

HANS-HERMANN BARTENS, KARL-HEINZ BEST (Göttingen)

WORTLÄNGEN IN ERZAMORDWINISCHEN TEXTEN

0. In dieser Untersuchung geht es um die Frage, welchen Regularitäten die Häufigkeit folgt, mit der Wörter unterschiedlicher Länge in erzamordwinischen Texten verwendet werden. Zu diesem Zweck kann man sich inzwischen auf weit verbreitetes linguistisches Wissen stützen, wie es sich in Fucks (1968 : 80, *passim*) oder Crystal (1993 : 86—87) niedergeschlagen hat: Immer wieder hat sich gezeigt, daß die längeren Wörter die selteneren, die kürzeren die häufigeren sowohl in den Sprachen als auch in den einzelnen Texten sind. In Häufigkeitspyramiden (z.B. König 1978 : 114) kommt dies besonders eindrucksvoll zum Ausdruck: Unter den 30 häufigsten Wörtern in geschriebenen deutschen Texten um 1900 kommt nur ein einziges zweisilbiges Wort vor, während am Fuß der Pyramide bereits dreisilbige Wörter erscheinen. Ganz ähnliche Ergebnisse zitiert Braun (1993 : 160—161) aus drei verschiedenen Untersuchungen zum Deutschen. Meier (1967 : 291) zeigt in einer eindrucksvollen Graphik, welche relativen Anteile den ein- bis fünfsilbigen Wörtern in einem deutschen Textkorpus von 50 000 Wörtern zukommen. Es müssen allerdings nicht, wie im Deutschen, immer die einsilbigen Wörter die häufigsten sein; in anders strukturierten Sprachen wie dem Erza- und Mokschamordwinischen kann der Häufigkeitsgipfel bei den zwei- oder gar dreisilbigen Wörtern liegen. Danach sinken die Häufigkeiten wie im Deutschen.

Man darf also feststellen: Es liegen gut dokumentierte Erkenntnisse dazu vor, daß zwischen der Länge und der Häufigkeit von Wörtern ein enger Zusammenhang besteht, der aber nicht unbedingt für das einzelne Wort, sondern für die unterschiedlichen Längenklassen der Wörter gilt.

Was bis hierher angeführt wurde, waren im wesentlichen Erhebungen rein statistischer Natur ohne explizite Hypothesen und ohne höheren theoretischen Anspruch. Es gibt nun aber einen relativ schmalen Strang linguistischer Forschung, den man mit den Arbeiten Thorndikes und Zipfs beginnen lassen kann. So kann man bereits bei Thorndike (1938 : 61) den interessanten Versuch einer Deutung der Kürzungstendenzen bei Wörtern nachlesen: "Such abbreviations will be more liked by the speaker than the original words not only because of their ease, but also because of the mental atmosphere of intimacy, security, and power which goes with nicknames in general." Daß diesen Bedürfnissen andere entgegenstehen, wird ebenfalls betont: "The reduction to a minimum of easily articulated phonemes might not be reached even among words used very often by very many people, because ease in the acts of the organs of speech may be counterbalanced by difficulty in distinguishing the words and connecting each with its meaning." (Thorndike 1938 : 63).

Während Thorndike sich also mit den Sprachbedürfnissen der Sprecher/Hörer befaßt und damit begründet, welche Motive zur Kürzung von Lexemen führen können bzw. diesen Kürzungstendenzen entgegenwirken, hat Zipf (1932 : 1) eine recht modern anmutende Hypothese, sein *Principle of Relative Frequency* formuliert: "The ... degree of conspicuousness of any word, syllable, or sound is inversely proportionate to the relative frequency of that word, syllable, or sound, among its fellow words, syllables, or sounds in the stream of spoken language. As any element's usage becomes more frequent, its form tends to become less accented, or more easily pronounceable, and *vice versa*."

Es ist nun kein Problem, wesentliche Aspekte der Auffassungen Thorndikes und Zipfs in aktuellen Versuchen einer linguistischen Theoriebildung wiederzuerkennen. So hat vor allem R. Köhler (1986) in seinem Konzept einer linguistischen Synergetik verschiedene sprachliche Entitäten ("Systemgrößen" genannt), darunter die Worthäufigkeit und -länge, in einem systematischen Zusammenhang mit Bedürfnissen der Sprecher ("Systembedürfnisse") gebracht und einen selbstregulierenden Regelkreis (Köhler 1986 : 74) entwickelt, in dem jede Systemgröße funktional von andern Systemgrößen und/oder -bedürfnissen abhängig ist. So ist die Länge eines Wortes abhängig von seiner Frequenz, aber auch vom Umfang des Phoneminventars der betreffenden Sprache und der Größe ihres Lexikons. Die Zusammenhänge sind klar: Will man z.B. 10 000 Wörter in einer Sprache differenzieren können, die nur 20 Phoneme im Inventar zur Verfügung stellt, so müssen diese im Durchschnitt länger sein als bei einer anderen Sprache, die ein Inventar von 50 Phonemen bereit hält, vorausgesetzt, alle anderen Bedingungen wie Phonetik, phonologische Relevanz der Akzente oder Töne, etc. sind gleich.

1. In dieser Untersuchung geht es um einen anderen Aspekt des Zusammenhangs zwischen Länge und Frequenz von Wörtern: Wie gestaltet sich die Interaktion zwischen Wörtern verschiedener Längen innerhalb von einzelnen Texten? Die oben erwähnten Häufigkeitspyramiden beziehen sich auf textübergreifende Korpora; es wäre ja immerhin denkbar, daß dieser Zusammenhang in einzelnen Texten nicht oder zumindest weniger deutlich zum Ausdruck kommt als bei größeren Textsammlungen insgesamt.

Zu diesen Problemen wurden bisher zu über 30 Sprachen (vgl. Best, Altmann 1996), darunter auch etliche finnisch-ugrische Sprachen (s.u.), entsprechende Untersuchungen durchgeführt. Ausgangspunkt dieser Arbeit war die Theorie von Wimmer u.a. (1994), die von der Überlegung ausgehen, daß die Häufigkeit der Wörter einer bestimmten Länge P_{x+1} abhängig sein sollte von der Häufigkeit der Wörter der vorhergehenden Länge P_x . Es ist klar: Wenn in einem Text eine sehr hohe Zahl einsilbiger Wörter vorkommt, so können bei gleicher Gesamtwortzahl des Textes die zweisilbigen nur weniger häufig vertreten sein als in einem Text mit deutlich weniger einsilbigen. Dies kann man nun für alle vorkommenden Wortlängensklassen entsprechend verallgemeinern. Wimmer u.a. (1994 : 101) kommen deshalb zu dem Ansatz: $P_x = g(x) P_{x-1}$. Je nachdem, welche genaueren Bedingungen für die Funktion $g(x)$ bestimmt werden, kommt man zu einer von mehreren möglichen Funktionen. Setzt man z.B. $g(x) = a/(c + x)$, so erhält man als Verteilungsmodell die Hyperpoisson-Verteilung (Wimmer, Köhler, Grotjahn, Altmann 1994 : 102), deren Formel wie folgt lautet:

$$P_x = \frac{a^{x-1}}{b^{(x-1)} {}_1F_1(1; b; a)}, \quad x = 1, 2, \dots$$

Setzt man für $g(x)$ andere Bedingungen an, kommt man zu weiteren Verteilungsmodellen. Schon Fucks (1955 : 205) hat die Poisson-Verteilung als geeignetes Mo-

dell für Wortlängenverteilungen vorgeschlagen, während Grotjahn (1982 : 55ff.) diese nur als einen Spezialfall betrachtet und stattdessen die zusammengesetzte Poisson-Verteilung, d.h. die negative Binomialverteilung, als das geeignetere Modell ansieht.

Diese drei Modelle und noch einige andere haben sich bei der Bearbeitung von Wortlängenhäufigkeiten in Texten immer wieder bewährt. Es wäre aber aus theoretischen Gründen wünschenswert, daß man bei der Bearbeitung von Texten in einer Sprache aus allgemeinen Erwägungen heraus bereits sagen könnte, welches dieser Modelle das geeignete sein müßte. Also etwa so: Untersucht man in Sprache x einen Text der Textsorte y des Autors z — und evt. noch weitere Bedingungen —, daß m u β eine ganz bestimmte Verteilung das geeignete Modell sein. Das Problem besteht nun darin, daß solche spezifischen Hypothesen (noch?) nicht aufgestellt werden können. Die Forschungsstrategie enthält deshalb hier ein induktives Moment: Man untersucht, wie oft Wörter einer bestimmten Länge in einem ausgewählten Text vorkommen, tut dies für alle Wortlängenklassen, und sucht dann mit Hilfe einer geeigneten Software (hier: Altmann-Fitter 1994) im Rahmen der theoretisch begründeten Verteilungsfunktionen nach dem Modell, das sich für den betreffenden Text am besten eignet. Für solche Untersuchungen ist es allerdings notwendig, eine größere Zahl von Texten zu berücksichtigen, da an einzelne Texte oft eine Vielzahl von Funktionen angepaßt werden kann. Will man also Erkenntnisse über die Sprache gewinnen, kommt man mit einem oder wenigen Texten nicht aus. Als unterste Grenze für eine sinnvolle Untersuchung kann man von ca. 10 Texten ausgehen.

2. Für diese Untersuchung wurden insgesamt 28 Texte (13 Gedichte und Poeme sowie 15 Prosatexte) bearbeitet; außerdem wurde ein von Veenker (1981 : 49) mitgeteilter erzamordwinischer Text (Text 29) berücksichtigt. Bei diesem Text handelt es sich um ein künstlich ausgestaltetes Tiermärchen, dessen russischer Grundtext von G. I. Jermuškin verfaßt wurde, der auch die erzamordwinische Version entwickelte. Diese Version ist "im Hinblick auf die Wortwahl ein wenig "manipuliert", so daß möglichst viele Wörter finnisch-ugrischen Ursprungs vorkommen" (Veenker 1981 : 36). In der gleichen Arbeit gibt Veenker (1981 : 60) die Daten des analogen mokschamordwinischen Textes wieder, die hier als Text 30 bearbeitet werden. Diese beiden von Veenker behandelten Texte sind nicht ganz ideal für eine Untersuchung wie die vorliegende, da es sich um Übersetzungen/Übertragungen mit zusätzlichen lexikalischen Anpassungen an die Zielsprache handelt und nicht um ursprünglich in erza- oder mokschamordwinischer Sprache verfaßte Texte.

Bei der Bearbeitung der einzelnen Texte wurden folgende Prinzipien gewahrt: Es wird immer nur der laufende Text ohne Überschrift und etwaige andere Textbestandteile ausgewertet. Für jeden Text wird ausgezählt, wieviele einsilbige, zweisilbige, etc. Wörter er enthält. "Wort" wird als das orthographische Wort bestimmt, das im Druckbild also von andern Wörtern durch Leerstellen und/oder Interpunktionszeichen getrennt wird. Trennungsstrich, Bindestrich und Apostroph werden nicht als Interpunktionszeichen gewertet; sie verbinden Wortteile zu einem Wort.

Als Anzeiger für die "Silbe" werden Vokale und — wo vorhanden — Diphthonge gewertet. Die Zahl der Silben eines Wortes entspricht im Erzamordwinischen der Zahl seiner Vokale. Man muß also nicht unbedingt exakt die Silbengrenzen bestimmen, um zu entscheiden, wieviele Silben ein Wort hat.

Nachdem die Zahl der ein- bis n-silbigen Wörter für jeden Text festgestellt ist, werden die Daten der Texte in einen Rechner eingelesen; anschließend wird mit der Software Altmann-Fitter (1994) geprüft, welche der infrage kommenden Verteilun-

gen an diese Textdateien am besten angepaßt werden kann. Da bereits bei mehreren finnisch-ugrischen Sprachen die o.a. Hyperpoisson-Verteilung erfolgreich angewendet wurde, lag es nahe, dies auch bei Erzamordwinisch zu versuchen. Die Ergebnisse dieser — erfolgreichen — Anpassung finden sich in den folgenden Tabellen.

3. Die Tabellen liefern folgende Informationen:

x bedeutet die Silbenzahl pro Wort;

n_x die Zahl der beobachteten Wörter der Länge x ;

NP_x die aufgrund der Hyperpoisson-Verteilung erwartbare, berechnete Anzahl der Wörter der Länge x ;

a, b sind die Parameter der Hyperpoisson-Verteilung;

χ^2 ist das Chiquadrat;

der Index r in χ_r^2 gibt die Zahl der Freiheitsgrade an;

P ist die Wahrscheinlichkeit dafür, daß der berechnete Wert des Chiquadrats unterschritten wird; P ist umso kleiner, je größer χ^2 ; dabei soll das Kriterium $P \geq 0.05$ erfüllt werden; noch akzeptabel ist $0.01 \leq P < 0.05$.

$C = N/\chi^2$ ist der Diskrepanzkoeffizient, der dann verwendet wird, um die Güte der Anpassung zu beurteilen, wenn P mangels Freiheitsgraden nicht bestimmt werden kann. Bedingung: $C \leq 0.01$; noch akzeptabel: $C \leq 0.02$. Die Anpassung einer Verteilung ist umso besser, je kleiner C . Der Diskrepanzkoeffizient C wird vernachlässigt, wenn $P \geq 0.05$ ist.

4. Die Ergebnisse

a) Gedichte und Poeme

x	Text 1		Text 2		Text 3	
	n_x	NP_x	n_x	NP_x	n_x	NP_x
1	56	55.03	35	35.13	34	34.20
2	97	99.86	83	83.32	93	93.56
3	44	40.31	38	36.23	40	38.93
4	9	10.80	7	8.67	9	8.76
5			2	1.65	1	1.55
	$a = 0.5192$		$a = 0.5325$		$a = 0.4907$	
	$b = 0.2861$		$b = 0.2245$		$b = 0.1794$	
	$\chi_1^2 = 0.735$		$\chi_2^2 = 0.493$		$\chi_2^2 = 0.223$	
	$P = 0.39$		$P = 0.78$		$P = 0.89$	
	$C = 0.004$		$C = 0.003$		$C = 0.001$	

Text 1: Моро: Седейшкава (Herzlich), S. 7f.

Text 2: ders.: Мон кунсолинъ Фидельнъ (Ich hörte Fidel), S. 34—36.

Text 3: ders.: Саранск (Saransk), S. 66f.

x	Text 4		Text 5		Text 6	
	n_x	NP_x	n_x	NP_x	n_x	NP_x
1	36	35.73	30	29.71	31	31.12
2	85	84.36	76	77.03	98	95.03
3	17	18.57	24	22.40	31	35.64
4	2	2.14	3	3.86	10	8.21
5	1	0.20				

Text 4

a = 0.2428
b = 0.1028
 $\chi_1^2 = 0.338$
P = 0.56
C = 0.002

Text 5

a = 0.3276
b = 0.1264
 $\chi_1^2 = 0.316$
P = 0.57
C = 0.002

Text 6

a = 0.4276
b = 0.1400
 $\chi_1^2 = 1.093$
P = 0.30
C = 0.006

Text 4: ders.: Мазый течись (Das schöne Heute), S. 74f

Text 5: ders.: Одс вандолди (Aufs neue strahlend), S. 76f.

Text 6: ders.: Парксо (Im Park), S. 91f.

Text 7

x	n_x	NP_x
1	28	27.92
2	68	67.80
3	28	28.52
4	8	7.76
5		
6		

a = 0.5088
b = 0.2095
 $\chi_1^2 = 0.019$
P = 0.89
C = 0.0001

Text 8

x	NP_x
18	17.81
65	63.91
26	28.34
8	6.69
1	1.25

a = 0.5060
b = 0.1411
 $\chi_2^2 = 0.508$
P = 0.78
C = 0.004

Text 9

n_x	NP_x
189	194.89
315	324.82
221	173.27
34	55.02
3	12.43
1	2.57

a = 0.7846
b = 0.4707
C = 0.002

Text 7: ders.: Сюпав инже (Der reiche Gast), S. 96f.

Text 8: Надькин: Зярдс стакс (Wenn man es schwer hat), S. 33f.

Text 9: ders.: Тетян сёрма (Der Brief meines Vaters), S. 46—48.

Text 10

x	n_x	NP_x
1	68	68.10
2	155	155.24
3	55	54.69
4	11	10.44
5	1	1.53
6		

a = 0.4167
b = 0.1828
 $\chi_2^2 = 0.207$
P = 0.90
C = 0.0007

Text 11

x	NP_x
107	106.51
281	279.72
133	136.29
44	36.59
1	6.78
1	1.11

a = 0.5982
b = 0.2278
 $\chi_2^2 = 5.957$
P = 0.05
C = 0.0105

Text 12

n_x	NP_x
22	21.26
55	53.16
24	28.18
12	10.40

a = 0.6726
b = 0.2691
 $\chi_1^2 = 0.961$
P = 0.33
C = 0.009

Text 10: ders.: Певтеме рисьме (Die endlose Kette), S. 49—53.

Text 11: ders.: Кустемат (Treppen), S. 37—45.

Text 12: Гайни: Телень пакс (Winterliches Feld), S. 36—38.

Text 13

x	n_x	NP_x
1	74	73.41
2	150	148.81
3	56	59.36
4	16	13.13
5	1	2.29

a = 0.4967
b = 0.2450
 $\chi_2^2 = 1.541$
P = 0.46
C = 0.005

Text 13: Гайни:
Тонеть, грудница ава
(Dir, werktätige Mutter),
S. 39—41.

b) Prosa (Kurzgeschichten, Erzählungen, Novellen, Humoresken)

Text 14			Text 15		Text 16	
x	n_x	NP_x	n_x	NP_x	n_x	NP_x
1	57	58.60	38	37.31	60	58.48
2	115	118.23	117	114.89	129	125.75
3	73	64.45	39	42.95	61	65.02
4	21	20.31	9	8.54	16	19.10
5	1	5.41	2	1.31	6	3.92
					1	0.73
		$a = 0.7470$		$a = 0.4256$		$a = 0.6808$
		$b = 0.3702$		$b = 0.1382$		$b = 0.3167$
		$X_2^2 = 4.867$		$X_2^2 = 0.831$		$X_2^2 = 2.091$
		$P = 0.09$		$P = 0.66$		$P = 0.35$
		$C = 0.018$		$C = 0.004$		$C = 0.008$

Text 14: Седойкин: Кажто пуворькст (Zwei Schwielen), S. 222f.

Text 15: ders.: Сөксень вайгель (Stimme des Herbstes), S. 225.

Text 16: ders.: Оймсема (Ruhe), S. 226f.

Text 17			Text 18		Text 19	
x	n_x	NP_x	n_x	NP_x	n_x	NP_x
1	54	54.29	42	40.57	39	40.00
2	96	96.52	106	102.40	78	81.58
3	50	49.04	44	51.99	56	48.01
4	15	14.53	17	14.67	17	16.50
5	1	3.04	3	2.86	1	4.91
6	1	0.49	1	0.51		
7	0	0.06				
8	1	0.03				
		$a = 0.7114$		$a = 0.6356$		$a = 0.8272$
		$b = 0.4002$		$b = 0.2519$		$b = 0.4057$
		$X_2^2 = 0.140$		$X_2^2 = 1.898$		$X_2^2 = 4.626$
		$P = 0.93$		$P = 0.39$		$P = 0.10$
		$C = 0.0006$		$C = 0.009$		$C = 0.024$

Text 17: ders.: Эряви ломань (Man braucht einen Menschen), S. 228f.

Text 18: ders.: Аволь есень кель (Es war nicht meine Sprache), S. 229f.

Text 19: Буянов: Манявсь... (Er täuschte sich...), S. 32.

Text 20			Text 21		Text 22	
x	n_x	NP_x	n_x	NP_x	n_x	NP_x
1	67	67.53	75	73.01	137	140.32
2	139	133.92	147	152.12	214	219.19
3	69	76.60	96	90.76	109	103.78
4	28	25.60	34	31.59	33	28.96
5	7	6.04	3	7.76	3	5.72
6	1	1.31	1	1.47	1	0.87
7			1	0.29	1	0.11
8					1	0.05

Text 20

a = 0.8038
b = 0.4053
 $\chi_3^2 = 1.393$
P = 0.71
C = 0.005

Text 21

a = 0.8359
b = 0.4012
 $\chi_3^2 = 3.677$
P = 0.30
C = 0.01

Text 22

a = 0.6794
b = 0.4350
 $\chi_2^2 = 1.107$
P = 0.58
C = 0.002

Text 20: ders.: Таркат... (Dein Platz...), S. 32f.

Text 21: ders.: Данилка (Danilka), S. 33f.

Text 22: ders.: Паня баба (Die alte Panja), S. 34f.

Text 23

x	n_x	NP_x
1	51	49.87
2	138	134.94
3	50	56.87
4	18	15.32
5		
6		
7		
8		

a = 0.4993
b = 0.1845
 $\chi_1^2 = 1.401$
P = 0.24
C = 0.006

Text 24

x	n_x	NP_x
1	76	75.60
2	204	202.93
3	113	116.25
4	42	37.27
5	5	8.29
6	2	1.66
7		
8		

a = 0.7283
b = 0.2713
 $\chi_3^2 = 2.090$
P = 0.55
C = 0.005

Text 25

x	n_x	NP_x
1	50	50.71
2	126	127.80
3	72	68.17
4	20	20.33
5	2	4.20
6	1	0.66
7	0	0.08
8	1	0.05

a = 0.6766
b = 0.2685
 $\chi_2^2 = 0.446$
P = 0.80
C = 0.002

Text 23: ders.: Алонь порксамо — кудов (Die Eier schälen nach Hause), S. 35f.

Text 24: Рузавина: Кежтньень таркас — превть-тевть (Statt Zorn vernünftige Taten), S. 47f.

Text 25: dies.: Куяр пире (Der Gurkengarten), S. 48f.

Text 26

x	n_x	NP_x
1	91	90.17
2	242	255.16
3	134	125.55
4	39	33.83
5	3	6.27
6	3	1.02
7		
8		

a = 0.5956
b = 0.2105
 $\chi_2^2 = 2.268$
P = 0.32
C = 0.004

Text 27

x	n_x	NP_x
1	99	98.91
2	238	237.78
3	125	125.79
4	38	37.38
5	8	7.72
6	0	1.22
7	0	0.15
8	1	0.05

a = 0.6783
b = 0.2821
 $\chi_3^2 = 0.139$
P = 0.99
C = 0.0003

Text 28

x	n_x	NP_x
1	89	88.18
2	202	200.15
3	98	100.64
4	27	28.45
5	5	5.59
6	3	0.99
7		
8		

a = 0.6459
b = 0.2846
 $\chi_2^2 = 0.487$
P = 0.78
C = 0.001

Text 26: Втулкин: Аванть питне. Юмореска (Der Preis der Frau. Humoreske), S. 90f.

Text 27: Новичков: Сёрма. Новелла (Ein Brief. Novelle), S. 91f.

Text 28: Арапов: Манчигай. Юмореска (Ein Betrüger. Humoreske), S. 92f.

Text 29			Text 30	
x	n_x	NP_x	n_x	NP_x
1	122	129.49	150	146.72
2	187	198.49	182	191.40
3	122	91.00	101	91.43
4	16	24.53	27	26.73
5	2	5.49	3	6.72
	a	= 0.6542	a	= 0.7538
	b	= 0.4268	b	= 0.5778
	c	= 0.009	χ_2^2	= 3.582
			P	= 0.17
			C	= 0.008

Text 29: Veenker 1981, S. 49

Text 30: Veenker 1981, S. 60

5. Die Untersuchung erzamordwinischer Texte hat damit folgende Ergebnisse erbracht: An alle 29 Texte kann die Hyperpoisson-Verteilung angepaßt werden; dasselbe gilt für den einen mokschamordwinischen Text. Nur bei einem Gedicht (Text 9) und dem nicht ganz unproblematischen Übersetzungstext von Jermuškin (Text 29) muß auf den Diskrepanzkoeffizienten als Kriterium zurückgegriffen werden, der in beiden Fällen gute Übereinstimmung des Modells mit den Daten signalisiert.

Das Ergebnis zum Mokschamordwinischen darf natürlich nicht überbewertet werden; um feststellen zu können, daß die Hyperpoisson-Verteilung ein gutes Modell auch für diese Sprache ist, muß ein umfangreicheres Korpus ausgewertet werden. Die Auswertung der 30 Texte unterstützt die Annahme, daß auch Wortlängenhäufigkeiten stochastischen Gesetzen unterliegen.

6. Abschließend sei ein Blick auf andere finnisch-ugrische Sprachen geworfen. Ebenso wie im Erzamordwinischen stellt die Hyperpoisson-Verteilung auch im Estnischen (Bartens, Best 1996) und Lappischen (Bartens, Best 1997) ein geeignetes Modell für die Wortlängenverteilung in Texten dar. Damit ist die Hyperpoisson-Verteilung ein wesentliches Modell der Wortlängen in finnisch-ugrischen Sprachen. Allerdings konnte sie bei finnischen (Vettermann, Best 1996) und ungarischen (Bartens, Zöbelin 1996) Texten nicht angepaßt werden; bei finnischen Gedichten bewährte sich die Dacey-Poisson-Verteilung, bei Prosatexten die Hyperpascal-Verteilung. Im Ungarischen ließ sich an alle Texte die positive Poisson-Verteilung anpassen.

Dies zeigt: Die verwendbaren Modelle können über die Grenzen der Einzelsprachen hinweg genutzt werden wie im Erzamordwinischen, Estnischen und Lappischen; eine Sprache der gleichen Sprachfamilie kann ein anderes Modell bevorzugen wie Finnisch und Ungarisch; und in einer Sprache können verschiedene Verteilungen je nach Textgruppe wie im Finnischen Verwendung finden.

Zu einer besseren Einschätzung der finnisch-ugrischen Sprachen kann man nur kommen, wenn 1. weitere Sprachen untersucht werden und dabei 2. innerhalb der Sprachen Textgruppen bearbeitet werden, die nach Zeit, Stil und Autor unterschiedlich sind; womöglich müssen auch noch weitere Faktoren in Betracht gezogen werden, die bisher noch nicht als wichtig für die Längenverteilungen erkannt wurden.

Quellen

Арапов — В. Арапов, Манчигай. Юмореска. — Сятко 1990, № 6, Саранск; Буянов — И. Буянов в Поваделень эрямосто. Нурька ёвтнемат. — Сятко 1990, № 6, Саранск; Втулкин — М. Втулкин, Аваньть питне. — Сятко 1990, № 6, Саранск; Гайни —

П. Г а й н и, Нудей, Саранск 1990; М о р о — А. М о р о, Кенярдома, Саранск 1964; **Надькин** — Д. Н а д ь к и н, Кустемат, Саранск 1977; **Новичков** — Н. Н о в и ч к о в, Сёрма. Новелла. — Сятко 1990, № 6, Саранск; **Рузавина** — О. Р у з а в и н а, Эрямонь йант. Ёвтнемат. — Сятко 1990, № 6, Саранск; **Седойкин** — Л. С е д о й к и н, Нурька ёвтнемат. — Откстомозь мастор, Саранск 1987.

Software

Altmann-Fitter. 1994. Lüdenschaid.

L I T E R A T U R

- B a r t e n s, H.-H., B e s t, K.-H. 1996, Wortlängen in estnischen Texten. — UAJb., Neue Folge 14 (erscheint).
 — 1997, Word Length Distribution in Sámi Texts. — Linguistic Structures (erscheint).
 B a r t e n s, H.-H., Z ö b e l i n, T. 1996, Wortlängenhäufigkeiten im Ungarischen. — Glottometrika 16 (erscheint).
 B e s t, K.-H., A l t m a n n, G. 1996, Project Report. — Journal of Quantitative Linguistics 3 (erscheint).
 B r a u n, P. 1993, Tendenzen in der deutschen Gegenwartssprache, Stuttgart.
 C r y s t a l, D. 1993, Die Cambridge Enzyklopädie der Sprache, Frankfurt—New York.
 F u c k s, W. 1995, Theorie der Wortbildung. — Mathematisch-Physikalische Semesterberichte 4, 195—212.
 — 1968, Nach allen Regeln der Kunst, Stuttgart.
 G r o t j a h n, R. 1982, Ein statistisches Modell für die Verteilung der Wortlänge. — Zeitschrift für Sprachwissenschaft 1, 44—75.
 K ö h l e r, R. 1986, Zur linguistischen Synergetik. Struktur und Dynamik der Lexik, Bochum.
 K ö n i g, W. 1978, dtv-Atlas zur deutschen Sprache. Tafeln und Texte, München.
 M e i e r, H. 1967, Deutsche Sprachstatistik, Hildesheim.
 T h o r n d i k e, E. L. 1938, Studies in the Psychology of Language, New York (Archives of Psychology 231).
 V e e n k e r, W. 1981, Zur phonologischen Statistik der mordwinischen Schriftsprachen. — UAJb., Neue Folge 1, 33—72.
 V e t t e r m a n n, A., B e s t, K.-H. 1996, Wortlängen im Finnischen (Manuskript).
 W i m m e r, G., K ö h l e r, R., G r o t j a h n, R., A l t m a n n, G. 1994, Towards a Theory of Word Length Distribution. — Journal of Quantitative Linguistics 1, 98—106.
 Z i p f, G. K. 1932, The Principle of Relative Frequency in Respect to Chinese Phonology. — G. K. Z i p f, Selected Studies of the Principle of Relative Frequency in Language, Cambridge, 1—8.

ХАНС-ХЕРМАНН БАРТЕНС, КАРЛ-ХЕЙНЦ БЕСТ (Гёттинген)

РАСПРЕДЕЛЕНИЕ СЛОВ ПО ДЛИНЕ В ЭРЗЯ-МОРДОВСКИХ ТЕКСТАХ

Исследование осуществлено в рамках международного научного проекта, цель которого — выявить соответствующие стохастические закономерности в разных языках. Уже получены результаты исследований более чем по 30 языкам, в том числе по большинству финно-угорских. Как в отношении других языков, так и в случае с эрзянским можно констатировать, что в основе распределения встречаемости слов по их длине лежит определенный закон. Результат данной работы получен в ходе анализа 28 текстов — стихотворений и поэм, а также беллетристики (короткая проза). В дальнейшем он должен быть подтвержден анализом текстов иных типов.