112	.12o5	.1313	.1246				
.2o35	.1233	.1371	.1289				
.23125	.1261	.1432	.1332				
r =	.8322	.9393	.93o8				

Die Abweichungen von der ln-Funktion sind teilweise durch niedrige n'-Werte entstanden (z.B. [ø] der Sprecherin, [œ p ʁ] des Sprechers), durch die zu wenig Samples zur Kurvenbestimmung verarbeitet werden können. Teilweise sind die Korrelationskoeffizienten so gering voneinander unterschieden, daß der Unterschied praktisch vernachlässigt werden kann. Das trifft natürlich auch im umgekehrten Fall zu, wenn die ln-Funktion nur geringfügig höher korreliert als andere Kurvenformen. Insgesamt gesehen sind wohl beide Funktionen (ln und 3. Wurzel) gut zur Beschreibung des Längenanstiegs geeignet. Durch die Funktion $y = \sqrt[3]{ax^3+bx^2+cx+d}$ sind im vorliegenden Material die Laute [ɔ h l ŋ g] für beide Sprecher, [ɪ e a ɑ y b ʁ n] der Sprecherin und [o ʔ ʃ] sowie Pausen des Sprechers adäquater bezeichnet. Worin die Ursachen dafür liegen, daß diese oder jene Funktion besser angepaßt ist, d.h., welche phonetischen oder perzeptuellen Gründe zu einem der beiden Kurvenverläufen führen, läßt sich im Material nicht ermitteln.

Die Pausen- und Staupausenlängen beziehen sich sehr viel stärker auf die Zeitverläufe des Gesprächs; sie dokumentieren weniger die apperzeptive Zeitstruktur als vielmehr die Gesprächsdynamik. Das interaktionale Arrangement ist - so können die Zeitrelationen interpretiert werden - demzufolge nicht linear, sondern biologisch-rhythmisch eingeteilt. Die Zeitangaben für Staungen und Pausen geben die phänomenale Zeitstruktur der Sprecher an, die der Turnorganisation und zeitlichen Gestaltung des Gesprächs zugrundeliegt.

Praktischer Nutzen aus der Extensionsanalyse ist beispielsweise für eine hörgerechte Sprachsynthese oder für das Verstehen von Notaten zu ziehen. Wie die phonetische Notation verwenden auch andere (z.B. konversationsanalytische) Transkriptionssysteme Längenangaben (" ˙ : ::" usw.), die sich der Leser des Notats nicht linear, sondern logarithmisch steigend vorzustellen hat. Die Transkription umschreibt nicht "getaktete", sondern rhythmisch aufeinanderfolgende Segmente.

2.2 LAUTHÄUFIGKEIT UND -KOMBINATORIK UMGANGSSPRACHLICH BEEINFLUSSTER AUSSPRACHE

Die Auswertung von Notaten nach phonologischen Kriterien ist methodologisch gerechtfertigt, weil sie nicht im Widerspruch zum Gehalt von Transkriptionen steht. Für phonematische Analysen müssen die Notate nicht unbedingt lautsprachlich oder artikulatorisch orientiert sein. Es gibt dennoch einen Grund, trotzdem auf phonetische Transkripte zurückzugreifen. Mit RICHTER soll zwischen präund postrealisatorischen Notaten unterschieden werden; das sind zwei verschiedene Funktionen, mit denen unterschiedliche Verwendungen verbunden sind. ORTMANN (1975) hatte für eine statistische Beschreibung der deutschen Laute und Lautkombinationen das Wörterbuch von KAEDING phonetisch transkribiert. Die Transkripte waren prärealisatorisch und die Resultate sind präskriptiv, nicht deskriptiv aufzufassen. Denn durch die Umwandlung von Schrifttexten in phonetische Symbole ist nicht die lautsprachliche, sondern die schriftkonstituierte Frequenz und Kombinatorik des Deutschen erfaßt, wenngleich die graphematische Komponente weitgehend getilgt ist. Der Status des prärealisatorischen Notats bleibt auch dann erhalten, wenn Schrifttexte so phonetisch transformiert werden, wie die allgemeine Aussprache einer Kombination wäre (z.B. Endsilbe -er als [ɐʊ]). Ebenso kann eine Analyse der (nicht empirisch entwickelten) Ausspracheempfehlungen des Aussprache-Dudens nicht die in praxi gesprochenen Lautverbindungen beschreiben. Für eine Deskription sind postrealisatorische Notate zu verwenden, die allerdings mit jeglichen phänomenalen und apperzeptiven Vorzügen und Nachteilen belastet sind. Die Analyse befindet sich gewissermaßen auf einem Grenzgebiet, auf dem Kompromisse gemacht werden müssen. Schwer zu entkräften wird ein Argument sein, das aus der wissenssoziologischen Theorie abgeleitet ist: Mit der Konstitution von Transkriptionssystemen ist das gemeinschaftliche Wissen über Lauthäufigkeit und -kombinatorik bereits objektiviert. Jeder Sprachbenutzer weiß - der Fachmann noch auf eine besondere Weise - , daß eine dreifache Konsonantenverbindung nicht "typisch deutsch" ist. Eine statistische Analyse beschreibe nur noch - hyperbolisch formuliert - die Regeln des Notationssystems und ob der Transkribent die Regeln einhält. Der Transkribent weiß vorher, daß sich

das Rundungszeichen [◦] nur mit Vokalen kombinieren kann; er verwendet auch keine anderen Kombinationen und die statistische Analyse bestätigt, daß es keine "gerundeten Konsonanten" gibt. Die Notwendigkeit und der Einfluß des Vorwissens werden in der Phonetik nicht nur akzeptiert, sondern gewöhnlich auch gefordert: Der Transkribent muß die Sprache beherrschen, d.h. Lauthäufigkeit und -kombinatorik per erworbener Kompetenz kennen, und gelernt haben, was mit einem Symbol gemeinschaftlich konstituiert und objektiviert worden ist, wie es zu "klingen" hat usw., bevor er phonetisch transkribieren kann. Eine Frequenz- und Kombinationsanalyse ist in diesem Sinne keine "Vorstufe für experimentelle Untersuchungen des Sprechbewegungsablaufes" (LINDNER 1975, 77), sondern eine Nachbereitung des interindividuellen Wissens über statistische Regelhaftigkeiten von Artikulationsbewegungen, die der gesprochenen Sprache zugrundeliegen.

Der Gewinn bei der Analyse postrealisatorischer Notate liegt vor allem in jenen Abweichungen, die sich beim "Auflegen" der intersubjektiv verbindlichen Wissensfolie über idealtypische Lautverbindungen auf konkrete, empirische und teilweise vortypische Varianten ergeben. Die statistische Analyse regt zur Überprüfung an, ob der allgemeine Wissensvorrat auch spezielle Unterarten und Besonderheiten umfaßt, ob er erweitert werden muß, wie hoch das Idealisierungsniveau ist und welches Resultat entsteht, wenn eine objektive Wissensstruktur mit empirischen Befunden in Verbindung gebracht wird (beispielsweise durch Transkribieren auf das empirische Sprachmaterial "aufgeprägt" wird).

Am günstigsten ist demnach postrealisatorisches Transkriptmaterial, das mit einem gemeinschaftlich akzeptierten und prototypischen Notationssystem angefertigt worden ist und bei dem der Transkribent zwar korrekt, aber "unorthodox" verfahren ist (also auch "ungewöhnliche" Laute und Kombinationen zugelassen hat). Das phonematische Vorwissen kann zwar nicht ausgeschaltet werden (soll es auch nicht), es kann aber in gewisser Weise relativiert werden, indem phonetische Transkripte für phonologische Analysen verwendet werden. Die phonetische Notation nimmt weniger Rücksicht auf systematische Kobinationen. Phonetisch werden z.B. auch Folgen mehrerer Glottiseinsätze, mehrerer Konsonanten oder VVV-Ketten (statt der standardgemäßen VCV-/CVC-Verbindungen) notiert, wie sie in realen Sprachsequenzen auftreten. Die phonetische Intention des Transkribierens (neben phonematischen auch lautsprachliche und sprechphysiologische Einheiten zu dokumentieren) wird zum methodischen Vorteil für die phonologische Analyse. Der Transkribent muß zu diesem Zweck nicht unabdingbar native speaker sein; er muß nur das Notationssystem anwenden können (in wissenssoziologischer Formulierung: er muß Teilhaber des gesellschaftlichen Wissens sein). Das statistische Resultat - das zeigt den Kompromiß - ist sowohl phonetisch als auch phonologisch und, im größeren methodologischen Zusammenhang, auch wissenstheoretisch interpretierbar; wie jede Statistik keine bestimmte inhaltliche Auslegung zwingend vorschreibt.

Die Aufnahmen des TAKE D wurden phonetisch mit dem IPA-System[1] computerunterstützt transkribiert. Der Transkribent beherrschte weder die berliner noch die schwäbisch-alemannische Umgangssprache. Alle Lautvarianten, die mit der speziellen Technik des auditiv-visuellen Segmentierens und Transkribierens zu identifizieren waren, wurden transkribiert, auch wenn sie ungewohnt anmuteten. Die Aufnahmen wurden in zwei Versionen transkribiert: auditiv oder akustisch (d.i. visuell anhand von Kurvenformen) orientiert. Beide Versionen werden zur Analyse herangezogen: Die Frequenzanalyse stützt sich auf die auditiv, die Distributionsanalyse auf die akustisch orientierte Transkription.

Grundlage sind alle verwendeten Transkriptionssymbole, Sonderzeichen mit rein phonetischer Bedeutung inbegriffen. Die statistische Analyse verwendet die Verfahren, die ALTMANN/LEHFELDT (1980) entwickelt haben und die eine umfassende Beschreibung und Klassifikation von Lautfrequenzen und -kombinationen ermöglichen.

Das Prozedere statistischer Auswertungen ist normalerweise sehr aufwendig. Die Häufigkeiten und Kombinationen müssen manuell ausgezählt und verrechnet werden; bei Verwendung einer Rechenanlage (das besagt üblicherweise der Begriff "computerunterstützt") sind alle Auszählungen einzulesen, wie z.B. bei dem Verfahren von WOTHKE (1980). Das vom Münchener Institut entwickelte Transkriptionsprogramm enthebt den Analytiker aller zeitraubenden Dateneingaben; während des Segmentierens wird automatisch ein Transkriptions-File angelegt, der problemlos ausgewertet werden kann. Solche Programme sind jedoch noch nicht sehr verbreitet, weil vor allem die kostspielige Hardware fehlt. Um zu demonstrieren, wie eine phonologische Statistik mit einfacher Hardware durchgeführt werden kann, wurden die nachfolgenden Analysen mit einem programmierbaren Taschenrechner (TI 59 mit PC 100C) durchgerechnet. Die zugehörige Software, die viele Parameter aus der "Quantitativen Phonologie" von ALTMANN/LEHFELDT (1980) berechnet, ist im Anhang aufgelistet.

Die Transkriptions-Files des TAKE D wurden zu diesem Zweck teilweise maschinell vorausgewertet (z.B. automatische Umsetzung des computerinternen phonetischen Codes in normale phonetische Symbole), teilweise manuell ausgezählt (z.B. nach dem Schema S. 245 in ALTMANN/LEHFELDT 1980).[2]

2.2.1 Frequenzanalyse

In Tabelle 5 sind die statistischen Parameter der Häufigkeiten der umgangssprachlich beeinflußten Aussprache aufgeführt. Die verwendete Stichprobe hatte "natürliche" Grenzen und konnte nicht beliebig erweitert werden. Schätzt man die benötigte Größe der Stichprobe aus der geringsten Häufigkeit (z.B. $[ʒ]_{Sprecher} = .00035$), dann müßte N = 1085252.128 betragen. Nach der Faustregel: 1000 x (2 x Inventar) wären 1000000 Messungen nötig.

Die Parameter bedeuten (vgl. ALTMANN/LEHFELDT 1980, 142-182):

		verwendete Formel
K	Symbolinventar	
N	Gesamthäufigkeiten	
$\hat{p}_i$	relative Häufigkeit	f_i/N
R	Repeatrate	$(\chi^2+N)/NK$
D	relativiertes R bzw. χ^2	$\chi^2/N(K-1)$
V	Variationsquotient	$\sqrt{\chi^2/N}$
H_o, H_1	Entropie	ldK; $\approx H_o - (\chi^2/(2N \ln 2))$
h	relative Entropie	H_1/H_o
Re	relative Redundanz	$1-h$
I_L	Index Funktionalmittelwerte	2^{-H_1}

Im Vergleich des Parameters $\hat{R}$ (geschätzter R-Wert) mit den Ergebnissen anderer Analysen der deutschen Sprache ergeben sich Differenzen, die durch die verschiedenen Beträge von K entstehen, worin sich der Unterschied von phonetischen und phonematischen Einheiten zeigt (die ersten drei Werte sind der Zusammenstellung von ALTMANN/LEHFELDT 1980, 154f. entnommen):

Material	K	$\hat{R}$	Autor
deutscher Standard	29	.072007	Ovčinnikov 1962
deutscher Standard	30	.073938	Fitialova 1965
deutscher Standard	33	.059241	Kučera,Monroe 1968
umgangssprachlich beeinflußt	50	.035951	
berlinisch beeinflußt	48	.025182	
schwäbisch-alemannisch beeinfl.	47	.037431	

Tabelle 5: Parameter und Häufigkeiten der umgangssprachlich beeinflußten Aussprache

Symbol	berliner und schwäbisch-alemannischer Einfluß		schwäbisch-alemannisch		berliner Einfluß	
	f_i	$\hat{p}_i$	f_i	$\hat{p}_i$	f_i	$\hat{p}_i$
ʒ	1	.0001298533	-	-	1	.000354
χ	2	.0002597065	2	.00041o	-	-
ɾ	2	.0002597065	-	-	2	.00071o
ʁ	4	.0005194131	-	-	4	.001416
˜	5	.0006492663	3	.000615	2	.00071o
∘	12	.0015582392	12	.002461	-	-
Y	14	.0018179457	1o	.002o51	4	.001416
ø	15	.0019477990	8	.001641	7	.002478
p	16	.002o776523	6	.001231	1o	.00354o
y	19	.0024672121	7	.001436	12	.004248
œ	22	.0028567718	19	.003897	3(-1)	.001062
ŋ	47	.0061o31o35	34	.006973	13	.0046o2
u	53	.0068822231	33	.006768	2o	.007o79
k	59	.0076613427	47	.009639	12	.004248
x	73	.0094792884	43	.008819	3o	.01o619
j	75	.0097389949	52	.01o664	23	.008142
ʃ	81	.01o5181145	61	.01251o	2o	.007o79
ɔ	86	.0111673809	69	.014151	17	.006018
ʁ	88	.011427o874	63	.01292o	25	.008849
b	1o5	.0136345929	55	.011279	5o	.017699
z	1o5	.0136345929	48	.009844	57	.02o177
h	1o9	.014154oo6	88	.018o48	21	.007434
g	115	.0149331256	78	.015997	37	.013o97
l	121	.0157122452	63	.01292o	58	.02o531
ɛ	122	.015842o984	77	.015792	45	.015929
.	127	.0164913648	75	.015381	52	.018407
ɑ	129	.016751o713	8o	.016407	49	.017345
ɒ	132	.0171406311	6o	.0123o5	72	.025487
ɔ	138	.0179197507	88	.018o48	5o	.017699
o	149	.0193481366	86	.017637	63	.0223o1
f	15o	.0194779899	1oo	.02o5o9	5o	.017699
ll	158	.02o5168160	82	.016817	76	.026903

Tabelle 5 (Fortsetzung)

Symbol	berliner und schwäbisch-alemannischer Einfluß		schwäbisch-alemannisch		berliner Einfluß	
	f_i	$\hat{p}_i$	f_i	$\hat{p}_i$	f_i	$\hat{p}_i$
h	16o	.o2o7765225	92	.o188679	68	.o24o71
v	178	.o231138812	122	.o25o2o5	56	.o19823
l	192	.o24931827o	1o5	.o21534o	87	.o3o796
ç	193	.o25o6168o3	128	.o26251o	65	.o23oo9
t	215	.o279184521	151	.o3o968o	64	.o22655
ɐ	222	.o28827425o	159	.o326o87	63	.o223o1
e	225	.o292169848	133	.o272765	92	.o32567
·	231	.o299961o44	154	.o315833	77	.o27257
:	234	.o3o3856642	1oo	.o2o5o86	134	.o47434
/	235	.o3o5155175	126	.o2584o9	1o9	.o38584
ɪ	251	.o325931697	164	.o336341	87	.o3o796
m	253	.o328528762	153	.o313782	1oo	.o35389
ə	367	.o476561486	258	.o529122	1o9	.o38584
d	411	.o533696922	253	.o518868	158	.o55929
s	483	.o672191274	324	.o664479	159	.o56283
a	488	.o633683937	314	.o64397o	174	.o61593
n	489	.o63498247o	315	.o646o21	174	.o61593
° dev.	54o	.o7o12o7635	376	.o771124	164	.o58o53
K	5o		47		48	
N	77o1		4876		2825	
Σf^2	21321o1		88994o		28o773	
χ^2	6142.o13894		37o2.174733		1945.656283	
R	.o359512113		.o374311641		.o35181831	
D	.o162767462		.o165o57546		.o146537848	
V	.893o62463		.8713579698		.8298963115	
H_o	5.64385619		5.554588852		5.5849625o1	
H_1	5.o68537855		5.oo6895135		5.o8815o347	
h	.898o628997		.9o13979735		.911o44675	
Re	.1o19371oo3		.o986o2o265		.o88955325	
I_L	.o298oo124		.o311o1oo19		.o293977521	

NB: Diphtonge und Affrikata sind nach ihren Bestandteilen gezählt
dev. = devoiced

Merkmale der Umgangssprache sind häufig auftretende Entstimmlichungen (54O; aus 376 und 164) und Zusammenziehungen von Silben, wobei der Restlaut (zumeist Nasal) silbisch gesprochen wird (mittlere Häufigkeit, 75). Schwäbisch-alemannisch sind die Vokalrundungen (12), die nicht im berliner Einfluß auftreten. Die berlinisch beeinflußte Aussprache zeigt sich u.a. in der höheren Zahl von vokalischen Ersatzbildungen des R-Lautes (72 gegenüber 6O).

2.2.2 Distributionsanalyse

Mit den statistischen Maßen der Distribution werden die Kombinationen von Transkriptionssymbolen ohne Rücksicht auf die Frequenz beschrieben. Es wird die Menge von Symbolen ausgewertet, die als Vorgänger oder Nachfolger eines Symbols i auftritt. Es hängt vom methodologischen Standpunkt ab, ob diese Kombinatorik als phonetisch-artikulatorische oder phonematische oder aber als rein transkriptionelle Relation interpretiert wird. Die statistischen Begriffe wie Attraktivität oder Aggressivität besagen lediglich, daß sich bestimmte Symbole mit größeren oder kleineren Mengen anderer Symbole kombinieren. In einem begrenzten Material lassen sich nicht alle statistisch möglichen Kombinationen finden, auch nicht alle Verbindungen, die von einem Transkriptionssystem vorgesehen sind. Die Maßzahlen geben in jeglicher Hinsicht nur Ausschnitte an; selektiv aufgetretene Beziehungen innerhalb von Transkriptionsregeln, innerhalb des Phonemsystems oder lautsprachlichen Realisierung. Die Schätzungen und Klassifikationen aufgrund des vorliegenden Materials können nur Annäherungen sein. Ein Problem der statistischen Auswertungen, das auch von ALTMANN/LEHFELDT (198O, 217-22O) diskutiert wird, entfällt durch die Vorentscheidung für ein bestimmtes Transkriptionssystem: die Wahl von Rahmeneinheiten wie Phonem, phonotaktische Einheit, Silbe etc.; mit den phonetischen Symboleinheiten ist die Ebene der Segmentierung vorgegeben. Auch aus diesem Grund ist es angebracht, die statistischen Resultate zunächst nur als Ausdruck vom Symbol-Relationen aufzufassen.

In den Tabellen 6.1 - 6.3 sind die Distributionsmaße für die Symbolfolgen angeführt, die bei der Transkription der Umgangssprache verwendet worden sind. Von den Parametern und Algorithmen, die

ALTMANN/LEHFELDT (1980, 228ff.) vorschlagen, sind diejenigen ausgewählt worden, die für Einzelfälle am ehesten geeignet sind (Totalmaße für die Beschreibung einer Sprache insgesamt sind nicht verwendet worden) bzw. die den Aufwand an Datenarrangement minimal halten. Die verwendete Software ist so eingerichtet, daß aus drei Werten und K alle anderen Parameter berechnet werden können. Bei alternativen Rechenwegen bzw. Algorithmen wurden die für das Rechenprogramm ökonomischeren benutzt (siehe folgende Aufstellung):

Bezeichnung		Algorithmus
Ai	Menge der Vorgänger	
iB	Menge der Nachfolger	
VE	= Ti; Vereinigungsmenge von Ai und iB	
DU	A∩B	
SD	A⊕B	
Maße der Assoziativität		
At	Attraktivität	At(i)=[Ai]/K
Ag	Aggressivität	Ag(i)=[iB]/K
As	Assoziativität	As(i)=[Ai∪iB]/K
As*	Maß der Paarbildung von i	As*=([Ai]+[iB]-G(i))/(2K-1)
At_i	interne Attraktivität	At_i=[Ai]/VE
Ag_i	interne Aggressivität	Ag_i=[iB]/VE
Maße der Symmetrie		
S	Symmetrie	S=[Ai∩iB]/K
S_i	interne Symmetrie	S_i=[Ai∩iB]/VE
S_i*	potentielle Symmetrie	S_i*=2[Ai∩iB]/[Ai]+[iB]
S'	Antisymmetrie	S'=[Ai⊕iB]/K
S_i'	interne Antisymmetrie	S_i'=[Ai⊕iB]/VE
S_i'*	potentielle Antisymmetrie	S_i'*=[Ai⊕iB]/[Ai]+[iB]
Reflexivität:		
G	Kombination i mit i	= 1, wenn ii auftritt, sonst = 0

Die Symbolfolgebeziehungen wurden nach dem Kriterium für K = 50 klassifiziert; Werte zwischen 18 und 32 gelten als semiassoziativ (siehe Tabellen 7.1 bis 7.3).

Tabelle 6.1: Distributionsmaße für Textsyntagmen

Sy	Ai	iB	VE	DU	SD	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S_i'	S_i'*	G
ζ	1	1	2	0	2	.02	.02	.04	.02020	.05	.05	0	0	0	.04	1	1	0
χ	3	2	5	0	5	.06	.04	.01	.05051	.06	.04	0	0	0	.01	1	1	0
‹	2	2	2	2	0	.04	.04	.04	.04040	1	1	.04	1	1	0	0	0	0
ε	6	5	10	1	9	.12	.01	.02	.11111	.06	.05	.02	.01	.18181	.18	.9	.81818	0
~	4	4	8	0	8	.08	.08	.16	.08081	.05	.05	0	0	0	.16	1	1	0
∘	4	10	12	2	10	.08	.2	.24	.14141	.33333	.83333	.04	.16667	.28571	.2	.83333	.71428	0
γ	11	10	16	5	11	.22	.2	.32	.21212	.6875	.625	.1	.3125	.47619	.22	.6875	.52380	0
φ	8	8	15	1	14	.16	.16	.3	.16162	.53333	.53333	.02	.06667	.125	.28	.93333	.875	0
ρ	11	10	18	3	15	.22	.2	.36	.21212	.61111	.55556	.06	.16667	.28571	.3	.83333	.71428	0
y	15	17	25	7	18	.3	.34	.5	.31313	.6	.68	.14	.28	.4375	.36	.72	.5625	1
ιη	13	16	24	5	19	.26	.32	.48	.29293	.54166	.66666	.1	.20833	.34482	.38	.79166	.65517	0
η	19	16	27	8	19	.38	.32	.54	.34343	.70370	.59259	.16	.29629	.45714	.38	.70370	.54285	1
υ	26	13	29	10	19	.52	.26	.58	.39394	.89655	.44827	.2	.34482	.51282	.38	.65517	.48717	0
κ	22	23	35	10	25	.44	.46	.7	.45455	.62857	.65714	.2	.28571	.44444	.5	.71428	.55555	0
χ	32	16	34	14	20	.64	.32	.68	.48484	.94117	.47058	.28	.41176	.58333	.4	.58823	.41666	0
ϳ	22	25	30	17	13	.44	.5	.6	.46464	.73333	.83333	.34	.56666	.72340	.26	.43333	.27659	1
ʃ	30	20	37	13	24	.6	.4	.74	.50505	.81081	.54054	.26	.35135	.52	.48	.64864	.48	0
ο	14	24	29	9	20	.28	.48	.58	.38383	.48275	.82758	.18	.31034	.47368	.4	.68965	.52631	0
в	22	27	35	14	21	.44	.54	.7	.49494	.62857	.77142	.28	.4	.57142	.42	.6	.42857	0
b	30	27	38	19	19	.6	.54	.76	.57575	.78947	.71052	.38	.5	.66666	.38	.5	.33333	0
z	21	23	33	11	22	.42	.46	.66	.44444	.63636	.69696	.22	.33333	.5	.44	.66666	.5	0
h	29	17	37	9	28	.58	.34	.74	.46464	.78378	.45945	.18	.24324	.39130	.56	.75675	.60869	0
g	28	30	39	19	20	.56	.6	.78	.58585	.71794	.76923	.38	.48717	.65517	.4	.51282	.34482	0
i	35	21	38	18	20	.7	.42	.76	.55555	.92105	.55263	.36	.47368	.64285	.4	.52631	.35714	1

Tabelle 6.1 (Fortsetzung)

Sy	Ai	iB	VE	DU	SD	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S_i'	S_i'*	G
ɛ	3o	32	37	25	12	.6	.64	.74	.616162	.81o811	.864865	.5	.675676	.8o6452	.24	.324324	.193548	1
.	7	32	33	6	27	.14	.64	.66	.393939	.212121	.969697	.12	.181818	.3o7692	.54	.818182	.6923o8	o
ɑ	36	27	42	21	21	.72	.54	.84	.636364	.857143	.642857	.42	.5	.666667	.42	.5	.333333	o
ɒ	32	21	38	15	23	.64	.42	.76	.535353	.8421o5	.552632	.3	.394737	.566o38	.46	.6o5263	.433962	o
ɔ	27	27	38	16	22	.54	.54	.76	.535353	.71o526	.71o526	.32	.421o53	.592593	.44	.578947	.4o74o7	1
o	37	31	4o	28	12	.74	.62	.8	.676768	.925	.775	.56	.7	.823529	.24	.3	.176471	1
f	35	33	41	27	14	.7	.66	.82	.676768	.853659	.8o4878	.54	.658537	.794118	.28	.341463	.2o5882	1
ˈˈ	32	26	39	19	2o	.64	.52	.78	.585859	.82o513	.666667	.38	.487179	.655172	.4	.512821	.344828	o
h	35	33	42	26	16	.7	.66	.84	.676768	.833333	.785714	.52	.619o48	.7647o6	.32	.38o952	.235294	1
v	27	31	39	19	2o	.54	.62	.78	.575758	.6923o8	.794872	.38	.487179	.655172	.4	.512821	.344828	1
l	33	29	37	25	12	.66	.58	.74	.616162	.891892	.783784	.5	.675676	.8o6452	.24	.324324	.193548	1
ç	37	32	43	26	17	.74	.64	.86	.686869	.86o465	.744186	.52	.6o4651	.753623	.34	.395349	.246377	1
t	37	31	42	26	16	.74	.62	.84	.676768	.88o952	.738o95	.52	.619o48	.7647o6	.32	.38o952	.235294	1
ʔ	25	37	44	18	26	.5	.74	.88	.616162	.568182	.84o9o9	.36	.4o9o91	.58o645	.52	.59o9o9	.419355	1
e	36	32	42	26	16	.72	.64	.84	.676768	.857143	.7619o5	.52	.619o48	.7647o6	.32	.38o952	.235941	1
·	37	22	43	16	27	.74	.44	.86	.59596o	.86o465	.511628	.32	.372o93	.542373	.54	.6279o7	.457627	o
:	41	36	43	34	9	.82	.72	.86	.767677	.953488	.8372o9	.68	.79o698	.883117	.18	.2o93o2	.116883	1
/	4o	3o	43	27	16	.8	.6	.86	.69697o	.93o233	.697674	.54	.6279o7	.771429	.32	.372o93	.228571	1
ɪ	26	35	38	23	15	.52	.7	.76	.616162	.684211	.921o53	.46	.6o5263	.754o98	.3	.394737	.2459o2	o
m	38	37	43	32	11	.76	.74	.86	.747475	.883721	.86o465	.64	.744186	.853333	.22	.255814	.146667	1
ə	39	4o	47	32	15	.78	.8	.94	.787879	.829787	.851o64	.64	.68o851	.81o127	.3	.319149	.189873	1
d	39	4o	45	34	11	.78	.8	.9	.787879	.866667	.888889	.68	.755556	.86o759	.22	.244444	.139241	1
s	4o	34	43	31	12	.8	.68	.86	.737374	.93o233	.79o698	.62	.72o93o	.837838	.24	.279o69	.162163	1
a	37	37	42	32	1o	.74	.74	.84	.747475	.88o952	.88o952	.64	.7619o5	.864865	.2	.238o95	.135135	o
n	43	42	46	39	7	.86	.84	.92	.848485	.934783	.913o43	.78	.847826	.917647	.14	.152174	.o82353	1
ˌ	14	4o	4o	14	26	.28	.8	.8	.545455	.35	1	.28	.35	.518519	.72	.65	.481481	o

Sy	Ai	iB	VE	DU	SD	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S '	S_i'	S_i'*	G
ˈ	3	2	5	o	5	.o6383o	.o42553	.1o6383	.o53763	.6	.4	o	o	o	.1o6383	1	1	o
˜	4	3	7	o	7	.o851o6	.o63829	.148936	.o75269	.571429	.428571	o	o	o	.148936	1	1	o
p	8	7	14	1	13	.17o213	.148936	.297872	.16129o	.571429	.5	.o21277	.o71429	.133333	.276596	.928571	.866667	o
y	8	8	11	5	6	.17o213	.17o213	.234o43	.16129o	.727273	.727273	.1o6382	.454545	.625	.127659	.545455	.375	1
ø	7	6	12	1	11	.148936	.127659	.255319	.139785	.583333	.5	.o21277	.o83333	.153846	.234o43	.916667	.846154	o
Y	8	6	12	2	1o	.17o213	.127659	.255319	.15o537	.666667	.5	.o42553	.166667	.285714	.212766	.833333	.714286	o
°	4	1o	13	1	12	.o851o6	.212766	.276596	.15o538	.3o7692	.769231	.o21277	.o76923	.142857	.255319	.923o77	.857142	o
œ	13	15	23	5	18	.276586	.319149	.489362	.3o1o75	.565217	.652174	.1o6383	.217391	.357142	.382979	.7826o9	.642857	o
u	21	11	26	6	2o	.4468o9	.234o43	.553191	.344o86	.8o7692	.423o77	.127659	.23o769	.375	.425532	.769231	.625	o
ŋ	18	14	25	7	18	.382979	.297872	.531915	.344o86	.72	.56	.148936	.28	.4375	.382979	.72	.5625	o
×	26	12	3o	8	22	.553191	.255319	.638298	.4o86o2	.866667	.4	.17o213	.266667	.421o53	.468o85	.733333	.578947	o
k	24	19	35	8	27	.51o638	.4o4255	.744681	.462366	.685714	.542857	.17o213	.228571	.372o93	.574468	.771429	.6279o7	o
z	16	15	25	6	19	.34o426	.319149	.531915	.333333	.64	.6	.127659	.24	.387o97	.4o4255	.76	.6129o3	o
j	19	19	27	11	16	.4o4255	.4o4255	.574468	.4o86o2	.7o37o4	.7o37o4	.234o43	.4o74o7	.578947	.34o426	.592593	.421o53	o
b	24	22	31	15	16	.51o638	.468o85	.659574	.494624	.774194	.7o9677	.319149	.483871	.652174	.34o426	.516129	.347826	o
ɒ	25	17	36	6	3o	.531915	.3617o2	.765957	.451613	.694444	.472222	.127659	.166667	.285714	.638298	.833333	.714286	o
ʃ	25	16	32	9	23	.531915	.34o426	.68o851	.44o86o	.78125	.5	.191489	.28125	.439o24	.489362	.71875	.56o976	o
i	31	18	34	15	19	.659574	.382979	.7234o4	.526882	.911765	.529412	.319149	.441176	.612245	.4o4255	.558824	.387755	o
ʁ	2o	25	33	12	21	.425532	.531915	.7o2128	.483871	.6o6o61	.757576	.255319	.363637	.533333	.4468o9	.636364	.466667	o
ɑ	14	23	28	9	19	.297872	.489362	.595745	.397849	.5	.821485	.191489	.321429	.486487	.4o4255	.678571	.513512	o
ˌ	13	32	34	11	23	.276596	.68o851	.7234o4	.483871	.382353	.941176	.234o43	.323529	.488889	.489362	.676471	.511111	o
c	23	23	31	15	16	.489362	.489362	.659574	.494624	.741935	.741935	.319149	.483871	.652174	.34o426	.516129	.347826	o
g	28	24	34	18	16	.595745	.51o638	.7234o4	.559139	.823529	.7o5882	.382979	.529412	.6923o8	.34o426	.47o589	.3o7692	o
ɑ	31	22	38	15	23	.659574	.468o85	.8o8511	.569892	.815789	.578947	.319149	.394737	.566o38	.489362	.6o5263	.433962	o

Tabelle 6.2 (Fortsetzung)

Sy	Ai	iB	VE	DU	SD	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S_i'	S_i'*	G
ɪ	26	21	33	14	19	.553191	.4468o9	.7o2128	.5o5376	.787879	.636364	.297872	.424242	.595744	.4o4255	.575758	.4o4255	o
o	32	28	39	21	18	.68o851	.595744	.829787	.6344o9	.82o513	.717949	.4468o9	.538462	.7	.382979	.461538	.3	1
h	28	13	38	3	35	.595745	.276596	.8o8511	.44o86o	.736842	.3421o5	.o63829	.o78947	.146341	.744681	.921o53	.853659	o
ɔ	26	22	35	13	22	.553191	.468o85	.744681	.5o5376	.742857	.628571	.276595	.371429	.541667	.468o85	.628571	.458333	1
h	29	26	36	19	17	.617o21	.553191	.765957	.58o645	.8o5556	.722222	.4o4255	.527778	.69o9o9	.3617o2	.472222	.3o9o91	1
f	32	33	41	24	17	.68o851	.7o2128	.87234o	.688172	.78o488	.8o4878	.51o638	.585366	.738462	.3617o2	.414634	.261538	1
:	36	28	39	25	14	.765957	.595745	.829787	.688172	.923o77	.717949	.531915	.641o26	.78125	.297872	.358974	.21875	o
l	27	23	31	19	12	.574468	.489617	.659574	.537634	.87o968	.741935	.4o4255	.6129o3	.76	.255319	.387o97	.24	o
v	27	3o	39	18	21	.574468	.638298	.829787	.6o2151	.6923o8	.769231	.382979	.461538	.631579	.4468o9	.538462	.368421	1
/	3o	27	37	2o	17	.638298	.574468	.787234	.6o2151	.81o811	.72973o	.425532	.54o541	.7o1754	.3617o2	.45946o	.298246	1
ç	33	29	4o	22	18	.7o2128	.617o21	.851o64	.655914	.825	.725	.468o85	.55	.7o9677	.382979	.45	.29o32	1
e	3o	22	36	16	2o	.638298	.468o85	.765957	.548387	.833333	.611111	.34o426	.444444	.615385	.425532	.555556	.384615	1
t	36	3o	42	24	18	.765957	.638298	.893617	.698925	.857143	.714286	.51o638	.571429	.727273	.382979	.428571	.272727	1
m	36	37	45	28	17	.765957	.787234	.957447	.774194	.8	.822222	.595745	.622222	.767123	.3617o2	.377778	.232877	1
·	33	2o	4o	13	27	.7o2128	.425532	.851o64	.569892	.825	.5	.276596	.325	.49o566	.574468	.675	.5o9434	o
?	26	32	4o	18	22	.553191	.68o851	.851o64	.6129o3	.65	.8	.382979	.45	.62o689	.468o85	.55	.37931o	1
ɪ	25	27	33	19	14	.531915	,574468	.7o2128	.559139	.757576	.818182	.4o4255	.575758	.73o769	.297872	.424242	.2693o8	o
d	35	36	4o	31	9	.744681	.765957	.851o64	.752688	.875	.9	.659574	.775	.873239	.191489	.225	.126761	1
ə	36	37	44	29	15	.765957	.787234	.93617o	.774194	.818182	.84o9o9	.617o21	.659o91	.794521	.319149	.34o9o9	.2o5479	1
a	32	37	4o	29	11	.68o851	.787234	.851o64	.741935	.8	.925	.617o21	.725	.84o579	.234o43	.275	.15942o	o
n	4o	39	44	35	9	.851o64	.829787	.93617o	.83871o	.9o9o91	.886364	.744681	.795455	.886o76	.191489	.2o4545	.113924	1
s	39	33	44	28	16	.829787	.7o2128	.93617o	.763441	.886364	.75	.595745	.636364	.777778	.34o425	.363636	.222222	1
₀	11	37	39	9	3o	.234o43	.787234	.829787	.516129	.282o51	.948718	.191489	.23o769	.375	.638298	.769231	.625	o

- 26 -

Sy	Ai	iB	VE	DU	SD	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S_i'	S_i'*	G
ʒ	1	1	2	o	2	.o2o833	.o2o833	.o41667	.o21o53	.5	.5	o	o	o	.o41667	1	1	o
~	1	1	2	o	2	.o2o833	.o2o833	.o41667	.o21o53	.5	.5	o	o	o	.o41667	1	1	o
←	2	2	2	2	o	.o41667	.o41667	.o41667	.o421o5	1	1	.o41667	1	1	o	o	o	o
ɐ	3	3	6	o	6	.o625	.o625	.125	.o63158	.5	.5	o	o	o	.125	1	1	o
ʁ	5	5	9	1	8	.1o4167	.1o4167	.1875	.1o5263	.555556	.555556	.o2o833	.111111	.2	.166667	.888889	.8	o
Y	5	4	8	1	7	.1o4167	.o83333	.166667	.o94737	.625	.5	.o2o833	.125	.222222	.145833	.875	.777778	o
ø	5	6	1o	1	9	.1o4167	.125	.2o8333	.115789	.5	.6	.o2o833	.1	.181818	.1875	.9	.818182	o
p	6	4	9	1	8	.125	.o83333	.1875	.1o5263	.666667	.444444	.o2o833	.111111	.2	.166667	.888889	.8	o
k	13	1o	19	4	15	.27o833	.2o8333	.395833	.2421o5	.684211	.526316	.o83333	.21o526	.347826	.3125	.789474	.652174	o
y	12	9	19	2	17	.25	.1875	.395833	.221o53	.631579	.473684	.o41667	.1o5263	.19o476	.354167	.894737	.8o9524	o
ŋ	7	12	17	2	15	.145833	.25	.354167	.189474	.411765	.7o5882	.o41667	.117647	.21o526	.3125	.882353	.789474	1
o	9	7	12	4	8	.1875	.145833	.25	.168421	.75	.583333	.o83333	.333333	.5	.166667	.666667	.5	o
ʃ	11	18	24	5	19	.229167	.375	.5	.3o5263	.458333	.75	.1o4167	.2o8333	.344828	.395833	.791667	.655172	o
u	1o	13	2o	3	17	.2o8333	.27o833	.416667	.2421o5	.5	.65	.o625	.15	.26o869	.354167	.85	.73913o	o
h	14	14	24	4	2o	.291667	.291667	.5	.294737	.583333	.583333	.o83333	.166667	.285714	.416667	.833333	.714286	o
j	19	13	25	7	18	.395833	.27o833	.52o833	.326316	.76	.52	.145833	.28	.4375	.375	.72	.5625	1
ʁ	12	11	2o	3	17	.25	.229167	.416667	.2421o5	.6	.55	.o625	.15	.26o869	.354167	.85	.73913o	o
×	22	1o	25	7	18	.458333	.2o8333	.52o833	.336842	.88	.4	.145833	.28	.4375	.375	.72	.5625	o
g	22	16	27	11	16	.458333	.333333	.5625	.4	.814815	.592593	.229167	.4o74o7	.578947	.333333	.592593	.421o53	o
ɛ	24	23	31	16	15	.5	.479167	.645833	.484211	.774194	.741935	.333333	.516129	.68o851	.3125	.483871	.319149	1
ɑ	19	25	32	12	2o	.395833	.52o833	.666667	.463158	.59375	.78125	.25	.375	.545455	.416667	.625	.454545	o
b	21	21	3o	12	18	.4375	.4375	.625	.4421o5	.7	.7	.25	.4	.571429	.375	.6	.428571	o
f	22	27	32	17	15	.458333	.5625	.666667	.5o5263	.6875	.84375	.354167	.53125	.693878	.3125	.46875	.3o6122	1
ɔ	19	2o	29	1o	19	.395833	.416667	.6o4167	.41o526	.655172	.689655	.2o8333	.344828	.512821	.395833	.655172	.487179	o

- 27 -

Tabelle 6.3 (Fortsetzung)

Sy	Ai	iB	VE	DU	SD	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S_i'	S_i'*	G
,	23	4	26	1	25	.479167	.083333	.541667	.284211	.884615	.153846	.020833	.384615	.074074	.520833	.961538	.925926	o
v	21	13	25	9	16	.4375	.270833	.520833	.357895	.84	.52	.1875	.36	.529412	.333333	.64	.470588	o
z	21	17	26	12	14	.4375	.354167	.541667	.4	.807692	.653846	.25	.461538	.631579	.291667	.538462	.368421	o
i	13	29	32	10	22	.270833	.604167	.666667	.431579	.40625	.90625	.208333	.3125	.476190	.458333	.6875	.523809	1
ʔ	20	19	26	13	13	.416667	.395833	.541667	.4	.769231	.730769	.270833	.5	.666667	.270833	.5	.333333	1
o	18	29	36	11	25	.375	.604167	.75	.494737	.5	.805556	.229167	.305556	.468085	.520833	.694444	.531915	o
t	23	26	33	16	17	.479167	.541667	.6875	.515789	.696969	.787879	.333333	.484849	.653061	.354167	.515152	.346939	o
ç	19	27	30	16	14	.395833	.5625	.625	.473684	.633333	.9	.333333	.533333	.695652	.291667	.466667	.304348	1
h	25	24	32	17	15	.520833	.5	.666667	.505263	.78125	.75	.354167	.53125	.693878	.3125	.46875	.306122	1
p	18	30	37	11	26	.375	.625	.770833	.505263	.486487	.810811	.229167	.297297	.458333	.541667	.702703	.541667	o
ŋ	22	25	34	13	21	.45833	.520833	.708333	.494737	.647059	.735294	.270833	.382353	.553191	.4375	.617647	.446809	o
ˈ	13	28	31	10	21	.270833	.583333	.645833	.431579	.419355	.903226	.208333	.322581	.487805	.4375	.677419	.512195	o
ɪ	27	19	34	12	22	.5625	.395833	.708333	.484211	.794118	.558824	.25	.352941	.521739	.458333	.647059	.478261	o
l	24	26	33	17	16	.5	.541667	.6875	.515789	.727273	.787879	.354167	.515152	.68	.333333	.484848	.32	1
e	24	27	33	18	15	.5	.5625	.6875	.526316	.727273	.818182	.375	.545455	.705882	.3125	.454545	.294118	1
m	28	32	40	20	20	.583333	.666667	.833333	.621053	.7	.8	.416667	.5	.666667	.416667	.5	.333333	1
ə	28	36	39	25	14	.583333	.75	.8125	.663158	.717949	.923077	.520833	.641026	.78125	.291667	.358974	.21875	1
/	25	31	38	18	20	.520833	.645833	.791667	.578947	.657895	.815789	.375	.473684	.642857	.416667	.526316	.357143	1
:	24	33	38	19	19	.5	.6875	.791667	.589474	.631579	.868421	.395833	.5	.666667	.395833	.5	.333333	1
d	37	35	43	29	14	.770833	.729167	.895833	.747368	.860465	.813953	.604167	.674419	.805556	.291667	.325581	.194444	1
s	30	35	40	25	15	.625	.729167	.833333	.673684	.75	.875	.520833	.625	.769231	.3125	.375	.230769	1
ˌ	33	7	33	7	26	.6875	.145833	.6875	.421053	1	.212121	.145833	.212121	.35	.541667	.787879	.65	o
n	39	39	44	34	10	.8125	.8125	.916667	.810526	.886364	.886364	.708333	.772727	.871795	.208333	.227273	.128205	1
a	28	28	39	17	22	.583333	.583333	.8125	.589474	.717949	.717949	.354167	.435897	.607143	.458333	.564103	.392857	o

Vergleicht man die Klassifikation in der Tabelle 7.1. mit der Kategorisierung, die ALTMANN,LEHFELDT (1980,263) für Transkriptionen des amerikanischen Englisch (aE) auf der Basis einer Studie von ROBERTS aus dem Jahre 1965 vorgenommen haben, so ergibt sich:

(a) In die gleiche Klasse gehören

[m ə d s n] = aggressiv + attraktiv;

[j ʃ b g ɔ v] = semiaggressiv + semiattraktiv ("Mittelfeld").

(b) Um nur eine Klasse verschieden sind (z.B. ist [ʒ] im aE nichtaggressiv + semiattraktiv; im Deutschen nichtaggressiv + nichtattraktiv)

[z ʒ t a k l e u h i].

(c) Unterschiedliche statistische Eigenschaften haben

[f] (aE: semiaggressiv + semiattraktiv; Deutsch: aggressiv + attraktiv);

[p] (aE: aggressiv + semiattraktiv; Deutsch: nichtaggressiv + nichtattraktiv).

(Es wurden nur solche Symbole verglichen, die für deutsche und amerikanische Aussprache phonetisch adäquat sind).

Für den F- und P-Laut lohnt sich ein Vergleich der einzelnen Parameter, um die Unterschiede spezifizieren zu können:

	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S'_i	S'_i*
F_{aE}	.69	.50	.69	.60	1.	.73	.5	.73	.84	.19	.27	.16
F_D	.70	.66	.82	.68	.85	.80	.54	.66	.79	.28	.34	.21
P_{aE}	.53	.72	.75	.64	.71	.96	.50	.67	.80	.25	.33	.20
P_D	.22	.20	.36	.21	.61	.56	.06	.17	.29	.30	.83	.71

Als unterschiedlich sollen Maßzahlen gelten, deren Differenz $|>.10|$ ist. P-und F-Laute unterscheiden sich aus verschiedenen Gründen in beiden Transkriptionen. Die deutsche Symbolverwendung [f] ist assoziativer (As), weil mehr Nachfolgesymbole möglich sind (Ag); bezogen auf die Zahl der Symbole, mit denen F als Vorgänger und/oder Nachfolge Paare bildet (At_i), ist die deutsche Verwendung weniger attraktiv.

Der Gebrauch von [p] ist in beiden Transkriptionen ähnlich

Tabellen 7.1 bis 7.3: Kombinierte assoziative Klassifikation für Transkriptionssymbole bei Aussprache mit schwäbisch-alemannischem und berliner Einfluß (7.1), schwäbisch-alemannischem Einfluß (7.2) und berliner Einfluß (7.3)

7.1	nichtattraktiv	semiattraktiv	attraktiv
nichtaggressiv	ʒ ɣ ɬ ʁ ˜ ˈ ʏ ɸ p y œ	ŋ u x ʰ	
semiaggressiv	ɑ ˌ	k j ʃ ʁ b z g ɛ ɔ ɔ v	i ɑ o ˈˈ ˌ ç t ˑ /
aggressiv	ˌdev	ʔ ɪ	f h : m ə d s a n

7.2	nichtattraktiv	semiattraktiv	attraktiv
nichtaggressiv	χ ˜ p y ɸ ʏ ˈ œ ŋ z	u x ɔ ʃ i ʰ	
semiaggressiv	ɑ ˌ	k j b ʁ ɛ g ɑ ˈˈ o ɔ h ɪ v / e ʔ ɪ	: ç t ˑ
aggressiv	ˌdev	f a	m d ə n s

7.3	nichtattraktiv	semiattraktiv	attraktiv
nichtaggressiv	ʒ ˜ ɬ ɤ ʁ ʏ ɸ p k y ŋ ɑ ʃ u ʰ ʁ	j x g ˌ v z	ˌdev
semiaggressiv	ɪ o ˑ	ɛ ɑ b f ɔ ɔ ʔ t ç h ˈˈ ɪ l e m / a	
aggressiv		ə : s	d n

in der Anzahl der Symbole, mit denen [p] als Vorgänger und/oder als Nachfolger Paare bildet sowie in der Eigenschaft, entweder Vorgänger oder Nachfolger eines anderen Symbols, niemals beides gleichzeitig zu sein (S').

Diese Parameter sind geeignet, die Symbolrelationen innerhalb ein- und desselben Systems zu beschreiben. Phonologisch aufgefaßt ist das System das Phonemsystem; methodisch gesehen das einheitliche Transkriptionssystem. Anhand des Resultats kann noch einmal das wissenssoziologisch gestützte Argument aufgegriffen werden, ob und wie stark der Vollzug von Transkriptionsregeln in das Ergebnis einfließt. Es haben sich bei gleichem Transkriptionssystem, aber unterschiedlichem Material, andere Verwendungseigenschaften herausgestellt. Die Symboldistribution bildet in diesem Beispiel mehr als nur die Kombinatorik des Notationssystems ab; die Regeln des Systems sind weit genug gefaßt, um phonematische oder lautsprachliche Kombinationen abbilden zu können. Sachlich gesehen, d.h. wenn die statistischen Unterschiede der Verwendung von [p] als Ausspracheunterschied interpretiert wird, ist es evident, daß in der deutschen Umgangssprache [p] weit häufiger wie [p̥] ausgesprochen wird als in der Standardaussprache des aE bzw. als phonologisch überhaupt erfaßt werden kann. Es läßt sich auch das Argument vom Einfluß des Vorwissens auf die Transkription relativieren: Phonetisch erwartbar ist, daß alle stimmlosen Fortis-Klusile umgangssprachlich wie entstimmlichte stimmhafte Lenis-Explosivlaute ausgesprochen werden, aber nur [p] ist auffällig weniger notiert worden als [k] oder [t].

Nicht nur in der distributionellen Eigenschaft eines einzelnen Symbols, sondern auch in der Symbolähnlichkeit läßt sich ein Unterschied zwischen verschiedenen Sprachen (bei Verwendung des gleichen Transkriptionssystems) nachweisen. Zwei Symbole können sich aufgrund ihrer distributionellen Eigenschaften ähnlich sein; wenn sich die Paarähnlichkeit von Material zu Material ändert, kann die Differenz nicht auf Eigenschaften des Transkriptionssystems beruhen. Exemplarisch kann das am Symbolpaar [t] und [v] gezeigt werden, das sich im aE wie im Deutschen klassifikatorisch unterscheidet und zusätzlich durch [t] das aE und das Deutsche um eine Klasse trennt (Berechnung s. Tab. 8). Der Wert von $S_{attraktiv}$ drückt aus, ob Symbole, die vor [t] stehen, auch vor [v] stehen können (der Parameter $S_{aggressiv}$ im umgekehrten Sinne; vergleiche

Tabelle 8: Vergleich der Distribution der Symbole [t] und [v]

ALTMANN, LEHFELDT 1980, 248ff.). Die Verwendung von [t] und [v] zum Transkribieren des aE ist ähnlich wie die bei deutscher Umgangssprache in bezug auf die Zahl von Symbolen, die einerseits vor [t] und [v], andererseits nach [t] und [v] stehen können (S'_d). Sie ist es nicht in bezug auf die Zahl der Symbole, die überhaupt vor und nach beiden Symbolen auftreten (S_d). Die assoziative Ähnlichkeit von [t] und [v], die aus Attraktivität und Aggressivität kombiniert ist, ist im Deutschen konträr zum aE zusammengesetzt. [t] und [v] haben im Transkriptionssystem den gleichen distributionellen Stellenwert (zumindest beim Vergleich englischer und deutscher Transkriptionen); unterscheiden sich aber sprachspezifisch durch den Grad der Ausnutzung der maximal möglichen Übereinstimmung:

	aE	Deutsch
$S_{attraktiv}$	.63	.54
$S_{aggressiv}$	.56	.68
S_d	.61	.42
S'_d	.59	.60

(Werte für aE entnommen aus ALTMANN,LEHFELDT 1980,249 und 251)

2.2.3 Distributionelle Frequenz

Wenn auch die Häufigkeit, mit der sich Symbole kombinieren, berücksichtigt wird, dann können probabilistische Aussagen über die Kopplungstendenz gemacht werden. Mit zwei Kriterien - der Wahrscheinlichkeit der Kombination und der Existenz der Kopplung im Material - sind die Symbole in fünf Kategorien klassifizierbar:

	assoziativ	neutral	dissoziativ
$n_{ij}>0$	P (bevorzugt)	A (aktuell)	M (marginal)
$n_ij=0$	---	V (virtuell)	I (unzulässig)

Die Matrix der Tabellen 9.1 und 9.2 sowie deren gemeinsame Matrix wurden statistisch klassifiziert (Programm s. Anhang; Resultat in Tabellen 10.1 bis 10.3). Folgende Berechnungen werden durchgeführt (nähere Beschreibung s. ALTMANN,LEHFELDT 1980,295ff.):

Tabelle 9.1: Distributionelle Frequenz der Transkriptionssymbole f
schwäbisch-alemannisch beeinflußte Aussprache

	ˈˌ	h	ŋ	ə	˜	ʁ	œ	ˌ	ˑ	/	ø	ː	·	ʔ	ɒ	a	ç	ɛ	ɪ	ɔ	ʀ	ʃ	ʊ	ʏ	ɣ
ˈˌ	–	1	–	6	–	1	3	4	11	–	–	4	2	31	–	5	3	1	3	–	–	–	2	–	–
h	–	–	–	–	–	1	–	1	13	–	–	2	–	–	–	–	–	–	–	–	–	–	–	–	–
ŋ	2	1	–	8	–	1	–	3	2	1	–	–	–	1	–	1	–	–	1	–	–	–	–	–	–
ə	1o	12	1	8	–	9	3	2	29	12	–	7	7	9	9	1o	4	3	5	1	–	13	–	–	–
˜	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–
ʁ	–	–	–	5	–	–	–	1	–	–	–	3	1o	3	–	1o	2	7	1o	1	–	–	5	–	1
œ	–	1	5	2	–	–	–	3	4	–	–	1	–	–	1	–	1	–	–	–	–	–	–	–	–
ˌ	–	–	7	–	–	–	–	–	–	–	–	2	1	–	–	1	–	–	–	–	–	–	–	–	–
ˑ	–	1	–	–	–	–	–	–	–	–	–	1	–	–	–	1	–	–	–	–	–	–	–	–	–
/	2	8	–	11	–	–	1	1	33	6	–	3	7	24	–	11	4	–	4	2	–	1	–	–	–
ø	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
ː	–	–	1	3	–	–	1	2	1	2	1	–	1	1	–	3	2	2	–	–	–	–	–	–	–
·	–	–	1	5	–	–	1	–	–	–	–	1	–	–	1	1	3	1	1	–	–	–	1	–	–
ʔ	–	–	–	24	–	–	1o	3	1	22	–	5	6	3o	–	44	1	9	2o	4	–	–	9	–	1
ɒ	2	6	–	–	–	1	–	1	14	2	–	1	–	4	–	4	2	–	2	–	–	2	1	–	–
a	1	1	3	2	–	2	2	6	11	3	–	3	4	5	1	–	1	–	3	36	–	4	–	–	–
ç	5	5	–	7	–	2	1	1	29	5	–	5	–	8	–	3	1	3	7	2	–	2	1	–	1
ɛ	–	–	2	1	–	2	–	–	7	–	–	1	2	–	13	8	5	–	2	–	–	–	–	1	–
ɪ	–	3	3	–	–	2	–	–	1o	1	–	5	4	–	2	1	76	–	–	–	–	4	–	–	–
ɔ	1	1	–	5	–	–	1	1	13	1	5	1	2	1	1	–	–	–	1	1	–	–	–	–	–
ʀ	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
ʃ	–	1	–	3	–	–	–	1	14	–	1	–	3	2	–	1o	–	2	2	2	–	–	2	–	–
ʊ	1	–	7	2	–	2	–	1	–	–	–	3	2	–	2	–	–	–	–	–	–	–	–	–	–
χ	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
ʏ	–	–	–	2	–	–	–	–	1	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–
ʒ	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
a	2	5	3	3	–	2	–	1	13	5	–	1	–	6	2	1	1	1	4	2	–	–	–	–	–
b	1	–	–	1o	–	3	–	3	3	–	–	–	6	2	–	3	–	–	4	1	–	1	–	–	–
d	6	–	2	46	1	5	–	7	13	7	–	4	19	1o	1	35	2	13	9	5	–	5	6	–	–
e	–	1	1	–	–	3	–	5	9	1	–	2	8	5	8	2	4	2	3	–	–	1	2	–	–
f	2	1	–	14	–	4	1	5	12	3	–	2	1	7	–	14	–	1	1o	9	–	–	1	–	5
g	–	2	5	18	–	3	–	6	3	1	1	3	6	2	–	8	–	2	1	3	–	–	1	–	–
h	9	3	–	7	–	–	2	2	9	16	–	6	8	9	–	26	3	–	5	1	–	–	3	–	–
i	1	3	–	5	–	2	–	–	9	2	–	–	1	3	7	6	2	1	1	–	–	1	1	–	–
j	2	1	–	1	–	–	–	1	2	2	–	8	8	–	1	9	1	6	–	–	–	–	5	–	–
k	–	–	–	1	–	3	4	2	1	–	–	2	2	1	–	7	1	1	1	6	–	–	1	–	–
l	–	2	–	6	–	–	–	–	14	1	–	–	7	2	–	11	7	3	9	3	–	–	1	–	–
m	1o	2	–	16	–	–	1	4	18	6	–	7	8	7	–	3o	1	2	12	2	–	5	12	–	–
n	7	16	–	26	–	1	2	7	42	1o	1	8	9	25	–	18	5	4	29	5	–	12	1o	1	–
o	7	2	–	4	–	1	–	2	14	5	–	4	–	8	3	2	2	2	4	–	–	1	3	–	–
p	–	–	–	1	–	1	–	–	–	–	–	1	1	–	–	–	–	–	–	1	–	–	–	–	–
s	11	1o	1	15	–	6	2	6	57	27	1	4	15	27	–	25	2	5	19	8	–	4	6	–	–
t	2	5	–	16	–	5	–	1	5	8	–	6	9	7	–	1o	3	4	7	–	–	4	5	–	1
u	–	1	–	1	–	–	–	–	2	–	–	1	1	2	5	2	1	–	1	–	–	–	1	–	–
v	–	–	–	14	1	1	–	1	2	1	–	2	2	1	2	52	1	13	6	1	–	–	–	–	1
x	1	1	–	2	–	1	–	1	6	4	–	1	1	3	–	1	–	–	1	1	–	1	2	–	–
y	–	–	–	–	–	–	–	–	–	1	–	–	1	1	2	–	–	–	–	–	–	–	–	–	–
z	–	1	–	6	1	–	–	1	2	–	–	4	9	–	–	3	–	–	6	–	–	–	–	–	–
←	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
∘	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	7	1	–	–	–	–	–	–	–	–
Σ	85	98	22	3o7	3	64	35	86	435	155	1o	115	173	248	61	386	142	88	193	97		61	81	2	1o

	ʒ	a	b	d	e	f	g	h	i	j	k	l	m	n	o	p	s	t	u	v	x	y	z	←	∘	Σ
ˈˌ	–	1	1	4	–	1	–	1	–	4	–	–	12	18	–	–	5	2	–	9	–	–	–	–	1	ˈˌ	136
h	–	–	–	11	–	–	6	–	–	–	38	–	1	1	1	1	1	27	–	–	–	–	–	–	–	h	1o4
ŋ	–	–	–	4	–	–	2	–	–	–	1	–	1	3	–	–	2	4	–	6	–	–	–	–	–	ŋ	44
ə	–	–	7	22	1	14	6	13	–	2	8	7	11	17	–	–	51	8	–	9	1	–	16	–	1	ə	348
˜	–	1	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	˜	4
ʁ	–	6	–	–	4	–	–	–	2	–	–	–	1	3	1	–	1	–	2	–	–	–	–	–	–	ʁ	78
œ	–	–	–	3	–	1	2	–	–	–	–	–	–	4	–	–	–	–	–	1	–	–	–	–	–	œ	29
ˌ	–	–	–	1	–	–	–	–	–	–	–	–	11	54	–	–	–	–	–	–	–	–	–	–	–	ˌ	77
ˑ	–	–	44	213	–	–	–	2	–	–	–	–	–	1	–	1	–	–	1	56	–	–	5	–	–	ˑ	326
/	–	1	4	23	1	2	4	7	–	–	4	–	7	8	2	–	9	6	–	6	–	1	–	–	–	/	2o3
ø	–	–	1	–	–	1	–	–	–	–	–	–	1	2	–	–	–	3	–	–	–	–	1	–	–	ø	1o
ː	–	11	–	1	4	4	2	2o	3	–	–	–	7	9	7	–	16	1	1	1	1	–	–	–	1	ː	1o9
·	–	29	–	–	3o	8	–	2	19	–	–	–	–	7	14	–	2o	–	8	–	–	1	–	–	·	194
ʔ	–	4	–	1	5	–	–	2	1	–	–	–	11	6	2	–	–	–	–	–	1	3	–	–	1	ʔ	226
ɒ	–	–	–	8	–	4	5	2	–	2	3	3	1	3	–	–	3	8	–	5	–	–	–	–	–	ɒ	89
a	–	1	3	7	78	8	–	1	–	3	–	27	5	44	–	–	46	9	–	6	21	–	–	–	2	a	349
ç	–	–	2	15	–	5	5	8	–	4	3	1	11	4	–	–	11	1o	1	11	–	1	–	–	–	ç	18o
ɛ	–	–	–	4	–	2	3	3	–	2	1	7	3	18	–	–	3	2	–	–	1	–	–	–	–	ɛ	93
ɪ	–	–	1	6	–	2	5	1	1	2	5	6	9	16	–	–	17	5	–	1	–	–	–	–	–	ɪ	188
ɔ	–	–	5	9	–	8	–	1	–	–	–	3	1o	13	–	–	1o	2	–	4	11	–	1	–	–	ɔ	112
ʀ	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	ʀ	1
ʃ	–	–	2	7	–	1	–	1	1	–	–	1	1	4	3	2	–	1o	1	4	–	–	–	–	–	ʃ	81
ʊ	–	1	–	–	–	–	–	–	–	–	–	1	1o	28	1	–	11	–	–	–	–	–	–	–	–	ʊ	72
χ	–	–	–	–	1	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	χ	3
ʏ	–	–	–	–	–	–	–	–	–	–	–	1	–	1	–	–	1	–	–	1	–	–	3	–	–	ʏ	11
ʒ	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	ʒ	–
a	–	–	3	13	1	1	8	2	–	1	2	8	1	6	1	–	3	7	–	2	2	–	–	–	–	a	118
b	–	–	–	2	–	3	2	–	7	1	–	4	1	6	1	–	2	1	–	2	1	–	–	–	–	b	7o
d	–	1o	–	4	12	5	2	5	24	1	–	4	7	13	1	–	29	–	8	7	–	–	3	–	–	d	331
e	–	–	3	12	1	3	7	–	–	1	2	9	14	29	1	–	23	6	–	–	–	–	1	–	–	e	169
f	–	–	2	6	–	2	4	–	4	–	1	1	2	7	1	–	3	2	–	1	–	2	2	–	–	f	132
g	–	6	–	2	3	1	–	–	2	1	–	5	2	7	2	–	4	2	–	–	–	–	–	–	–	g	1o2
h	–	3	2	7	–	2	1	6	2	2	–	–	3	8	2	–	2	3	–	–	–	–	–	–	2	h	154
i	–	1	4	2	–	2	4	–	–	1	1	1	4	2	1	1	2	4	–	2	–	–	3	–	–	i	8o
j	–	18	–	2	4	–	–	–	–	–	–	–	2	–	4	–	–	–	–	–	–	–	–	–	3	j	8o
k	–	–	–	–	3	1	–	1	–	–	–	1	2	3	2	–	6	–	–	1	–	1	–	–	–	k	54
l	–	2	1	15	5	2	1	1	2	–	–	–	2	5	5	–	6	8	–	2	–	–	9	–	–	l	132
m	–	7	6	11	3	1	1	5	3	3	–	4	7	6	2	1	9	5	–	1	1	–	1	–	–	m	22o
n	–	7	3	46	4	7	5	19	1	8	8	4	17	9	1	–	12	28	1	13	–	–	5	–	1	n	437
o	–	–	2	8	–	5	4	2	–	1	–	3	6	13	1	–	1	2	–	12	1	–	3	–	1	o	129
p	–	–	–	–	1	–	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	p	8
s	–	4	6	3o	4	11	11	5	8	14	5	3	18	26	3o	1	7	23	23	2o	–	–	–	–	–	s	5oo
t	–	5	–	1	6	1	2	7	3	3	2	1	1	1	3	–	48	3	6	4	1	2	–	–	–	t	198
u	–	–	–	2	–	2	–	6	–	–	1	3	2	–	–	–	–	1	–	1	3	–	2	–	–	u	41
v	–	5	–	2	3	1	–	–	11	–	–	–	–	1	2	–	5	–	–	1	–	–	1	–	3	v	135
x	–	–	1	4	1	3	–	–	–	–	1	–	1	6	–	1	5	8	–	1	–	–	–	–	–	x	59
y	–	–	–	–	–	1	–	–	–	–	–	–	–	–	1	–	–	–	–	3	–	1	–	–	–	y	11
z	–	4	–	1	–	–	1	–	–	–	–	–	–	1	24	–	–	1	3	–	–	–	–	–	–	z	68
←	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	←	–
∘	–	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	∘	11
Σ	–	128	1o4	5o9	175	115	93	124	95	56	88	1o7	2o6	412	117	8	374	2o1	55	193	45	12	56	–	16		6266

Tabelle 9.2: Distributionelle Frequenz der Transkriptionssymbole für berlinisch beeinflußte Aussprache

	ll	h	ŋ	ə	~	ʁ	æ	.	◦	/	ø	:	·	ʔ	ɒ	a	ç	ɛ	ɪ	ɔ	ʀ	ʃ	ʊ	χ	ʏ
ll	-	1	-	-	-	-	-	-	6	-	-	3	1	3	1	4	1	-	4	-	-	1	1	-	-
h	1	-	-	-	-	-	-	-	3	-	-	1	1	-	-	-	-	-	1	-	-	-	-	-	-
ŋ	1	-	1	1	-	-	-	-	2	2	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ə	3	2	1	1	-	6	1	2	8	4	-	3	4	2	15	6	1	1	1	-	-	1o	-	-	-
~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʁ	-	-	-	3	-	-	-	-	1	-	-	4	4	-	-	1	-	4	4	3	-	-	-	-	-
æ	-	-	-	-	-	-	-	1	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-
.	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
◦	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
/	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ø	1	3	-	3	-	-	-	-	25	5	-	6	1	14	-	1o	2	4	-	-	1	-	-	-	1
:	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
·	1	2	-	2	-	-	-	4	2	-	-	1	-	-	1	-	1	1	4	-	-	-	-	-	-
ʔ	-	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	3	2	-	-	-	-	-	-
ɒ	-	-	-	6	-	-	-	-	-	2	-	3	2	1	-	17	1	1o	1o	1	-	-	-	-	-
a	5	3	-	-	-	-	-	2	9	13	-	2	2	3	-	6	-	3	-	2	-	1	1	-	-
ç	-	1	-	-	-	-	-	3	9	2	1	7	1	2	2	-	1	1	1	14	-	-	1	-	-
ɛ	2	1	-	1	-	2	-	4	9	3	-	2	-	4	-	1	1	1	5	-	1	2	-	-	-
ɪ	2	-	-	-	-	2	-	1	2	3	-	2	1	4	1	3	6	2	2	-	-	-	-	-	-
ɔ	-	-	4	-	-	-	-	3	7	1	-	4	-	-	5	-	3o	-	-	-	-	-	-	-	-
ʀ	-	-	-	1	-	-	-	1	7	1	3	3	-	-	1	-	-	-	1	-	-	-	-	-	-
ʃ	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	1	1	-	-	-	-	-	-
ʊ	-	-	-	1	-	-	-	2	3	-	1	1	1	-	1	-	2	-	1	1	-	-	-	-	-
χ	-	-	4	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʏ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʒ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
a	-	-	-	-	-	1	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
e	8	-	-	-	-	-	-	1	9	3	1	-	1	1	3	1	-	-	2	-	-	1	-	-	-
f	1	-	-	11	-	-	1	3	1	-	-	1	2	-	1	6	1	2	3	-	-	1	-	-	-
g	7	-	-	19	-	1	-	2	7	4	-	2o	1o	1	2	27	5	1	4	4	1	1	1	-	-
h	-	-	1	1	-	-	-	1	3	2	-	3	1	4	24	3	4	-	-	1	-	-	-	-	-
i	1	1	-	2	-	-	-	1	2	3	-	3	5	1	-	5	-	3	4	4	-	-	-	-	1
j	-	-	-	15	-	2	-	2	1	-	-	1	2	-	-	4	-	-	1	1	-	-	-	-	-
k	6	2	-	7	2	1	-	-	-	14	-	1o	4	9	-	11	1	4	-	-	-	-	-	-	-
l	1	3	-	2	-	1	-	-	3	4	-	6	2	4	4	7	-	2	2	1	-	-	2	-	-
m	-	-	-	2	-	-	-	-	1	3	-	5	3	-	-	1	-	5	-	-	-	-	1	-	-
n	-	-	1	1	-	-	1	1	-	-	-	-	-	-	-	2	-	2	-	2	-	-	-	-	-
o	2	-	-	3	-	-	-	-	8	3	-	4	4	2	-	8	2	-	4	2	-	1	-	-	-
p	19	1	-	4	-	-	-	-	8	5	-	7	2	-	-	16	2	1	9	2	-	-	5	-	1
s	7	1	-	11	-	3	-	3	19	1o	1	13	4	1o	-	14	1	4	9	2	1	2	-	-	1
t	5	-	2	1	-	-	-	3	4	4	-	5	2	4	2	5	-	-	5	-	-	1	-	-	-
u	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
v	5	5	-	5	-	-	-	6	25	18	1	12	7	8	-	8	4	3	11	1	-	-	1	-	1
x	1	1	-	9	-	1	-	4	2	1	-	3	3	-	-	3	4	1	3	-	-	-	4	-	-
y	1	-	-	-	-	-	-	-	1	-	-	1	-	1	5	-	-	-	2	-	-	-	-	-	-
z	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
←	1	-	-	2	-	-	-	4	7	2	-	-	1	1	-	2	-	2	-	1	-	-	-	-	-
◦	-	-	-	4	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	3	2	1	-	3	8	-	-	7	-	1	5	4	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Σ	85	16	2	3	198	7	84	73	7o	1o5	5	17	5	27	121	25	57	116	149	8o	2o2	64	51	22	-

	ɐ	b	d	e	f	g	h	i	j	k	l	m	n	o	p	s	t	u	v	x	y	z	←	◦		Σ
ll	-	1	6	1	1	1	8	-	7	-	2	13	13	-	-	6	1	-	5	-	1	-	-	-	ll	92
h	-	-	4	1	-	1	-	-	1	3	-	-	2	-	-	-	6	1	-	-	1	-	-	-	h	27
ŋ	-	-	1	-	-	3	-	-	1	1	-	2	1	-	-	2	-	-	-	-	-	-	-	-	ŋ	19
ə	1	1	9	2	5	2	6	-	1	2	2	3	4	-	-	1o	4	-	5	1	1	6	-	-	ə	136
~	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	~	2
ʁ	6	-	1	-	-	-	-	-	-	-	-	-	2	-	-	-	1	-	-	-	-	-	-	-	ʁ	34
æ	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	æ	5
.	-	-	-	-	-	-	1	-	-	-	-	12	37	-	-	-	-	-	-	-	-	-	-	-	.	52
◦	-	22	84	1	-	12	-	-	-	-	-	-	1	-	-	1	-	-	25	-	-	19	-	-	◦	165
/	-	6	2o	1	3	5	6	-	1	2	2	7	9	1	5	4	7	-	3	1	-	4	-	-	/	163
ø	-	-	2	-	-	-	-	-	-	-	-	1	3	-	-	-	-	-	-	-	-	-	-	-	ø	11
:	9	-	2	11	3	-	11	27	-	1	1	19	9	12	-	11	1	4	-	-	4	-	-	-	:	144
·	17	-	1	1o	4	-	-	11	-	-	-	-	3	14	-	1	-	3	-	-	6	-	-	-	·	77
ʔ	1	-	1	4	-	-	2	2	-	-	-	3	3	1	-	-	-	-	-	1	1	-	-	-	ʔ	72
ɒ	-	2	9	1	4	1	2	-	1	2	3	7	8	1	-	5	1	-	6	-	1	3	-	-	ɒ	1o9
a	-	4	7	36	1	1	-	-	1	-	26	4	13	-	-	41	2	-	3	13	-	2	-	-	a	2oo
ç	-	-	6	1	5	-	2	-	-	2	-	4	7	1	-	6	7	1	6	-	-	2	-	-	ç	89
ɛ	-	-	2	-	2	-	2	-	-	-	6	3	7	1	-	4	2	-	-	-	-	1	-	-	ɛ	61
ɪ	-	1	5	-	1	1	1	-	1	-	5	4	12	-	-	1o	5	-	1	-	-	-	-	-	ɪ	1o1
ɔ	-	7	4	-	5	-	-	-	-	1	7	3	1	-	-	4	2	-	-	8	-	1	-	-	ɔ	61
ʀ	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ʀ	5
ʃ	-	-	1	-	-	-	-	-	1	-	-	2	6	-	-	3	-	-	-	-	-	-	-	-	ʃ	29
ʊ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ʊ	19
χ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	χ	-
ʏ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	ʏ	5
ʒ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ʒ	1
a	-	4	5	-	1	2	3	-	2	-	2	4	3	-	-	4	4	-	3	1	-	2	-	-	a	71
b	-	-	2	-	-	-	3	5	-	-	3	6	1	-	-	4	-	-	-	-	-	1	-	-	b	59
d	1	5	1	3	16	-	1	2	24	3	1	4	4	4	1	-	6	-	8	5	-	-	4	-	d	21o
e	-	1	2	4	1	1	2	1	1	1	-	5	9	21	-	-	4	-	-	1	-	1	2	-	e	1o5
f	-	-	-	1	-	1	2	3	6	-	-	2	-	2	2	-	4	2	-	1	1	4	-	-	f	67
g	-	3	-	1	4	-	-	-	-	-	-	2	1	2	-	-	1	-	-	-	-	-	-	-	g	43
h	-	6	-	3	1	1	-	9	1	1	-	-	7	12	-	-	2	1	-	-	-	-	1	1	h	117
i	-	-	-	2	-	4	1	-	1	1	1	6	4	-	2	-	4	3	-	1	-	3	5	-	i	82
j	-	8	1	-	-	-	-	1	-	1	-	-	-	2	1	-	-	-	-	-	-	-	-	-	j	35
k	-	-	-	-	-	-	-	1	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	k	13
l	-	-	5	9	5	-	1	1	-	1	-	2	2	4	5	-	14	6	-	2	1	-	9	-	l	11o
m	-	6	4	3	2	2	3	5	1	2	1	1	18	5	-	1	4	4	-	3	-	-	1	-	m	148
n	-	4	7	27	6	3	3	2o	-	1	1	1	8	7	2	1	6	5	3	6	-	1	8	1	n	237
o	-	1	4	5	1	6	3	2	-	-	-	4	2	11	-	-	2	1	-	1	1	1	1	-	o	89
p	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	3	-	-	-	-	p	1o
s	-	-	2	3o	7	4	7	3	5	2	2	1	5	14	18	2	2	8	3	7	-	-	3	-	s	246
t	-	1	1	4	3	1	-	1	2	2	-	1	5	-	-	15	-	3	-	-	-	-	-	-	t	79
u	-	-	1	3	1	2	2	-	-	-	-	-	2	-	-	-	-	-	-	-	1	-	-	-	u	23
v	-	3	-	1	3	-	-	-	13	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	v	67
x	-	1	-	4	-	1	1	1	-	-	-	-	3	6	-	-	1	1	1	2	-	-	1	-	x	46
y	-	1	2	1	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	3	-	-	-	-	y	12
z	-	2	-	2	1	-	-	-	-	-	1	-	-	3	24	1	-	-	-	-	-	-	-	-	z	72
←	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	←	2
◦	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	◦	-
Σ	1	78	79	276	121	61	56	98	1oo	32	2o	91	16o	249	91	15	184	77	28	9o	31	25	77	2	-	3622

$E_{ij} = \Sigma_i \cdot \Sigma_j / \Sigma_{ij}$

wenn $E_{ij} \leq 30$:

wenn $n_{ij} < E_{ij}$:

$$P(X \leq n_{ij}) = \sum_{k=n_{ij}+1}^{\infty} \frac{e^{-E_{ij}} E_{ij}^{k}}{k!}$$

P>o.o5 neutral
P≤o.o5 dissoziativ

wenn $n_{ij} > E_{ij}$:

$$P(X \geq n_{ij}) = \sum_{n_{ij}}^{\infty} = (e^{-E_{ij}} E_{ij}^{k}/k!)$$

P>o.o5 neutral
P≤o.o5 assoziativ

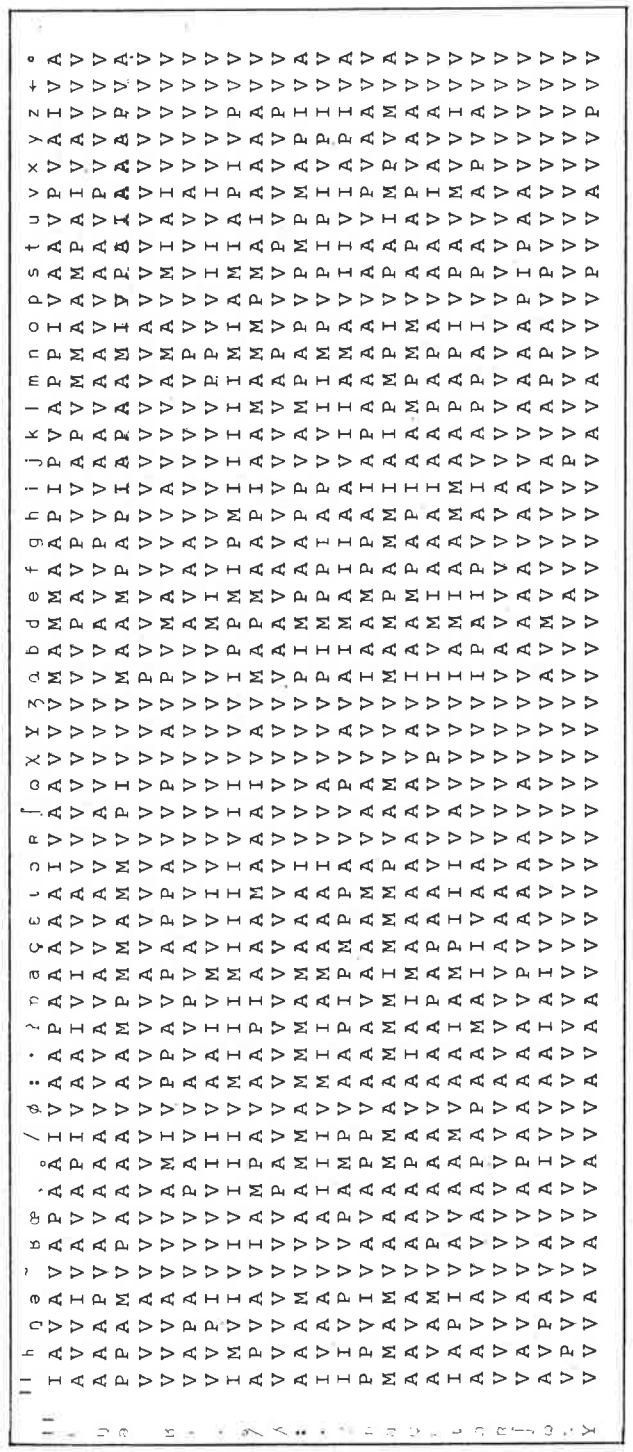
wenn $E_{ij} > 30$:

$$z = (n_{ij} - E_{ij}) / \sqrt{n_{i.} n_{.j} (n-n_{i.})(n-n_{.j}) / n^2 (n-1)}$$

z≥1.645 assoziativ
z≤-1.645 dissoziativ
-1.645< z <1.645 neutral

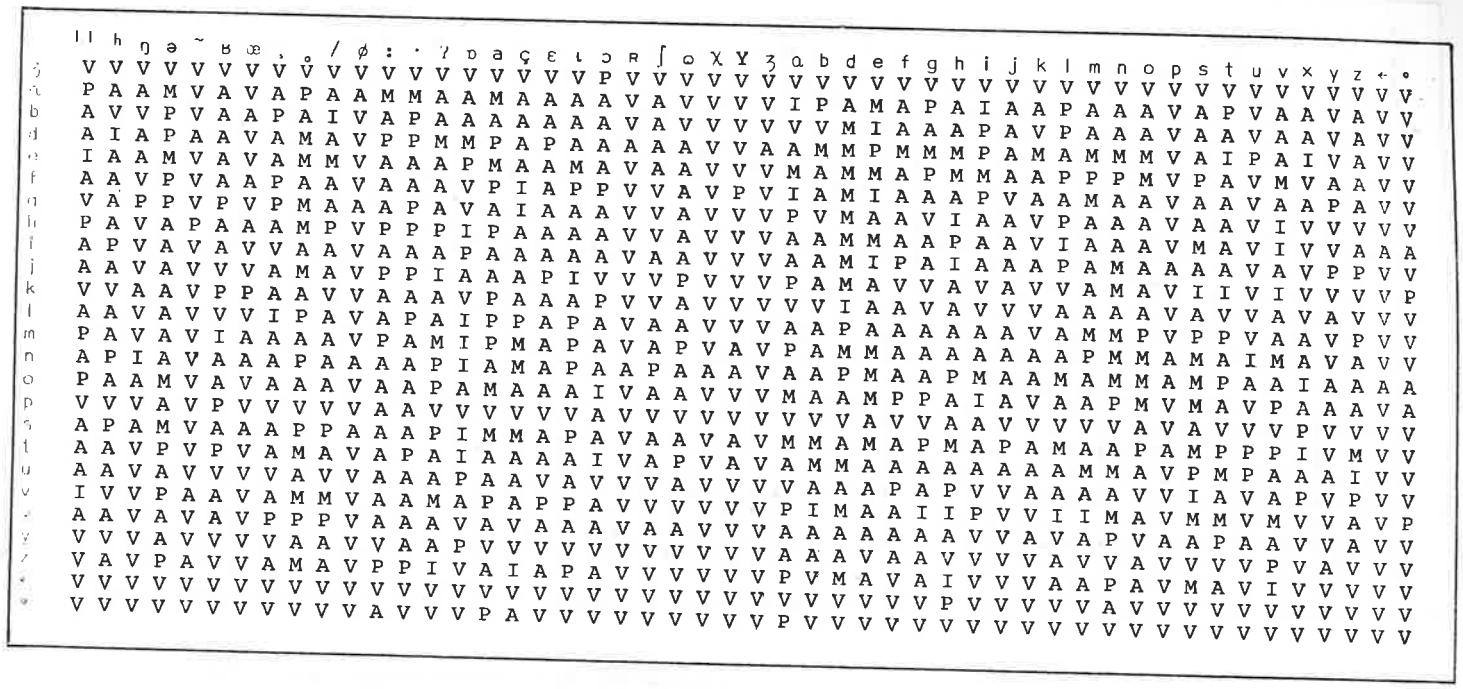
Die phonetischen Symbole [ʰ ˜ ː · ˳ ˚ ˌ ←] stehen computertechnisch bedingt vor dem zugehörigen Transkriptionszeichen; die Kombinationen mit diesen Zeichen sind in den Tabellen 9 bis 11 als "gleichzeitig" oder "nachfolgend" notiert zu lesen. [ʰ] vor [t] ist identisch mit [tʰ]; [a] nach [˜] mit [ã]. Weil die Kombinationsanalyse nur Paarbeziehungen berücksichtigt, ist ein gewisser Verlust an phonologisch interessierenden Beziehungen entstanden: Folgen wie [˳] nach [o] bedeuten, daß irgendein entstimmlichter Laut dem Vokal folgt, wobei nicht erfaßt wird, welcher.

Tabelle 1o: Klassifikation der Transkriptionssymbole nach ihrer distributionellen Frequenz
1o.1: schwäbisch-alemannisch und berlinisch beeinflußte Aussprache
1o.2: schwäbisch-alemannisch beeinflußte Aussprache
1o.3: berlinisch beeinflußte Aussprache

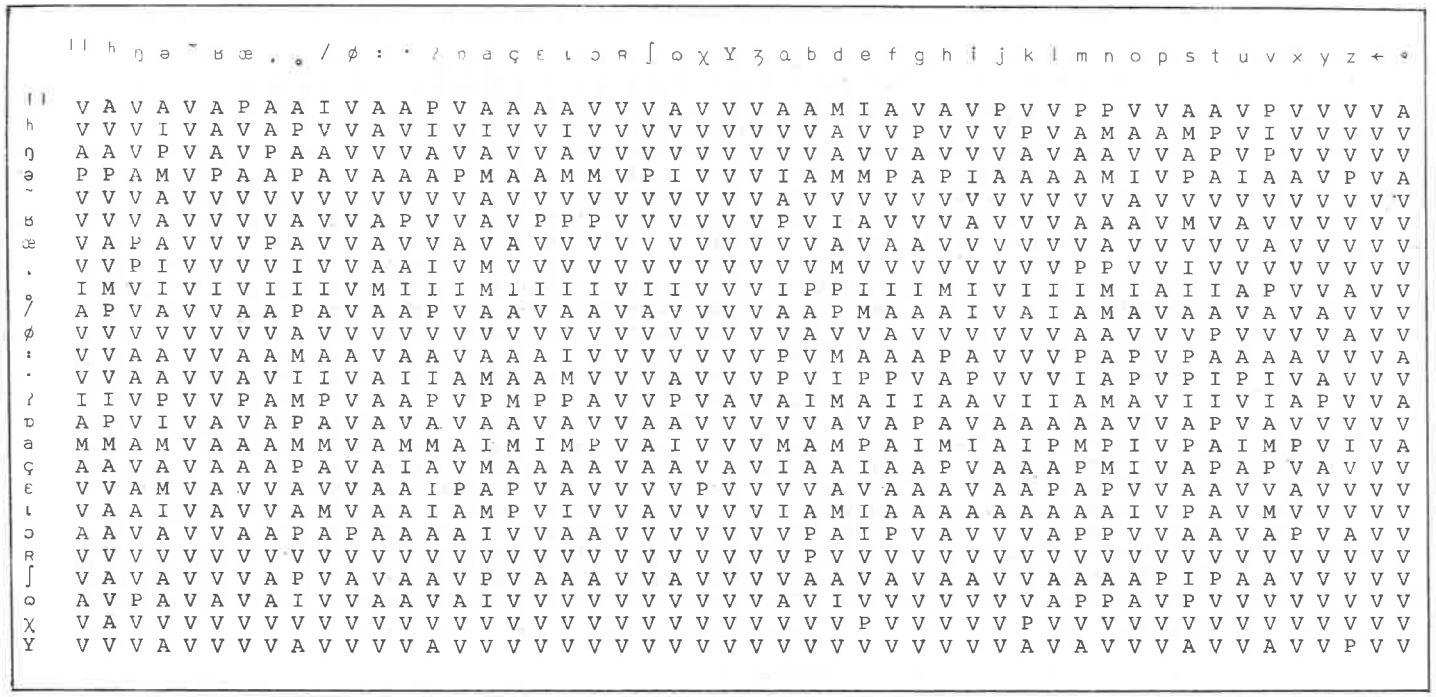


P = bevorzugt (= assoziativ, existent), A = aktuell (= neutral, existent), M = marginal (= dissoziativ, existent), V = virtuell (= neutral, nichtexistent), I = unzulässig (= dissoziativ, nichtexistent)

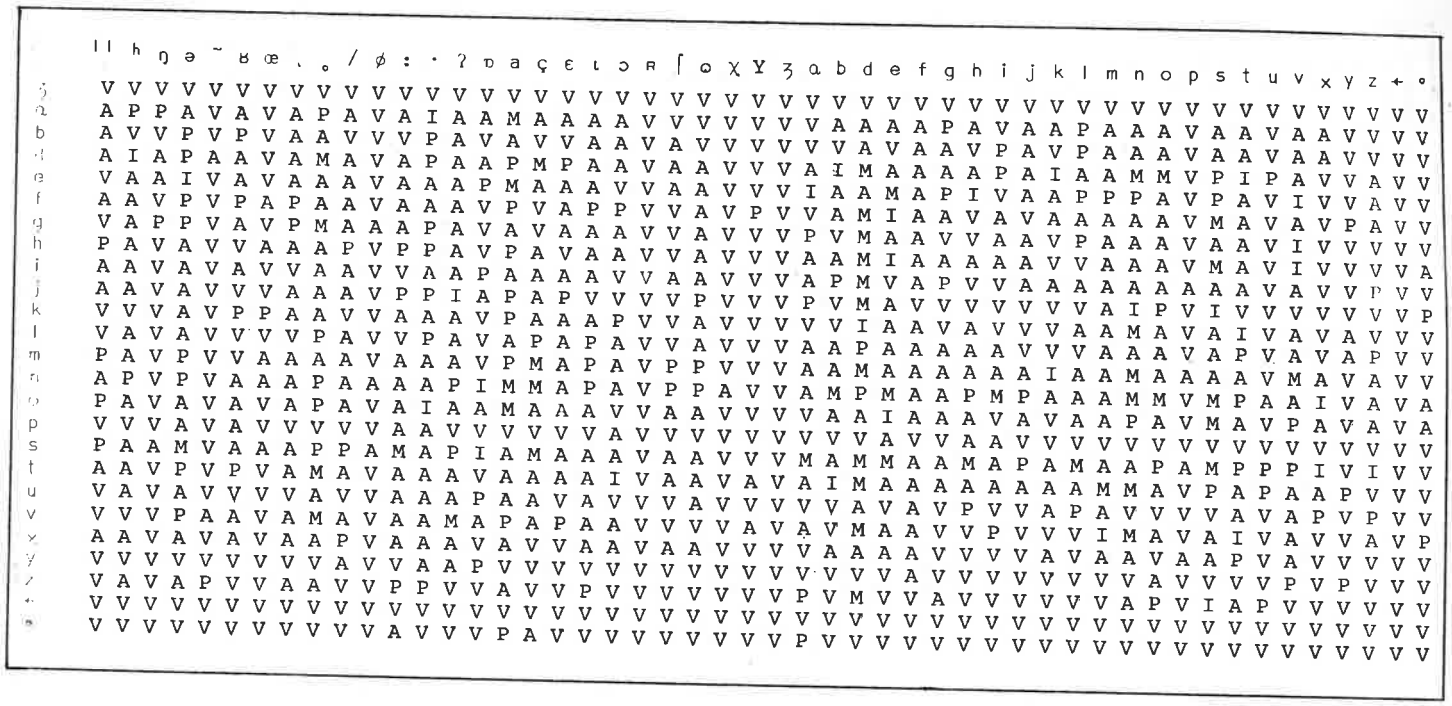
1o.1 (Fortsetzung)



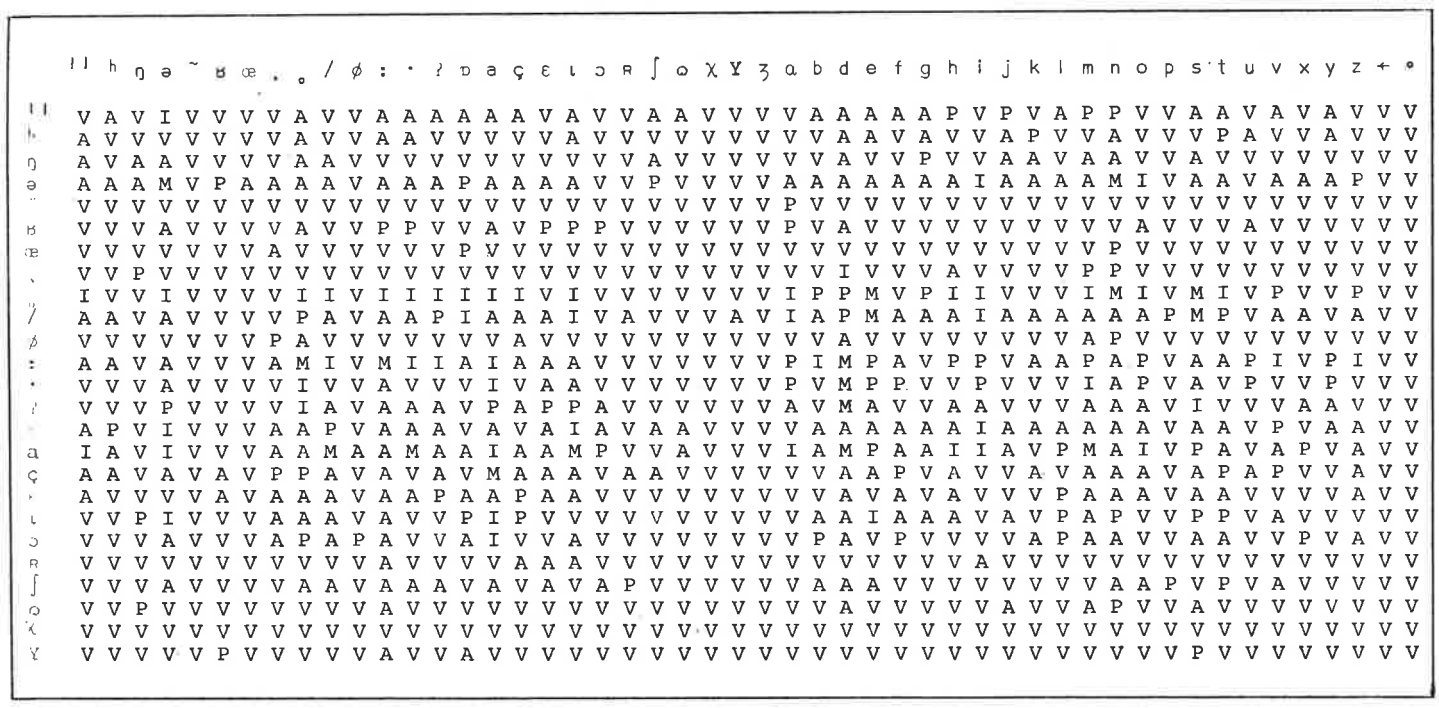
1o.2

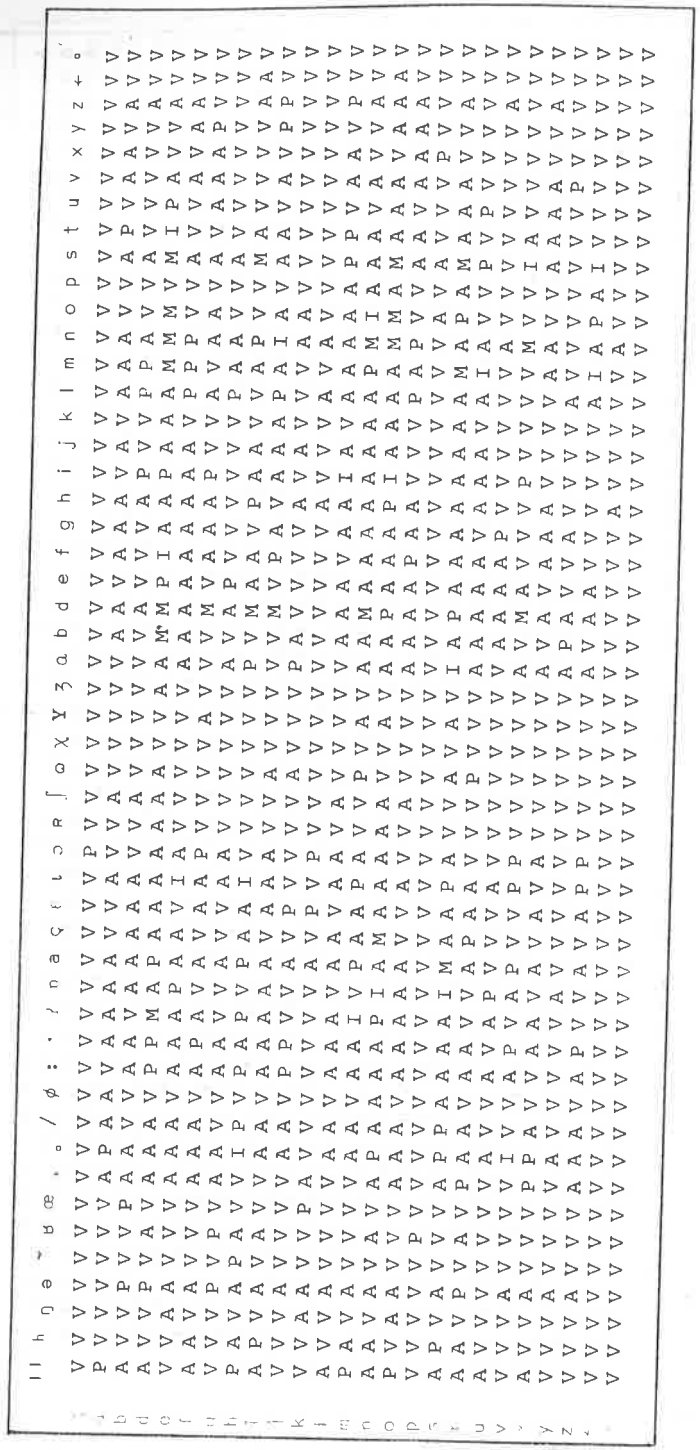


1o.2 (Fortsetzung)



1o.3





1o.3 (Fortsetzung)

Mit den statistischen Beziehungen werden - soweit im Transkript als phänomenale und hermeneutische Dokumente niedergelegt - mehrere Relationen ausgedrückt:

- Transkriptionsregeln: Kombination [a] und [i] ist unzulässig (I), weil der Diphtong [ae] notiert wird;
- Vorwissen (im Zusammenhang mit Transkriptionsregeln): [ɔ]+[ː] ist unzulässig (offne Vokale sind kurz);"Selbstverständlichkeiten" wie:[ˌ]+[t] ist unzulässig, ebenso [·]+[g] ;
- standardmäßige deutsche Artikulationstendenzen: [ʰ]+[k] oder [ˌ] + [b]= bevorzugt; [ʃ]+[ʁ]=unzulässig;
- umgangssprachliche Lautung: [œ]+[ŋ] = bevorzugt (aus dem Wortteil "irgend...");
- koartikulatorische Verbindungen: [j] + [ʔ]= unzulässig; [d]+[t] = unzulässig;
- sprachliche (phonematische, morphologische) Relationen: [ə]+[ˈˈ] = bevorzugt (Endsilben) oder [o] + [ɔ]= unzulässig;
- Sprechdynamik (Pausen, Lautlängen, nichtsprachliche phonetische Elemente wie Lachen usw.): [ʔ]+[/] = bevorzugt (beginnende Phonation durch eine rhythmische Stauung nach dem Glottiseinsatz abgebrochen); [n] + [ʔ]= bevorzugt (die rhythmische Verzögerung wird durch einen Glottisschlag ausgedrückt, der koartikulatorisch nicht notwendig oder erwartbar ist; oder als Relikt aus "na ich...", "na aber" usw., wobei der Vokaleinsatz stellvertretend den nicht realisierten Laut andeutet; für solche Fälle spricht, daß die Verbindung [m] + [ʔ] statistisch unzulässig ist);
- konversationsanalytische Auffälligkeiten: Redeabbrüche mitten in der phonetischen Realisation, z.B. [ʔ] oder [h] vor Stauungen = statistisch bevorzugt;
- sprechertypische (dialektale und paralinguistische) Verwendungen der potentiellen Kombinatorik: [ʔ]+[ʔ] = bevorzugt; der Vergleich zwischen beiden Sprechern und beide jeweils zum Gesamt der umgangssprachlichen Artikulation in Tab. 12 gibt dazu nähere Informationen.

Zum besseren Verständnis der Tabellen 1o.1 bis 1o.3 sind die statistischen Beziehungen in Tabelle 11 - jeweils die Extreme P und I - nach phonetischen Gesichtspunkten geordnet nochmals aufgelistet (die acht Sonderzeichen sind "umgepolt" zu lesen).

Tabelle 11: Zusammenstellung der statistischen Klassifikationen P (bevorzugt) und I (unzulässig) gemäß Tab. 1o.1

vor * steht * nach * folgt bevorzugt:	vor * ist * nach * ist unzulässig:
1. Vokale	
[: · b d f v] ⇒ [i] ⇐ [ʰ ɒ f l y z]	[ç ɪ̯ ɛ ɑ a ɔ o ₒ ə ɒ ø] ⇐ [i] ⇒ [e]
[ʁ ʔ f l m n s v z] ⇒ [ɪ] ⇐ [ŋ ç l n s]	[· ₒ i j] ⇐ [ɪ] ⇒ [ˈˈ ə · ɛ i ɔ ɑ e o z]
[: · a d] ⇒ [e] ⇐ [ɒ g l m n s]	[ɔ b f ɪ ɪ ɛ ˑ] ⇐ [e] ⇒ [ˈˈ]
[ʁ ʔ d j v] ⇒ [ɛ] ⇐ [ʁ ɒ ç χ l h]	[ₒ ɪ] ⇐ [ɛ] ⇒ [ɑ e i v]
[g t v z f d ŋ ʔ] ⇒ [ə] ⇐ [ʰ ʁ k ˈˈ ɒ f h s z]	[. ₒ ɒ ɪ] ⇐ [ə] ⇒ [ɷ u o]
[ʁ : · g ₒ ~ j m v z] ⇒ [ɑ] ⇐ [ˈˈ ₒ d b t g l]	[ɒ ɛ ɪ ɔ ₒ ç ɑ f] ⇐ [ɑ] ⇒ [ɑ i]
[ʁ ʔ ʃ d f h k l m v ₒ] ⇒ [a] ⇐ [ɔ e l m s x]	[ʰ a ɔ ɷ] ⇐ [a] ⇒ [ɑ i k o u]
[: · l s] ⇒ [o] ⇐ [ˈˈ ʔ f g n v]	[ˈˈ əₒ a ɪ ɔ] ⇐ [o] ⇒ [ɔ i]
[a ʒ f k] ⇒ [ɔ] ⇐ [ₒ ø b f m l t x]	[ˈˈ ˈ · ɪ o t̊] ⇐ [ɔ] ⇒ [a ç ɑ e i]
[: · d s t] ⇒ [u] ⇐ [ɒ f h x z]	[ə / a m] ⇐ [u] ⇒ [s]
[ʁ ʔ j m t] ⇒ [ɷ] ⇐ [ŋ m n s]	[ə ₒ /] ⇐ [ɷ] ⇒ [ʔ a]
[ɔ] ⇒ [ø] ⇐ [. n t z]	⇐ [ø] ⇒

abelle 11 (1. Fortsetzung)

or * steht * nach * folgt bevorzugt:	vor * ist * nach * ist unzulässig:
[ˈˈ k ʔ] ⇒ [æ] ⇐ [ŋ , ɔ n]	⇐ [ɐ] ⇒
[: · ʔ f i] ⇒ [y] ⇐ [ɔ v]	⇐ [y] ⇒
[f] ⇒ [Y] ⇐ [s z]	⇐ [Y] ⇒
[œ ɛ e i u y] ⇒ [ɒ] ⇐ [ˈˈ h ₒ / f v g k]	[ₒ ʔ ç h l . m n s t] ⇐ [ɒ] ⇒ [ŋ ɑ i]
2. Klusile	
[ₒ ɔ ɑ] ⇒ [ɒ] ⇐ [ə . · i ɪ]	[: · ʔ v] ⇐ [b] ⇒ [/ e]
[ʰ ₒ / l n] ⇒ [d] ⇐ [ə : ˑ a ɛ e i u]	[k] ⇐ [d] ⇒ [ʰ t x]
[ɑ e o s ɒ ŋ ʰ ₒ] ⇒ [g] ⇐ [ŋ ə ʁ ɪ ɑ . ˑ]	[· ʔ v] ⇐ [g] ⇒ [ç h v]
[/ ʃ] ⇒ [p] ⇐ [ʁ v]	⇐ [p] ⇒
[ɑ ɪ n s x ʰ ø ç ʃ] ⇒ [t] ⇐ [ə ʁ · ɷ s]	[· ʔ d j ₒ . ʁ] ⇐ [t] ⇒ [ɒ ɔ ʒ]
[ʰ ə ɒ] ⇒ [k] ⇐ [ʁ œ a ɔ]	[ₒ ʔ a] ⇐ [k] ⇒ [d]
[ʰ n o s ʔ / ˈˈ] ⇒ [ʔ] ⇐ [ə ʁ / ʔ a ɛ i ɷ y]	[ɪ ɷ ʒ j . ʰ ˑ ₒ] ⇐ [ʔ] ⇒ [ˈˈ h ɒ b f ʔ g k l s t ʒ]
3. Nasale	
[ɔ ɷ e m ç : . ˈˈ] ⇒ [m] ⇐ [ˈˈ : a ɪ ɷ ɑ m]	[ₒ · v] ⇐ [m] ⇒ [ʁ ɒ u]

Tabelle 11 (2. Fortsetzung)

vor * steht * bevorzugt:	nach * folgt bevorzugt:	vor * ist unzulässig:	nach * ist unzulässig:
[ɪ ɑ e o x z ; ɛ a ø ˌ œ ˈˌ]	⇒ [n] ⇐	[ʃ d h t ; ₒ h ˀ ɪ]	⇐ [n] ⇒ [ŋ ɾ x]
[œ ˌ g ɪ ɑ]	⇒ [ŋ] ⇐ [ə v g]	[n] ⇐ [ŋ] ⇒	
4. Lateral			
[ɑ b e g i ; ɔ ɪ ɛ a]	⇒ [l] ⇐	[ₒ ˑ a ç ɪ ; o d s t z]	[ˑ ₒ ˀ v] ⇐ [l] ⇒ [ˌ ˑ]
5. Vibranten			
	⇒ [ʁ] ⇐		⇐ [ʁ] ⇒
[ə ɛ g k p t]	⇒ [ʁ] ⇐ [: ˑ ɛ ɪ ɑ ɑ]	[ₒ / m] ⇐ [ʁ] ⇒	[/ t v]
6. Konstriktive			
[ç o s y ; ŋ ₒ ˈˈ ŋ]	⇒ [v] ⇐ [ᵒi ɑ ɪ ; ɛ a ə]	[ɛ g h j z ; h ʁ ˌ ˑ ˀ]	⇐ [v] ⇒ [ˈˈ d g ; h l m]
[Y i l u ; ə ₒ ˑ]	⇒ [z] ⇐ [ə ɪ ˑ ɪ ɑ ɑ]	[ˈˈ : ˑ ʔ ɪ t]	⇐ [z] ⇒ [ˀ ç h v]
	⇒ [ʒ] ⇐ [ɔ]		⇐ [ʒ] ⇒
[ˈˈ χ s]	⇒ [j] ⇐ [: ˑ ɛ ɑ ɑ ᵉ]	[ₒ]	⇐ [j] ⇒ [ʔ ɪ s t v]
[ɔ i o u ; ɒ ˑ ə ç]	⇒ [f] ⇐ [ə ȯ a ; ɔ Ẏ i y]	[ₒ ˀ]	⇐ [f] ⇒ [ç ɑ e]

abelle 11 (3. Fortsetzung)

or * steht * evorzugt:	nach * folgt bevorzugt:	vor * ist unzulässig:	nach * ist unzulässig:
[ɒ ɔ Y e ɪ t ; ɪ a ˑ ː ə]	⇒ [s] ⇐ [h ₒ / ˀ ɪ ; g j t v u]	[ˌ ˀ ʃ u j]	⇐ [s] ⇒ [ɒ x]
[ə n]	⇒ [ʃ] ⇐ [a p t]	[ₒ]	⇐ [ʃ] ⇒ [s]
[ɛ ɪ l]	⇒ [ç] ⇐ [ₒ h m t v]	[ₒ ɔ f g z]	⇐ [ç] ⇒ [ˑ ɔ ɑ i]
[a ɔ p u]	⇒ [x] ⇐ [ˌ ₒ / n t]	[ₒ d n s]	⇐ [x] ⇒
[ɛ]	⇒ [χ] ⇐ [h j]		⇐ [χ] ⇒
[ː ç h n u ; ˈˈ ə / ←]	⇒ [h] ⇐ [ˈˈ ˜ / ː ˑ ; ˀ a h]	[g i v z]	⇐ [h] ⇒ [ɒ l v]
7. Pausen			
[ə ɒ ɑ h m o]	⇒ [ˈˈ] ⇐ [æ ʔ h j ; m n v]	[ˀ ɪ e ˑ ; ˈˈ ₒ ˀ]	⇐ [ˈˈ] ⇒ [ˈˈ / ɔ i o a]
[ʔ ɔ h s x]	⇒ [/] ⇐ [h ₒ ˀ d h p]	[h ʁ ˑ b ; ˈˈ . ₒ]	⇐ [/] ⇒ [ʁ ɔ ɑ i u]
8. Sonderzeichen			
[i n χ / ; ɒ s ə]	⇒ [h] ⇐ [g d ₒ k p]	[ʔ d]	⇐ [h] ⇒ [ʔ / v ə a]
[h]	⇒ [˜] ⇐ [ɑ]		⇐ [˜] ⇒

Tabelle 11 (4. Fortsetzung)

vor * steht bevorzugt:	*	nach * folgt bevorzugt:	vor * ist unzulässig:	*	nach * ist unzulässig:
[ʁ d h j m z] ⇒		[ː] ⇐ [ɑ e h i m o s u y]		⇐ [ː] ⇒	[ɔ b z]
[ʰ j l t z g b d ʁ] ⇒		[ˑ] ⇐ [ɑ e f i o s u y]	[ₒ ˑ ç] ⇐	[ˑ] ⇒	[ˈˈæ,ₒˑtv ʔɔbgɪmz]
[æ ø b f g x] ⇒		[.] ⇐ [n m ŋ]	[ₒ ˑ ɪ] ⇐	[.] ⇒	[ə ₒ / ɪ e s t]
⇒		[←] ⇐ [h]	⇐	[←] ⇒	
[j v] ⇒		[°] ⇐ [ɑ a]	⇐	[°] ⇒	
[ʰ / ɒ ç ɔ ʃ ɑ ɪ n s x] ⇒		[ₒ] ⇐ [b d g v z]	[ₒ . ˑ o] ⇐	[ₒ] ⇒	[ˈˈ ,ₒˑʁə/ çɛɪɔʃɑɑfɪ jklmot]

In der Tabelle 12 werden die schwäbisch-alemannisch und berlinisch beeinflußte Aussprache untereinander und mit dem beiden gemeinsamen Klassifikationsfeld verglichen; jedes Symbol, das zur bevorzugten oder unzulässigen Kombination gehört, wird einzeln aufgeführt. Um einen globaleren Überblick zu haben, wird in Tabelle 13 die statistische Relation von Lautgruppen angegeben. Die Lautgruppen sind so zusammengestellt und summiert worden,daß ein Vergleich mit der Klassifikation von ALTMANN, LEHFELDT (198o,316ff.) des Materials von LINDNER (1975) möglich ist.

Das Material von LINDNER bestand in Schrifttexten, so daß mit ihm der Unterschied zur Umgangssprache herausgestellt werden kann. Die vorliegenden Transkriptionen wurden auf 36 Symbole reduziert; Sonderzeichen, Pausen, [ə],[ɒ],[ɾ] und [χ] wurden nicht berücksichtigt. Diphtonge sind im Gegensatz zur LINDNERschen Auflistung nicht gesondert ausgewertet worden, sie sind in den Vokalkombinationen enthalten.

In 2o von 64 möglichen Kombinationen weicht die Transkription der Umgangsaussprache von der schriftsprachlichen Kombinatorik ab.

- Kurze Vokale unterscheiden sich nicht vom normalsprachlichen distributionellen Gebrauch; lange Vokale gehen bevorzugt(⇔ aktuell) Nasalen voran und aktuell (⇔ bevorzugt) stimmlosen Verschlußlauten.
- Stimmhafte Verschlußlaute können mit sich selbst und mit stimmlosen Frikativen verbunden werden, wenngleich nur marginal (⇔ unzulässig), worin sich das in der Umgangssprache gewöhnliche Auslassen von Zwischenvokalen zeigt. Mit stimmhaften Frikativen sind sie aktuell (⇔ marginal), mit Liquiden bevorzugt (⇔ marginal) gekoppelt; in beiden Fällen ein Anstieg der Kombinationsmöglichkeit.
- Stimmlose Verschlußlaute stehen nur aktuell (⇔ bevorzugt) vor langen oder kurzen Vokalen; sie verbinden sich bevorzugt mit Liquiden (⇔ marginal), aber nur marginal mit stimmhaften Reibelauten (⇔ aktuell).
- Die Kombination der stimmhaften Reibelaute weicht stark vom Schrifttext ab; sie treten vor stimmhaften Verschlußlauten, stimmlosen Verschlußlauten, stimmlosen Reibelauten, Nasalen und Liquiden marginal auf (⇔ unzulässig in allen Fällen).
- Stimmlose Frikative traten weniger häufig vor langen Vokalen (aktuell ⇔ bevorzugt), aber bevorzugt vor stimmhaften Frikativen

Tabelle 12: Vergleich der Klassifikationen P und I gemäß Tabellen 1o.1 bis 1o.3
(sa = schwäbisch-alemannisch; be = berlinisch; ug = sa und be)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
1. Vokale			
i ug ː·bdfv sa -·bd-v be ː·bdfv	ʰɑfɾlyz -----zɒbg ʰ-flyz	çɛɑaɔoəɒø‖ ---ə--ə-- -ₒ/ ---a--əɒ- -ₒ/ln	e - -n
ɪ ug ʁɾflmnsvz sa ʁɾflmn--z be ʁɾ--m-svz	ŋçlns -ç--s ŋçlnstɒ	·ₒij ~ₒ-jɪʰ -ₒ----he/ɒ	‖ə·ɛiɔɑeoz -ə----ɑeo-ɪʔ -ə- ---e----a
ɛ ug ː·ed sa -·ɛ-aχ be ː·-da-g	ɒglmns ɒglmns ɒ-lmn-	ɔbfiɪɛ· ɔ-f-ɪ--hoₒç‖ ----ɪ------ -	‖ -əɑhv -----ɪ
ɾ ug ʁɾdjv sa -ʔdjv be ---j-k	ʁɒçχlh -ɒçχl-n ----l--?	ₒᵗ ₒ-a ---	ɑeiv ----? -----
ə ug gtvzfdŋʔ sa gtv-fdŋɾbmn be gt---d-ɾb--	ʃʰʁkɑfhszˈ‖ ʃʰʁ--fhszˈ‖ₒɒ ʃ-ʁ------z --ɒ	·ₒɑɪ -ₒ-ɪʰˌɒe -ₒ-ɪ--ɒ-aˈ‖	ɑuo ɑuoɑi --o-i
ɐ ug ʁː·gᵊ˜jmvz sa -ː·gᵊ-j--z be ʁɪ·--˜j---h	‖ₒdbtgl -ˉ---glŋʰ ‖ᵊ--t----	ɒɛɪɔₒçɑf --ɪ-ₒç--eə ----------asˌ/	ɑi --· ---
a ug ʁɾʃdfhklmvᵊ sa -ɾʃdfhk-mvᵊj be -ɾ-d-h--mv--	ɔelmsx ɔel-sxn ɔel-sx-	ʰaɔɑ ʰaɔɷʁ -aɔ--ₒː·ɪ	ɑikou -ikouaɛɷgz ɑi-o-a----əhˈ‖
o ug ː·ls sa ː·-sjz be ː·-s-z	‖ɾfgnv ‖---n-xₒ ‖-f-n---l	‖əₒaɪɔ -əₒaɪ-ç -əₒa---m	ɔi --·ə ----

Tabelle 12 (1. Fortsetzung)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
ɔ ug aʒfk sa a-fkʁ be aʒfkʁvz	ₒøbfmltx ₒøb-m--xgn ₒ-bf-l-x--/	‖ː·ₒɪot ---ₒ--t -------	açɑɾi a--e- a----
u ug ː·dst sa -·dstz be ː·d-t-	ɒfhxz ɒ-hxzl ɒ-----g	ə/am ə-a- ----	s - -
ɷ ug ʁɾjmt sa -ɾjm-n be ---mt-	ŋmns ŋmn-t ŋ-n--	əₒ/ əₒ-a ----	ʔa -aₒd ----
ø ug ɔ sa ɔ be ɔ	ˌntz --t- ˌn--		
æ ug ‖kɾ sa ‖-ʔ -k-b	ŋˌɒn ŋˌ-- --ɒn		
y ug ː·ʔfi sa --ʔf-ty be ː·-fi--	ɒv ɒvy -v-b		
Y ug f sa f be -	sz -z s-ʁ		
ɒ ug œɛeiuy sa -ɛeiuy be œ-e-u-əɪ	‖ʰₒ/vfkg -ʰₒ----gt ---/v----h	ₒʔçhlˌmnst ₒ------ns- ₒ------ns-/	ŋɑi ---ə --iəɪ

Tabelle 12 (2.Fortsetzung)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
2. Klusile			
b ug ɒɔɳ sa ɒɔ-ʁʁi be ɒɔ----y	ə.·il --·ilʁ ə--il-œm	:·ʔv --ʔ-dt :-----	øə -- --
d ug ʰɒ/ln sa --/lnɒ be -ɒ/-n-s	ə:·aɛeiu ə-·-ɛ-ɪus ə:·a-eiu-	k kʁ·ɑ ----.	ʰtx ʰt-bk -t---f
g ug ɑeosɴŋʰɒ sa ɑe--ɒ-ʰɒi be -----ŋ-ɒ-u	ŋəʁlɑ.· ŋə-lɑ.· -əʁl---e	·ʔv ·?-a ----	çhv --v ---
p ug /ʃ sa -ʃ be /ʃ	ʁv -- ʁ-x		
t ug ɑlmsxʰøçʃ sa -l-sxʰøçʃŋɒn be ɑl---ʰ-çʃ---ɪ/	əʁ·ɑs əʁ--sy ʁ--ɑs-.uç	·ʔdjɒ.ʁ ·ʔd-ɒ--kv ·-d------	ɒɔʒ -ɔ-b ----m
k ug ʰəɒ sa ʰ--χ be ʰ---	ʁœaɔ ʁœaɔ -œ-ɔɛ	ɒʔa ---dm -----	d dt --
ʔ ug hnosʔ//ll sa -n-sʔ//llʁ be hn---/ --ɛ	əʁ/ʔaɛiɑy ə-/ʔaɛ-ɑyœɪ ə---aɛ----ɪ	ɪɑʒj·hɒ.ɛ ɪ--j·hɒ.ɛ --------ɒ-m:	llhɒbfvgklstʒ llh-bfvgklst- ---------s--ɒ

Tabelle 12 (3.Fortsetzung)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
3. Nasale			
m ug ɔɷemç:·ll sa ɔɷe-ç:-ll be --em-:-ll.	ll:aɪɷɑm ll-aɪɷ--ʃə ll-aɪɷ----	ɒ·v ɒ·v ɒ·-tz	ʁɴu ---k ----ʔo
n ug ɪɷeoxʔɛəø.ɒll sa -ɷeo--ɛa-.-llɔ be ɪ-eox---ø.œll-h	ʃdhtʰɒʔɪ ʃdhtʰɒʔɪəɷ -dh-ɒ-ʔ---	- j -i	ŋɒx -ɒx -ɒ-i
ŋ ug ʁ.gɪɷ sa ɴ.g-ɷɑ be -.-ɪ--x	əvg əv-t. --g--	n - -	
4. Lateral			
l ug ɑbegiɔɪɛa sa ɑb-g---ɛau be -begiɔɪ.a-o	ɒ·açɪodstz ɒ·-çɪ-d--zu -------stz-	·ɒʔv -ɒʔ-/ -----	.ɒ -- --i
5. Vibranten			
ʁ ug - sa - be ʃ	- b -		
ʁ ug əɪgkpt sa ə--k-tbf be ə-g-p---Y	:·ɛɪɷɑ -·ɛɪ-ɑɔ :·ɛɪ-ɑɔ	ɒ/m ɒ-- ---	øtv ---d ----

Tabelle 12 (4. Fortsetzung)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
6. Konstriktive			
v ug çosyŋ$_{\circ}$ɒ^{ll} sa çosyŋ$_{\circ}$-^{ll} be ç--y-$_{\circ}$ɒ-	$^{\circ}$ɪɑɪɛəə $^{\circ}$i--ɛəə -ɪ-ɪ-ə-ɔ	ɛghjz^{h}ʁ$_{,}$$^{\cdot}$ʔ -gh--^{h}---$^{\cdot}$ʔe ----------:	^{ll}dghlm -----mu -------$_{\circ}$t
z ug Yiɪuə$_{\circ}$ø sa Yiɪuə-- be -iɪ-ə$_{\circ}$-	ə:$^{\cdot}$ɪɑu ə:$^{\cdot}$ɪɑuo --$^{\cdot}$ɪ--oɔ	^{ll}:$^{\cdot}$ʔɪt ------as -:------	ʔçhv ----s ----sm
ʒ ug sa be	ɔ - ɔ		
ʃ ug ^{ll}χs sa ^{ll}-sn be ^{ll}---	:$^{\cdot}$ɛɷɑ$^{\circ}$ $^{\cdot}$$^{\cdot}$ɛɷɑ$^{\circ}$ao :$^{\cdot}$ɛ-ɑ	$^{\circ}$ -	ʔɪstv ʔ-s--n ------
f ug ɔɷiuɒ$^{\cdot}$əç sa ɔ-----$^{\cdot}$ə-ʁ be ɔ-i--$^{\cdot}$-ç-o	ə$_{,}$acY-yɪʁ ---ɔ-iy--	$^{\circ}$ʔ $^{\circ}$ʔ --d	çɑe --e ---
s ug ɷɔYeltɪa$^{\cdot}$:ə sa ɷ--e-tɪa$^{\cdot}$:əd be --Y-ltɪa----	^{h}$_{\circ}$/ʔɪgjtvu -$_{\circ}$/ʔ--jtvuo^{ll} ^{h}$_{\circ}$/-ɪ-----o -d	$_{,}$ʔʃuj $_{,}$ʔʃ-jzʁ$_{\circ}$ -ʔ---z--v	ɒx ɒxz ɒ--a
ʃ ug ən sa ənm be ə--	apt ap-u$_{\circ}$ -pt--ʁ	$^{\circ}$ $^{\circ}$	s s -

Tabelle 12 (5.Fortsetzung)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
ç ug ɛɪl sa ɛɪlʁ be ɛɪ--t	$^{\circ}$hmtv $^{\circ}$hmtv $^{\circ}$--tv$_{,}$	$_{\circ}$ɔfgz $^{\circ}$----- -----	$^{\cdot}$ɒɑi $^{\cdot}$-ɑ-eo ------
x ug aɔpu sa aɔ-u be aɔp-	$_{,}$$^{\cdot}$/nt --/-t $_{,}$----o	$_{\circ}$dns --ns ----	
χ uŋ ɛ sa ɛ be -	^{h}j --el ----		
h ug :çhnuə/^{ll}← sa :ç-nuə- -- be :-hn---^{ll}-	^{ll}~/:$^{\cdot}$ʔah ^{ll}-/:$^{\cdot}$-a- ^{ll}~/:-ʔahɑn	givz ----e -----$_{\circ}$a	ɒlv --ve ----$_{\circ}$ɬ
7. Pausen			
^{ll}ug əɒɑhmo sa ə--hmos be --ɑhmo-	œʔhjmnv œʔ-jmnvu --hjmn--	vɪe$^{\cdot}^{ll}_{\circ}$ʔ ---- -$_{\circ}$ʔ ---- -$_{\circ}$-a	^{ll}/ɔioa -/-----e --------ə
/ ug ʔɒhsx sa ʔ-hsx be -ɒhs-	^{h}$_{\circ}$ʔdhp ^{h}$_{\circ}$ʔd-- -$_{\circ}$ʔd-pt	^{h}ʁ$^{\cdot}$b^{ll} ---$^{\cdot}$-^{ll}-$_{\circ}$ ---- --$_{\circ}$:	ʁɒɑiu ---i-ɪ -ɒ-i--ɪɑ
8. Sonderzeichen			
^{h} ug inχ/ɒsə sa -n-/ɒ-əɑ be i---ɒs--	gd$_{\circ}$kp g-$_{\circ}$k-t ---k-t	ʔd ʔd --	ʔ/vəa ʔ-və-ɪ ------

Tabelle 12 (6. Fortsetzung)

* bevorzugt vor *	nach *	unzulässig vor *	nach *
ɔ ug h be h sa -z	ɑ ɑ -	-	ɔbz -bzvaɾ·/ --------ɹ
ɜ ug ʁdhjmz be ʁdhj-- sa --hj-z	ɑehimosuy ɑehi--uyln ɑ-h-mos---n	·ç -·: ·ç-ɑo	ˈˈæ.ₒ·tvʔɔbglmz ----------m-a ---ₒ·tvɾ----m--d
· ug hjltzgbdʁ be -j--z--dʁv sa -jl-z-bd--gh	ɑefiosuy ɑefio-uy ɑ-fiosu-g	·ɪ -·: · --	aₒ/ʔɪestv ----------d eₒ-ɾ--s--
ɐ ug ɸæbfgx be ɸ----xçt sa -œ-fg---ŋ	nmŋ nmŋ nmŋ		
ɬ ug be sa	h - -		
ɷ ug jv be -- sa jv	ɑa -- ɑa		
ɷ ug h/ɔçɔʃɑlnsx be -/-çɔ-ɑ-nsx sa h/ɒçɔʃɑlns-ə	bdgvz bdgvz bd-v-	·ɑ -:--?hv ·ɑ---ʁ	·ʁə/ʔnçɛɪɔʃɑɑfix -·ə/ʔɒç-ɪ-ɑ-i-anmot ·ʁə/ʔnçɛɪɔʃɑɑfi--nmotsekl

Tabelle 13: Häufigkeitsmatrix und statistische Klassifikationen der nach Lautgruppen zusammengefaßten Transkriptionssymbole für schwäbisch-alemannisch und berlinisch beeinflußte Umgangsaussprache (reduziertes K = 36)

	kurze Vok	lange Vok	stih Klu	stil Klu	stih Kon	stil Kon	Nasal	Liqui	Σ
kuVo	79	126	81	33	35	57o	229	99	1252
laVo	74	23	119	39	58	116	147	51	627
shKl	146	143	2o	4	29	82	61	37	522
slKl	65	43	8	5	1o	1oo	14	19	264
shKo	159	126	1o	3	3	9	38	1	349
slKo	255	154	163	1o3	85	138	148	27	1o73
Na	192	66	129	61	59	127	85	16	735
Li	89	49	34	14	23	39	17	2	267
Σ	1o59	73o	564	262	3o2	1181	739	252	5o89

kuVo	M *M*	M *M*	M *M*	M *M*	M *M*	P *P*	P *P*	P *P*
laVo	M *M*	M *M*	P *P*	A *P*⇐	P *P*	M *M*	P *A*⇐	P *P*
shKl	P *P*	P *P*	M *I*⇐	M *M*	A *M*⇐	M *I*⇐	M *M*	P *M*⇐
slKl	A *P*⇐	A *P*⇐	M *M*	M *M*	M *A*⇐	P *P*	M *M*	P *M*⇐
shKo	P *P*	P *P*	M *I*⇐	M *I*⇐	M *M*	M *I*⇐	M *I*⇐	M *I*⇐
slKo	P *P*	A *P*⇐	P *P*	P *P*	P *A*⇐	M *M*	A *M*⇐	M *M*
Na	P *P*	M *M*	P *P*	P *P*	P *P*	M *M*	M *M*	M *M*
Li	P *P*	P *P*	A *P*⇐	A *P*⇐	P *P*	M *A*⇐	M *M*	M *M*

(kursiv: Klassifikation des schrifttextlichen Materials von LINDNER 1975 durch ALTMANN,LEHFELDT 198o, 316)

Abkürzungen:

kuVo	kurze Vokale
laVo	lange Vokale
shKl	stimmhafte Klusile
slKl	stimmlose Klusile
shKo	stimmhafte Konstriktive
slKo	stimmlose Konstriktive
Na	Nasale
Li	Liquide

(⇔ aktuell) und aktuell vor Nasalen (⇔ marginal) auf.
- Nasale als Vorgänger aller anderen Lautgruppen weichen in der Häufigkeit nicht vom Schriftstandard ab.
- Liquide werden weniger oft mit stimmhaften und stimmlosen Verschlußlauten (aktuell ⇔ bevorzugt) und stimmlosen Reibelauten (marginal ⇔ aktuell) nachfolgend verbunden.

Insgesamt nehmen die Kombinationen stimmhafter Konsonanten mit anderen Lautgruppen zu, was man in der älteren Literatur als "Vorherrschen des stimmlichen Elementes" bezeichnet hätte. Der phonatorisch-artikulatorische Impetus ist mehr legato als staccato, eher fließend als abrupt; der Artikulationsstil ist verbindlicher, "intimer" und vermeidet "hart" klingende Verbindungen. So werden stimmlose Verschlußlaute (bis auf die Kopplung mit den "fließenden" Lauten, d.h. Liquiden) weniger oft kombiniert; die Liquide selbst werden ebenfalls reduziert (Verschleifungen der R- und L-Laute) und der normalsprachliche Gebrauch von stimmlosen Frikativen zugunsten stimmhafter Frikative und Nasale herabgesetzt.

2.3 INDIVIDUELLE ARTIKULATIONSEIGENHEITEN ZWEIER SPRECHER WÄHREND EINES DYADISCHEN GESPRÄCHS IM COLLOQUIAL STYLE

Mit der Auswertung phonetischer Transkriptionen ist - im Kontrast zu rein phonologischen Statistiken - die Möglichkeit gegeben, auch individuelle, paralinguistische Artikulationsbesonderheiten statistisch zu beschreiben. Tabelle 12 enthält bereits einige solcher Unterschiede, seien sie durch dialektale, umgangssprachliche oder individuelle Komponenten bedingt; jedoch sind nur die statistisch bevorzugten oder unzulässigen Distributionen aufgeführt. ALTMANN, LEHFELDT (1980, 266ff.) schlagen einige Verfahren zum statistischen Vergleich verschiedener Sprachen vor, die - in Auswahl - auf einen Vergleich zweier Sprecher angewendet werden sollen.

Sprecherspezifische Artikulationsstile machen einen Teil der paralinguistischen Informationen aus, die dem Hörer den typischen "sound" eines Sprechers vermitteln. Der Höreindruck, soweit er sich an der Aussprache festmachen läßt, kann mit dem IPA-System notiert werden; der Satz von POYATOS (1975), daß alles paralinguistisch

genannt wird, was nicht mit IPA-Symbolen transkribierbar ist (295/296), ist nur ein ironisches Aperçu (zur Notation paraphonetischer features s. WINKLER 1979 und 1982).

Die statistische Analyse soll auf solche Symbole beschränkt werden, die der Transkribent möglicherweise paralinguistisch gemeint hat. Die Selektion ist ein interpretativer Eingriff, der m.E. berechtigt ist, weil ihm eine vollständige statistische Beschreibung vorangegangen ist. In den Tabellen 5 und 7.2/7.3 sind eine Reihe von Symbolen durch die Assoziativitäts-Parameter unterschieden; dort sind sie als Indiz für dialektale und paraphonetische Beschreibungen gewertet worden. Aus diesen klassifikatorisch verschiedenen Symbolen (das sind: [χ ʒ ← ʁ k ɷ ʃ n ʰ ʁ j o z ˈ j g ˌ v ˌ f ɒ t ç m a ə ː s]), den nicht unterschiedlich verwendeten Zeichen [/ ‖ ~ °] und dem Glottiseinsatz werden anhand des folgenden Fragekataloges 16 Symbole ausgewählt:
Sind unterschiedliche Häufigkeiten und Kombinationen von Symbolen entstanden, mit denen der Transkribent die Sprecher in bezug auf
- den Gebrauch des Glottiseinsatzes [ʔ],
- die Behauchung [ʰ],
- die Entstimmlichungstendenz [ˌ],
- Rundungen und Nasalierungen von Vokalen [° ~],
- die silbische Realisierung [ˌ],
- die Länge von Lauten [ː ˑ],
- inverse Atmung [←],
- den Gebrauch des schwachtonigen E-Lautes [ə],
- die Allophone des R-Lautes bzw. dessen Ersatzrealisation [ʁʁɒχ],
- Stauungen und Pausen [/ ‖]

charakterisiert hat?

Möglicherweise überlagern sich dialektaler und paralinguistischer Gebrauch von [~ ° χ ɒ ʰ], der statistisch nicht zu trennen ist. Zumindest gehen die Verwendungseigenschaften von [ə ˌ ː ˑ], vielleicht auch von [ʰ], eindeutiger auf den colloquial style des Gesprächs zurück.

Zur Unterscheidung von den vorangehenden Tabellierungen werden die Sprecher mit "SIE" (vorher: schwäbisch-alemannische Aussprache) und "ER" (vorher: berlinische Aussprache)gekennzeichnet.

Die statistischen Parameter werden nunmehr mit reduziertem K (=16) berechnet, die Kombinatorik jedoch mit K = 50, wobei die Symbole zu Lautgruppen zusammengefaßt sind (ähnlich Tab. 13).

Tabelle 14: Häufigkeitsparameter für 16 Transkriptionssymbole, Divergenz und Korrelation zweier Sprecher

Symbol	SIE f_i	$\hat{p}_i$	ER f_i	$\hat{p}_i$
ʁ	-	-	4	.oo439
←	-	-	2	.oo219
ɻ	2	.oo128	-	-
~	3	.oo193	2	.oo219
ə	12	.oo77o	-	-
ɛ	6o	.o3851	72	.o7912
ɐ	63	.o4o44	25	.o2747
.	75	.o4814	52	.o5714
||	82	.o5263	76	.o8352
ɦ	88	.o5648	21	.o23o8
:	1oo	.o6418	134	.14725
/	126	.o8o87	1o9	.11978
ˈ	154	.o9884	77	.o8462
ɾ	159	.1o2o5	63	.o6923
ə	258	.16559	1o9	.11978
ˌdev	376	.24134	164	.18o22
N	1558		91o	
Σf^2	31o632		93266	
χ^2	1632.o59		729.842	
R	.12797		.11263	
D	.o698		.o535	
V	1.o235		.8956	
H_o	4		4	
H_1	8.2444		8.4215	
h	.8111		.8554	
Re	.1889		.1446	
I_L	.1o55		.o933	

r = .838

D = .1399

χ^2 = 139.85 (G = 145.58)

Tabelle 14 enthält die Häufigkeitsangaben für beide Sprecher; die Verwendung der 16 Symbole korreliert zwischen beiden Stichproben recht hoch (r = .838; Signifikanz t = 5.75 bei Tafelwert $t_{.o5;14}$ = 2.14). So ist die Divergenz mit .14 nicht sehr groß (D kann sich zwischen o und 1.41 bewegen; Berechung aus $\sqrt{\Sigma(\hat{p}x_i - \hat{p}y_i)^2}$). Die Stichproben sind jedoch signifikant inhomogen.

In den distributionellen Eigenschaften unterscheiden sich beide Symbolverwendungen (gemäß Tabellen 6.2 und 6.3):

Tabelle 15: Differenzen der Distributionsmaße beider Sprecher

Sy	At	Ag	As	As*	At_i	Ag_i	S	S_i	S_i*	S'	S_i'	S_i'*	$\sqrt{\Sigma(a-b)^2}$
h	.31	.o1	.31	.15	.16	.24	.o2	.o9	.14	.32	.o9	.14	.68
ˌdv	.46	.64	.14	.o9	.72	.74	.o4	.o2	.o3	.1o	.o2	.o2	1.31
ə	.o9	.21	.28	.15	.31	.77	.o2	.o8	.14	.26	.92	.86	1.59
~	.o7	.o4	.11	.o1	.o7	.o7	.oo	.oo	.oo	.11	.oo	.oo	.2o
.	.2o	.6o	.18	.2o	.5o	.79	.16	.o6	.42	.o3	.28	.42	1.34
:	.27	.o9	.o4	.1o	.29	.17	.13	.o4	.11	.1o	.14	.11	.54
ˈ	.43	.17	.2o	.14	.41	.4o	.o7	.11	.oo	.13	.oo	.oo	.8o
||	.o9	.o7	.o1	.o2	.14	.1o	.o2	.o4	.o5	.o4	.o4	.o5	.23
←	.o4	.o4	.o4	.o4	1.	1.	.o4	1.	1.	.oo	.oo	.oo	2.oo
ə	.19	.o4	.13	.11	.1o	.o8	.1o	.o2	.o1	.o3	.o2	.o1	.31
ɐ	.18	.3o	.28	.24	.o1	.21	.2o	.21	.27	.1o	.21	.27	.77
ʁ	.1o	.1o	.19	.11	.56	.56	.o2	.11	.2	.17	.89	.8o	1.49
ɛ	.15	.27	.oo	.o6	.2o	.34	.1o	.13	.17	.1o	.13	.17	.61
/	.12	.o8	.o1	.o2	.15	.o9	.o5	.o7	.o6	.o6	.o7	.o6	.27
χ	.o7	.o4	.11	.o5	.6o	.4o	.oo	.oo	.oo	.11	1.	1.	1.6o
ɾ	.13	.28	.31	.21	.12	.o7	.11	.o5	.o5	.2o	.o5	.o5	.56

D = 4.22

Die größten Unterschiede zwischen beiden Sprechern bestehen in der Verwendung der ingressiven Lautung, im Ersatzes des R-Lautes durch einen Reibelaut, in Vokalrundungen, im Gebrauch des uvularen R-Lautes, der Silbigkeit und der Entstimmlichung. Am wenigsten differiert die Distribution bei der Nasalierung, den Pausen und Staupausen. Eine mittlere Distanz nimmt die Kombinatorik des schwachtonigen E-Lautes, des Glottiseinsatzes, der

Lautlänge, der vokalischen und konstriktiven (stimmhafter uvularer Reibelaut) Ersatzlautung des /R/ und der Aspiration ein. [ʁ ̥ χ] und [ɹ̝] werden jeweils nur von einem Sprecher gesprochen.

Die silbige Aussprache wird vom Sprecher weniger symmetrisch kombiniert als von der Sprecherin; die Assoziativität ist dagegen bei der Sprecherin höher. Der Sprecher kombiniert die Silbigkeit vor allem attraktiv, die Sprecherin stärker aggressiv als attraktiv. Demgemäß ist die interne Aggressivität der silbigen Verbindungen bei der Sprecherin höher, die interne Attraktivität niedriger als beim Sprecher.

Entstimmlichte Laute werden vom Sprecher mehr attraktiv als aggressiv kombiniert; konträr zur Verwendung durch die Sprecherin. Die Symmetrie ist bei beiden gleich hoch (anders als die qualitative und quantitative Symmetrie, die später noch dargestellt wird).

Schwachtoniges /E/ ist in Symmetrie und Assoziativität zwischen den Sprechern um nur .1bzw. .13 unterschiedlich; es wird von der Sprecherin eher aggressiv kombiniert.

Anders ist die Relation beim Glottisschlag:Die Symmetrie unterscheidet sich um .11, die Assoziativität um .21; die Sprecherin verwendet den Stimmeinsatz eher aggressiv-kombinatorisch, der Sprecher gleichermaßen attraktiv als auch aggressiv.

Von beiden Sprechern wird die Behauchung symmetrisch kombiniert (vgl. jedoch die qualitative Symmetrie). Der Grad der Assoziativität ist um .31 unterschiedlich. Die Sprecherin verwendet vorwiegend attraktive Kopplungen, aggressive Kombinationen treten bei beiden Sprechern gleich häufig auf.

Die Bewertung der Symmetrie bezog sich bisher auf die Relation eines Symbols i zum Symbol j, d.h. wenn die beiden Paare ij und ji im Material auftreten. Wird ein Symbol i danach klassifiziert, wieviel gemeinsame Elemente in Ai und iB existieren, erhält man eine qualitativ-symmetrische Einordnung (ALTMANN,LEHFELDT 198o, 266ff.). Die quantitativ-symmetrische Klassifikation geht aus den Differenzen der Vorgänger- und Nachfolgermenge zum Symbol i hervor. Die Artikulationstendenzen der beiden Sprecher wiesen bestimmte symmetrische Kombinationen auf, die aber nichts aussagen über die generelle Behauchung vieler Laute (d.h. etwa ein habitueller Trend) oder über Konzentrationen

auf bestimmte Laute, wie es tendentiell in der deutschen Standardaussprache der Fall ist. Die individuelle Behandlung der Kombinatorik kann durch die Zahl der Vorgänger- und Nachfolgersymbole und der in den Mengen enthaltenen Lauttypen bestimmt werden.

Qualitative und quantitative Symmetrie wurden nach folgenden Algorithmen bestimmt:

AS (qualitativ-asymmetrisch):
wenn $[Ai \cap iB] < E(Y)$ und $P(Y \leq [Ai \cap iB]) \leq \alpha$

SS (qualitativ-semisymmetrisch):
wenn $[Ai \cap iB] < E(Y)$ und $P(Y \leq [Ai \cap iB]) > \alpha$ oder
wenn $[Ai \cap iB] > E(Y)$ und $P(Y \geq [Ai \cap iB]) > \alpha$

S (qualitativ-symmetrisch)
wenn $[Ai \cap iB] > E(Y)$ und $P(Y \geq [Ai \cap iB]) \leq \alpha$

$$E(Y) = (([iB]-G)([Ai]-G))/(K-1)$$

$$P(Y \leq [Ai \cap iB]) = \sum_{y_{min}}^{[Ai \cap iB]} \left[\binom{[Ai]-G}{y-G} \binom{K-1-[Ai]+G}{[iB]-y} \right] / \binom{K-1}{[iB]-G}$$

$$y_{min} = [Ai]+[iB] \text{ wenn } [Ai]+[iB] > K$$
$$y_{min} = G \text{ wenn } [Ai]+[iB] \leq K$$

$$P(Y > [Ai \cap iB]) = \sum_{y=[Ai \cap iB]}^{\min([Ai],[iB])} \left[\binom{[Ai]-G}{y-G} \binom{K-1-[Ai]+G}{[iB]-y} \right] / \binom{K-1}{[iB]-G}$$

Die Klassifikation in die Kategorien quantitativ-asymmetrisch (QAS) oder quantitativ-symmetrisch (QS) erfolgte anhand der kritischen Werte für u_α (Tabelle in ALTMANN,LEHFELDT 198o, 276).

Tabelle 16: Klassifikationen der qualitativen und quantitativen Symmetrie

	SIE		ER	
h	AS	QAS	SS	QS
ɐ	SS	QAS	S	QAS
ɔ	SS	QS	-	-
~	SS	QS	SS	QS
ˌ	SS	QAS	SS	QAS
ː	S	QS	SS	QS
ˑ	SS	QAS	SS	QAS
ɾ	-	-	SS	QS
ə	SS	QS	S	QS
ʁ	-	-	SS	QS
ʁ̥	SS	QS	SS	QS
ɒ	SS	QS	SS	QAS
χ	SS	QS	-	-
ʔ	SS	QS	S	QS
ˈ	SS	QS	SS	QS
/	S	QS	SS	QS

Nicht sprecherspezifisch werden Nasalierungen, Silbigkeit, Halblänge, Pausen und [ʁ] kombiniert (vgl. dagegen die distributionelle Frequenz).

Allein die Aspiration tritt asymmetrisch auf (Sprecherin; die qualitative Asymmetrie ist sehr selten, wie ALTMANN,LEHFELDT 1980, 271 bei der Analyse von 25 Sprachen festgestellt haben). Der Sprecher behaucht globaler (qualitativ semisymmetrisch, quantitativ symmetrisch) als die Sprecherin.

Die Entstimmlichungstendenz ist (anders als bei der Assoziativität) beim Sprecher qualitativ symmetrischer als bei der Sprecherin. Auch der Glottiseinsatz wird vom Sprecher qualitativ symmetrischer verwendet.

Stauungen werden von der Sprecherin qualitativ symmetrisch gesprochen, der sprecher kombiniert sie semisymmetrisch. In quantitativer Hinsicht treten sie bei beiden Sprechern symmetrisch auf.

Das schwachtonige /E/ wird vom Sprecher qualitativ symmetrisch und von der Sprecherin semisymmetrisch kombiniert; ein Unterschied, der in der Assoziativität (und deren Spezialfall Symmetrie) nicht deutlich wurde.

Die vokalische Ersatzlautung des /R/ wird qualitativ von beiden Sprechern gleich gekoppelt; quantativ ist beim Sprecher eine Asymmetrie festzustellen, die sicher dialektal bedingt ist.

In der distributionellen Frequenz sind die Sprecher typisch voneiander verschieden.

In Tabelle 17 sind die Kombinationen der Symbole der linken Spalte als Nachfolger der 16 Symbole aufgelistet. Die Kombination der Sonderzeichen ist in diesem Fall eingeschränkt aussagekräftig, weil sie vor allem die Einhaltung der Reihenfolge Sonderzeichen-zugehöriges Symbol dokumentiert; für einen Koder-Reliabilitätstest wären diese statistischen Berechnungen gerade sinnvoll.

In Tabelle 18 ist die Kombinatorik der Symbole der linken Spalte als Vorgänger aufgeführt.

Der Sprecher aspiriert häufiger und mehr Laute als die Sprecherin: offene Vokale, stimmhafte Konstriktive, Stauungen (aktuell ⇔ unzulässig bei der Sprecherin); auch geschlossene Vokale und Nasale (aktuell ⇔ marginal). Statistisch möglich (virtuell) wäre auch ein Glottiseinsatz nach[ʰ] (Sprecherin: unzulässig), der in der Standardaussprache nicht zu erwarten ist.

[ₒ] in der Kombination mit stimmlosen Verschluß- und Reibe-

Tabelle 17: Klassifikation der Nachfolgezeichen (1.: SIE, 2.: ER)

nach ->ʰ folgt↓	°	~ₒ	·	: ·	+	ə	ʁʁɒχ	?	11/
ofVo I A	M I	P V	I I	M M	V V	M A	P P	P P	P A
geVo M A	M M	P A	I I	P P	V V	M M	A M	A M	M M
shKl A A	P P	V V	M I	M M	V V	M M	M M	M M	A A
slKl P P	M I	V V	I V	M M	V V	A A	A A	I V	A P
shKo I A	P P	V V	I I	M M	V V	P A	A A	I I	P P
slKo M V	M M	A V	M I	P A	V A	P P	A A	M A	P P
Nas M A	M M	V V	P P	A M	V A	A M	M A	M M	P P
ʁʁɒχ A V	I I	V V	V V	A I	V V	P P	A V	V V	A A
? I V	I I	V V	I V	M I	V V	M A	A A	P A	P P
11/ I A	I I	V V	V V	M M	V V	P A	A P	P A	A A

ofVo: offene Vokale geVo: geschlossene Vokale shKl: stimmhafte Klusile slKl: stimmlose Klusile
shKo: stimmhafte Konstriktive slKo: stimmlose Konstriktive Nas: Nasale

lauten und Vokalen ist zwar "ungewöhnlich", kennzeichnet in seltenen Fällen jedoch ein koartikulatorisches "tonloses Sprechen", vor allem bei Vokalen ohne Phonation. Bei offenen Vokalen und stimmlosen Verschlußlauten des Sprechers sind diese Kombinationen nicht aufgetreten.

Vokale werden bevorzugt von der Sprecherin nasaliert und gerundet (Sprecher: aktuell bzw. virtuell). Stimmlose Klusile sind statistisch aktuell "gerundet oder nasaliert" notiert worden, was wiederum unüblich ist und als Versuch interpretiert werden kann, eine typische, koartikulatorische Überformung der Vokalnachfolger zu kennzeichnen.

Die Sprecherin realisiert im Gegensatz zum Sprecher auch stimmhafte Klusile und stimmlose Reibelaute silbig (⇔ unzulässig). Erwartbar (virtuell) sind sogar silbig verwendete Glottiseinsätze und stimmlose Klusile (Sprecher).

Längenunterschiede treten bei stimmlosen Konstriktiven (SIE: bevorzugt, ER: aktuell), Nasale (aktuell ⇔ marginal), R-Allophonen (aktuell ⇔ unzulässig) und Glottiseinsätzen (marginal ⇔ unzulässig) auf; die Sprecherin spricht in allen Fällen etwas "gedehnter".

Ingressive Lautbildungen treten bei stimmlosen Reibelauten und Nasalen auf (nur Sprecher).

Das schwachtonige /E/ wird im Unterschied zur Sprecherin vom Sprecher vor offenen Vokalen, Glottiseinsätzen (jeweils aktuell ⇔ marginal) häufiger, seltener vor stimmhaften Reibelauten (aktuell ⇔ bevorzugt), Nasalen (marginal ⇔ aktuell) und Pausen (aktuell ⇔ bevorzugt) gesprochen.

R-Allophone spricht ER bevorzugt vor Pausen (Dialekteinfluß), aktuell (⇔ marginal) vor stimmlosen Konsonanten, aber nur marginal (⇔ aktuell) vor geschlossenen Vokalen.

Geschlossene Vokale werden nur marginal mit [ʔ] begonnen (ER), (⇔ aktuell); stimmlose Konsonanten haben aktuell (⇔ marginal) "glottale Stauungen" vor dem eigentlichen Lautbeginn, sie wären auch vor stimmlosen Klusilen zu erwarten (virtuell).

Pausen werden von beiden Sprechern ähnlich plaziert; der Sprecher bevorzugt (⇔ aktuell) sie vor stimmlosen Klusilen und verwendet sie aktuell (⇔ bevorzugt) vor offenen Vokalen.

Tabelle 18: Klassifikationen der Vorgängersymbole (1.: SIE, 2.: ER)

vor → ʰ steht ↓	°	~°	.	: ˑ	+	ə	ʀʁɒχ	ʔ	ǁ/
ofVo	A A	P P	A V	A P	A A	V V	M M	P P	M A	M M
geVo	A A	A A	A V	A A	A A	V V	M M	P P	A A	A M
shKl	M V	M M	A V	P A	A P	V V	P P	A A	M M	M M
slKl	A A	M M	V V	A P	P A	V V	A P	A P	A V	A M
shKo	A V	A A	A P	A A	P P	V V	A A	A M	M M	M M
slKo	A P	P A	A V	A A	M A	V A	M M	M M	A P	P P
Nas	P I	A A	A V	A M	A A	V A	A A	M M	A A	A P
ʀʁɒχ	P A	A A	V V	A A	P A	V V	A A	A I	A A	A P
ʔ	I V	M V	A V	A V	A A	V V	P P	I V	P A	P A
ǁ/	A A	A P	A V	A I	M M	V V	M M	M M	P P	M M

ofVo: offene Vokale geVo: geschlossene Vokale shKl: stimmhafte Klusile slKl: stimmlose Klusile
shKo: stimmhafte Konstriktive slKo: stimmlose Konstriktive Nas: Nasale

Der Sprecher spricht keine Nasale vor einem behauchten Laut (⇔ bevorzugt), beginnt nach einer Pause oder Stauung nicht mit silbigen Realisationen (⇔ aktuell) und kombiniert keine R-Allophone miteinander (⇔ aktuell). Er spricht bevorzugt stimmhafte Reibelaute vor nasalierten Vokalen (⇔ aktuell), stimmlose Verschlußlaute vor silbiger Aussprache (⇔ aktuell), stimmhafte Klusile vor langen Lauten, schwachtoniges /E/ nach stimmlosen Klusilen (⇔ aktuell ; der einzige Unterschied im Gebrauch von [ə] zwischen den Sprechern), R-Allophone nach stimmlosen Verschlußlauten (⇔ aktuell), Glottiseinsätze nach stimmlosen Verschlußlauten und beginnt nach Pausen häufig mit Nasalen oder R-Allophonen (⇔ aktuell).

Die Sprecherin kombiniert bevorzugt behauchte Laute mit stimmlosen Klusilen und R-Allophonen (⇔ aktuell), Entstimmlichungen nach stimmlosen Reibelauten und Pausen (⇔ aktuell), silbige Aussprache nach stimmhaften Klusilen, Glottiseinsätze nach Pausen und mit weiteren Glottisschlägen (jeweils ⇔ aktuell).

3. Zusammenfassung

Die statistische Analyse von auditiven Segmentierungen, der Häufigkeit und Kombination sowie der Häufigkeiten von Kombinationen hat gezeigt, daß mit dem IPA-System Unterschiede zwischen Sprachen, Formstufen und individuellen Artikulationsstilen dokumentiert werden können. Die Transkriptionsregeln sind weit genug gefaßt, um Höreindrücke von lautsprachlichen Sequenzen objektivieren zu können. Die Segmentation von Lauten ist typisch für die phänomenale Verarbeitung des akustischen Signals; auditive Zeitquantelungen sind Ausdruck der apperzeptiven Zeitstruktur. Phonetische Erwartung und fachliches Vorwissen beeinflussen nicht übermäßig stark die Transkription.

Die Umgangssprache unterscheidet sich von der schriftsprachlichen Kombinatorik und Lauthäufigkeit durch einen Anstieg der Kombination stimmhafter Konsonanten mit anderen Lautgruppen und durch eine Abnahme von Kombinationen mit stimmlosen Konsonanten, vor allem mit stimmlosen Verschlußlauten. Liquide werden häufig reduziert und verschliffen ausgesprochen; Nasale und Vokale sind am wenigsten beeinflußt.

Sprechertypisch und individuell war der Gebrauch der Aspiration, der Entstimmlichung, der Glottiseinsätze, der Nasalierung und der Reduktionen unbetonter Endsilben. Pausengestaltung und die Silbigkeit von Lauten unterschieden sich nicht sprecherspezifisch.

Fußnoten:

1 Zwei Unterschiede zum IPA-Standard sind zu beachten: [ɷ] steht für [ʊ̯]; [ɒ] für [ɐ̯].

2 Herrn Wolfgang Brudler danke ich herzlich für die Auszählung und Tabellierung der Daten.

Anhang

Programm Distributionsanalyse (TI 59, PC 1oo C, ML-Modul)

Labels: oo1 E'
o13 A'
o19 A
o24 B
o29 C
o34 C'
o39 E

User defined labels and user instruction:

1 Distribution				2
ii?		K (1x)		Entry
Ai	iB	A∩B		Code Sy

Use order E' - C' (first)
(A' if ii)
A→ B→ C→ E (starts program)

Entry		Store K		(SD in Reg o3)		(SD)	
ooo	LBL	o33	LBL	o67	=	1o4	3
oo1	E'	o34	C'	o68	-	1o5	6
oo2	PGM	o35	STO	o69	(	1o6	1
oo3	01	o36	o7	o7o	2	1o7	6
oo4	SBR	o37	R/S	o71	x	1o8	OP
oo5	CLR			o72	RCL	1o9	o4
oo6	O	Store code/		o73	o2	11o	RCL
oo7	STO	run pgm		o74	)	111	o3
oo8	OO	o38	LBL	o75	=	112	OP
oo9	OP	o39	E	o76	STO	113	o6
o1o	OO	o4o	OP	o77	o3	114	ADV
o11	R/S	o41	o2				
		o42	OP	(VE)		(At)	
Store G		o43	o5	o78	4	115	1
o12	LBL			o79	2	116	3
o13	A'	(Ai)		o8o	1	117	3
o14	1	o44	1	o81	7	118	7
o15	STO	o45	3	o82	OP	119	OP
o16	o5	o46	7	o83	o4	12o	o4
o17	R/S	o47	4	o84	RCL	121	RCL
		o48	OP	o85	o2	122	oo
Store Ai		o49	o4	o86	+	123	÷
o18	LBL	o5o	RCL	o87	RCL	124	RCL
o19	A	o51	oo	o88	o3	125	o7
o2o	STO	o52	OP	o89	=	126	=
o21	oo	o53	o6	o9o	STO	127	OP
o22	R/S			o91	o4	128	o6
		(iB)		o92	OP		
Store iB		o54	7	o93	o6	(Ag)	
o23	LBL	o55	4			129	1
o24	B	o56	1	(DU)		13o	3
o25	STO	o57	4	o94	1	131	2
o26	o1	o58	OP	o95	6	132	2
o27	R/S	o59	o4	o96	4	133	OP
		o6o	RCL	o97	1	134	o4
Store A∩B		o61	oo	o98	OP	135	RCL
o28	LBL	o62	+	o99	o4	136	o1
o29	C	o63	RCL	1oo	RCL	137	÷
o3o	STO	o64	o1	1o1	o2	138	RCL
o31	o2	o65	OP	1o2	OP	139	o7
o32	R/S	o66	o6	1o3	o6	14o	=
						141	OP
						142	o6

(As)		195	OP
143	1	196	o6
144	3	197	RCL
145	3	198	oo
146	6	199	÷
147	OP	2oo	RCL
148	o4	2o1	o4
149	RCL	2o2	=
15o	o4	2o3	OP
151	÷	2o4	o6
152	RCL		
153	o7	(Ag_i)	
154	=	2o5	1
155	OP	2o6	3
156	o6	2o7	2
		2o8	2
(As*)		2o9	2
157	1	21o	4
158	3	211	OP
159	3	212	o4
16o	6	213	RCL
161	5	214	o1
162	1	215	÷
163	OP	216	RCL
164	o4	217	o4
165	RCL	218	=
166	oo	219	OP
167	+	22o	o6
168	RCL		
169	o1	(S)	
17o	=	221	3
171	-	222	6
172	RCL	223	OP
173	o5	224	o4
174	=	225	RCL
175	÷	226	o2
176	(	227	÷
177	(	228	RCL
178	2	229	o7
179	x	23o	=
18o	RCL	231	OP
181	o7	232	o6
182	)		
183	-	(S_i)	
184	1	233	3
185	)	234	6
186	=	235	2
187	OP	236	4
188	o6	237	OP
		238	o4
(At_i)		239	RCL
189	1	24o	o2
19o	3	241	÷
191	3	242	RCL
192	7	243	o4
193	2	244	=
194	4	245	OP
		246	o6

(S_i*)		299	5
247	3	3oo	2
248	6	3o1	4
249	2	3o2	OP
25o	4	3o3	o4
251	5	3o4	1
252	1	3o5	-
253	OP	3o6	(
254	o4	3o7	RCL
255	RCL	3o8	o2
256	o2	3o9	÷
257	x	31o	RCL
258	2	311	o4
259	=	312	)
26o	÷	313	=
261	(	314	OP
262	RCL	315	o6
263	oo		
264	+	(S_i'*)	
265	RCL	316	3
266	o1	317	6
267	)	318	6
268	=	319	5
269	STO	32o	2
27o	o8	321	4
271	OP	322	5
272	o6	323	1
		324	OP
(S')		325	o4
273	3	326	1
274	6	327	-
275	6	328	RCL
276	5	329	o8
277	OP	33o	=
278	o4	331	OP
279	RCL	332	o6
28o	o4		
281	÷	(G)	
282	RCL	333	2
283	o7	334	2
284	=	335	OP
285	-	336	o4
286	(	337	RCL
287	RCL	338	o5
288	o2	339	OP
289	÷	34o	o6
29o	RCL	341	ADV
291	o7	342	ADV
292	)	343	ADV
293	=	344	ADV
294	OP	345	GTO
295	o6	346	E'
(S_i')		Register	Contents
296	3	o1 Ai	o5 G
297	6	o2 iB	o6
298	6	o3 DU	o7 K
		o4 VE	o8 used

Programm Klassifikation quantitative und qualitative Symmetrie

(NB: im Anschluß an die Distributionsanalyse kann das Programm eingelesen werden und mit Lbl E gestartet werden. Die Rechenzeit kann stark variieren, je nach Größe der Zahlen, für die eine Kombinatorik berechnet werden muß.)

```
ooo  LBL     o5o  GE      1oo  RCL     15o  STO
oo1  E       o51  E'      1o1  o8      151  13
oo2  RCL     o52  1       1o2  )       152  IFF
oo3  o2      o53  SUM     1o3  x       153  o2
oo4  STO     o54  15      1o4  (       154  B'
oo5  12      o55  3       1o5  RCL     155  RCL
oo6  RCL     o56  o       1o6  1o      156  1o
oo7  oo      o57  GE      1o7  -       157  +
oo8  STO     o58  E'      1o8  RCL     158  RCL
oo9  1o      o59  1       1o9  o8      159  11
o1o  RCL     o6o  SUM     11o  )       16o  =
o11  o1      o61  15      111  =       161  x⇔t
o12  STO     o62  3       112  ÷       162  RCL
o13  11      o63  8       113  RCL     163  o7
o14  RCL     o64  GE      114  o7      164  GE
o15  o5      o65  E'      115  =       165  SIN
o16  STO     o66  1       116  STO     166  x⇔t
o17  o8      o67  SUM     117  o9      167  -
o18  RCL     o68  15      118  x⇔t     168  (
o19  o7      o69  4       119  RCL     169  RCL
o2o  x⇔t     o7o  8       12o  12      17o  o7
o21  PGM     o71  GE      121  EQ      171  +
o22  o1      o72  E'      122  TAN     172  1
o23  SBR     o73  1       123  GE      173  )
o24  CLR     o74  SUM     124  ENG     174  =
o25  4       o75  15      125  GTO     175  STO
o26  STO     o76  5       126  SUM     176  o5
o27  15      o77  8       127  LBL     177  GTO
o28  7       o78  GE      128  ENG     178  COS
o29  GE      o79  E'      129  STF     179  LBL
o3o  E'      o8o  1       13o  o2      18o  SIN
o31  1       o81  SUM     131  LBL     181  RCL
o32  SUM     o82  15      132  SUM     182  o8
o33  15      8o3  7       133  RCL     183  STO
o34  1       o84  o       134  o7      184  o5
o35  1       o85  GE      135  PGM     185  LBL
o36  GE      o86  E'      136  16      186  COS
o37  E'      o87  1       137  A       187  RCL
o38  1       o88  SUM     138  RCL     188  12
o39  SUM     o89  15      139  11      189  -
o4o  15      o9o  LBL     14o  -       19o  RCL
o41  1       o91  E'      141  RCL     191  o5
o42  6       o92  1       142  o8      192  +
o43  GE      o93  INV     143  =       193  1
o44  E'      o94  SUM     144  PGM     194  =
o45  1       o95  o7      145  16      195  STO
o46  SUM     o96  (       146  B       196  o6
o47  15      o97  RCL     147  PGM     197  GTO
o48  2       o98  11      148  16      198  D
o49  3       o99  -       149  E       199  LBL
```



```
2oo  B'      256  STO     312  OP      369  3
2o1  RCL     257  14      313  o3      37o  6
2o2  12      258  RCL     314  GTO     371  OP
2o3  STO     259  o7      315  C       372  o3
2o4  o5      26o  -       316  LBL     373  LBL
2o5  RCL     261  RCL     317  NOP     374  PRT
2o6  1o      262  1o      318  IFF     375  OP
2o7  x⇔t     263  +       319  o2      376  o5
2o8  RCL     264  RCL     32o  GRD     377  ADV
2o9  11      265  o8      321  1       378  OP
21o  GE      266  =       322  3       379  oo
211  FIX     267  PGM     323  3       4oo  CLR
212  STO     268  16      324  6       4o1  R/S
213  o6      269  A       325  OP      4o2  LBL
214  GTO     27o  RCL     326  o3      4o3  DEG
215  OP      271  11      327  GTO     4o4  3
216  LBL     272  -       328  C       4o5  4
217  FIX     273  RCL     329  LBL     4o6  1
218  x⇔t     274  o5      33o  GRD     4o7  3
219  STO     275  =       331  3       4o8  3
22o  o6      276  PGM     332  6       4o9  6
221  LBL     277  B       333  OP      41o  OP
222  OP      278  PGM     334  o3      411  o3
223  RCL     279  16      335  LBL     412  GTO
224  o6      28o  E       336  C       413  PRT
225  -       281  x       337  1
226  RCL     282  RCL     338  SUM
227  o5      283  14      339  o7
228  +       284  =       341  OP
229  1       285  ÷       342  o5
23o  =       286  RCL     343  PGM
231  STO     287  13      344  o1
232  o6      288  =       345  SBR
233  LBL     289  SUM     346  CLR
234  D       29o  oo      347  o
235  RCL     291  1       348  STO
236  1o      292  SUM     349  oo
237  -       293  o5      35o  STO
238  RCL     294  DSZ     351  o8
239  o8      295  o6      352  INV
24o  =       296  D       353  STF
241  PGM     297  .       354  o2
242  16      298  o       355  RCL
243  A       299  5       356  15
244  RCK     3oo  x⇔t     357  x⇔t
245  o5      3o1  RCL     358  RCL
246  -       3o2  oo      359  1o
247  RCL     3o3  INV     36o  -
248  o8      3o4  GE      361  RCL
249  =       3o5  NOP     362  11
25o  PGM     3o6  LBL     363  =
251  16      3o7  TAN     364  IxI
252  B       3o8  3       365  GE
253  PGM     3o9  6       366  DEG
254  16      31o  3       367  3
255  E       311  6       368  4
```

Programm Klassifikation nach distributioneller Frequenz (ML-Modul)

(NB: Bei sehr kleinem P kann der Rechner blinken ohne das Programm zu unterbrechen; vor Neueingabe CLR drücken. Lbl E: für "O" vorgesehen; bei mehreren gleichen Werten kann die Zahl aktuell in das Rechenprogramm eingesetzt werden - manuell GTO o22, LRN, Zahl ändern, LRN. Σij bei Ersteingabe mittels B speichern, sonst nur C.
Zeitbedarf: ca. 11 Sek.; bei Matrix 5ox5o: 1 Spalte 9,1 Minuten)

Labels:

o19	E
o32	C
198	A'
2o8	A
212	E'
218	B
223	C'

User defined labels and user instruction:

1 Distributional Frequency 7op17 2				
Symbol		n		Entry
Σi.	Σ.j	Σij		Σij=O?

Use order C'→ A→ A'→C_1 (or $C_1 \to B_1$
C_2 $C_2 \to B_2$
C_3 $C_3 \to B_3$
etc. etc.)

(SBR categories)			start
ooo	RC✱	o31	LBL
oo1	69	o32	C
oo2	OP	o33	STO
oo3	o4	o34	56
oo4	RCL	o35	CP
oo5	57	o36	INV
oo6	-	o37	EQ
oo7	4	o38	oo
oo8	=	o39	45
oo9	OP	o4o	1
o1o	o6	o41	SUM
o11	1	o42	69
o12	SUM	o43	STF
o13	57	o44	o1
o14	o	o45	RC✱
o15	STO	o46	57
o16	69	o47	x
o17	RTN	o48	RCL
		o49	58
Store counter		o5o	=
o18	LBL	o51	÷
o19	E	o52	RCL
o2o	STO	o53	59
o21	55	o54	=
o22	o	o55	STO
o23	SBR	o56	67
o24	oo	o57	NOP/PRT
o25	33	o58	x⇔t
o26	DSZ	o59	2
o27	55	o6o	9
o28	oo	o61	GE
o29	22	o62	o1
o3o	R/S	o63	41

o64	(	o98	RCL
o65	RCL	o99	59
o66	56	1oo	-
o67	-	1o1	1
o68	x⇔t	1o2	)
6o9	)	1o3	)
o7o	÷	1o4	)
o71	(	1o5	$\sqrt{x}$
o72	(	1o6	)
o73	(	1o7	=
o74	RCL	1o8	NOP/PRT
o75	59	1o9	x⇔t
o76	-	11o	RCL
o77	RCL	111	66
o78	58	112	INV
o79	)	113	GE
o8o	x	114	o1
o81	(	115	27
o82	RCL	116	+/-
o83	59	117	GE
o84	-	118	o1
o85	RC✱	119	34
o86	57	12o	6
o87	)	121	2
o88	x	122	SUM
o89	RCL	123	69
o9o	67	124	GTO
o91	)	125	oo
o92	÷	126	oo
o93	(	127	6
o94	RCL	128	O
o95	59	129	SUM
o96	x	13o	69
o97	(	131	GTO

132	oo	189	RCL
133	oo	19o	56
134	6	191	GE
135	4	192	o1
136	SUM	193	27
137	69	194	GTO
138	GTO	195	o1
139	oo	196	34
14o	oo		
141	IFF	Store Code	
142	o1	197	LBL
143	o1	198	A'
144	6o	199	OP
145	RCL	2oo	oo
146	56	2o1	OP
147	PGM	2o2	o1
148	16	2o3	ADV
149	A	2o4	OP
15o	PGM	2o5	o5
151	16	2o6	R/S
152	C		
153	RCL	Store Σi.	
154	67	2o7	LBL
155	y^x	2o8	A
156	RCL	2o9	STO
157	56	21o	58
158	=	211	LBL
159	x	212	E'
16o	RCL	213	5
161	67	214	STO
162	+/-	215	57
163	INV	216	R/S
164	LNX		
165	=	Store Σ.j	
166	IFF	217	LBL
167	o1	218	B
168	o1	219	ST✱
169	75	22o	57
17o	÷	221	R/S
171	RCL		
172	o4	Store n	
173	=	222	LBL
174	NOP/PRT	223	C'
175	x⇔t	224	STO
176	INV	225	59
177	STF	226	R/S
178	o1		
179	.		
18o	o		
181	5		
182	INV		
183	GE		
184	o1		
185	2o		
186	RCL		
187	67		
188	x⇔t		

Register Contents:
oo - o4 used (PGM 16)
o4 n!
o5 - 54 Σj
55 counter
56 Σij
57 counter
58 Σi
59 n
6o-65 code categories
66 1.645
67 E
68 z
69 counter for 6o-65

flag o1
x⇔t

Store following codes
and date up on card side 2:

6o	33oooo
61	17353532
62	13oooo
63	42oooo
64	3ooooo
65	24oooo
66	1.645

LITERATURVERZEICHNIS

ALTMANN, G.
1971 Introduction to Quantitative Phonology. Habilitationsschrift Bochum.

ALTMANN, G., LEHFELDT, W.
1972 Typologie der phonologischen Distributionsprofile. Beiträge zur Linguistik und Informationsverarbeitung 22, 8-32.
1980 Einführung in die Quantitative Phonologie. Bochum: Brockmeyer.

ANDREEV, N.D. (Hrsg.)
1965 Statistiko-kombinatornoe modelirovanie jazykov. Moskva-Leningrad: Nauka.

DENES, P.D.
1964 On the statistics of spoken English. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 17, 51-71.

KARLGREN, H.
1974[3] Statistical methods in Phonetics. In MALMBERG, B. (Ed.): Manual of Phonetics. Amsterdam, London, New York, 129-154.

LADEFOGED, P.
1975 A Course in Phonetics. New York.

LEHFELDT, W.
1973 Distributionelle Phonemähnlichkeit. Phonetica 27, 82-99.
1975 Die Verteilung der Phonemanzahl in den natürlichen Sprachen. Phonetica 31, 274-287.
1976 Zur Messung der phonetischen Lautdifferenz. Eine begriffskritische Untersuchung. Glottometrika 1, 26-45.

LEHFELDT, W., ALTMANN, G.
1975 Begriffskritische Untersuchungen zur Sprachtypologie. Linguistics 144, 49-78.

LINDNER, G.
1975 Der Sprechbewegungsablauf. Eine phonetische Studie des Deutschen. Berlin: Akademie-Verlag.
1980 Lautfolgestrukturen im Deutschen. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 33, 468-477.

OLLER, D.K., EILERS, R.E.
1975 Phonetic expectation and transcription validity. Phonetica 31, 288-304.

ORTMANN, W.D.
1975 Beispielwörter für deutsche Ausspracheübungen. München.

POYATOS, F.
1975 Cross-Cultural Study of Paralinguistic "Alternants" in Face-to-Face-Interaction. In KENDON, A., HARRIS, R.M., KEY, M.R. (Eds.): Organization of Behaviour in Face-to-Face-Interaction, 285-313.

RICHTER, B., RICHTER, H.
1981 Zur Abbildungstreue von Transkriptionen. In LANGE-SEIDL, A. (Hrsg.): Zeichenkonstitution II, 110-119.

SEGAL, M.
1972 Osnovy fonologičeskoj statistiki. Moskva-Leningrad: Nauka.

TIFFANY, W.R., CARRELL, J.
1977[2] Phonetics. New York.

WOTHKE, K.
1980 Computerunterstützte Phonemanalyse. In HALLER, J. (Hrsg.): Maschinelle Sprachverarbeitung, 36-58.

WINKLER, P.
1979 Notationen des Sprechausdrucks. Zeitschrift für Semiotik 1, 211-224.
1981 Anwendungen phonetischer Methoden für die Analyse von Face-to-Face-Situationen. In WINKLER, P. (Hrsg.): Methoden der Analyse von Face-to-Face-Situationen. Stuttgart: Metzler, 9-46.
1982 Notation of Paraphonetic Features. In HESS-LÜTTICH, E.W.B. (Ed.): Multimedial Communication 1, Tübingen: Narr.

GRAPHEME UND GRAPHEMKOMBINATORIK DER RUSSISCHEN FACHSPRACHE

EINE PHONOSTATISTISCHE UNTERSUCHUNG

J. Dietze, Halle (Saale)

Die Phonostatistik beschäftigt sich mit dem frequentativen Verhalten von Phonen (Lauten), Phonemen sowie distinktiven Phonemmerkmalen, um entsprechende quantitative Regularitäten zu ermitteln (vgl. ROCŁAWSKI 1876, 21). Von ihr zu unterscheiden ist die Phonotaktik, deren Aufgabenbereich APRESJAN (1971, 52) wie folgt definiert: "Innerhalb der Phonologie sei die Phonotaktik herausgehoben, die Wissenschaft von den Gesetzen der Phonemverbindung, welche die strukturbedingten Beschränkungen der Kombinationsmöglichkeiten der Phoneme erforscht." Für die Untersuchung phonotaktischer Fragestellungen lassen sich statistische Methoden erfolgreich einsetzen. Phoneme sind die kleinsten phonologischen Einheiten, die nicht in noch kleinere aufgespalten werden können (vgl. KOSCHMIEDER 1977, 13). Strukturell gesehen, entspricht dem Phonem auf der graphematischen Ebene das Graphem. Das Graphem ist das Bezeichnende, das Phonem das Bezeichnete. Zwischen dem Graphemsystem einer Sprache und deren Phonemsystem bestehen Beziehungen, die auch direkte Schlüsse vom Graphemsystem auf das Phonemsystem gestatten können; denn die phonologische Struktur determiniert die graphematische, wenn auch vermittelt. Es wäre deshalb korrekt, die folgende Untersuchung "graphostatistisch" zu nennen. Wegen der Ungewöhnlichkeit dieses Terminus wollen wir sie jedoch - unter Beachtung der genannten Einschränkungen - der Phonostatistik zuordnen.

Die Buchstaben des Russischen nennen wir "Grapheme". Da in den von uns ausgewerteten Texten keine Allographe vorkommen (Großbuchstaben und ē wurden als Kleinbuchstaben bzw. e abgelocht), brauchen wir die begriffliche Unterscheidung von Graphemen und Graphen in unserem Falle nicht zu beachten.[1]

Die erste umfassende Arbeit zur Phonostatistik des Russischen haben BELONOGOV und FROLOV (1963) veröffentlicht. Diesen Autoren hatten russische Texte der Geschäftssprache mit einem Umfang von 30 000 Wörtern bzw. 200 000 Buchstaben als Materialgrundlage gedient. TOPOROV (1966, 110) hält den Umfang dieser Stichprobe für zu klein - zu Unrecht, wie wir meinen. Jedenfalls wird für sie keine nachprüfbare Begründung angeführt. Ferner sind TOPOROVs eigene Untersuchungen ohne textstatistische Parameter, da sein Material erklärenden Wörterbüchern entstammt, Quellen also, in denen nichts über die Auftretenshäufigkeit der verzeichneten Einheiten zu finden ist. Um festzustellen, ob eine Stichprobe für sprachstatistische Forschungen hinreichend groß ist, kann man nach der Wahrscheinlichkeitsrechnung den relativen Fehler approximativ ermitteln. Nach FRUMKINA (1973, 282) benützt man hierfür folgenden Ausdruck:

$$\frac{z_\rho}{\sqrt{NP}}$$

N = Stichprobenumfang (beispielsweise Anzahl der Buchstaben)

P = relative Häufigkeit der zu untersuchenden Einheit

z_ρ = Konstante, die die Größe der Wahrscheinlichkeit angibt. Wir setzen hier z = 1.96, damit die Konfidenz 95 % beträgt (vgl. KALININA 1973, 94).

Wenn man P als arithmetisches Mittel der Vorkommenshäufigkeit eines russischen Graphems in der relativen Form (bezogen auf die Gesamtzahl der Buchstaben) auffaßt, dann beträgt der mittlere relative Fehler 0.0248, d.h. rund 2.5 %. Bei sprachstatistischen Untersuchungen sollte der relative Fehler 30 % nicht übersteigen.

Unserer phonostatistischen Untersuchung wurden 500 russischsprachige Referate einschließlich ihrer Titel zugrundegelegt, die der Referatezeitschrift (Referativnyj žumal) zum Bereich "wissenschaftliche und technische Information" entstammen. Dieses Textkorpus enthält 57 566 Wortformen mit insgesamt 429 257 Buchstaben, wobei hier, wie erwähnt, der Unterschied zwischen

Groß- und Kleinbuchstaben vernachlässigt wurde. Das arithmetische Mittel der Frequenz der russischen Grapheme beträgt in unserem Falle 13 414.28. Diese Zahl wird in die obige Formel eingesetzt, so daß sich der mittlere relative Fehler auf 0.0169, d.h. rund 1.7 %, beläuft. Die Größe der von uns gewählten Stichprobe dürfte in Anbetracht des Untersuchungsgegenstandes statistisch repräsentativ sein; denn die kleinste Frequenz von 156 (für ь) weist nur einen relativen Fehler von 0.157, d.h. 15.7 % auf. Zum Vergleich sei auf die Arbeit von GVOZDOVIČ (1976c) hingewiesen. Dieser Autor hat deutschsprachige Texte mit 50 000 Buchstaben bezüglich der Distribution untersucht und gibt eine mögliche Fehlerquote von 5 % an.

GVOZDOVIČ (GWOSDOWITSCH 1976a) hat in gleicher Weise wie TOPOROV neben ausländischen Ortsnamen auch Abkürzungen in seine graphemkombinatorischen Untersuchungen einbezogen. Soweit es sich bei den Abkürzungen um Kurzwörter handelt, die nach dem Vorbild der Akronyme aus den Anfangssilben der abzukürzenden Wortformen gebildet werden, ist gegen dieses Verfahren nichts einzuwenden, da bei solchen Abbreviaturen das phonotaktische System nicht verletzt wird. Werden jedoch Akronyme aus den Anfangsbuchstaben der abzukürzenden Wortformen zusammengesetzt, so sollten sie bei phonostatistischen Forschungen insofern nicht berücksichtigt werden, als sie die Untersuchungsergebnisse verfälschen können. Beim Ablochen der Texte haben wir bereits gängige fachliche Akronyme aufgelöst.

In Tabelle 1 führen wir die Vorkommenshäufigkeit der einzelnen russischen Grapheme an, und zwar in absoluten sowie in Prozentzahlen. Ergänzend dazu bringen wir auch die in Prozentzahlen umgerechneten Werte von BELONOGOV und FROLOV (1963, 287).[2] Die Reihenfolge der Buchstaben ergibt sich aus der absteigenden Frequenz.

Die im rückläufigen Wörterbuch des Russischen von 1974 (Obratnyj slovar') aufgeführten phonostatistischen Angaben konnten wir leider nicht zum Vergleich heranziehen, weil sie nur den Auslaut betreffen. Gleiches gilt auch für die noch zu behandelnden russischen Graphemkombinationen.

Den phonetischen Gegensatz "hart ~ palatalisiert" der in

Tabelle 1: Absolute und relative Häufigkeit der russischen Grapheme in der Textstichprobe

Graphem	absolute Frequenz	relative Frequenz in %	
		DIETZE	BELONOGOV, FROLOV
о	44 172	10.29	10.47
и	42 024	9.79	7.00
е	35 662	8.31	8.36
а	33 967	7.91	8.08
н	29 877	6.96	7.23
т	27 447	6.39	6.25
с	26 034	6.06	4.66
р	22 279	5.19	5.84
в	17 586	4.10	5.69
л	14 613	3.40	2.48
к	14 189	3.31	2.64
м	13 890	3.24	3.37
п	12 736	2.97	3.71
д	11 079	2.58	3.88
я	9 893	2.30	2.49
ы	8 632	2.01	1.79
у	8 413	1.96	2.02
з	7 000	1.63	1.47
б	6 464	1.51	1.70
ч	6 005	1.40	1.80
х	5 390	1.26	1.32
й	4 852	1.13	1.58
г	4 716	1.10	1.11
ц	4 491	1.05	0.29
ь	4 389	1.02	1.68
ф	3 912	0.91	0.17
ю	2 904	0.68	0.63
ж	2 537	0.59	0.96
щ	1 670	0.39	0.35
ш	1 224	0.29	0.50
э	1 054	0.25	0.17
ъ	156	0.04	0.03

Frage kommenden russischen Konsonanten haben wir statistisch ebenfalls auf maschinellem Wege approximativ zu erfassen versucht.[3] Als palatalisiert gelten hierbei die unten angeführten Konsonanten in folgender Position: Konsonant + е, и, ь, ю, я. Harte Konsonanten stehen in folgenden Positionen: 1. Konsonant + а, о, у, ъ, э. 2. Konsonant im absoluten Auslaut. 3. Konsonant + Konsonant(en). Bei Konsonantenverbindungen konnten wir mögliche Palatalisierungen, die im Russischen meist regressiv wirken und Assimilationen, die die statistischen Werte verschieben, nicht berücksichtigen. Die von uns beigebrachten Prozentangaben beziehen sich jeweils auf die Gesamtfrequenz des betreffenden Konsonantengraphems.

Tabelle 2:

Graphem	Palatalisiert		Hart	
	absolute Frequenz	/ in %	absolute Frequenz	/ in %
б	1243	19.2	5221	80.8
в	2655	15.0	14931	85.0
г	596	12.6	4120	87.4
д	3577	32.3	7502	67.7
э	684	9.8	6316	90.2
к	2354	16.6	11835	83.4
л	10175	69.6	4438	30.4
м	4534	32.6	9356	67.4
н	8471	28.4	21406	71.6
п	2384	18.7	10352	81.3
р	6782	30.4	15497	69.6
с	5227	20.1	20807	79.9
т	9534	34.7	17913	65.3
ф	1329	34.0	2583	66.0
х	199	3.7	5191	96.3

Addiert man zu den palatalisierten russischen Konsonanten (59 744) noch die Frequenz der Grapheme der palatalen Konsonanten (ч, щ) so ergibt sich eine Frequenz von insgesamt 67 419. Addiert man in gleicher Weise die Frequenz der eben erfaßten harten Konsonanten (176 676) zu der Frequenz der Grapheme für positionsunabhängig harte Konsonanten (ж, ц, ш), so erhalten wir als Summe 185 928. Von den 253 347 Konsonantengraphemen bezeichnen mithin 26.6 % palatalisierte bzw. palatale Konsonanten und 73.4 % harte Konsonanten. Diese Angaben sind unter Beachtung der oben gemachten Einschränkungen lediglich als Näherungswerte zu verstehen.

Bei der nun folgenden Behandlung von Graphemverbindungen beschränken wir uns bewußt auf die Verbindungen von Konsonantengraphemen untereinander, unter Einbeziehung des Weichheitszeichens. BELONOGOV und FROLOV (1963) haben die bigraphemischen Verbindungen unter Einschluß der Vokalgrapheme untersucht, so daß deren statistische Frequenzangaben mit den unsrigen nicht vergleichbar sind. In gleicher Weise ist auch TOPOROV (1966) vorgegangen. Beide Arbeiten zeigen aber bei den bigraphemischen Konsonantenverbindungen mehr Kombinationsmöglichkeiten als in unseren Texten. Das trifft vor allem für die Angaben bei TOPOROV zu und ist damit zu erklären, daß TOPOROV Lexika ausgewertet hat, und zwar unter Einschluß von Akronymen (Bildung mit Hilfe der Initialbuchstaben). Das Besondere der Untersuchung TOPOROVs besteht in der Klassifizierung der Graphempaare nach Distributionsmerkmalen (Valenz, Symmetrie, Transitivität), wobei er der von HARARY und PAPER (1957) entwickelten Methode folgt, die auf der Relationstheorie beruht und mit deren Hilfe die "interaction" von Graphempaaren bestimmt wird.[4] Wir müssen jedoch einwenden, daß damit von der Graphemebene aus kaum weiterführende Einsichten in höher liegende Strukturebenen wie die der Phonetik oder Phonologie oder gar der Morphologie gewonnen werden können. Wir haben uns daher entschlossen, die konsonantischen Graphempaare im Anhang lediglich in Form eines Überblicks darzubieten. Dies geschieht mittels einer Tabelle, in der auf der ersten Spalte die Konsonantengrapheme in erster Position und in der ersten Zeile die Konsonantengrapheme in zweiter Position eingezeichnet sind. Diese Tabelle gestattet es, den Sättigungsgrad der Valenz (= Anzahl der Grapheme, die

mit dem zu untersuchenden Graphem eine Verbindung eingehen) sowie die Symmetrie (= Fähigkeit eines Graphems, die 1. und/oder die 2. Position einzunehmen) der Grapheme zu bestimmen. Daraus lassen sich wiederum Distributionsklassen aufstellen.[5] Unsere Frequenzangaben sind absolut, beim Errechnen der relativen Frequenz müßte man sich auf die Summe aller von uns gezählten konsonantischen Graphempaare beziehen (52 454), vgl. Anhang I.

Vom Typ her konnten wir 214 bigraphemische Konsonantenverbindungen von 400 (= 20^2) mathematisch möglichen feststellen; das sind 53.5 %. TOPOROV (1966, 66) hat ermittelt, daß der Anteil der bigraphemischen Verbindungen von den rechnerisch möglichen unter Berücksichtigung aller Grapheme bei 75 % liegt.

Die Kombinatorik der Konsonantengrapheme läßt u.E. erst dann Einsichten auf der höherliegenden Strukturebene zu, wenn die ermittelten Graphemverbindungen typisiert werden. Wir folgen hier dem Versuch von ŠABROVA (1978), die die russischen Konsonantengrapheme zu phonetisch definierten Gruppen zusammengefaßt hat. Dieses Vorgehen hat den Vorteil, daß Palatalisierungserscheinungen und vor allem regressive Stimmassimilationen, die auf der phonetischen Ebene ablaufen, auf der Graphemebene neutralisiert werden. Für die Gruppenbildung werden artikulatorische Merkmale unter Beachtung der Klassifikation nach Geräuschkonsonanten und Sonoren herangezogen: T = Klusile (б, п; г, к; д, т); S = Spiranten (з, с; ж, ш; в, ф; х); S = Sonore (м, н, р, л); A = Affrikaten (ч, ц, щ). Die kodierte Bezeichnung der Gruppen haben wir ebenfalls von ŠABROVA übernommen, um Vergleiche des Materials zu erleichtern. ŠABROVA bietet jedoch keine Frequenzangaben, sondern nur distributive Gruppierungen. In diesem Zusammenhang ist der Hinweis wichtig, daß alle bigraphemischen Gruppen mit zwei Ausnahmen zu dem sogenannten gemischten Typ gehören, d.h. die konsonantischen Graphempaare stehen sowohl auf der Morphemgrenze als auch innerhalb eines Morphems. Nur die Gruppen TA und AS kommen ausschließlich auf der Morphemgrenze vor.

Bei der maschinellen Aufbereitung unserer Texte haben wir jeweils den Wortlaut gesondert gekennzeichnet, so daß wir bei

den einzelnen Gruppen der Graphempaare die im Anlaut stehenden Verbindungen angeben werden. Hervorheben möchten wir diejenigen konsonantischen Graphempaare, die allein im Anlaut vorkamen:[6] сг-, сд-, сж-, сф-, тщ-, фл-, шв-, шп-, шр-. Zur Illustration unserer obigen Behauptung, daß Akronyme des von uns ausgeschlossenen Bildungstyps in der Regel die graphisch reflektierten phonetischen Gesetzmäßigkeiten durchbrechen, seien die entsprechenden Abkürzungen hier angeführt, da sie zum Anlaut gehören: нм, мм, р(ефератный) ж(урнал), ц(ентральный) к(аталог), Ц(ентральный) Н(аучно-) И(сследовательский) И(нститут) П(атентной) И(нформации).

Von den 16 mathematisch möglichen Gruppen der bigraphemischen Konsonantenverbindungen sind in unseren Texten 15 realisiert (93.8 %). Wir geben die absolute Frequenz der einzelnen Graphempaare, die zu der Gruppe gehören, an, darauf folgen die relative Frequenz in Prozenten (bezogen auf die Gesamtzahl der konsonantischen Graphempaare 52 454) und abschließend die Anzahl der anlautenden Graphempaare (absolute Frequenz).

Tabelle 3:

Graphemgruppe	absolute Gesamtfrequenz	/ in %	Anlaut
TT	2125	4.1	13
TS	3981	7.6	220
TR	11820	22.5	5083
TA	419	0.8	4
ST	9474	18.1	997
SS	2728	5.2	1055
SR	6559	12.5	1667
SA	108	0.2	26
RT	3097	5.9	-
RS	3000	5.7	-
RR	7128	13.6	104
RA	255	0.5	-
AT	420	0.8	309
AS	43	0.1	-
AR	1194	2.3	16

An dreigraphemischen Konsonantenverbindungen haben wir insgesamt 4 438 in unseren Texten gezählt. Vom Typ her wären rechnerisch 20^3 = 8 000 Verbindungen möglich, 142 kommen jedoch nur vor (1.8 %), wobei die ermittelten folgenden Abkürzungen unberücksichtigt blieben: б(иблиотечно-) б(иблиографическая) к(лассификация), В(енгерская) Н(ародная) Р(еспублика), ГДР, Н(ародная) Р(еспублика) Б(олгария), П(ольская) Н(ародная) Р(еспублика), р(аскладочно-) п(одборочная) м(ашина), с(четно-) п(ерфорационная) м(ашина), С(оциалистическая) Р(еспублика) Р(умыния), С(оветская) С(оциалистическая) Р(еспублина), Ф(едеративная) Р(еспублика) Г(ермании), (электронная) ц(ифровая) в(ычислительная) м(ашина). Demgegenüber finden sich folgende dreigraphemische Konsonantenverbindungen, die keine Abkürzungen darstellen, nur im Anlaut: взв-, взр-, вкл-, впл-, всл-, вст-, эдр-, мгн-, сбл-, сгл-, сгр-.[6]

Von 64 (= 4^3) mathematisch möglichen Gruppen dreigraphemischer Konsonantenverbindungen finden sich 28 (43.8 %) in den Texten. Die Prozentangaben bei der Frequenz beziehen sich auf die Gesamtzahl aller dreigraphemischen Konsonantenverbindungen (4 438).

Die Summe aller viergraphemischen Konsonantenverbindungen ist 795. Vom Typ her wären rechnerisch 20^4 = 160 000 Verbindungen möglich, 34 kommen jedoch nur vor (0.02 %); dies ohne Berücksichtigung der Abkürzungen СССР und ЧССР. Echte, ausschließlich im Anlaut stehende Verbindungen sind взгл- sowie вскр-.[6]

Von 256 (= 4^4) mathematisch möglichen Gruppen viergraphemischer Konsonantenverbindungen finden sich 15 (5.9 %). Die Prozentangaben bei der Frequenz beziehen sich auf die Gesamtzahl aller viergraphemischen Konsonantenverbindungen (795).

TOPOROV (1966, 126) führt noch eine Reihe von Verbindungen mit 5 bzw. 6 Konsonantengraphemen an, es handelt sich dabei aber mit folgenden Ausnahmen immer um Fremdwörter: рхвкл (сверхвключение), рхвкр (сверхвкрадчивый); dieser Typ RS/STR liegt auf der Morphemgrenze. Zu beachten wäre bei einem Vergleich mit TOPOROVs Ergebnissen, daß wir das Weichheitszeichen bei der Konstituierung der Gruppen aus phonetischen Gründen nicht mitgezählt haben. Daher wird die Verbindung льств (z.B. довольство) von TOPOROV zu den fünfgraphemischen gerechnet, während wir sie als viergraphemisch (RSTS) ansehen. Als fünfgraphemische Konsonantenverbindung konnten wir lediglich die Abkürzung РСФСР ermitteln.

Tabelle 4:

Graphemgruppe	absolute Gesamtfrequenz	/ in %	Anlaut
TTS	9	0.2	-
TTR	167	3.8	-
TST	361	8.1	-
TSS	3	0.1	-
TSR	104	2.3	-
TSA	8	0.2	-
TRR	5	0.1	-
TAR	3	0.1	-
STT	20	0.5	-
STS	640	14.4	1
STR	1235	27.8	451
SST	112	2.5	17
SSS	2	0.05	2
SSR	576	13.0	1
SSA	5	0.1	-
SRR	11	0.3	1
SAR	2	0.05	-
RTT	11	0.3	-
RTS	27	0.6	-
RTR	548	12.4	1
RTA	64	1.4	-
RST	386	8.7	-
RSS	8	0.2	-
RSR	116	2.6	-
RRT	1	0.02	-
RRR	8	0.2	-
RAT	2	0.05	-
AST	1	0.02	-

Tabelle 5:

Graphemgruppe	absolute Gesamtfrequenz	/ in %	Anlaut
TSTS	373	46.9	-
TSTR	44	5.5	-
TSST	2	0.25	-
TRTR	1	0.125	-
TRST	3	0.4	-
STST	3	0.4	-
STSR	73	9.2	-
SSTS	7	0.9	-
SSTR	39	4.9	26
RTST	11	1.4	-
RSTS	200	25.2	-
RSTR	35	4.4	-
RSST	1	0.125	-
RRST	1	0.125	-
RAST	2	0.25	-

Anmerkungen

1 Vgl. für das Altrussische DIETZE (1975, 39).

2 In diesem Zusammenhang soll auch auf die Frequenzangaben zur russischen Fachsprache der Radioelektronik bei ANDREEVA et al. (1965, 50) verwiesen werden.

3 Für das Programmieren und die rechentechnische Betreuung schulde ich Frau Chr. Schleiff vom Organisations- und Rechenzentrum der Martin-Luther-Universität in Halle/S. herzlichen Dank.

4 Der Klassifikationsmethode von HARARY und PAPER (1957) folgt u.a. auch GVOZDOVIČ (vgl. 1976a, 1976b, 1976c) für das Deutsche.

5 Diese Ergebnisse s. bei TOPOROV (1966, 84-116).

6 Vgl. dazu TOPOROV (1966, 127-133).

Anhang I: Frequenzen zweigliedriger Konsonantengraphemkombinationen im Russischen

	б	в	г	д	ж	з	н	л	м	п	р	с	т	ф	х	ц	ч	ш	щ	ni.
б		2			10	37	14	1178	22	210		675	42			145		3	210	2548
в		118		7		43	163	833	202	940	18	246	180	397		36		54	2	3239
г	1			37			16	112	25	113		588					16			908
д	14	209	81	10	10	3	179	879	166	570	83	269	8	1	3	19	7	45	14	2570
ж	105	1		441			2			419										968
з	63	405	15	575		3	44	216	251	511	2	269					12	12		2378
н		77	2		135	5	1	547		99		338	351	825			67		4	2451
л	2	2	41	2	37	383	155	112	54	1463			20	171	78		3	2	136	2661
м	31	14		1			35	15	141	155	33		5							430
н		23	17	535	26	11	140	2		2753			31	1033	1537		208	15	40	1	6372
п							3	147		54	109	3826	315	188	4		4			4650
р	5	122	252	69	190		76	10	2002	506	33	19	71	466	173	41	2	24	60	4121
с	42	532	2	14	11		2251	1209	143	478	1090	254	874	4194	13	86	19	61	124	2	11399
т	31	293		99			420	34	123	673	6	1171	2292	5				66		4	5217
ф								1	2			105	1	20	138						267
х	1	50	1				2	5	1	297		87	55	2							501
ц							16														16
ч							79	16		1152		17		325				43			1632
ш		4					13	61		27	1	2		9							117
щ									1	8											9
n.j	295	1852	411	1790	419	485	3609	5377	3133	10428	1375	7866	4245	7636	1946	327	322	241	478	219	52454

Anhang II: Status zweigliedriger Konsonantengraphemkombinationen im Russischen (nach ALTMANN, G. (1973): Probabilistische Klassifikation von Konsonantenverbindungen des Indonesischen. In: Zeitschrift der Deutschen Morgenländischen Gesellschaft 123, 98-116)

	б	в	г	д	ж	з	к	л	м	н	п	р	с	т	ф	х	ц	ч	ш	щ
б	I	M	I	I	M	P	M	P	M	M	I	P	M	I	I	P	I	I	M	P
в	I	A	I	M	I	P	M	P	A	P	M	M	M	M	I	P	I	I	P	M
г	A	I	I	A	I	I	M	P	M	M	I	P	I	I	I	I	I	P	I	I
д	A	P	P	M	M	M	A	P	A	P	P	M	M	M	M	A	M	P	M	I
ж	P	M	I	P	I	I	M	I	I	P	I	I	I	I	I	I	I	I	I	I
з	P	P	A	P	I	M	M	A	P	P	M	M	I	I	I	I	A	A	I	I
к	I	A	M	I	P	M	M	P	I	M	I	A	P	P	I	I	P	I	M	I
л	M	M	P	M	P	P	M	M	M	P	I	I	M	M	M	I	M	M	P	I
м	P	A	V	M	V	I	A	M	P	P	P	I	M	I	I	V	V	V	I	V
н	I	M	M	P	M	M	M	M	I	P	I	I	M	P	P	I	P	M	M	M
п	I	I	I	I	I	I	M	M	I	M	A	P	M	M	M	I	M	I	I	I
р	M	M	P	M	P	I	M	M	P	M	M	M	M	M	A	P	M	A	P	I
с	M	P	M	M	M	I	P	A	M	M	P	M	A	P	M	P	M	A	P	M
т	A	P	I	M	I	I	P	M	M	M	M	P	P	M	I	I	I	P	I	M
ф	V	I	V	I	V	V	I	M	M	I	I	P	M	M	P	V	V	V	V	V
х	A	P	A	I	I	I	M	M	M	P	I	A	P	M	I	V	V	V	I	V
ц	V	V	V	V	V	V	P	V	V	I	V	V	V	V	V	V	V	V	V	V
ч	I	I	I	I	I	I	M	M	I	P	I	M	I	P	I	I	I	I	P	I
ш	V	A	V	I	V	V	A	P	I	A	A	M	I	M	I	V	V	V	V	V
щ	V	V	V	V	V	V	V	V	A	P	V	V	V	V	V	V	V	V	V	V

Legende: P = "präferiert", d.h., existent und signifikant häufig;

A = "aktual", d.h., existent und normal häufig;

M = "marginal", d.h., existent, aber signifikant selten;

V = "virtuell", d.h., nichtexistent, aber eine "zufällige Lücke";

I = "inadmissible", d.h., nichtexistent und zugleich eine "strukturelle Lücke".

LITERATUR

ANDREEVA, L.D. et al.

1965 Polučenie pervogo morfologičeskogo tipa russkogo jazyka v pod"jazyke radioėlektroniki posredstvom algoritma statistiko-kombinatornogo modelirovanija. In: Statistiko-kombinatornoe modelirovanie jazykov. Moskva-Leningrad, 49-64.

APRESJAN, Ju.D.

1971 Ideen und Methoden der modernen strukturellen Linguistik. Kurzer Abriß. Berlin (Sammlung Akademie-Verlag, Sprache, Bd. 12).

BELONOGOV, G.G., FROLOV, G.D.

1963 Ėmpiričeskie dannye o raspredelenii bukv v russkoj pis'mennoj reči. In: Problemy kibernetiki 9, 287-305.

DIETZE, J.

1975 Die Sprache der Ersten Novgoroder Chronik. Poznań (Uniwersytet im. Adama Mickiewicza w Poznaniu, Serie Filologia rosyjska, 6).

FRUMKINA, R.A.

1973 Die Anwendung statistischer Methoden in der Sprachforschung. In: Sprachstatistik. Berlin (Sammlung Akademie-Verlag, Sprache, Bd. 22), 272-298.

GWOSDOWITSCH, B.N.

1976a Deutsche Grapheme. In: Wissenschaftliche Zeitschrift der Martin-Luther-Universität Halle-Wittenberg, Gesellschafts- und sprachwiss. Reihe 25/1, 61-81.

1976b Deutsche Graphempaare. In: Wissenschaftliche Zeitschrift der Martin-Luther-Universität Halle-Wittenberg, Gesellschafts- und sprachwiss. Reihe 25/1, 83-90.

GVOZDOVIČ, B.N.

1976c Zakony raspredelenija častot nemeckich grafem. In: Voprosy terminologii i lingvističeskoj statistiki. Voronež, 105 f.

HARARY, F., PAPER, H.
1957 Toward a general calculus of phonemic distribution. In: Language 33, 143-169.

KALININA, E.A.
1973 Untersuchung lexikostatistischer Gesetzmäßigkeiten auf der Grundlage eines Wahrscheinlichkeitsmodells. In: Sprachstatistik. Berlin (Sammlung Akademie-Verlag, Sprache, Bd. 22), 89-143.

KOSCHMIEDER, E.
1977 Phonationslehre des Polnischen. München (Münchener Universitäts-Schriften, Reihe der Philosoph. Fakultät, Bd. 13).

OBRATNYJ SLOVAR'
1974 Obratnyj slovar' russkogo jazyka. Moskva.

ROCŁAWSKI, Br.
1976 Zarys fonologii, fonetyki, fonotaktyki i fonostatystiki współczesnego języka polskiego. Gdańsk.

ŠABROVA, V.A.
1978 Sočetanija dvuch soglasnych na morfemnom styke. In: Russkij jazyk v škole 4, 85-89.

TOPOROV, V.N.
1966 Materialy dlja distribucii grafem v pis'mennoj forme russkogo jazyka. In: Strukturnaja tipologija jazykov. Moskva, 65-143.

ZUR ÜBERPRÜFUNG DES MENZERATH'SCHEN GESETZES

IM BEREICH DER MORPHOLOGIE

Rainer Gerlach, Göttingen

In der vorliegenden Arbeit soll an deutschem Lexikonmaterial die Hypothese überprüft werden, ob zwischen der Wortlänge (in Phonemen) und der durchschnittlichen Morphlänge (in Phonemen) der stochastische Zusammenhang:

'Je länger das Wort, um so kürzer seine Morphe.'

besteht; d.h. das Menzerath'sche Gesetz, wie von ALTMANN (1980) abgeleitet, soll im Bereich der Morphologie getestet werden.

Zu diesem Zweck wurde das gesamte Stichwortinventar von WAHRIG (1978) untersucht. Es handelt sich um ca. 16 000 Lemmata. Von jedem Stichwort wurde die Länge, die Zahl seiner Morphe und deren Durchschnittslänge festgestellt. Der phonologischen Transkription der Stichwörter wurde das Phoneminventar des Deutschen von WERNER (1972, 22;40) zugrundegelegt. Die Frikative [s], [x], [š], [ž], der velare Nasal [ŋ] sowie [h] und [j] wurden als *ein Phonem* gewertet. Das Problem der mono- bzw. biphonematischen Wertung der Affrikaten und Diphthonge wurde zugunsten des Biphonematischen entschieden (vgl. dazu MORCINIEC 1958 und 1968).

Die Stichwörter in Form von Phonemsequenzen wurden morphologisch analysiert, wobei das Morpheminventar von AUGST (1975) als Basis benutzt wurde. Leider konnte dieses Inventar aber nicht generell verwendet werden, weil es ausschließlich auf Informantenbefragung beruht und daher für linguistische Analysen oft zu ungenau ist. Deshalb wurde ein großer Teil der Stichwörter noch weiter segmentiert, nachdem dies auf Grund eines Substitutionstests angezeigt war (vgl. GREENBERG 1960,

188). Ø-Elemente wurden nicht verwendet; unikale Morphe wurden als eigenständige Segmente behandelt; Eigennamen wurden nicht verwertet; Homonyme bzw. Homographe wurden - je nachdem wie oft diese im Lexikon verzeichnet waren - doppelt oder mehrfach behandelt und gezählt.

Insgesamt wurden 15011 Stichwörter untersucht, die 110672 Phoneme enthielten und sich auf 35346 Morphe verteilten. Die Durchschnittslänge des deutschen Wortes beträgt demnach 7.37 Phoneme (MENZERATH: 7.29), die sich auf durchschnittlich 2.35 Morphe pro Wort verteilen.

Tabelle 1 zeigt die Verteilung der 15011 Einheiten auf die einzelnen Phonem-/Morphtypen; Tabelle 2 verzeichnet dieselbe Verteilung in relativen Zahlen. Hier erkennt man, daß die relativ häufigsten deutschen Wortverbindungen der sechs- und siebenphonemige Zweimorpher und der acht- und neunphonemige Dreimorpher sind (= 40.21 % des Gesamtmaterials). Nimmt man noch die Kombinationen 4/1, 5/1, 5/2, 8/2, 7/3 und 10/3 hinzu, so bilden diese 10 Kombinationen (= 5.95 % der 168 möglichen Kombinationen, bei 21 Phonemen und 7 Morphen) 73.22 % des gesamten Materials.

Die deutsche Gegenwartssprache operiert also in lexikalischer Hinsicht zu ca. 75 % mit nur 10 Phonem-/Morphtypen!

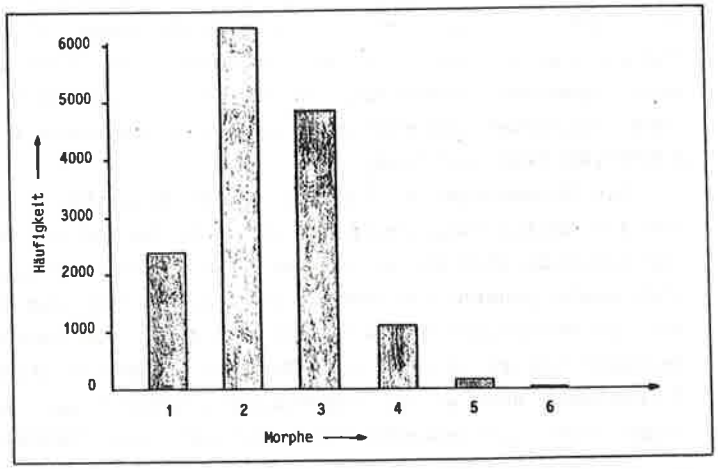


Abbildung 1: Häufigkeitshistogramm der Wortlänge (in Morphzahl)

Abbildung 1 zeigt die Häufigkeitsverteilung auf die einzelnen Morphe; Abbildung 2 veranschaulicht die Häufigkeitsverteilung auf die Phoneme; Abbildung 3 zeigt beide Häufigkeitsverteilungen in integrierter, dreidimensionaler Form.

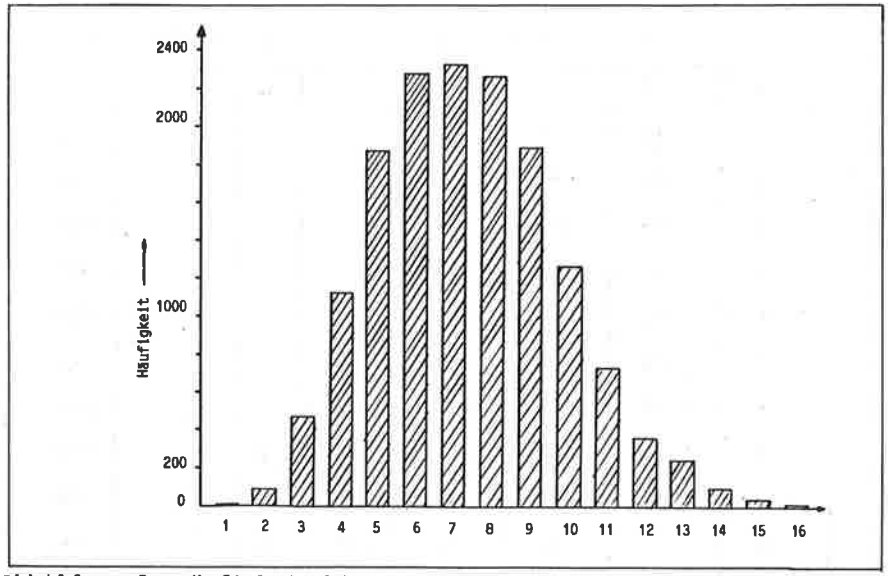


Abbildung 2: Häufigkeitshistogramm der Wortlänge (in Phonemzahl)

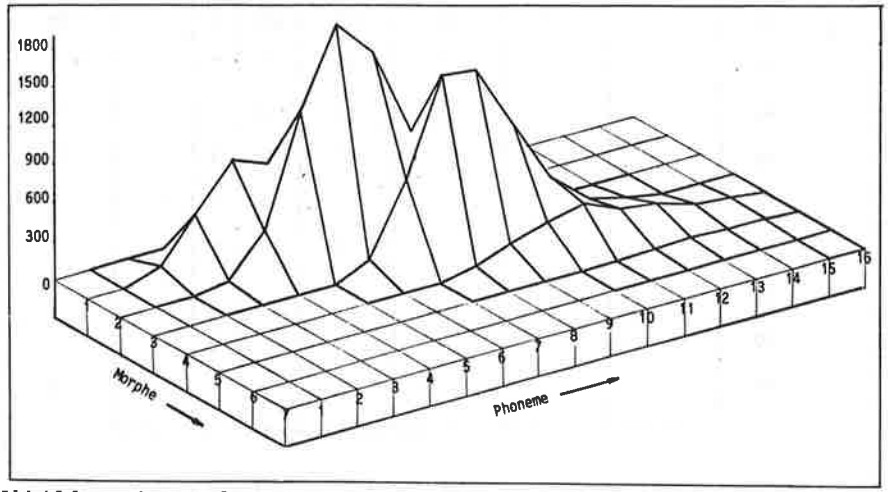


Abbildung 3: Häufigkeitsverteilungen in integrierter (dreidimensionaler) Form

Tabelle 1: Verteilung der 15 011 Stichwörter auf die einzelnen Phonem-/Morphkombinationen

Phoneme ↓ / Morphe →	1	2	3	4	5	6	7	Summe
1	3							3
2	78	2						80
3	422	38						460
4	745	370						1115
5	645	1205	12					1862
6	336	1807	128					2271
7	124	1489	702	7				2322
8	27	812	1380	42	1			2262
9	9	391	1360	142	2			1904
10	2	165	845	242	4			1258
11		55	372	293	16			736
12		8	133	186	14			341
13		1	46	144	26	2		219
14			10	68	20	2		100
15				23	18	3		44
16			1	11	3	3	1	19
17					7	1	1	9
18				1	1	1		3
19						1		1
20								
21							2	2
Summe:	2391	6343	4989	1159	112	13	4	15011

Tabelle 2: Relative Häufigkeitsverteilung der 15011 Stichwörter auf die einzelnen Phonem-Morphkombinationen

Phoneme ↓ / Morphe →	1	2	3	4	5	6	7	Summe
1	0.02							0.02
2	0.52	0.01						0.53
3	2.81	0.25						3.06
4	4.96	2.46						7.43
5	4.30	8.03	0.08					12.40
6	2.24	12.04	0.85					15.13
7	0.83	.9.92	4.68	0.o5				15.47
8	0.18	5.41	9.19	0.28	0.007			15.07
9	0.06	2.60	9.06	0.94	0.01			12.68
10	0.01	1.10	5.63	1.61	0.026			8.38
11		0.37	2.48	1.95	0.11			4.90
12		0.05	0.88	1.24	0.09			2.27
13		0.007	0.31	0.96	0.17	0.013		1.46
14			0.07	0.45	0.13	0.013		0.67
15				0.15	0.12	0.02		0.29
16			0.007	0.07	0.02	0.02	0.007	0.13
17					0.047	0.007	0.007	0.06
18				0.007	0.007	0.007		0.02
19						0.007		0.007
20								
21							0.013	0.013
Summe:	15.93	42.26	33.24	7.72	0.75	0.09	0.027	100%

Um zu überprüfen, ob der vermutete Zusammenhang zwischen Wortlänge und durchschnittlicher Morphlänge aufweisbar ist, reicht es im Grunde aus, die durchschnittliche Länge jedes Morphtyps in Betracht zu ziehen. Die beobachteten Werte sind in Tabelle 3 angegeben. Die Wortlänge x = 7 wurde wegen der kleinen Zahl der Belege ausgeschlossen.

Tabelle 3: Durchschnittliche Morphlänge in deutschen Wörtern

Wortlänge in Morphen x	1	2	3	4	5	6
Morphdurchschnittslänge y_e	4.53	3.25	2.93	2.78	2.65	2.58

Die durchschnittliche Morphlänge nimmt also ab, je höhermorphig das Konstrukt wird. Die Abnahmerate ist dabei aber nicht konstant, sondern verringert sich mit zunehmender Morphzahl. Beides veranschaulicht der graphische Abnahmeverlauf (y_e) in Abbildung 4.

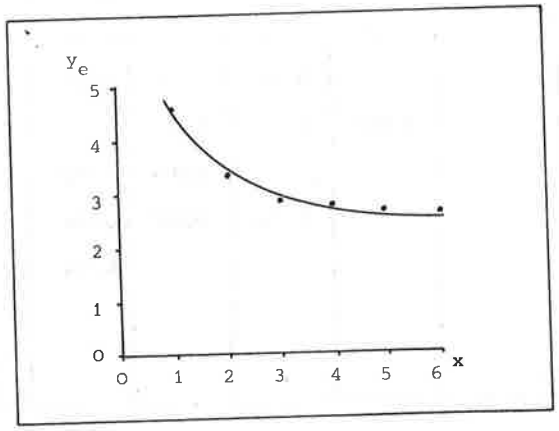


Abbildung 4: Abnahme der Morphlänge

Der vermutete Zusammenhang zwischen beiden Untersuchungsvariablen ist damit prinzipiell bekräftigt. Es bleibt nur noch zu prüfen, ob dieser Verlauf der von ALTMANN (1980, 3) abgeleiteten Funktion

$$y = A\, x^b\, e^{-cx},$$

die das Menzerath'sche Gesetz ausdrückt, entspricht. A, b und c sind Koeffizienten, die vorläufig aus den Daten geschätzt werden müssen (Formeln siehe ALTMANN 1980, 4-5). Aus den obigen Daten ergeben sich folgende Werte für die Konsonanten:

A = 4.0959
b = -0.5568
c = -0.0911

Die 'theoretischen' Werte der Morphlängen (y_t) lassen sich also aus der Funktion

$$y_t = 4.0959 \cdot x^{-0.5568} \cdot e^{0.0911x}$$

berechnen. Der theoretische Funktionsverlauf (y_t) stimmt mit den empirischen Durchschnittswerten (y_e) weitgehend überein. Das zeigt einmal die folgende Übersicht:

Wortlänge in Morphen	1	2	3	4	5	6
y_e	4.53	3.25	2.93	2.78	2.65	2.58
y_t	4.49	3.34	2.92	2.73	2.64	2.61

vor allem aber das Ergebnis des F-Tests (vgl. Anmerkung 1):

$$F_{(2,3)} = 163.55$$

Vergleicht man dieses Ergebnis mit dem tabellierten Wert $F_{0.99(2,3)} = 30.8$, so sieht man, daß die Kurve einen signifikant großen Teil der Datenvariabilität erklärt. Man kann die Gestaltung der deutschen Morpheme als eine Bestätigung des Menzerath'schen Gesetzes betrachten.

Anmerkungen

1 Die Koeffizienten müssen aus den Daten geschätzt werden, weil bisher keine theoretischen Ableitungen vorliegen. Der F-Test wird mit der linearisierten Funktion:

$$\ln y = \ln A + b \cdot \ln x - cx$$

durchgeführt. Man nimmt an, daß die Bedingungen für eine multiple Regression erfüllt sind. Da nur mittlere Längen von Interesse sind, verliert man zwar viele Freiheitsgrade und daher auch eine höhere Signifikanz, erspart sich aber auch umfangreiche Rechenarbeit.

ALTMANN, G.
1980 Prolegomena to MENZERATH's Law. In: GROTJAHN, R. (Ed.), Glottometrika 2. Bochum, Brockmeyer, 1-10.

AUGST, G.
1975 Lexikon zur Wortbildung, 3 Bde. Tübingen, Narr.

GREENBERG, J.H.
1960 A quantitative approach to the morphological typology of languages. IJAL, 26, 178-194.

MORCINIEC, N.
1958 Zur phonologischen Wertung der deutschen Affrikaten und Diphthonge. ZPSK, 11, 49-66.

1968 Zur Ein- und Zweiphonemigkeit in der deutschen Sprache. Linguistics, 41, 64-74.

WAHRIG, G. (Hg.)
1978 dtv-Wörterbuch der deutschen Sprache. München, Deutscher Taschenbuch Verlag.

WERNER, O.
1972 Phonemik des Deutschen. Stuttgart, Metzler.

DAS MENZERATHSCHE GESETZ AUF SATZEBENE

R. Köhler, Essen

1. Das von ALTMANN in [1] formulierte und von ihm so getaufte Menzerathsche Gesetz besagt, daß die Länge von linguistischen Einheiten (Komponenten) eine Funktion der Länge des aus ihnen gebildeten sprachlichen Konstrukts ist. Als Differentialgleichung hat das Gesetz die Form:

$$\frac{y'}{y} = - c + \frac{b}{x} , \qquad (1)$$

wobei mit x die Länge des Konstrukts und mit y die Länge seiner Komponenten bezeichnet wird. y' ist die erste Ableitung von y.

Die Differenzialgleichung (1) steht für ein *stochastisches* Gesetz; es erfaßt also Tendenzen. Bei der Datengewinnung für empirische Untersuchungen zum Menzerathschen Gesetz müssen daher Mittelwerte verwendet werden. Wir verfahren dabei folgendermaßen:

Eine Stichprobe von N Sprachkonstrukten (Silben, Wörtern, Sätzen ...) besteht aus n_1 Konstrukten der Länge x_1, n_2 Konstrukten der Länge x_2, usw. bis n_k Konstrukten der maximalen beobachteten Länge x_k, wobei die Länge als Zahl der unmittelbaren Komponenten angegeben wird (also z.B. Wortlänge als Anzahl der Silben, und nicht als Anzahl der Laute). Die mittlere Länge y_i der Komponenten eines Konstrukts der Länge x_i wird als mittlere Anzahl direkter Bestandteile der Komponenten dieser Konstrukte berechnet:

$$y_i = \sum_{j=1}^{n_i} \frac{y_{ij}}{n_i x_i} ; \qquad 1 \leq i \leq k , \qquad (2)$$

wo y_{ij} die beobachtete Anzahl von direkten Bestandteilen der Komponenten eines x_i-langen Konstrukts ist.

2. Auf der Satzebene folgt aus dem Menzerathschen Gesetz die Hypothese, daß

> *"in long sentences (measured in number of clauses) the clauses are shorter and vice versa ..."* (i)
> (s. [1], S. 9).

Anhand englischer Texte[1] wurde die Hypothese (i) auf folgende Weise empirisch überprüft: Aus den Sätzen der nichtwörtlichen Rede wurde eine zufällige Stichprobe gezogen. Die Satzlänge wurde in Teilsätzen (als Kriterium für einen Teilsatz wurde das Vorhandensein eines finiten Verbs gewählt[2]), die Teilsatzlänge in Wörtern gemessen. Der Umfang der Stichprobe betrug N = 843 Sätze; zwei beobachtete Sätze der Länge 7 wurden wegen zu geringer Frequenz von der Betrachtung ausgeschlossen, so daß $1 \leq x_i \leq 6$ und N = 841.

Aus den sechs Einzelverteilungen (vgl. Abb. 1) wurden die Mittelwerte (s. Tab. 1) errechnet. Diese sechs Messungen wurden

Tabelle 1: Mittelwerte des Datenmaterials

x_i	y_i
1	9.7356929779
2	8.3773231506
3	7.3511447906
4	6.7656250000
5	6.1466665268
6	6.2424242424

als empirische Werte für die mittleren Teilsatzlängen in Sätzen mit x_i Teilsätzen verwendet.

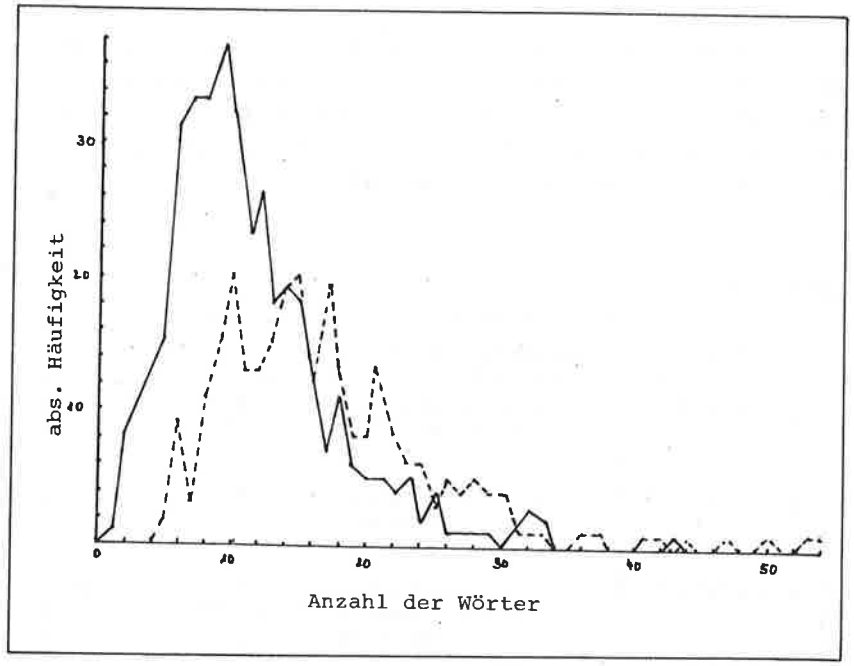


Abbildung 1: Verteilung der Häufigkeit von Teilsatzlängen in Sätzen mit einem (——) und mit zwei (----) Teilsätzen

ALTMANN gibt in [1] die Lösung der Differentialgleichung (1)

$$y = Ax^b e^{-cx} \tag{3}$$

mit den Sonderfällen (4) und (5) für b = 0 bzw. c = 0 an:

$$y = Ae^{-cx} \tag{4}$$

$$y = Ax^b \ . \tag{5}$$

Obwohl die Vermutung naheliegt, daß auf der Satzebene die monoton fallende hyperbolische[3] Funktion (5) zutrifft (Null-Längen kommen nicht vor; auch gibt es keinen Hinweis darauf, daß besonders lange Sätze aus wieder längeren Teilsätzen ge-

bildet werden), wurden alle drei Funktionen nach [1] mit der Methode der kleinsten Fehlerquadrate angepaßt und mit dem F-Test überprüft. Die Ergebnisse (s. Tab. 2) bestätigen die Hypothese (i), aber auch unsere Vermutung, daß (5) gegenüber (3) völlig ausreicht (s. dazu auch Abb. 2).

Tabelle 2:

Funktion	abgeschätzte Parameter	F-Test
$y = e^{\alpha}e^{-cx}$	$\alpha = 2.31595370$ $c = 0.09240026$	$\hat{F}_{1,4} = 48.988$
$y = e^{\alpha}x^{b}$	$\alpha = 2.28736452$ $b = -0.26885588$	$\hat{F}_{1,4} = 217.81$
$y = e^{\alpha}x^{b}e^{-cx}$	$\alpha = 2.28939701$ $b = -0.25784883$ $c = 0.00402919$	$\hat{F}_{2,3} = 82.199$
$F_{1,4(0.05)} = 7.709$; $F_{2,3(0.05)} = 9.552$		
$A = e^{\alpha}$		

3. Besteht ein gesetzmäßiger Zusammenhang zwischen zwei oder mehreren Größen, so gilt er natürlich unabhängig davon, in welchen Einheiten diese Größen gemessen werden. Auch linguistische Einheiten wie Satz, Teilsatz, Wort, Morphem usw. sind konventionell festgelegt (bzw. festlegbar). Zur Illustration können Tabelle 3 und Abbildung 3 dienen, die auf den gleichen Daten wie Tabelle 2 und Abbildung 2 beruhen, nur daß neben den finiten Verben jetzt auch Partizipien mit Objekt oder Präpositionalphrase als Teilsatzindikatoren galten. Die Ergebnisse bestätigen wieder die Hypothese (i) und die Funktion (5).

Solange die festgelegten Einheiten konsistent verwendet werden, tangiert ihre Auswahl nicht die Anwendbarkeit von Ge-

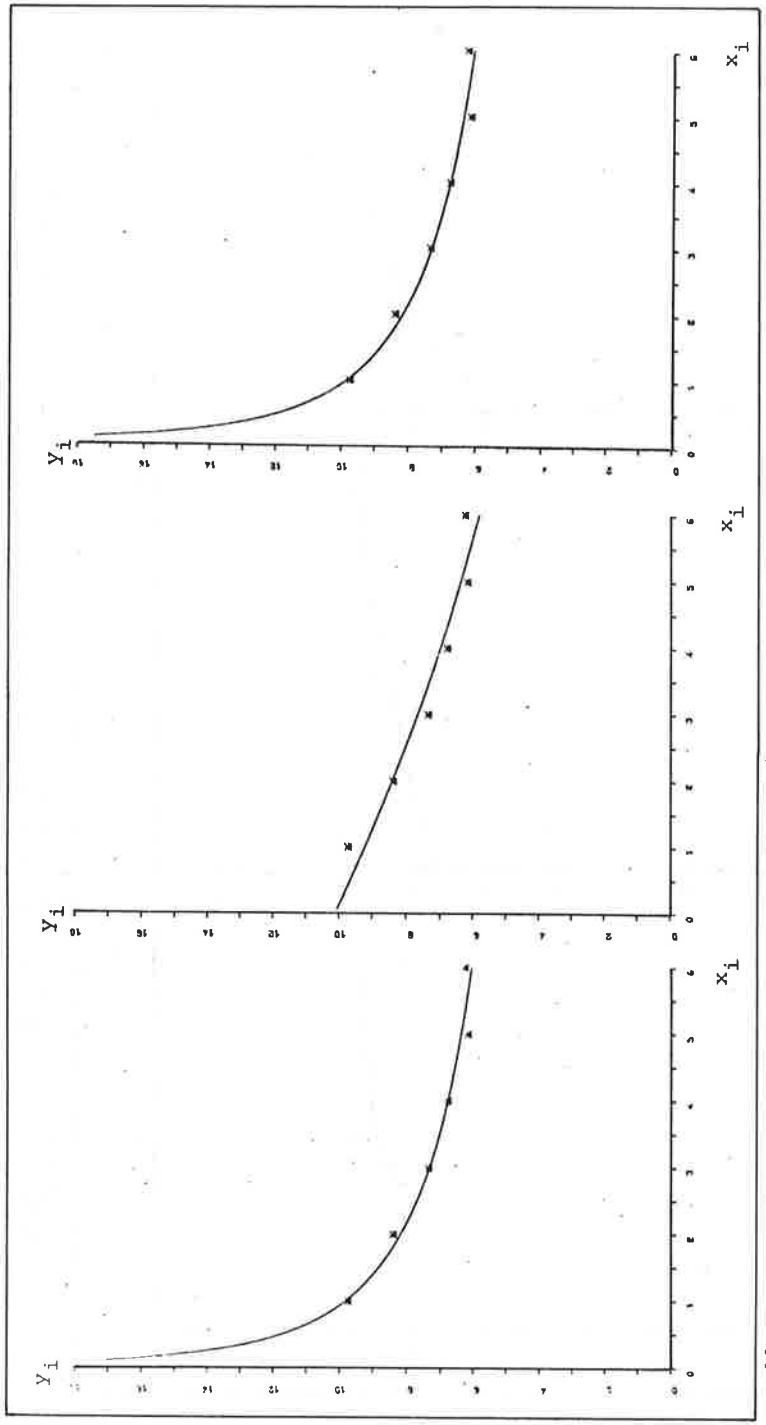


*Abb. 2: Stichprobe Crispin; die Graphen der Funktionen (3), (4) und (5) mit den Parametern nach Tab. 2 . Die empirischen Werte wurden mit * gekennzeichnet.*

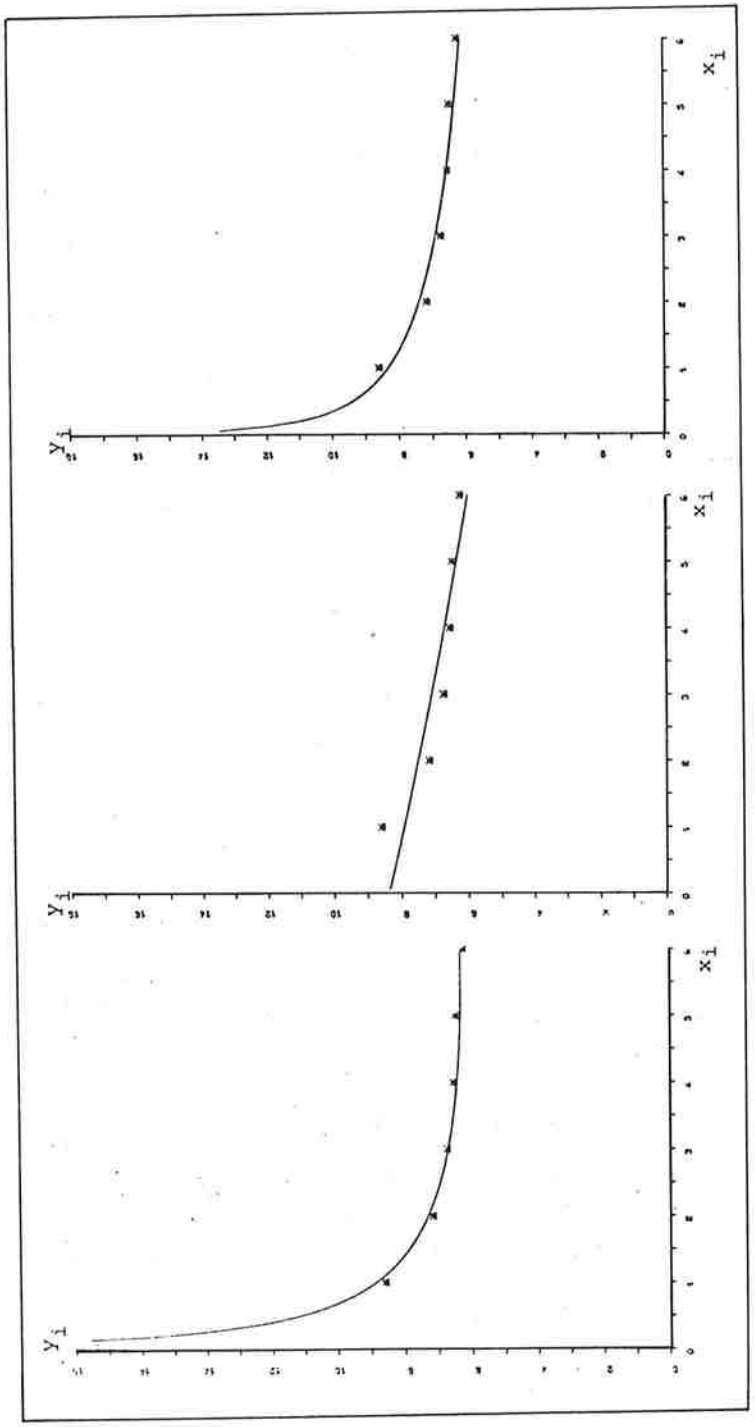


*Abb.3: Stichprobe Crispin; die Graphen der Funktionen (3), (4) und (5) mit den Parametern nach Tab.3 . Die empirischen Werte wurden mit * gekennzeichnet.*

Tabelle 3:

Funktion	abgeschätzte Parameter	F-Test
$y = e^{\alpha}e^{-cx}$	α = 2.21828914 c = 0.05537468	$\hat{F}_{1,4}$ = 15.770
$y = e^{\alpha}x^{b}$	α = 2.12151069 b = -0.17056614	$\hat{F}_{1,4}$ = 75.153
$y = e^{\alpha}x^{b}e^{-cx}$	α = 2.11722649 b = -0.37297114 c = -0.06659070	$\hat{F}_{2,3}$ = 54.177
$F_{1,4(0.05)}$ = 7.709; $F_{2,3(0.05)}$ = 9.552 $A = e^{\alpha}$		

setzen. Man kann prinzipiell Umrechnungsformeln für die Konvertierung der Einheiten angeben. Es kann aber vermutet werden, daß die Differentialgleichung (1) und die Funktionen (3) bis (5) nicht mehr in dieser Form gelten, wenn in Einheiten gemessen wird, die einer anderen linguistischen Ebene angehören, d.h. wenn Ebenen übersprungen werden (Messung der Satzlänge in Morphemen o.ä.). Ließe sich diese Vermutung empirisch erhärten, könnte dies als Nachweis einer psycholinguistischen Realität der linguistischen Ebenen gelten, da in die Definition der Einheiten ihre Länge in keiner Weise eingegangen ist. Die Befunde von GERŠIĆ und ALTMANN in [3] bezüglich der Abhängigkeit der mittleren Lautdauer von der Wortlänge (in Lautanzahl), wo also die Silbenzahl nicht berücksichtigt wurde, sind bereits eine solche Bestätigung.

4. Eine wesentliche Aufgabe nach der Aufstellung funktionaler Beziehungen wie (3) bis (5) besteht in der *allgemeinen* Bestimmung bzw. in der linguistischen Interpretation der Parameter. Im Fall der Funktion (5), die wir für die Satzebene postulieren, können wir den Parameter α bzw. $A(A = e^{\alpha})$ direkt

interpretieren:

$$A = y_1 \quad . \tag{6}$$

Der Parameter A steht für die *primäre Teilsatzlänge*, für die Länge, die der Kürzungstendenz unterliegt; d.h. A ist die mittlere Teilsatzlänge in Sätzen mit nur einem Teilsatz. Mit zunehmender Teilsatzzahl wird von ihr nach unten abgewichen.

Ein Blick auf die Funktionsgleichung von (5) nach Einsetzen der Teilsatzzahl 1 für x_1 zeigt die formale Richtigkeit der Interpretation:

$$y_1 = A \cdot 1^b = A \quad . \tag{7}$$

Der Parameter b, der die Steilheit der Kürzung beschreibt, muß weiterhin als empirische Konstante angesehen werden also als sprach- und evtl. textspezifisch.

5. Die bisher gewonnenen Ergebnisse bedürfen weiterer Überprüfung; daher wurden anhand von zwei Texten in deutscher Sprache, ebenfalls auf Satzebene, Daten erhoben[4]. Aus den Anfangskapiteln von U. JOHNSONs "Das Dritte Buch über Achim"[5] und O. F. BOLLNOWs "Maß und Vermessenheit des Menschen"[6] wurden die mittleren Teilsatzlängen y_i für die Satzlängen x_i aus den Sätzen errechnet, die keine wörtliche Rede, Verse o.ä. enthalten. Von der Betrachtung ausgeschlossen wurden Sätze der Länge $x_i > 6$ (JOHNSON) bzw. der Länge $x_i > 5$ (BOLLNOW), da sie (pro Satzlänge) weniger als 10 Vorkommen aufweisen ($n_i < 10$). Tabelle 4 und Tabelle 5 enthalten die so gewonnenen Daten, die abgeschätzten Parameter und die Ergebnisse der F-Tests[7]; die Anpassung der Funktion ist in den Abbildungen 4 und 5 dargestellt.

Tabelle 4: (JOHNSON)

x_i	y_i (beobachtet)	$\hat{y}_i$ (berechnet)	n_i
1	9.26443672	9.28773880	329
2	8.57024765	8.42374992	242
3	7.98333263	7.95610332	120
4	7.35087641	7.64013386	57
5	7.35714245	7.40371704	28
6	7.41176414	7.21599007	17
α = 2.22869539 b = -0.14086473			N = 793
$\hat{F}_{1,4} = 68.036 > F_{1,4(0.05)} = 7.709$			

Tabelle 5: (BOLLNOW)

x_i	y_i (beobachtet)	$\hat{y}_i$ (berechnet)	n_i
1	13.95406342	13.50633430	283
2	10.82352829	11.28818893	204
3	10.01626015	10.16360855	123
4	9.16326523	9.43432999	49
5	9.39130402	8.90489197	23
α = 2.60315895 b = -0.25882214			N = 682
$\hat{F}_{1,3} = 48.298 > F_{1,3(0.05)} = 10.128$			

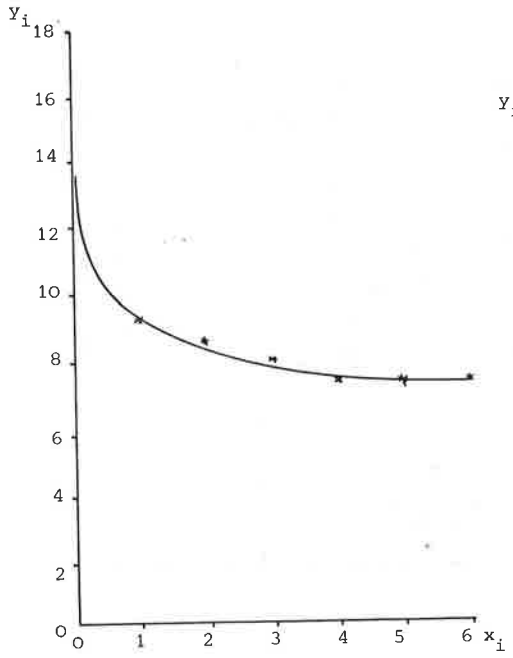


Abbildung 4: Stichprobe JOHNSON; Graph der Funktion (5) mit den Parametern aus Tab. 4.

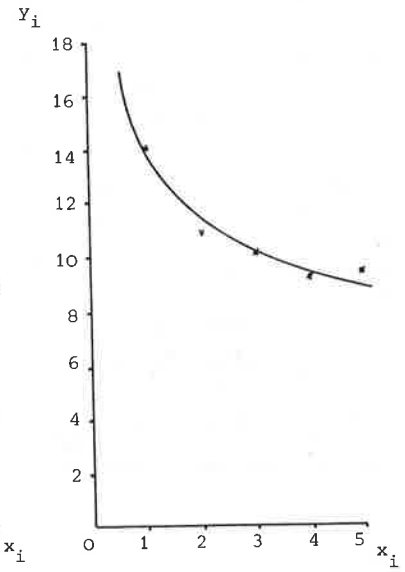


Abbildung 5: Stichprobe BOLLNOW; Graph der Funktion (5) mit den Parametern aus Tab. 5.

Die empirischen Werte sind in beiden Abbildungen mit * gekennzeichnet.

Wiederum wurdem alle Annahmen bestätigt, womit auch die Interpretation des Parameters A gestützt ist. Der Parameter b erwies sich als textabhängig.

6. Mit der Annahme der Hypothese, daß die Funktion (5) die Kürzungstendez auf der Satzebene beschreibt, ist die Notwendigkeit verbunden, die Differentialgleichung (1) zu reinterpretieren. ALTMANN ging bei der Aufstellung der Gleichung in [1] davon aus, daß das Glied -c die Kürzung repräsentiert, und das Glied $\frac{b}{x}$ mögliche entgegengesetzte Einflüsse erfaßt. Die Funktion (5) entspricht aber gerade dem zweiten Glied, so daß die vorliegenden Resultate gegen eine konstante Kürzungsrate, wenigstens auf der Satzebene, sprechen. Weitere empirische Arbeiten, besonders auf anderen linguistischen Ebenen, scheinen zur Klärung nötig zu sein.

Anmerkungen

1 Es wurden die folgenden vier Bände ausgewertet: CRISPIN (Ed.): Best SF 2 (1956), 3 (o.J.), 4 (1960) und 5 (1971), London.

2 Zur Definition des Teilsatzes vgl. [2], S. 27.

3 Wegen b < 0, da ja Kürzung angenommen wird.

4 Für ihre Unterstützung bei der Datenverarbeitung danke ich Frau B. Stöckmann.

5 JOHNSON, U., Das dritte Buch über Achim, Frankfurt a.M., 1961.

6 BOLLNOW, O.F., Maß und Vermessenheit des Menschen. Philosophische Aufsätze, Göttingen 1962.

7 Alle Tests wurden ungewichtet ausgeführt, da die Ergebnisse auch ohne Berücksichtigung der n_i hohe Signifikanz aufwiesen.

LITERATUR

[1] ALTMANN, G.
1980 Prolegomena to Menzerath's Law. In: GROTJAHN, R. (Ed.), Glottometrika 2. Bochum, 1-10.

[2] PIEPER, U.
1979 Über die Aussagekraft statistischer Methoden für die linguistische Stilanalyse. Tübingen.

[3] GERŠIĆ, S., ALTMANN, G.
1980 Laut - Silbe - Wort und das Menzerathsche Gesetz. In: Forum Phoneticum 21. Frankfurt a.M., 115-123.

Phonological Patterning in English and German Verse: A Computer-assisted Approach

David H. Chisholm, Arizona

Introduction

A few years ago I investigated the sound repetition patterns in a corpus of nineteenth century German iambic pentameter verse with the aid of two computer programs written in SNOBOL 4. What distinguished this investigation from previous phonological studies of German verse was a procedure which takes into account not only conventional manifestations of phonological recurrence, such as end-rhyme, alliteration and assonance, but *all* recurrence patterns in a given poem or prose selection. This procedure treats all patterns of equivalence in binary terms, and distinguishes recurrence pairs according to the size of the gap between the two syllables in which the sound occurs. The analysis of the German corpus, which has since been expanded to include early twentieth century as well as nineteenth century poets, yielded some striking results which distinguish the sound patterning of metered verse, free verse and prose, phonostylistic characteristics of the work of individual poets and diachronic changes in nineteenth and twentieth century verse.

The fruitfulness of this approach to German verse suggested its application to English verse of roughly the same period to determine to what extent the predominant phonological patterns of English verse differ from, or are similar to, those which prevail in German. Application of this type of computer-assisted approach to English verse also enables the investigator to isolate those phonological patterns which distinguish the verse of individual poets and periods. Moreover, it provides a reliable basis for evaluation of literary studies which discuss the sound structure of English verse.

1. Phonological Frames

The procedure, which has been described in CHISHOLM (1976), may be summarized as follows: Pairs of equivalent sounds are distinguished on the basis of proximity by assigning each pair to the appropriate *phonological frame*. Equivalence between contiguous syllables is assigned to Frame One, between alternate syllables to Frame Two, and so forth. In each phonological frame, all pairs of syllables are scanned for equivalence. In the line given in (1), the three instances of prevocalic /m/ generate *three pairs* of equivalent sounds which are assigned to different frames:

(1) Maiden and mistress of the months and stars

Frame 3: Mai-/ mis-
Frame 4: mis-/ months
Frame 7: Mai-/ months

In this way, patterns of equivalence are identified in binary terms within a positional framework. One restriction should be noted: equivalence is recognized only between *corresponding parts* of syllables; in the sequence "ma*iden and*," for example, the postvocalic /n/'s constitute an equivalence, while the two /d/'s, one being prevocalic and the other postvocalic, are not recognized as a pair. Each poem is viewed as a single long string of syllables, so that the two members of a pair may be located in different lines. The patterning in the German poems attests to the theoretical soundness of distinguishing *vertical equivalence between lines* from all other equivalence. The frames which include vertical equivalence are called *primary frames*; all other frames are *secondary frames*. Primary Frame 1 includes all vertical correspondence between successive lines, Primary Frame 2 all vertical correspondence between alternate lines, and so forth. This distinction, of course, is meaningful only for verse, in which the "line" functions as a rhythmic entity.

2. The Data

The corpus of English iambic pentameter verse selected for this investigation consists of forty sonnets by four nineteenth century English poets - Wordsworth, Keats, Rossetti and Swinburne - and 148 lines of blank verse from Swinburne's *Atalanta in Calydon* which is divided into ten fourteen- and sixteen-line segments. The data thus provide a fairly representative sampling of English iambic pentameter verse from both the Romantic period (Wordsworth, Keats) and the Victorian period (Rossetti, Swinburne). The inclusion of the blank verse passages from *Atalanta in Calydon* (lines 1-64 and 121-204) enables us to compare rhymed and unrhymed iambic pentameter verse. Furthermore, it is the text which B.F. SKINNER (1941) used in support of his theory of "formal perseveration" and it can therefore be used as a means of testing that theory. A corpus of fifty nineteenth century German sonnets be Goethe, Schlegel, Brentano, Heine and Mörike is included for a general comparison of phonological patterning in nineteenth century German and English verse.

3. Transcription of the Data

The phonemic transcription of the English data follows the phonetic description of English Received Pronunciation given by Daniel JONES (1956) and A.C. GIMSON (1962). The forty-five phonemes of English described by JONES and GIMSON are given in (2) with examples of words in which they occur, and the keypunch character used for each phoneme. In this investigation the diphthongs /ei, ai, oi, ou, au/ and the affricates /tš, dž/ are treated as single phonemes. The transcription of words such as *tire*, and *tower*, which contain a diphthong followed by schwa, depends on their syllabic relation to the underlying meter: if the sequence /aiə/ or /auə/ occupies two metrical positions it is assumed to be disyllabic and is transcribed accordingly; if it occupies only a single metrical position, it is treated as monosyllabic and reduced to the diphthongal glide /a:ə/, with

(2) Phonemes of Britisch English

a)

Vowel Phonemes	Example	Keypunch Character
/i:/	see	1
/ɪ/	bit	I
/e/	get	E
/æ/	cat	A
/ɑ:/	father	6
/ɔ/	hot	O
/ɔ:/	saw	8
/u/	put	U
/u:/	too	7
/ʌ/	but	^ (upper case Q)
/ə:/	bird	9
/ə/	china	ə
/eɪ/	day	EI
/aɪ/	die	AI
/oɪ/	boy	OI
/ou/	go	OU
/au/	how	AU
/iə/	here	Iə
/ɛə/	there	Eə
/uə/	tour	Uə
/ɑ:ə/	tire, hour	6ə (monosyllabic form of /aiə, auə/)

(N.B. ə corresponds to ɑ in (3) b) and to ö in Figure 1(cont.).)

b)

Consonant Phonemes	Example	Keypunch Character
/p/	pole	P
/b/	bowl	B
/t/	toll	T
/d/	dole	D
/k/	coal	K
/g/	goal	G
/tš/	cheap	C
/dž/	jeep	X
/f/	feel	F
/v/	veal	V
/θ/	thigh	Q
/ð/	thy	]
/s/	seal	S
/z/	zeal	Z
/š/	vicious	(
/ž/	vision	)
/h/	heal	H
/m/	meat	M
/n/	neat	N
/ŋ/	sing	/
/l/	leap	L
/r/	reap	R
/w/	weep	W
/j/	year	J

the result that *tire* and *tower* become homophones. This approach is consistent with GIMSON's interpretation of /aiə/ and /auə/ and the levelling of /a:ə/ and /ɑ:ə/ into a single phoneme (1962: 133).

Stress is treated as a binary opposition + STRESS/-STRESS.
+ Stress is assigned to:

(1) all syllables bearing primary stress in polysyllabic words and all monosyllabic nouns, non-auxiliary verbs and adjective

(2) all syllables bearing primary stress *relative to their immediate morphological context* in words of three or more syllables (e.g. *grandfa*ther, apple*tree*, *work*man*ship*, over*hear*).

In the input data, all stressed syllables are preceded by a plus sign.

The *syllable* may be identified as a vowel nucleus optionally preceded and/or followed by one or more consonants; when two vowels are separated by a cluster of consonants, however, it is often difficult to determine where one syllable ends and the next begins. GIMSON gives as an example the word *extra* /'ekstrə/, in which the phoneme /s/ could belong either to the first or second syllable and still be in accord with the phonological rules of the English language. For computer-assisted analysis syllable-boundary is determined in the following necessarily somewhat arbitrary way. It is located:

(1) between words (e.g. *first day*)

(2) immediately before all free lexical items (e.g. *first-born*) and derivational suffixes beginning with a consonant (e.g. *first-ly*)

(3) in the middle of consonant clusters where neither (1) nor (2) apply. Such clusters are divided evenly, with any extra consonants placed to the right of the syllable boundary (e.g. /'eks-trə, 'læs-tiŋ, 'rai-diŋ, ik-'spænd/).

If word-final orthographic *r* is followed by a word-initial vowel (so called "linking-r"), it is transcribed as a post-vocalic /r/ (e.g. *far away* /fɑ:r ə'wei/ and the word-boundary serves as the syllable boundary.[1] In the input data, syllable-boundaries are marked by an asterisk. As an illustrative example, the machine-readable transcription of a sonnet by Keats is given in (3).

(3) John Keats, "Keen, fitful gusts ..."

a) Orthographic Representation:

Keen, fitful gusts are whispering here and there
  Among the bushes half leafless, and dry;
  The stars look very cold about the sky,
And I have many miles on foot to fare.
Yet feel I little of the cool bleak air,
  Or of the dead leaves rustling drearily,
  Or of those silver lamps that burn on high,

Or of the distance from home's pleasant lair:
For I am brimful of the friendliness
  That in a little cottage I have found;
Of fair-haired Milton's eloquent distress,
  And all his love for gentle Lycid drowned;
Of lovely Laura in her light green dress,
  And faithful Petrarch gloriously crowned.

b) Phonological Transcription:

```
&JOHN KEATS 06, PAGE 297B "KEEN, FITFUL GUSTS..."
&
1.  +KIN, +FIT*FUL +G^STS 6 +WIS*PRI/ HIα AND JEα
2.  α**M^/ Jα +BU*CIZ, +H6F +LIF*LIS AND +DRAI;
3.  Jα +ST6Z +LUK +VE*RI +KOULD α**BAUT Jα +SKAI,
4.  AND AI HAV +ME*NI +MAILZ ON +FUT T7 +FEα,
5.  JET +F1L AI +LI*TαL OV Jα +K7L +BLIK +Eα,
6.  8 OV Jα +DED +L1VZ +RUS*LI/ +DRIα*RI*LI,
7.  8 OV JOUZ +SIL*Vα +LAMPS JAT +B9N ON +HAI,
8.  8 OV Jα +DIS*TαNS FROM +HOUMZ +PLE*SαNT +LEα:
9.  F8 AI AM +BRIM*FUL OV Jα +FREND*LI*NIS
10. JAT IN α +LI*TαL +KO*TIX AI HAV +FAUND;
11. OV +FEα*HEαD +MIL*TαNZ +E*LO*KWαNT DIS**TRES
12. AND 8L HIZ +L^V F8 +XEN*TαL +LI*SID +DRAUND,
13. OV +L^V*LI +L8*Rα IN H9 +LAIT +GRIN +DRES,
14. AND +FEIQ*FUL +PI*TR6K +GL8*RI*αS*LI +KRAUND.
```

The transcription of the German texts, based on Moulton's *Sounds of English and German*, Siebs' *Deutsche Aussprache*, and the Duden *Aussprachewörterbuch*, has been described elsewhere (CHISHOLM 1976: 7).

4. The Programs

The two SNOBOL 4 programs initially used for German verse and prose may be utilized for English by simply changing the inventory of consonant and vowel phonemes. (In this way the programs can be used for phonological studies of virtually any language.) The *Phoneme Frequency Program* computes the absolute and relative frequency of all sounds in prevocalic, vocalic,

and postvocalic position in stressed and unstressed syllables. The *Phonological Frame Program* identifies all recurrence-pairs in a given number of frames and prints the results in graphic and statistical form. A sample of the output of the Phonological Frame Program is given in Figure 1 (a and b), which shows the graphic representation and statistics for Frame Two of the transcribed Keats sonnet in (3). The graphic representation shows only those syllables which contain equivalences in this frame. Comparison of the graphic representations of the different frames of a poem gives a general impression of its phonological development from beginning to end. The statistics, which can be derived from the graphic representation, show the frequency of sound repetition pairs in stressed and unstressed syllables (labelled S and U) in prevocalic, vocalic and postvocalic position. The amount of phonological equivalence in each frame is expressed as a ration of the number of sound repetition pairs to the numbers of syllable pairs in that frame.

5. The Consonant-Vowel Ratio in German and English

Table 1, based in part on the output of the frequency program, gives for both English and German the absolute and relative frequency of vowels and consonants in (a) the phoneme inventory and (b) in stressed and unstressed syllables in nineteenth century verse. Since the phonemic inventory of English is substantially larger than that of German, we can expect *less phonological recurrence* in English. Note also that English, due largely to the loss of the unstressed vowel /ə/ in the thirteenth and fourteenth centuries and concomitant increase in lexical monosyllables, has a much lower proportion of *unstressed syllables* than German. Furthermore, due in part to the same type of historical process (loss of unstressed inflectional morphemes) English has a lower proportion of unstressed *consonants* than German. Comparison of the English sentence "*We* stay *with our* old friends" with its German equivalent "*Wir* blei*ben bei unseren alten* Freun*den*" (unstressed syllables underlined)

Figure 1: John Keats, "Keen fitful gusts ..."

a) Statistics for Frame Two

	Prevocalic			Vocalic			Postvocalic			Total		
	R	Syl	R/S	R	Syl	R/S	R	Syl	R/S	R	Syl	R/S
S-S:	7	32	.219	0	34	.000	2	11	.182	9	34	.265
S-U:	1	9	.111	1	23	.043	1	10	.100	3	23	.130
U-S:	0	10	.000	1	22	.045	0	11	.000	1	22	.045
U-U:	2	36	.056	9	59	.153	1	16	.063	12	59	.203
TOTAL:	10	87	.115	11	138	.080	4	48	.083	25	138	.181

R = Number of Sound Repetition Pairs
Syl = Number of Syllable Pairs
R/Syl = Ratio of Sound Repetition Pairs to Syllable Pairs
S = Stressed Syllable
U = Unstressed Syllable


```
GRAPHIC REPRESENTATION FOR FRAME 2
**********************************

      1        2        3        4        5        6        7        8        9        10       11
*****************************************************************************************************
1               +FIT             +G^STS           +WIS
2     Ö                  JÖ
3                                                                 Ö                 JÖ       +SKAI
4     AND       AI       HAV      +ME              +MAILZ           +FUT             +FEÖ
5               +FIL                       TÖL              JÖ
6                                          +L1VZ    +RUS    LI/      +DRIÖ    RI
7
8                        JÖ               TÖNS                       +PLE             +LEÖ
9                                                                                    NIS
10              IN       Ö        +LI      TÖL              TIX              HAV      +FAUND
11    ÖV        +FEÖ
12
13              +L^V             +L8                                          +GRIN
14    AND                                   TR6K             RI               LI
*****************************************************************************************************
```

Figure 1 (cont.): Keats, "Keen, fitful gusts ..."

Table 1: Vowels and Consonants in English and German Verse

A. Phonemic Inventory of the Language

	English		German	
Vowels	21	46.67 %	18	47.37 %
Consonants	24	53.33 %	20	52.63 %

B. Frequency in Verse

	Four English Poets		Five German Poets	
Stressed Vowels	2380	32.66 %	2774	34.52 %
Stressed Consonants	4908	67.34 %	5262	65.48 %
Stressed Phonemes	7288	100.00 %	8036	100.00 %
Unstressed Vowels	3254	40.62 %	4807	36.53 %
Unstr. Consonants	4757	59.38 %	8351	63.47 %
Unstressed Phonemes	8011	100.00 %	13158	100.00 %
Total Vowels	5634	36.83 %	7581	35.77 %
Total Consonants	9665	63.17 %	13613	64.23 %
Total Phonemes	15299	100.00 %	21194	100.00 %
Total Stressed Syllables	2380	42.24 %	2774	36.59 %
Total Unstressed Syllables	3245	57.76 %	4807	63.41 %
Total Syllables	5634	100.00 %	7581	100.00 %

illustrates the difference. The English sentence has three unstressed syllables containing three consonant phonemes, while the German equivalent has seven unstressed syllables containing *twelve* consonant phonemes. In stressed syllables, on the other hand, English shows a higher proportion of consonants than does German. In the sentence given above, the three stressed English syllables contain nine consonant phonemes, while the four stressed German syllables contain only seven. This is reflected in verse, where English has a higher proportion of stressed consonants than German.

6. Phonological Patterning in English Verse and Prose

Table 2 (and the accompanying graph in Figure 2) gives the frequency and distribution of phonological equivalence in Frames 1-10 of forty sonnets by Wordsworth, Keats, Rossetti and Swinburne, ten 14-16 line segments from the blank verse passages of Swinburne's tragedy *Atalanta in Calydon*, and a 420-syllable prose sample from his essay on William Blake. The frequency and distribution of fifty nineteenth century German sonnets and twenty-five fourteen-line segments from Goethe's blank verse drama *Iphigenie auf Tauris* are included for further comparison. In the four verse corpora the even frames (2, 4, 6, 8, 10) are consistently higher than the odd frames (1, 3, 5, 7, 9) which means that there is a significantly greater amount of equivalence between two metrically strong or two metrically weak positions in the iambic pentameter line than between a strong and a weak position. This contrast between even and odd frame equivalence reflects the principle of alternation which inheres in both English and German iambic verse. Frame Ten (which includes rhymes in successive lines) is higher than the other even frames in the German and English sonnets. In the *unrhymed* blank verse samples, Frame Ten is significantly higher in German (with Frame Six in second place); in English, on the other hand, Frames 6, 4 and 2 are higher than Frame Ten. This reflects a major difference in the extent of vertical patter-

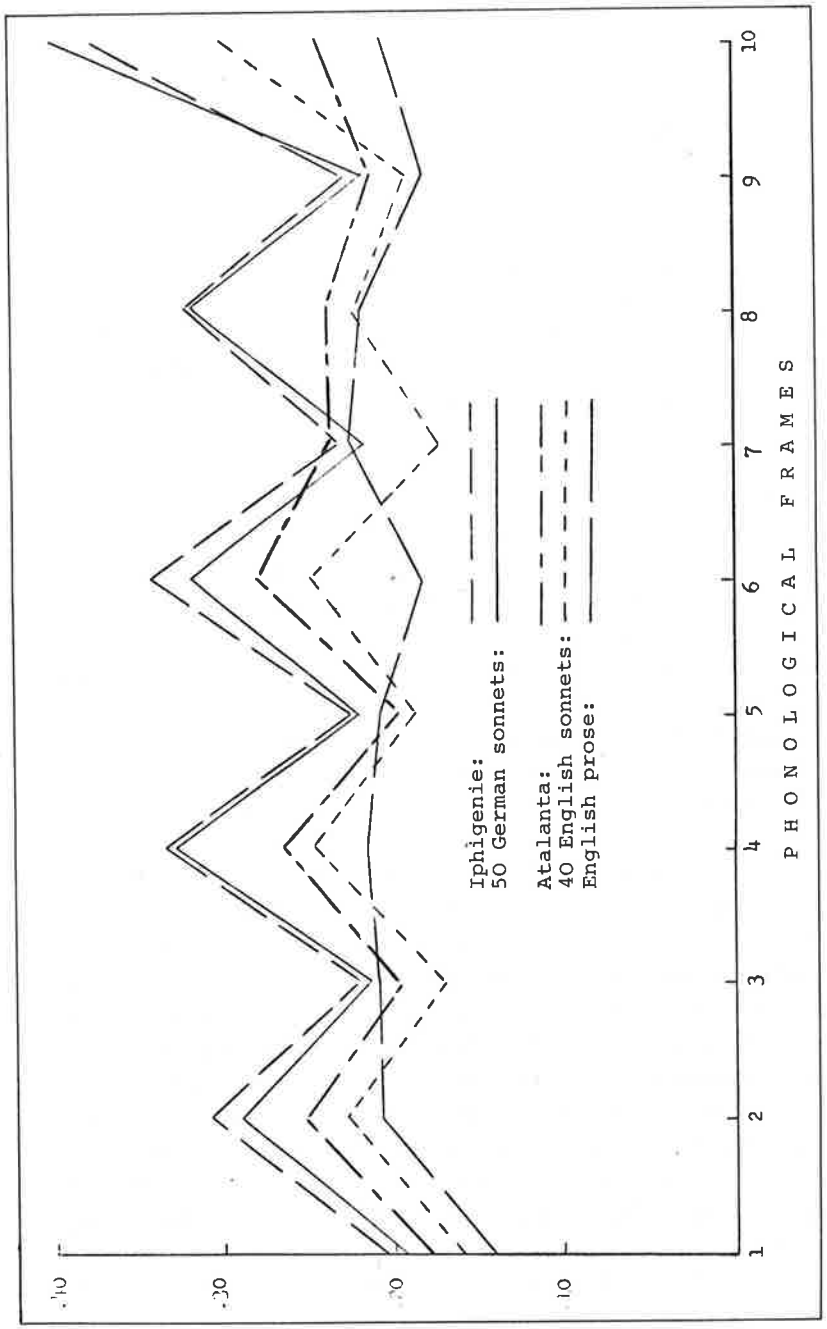


Figure 2: Ratio of Sound Pairs to Syllable Pairs in Frames 1-10 in English Poetry and Prose

Table 2: Ratio of Sound Pairs to Syllable Pairs in Frames 1-10 in German and English Iambic Pentameter Verse and a Sample of English Prose

FRAMES

	1	2	3	4	5	6	7	8	9	10	Total
GERMAN											
Five Poets	.198	.292	.217	.330	.220	.320	.219	.320	.221	.398	.273
Iphigenie	.205	.310	.221	.334	.223	.343	.235	.321	.226	.375	.279
ENGLISH											
Four Poets	.165	.229	.173	.247	.189	.244	.174	.225	.193	.299	.213
Atalanta	.182	.257	.199	.265	.197	.282	.237	.239	.213	.245	.232
Prose	.141	.208	.209	.218	.210	.182	.226	.260	.183	.208	.204

ning in English and German verse: in German verse, vertical patterning predominates *throughout the verse line*, so that unrhymed as well as rhymed iambic pentameter verse consistently shows more equivalence in Frame Ten (= Primary Frame One) than in Frames 2, 4, 6 and 8. The English blank verse of Swinburne's *Atalanta* on the other hand, does not show this tendency.

A measure of equivalence for comparing one poem or group of poems with another may be obtained by dividing the total number of sound repetition pairs in a given number of frames by the total number of syllable pairs in those frames. This measure is called the *coefficient of phonological equivalence*. Computation of the total equivalence in Frames 1-10 in the blank verse and the sonnets shows that this coefficient is somewhat *greater* in the blank verse passages than in the sonnets for both English (.299 versus .213) and German (.279 versus .273)[2]. In contrast to the graphs of the iambic pentameter poems, the graph of the prose sample shows no regular pattern

of alternation. Table 2 and Figure 2 do show one feature that both the verse and prose samples share in both languages: there is a marked tendency - particularly in prose - to avoid equivalence in Frame One. This seems to reflect a feeling among writers of English and German that excessive phonological equivalence between contiguous syllables is "bad style" and should be avoided[3].

7. Phonological Frames and Skinner's "Formal Perseveration"

In an article entitled "A Quantitative Estimate of Certain Types of Sound Patterning in Poetry", the behaviorist B.F. SKINNER argues that sound repetition in verse is merely an example of what he calls "formal perseveration" or "formal strenghtening". He describes this principle by saying that "when a speech-sound is once emitted, the probability of emission of that sound is temporarily raised" (1941: 64), but then falls to the chance level after a certain number of intervening syllables. To support this general claim, SKINNER analyzed alliteration and assonance in the first 500 iambic pentameter lines of Swinburne's "Atalanta in Calydon". He restricted his investigation to accented syllables, thereby omitting over 40 % of the linguistic material, and claimed that "the omitted material may be regarded either as not contributing in any effective way to the sound pattern or as not permitting any variation in the behavior of the writer" (p. 65). It is easy to demonstrate, however, that unaccented syllables do contribute to the sound pattern and that they also allow - albeit to a lesser degree than accented syllables - stylistic variation on the part of the poet. Thus SKINNER prejudiced his results from the outset by excluding from consideration over 2000 of the 5000 syllables in his sample. Furthermore he limited his investigation of alliteration to only ten of the 24 consonant phonemes of English. The first part of his investigation merely shows that here is a lot of line-internal alliteration in Swinburne's blank verse, which is hardly a startling fact. The

second part, despite unreliability of results, is methodologically of considerable interest. In order to compute the number of syllables between alliterating sounds, SKINNER placed the initial consonants of all selected syllables in order on a long tape:

> "The tape was then passed under cards in which windows were cut to display consonants with any desired number of intervening syllables, beginning with adjacent pairs. The numbers of pairs of each consonant at distances up to that of eight intervening syllables were tallied by moving the type under each card one space at a time throughout its length." (p. 67 f.).

This procedure is essentially the same as that of the previously mentioned Phonological Frame Program, the difference being that the latter takes *all* syllables (both stressed and unstressed) and *all* phonemes (prevocalic, vocalic and postvocalic) into consideration. The essential weakness of SKINNER's procedure is that the size of the gaps between repeated sounds in his data bears no constant relationship to that of the original poem, as illustrated by three lines from *Atalanta in Calydon* given in (5):

(5)

1	2	3	4	5	6	7	8	9
For	if	*sleep*	have	no	*mer*cy	and	*mans*	*dreams*
11	12	13	14	15	16	17	18 19	20
Bite	to	the	*blood*	and	*burn*	into	the	*bone*
21	22	23	24	25	26 27	28	29	30
What	*shall*	this	*man*	do	*wa*king?	By	the	*gods*

Since SKINNER's long tape contains only the underlined syllables, his pairs having "0 intervening syllables" would have 0, 1, 2 and 3 intervening syllables in the original poem (Frames 1, 2, 3, 4), those having "1 intervening syllable" would have 1, 3, 4 and 5 (Frames 2, 4, 5 and 6), those having "2 intervening syllables" would have 4, 5, 6, 7 and 8 (Frames 5, 6, 7, 8, 9) and those having "8 intervening syllables (the largest gap considered by SKINNER) would have 19 and 20 intervening syllables (Frames 20, 21). The Phonological Frame Program gi-

ves a much more accurate picture of the distribution of sound patterns in verse and prose, since no part of the text is excluded from consideration.

Even if SKINNER had taken all syllables into account, however, there is still a fundamental flaw in his reasoning. In *Atalanta*, Swinburne is consciously making extensive use of the *conventional device* of alliteration. In order for the listener to perceive the alliterative pattern, the repeated sound must occur before the impression created by its antecedent is blotted out by the intervening phonological material. In isolation the conventions of alliteration and assonance, which require repetition of only a single phoneme, appear to be consciously perceivable as such over a span of not much more than about ten syllables, roughly the length of the iambic pentameter line. The convention of end-rhyme, on the other hand, which is positionally bound and requires the repetition of more phonetic material (usually two or more phonemes), can be perceived over a span of 2 or 3 *lines* (as in the sonnet form). Similarly, the repetition of the *refrain* in stanzaic forms may be separated from its initial occurrence by many lines, since the repeated event consists of a span of phonemes (and words) repeated in the same sequence.

All of these and similar conventions, however, are not examples of "formal perseveration" as defined by SKINNER. To determine whether sound repetition patterns in poetry reflect what SKINNER calls "formal perseveration", we need to consider non-conventional as well as conventional patterns of phonological equivalence. The Phonological Frame Program does this by taking all phonemes into consideration.

One way of determining whether "formal perseveration" is an operative principle in English verse is to compare the amount of phonological equivalence in each frame. If SKINNER's assertion is correct, we should expect the greatest amount of equivalence in Frame Two, and successively less in Frames 4, 6, 8, 10 etc.; similarly, we would expect a gradual decrease in equivalence in odd frames after Frame One. This assumption was tested by computing the equivalence in each of the first

Table 3: Highest Coefficients of Equivalence in 148 Iambic Pentameter Lines of *Atalanta in Calydon*

A. Highest Coefficients of Frames 1-20

Coefficient of Equivalence	Frame	Lines of Text
.399	2	17- 22
.373	10	49- 63
.365	4	49- 64
.350	20	1- 16
.336	20	17- 32
.328	6	163-176
.325	6	49- 64
.321	4	17- 32
.318	6	1- 16
.316	14	1- 16
.312	20	163-176
.311	9	1- 16
.311	12	1- 16
.309	8	33- 48
.308	7	135-148
.304	2	33- 48
.300	10	135-148

B. Highest Coefficients of Equivalence in Odd Frames 1-19

Coefficient of Equivalence	Frame	Lines of Text
.31	9	1- 16
.31	7	135-148
.29	13	17- 32
.29	11	49- 64
.27	13	163-176
.26	15	33- 48
.25	13	149-162
.25	1	191-204
.24	7	121-134

twenty frames in the first 148 unrhymed iambic pentameter lines of Swinburne's *Atalanta in Calydon*. As mentioned previously, the text was divided into ten segments to facilitate comparison with the sonnets. Table 3 lists in rank order those frames which show the highest equivalence (from .399 - .300) in these segments. (The frames with less equivalence than those included here range from .101 - .292.) This data offers strong counter-evidence to SKINNER's theory of "formal perseveration", since the frames which occur most frequently in this group are Frames 20 and 6, followed by 4, 2 and 10. If SKINNER's theory were correct, there would be numerous occurrences of Frames 2 and 4 at the top of the list, and Frames 10 and 20 would not even occur in this range. Taken separately, the *odd* frames yield the same kind of evidence; Frame 13 is the highest of the odd frames in three of the segments (equivalence ratio = .29, .27, .25) and Frame 7 in three segments (.31, .24, .19) while Frames 9, 11, 15 and 1 are each highest in only one segment and Frames 3 and 5 are not highest in any segment. Table 2 as well indicates that SKINNER's general assumption is incorrect: in the first ten frames of the forty English sonnets, Frames 10, 4 and 6 each have more equivalence than Frame 2. In the fifty German sonnets and in the twenty-five segments from *Iphigenie*, Frame 2 shows the lowest equivalence of the first five even frames. Nor does SKINNER's assumption hold for Swinburne's *prose*; among the first twenty frames of the prose sample from his essay on William Blake, Frames 8, 7, 15, 4 and 13 show the greatest amount of equivalence.

8. Coefficients of Equivalence in Even and Odd Frames

The graph in Figure 2 emphasizes the importance of the distinction between even and odd-numbered frames; for phonological equivalence in even frames tends to support the underlying meter, while odd-frame equivalence tends to disrupt it. Table 4 gives the coefficients of equivalence in even and odd frames for the first ten frames in the forty sonnets by Wordsworth,

Table 4: Coefficients of Equivalence in the Sonnets of Wordsworth, Keats, Rossetti and Swinburne

A. Even Frames (2, 4, 6, 8, 10)

	Sound Pairs	Syllable Pairs	Coefficient of Equivalence
Swinburne	1855	6700	.277
Rossetti	1649	6700	.246
Keats	1631	6700	.243
Wordsworth	1524	6700	.227
Total	6661	26800	.249

B. Odd Frames (1, 3, 5, 7, 9)

Rossetti	1306	6750	.193
Swinburne	1242	6750	.184
Wordsworth	1191	6750	.176
Keats	1096	6750	.162
Total	4835	27000	.179

C. Frames 1-10

Swinburne	3097	13450	.230
Rossetti	2955	13450	.220
Keats	2727	13450	.203
Wordsworth	2715	13450	.202
Total	11494	53800	.214

D. Difference between Even and Odd Frames

Swinburne	.093
Keats	.081
Rossetti	.053
Wordsworth	.052
Total	.070

Keats, Rossetti and Swinburne, as well as the difference between even- and odd-frame equivalence for each poet. Swinburne is distinguished from the other English poets by a very high coefficient of equivalence in even frames, and Keats is distinguished by a low coefficient in odd frames. The *total* equivalence in Frames 1-10 clearly distinguishes the Victorian poets Swinburne and Rossetti from the Romantic poets Wordsworth and Keats. The contrast between even and odd frames is greatest in Swinburne and Keats and lowest in Rossetti and Wordsworth. A poem which shows an extremely high degree of contrast between even and odd-frame equivalence is Keats' sonnet "Great spirits now on earth ...". Figure 3 shows the distribution of equivalence in the first ten frames of this poem compared with that in all forty English sonnets. On the other end of the scale are those poems in which the contrast between even and odd frame equivalence is so *low* that the ponology threatens to drastically disrupt the abstract metrical framework. Two examles of such patterning are Wordsworth's "To Sleep (II)" and Rossetti's "Body's Beauty". The distribution of equivalence in these poems is compared with the forty English sonnets in Figures 4 and 5. Frames 1-3 in the Wordsworth sonnet, Frames 2-5 in the Rossetti sonnet and Frames 7-9 in *both* sonnets show exactly the *opposite* of the expected patterning. In the Rossetti poem in particular, the distribution of even and odd frame equivalence approximates to some extent the type of distribution we might expect to find in English *prose*. Particularly striking is the unusually high degree of equivalence in Frames 5, 7 and 9 of this poem. Frame 7 even exceeds the amount of equivalence in Frame 10, despite the fact that four of the poem's rhyme-pairs occur in the latter. One of the factors contributing to the obscuration of the expected contrast between even and odd frame equivalence in this poem is its monosyllabicity. Over 91 % of the words are monosyllables, while only 69.1 % of those in the sonnet by Keats are monosyllabic. In German, this degree of obscuration of contrast, with its attendant disruption of the meter, virtually never occurs in the nineteenth century, and even in the twentieth century it ap-

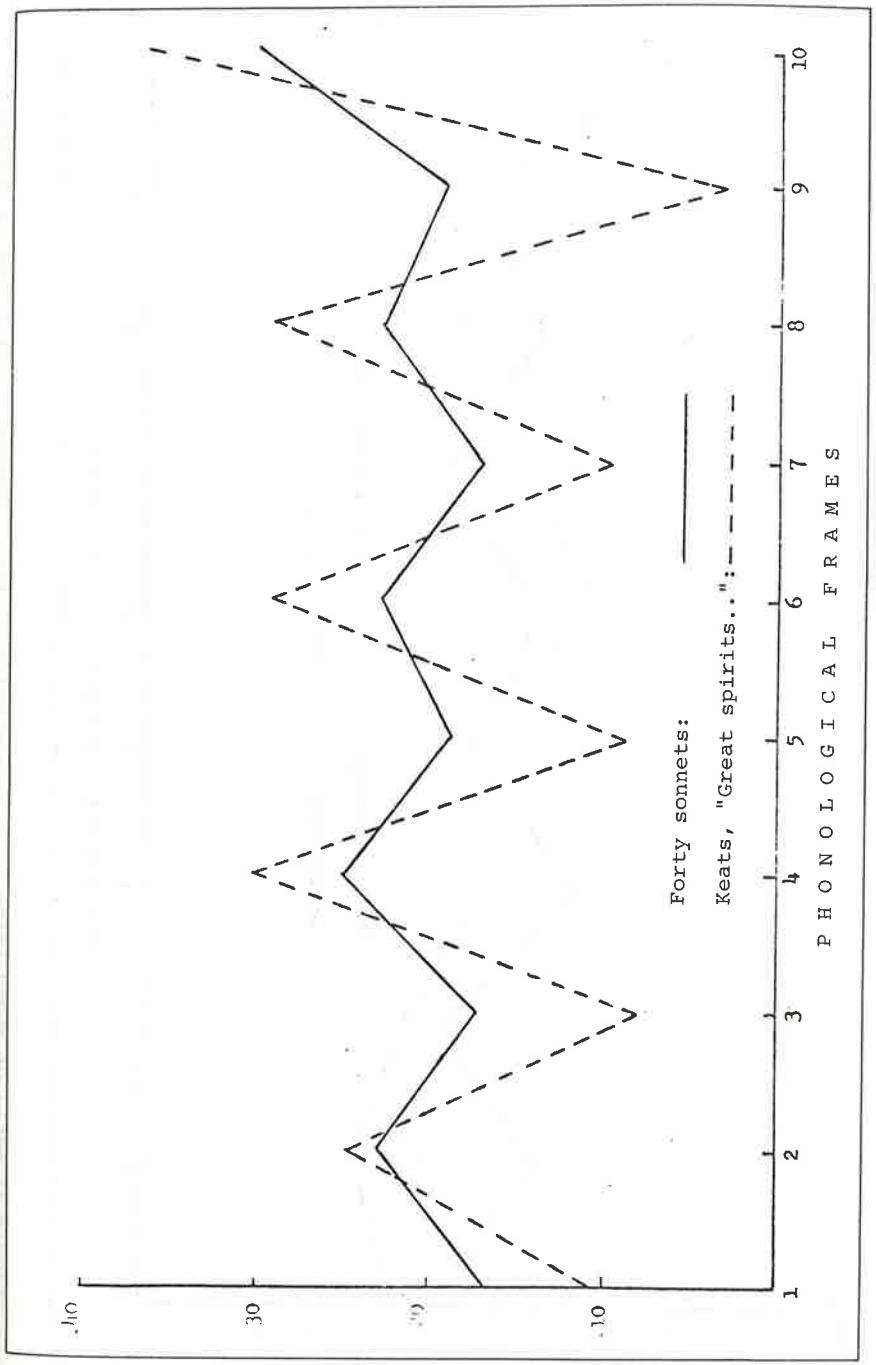


Figure 3: Ratio of Sound Pairs to Syllable Pairs in Forty English Sonnets and in Kents' "Great spirits now on earth ..."

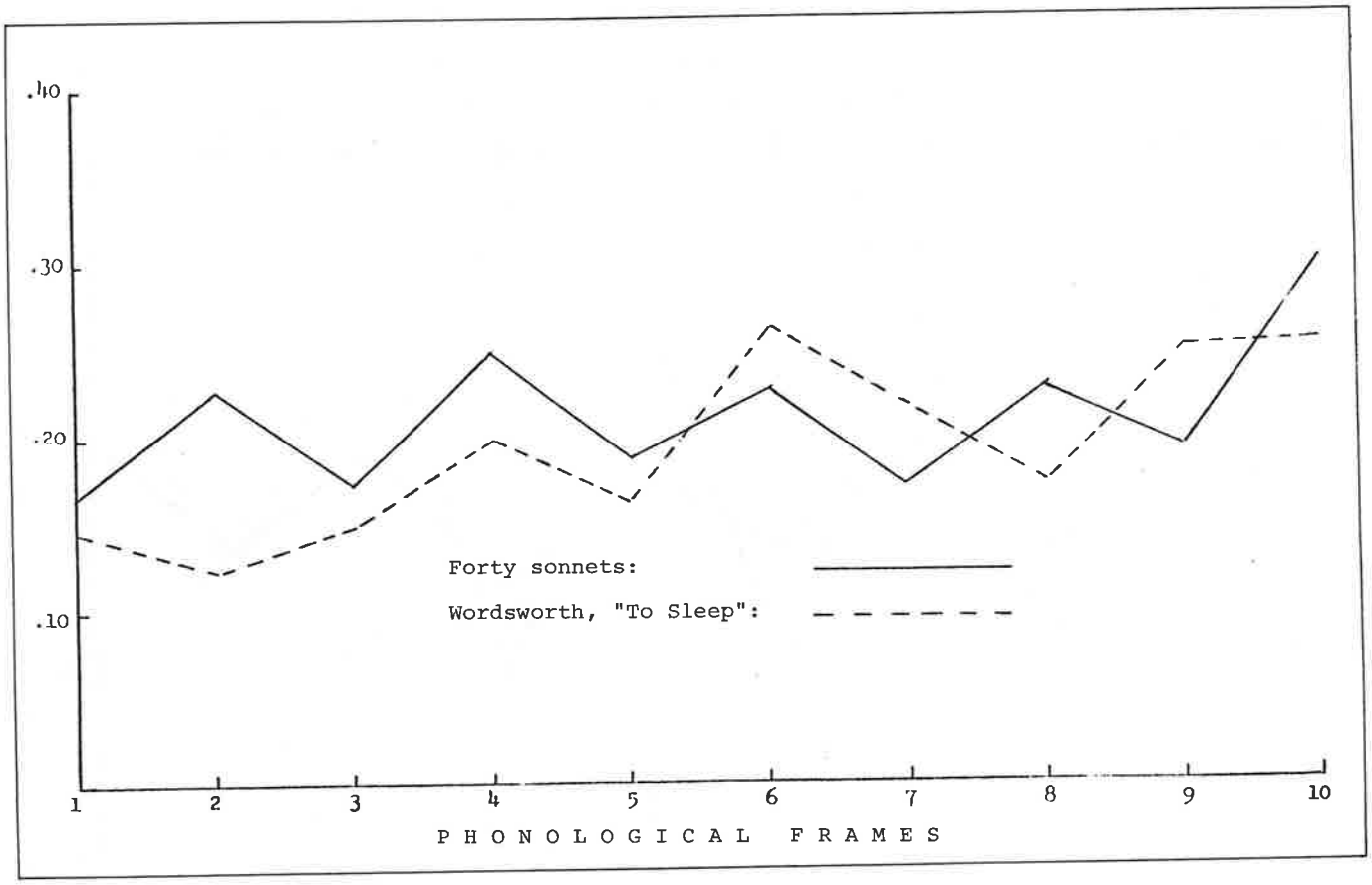


Figure 4: Ratio of Sound Pairs to Syllable Pairs in Forty English Sonnets and in Wordsworth's "To Sleep (II)

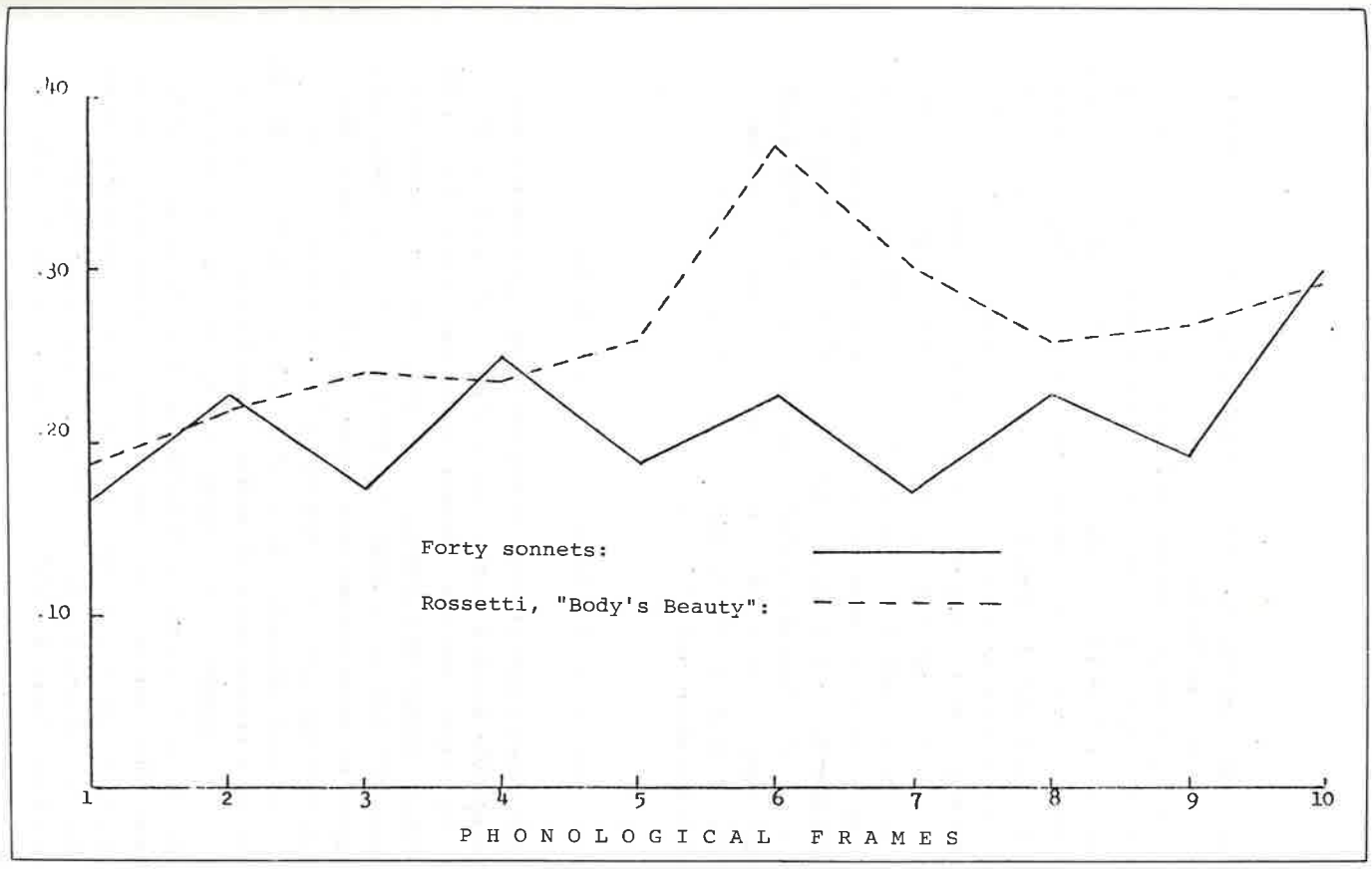


Figure 5: Ratio of Sound Pairs to Syllable Pairs in Forty English Sonnets and in Rossetti's "Body's Beauty"

pears to be limited to a very few poets such as Rainer Maria Rilke and Bertold Brecht.

Part C of Table 4 showed that the total equivalence in Frames 1-10 in the sonnets by the late nineteenth century poets Swinburne and Rossetti significantly exceeds that of the early nineteenth century poets Wordsworth and Keats. Table 5 provides a closer look at each poet by giving, in rank order, the coefficient of equivalence in Frames 1-10 for *each* of the forty sonnets. The ten segments from Swinburne's tragedy *Atalanta in Calydon* are included for comparison, and the poets are arranged chronologically from left to right according to the approximate dates of composition of the poems. While the sonnets of Wordsworth show a fairly wide range of values, those by Keats are all clustered together at the low end of the scale with a range of only .042. The sonnets of Rossetti and Swinburne, as well as the passage from *Atalanta in Calydon*, tend toward the upper end of the scale.

9. *Vertical Patterning in German and English Verse*

I have shown elsewhere that nineteenth century German verse consistently shows a higher equivalence coefficient in the primary frames (vertical patterning between lines) than in the secondary frames (horizontal patterning) even when end-rhyme is excluded from consideration (CHISHOLM 1976: 13). Subsequent analysis of other German poets shows that this tendency prevails into the twentieth century as well although the diffefence becomes markedly less in the sonnets of Rilke and Brecht. Since this predominence of vertical patterning *throughout the verse line* appears to be a general characteristic of German iambic verse since the eighteenth century, the question arises whether this feature might not be common to the iambic meters of other Germanic languages, such as English. To test this hypothesis, the primary and secondary frames were compared in each of the forty English sonnets. Table 6 shows the coefficient of equivalence in primary and secondary frames both with and without end-rhyme in both the German and English sonnets.

Table 5: Coefficients of Equivalence in Frames 1-10 for Four English Poets

	Wordsworth	Keats	Rossetti	Swinburne, Sonnets	Swinburne, Atalanta		Wordsworth	Keats	Rossetti	Swinburne, Sonnets	Swinburne, Atalanta
.28					.282 (IV	.22	.217 (VIII) .216 (IV) .214 (V)	.219 (III) .217 (VIII)			.215 (VII) .213 (VI)
.27				.268 (III)		.21		.214 (IV) .213 (VII) .210 (VI) .207 (X)	.213 (III)	.213 (IX) .209 (V)	
.26			.261 (IX) .257 (X)		.261 (I)	.20	.200 (VI)		.209 (II) .208 (VII) .200 (I) .200 (VIII)	.207 (VI) .200 (VII)	.207 (X) .205 (III) .202 (V)
.25				.252 (IV) .247 (X)		.19	.192 (III)	.194 (I) .189 (IX) .187 (V)	.193 (IV)		.190 (IX)
.24	.239 (IX) .236 (I)		.235 (VI)	.240 (II) .234 (VII) .232 (I)	.239 (VIII)	.18	.178 (II) .177 (X)	.177 (II)			
.23						.17					
.22			.222 (V)			.16	.158 (VII)				
						.15					

Table 6: Phonological Equivalence in Primary (with and without End-rhyme) and Secondary Frames in English and in German Sonnets

	(Primary Frames 1-4)		(Secondary Frames 2,4,6,8)
I. English	Slots 1-10(E)	Slots 1-8(-E)	Slots 1-10
Wordsworth	.262	.224	.215
Keats	.305	.248	.231
Rossetti	.292	.245	.236
Swinburne, *Sonnets*	.320	.265	.263
(Swinburne, *Atalanta*	.249(-E)		.261)
II. German			
Goethe, *Sonnets*	.37	.34	.30
Goethe, *Iphigenie*	.38(-E)		.33
Schlegel	.42	.37	.32
Brentano	.42	.37	.31
Heine	.44	.44	.35
Mörike	.42	.36	.30
Meyer	.44	.41	.37
Keller	.46	.38	.36
Heym	.48	.44	.37
Rilke	.38	.36	.32
Brecht	.38	.33	.31

For each poet the number of syllable-pairs available for comparison in the primary frames (1-4) was 4600 for Slots 1-10 and 3680 for Slots 1-8; in the first four secondary even frames (2, 3, 6, 8) 5400 syllables were available for comparison. The table shows clearly that vertical patterning, other than that attributable to endrhyme, is *not* a dominant feature in the sonnets of Wordsworth, Keats, Rossetti and Swinburne. Whereas in German the differences between the coefficients of equivalence for primary frames in Slots 1-8 and the secondary frames are highly significant, the slight differences in the English sonnets are not statistically significant at the 5 % level of probability. Only when end-rhyme is included does English verse show significant increase in the primary frames.

10. Phonological Development in English and German Sonnets

A poem is perceived by the listener or reader not all in one instant, but rather as a progression of shifting relationships from beginning to end. This development can be analyzed on three levels: the verse line, the stanza and the whole poem. The phonology participates in this process, and for certain types of poetry the amount of phonological equivalence in a given frame or group of frames may differ significantly from one stage of the poem's development to another. Analysis of nineteenth century German sonnets, for example, showed that metrically disruptive odd frame equivalence decreases in the second half of sonnets by Brentano and Mörike, but *increase* in sonnets by Heine and C.F. Meyer. It also showed a considerable decrease in Frame 1 equivalence in the second half of the *line* in the sonnets by Goethe, and to a lesser extent Brentano. Table 7 gives the total coefficient of equivalence in both odd and even frames in the first and second half (lines 1-7 and 8-14 respectively) of the English sonnets by Wordsworth, Keats, Rossetti and Swinburne.

Whereas the Romantic poets Wordsworth and Keats both show a decrease in both odd and even frame equivalence in the se-

Table 7: Coefficient of Equivalence in Odd and Even Frames in the First and Second Half of Sonnets by Four English Poets

	First Half (lines 1-7)	Second Half (lines 8-14)	% Increase
1. Wordsworth			
Odd Frames :	605	536	- 11.40 %
Even Frames :	822	660	- 19.71 %
Total (Frames 1-10) :	1427	1196	- 16.19 %
2. Keats			
Odd Frames :	561	485	- 13.55 %
Even Frames :	791	762	- 3.67 %
Total (Frames 1-10) :	1352	1247	- 7.77 %
3. Rossetti			
Odd Frames :	585	673	+ 15.04 %
Even Frames :	758	808	+ 6.60 %
Total (Frames 1-10) :	1343	1481	+ 10.28 %
4. Swinburne			
Odd Frames :	656	780	+ 18.90 %
Even Frames :	847	907	+ 7.08 %
Total (Frames 1-10) :	1503	1687	+ 12.24 %

cond half of the sonnet, the Victorian poets Rossetti and Swinburne show a sizable *increase* in the second half, most of which is attributable to metrically disruptive odd frame equivalence. Since the "line" functions as an important rhythmic entity in verse, it is of interest to determine whether certain types of equivalence occur more or less frequently toward the end than at the beginning of the line. It was noted earlier that both poetry and prose tend to avoid Frame 1 equivalence, so that this frame usually shows the least equivalence of all frames in a given text. Table 8 gives the distribution of syllables containing Frame 1 equivalence from the beginning to the end of the line in the forty English sonnets under consideration. The sonnets of Wordsworth show a considerable *decrease* in the second half of the line, while those of Swinburne and Rossetti show an increase. Because of the small number of equivalences for each poet, only the results for Wordsworth are statistically significant; further work will be necessary to determine whether Rossetti and Swinburne consistently increase metrically disruptive Frame 1 equivalences in the second half of the line.

The following generalizations about English and German verse are supported by the results of this study:

(1) Whereas vertical sound patterning predominates *throughout the verse line* in German iambic verse, the nineteenth century English sonnets show no statistically significant increase in vertical patterning other than that attributable to end-rhyme.

(2) English verse, in part because of its greater monosyllabicity, shows substantially less contrast between even and odd frame equivalence than does German verse.

(3) There is a marked tendency - due in part to stylistic factors - to avoid equivalence between contiguous syllables in both prose and poetry in English and German. Excessive Frame 1 equivalence would weaken an underlying alternation pattern which exists as a *tendency* in prose and as a conscious metrical device in verse.

Table 8: Distribution of Syllables Continuing Frame 1 Equivalence in the First and Second Half of the Verse Line

	Metrical Slots												
	First Half of Line						Second Half of Line						
	1	2	3	4	5	Total	6	7	8	9	10	Total	% Increase
Wordsworth	42	41	45	39	36	203	38	32	26	31	36	163	- 19.70 %
Keats	30	28	35	44	38	175	37	39	33	32	33	174	- 0.57 %
Swinburne	54	47	39	35	39	214	42	39	46	48	55	230	+ 7.48 %
Rossetti	40	42	32	32	38	184	33	31	43	50	44	201	+ 9.24 %

(4) The English Victorian poets Swinburne and Rossetti are distinguished from the Romantic poets Wordsworth and Keats by (a) a significantly greater coefficient of equivalence in Frames 1-10 and (b) an *increase* in even and especially odd frame equivalence in the second half of the sonnet.

(5) The results for both languages offer strong counterevidence to B.F. SKINNER's theory of "formal perseveration" in verse.

The usefulness of this computer-assisted approach to iambic verse in two languages suggests that its extension to other verse forms and other languages will not only reveal stylistic distinctions between poets and periods, but could also lead to the discovery of universal prosodic patterns which inhere in the poetic phonology of diverse languages.

Notes

1 An alternative solution would be to attach the linking-r to the beginning of the following word, so that the data would show only prevocalic /r/.

2 The significance of the difference can be checked by means of the chisquare test. For English, the difference between blank verse and the sonnets is highly significant at the 5 % level of probability ($\chi^2 = 12.40$), while in German the difference is not quite significant ($\chi^2 = 2.86$).

3 As early as the seventeenth century Martin OPITZ proscribed sequences such as "Die, die dir diese Dinge sagen." See M. OPITZ (1624).

References

Chisholm, David
1976 Phonological Patterning in German Verse. Computers and the Humanities 10, 5-20.

Gimson, A.G.
1962 An Introduction to the Pronunciation of English, London, Arnold.

Jones, Daniel
1956 An Outline of English Phonetics, 8th Edition, Cambridge, Heffer.

Opitz, Martin
1624 Buch von der deutschen Poeterey, Breslau, David Müller. (Reprinted in: Neudrucke deutscher Literaturwerke, Neue Folge 8, Tübingen, Niemeyer, 1963.)

Skinner, B.F.
1941 A Quantitative Estimate of Certain Types of Sound-Patterning in Poetry. American Journal of Psychology 54, 64-79.

TOWARDS A LAW OF WORD REPETITIONS IN TEXT-BLOCKS

G. Altmann, V. Burdinski, Bochum

Abstract

The distribution of text-blocks containing x words (or other text units) is derived in the form of the negative hypergeometric (Beta-binomial) distribution. Tests are performed on English, Russian, Hungarian, Bulgarian and Indonesian data. Factors that are probably in relation to the parameters and the possible uses of this law are considered.

1. It is impossible to construct theories without laws, since according to the conception of all empirical sciences theories are systems of laws. Laws are hypothetical statement of general validity in a particular domain of reality, containing no observational concepts, derived from plausible assumptions which can be regarded as axioms, or which can be derived from a theory (even an embryonic one), i.e. can be derived from other laws, and can be sufficiently corroborated empirically (cp. Bunge 1961; 1967; 305-279). This rather casually adopted definition causes immediate rejection of empirical generalizations, systems of concepts and ad hoc hypotheses (e.g. rules) as potential laws, and this is probably the reason why in (text)linguistics (erronesouly referred to as text theory) laws are at least for the time being still so rarely met with. Text is a complex entity in which extremely varied factors play a rôle (language, style, genre, dialect, age, status, etc.) so that it seems possible to set up only empirical generalizations. But if one considers that texts, independent of the language and personality of the author - represent a class of objects having common properties which come into existence by

virtue of the existence of such laws, one is encouraged to derive hypotheses about text construction from assumptions. At the moment it is only possible to set up and test "simple" hypotheses. In this article we shall attempt - without taking any factors into account at all - to derive a theoretical frequency distribution holding to all word distributions in text blocks, and to test it against data. We shall restrict ourselves to the distribution of a single word in a text block.

2. The first investigations along these lines were perfomed, as far as we have been able to discover, by Frumkina (1962). She examined the distribution of different words ("bez" without, "den'" day, "drugoj" another, "novyj" new, "očen'" very, "zdes'" here, "do" until, "meždu" between, "pri" at, "nad" above, "ili" or, "znat'" to know) in text-blocks of 1000 words in the work of Pushkin. Since the data are not easily accessible we present them in Table 1. Frumkina probably proceeded from the assumtion that for entities with very low probability of occurrence "the law of small numbers" held and tried to fit the Poisson distributions to these data. The χ^2 test for goodness of fit and the respective probabilities are presented in Table 2. In three cases (meždu, nad, ili) small frequencies below 1.00 have been considered because of the small number of degrees of freedom. The results show that in 5 cases out of 12 the hypothesis of the Poisson distribution can be rejected. The fact that the mean and the variance of the Poisson distributions are equal should have been sufficient for the rejection of the Poisson distribution as inadequate (as a general model).

Brainerd (1972) was more successful in applying the Poisson distribution to the frequencies of the English article in 39 different samples of text-blocks consisting of 50 words respectively. He demonstrated that with one exception (a sample from J. Cheevers Wapshot Chronicle) all data fit the Poisson distribution (this holds for samples not tested by Brainerd as well). He decided to use the Poisson distribution because in all samples the mean and the variance yielded approximately the same value. The only exception was explained by Brainerd with

Table 1. The distribution of Russian words in passages of 1000 words according to Frumkina (1962)

Number of words	Number of passages												
	bez without	den' day	drugoj another	novyj new	očen' very	zdes' here	do until	meždu between	pri at	nad above	ili or	znat' to know
0	36	36	28	88	59	70	44	104	76	97	101	33
1	42	26	34	28	26	25	35	5	24	11	8	27
2	23	16	28	11	13	11	17	1	9	2	1	14
3	6	17	15		10	2	6		1			12
4	3	11	4	7		0	6					7
5					4	2	1					4
6		4										
7												
8												3
9												
Σ	110	110	109	134	112	110	109	110	110	110	110	100
m_1'	1,0727	1,6091	1,3853	0,5821	0,9286	0,5727	1,0642	0,0636	0,4091	0,1364	0,0909	1,6150
m_2	0,9947	2,5590	1,2460	1,0343	1,7181	0,9174	1,4179	0,0778	0,4599	0,1541	0,1008	3,1093

Table 2. Fitting the Poisson distribution to the data of Frumkina

Wort	X^2	df	P≈
bez	0,40	2	0,82
den'	21,41	3	<0,0005
drugoj	1,72	3	0,63
novyj	8,26	2	0,018
očen'	17,60	2	<0,0005
zdes'	4,67	1	0,03
do	3,43	2	0,18
meždu	1,95	1	0,17
pri	2,81	1	0,096
nad	1,88	1	0,18
ili	0,90	1	0,35
znat'	19,55	3	<0,0005

the assumption that the passages sampled were heterogeneous and represented perhaps a superposition of several Poisson distributions. By compounding three Poisson distributions he arrived at a good fit. It would probably be possible to achieve better fits to Frumkina's data using the compound Poisson distribution if one assumes for example that a particular word has several meanings which must be examined separately. This would, however, make semantic complications unavoidable.

Piotrowski (1975: 111-119) has shown that several Soviet authors tried to fit a posteriori the Poisson, the normal and the lognormal distribution and discussed the results. The data were not accessible to us. In this contribution we shall try to pursue another avenue of approach.

3. Let us consider any linguistic entity A (affix, word, phrase etc.) occurring in a language generally with probability p and let us denote with n the number of positions in a passage in which A could occur. With context-free occurrence of A the p is constant, with context-dependent occurrence of A we have in most places of the passage p = 0, in other positions, depending on context, p > 0 but not necessarily constant. If p were constant then the probability to find A exactly x times in these n positions would obviously be

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,\ldots,n \qquad (1)$$

representing the binomial probability distribution. Since p is different in each of the potential n positions the resulting probability to find exactly x As depends on p, i.e. it is a conditional probability. The parameter p can be considered as a continuous random variable whose domain lies in the interval <0,1>. In order to find the common distribution of x and p we take

$$f(x,p) = f(x|p)f(p)$$

from which the marginal distribution of x can be computed as

$$f(x) = \int_0^1 f(x|p)f(p)dp. \qquad (2)$$

The distribution of p is not known. It cannot be derived theoretically because there is no such thing as the population of all texts of a language: there are merely sublanguages (a term introduced by Soviet authors) having their own (perhaps very different) distributions of p. Therefore in order to keep the theoretical background flexible we choose the Beta distribution as being general enough and allowing different text factors to be brought into relation to its parameters. We write the Beta distribution in the form

$$f(p) = \frac{1}{B(M,K-M)} p^{M-1}(1-p)^{K-M-1}, \qquad 0 < p < 1 \qquad (3)$$

where

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a-b)} \quad \text{and} \quad \Gamma(k) = (k-1)!$$

Substituting (1) and (3) into (2) and integrating we obtain

$$P(X=x) = f(x) = \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \frac{1}{B(M,K-M)} p^{M-1} (1-p)^{K-M-1} dp =$$

$$= \binom{n}{x} \frac{1}{B(M,K-M)} \int_0^1 p^{M+x-1} (1-p)^{K-M+n-x-1} dp =$$

$$= \binom{n}{x} \frac{B(M+x,\ K-M+n-x)}{B(M,K-M)} .$$

Writing the Beta expressions as above and arranging them in form of binomial coefficients we obtain

$$P(X=x) = \frac{\binom{M+x-1}{x}\binom{K-M+n-x-1}{n-x}}{\binom{K+n-1}{n}} =$$

$$= \frac{\binom{-M}{x}\binom{-K+M}{n-x}}{\binom{-K}{n}}, \qquad x = 0,1,\ldots,n \qquad (4)$$

$$0 < M < K$$

The probability of finding exactly x times A in a text-block equals (4) and the expected number of text blocks containing exactly x entities A is given as $NP(X=x)$, where N is the number of text blocks sampled. Distribution (4) is called the negative hypergeometric (NH) (or Beta-binomial) distribution and represents a special case of the generalized hypergeometric distribution (Type IIa according to Kemp & Kemp 1956a; cf. also Johnson & Kotz 1969; Ord 1972). The parameters M and K are positive, not necessarily integers, $K > M$.

The NH distribution has three important limiting cases:

(i) If $K \to \infty$, $M \to \infty$ so that $M/K \to p$ then it tends to the binomial distribution as given in (1).

(ii) If $K \to \infty$, $n \to \infty$ so that $K/(K+n) \to p$ then it tends to the negative binomial distribution

$$P(X=x) = \binom{M+x-1}{x} p^M (1-p)^x, \qquad x = 0,1,\ldots$$

(iii) If $M \to \infty$, $K \to \infty$, $n \to \infty$ so that $Mn/K \to \lambda$ than it tends to the Poisson distribution

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0,1,\ldots$$

In this way we obtain the original binomial distribution, the Poisson distribution, which was relatively well corroborated on English articles, and the negative binomial distribution as limiting cases. If we succeed in showing empirically that the NH distribution or its limiting cases agree with the data then we would have a candidate for the law of word distributions in text passages.

The following circumstances are to be considered here:

(a) The Beta distribution is a very general density function, the use of which in this case would be difficult to justify theoretically. The fact that it can be derived as a special case of the Pearson curves or as a function of the normal or the Gamma distributions could be of some help. It may also turn out that one can make do with one or several special cases of the NH distribution.

(b) The parameters of the distribution should also be derived theoretically from other factors of language or text or from the properties of the entitity examined or from other laws. Piotrowski (1975) mentions on the basis of examinations of several Soviet authors (Maškina 1968; Bektaev & Lukjanenkov 1971; Paškovskij & Srebrjanskaja 1971) the following possible factors: (i) the relative frequency of a word in a frequency dictionary; (ii) the length of the passage in which the word distribution is examined; (iii) the word class and the lexical-grammatical character of the word; (iv) the relevance of the word for the theme.

Standard linguistics does not provide any help along these

lines so that for the time being we are restricted to the collecting and testing of data.

If there is a priori no possibility of determining the parameters, one must estimate them from the data. One could take n as the greatest number of occurence of the examined entity in passages of the given length in the given language and thereby spare one degree of freedom with the test of fit. If this is not possible (preliminarily) then we set

$\hat{n}$ + 1 ≥ number of classes of the random variable.

If $f_o \neq 0$ then

$\hat{n} \geq x_{max}$;

if $f_o = 0$, $f_1 \neq 0$ then

$\hat{n} \geq x_{max}-1$ etc.

The other parameters can be estimated by means of the moments

$$\mu_1' = \frac{nM}{K}$$

$$\mu_2 = \frac{nm(k-M)(K+n)}{K^2(K+1)}$$

as

$$\hat{K} = \frac{n(n\bar{x}-\bar{x}^2-s^2)}{n(s^2-\bar{x})+\bar{x}^2}$$

and

$$\hat{M} = \frac{\hat{K}\bar{x}}{n}.$$

The better but much more lengthy maximum-likelihood method can be found in Kemp & Kemp (1956b). A chi-square test for fitting the NH distribution can be performed only if there are at least 5 classes since we must substract 4 degrees of freedom.

The fitting of the NH distribution to the data of Frumkina is presented in Table 3. This table contains the expected frequencies NP(X=x), where $N = \Sigma f_i$ = sample size and P(X=x) was computed from (4) by means of the recursion formula

Table 3. The NH distribution and the test of fit for Frumkina's data

Number of words in passage	Expected number of passages NP(X=x)												
	bez	den'	drugoj	novyj	očen'	zdes'	do	mežďu	pri	nad	ili	znat'
0	36,21	36,09	27,50	90,09	60,81	72,11	46,19	104	75,55	97	101	35,85
1	41,59	25,64	35,17	23,45	23,47	21,74	30,25	5	25,36	11	8	22,55
2	22,58	18,82	27,47	11,23	13,09	9,72	18,05	1	7,65	2	1	15,50
3	7,58	13,39	14,54	5,71	7,63	4,34	9,45		1,45			10,64
4	1,73	8,87	4,32	2,64	4,25	1,67	3,99					7,06
5	0,28	5,11		0,88	2,06	0,43	1,06					4,40
6	0,03	2,09			0,70							
n	9	6	4	5	6	5	5	2	3	2	2	8
K	150,62	3.2621	6,9788	2,9565	3,2050	3,9438	4,7752	2,8131	5,6295	3,6948	5,1766	3,9568
M	17,95	0,8748	2,4170	0,3442	0,4960	0,4517	1,0163	0,0895	0,7677	0,2519	0,2353	0,7988
X^2	0,05	1,47	0,10		2,35	1,98	2,92					1,66
df	1	2	1	*	1	1	1	*	*	*	*	2
P≈	0,83	0,48	0,76		0,13	0,17	0,09					0,44

$$P(X=0) = \frac{(K-M)(K-M+1)\ldots(K-M+n-1)}{K(K+1)\ldots(K+n-1)}$$

$$P(X=x+1) = \frac{(M+x)(n-x)}{(x+1)(K-M+n-1-x)} \cdot P(X=x)$$

With bez the highest class is to be understood as NP(X≥6) = 0.03, with znat' as NP(X≥5) =4.40. The estimated parameters, the computed X^2 values, the degrees of freedom (df) and the corresponding probabilities (P) are in the rows under the table. In samples marked with* (meždu, pri, nad, ili) the X^2 test could not be performed because of small df but it is evident that the fit is good. Several fits could be further improved if a greater n were chosen or if other estimations were performed. It was, however, not our aim to find the best fit but to test the adequacy of the model.

Since Brainerds data with one exception follow the Poisson distribution it is sufficient to examine this exception. The observed and the computed frequencies are given in Table 4. As may be seen, the fitting of the NH distribution is good even in this case, although the empirical distribution is somewhat "pathological".

5. For the sake of the further testing of the model data from 3 other languages have been collected. Our hypothesis can be considered as corroborated by any empirical case if the NH distribution or some of its limiting distributions can be adequately fitted. Since the fitting of the limiting distributions is easier, one can use them if there are favourable circumstances. When choosing among them one can use the following criteria:

if $\bar{x} > s^2$ binomial

$\bar{x} = s^2$ Poisson

$\bar{x} < s^2$ negative binomial

where $\bar{x} = m_1' = \frac{1}{N}\sum_i x_i f_i$ and $s^2 = m_2 = \frac{1}{N}\sum_i (x_i-\bar{x})^2 f_i$ (we used here

Table 4. The distribution of the article in a sample of Cheevers Wapshot Chronicle (WC 1) according Brainerd (1972)

Number of article in passage	Number of passages		
	observed	Poisson	NH
0	8	1,54	7,62
1	8	6,14	9,89
2	12	12,23	10,71
3	11	16,26	10,72
4	14	16,22	10,20
5	4	12,93	9,29
6	8	8,60	8,07
7	8	4,90	6,62
8	6	2,44	5,01
9	2	1,08	3,30
10	2	0,66	1,58
		λ=3,988	n=10 K=3,6401 M=1,4517
	X^2	18,30	5,61
	df	6	6
	P≈	0,0056	0,47

the biased estimation of σ^2). The estimation of the parameters can be performed e.g. as follows:

Binomial: $\hat{n} = x_{max}$

$$\hat{p} = 1 - \frac{\hat{n}}{\sqrt{f_o/N}}$$

where f_0/N is the relative frequency of the zero class.

Poisson: $\hat{\lambda} = \bar{x}$

Negative binomial: (1) $\hat{p} = \frac{\bar{x}}{s^2}\,2;\ \hat{k} = \frac{\bar{x}^2}{s^2-\bar{x}}$

or iteratively from (2) $\frac{\bar{x}}{-\log \frac{f_0}{N}} = \frac{\hat{p}}{\log(1+\hat{p})}$; $k = \frac{\bar{x}}{\hat{p}}$

where $\hat{p} = \frac{1}{1+\hat{p}}$.

Let us now turn to the use of the limiting distributions in a concrete case. In Table 5 the empirical distribution of the German article das is presented in 200 passages from "Deutschstunde" by S. Lenz (column 1). In columns 2 to 6 n will be successively increased and it can be seen that the fit improves. Since $\bar{x} < s^2$, we assume that with increasing n and K the negative binomial will yield the best fit. The estimation of p according to (1) gives $\hat{p} = 0.5663$ and by means of (2) $\hat{p} = 0.6254$. Computing the criteria $K/(K+n)$ for columns 2 to 6 one sees that it slowly tends to $\hat{p}$:

Column	2	3	4	5	6
K/(K+n)	0.4275	0.4449	0.4969	0.5364	0.5498

In column 7 the negative binomial distribution is presented. As can be seen, this fit is the best one.

For the German samples in Tables 6 and 7 we used only the negative binomial and in one case, namely with "bis" ("until" with temproray meaning) the binomial because of $\bar{x} > s^2$. The use of the χ^2-test was not possible in each case but the agreement is even optically very good.

The theoretical frequency of the highest class x_{max} of the negative binomial distribution was computed as $\sum_{x=x_{max}}^{\infty} NP_x$.

Table 5. The distribution of the article "das" in the nominative in passages from Lenz

Number of "das" in the passage	observed number of passages	NH						Negative binomial
0	95	102,64	102,04	100,45	99,28	99,13		95,00
1	58	47,49	48,38	50,72	52,28	52,64		56,44
2	28	25,23	25,34	25,60	25,83	25,78		27,34
3	11	13,32	13,11	12,60	12,33	12,20		12,24
4	3	6,66	6,46	5,98	5,72	5,64		5,26
5	2	3,04	2,94	2,72	2,59	2,57		2,20
6	1	1,19	1,20	1,18	1,14	1,15		0,90
7	1	0,37	0,41	0,48	0,49	0,51		0,37
≥ 8	1	0,07	0,12	0,27	0,34	0,38		0,25
$\hat{n}$		8	9	15	30	60		$\hat{p}$ = 0,6254
$\hat{K}$		5,9740	7,2121	14,8159	34,7090	73,2888		$\hat{k}$ = 1,5862
$\hat{M}$		0,7094	0,7613	0,9383	1,0991	1,1604		
χ^2		5,64	4,89	3,30	2,47	2,30		0,29
df		2	2	2	2	2		2
P ≈		0,06	0,09	0,19	0,29	0,32		0,87

N = 200; $m_1' = 0{,}9500$; $m_2 = 1{,}6775$.

Table 6. Distribution of "das" in passages from Lenz

number in the passage	article in the accusative	NB	demonstrative pronoun in the nominative	NB	demonstrative pronoun in the accusative	NB	relative pronoun in the nominative	NB	relative pronoun in the accusative	NB
0	46	46,00	85	86,67	122	122,40	150	149,88	170	170,01
1	57	52,32	66	61,50	51	49,08	42	42,32	27	26,43
2	31	40,62	28	30,64	15	18,24	7	6,86	2	3,18
3	31	26,64	11	13,10	9	6,60	1	0,94	1	0,38
4	16	15,87	8	5,14	2	2,36				
5	7	8,88	0	1,91	0	0,84				
6	7	4,76	2	1,04	1	0,48				
7	4	2,46								
≥ 8	1	2,45								
Σ	200	200	200	200	200	200	200	200	200	200
m'_1	1,9450		0,9950		0,6100		0,2950		0,1700	
m_2	3,2320		1,3950		0,9279		0,3080		0,1911	
$\hat{p}$		0,5848		0,7133		0,6574		0,9578		0,9146
$\hat{k}$		2,7394		2,4751		1,1705		6,6923		1,8201
χ^2		4,87		1,38		0,94		*		*
df		5		2		1				
P ≈		0,43		0,50		0,34				

Table 7. Distributions of "in" and "bis" in passages from Lenz

number in the passage	Number of passages							
	"in" + dative	NB	"in" + accusative	NB	"bis" spatial	NB	"bis" temporal	bin.
0	10	11,28	52	52,00	160	160	154	154
1	34	30,40	61	62,37	30	32	42	43
2	42	42,82	46	43,98	10	8	4	3
3	37	41,96	25	23,77				
4	38	32,10	8	10,89				
5	17	20,44	4	4,45				
6	12	11,25	4	2,54				
7	6	5,52						
8	2	2,44						
9	2	1,79						
	200	200	200	200	200	200	200	200
m'_1	3,0700		1,5200		0,2500		0,2500	
m_2	3,4951		1,8996		0,2875		0,2275	
$\hat{p}$		0,8784		0,7891		0,8000		0,1225
$\hat{k}$		22,1710		5,6865		1,0000		$\hat{n}$ = 2
χ^2		2,94		1,10				
df		6		3		*		*
P ≈		0,81		0,78				

Table 8. The distribution of Bulgarian words in text-blocks

number in the pass.	tova (demonstrative pronoun)				koeto (relative pronoun)				-to (definite article)			
	AM	NH	St	NH	AM	NH	St	NH	AM	NH	St	NH
0	43	44,97	79	79,24	58	58,99	84	84,12	1	2,37	9	9,65
1	26	24,88	17	16,17	24	21,90	14	13,76	8	5,31	16	16,46
2	19	14,25	3	3,95	11	11,33	2	2,12	7	8,00	18	18,86
3	5	7,97	1	0,64	4	5,65			8	10,06	20	17,70
4	3	4,25			3	2,12			9	11,29	17	14,43
5	1	2,13							13	11,69	11	10,39
6	2	0,97							17	11,32	4	6,59
7	1	0,58							13	10,32	1	3,62
8									5	8,88	2	1,64
9									8	7,19	1	0,56
10									3	5,44	1	0,11
11									3	3,79		
12									2	2,38		
13									1	1,28		
14									0	0,54		
15									2	0,14		
Σ	100	100	100	100	100	100	100	100	100	100	100	100
m'_1	1,15		0,26		0,70		0,18		5,28		3,01	
m'_2	2,09		0,31		1,03		0,19		9,77		4,11	
$\hat{n}$		10		3		4		2		15		10
$\hat{K}$		7,5426		5,5478		2,8287		5,2519		7,0323		8,4396
$\hat{M}$		0,8674		0,4808		0,4950		0,4727		2,7285		2,5403
X^2		2,94				1,07				7,81		2,05
df		1		*		1		*		7		4
P≈		0,09				0,30				0,35		0,73

St - Stanev, Antichrist; AM - Andrejčin, Mirčev, Slavističen Sbornik

Table 8 (cont.). The distribution of Bulgarian words in text-blocks

number in the pass.	do "up to"				v "in"			
	AM	NH	St	NH	AM	NH	St	NH
0	77	77,12	81	81,00				
1	14	13,76	17	17,00			2	2,72
2	9	9,12	2	2,00	2	0,90	9	6,51
3					0	2,58	14	9,92
4					5	4,73	11	12,15
5					6	6,99	18	12,98
6					9	9,06	15	12,56
7					13	10,65	9	11,23
8					14	11,59	10	9,39
9					14	11,76	3	7,39
10					10	11,18	4	5,48
11					7	9,90	4	3,83
12					7	8,11	0	2,53
13					4	6,01	1	3,31
14					4	3,88		
15					4	2,01		
15					1	0,66		
Σ	100	100	100	100	100	100	100	100
m'_1	0,32		0,21		8,82		5,28	
m'_2	0,40		0,2059		9,21		9,19	
$\hat{n}$		2		2		14		20
$\hat{K}$		1,0488		2,4708		6,9600		12,9209
$\hat{M}$		0,1678		0,9944		3,3905		3,4111
X^2						4,48		10,29
df		*		*		7		6
P ≈						0,72		0,11

Table 9. The distribution of Indonesian words in text-blocks

number in the pass.	dan "and"				dengan "with"			
	H	NH	A	NH	H	NH	A	NH
0	2	2,11	3	1,27	8	10,09	13	9,18
1	3	4,98	2	4,06	20	18,42	13	22,13
2	5	7,62	7	7,72	24	21,21	35	30,16
3	13	8,87	9	11,22	17	19,12	31	29,56
4	16	11,40	15	13,66	16	14,36	23	22,42
5	16	12,10	13	14,52	7	9,13	12	13,32
6	9	11,97	21	13,76	5	4,85	6	6,03
7	6	11,09	12	11,74	2	2,07	1	1,89
8	10	9,60	9	9,02	0	0,64	1	0,32
9	8	7,70	4	6,20	1	0,11		
10	3	5,61	1	3,76				
11	6	3,58	1	1,95				
12	2	1,85	1	0,82				
13	1	0,61	1	0,26				
14			1	0,04				
Σ	100	100	100	100	100	100	100	135
m'_1	5,77		5,46		2,71		2,81	
m'_2	8,56		6,69		3,25		2,70	
$\hat{n}$		13		14		9		8
$\hat{K}$		6,1964		11,8886		10,1739		13,5504
$\hat{M}$		2,7503		4,6365		3,0635		4,7596
X^2		11,34		5,90		1,87		6,36
df		7		6		3		3
$P\approx$		0,13		0,44		0,60		0,096

H - Hartwardojo, Orang Buangan: A - Adiwinarto, Bintangpun tiada

In Table 8 some distributions from Bulgarian are peresented. The differences between prose (E. Stanev, Antichrist, Sofia 1973) and scientific literature (L. Andrejčin & Mirčev (Eds.), Slavističen sbornik, Sofia 1973) can be observed in the differences of the parameters but the model itself holds true.

In Table 9 the Indonesian word "dan" (and) and "dengan" (with) in two novels (B. Adiwinarto, Bintangpun tiada, Bandung 1971 and H.S. Hartowardojo, Orang Buangan, Djakarta 1971) are examined.

6. The circumstances that word distributions in text-blocks are governed by a stochastic law shows that in spite of (stilistic and other) differences between text and in spite of rules which are responsible for the wellformedness of sentences, the construction of texts is governed by laws. In order to elaborate these laws the following procedures must be performed:

(a) In absence of a theory models for individual text phenomena must be derived from assumptions. At the beginning one will be forced to compensate for the lack of rigorous derivations at some stages by approximations or trial-and-error assumptions. In the course of the development they should be, of course, replaced with plausible, interpretable steps.

(b) The estimations of the parameters extracted at the beginning from the data must be later on replaced by derived quantities which are related to other properties of language or text.

(c) The lawlike statements should be interrelated, i.e. one must construct a system of text laws that can be later on derived from an axiomatic background. We find ourselves here on the threshhold of a "physics of text", which is developing into a very extensive discipline (cf. e.g. Orlov & Boroda & Nadarejšvili 1980).

A satisfactorily elaborated theory of word repetitions in text-blocks could be of help for the following procedures (cf. Piotrowski 1975):

(i) mechanical identification of the membership of a word to a word class;

(ii) identification of terminologically and semantically dominating units;
(iii) ascertaining of the stylistic relevance of the word;
(iv) measurement of the stylistic individuality of a text;
(v) diagnosing the foci of some psychic diseases;
(vi) construction of learning automata.

Literatur

Bektaev, K.B., Luk'janenkov, K.F., O zakonach raspredelenija edinic pis'mennoj reči. In: Piotrowski, R.G. (Ed.), Statistika reči i automatičeskij analiz teksta. Leningrad, Nauka 1971, 47-112.

Brainerd, B., Article Use as an Indicator of Style among English-Language Authors. In: Jäger, S. (Hrsg.), Linguistik und Statistik. Braunschweig, Vieweg 1972, 11-32.

Bunge, M., Kinds and Criteria of Scientific Laws. Philosophy of Science 28, 1961, 260-281.

Bunge, M., Scientific Research I. Berlin, Springer 1967.

Frumkina, R.M., O zakonach raspredelenija slov i klassov slov. In: Mološnaja, T.N. (Hrsg.), Strukturno-tipologičeskie issledovanija. Moskva, AN SSSR, 1962, 124-133.

Johnson, N.L., Kotz, S., Discrete Distributions. Boston, Houghton Mifflin 1969.

Kemp, C.D., Kemp, A.W., Generalised Hypergeometric Distributions. Journal of the Royal Statistical Society, Series B 18, 1956a, 202-211.

Kemp, C.D., Kemp, A.W., The Analysis of Point Quadrat Data. Australian Journal of Botany 4, 1956b, 167-174.

Maškina, L.E., O statističeskich metodach issledovanija leksiko-grammatičeskoj distribucii. Minsk 1968 (Diss.).

Ord, J.K., Families of Frequency Distribution. London, Griffin 1972.

Orlov, In.K., Boroda, M.G., Nadarejšvili, I.Š., Sprache, Text, Kunst. Bochum, Brockmeyer 1980 (in press).

Paškovskij, V.E., Srebrjanskaja, I.I., Statističeskie ocenki pis'mennoj reči bol'nych šizofreniej. In: Inženernaja lingvistika. Leningrad 1971.

Piotrowski, K.G., Tekst, mašina, čelovek. Leningrad, Nauka 1975.

KANONISCHE ANALYSE VON KONTINGENZTAFELN[1]

J. Lauber, J. Linhart, Prag

In diesem Aufsatz nöchten wir den Leser mit einer der weniger bekannten Analysetechniken von Kontingenztafeln bekannt machen. Formal geht es um ein Verfahren, das mittels einer Transformation der Klassifikationsmerkmale (bzw. Dimensionen) der Kontingenztafel in eine zweidimensional normalverteilte Zufallsvariable den Korrelationskoeffizienten zwischen den Merkmalen maximalisiert und das für jede Kategorie (Ausprägungsstufe) der beiden Merkmale einen Score bestimmt.

Die Technik kann im Hinblick auf das Ziel und den verwendeten mathematischen Apparat unter die multivariaten Methoden der Analyse von soziologischen Daten eingeordnet werden. Die Vorgehensweise wurde sehr allgemein formuliert, so daß sie für alle Typen von Kontingenztafeln verwendet werden kann - für Tabellen, die durch Klassifizieren aufgrund von ordinal- und intervallskalierten Merkmale entstanden sind - wobei die Tabellen beliebig groß sein können. Die Interpretation der Ergebnisse ist nicht immer einfach, sie erfordert das Verstehen der Grundlagen der Methode. Die vorgestellte Technik erlaubt uns vor allem einen kompakten Überblick über die Struktur der Zusammenhänge zwischen den Merkmalen der Kontingenztafel. Dieser Überblick sollte durch eine detailliertere Analyse vervollständigt werden.

MATHEMATISCHE GRUNDLAGE DER KANONISCHEN ANALYSE

Die Methode basiert vor allem auf dem Lancaster-Satz (LANCASTER 1957; KENDALL, STUART 1973):

Wir wollen annehmen, daß die Zufallsvariable F = (X,Y) eine zweidimensionale Normalverteilung mit der Kovarianzmatrix

$$\sum = \begin{bmatrix} \sigma_1^2, \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1, \sigma_2^2 \end{bmatrix},$$

den Varianzen $D(X) = \sigma_1^2$ und $D(Y) = \sigma_2^2$ und der Kovarianz, bzw. der Korrelation

$$\text{cov}\ (XY) = \rho\sigma_1\sigma_2 \text{ und } \text{corr}\ (XY) = \rho$$

darstellt.

Transformiert man die einzelnen Komponenten der gegebenen zweidimensionalen Verteilung F, d.h.

$$X' = X'(X) \quad \text{und} \quad Y' = Y'(Y),$$

dann gilt unter sehr allgemeinen Bedingungen (Endlichkeit der Streuungen der Größen X' und Y')

$$\text{corr}\ (X'Y') \leq |\rho|,$$

d.h. die Korrelation der transformierten Verteilung kann im absoluten Wert die Korrelation der ursprünglichen Normalverteilung nicht überschreiten.

Der Lancaster-Satz gilt auch für nicht kategorisierte Größen. Betrachten wir weiter eine Kontingenztabelle (Tab. 1), die die absoluten Vorkommenshäufigkeiten aller Paare von Kategorien (Ausprägungsstufen) der Merkmale (Dimensionen) X und Y zusammen mit Zeilen- und Spaltensummen enthält (Dimension X soll r Klassen haben, und Dimension Y soll c Klassen haben).

Gegenstand der weiteren Überlegungen ist die Frage, auf welche Art und Weise den einzelnen Ausprägungen der Dimensionen X und Y bestimmte Zahlenwerte (Scores) zugeschrieben werden sollen, damit der Korrelationskoeffizient bei gegebenen Häufigkeiten sein Maximum erreicht.

Tabelle 1.

Merkmal Y						
Klasse / Klasse		1	2	...	c	Σ
Merkmal X	1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
	.	.	.	...	.	.
	.	.	.	...	.	.
	.	.	.	...	.	.
	r	n_{r1}	n_{r2}	...	n_{rc}	$n_{r.}$
	Σ	$n_{.1}$	$n_{.2}$	...	$n_{.c}$	n

Mit anderen Worten: wir wollen die Werte der diskreten Zufallsgrößen X und Y (zwecks Eindeutigkeit standardisiert), d.h. Zahlen (Score) x_i und y_j $(i = 1,2,\ldots,r)$ und $(j = 1,2,\ldots,c)$ so finden, daß die Randverteilung

X:	x_i	x_1	x_2	...	x_r
	$P(X = x_i)$	$\frac{n_{1.}}{n}$	$\frac{n_{2.}}{n}$		$\frac{n_{r.}}{n}$

Y:	y_j	y_1	y_2	...	y_c
	$P(Y = y_j)$	$\frac{n_{.1}}{n}$	$\frac{n_{.2}}{n}$		$\frac{n_{.c}}{n}$

die Einheitsvarianzen

$$D(X) = \sum_{i=1}^{r} \frac{n_{i.}}{n} x_i^2 = 1,$$

$$D(Y) = \sum_{j=1}^{c} \frac{n_{.j}}{n} y_j^2 = 1 \text{ hat,}$$

wobei der Korrelationskoeffizient

$$\text{corr } (X,Y) = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{n_{ij}}{n} x_i y_j \longrightarrow \text{maximal ist.}$$

Eine unmittelbare Ausnutzung des angeführten Lancaster-Satzes, d.h. eine direkte Konstruktion der zweidimensionalen Normalverteilung, kommt nicht in Frage. Das Auffinden der optimalen Werte x_i und y_j (hinsichtlich der Maximalisierung des Korrelationskoeffizienten) können wir als eine Aufgabe der Suche nach einem gebundenen Extremum unter Randbedingungen lösen.

Bezeichnet man

$$f(x_1,x_2,\ldots,x_r,y_1,y_2,\ldots,y_c) = \sum_i \sum_j \frac{n_{ij}}{n} x_i y_j = \text{corr } (XY),$$

$$g_1(x_1,x_2,\ldots,x_r) = \sum_i \frac{n_{i.}}{n} x_i^2 - 1 = 0,$$

$$g_2(y_1,y_2,\ldots,y_c) = \sum_j \frac{n_{.j}}{n} y_j^2 - 1 = 0,$$

so bekommt man die Aufgabe, das Maximum von f unter den Bedingungen $g_1 = g_2 = 0$ zu suchen. Diese Aufgabe kann man mit Hilfe von Lagrange-Multiplikatoren lösen. Da die Kenntnis der Art der Bedingungen und Beziehungen, die von x_i und y_j erfüllt werden, zur Interpretation wichtig ist, führen wir die zur Lösung der Extremwertaufgabe verwendeten Hauptverfahren auf.

Zunächst bilden wir die Lagrange-Funktion:

$$\varphi = f(x_1,x_2,\ldots,x_r,y_1,y_2,\ldots,y_c) + \lambda g_1(x_1,x_2,\ldots,x_r) +$$

$$+ \mu g_2(y_1,y_2,\ldots,y_c).$$

Die für das Extrem nötigen Bedingungen sind:

$$(1)\quad \frac{\partial\varphi}{\partial x_i} = \sum_j \frac{n_{ij}}{n} y_j + 2\lambda \frac{n_{i.}}{n} x_i = 0 \quad (i = 1,2,\ldots,r)$$

$$(2)\quad \frac{\partial\varphi}{\partial y_j} = \sum_i \frac{n_{ij}}{n} x_i + 2\lambda \frac{n_{.i}}{n} y_j = 0 \quad (j = 1,2,\ldots,c)$$

Aus den Beziehungen (1) und (2) folgt die Gültigkeit der Gleichungen

$$\sum_i \frac{\partial\varphi}{\partial x_i} x_i = 0, \qquad \sum_j \frac{\partial\varphi}{\partial y_j} y_j = 0,$$

aus denen sich nach dem Einsetzen

$$\sum_i \sum_j \frac{n_{ij}}{n} x_i y_j + 2\lambda D(X) = 0,$$

$$\sum_i \sum_j \frac{n_{ij}}{n} x_i y_j + 2\mu D(Y) = 0$$

oder

$$\text{corr}\,(XY) = R = -2\lambda D(X) = -2\mu D(Y)$$

ergibt. Bei standardisierten Größen (d.h. $D(X) = D(Y) = 1$) muß dann gelten

$$(3)\quad \text{corr}\,(XY) = R = -2\lambda = -2\mu.$$

Aufgrund von (3) kann man die Bedingungen (1) und (2) in Form eines linearen Gleichungssystems schreiben:

$$(4)\quad \sum_j n_{ij} y_j = R n_{i.} x_i, \quad (i = 1,2,\ldots,r),$$

$$(5)\quad \sum_i n_{ij} x_i = R n_{.j} y_i, \quad (j = 1,2,\ldots,c).$$

Es kann bewiesen werden (Kendall 1973), daß das lineare Gleichungssystem (4) und (5) eine nicht triviale Lösung (ungleich Null) nur für die Werte R besitzt, die den Wurzeln der Eigenwerte der Matrix entsprechen, die durch eine einfache Transformation der ursprünglichen Kontingenztafel entsteht.

Für die Existenz einer nicht trivialen Lösung (4) und (5) muß konkret gelten:

$$(6)\quad NN'u = R^2 u,$$

wobei N die Matrix mit den Elementen

$$(7)\quad n^*_{ij} = \frac{n_{ij}}{\sqrt{n_{i.} n_{.j}}}, \qquad (i = 1,2,\ldots,r),\ (j = 1,2,\ldots,c)$$

N' die transponierte Matrix N,
u der Spaltenvektor mit den Komponenten

$x_i\sqrt{n_i}$ $(i = 1,2,\ldots,r)$ ist.

Aus der Beziehung (6) ist ersichtlich, daß für die Existenz einer nicht trivialen Lösung die Gültigkeit der Gleichung

$$\det[NN' - R^2 J] = 0 \qquad (J\text{: Einheitsmatrix})$$

nötig ist, d.h. R^2 ist der Eigenwert der Matrix NN'.

Der Rang der Matrix N beträgt maximal

$$m = \min(r,c),$$

so daß der Rang der Matrix NN' ebenfalls höchstens m ist. Im allgemeinen hat die Matrix NN' m verschiedene Eigenwerte

$$\lambda_0, \lambda_1, \ldots, \lambda_{m-1}, \text{ wobei immer } \lambda_0 = R^2 = 1$$

einen Eigenwert der Matrix NN' darstellt (und zwar den größten), weil das Gleichungssystem (4) und (5) für R = 1 immer eine Lösung hat, die ungleich Null ist:

$$x_i \equiv K,\ y_j \equiv K \qquad (i = 1,2,\ldots,r),\ \ (j = 1,2,\ldots,c);$$

mit Rücksicht auf die Bedingungen D(X) = D(Y) = 1 bekämen wir aber

$$x_i \equiv 1 \text{ und } y_j \equiv 1.$$

Diese Werte hängen von der ursprünglichen Kontingenztabelle überhaupt nicht ab, ihre Kenntnis beinhaltet demnach keine Information. Der zu dem Einheitseigenwert zugehörige Eigenvektor u^0 hat die Komponenten $\sqrt{n_{i.}}$, bzw. der normierte Eigenvektor hat die Komponenten $\sqrt{n_{i.}/n}$, die nur über die relativen Größen der Marginalwerte Auskunft geben.

Die weiteren Eigenwerte, in nicht steigender Sequenz angeordnet:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{m-1}$$

ergeben zusammen mit den ihnen entsprechenden Eigenvektoren $u^1, u^2, \ldots, u^{m-1}$ schon eine nützliche Information.

Da die Eigenvektoren die Komponenten

$$u_i = x_i\sqrt{n_{i.}}$$

haben, bekommen wir aufgrund von u^1

$$(8) \quad x_i = u_i^1/\sqrt{n_{i.}}\ .$$

Die Werte y_j bekommen wir dann unter Ausnutzung der Beziehung (5), oder als

$$(9) \quad y_j = \frac{1}{Rn_{.j}} \sum_i n_{ij}x_i \qquad (j = 1,2,\ldots,c).$$

In der gleichen Weise kann man zeigen, daß auch die Beziehung

$$N'Nv = R^2v$$

gilt, wo v ein Vektor mit den Komponenten $y_j\sqrt{n_{.j}}$ ist. Rechnerisch ist es günstig, die Aufgabe des Findens der Eigenwerte und Eigenvektoren für diejenige der Matrizen NN' und N'N zu lösen, die die kleinere Ordnung hat, und die Werte der übrigen kanonischen Größen mit Hilfe von (5) oder (6) zu ermitteln.

Nun zeigen wir einige Verbindungen der beschriebenen Technik zu "traditionelleren" Verfahren der Analyse von Kontingenztafeln. Die Diagonalelemente der Matrix NN' sind gleich

$$NN'_{ii} = \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}},$$

also gilt für die Spur der Matrix NN'

$$trNN' = \sum_i \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}}.$$

Da

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n} = n\left\{ \sum_i \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 \right\}$$

ist, und die Spur der Matrix die Summe der Eigenwerte darstellt,

können wir die Größe χ^2 in Form von

$$(10) \quad \chi^2 = n(\lambda_1 + \lambda + \ldots + \lambda_{m+1})$$

ausdrücken.

Die Signifikanz der Größe χ^2 ermitteln wir mit Hilfe der üblichen Tabellen der Chi-Quadrat Verteilung für ein im voraus gewähltes Signifikanzniveau und für die Anzahl der Freiheitsgrade

$$FG = (c - 1)(r - 1).$$

DIE ALGORITHMISIERUNG VON BERECHNUNGEN UND DIE GRUNDEIGENSCHAFTEN DER KANONISCHEN VARIABLEN

Die Berechnung der Werte x_i und y_j, die den einzelnen Kategorien der Kontingenztafel zugeordnet werden, benötigt einen solchen Umfang an Rechenarbeiten, daß es unerläßlich ist, auch für Tafeln kleineren Ausmaßes die EDV zu nutzen. Die Berechnung der Eigenwerte und Eigenvektoren der symmetrischen Matrix NN', bzw. N'N, ist am anspruchsvollsten. Zu diesem Zweck kann man zum Beispiel einen der von Wilkinson angegebenen Algorithmen empfehlen (Wilkinson 1976). Im erarbeiteten Programm[2] wurde die QR-Methode verwendet, die sich gut bewährt hat.

Wir zeigen einzelne Rechenschritte und einige Eigenschaften der Werte x_i und y_i anhand eines einfachen Beispiels, das auf verschiedene Weisen modifiziert wird.

Die Kontingenztafel (Tab. 2) enthält die Ausgangsdaten für die kanonische Analyse. Die Matrix N (sie entsteht aus Tab. 1 mit Anwendung der Beziehung (7)) hat die Form

$$N = \begin{bmatrix} 0.6227524 & 0.0668153 & 0.2425356 \\ 0.3636670 & 0.5518254 & 0.1669245 \\ 0.0862796 & 0.2715377 & 0.6160411 \end{bmatrix}$$

und die Matrix NN' beinhaltet die Werte

$$N = \begin{bmatrix} 0.4511083 & 0.3038300 & 0.2212856 \\ 0.3038300 & 0.4646288 & 0.2840508 \\ 0.2212856 & 0.2840508 & 0.4606835 \end{bmatrix}$$

Tabelle 2.

	1	2	3	Σ
1	110	10	40	160
2	70	90	30	190
3	15	40	100	155
Σ	195	140	170	505

Die Eigenwerte der Matrix NN' sind:

$$\lambda_0 = 1.0000, \quad \lambda_1 = 0.2366, \quad \lambda_2 = 0.1398.$$

Eigenvektoren sind:

u^0	u^1	u^2
0.5629	0.6328	0.5318
0.6134	0.1114	-0.7819
0.5540	-0.7663	0.3254

Aus dem Vektor u^1 ermitteln wir durch Ausnutzung der Beziehung (8) die Werte der Variablen x_i (i = 1,2,3):

$$x_1 = 0.05, \qquad x_2 = 0.0081, \qquad x_3 = -0.0615.$$

Die Werte der Komponenten von Eigenvektoren sind, bis auf eine multiplikative Konstante, immer festgelegt. Die errechneten Werte sind üblicherweise normiert, d.h. zum Beispiel

$$\sum_{i=1}^{3} (u_i^0)^2 = 0{,}5629^2 + 0{,}6134^2 + 0{,}5540^2 = 1.$$

Für die Gültigkeit der Beziehungen D(X) = D(Y) = 1, d.h. speziell für

$$D(X) = \sum_{i=1}^{r} \frac{n_{i.}}{n} x_i^2 = 1$$

ist es notwendig, daß

$$\sum_{i=1}^{r} \frac{n_{i.}}{n} \frac{(u_i^1)^2}{n_{i.}} = \sum_{i=1}^{r} \frac{(u_i^1)^2}{n} = 1$$

gilt, so daß die Komponenten der normierten Eigenvektoren mit dem Faktor $\sqrt{n}$ multipliziert werden müssen, bzw. daß die kanonischen Variablen, die aus normierten Vektoren erhalten worden sind, mit dieser Konstante multipliziert werden müssen. Nach dieser Berichtigung bekommen wir die Werte der kanonischen Variablen

$$x_1 = 1,1236, \quad x_2 = 0,1820, \quad x_3 = -1,3820.$$

Die Werte der Variablen y_j ermitteln wir aus der Beziehung (5), konkret:

$$y_1 = 1,2180, \quad y_2 = -0,4067, \quad y_3 = -1,0629.$$

Den Wert χ^2 bestimmen wir entsprechend der Beziehung (10)

$$\chi^2 = n(\lambda_1 + \lambda_2 + \ldots + \lambda_{m-1}) = 505\ (0,2366+0,1398) = 190,1$$

was mit dem auf "klassische" Weise berechneten Resultat

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - o_{ij})^2}{o_{ij}}$$

übereinstimmt.

Der Korrelationskoeffizient ist bei den Werten x_i und y_1 für die Angaben in der Tabelle 2 gleich $R^2 = 0,2366 = 0,4864^2$, was wiederum mit dem durch eine direkte Berechnung erhaltenen Wert,

in dem jedes Wertpaar (x_i, y_j) mit der Häufigkeit n_{ij} vertreten ist, übereinstimmt.

Die Beziehungen (4) und (5), die von den Werten x_i und y_j erfüllt werden, können bei ihrer Interpretation helfen. Aufgrund der Gleichungen (4) und (5) gilt nämlich

$$\sum_j \frac{n_{ij}}{n_{i.}} y_j = Rx_i \qquad (i = 1,2,\ldots,r)$$

$$\sum_i \frac{n_{ij}}{n_{.j}} x_i = Ry_j \qquad (j = 1,2,\ldots,c),$$

wobei $\frac{n_{ij}}{n_{i.}}$ und $\frac{n_{ij}}{n_{.j}}$ Elemente der Matrizen der relativen Häufigkeiten sind (zeilen- bzw. spaltenweise relativiert).

Für die Kontingenztafel (tab. 2) ist die Matrix der relativen "Zeilenhäufigkeiten" Tab. 3 und die Matrix der relativen "Spaltenhäufigkeiten" Tab. 4.

Tabelle 3.

	11	12	13	Σ
1	0,6875	0,0625	0,2500	1,000
2	0,3684	0,4737	0,1579	1,000
3	0,0968	0,2581	0,6451	1,000

Tabelle 4.

	1	2	3
1	0,5641	0,0714	0,2353
2	0,3590	0,6429	0,1765
3	0,0769	0,2857	0,5882
Σ	1,0000	1,0000	1,0000

Die Beziehungen (4) und (5) können wir dann symbolisch so ausdrücken:

$$y_1 \begin{bmatrix} \text{1. Spalte} \\ \text{Tab. 3} \end{bmatrix} + y_2 \begin{bmatrix} \text{2. Spalte} \\ \text{Tab. 3} \end{bmatrix} + y_3 \begin{bmatrix} \text{3. Spalte} \\ \text{Tab. 3} \end{bmatrix} = R \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

x_1[1. Zeile Tab. 4] + [2. Zeile Tab. 4] + [3. Zeile Tab.] 4 =

$$= R \cdot [y_1, y_2, y_3],$$

wobei

$$R = \{x_i\}^{max}, \quad \{y_j\} \cdot \sum_i \sum_j n_{ij} x_i y_j$$

ist.

Aus bereits oben angeführten Überlegungen ist ersichtlich, daß die Werte der kanonischen Variablen x_i und y_j vom linearen Wachstum der Häufigkeiten in der Kontingenztafel unabhängig sind, denn sie sind lediglich von relativen Häufigkeiten abhängig. Führen wir eine Permutation in der Reihenfolge der Zeilen und Spalten der Tabelle durch, so werden übereinstimmend die Reihenfolgen der Werte x_i und y_j verändert.

Vertauschen wir, zum Beispiel bei der Kontingenztafel mit den Werten der Tabelle 2, die zweite Zeile mit der ersten und gleichzeitig die zweite Spalte mit der ersten, was den Veränderungen in der Reihenfolge der Merkmalskategorien entsprechen würde, so bekommen wir die Kontingenztafel (Tab. 5).

Tabelle 5.

	1	2	3	Σ
1	90	70	30	190
2	10	110	40	160
3	40	15	100	155
Σ	140	195	170	505

Für die Angaben in Tab. 5 bekommen wir folgende Werte:

$$\lambda_0 = 1{,}0000, \quad \lambda_1 = 0{,}2366, \quad \lambda_2 = 0{,}1398$$

$$x_1 = 0{,}1820 \qquad y_1 = -0{,}4067$$
$$x_2 = 1{,}1236 \qquad y_2 = 1{,}2180$$
$$x_3 = -1{,}3820 \qquad y_3 = -1{,}0629$$

Aus einem Vergleich dieser Werte mit den Ergebnissen für Tab. 2 ist ersichtlich, daß sich die Eigenwerte nicht verändert haben und daß bei den Werten x_i und y_j nur eine Veränderung der Reihenfolge zustandegekommen ist:

$$x_1 \longleftrightarrow x_2, \qquad y_1 \longleftrightarrow y_2.$$

Die χ^2-Werte ebenso wie die maximalen Korrelationskoeffizienten für die Tabellen 2 und 5 stimmen überein.

Durch eine Analyse dieser Ergebnisse gelangen wir beispielsweise zu dem Schluß, daß die Merkmale in der Tab. 2 (vor der Permutation) im Sinne der linearen Regression enger miteinanderverknüpft sind, auch wenn der χ^2-Wert für beide Tabellen gleich ist.

Konkrete Beispiele für die Anwendung der kanonischen Analyse von Kontingenztafeln findet der Interessent im Artikel von Charvát, Linhart und Vecerník (1978).

Anmerkungen:

1 Im Aufsatz wird die im Artikel Charvát, Linhart, Večerník (1978) verwendete Methode beschrieben.

2 Das Programm wurde für den Rechner NE 4120 zusammengestellt. Dieser Rechner wurde im Rechenzentrum von VŠE in Prag installiert.

Literatur:

CHARVÁT, F., LINHART, J., VEČERNÍK, J., Sociální mobilita v socialistických a kapitalistických zemích. Sociologický časopis 1978, č. 1, 40

KENDALL, M.G., STUART, A., The Advanced Theory of Statistics. Vol. 2: Inference and Relationship. London, Griffin, 1973

LANCASTER, H.O., Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Tables. Biometrika 44, 1957

WILKINSON, J.H., REINSCH, C., Handbook for Automation Computation. Linear Algebra. Berlin, Springer Verlag 1973

CURRENT BIBLIOGRAPHY

ABBREVIATIONS

CML	Zampolli, A., Calzolari, N. (Eds.): Computational and Mathematical Linguistics. Proceedings of the International Conference on Computational Linguistics, Pisa, 27.8. to 1.9.1973, Firenze: Olschki, 1980.
NTI	Naučno-texničeskaja informacija. Serija 2: Informacionnye processy i sistemy
PSML	Prague Studies in Mathematical Linguistics 7, 1981.
Symposium	Symposium "Computational Linguistics and Related Topics", Tallinn, November 24-26, 1980. Summaries. Tallinn 1980.
Trudy 5,6,7	Trudy po lingvostatistike 5, 6 (1980); 7 (1981), Tartuskij gosudarstvennyj universitet, Tartu.

GENERAL

1. ARAPOV, M.V.: Teorija sistem i izučenie estestvennogo jazyka: ponjatija simmetrii i statusa [Systems theory and natural language investigation: concepts of symmetry and status]. Gvišiani, D.M. (Ed.): Sistemnye issledovanija. Metodologičeskie problemy, 1981. Moskva: Nauka 1981, 121-141.

2. ARAPOV, M.V., KRYLOV, Y.K.: Mathematical models of classification in application to some problems of statistical linguistics. Symposium, 14-16.

3. COLLIANDER, P.: Methoden zur Berechnung des Informationswerts einer in der Linguistik einsetzbaren Matrix. Kopenhagener Beiträge zur Germanistischen Linguistik 16, 1980, 1-33.

4. DZHUBANOV, A., KHASANOV, B.: Computational description of the Kazakh language. CML, Vol. 1, 75-77.

5. GERŠIČ, Sl., ALTMANN, G.: Laut - Silbe - Wort und das MENZERATHsche Gesetz. Wodarz, H.W. (Ed.): Frankfurter phonetische Beiträge III, Hamburg 1980, 115-123.

6. GOEBL, H.: Typologia quantitativa oder Cosi fan tutte. Grazer Linguistische Studien 11/12, 1980, 103-117.

7. KLAVINA, S.: Statistica valodniecībā: Mācību līdzeklis [Statistics in Linguistics]. Riga 1980, 145 pp.

8. LECOMTSEV, Y.K.: On some rules for linguistic polarity. Symposium, 59-61.

9. MOSKOVICH, W., CAPLAN, R.: Distributive-statistical techniques in linguistic and literary research. Advances in Computer-Aided Literary and Linguistic Research, Birmingham 1979, 245-263.

10. MULLER, C.: Langue française et linguistique quantitative. Genève: SLATKINE Reprints, 1979, 470 pp.

11. NOSENKO, I.A.: Načala statistiki dlja lingvistov [The Principles of Statistics for Linguists]. Moskva: Vyšaja škola, 1981, 157 pp.

12. PAPP, F.: Lingvostatistika i vengerskij jazyk [Linguistic statistics and the Hungarian language]. Trudy 5, 15-37.

13. POPOVA, G.V.: On the practical use of reliability criteria in manual and automatic abstracting. Symposium, 91-93.

14. SAUKKONEN, P.: Statistical linguistic research in Finland. Trudy 5, 5-14.

15. ŠUBIK, S.A.: Statističeskie methody v lingvistike [Statistical methods in linguistics]. Piotrovskij, R.G. et al. (Eds.): Statistika reči i avtomatičeskij analiz teksta, Leningrad: Nauka, 1980, 52-63.

16. SUKHOTIN, B.V.: Experiments on recognizing linguistic objects by means of algorithms of linguistic deciphering. Symposium, 102-103.

17. TĚŠITELOVÁ, M.: Využití statistických metod v gramatice [The Use of Statistical Methods in Grammar]. Praha: Academia, 1981, 220 pp.

18. TUZOV, V.A.: Matematičeskaja model' jazyka [A Mathematical Model of Language]. Leningrad 1980, 46 pp.

19. WILDGEN, W.: Differentielle Linguistik. Entwurf eines Modells zur Beschreibung und Messung semantischer und pragmatischer Variation. Tübingen: Niemeyer 1977, 293 pp.

PHONOLOGY

20. ČIKADZE, V.G.: Statističeskij analiz fonem v odnosložnyx slovax tureckogo jazyka [The statistical analysis of phonemes in one-syllable words in Turkish]. Soobščenija AN Gruzinskoj SSR 96, 1979, 201-209.

21. DŽUBANOV, A.K.: K voprosu o grafemnoj statistike kazaxskogo teksta [On the problem of graphemic statistics in a Kazakh text]. Voprosy kazaxskoj fonetiki i fonologii, Alma-Ata 1979 , 79-86.

22. FEER, B.B.: Distribucija soglasnyx v tonal'nom ketskom jazyke (Pakulixinskij govor) [The distribution of consonants in tonal Ket language (Pakulixinsk dialect)]. Fonetika sibirskix jazykov, Novosibirsk 1970 , 98-120.

23. GLUŠKOVA, I.P.: Fonologičeskaja paradigmatika i sintagmatika jazyka maratxi [The Phonological Paradigmatic and Syntagmatic of Marathi]. Avtoreferat kandidatskoj dissertacii, Moskva 1980, 26 pp.

24. GRIGORIEV, V.I.: Frequency distribution of letters and their ranks in a running text. Symposium, 43-47.

25. HORECKÝ, J., NEMCOVÁ, E.: The use of entropy in evaluating the degree of completeness in the phonological calculus. PSML, 47-58.

26. HUG, M.: La distribution des phonèmes en français. Die Phonemverteilung im Deutschen. Essais statistiques. Genève: SLATKINE Reprints, 1979, 192 pp.

27. ISENGEL'DINA, A.A.: Tipologija sočetanij soglasnyx kazaxskogo i russkogo jazykov v ix forme i funkcii [A typology of the alternations of consonants in the Kazakh and Russian languages in their form and function]. Voprosy kazaxskoj fonetiki i fonologii, Alma-Ata 1979, 99-130.

28. KAASIK, Ü., TULDAVA, Ju.: Sõnalõpu grafeemotaktikast eestikeelses tekstis. [On the graphemotactics of word endings in Estonian texts]. Trudy 5, 51-98.

29. KNOWLES, F.: The quantitative syntagmatic analysis of the Russian and Polish phonological systems. CML, Vol. 1, 79-81.

30. KURENOV, S., GOKLENOV, Dž.: Količestvennaja xarakteristika bezudarnyx glasnyx v mnogosložnyx slovax turkmenskogo jazyka [Quantitative characteristics of unstressed vowels in polysyllabic words in Turkmen]. Izvestija An Turkmenskoj SSR. Serija obščestvennyx nauk 1979/6, 50-56.

31. LIEPA, E.: Vokālisma un zilbju kvantitāte latviešu literārajā valodā [The Quantity of Vocalism and Syllables in the Latvian Literary Language]. Riga 1979.

32. PETRJANIKINA, V.I.: Častotnyj komponent intonacii kommunikativnyx tipov vyskazyvanij v jazyke kikongo [The frequency component of the intonation of communicative types of utterances in Kikongo language]. Intonacija i fonologija, Moskva 1980, 56-66.

33. PETRJANIKINA, V.I.: Rol' častotnogo komponenta intonacii v differenciacii kommunikativnyx tipov vyskazyvanij v russkom jazyke i jazykax Afriki [The role of the frequential component of intonation in the differentiation of communicative types of utterances in Russian and in African languages]. Intonacija i fonologija. Moskva 1980, 3-12.

34. ROCŁAWSKI, B.: System fonostatystyczny współczesnego języka polskiego [The Phonostatistical System of Contemporary Polish]. Wrocław: Zakład Narodowy im. Ossolińskich. Wydawnictwo Polskiej Akademii Nauk, 1981, 178 pp.

35. SABOL, J.: The prosodic structure of the Slovak word. PSML, 59-80.

36. SENDLMEIER, W.F.: Der Einfluß der Quantität und Qualität auf die Perzeption betonter Vokale des Deutschen. Phonetica 38, 1981/5-6, 291-308.

37. TULDAVA, Ju.: Eesti keele sõnavara foneetilis-grafeemilised mõõted [Phonetic-Graphemic measures of Estonian lexics]. Trudy 5, 51-98.

38. VEENKER, W.: Zur phonologischen Statistik der čeremissischen (marischen) Schriftsprachen. Sovremennoe finno-ugrovedenie, Tallin 1980/2, 106-134.

39. VERNER, G.N., TAMBOVCEV, Ju.A.: Nekotorye rezul'taty fonostatičeskogo analiza zvukovogo sostava jugskogo jazyka [Some results of the phonostatistical analysis of the tonal inventory of the Yug language]. Teoretičeskie voprosy fonetiki i grammatiki jazykov narodov SSSR, Novosibirsk 1979, vypusk 1, (Ėksperimental'nye issledovanija), 47-52.

MORPHOLOGY

40. ARAPOV, M.V., HERZ, M.M.: Frequency and age as characteristics of a word. CML Vol. 1, 3-7.

41. BERGER, T.: The formation of the imperative in modern Russian. Russian Linguistics 6, 1981, 65-80.

42. GRYAZNUKHINA, T.A., KOMAROVA, L.I., KRITSKAYA, V.I.: Functional peculiarities of the noun in scientific abstracts. Symposium, 48-51.

43. HOOGE, D.J.: O nekotoryx kvantitativnyx aspektax upotreblenija modal'nyx glagolov (na materiale nemeckogo jazyka) [On some quantitative aspects of the use of modal verbs

in German]. Linguistica 13, 1980, 55-63.

44. JAKUBOVKA, N.K.: Rasstanovka affiksov v uzbekskix slovoformax [The arrangement of affixes in Uzbek word forms]. Tăskent: Izdatel'stvo "Fan" Uzbekskoj SSR, 1979, 108 pp.

45. KAASIK, Ü, TULDAVA, Ju.: Some problems of the automatic morphological analysis of word forms in Estonian texts. Symposium, 54.

46. PETRAVIČIUS, L.: Zur Valenz des Verbs in der deutschen wissenschaftlichen Periodik des 19. - 20. Jh. Kalbotyra 31, 1980/4, 35-53.

47. POLJUŽKIN, M.M.: Rol' kvantitativnyx metodov v issledovanii funkcionirovanija affiksov [The role of quantitative methods in the study of the functioning of affixes]. Slovoobrazovanie i ego mesto v kurse obučenija inostrannogo jazyka 7, Vladivostock 1979, 72-89.

48. VAJŠNORAS, A.: Produktivnost' i častotnost' funkcionirovanija leksiko-sintaksičeskoj ėljativnoj formy v sovremennyx ital'janskom i francuzskom jazykax [The productivity and frequency of the functioning of the lexical-syntactic elative form in contemporary Italian and French]. Kalbotyra 31, 1980/5, 7-18.

SYNTAX

49. BEEBE, R.D.: The frequency distribution of English syntagms. CML, Vol. 1, 9-22.

50. NARO, A.J.: The social and structural dimensions of a syntactic change. Language 57, 1981, 63-98.

51. ŠTĚPÁN, J.: On the statistics of the complex sentence (The degree of dependence of subordinate clauses and thinking). PSML, 113-122.

SEMANTICS

52. ARAPOV, M.V.: Reguljarnost' semantičeskix otnošenij v estestvennom jazyke [The regularity of semantic relations in natural languages]. NTI 1980/9, 1-8.

53. HÉRAULT, D.: Compréhension automatique et spectre sémantique. Paris: Jean-Favard 1981, 289 pp.

54. KARAULOV, Ju.N.: Častotnyj slovar' semantičeskix množitelej russkogo jazyka [A Frequency Dictionary of Semantic Factors in Russian]. Moskva: Nauka 1980, 207 pp.

55. LECOMTSEVA, M.I., LECOMTSEV, Y.K.: Towards algebraic modelling of linguistic semantics. Symposium, 62-71.

56. MOSKOWICH, W.: Polysemy in natural and artificial (planned) languages. Statistical Methods in Linguistics 1, 1977, 5-28.

57. RIEGER, B.: Linguistic semantics and the problem of vagueness: on analysing and representing word meaning. Advances in Computer-Aided Literary and Linguistic Research, Birmingham 1979, 271-288.

58. STERNIN, I.A.: Iz nabljudenij nad verojatnostnymi semami v značenii slova [Observations on probability semes in word meaning]. Sopostavitel'no-semantičeskie issledovanija russkogo jazyka, Voronež 1980, 14-20.

59. TĚŠITELOVÁ, M.: Sémantika a statistika [Semantics and statistics]. Slovo a slovesnost 51, 1980, 100-105.

LEXICOLOGY

60. ALEKSEEV, P.M., KAŠIRINA, M.E., TARASOVA, E.M. (Eds.): Častotnyj anglo-russkij fizičeskij slovar'-minimum [A Frequency English-Russian Minimum Physics Vocabulary]. Moskva: Voenizdat, 1980, 288 pp.

61. ALEKSEEV, P.M., KOPYLOVA, M.E.: O serijnyx dvujazyčnyx učebnyx častotnyx slovarjax [On serial bilingual frequency dictionaries for school use]. Trudy 6, 3-13.

62. ARZIKULOV, X.A., ARZIKULOVA, Ž.X.: O dvujazyčnyx častotnyx slovarjax-minimumax [On bilingual frequency vocabularies]. Naučnye doklady vysšej školy Filologičeskie nauki, Moskva 1980/3, 78-82.

63. AZAROVA, I.V., GOROKHOVA, S.I., KUZNETSOVA, E.L.: Machine-produced word indexes and concordances. Symposium, 17-18.

64. BRUNET, E.: Le vocabulaire de Jean Giraudoux. Structures et évolution. Genève: SLATKINE Reprints, 1978, 681 pp.

65. DOLPHIN, B.: Vocabulaire et lexique. Modèles mathématiques pour une linguistique quantitative. Genève: SLATKINE Reprints, 1980, 156 pp.

66. DUGAST, D.: Vocabulaire et discours. Essai de lexicométrie organisationnelle. Fragments de lexicologie quantitative. Avec trois séries de 730 couples de données rendant compte en 18 pages du roman de Maupassant "Fort comme la mort". Genève: SLATKINE Reprints, 1979, 137 pp.

67. EGIAZARJAN, Ė.V.: Nekotorye voprosy postroenija avtomatičeskix slovarej dlja mašinnogo perevoda [Some problems in the construction of automatic vocabularies for machine translation]. Strukturnyj analiz teksta. Erevan 1979, 79-89.

68. GERD, A.S.: A computer-aided terminological data bank in the light of the theory of scientific and technical lexicography. Symposium, 29-30

69. GRIGORIEV, H.G., MARUSENKO, M.A.: An algorithm of compiling a terminological dictionary word-list by a computer. Symposium, 39-41.

70. GRIGOREVA, A.S.: O častotnom slovare russkoj obixodnoj pis'mennoj reči [On a frequency dictionary of everyday written Russian]. Trudy 6, 24-31.

71. GUROVA, N.V.: Computational parameters of basic vocabulary selection from a word count. Symposium, 52-53.

72. JOHNSON, R.L.: Measures of vocabulary diversity. Advances in Computer-Aided Literary and Linguistic Research, Birmingham 1979, 213-217.

73. KVJATNOVSKAJA, T.A.: O dvujazyčnyx častotnyx slovarjax [On bilingual frequency dictionaries]. Izvestija Akademii Nauk Latvijskoj SSR, Riga 1981/2, 66-70.

74. LECOMTSEV, Y.K.: Notes on an algebraic model of lexical fields. Symposium, 55-58.

75. MALAXORSKIJ , L.V. : Principy častotnoj stratifikacii slovarnogo sostava jazyka [The Principles of Frequential Stratification of a Language's Vocabulary]. Leningrad: Nauka 1980, 223 pp.

76. MAMEDOVA, M.G., SKOROKHODKO, E.F.: Computerized network analysis of lexical systems. Symposium, 72-73.

77. MELIKJAN, N.A.: Častotnyj mikroslovar' pod-jazyka telegramm [A frequency micro-dictionary of the sublang-usage of telegrams]. Strukturnyj analiz teksta. Erevan 1979, 97-116.

78. MULLER, C.: Le vocabulaire tu théâtre de Pierre Corneille. Étude de statistique lexicale. Genève: SLATKINE Reprints, 1979, 379 pp.

79. MURAKI, S.: Comparison of basic vocabularies of Japanese and German [In Japanese]. Mathematical Linguistics 12, 1981, 356-366.

80. NAKANO, H.: The number of words in the word list by semantic principles [In Japanese]. Mathematical Linguistics 12, 1981, 376-381.

81. SEEGEBRECHT, A.-L.: Wortfrequenzvergleiche zwischen dem Čechischen und Finnischen. Dissertation. Freiburg i. Br. 1981.

82. TANAKA, Y.: Analysis of technical terms [In Japanese]. Mathematical Linguistics 12, 1981, 367-376.

83. TULDAVA, J.: A method for calculating vocabulary growth as a function of text length. Symposium, 110-114.

84. TULDAVA, Ju.: O teoretiko-metodologičeskix osnovax kvantitativno-sistemnogo analiza leksiki (I) [On the theoretical-methodological foundations of the quantitative-systemic lexicon analysis]. Linguistica 13, 1980, 143-158.

85. TULDAVA, Ju.: O teoretiko-metodologičeskix osnovax kvantitativno-sistemnogo analiza leksiki (2): Lingvističeskie aspekty issledovanija [On the theoretical methodological foundations of the quantitative-systematic lexicon analysis (2): The linguistic aspects of the research]. Linguistica 14, 1981, 114-133.

86. TUŠIŠVILI, M.A.: Nekotorye metody statističeskogo analiza vremennoj struktury rečevogo signala [Some statistical methods for analyzing the tense structure of speech signals]. Trudy Instituta sistem upravlenija Akademii Nauk GSSR, Tbilisi, 19, 1980/2, 144-153.

87. VIKS, Ü.: An automatic system for generation of dictionaries. Symposium, 119-120.

88. XOFFMANN, L.: Častotnye slovari pod"jazykov nauki i texniki - ključ k ponimaniju special'nyx tekstov [Frequency dictionaries of scientific and technological sublanguages - a key to the understanding of technical texts]. Trudy 6, 145-152.

Text Analysis

89. ALEKSEEV, P.M.: Computational linguistics and quantitative typology of text. Symposium, 5-7.

90. ALEKSEEV, P.M.: O kvantitativnoj tipologii teksta [On quantitative typology of text]. Trudy 7, 3-13.

91. ALEXANDROVA, L.N.: The postulates of Propp-Revzin and the repeat rate of parts of speech in texts samples. Symposium, 8-10.

92. ARAPOV, M.V.: Sistemnyj analiz leksičeskoj struktury tekstov [The systems analysis of the lexical structure of texts]. Gvišiani, D.M. (Ed.): Sistemnye issledovanija. Metodologičeskie problemy, 1980. Moskva: Nauka 1981, 372-402.

93. BAILEY, R.W.: Authorship attribution in a forensic setting. Advances in Computer-Aided Literary and Linguistic Research, Birmingham 1979, 1-20.

94. BEKTAEV, K.B.: Ėntropija tjurkskix pečatnyx tekstov [The entropy of Turkic printed texts]. Voprosy kazaxskoj fonetiki i fonologii, Alma Ata 1979, 43-52.

95. BELETSKAJA, I.P.: A knowledge storage system in dialogue mode in natural language. Symposium, 19-21.

96. BORODA, M.G.: Ob odnoj količestvennoj zakonomernosti ritmiki vokal'noj melodii [On a quantitative regularity in the rhythm of vocal melody]. G. Toradze (Ed.): Konferencia molodyx jazykovedov. Tezisy i materialy. Tbilisi 1980, 18-21.

97. BORODA, M.G., NADAREISHVILI, G.Sh.: On certain statistical characteristics of relations between text and melody in vocal music. Symposium 22-26.

98. BUSA, R.: How quantitative information on words, forms, and lemmas is presented in the Index Thomisticus. CML, Vol. II 887-892

99. DOLPHIN, C.: Méthodes de la statistique linguistique et vocabulaire fantastique de Malpertuis. Genève: SLATKINE

Reprints, 1980, 301 pp.

100. DUGAST, D.: Vocabulaire et stylistique. Théâtre et dialogue. Genève: SLATKINE Reprints, 1980, 293 pp.

101. FREY, E.: Subjective word frequency estimates and their stylistic relevance in literature. Poetics 10, 1981, 395-407.

102. GASPAROV, M.L.: Ital'janskij stix: sillabika ili sillabotonika? (Opyt izpol'zovanija verojatnostnyx modelej v sravnitel'nom stixovedenii) [Italian verse: syllabic or syllabo-tonic? (An attempt to use probability models in comparative verse research)]. Grigor'ev, V.P. (Ed.): Problemy strukturnoj lingvistiki 1978. Moskva 1981, 198-218.

103. GINDIN, S.I.: Ritmika, intonacija i smyslovaja kompozicija v poéme VI. Lugovskogo "Kak čelovek plyl s Odisseem" [Rhythm, intonation, and semantic composition in VI. Lugovskij's epic poem "Kak čelovek plyl s Odisseem"]. Grigor'ev, V.P. (Ed.): Problemy strukturnoj lingvistiki 1978. Moskva 1981, 230-265.

104. GRIDNEVA, L.M., DARCHUK, N.P., MURAVITSKAYA, M.P., PEREBEYNOS, V.I.: Symmetry elements in texts of scientific abstracts. Symposium, 35-38.

105. GROŠEVA, A.V.: Statističeskaja odnorodnost' tekstov na sintaksičeskom urovne [The statistical homogeneity of texts on the syntactic level]. Piotrovskij, R.G. et al. (Eds.): Statistika reči i avtomatičeskij analiz teksta, Leningrad: Nauka, 1980, 72-98.

106. IBRAGIMOV, S.I.: O nekotoryx strukturno-verojatnostnyx parametrax kirgizskogo naučno-texničeskogo teksta [On some structural-probabilistic parameters of a scientific-technical text in Kirghiz]. Sovetskaja tjurkologija 1979/5, 72-79.

107. JAKUBAJTIS, T.A., SKLJAREVIČ, A.N.: Korreljacionnye xarakteristiki častej reči v svjaznyx texstax [Correlational Characteristics of Parts of Speech in Interrelated Texts]. Riga: Zinatne, 1980, 135 pp.

108. KJETSAA, G.: A quantitative norm for the use of epithets in the age of Puškin. Russian Romanticism. Studies in the Poetic Codes. Stockholm Studies in Russian Literature 10, 1979, 204-226.

109. KJETSAA, G.: Stil' i norma [Style and norm]. Linguistica 14, 1981, 48-67.

110. KLIMEŠ, L.: On some quantitative aspects of the sentence in Loewenbrugk's Chronicle (1685). PSML, 123-136.

111. KOKKOTA, V.: Častotnost' anglijskix glagol'nyx form v naučno-texničeskoj literature [Frequencies of English verb forms in scientific and technical literature]. Trudy 6, 59-72.

112. KRASNOVA, E.A.: Leksiko-semantičeskie osobennosti žanra (Opyt statističeskogo analiza na materiale anglijskix i šotlandskix narodnyx ballad) [Lexical and semantic distinctions of a genre (Attempt at a statistical analysis based on English and Scottish folk ballads)]. Trudy 6, 73-85.

113. KRAUS, J.: Ličnye mestoimenija v češskix publicističeskix tekstax (Kvantitativnyj analiz) [Personal pronouns in Czech journalistic texts (Quantitative analysis)]. PSML, 27-46.

114. LEVIN, Ju.I.: Zamečanija o priloženii matematičeskoj statistiki dlja izučenija zavisimostej i svjazej meždu xarakteristikami xudožestvennyx tekstov [Notes on application of mathematical statistics to investigation of dependences and relations between parameters of literary texts]. Trudy 7, 46-59.

115. LIŠKOVÁ, Z.: Two signals of readability in journalistic utterance. PSML, 81-92.

116. MANASJAN, N.S.: O raspredelenijax terminov v anglijskom naučno-texničeskom tekste [On distribution analysis in

English scientific texts (Sublanguage of active Oscillators)]. Trudy 7, 60-73.

117. MARTEMYANOV, Y.S.: A word-correlated syntax and text analysis. Symposium, 74.

118. MARTYNENKO, G.Y.: Automatic classification of texts according to syntactic features. Symposium, 75-77.

119. MARUSENKO, M.A.: O izmerenii svjazi otraslevyx terminosistem s primeneniem ĖVM [Computer measuring of lexical connection of terminological systems relating to a certain branch of industry]. Trudy 7, 74-81.

120. MELETTI, F.: Proportion between number of words and number of forms in the works of the "Index Thomisticus". CML, Vol. 1, 89-98.

121. MIZUTANI, S.: Calculation of the distribution of word freqrencies [In Japanese]. Mathematical Linguistics 12, 1981, 339-356.

122. NADAREIŠVILI, I.S. Statističeskij analiz morfologičeskoj struktury teksta (Na materiale prozy K. Gamsaxurdija) [A statistical analysis of morphological text structure (Based on the prose of K. Gamsaxurdij)]. Izvestija Akademii Nauk GSSR. Serija jazykov i literatury, Tbilisi, 1980/3, 143-148.

123. NADAREISHVILI, I.Sh., ORLOV, Y.K.: On the effect of the group of frequent words upon the frequency structure and lexical concentration of a text. Symposium, 81-86.

124. OSTAPENKO, V.E.: Paraboličeskie i giperboličeskie zakonomernosti tekstoobrazovanija [Parabolic and hyperbolic regularities of text formation]. Trudy 6, 106-112.

125. PEREBEJNOS, V.I.: Raspredelenie glagolov v naučno-referativnom tekste [Distribution of verbs in scientific abstracts]. Trudy 7, 91-100.

126. PIOTROVSKIJ, R.G. et al. (Eds.): Statistika reči i avtomatičeskij analiz teksta [The Statistics of Speech and the Automatic Analysis of Texts]. AN SSSR Institut Jazykoznanija. Leningrad: Nauka, 1980, 223 pp.

127. REMMEL, M.: On text compression and rank statistics. Symposium, 94-96.

128. SAUKA, L.I.: O količestvennom analize stixopesennyx vstavok litovskix narodnyx skazok. [On the quantitative analysis of rhymed song-inserts in Lithuanian folk tales]. Trudy AN Litovskoj SSR, Serija obščestvennyx nauk 1980/1, 117-123.

129. SLEPAK, B.A.: "Prolegomeny" k statističeskoj teorii teksta ["Prolegomena" to the statistical theory of text]. Trudy 7, 101-119.

130. SLIVNJAK, D.I.: Statističeskie xarakteristiki mežsegmentnyx granic v dialoge [Statistical characteristics of intersegmental boundaries in the dialogue]. Trudy 7, 120-135.

131. TĚŠITELOVÁ, M.: On the language of the present-day publicist prose from the quantitative point of view. PSML, 9-26.

132. TULDAVA, Ju.A.: Opyt klassifikacii tekstov s pomoč'ju klaster-analiza [On a possibility of the realization of cluster-analysis]. Trudy 7, 136-157.

133. VALK-FALK, M.: An attempt of statistical analysis of melodic structures in two-part keyboard music. Symposium, 115-118.

134. VAŠÁK, P.: Tekstologija - sistema - stemma [Textology - system - stemma]. PSML, 137-147.

135. ŽILINSKENE, V.Ju.: Statističeskij analiz morfologii litovskix gazetno-žurnal'nyx tekstov (po dannym častotnogo slovarja litovskix gazetno-žurnal'nyx tekstov [A Statistical Analysis of the Morphology of Lithuanian Magazine and Newspaper Texts, Based on a Frequency Dictionary of Lithuanian Magazine and Newspaper Texts]. Avtoreferat kandidatskoj dissertacii, Vil'njus 1979, 24 pp.

statistical approach). Linguistica 13, 1980, 132-142.

DIALECTOLOGY

152. KUKORIN, V.N.: Distribucija glasnyx v čalkanskom dialekte altajskogo jazyka [(The distribution of vowels in the Čalkan dialect of Altai language]. Issledovanija zvukovyx sistem sibirskix jazykov, Novosibirsk 1979, 14-28.

153. VERTE, L.A.: Nekotorye distributivnye xarakteristiki soglasnyx fonem kazymskogo dialekta xantyskogo jazyka [A few distributive characteristics of consonantal phonemes of the Kazym dialect in Khanty]. Issledovanija zvukovyx sistem sibirskix jazykov, Novosibirsk 1979, 144-147.