

**Peter Grzybek**  
**(Graz / Salzburg)**

**Zur Wortlänge und ihrer Häufigkeitsverteilung  
in Sprichwörtern**  
**(Am Beispiel slowenischer Sprichwörter,  
mit einer Re-Analyse estnischer Sprichwörter)**

**Erscheint in:**

**Christine Palm-Meister (Hrsg.):**

***Euophras 2000***

**Tübingen: Narr, 2001.**

## **Peter Grzybek (Graz/Salzburg)**

### **Zur Wortlänge und ihrer Häufigkeitsverteilung in Sprichwörtern**

**(Am Beispiel slowenischer Sprichwörter, mit einer Re-Analyse estnischer Sprichwörter)**

#### ***0. Vorbemerkung***

Die sprachliche Organisation von Sprichwörtern ist, bei nüchterner Betrachtung, kaum jemals ernsthaft im Hinblick auf Gesetzmäßigkeiten untersucht worden. Dieses Urteil mag auf den ersten Blick vielleicht als zu pauschal und provokant erscheinen; es lässt sich jedoch damit begründen, dass die Suche nach sprachlichen Regularitäten von Sprichwörtern im Grunde genommen kaum jemals über eine symptomatische Betrachtungsebene hinaus gekommen ist und eine systematische Ebene bislang eigentlich nicht erreicht hat (Grzybek 2000a, 2001a).

Nachdem an anderer Stelle (Grzybek 2000a,b) bereits systematisch die Frage der Satzlänge von Sprichwörtern und damit verbundene methodologische Probleme behandelt wurden, soll es im folgenden um einen anderen Aspekt sprachlicher Gesetzmäßigkeiten in Sprichwörtern gehen, nämlich um die Frage der Wortlänge in Sprichwörtern. Genauer gesagt, soll danach gefragt werden, ob sich ein theoretisches Beschreibungsmodell dafür finden lässt, wie viele ein-, zwei-, drei-, usw. -silbige Wörter in einem Sprichwortkorpus vorkommen. Ähnlich wie im Falle der Analysen von Satzlängen geht diese Frage somit deutlich über die "einfache" Berechnung von Durchschnittslängen hinaus. Vielmehr liegen der folgenden Untersuchung eine Reihe von Grundannahmen zugrunde, wie dies in analoger Weise auch bei der Untersuchung von Satzlängen und deren Häufigkeitsverteilung der Fall ist (Grzybek 2000b). Im Detail handelt es sich um die Annahmen,

1. dass die Häufigkeit, mit der Wörter einer bestimmten Länge in einem Sprichwortkorpus enthalten sind, nicht zufällig (chaotisch) ist, sondern einer bestimmten Regel- oder Gesetzmäßigkeit folgt;
2. dass sich diese Regelmäßigkeit nicht nur beschreiben, sondern auch formalisieren lässt;
3. dass sich diese Regelmäßigkeit so formalisieren lässt, dass sich Querverbindungen zu allgemeinen (nicht nur auf Sprichwörter bezogenen) Untersuchungen und theoretischen Annahmen zur Wortlängenverteilung herstellen lassen;
4. dass diese Querverbindungen Aussagen über die Spezifik von Sprichwörtern und ihre sprachliche Konstituenten erlauben.

Innerhalb der allgemeinen (quantitativen) Linguistik hat es in den vergangenen Jahre eine Reihe von Vorschlägen gegeben, diese Gesetzmäßigkeiten zu erfassen und zu beschreiben (Wimmer/Altmann 1996, Wimmer et al. (1994).

In der gesamten Geschichte der Sprichwortforschung hingegen hat es auf internationaler Ebene lediglich eine einzige Untersuchung gegeben, die jemals einer ähnlichen Frage nachgegangen ist, und zwar die Untersuchung von Krikmann (1967) zu estnischen Sprichwörtern. Auch wenn es in der vorliegenden Untersuchung um slowenische Sprichwörter gehen soll, scheint es sinnvoll, in aller gebotenen Kürze die Ergebnisse der Krikmann-Studie zunächst würdigend in Erinnerung zu bringen, zu kommentieren sowie dann aus heutiger Sicht neu zu bewerten, zumal sich aus einer entsprechenden Darstellung die methodologischen Voraussetzungen für unsere Fragestellung ergeben.

### ***1. Zur Wortlänge estnischer Sprichwörter***

Krikmann (1967) beschäftigte sich in seiner Studie u.a. mit der Häufigkeitsverteilung der in Silben gerechneten Wortlänge estnischer Sprichwörter. Untersucht wurden von ihm insgesamt 3.576 Sprichwörter auf verschiedene Aspekte der Satz- und Wortlänge. Für die Sprichwörter im Intervall von 3-10 Wörtern pro Satz – es handelt sich hierbei um insgesamt 3.362 Sprichwörter (d.h. 94.02% des gesamten Korpus) – führte Krikmann die Silbenanzahl pro Wort tabellarisch an; die Werte der übrigen 214 Sprichwörter (5.98%) hat er nicht angeführt. Tab. 1 (s.u.) enthält in der zweiten Spalte die empirisch beobachteten Häufigkeiten ( $F[i]$ ) der einzelnen Silbenlängen ( $X[i]$ ).

Krikmann versuchte bereits 1967, für diese beobachteten Häufigkeiten ein theoretisches Modell der Häufigkeitsverteilung zu finden. Dieses Vorgehen wird üblicherweise als "Anpassung" eines theoretischen Modells (hier: der Häufigkeitsverteilung) an die beobachteten Daten (hier: die Häufigkeiten der einzelnen Wortlängen) bezeichnet. Darunter ist eine solche mathematische Beschreibung zu verstehen, die es erlaubt, die tatsächliche Vorkommenshäufigkeit der einzelnen Wortlängen mit einem theoretisch berechenbaren Verteilungsmodell so zu vergleichen, dass anschließend mit Hilfe eines statistischen Tests (des sog. Chi-Quadrat-Anpassungstests) die Güte der Anpassung des theoretischen Modells an die beobachteten Ergebnisse geprüft werden kann. Solche Verteilungsmodelle (von denen es ca. 750-1000 verschiedene, wenn auch in bestimmten "Familien" zusammenfassbare gibt) entsprechen bestimmten Formeln mit (variierenden) Parametern; es kommt also darauf an, für jedes Modell zunächst die Parameter so zu bestimmen, dass die sich aufgrund der Formel ergebenden theoretischen Werte möglichst gut mit den beobachteten übereinstimmen, um dann zu sehen, welches der Modelle insgesamt der beobachteten Häufigkeitsverteilung am besten entspricht.

Ohne Frage hat Krikmann mit seinem Vorgehen Pionierarbeit geleistet, die seitens der Parömiologie kaum jemals gewürdigt worden ist, und auf die in der Folge praktisch nie Bezug genommen wurde. Allerdings hat er in seiner Studie zwei wichtige Schritte unterlassen:

- (a) er hat die Anpassungsgüte seines theoretischen Modells keinem statistischen Test unterzogen,
- (b) er hat nicht die Qualität verschiedener Verteilungsmodelle überprüft.

Vielmehr hat Krikmann ausschließlich die sog. Fucks-Verteilung an die oben beschriebenen Daten der estnischen Sprichwörter anzupassen versucht. Bei dieser Verteilung handelt es sich um eine um 1 verschobene Poisson-Verteilung, so dass sich die Formel als Modifikation der Poisson-Verteilung (1) wie in (2) darstellt.

$$(1) \quad P_x = \frac{e^{-a} a^x}{x!} \quad x = 0, 1, 2, \dots \quad a \geq 0$$

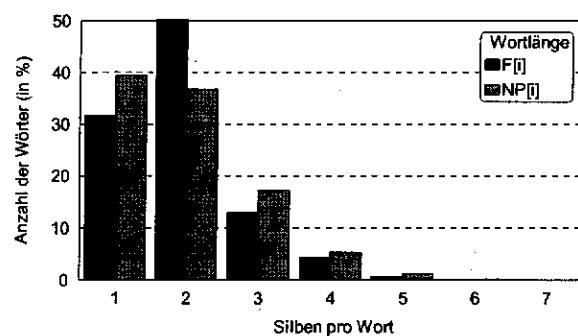
$$(2) \quad P_x = \frac{e^{-a} (a)^{x-1}}{(x-1)!} \quad x = 1, 2, \dots \quad a \geq 0$$

Die Verschiebung um 1 ist deswegen geboten, weil es im Estnischen (wie in vielen anderen Sprachen) keine Wörter mit 0 Silben gibt. Die Ergebnisse der theoretischen Berechnungen auf der Basis der Fucks-Verteilung ist der dritten Spalte der Tab. 1 zu entnehmen ( $NP[i]$ ) und in Abb. 1 graphisch dargestellt:

**Tab. 1: Wortlängenhäufigkeiten nach Krikmann (1967)**

X[i]	F[i]	NP[i]	F [i] (%)	NP [i] (%)
1	6648	8264,96	31,60	39,29
2	10573	7722,00	50,26	36,70
3	2730	3607,35	12,98	17,15
4	920	1123,46	4,37	5,34
5	149	262,41	0,71	1,25
6	16	49,03	0,08	0,23
7	2	8,79	0,01	0,04
	<i>a</i>	0,9343	→	1695,69
			<i>FG</i>	5
			<i>P</i>	< 0,0001
			<i>C</i>	0,0806

**Abb. 1: Graphische Darstellung der Wortlängenhäufigkeiten nach Krikmann (1967)**



Wie zu sehen ist, erweist sich die Anpassung als nicht gerade erfolgreich; dies bestätigt auch eine Nachberechnung des Kontingenzkoeffizienten, der mit  $C = 0.08$  den Unterschied zwischen den beobachteten und theoretischen Verteilungshäufigkeiten als signifikant ausweist.<sup>1</sup>

Im hier gegebenen Zusammenhang kommt es nicht darauf an, detaillierter auf das estnische Material einzugehen; dieser Frage soll statt dessen eine eigene Untersuchung auf erweiterter Textbasis gewidmet werden (Grzybek/Krikmann 2001). Es sei dennoch darauf hingewiesen, dass sich zur Anpassung an die Daten von Krikmann (1967) zwar die Fucks-Verteilung nicht besonders gut eignet, dass sich aber hervorragend ein anderes Verteilungsmodell anpassen lässt, nämlich die sog. **Hyperpascal-Verteilung**.<sup>2</sup>

Was bedeutet es nun, bzw. wie lässt es sich erklären, wenn die Häufigkeitsverteilung der Wortlängen dem Modell der Hyperpascal-Verteilung folgt? Fügen wir an dieser Stelle einige grundlegende Bemerkungen ein.

Grundannahme der Modellierung der Häufigkeitsverteilung der Wortlängen ist, dass die Verteilung der Vorkommenshäufigkeiten der jeweiligen Längenklassen – sei es in Texten, in Wörterbüchern oder, wie im gegebenen Fall, in einem Sprichwortkorpus – nicht chaotisch organisiert ist, sondern dass die einzelnen Längenklassen in einem Proportionalitätsverhältnis zueinander stehen. Dieses Verhältnis, das sich in allgemeiner Form als  $P_x \sim P_{x-1}$  verstehen lässt, besagt, dass der Umfang einer jeweils "höheren" Klasse in Abhängigkeit von der ihr vorausgehenden "niedrigeren" Klasse bestimmt wird. Das Proportionalitätsverhältnis lässt sich mit der Funktion  $g(x)$  beschreiben, so dass sich der folgende allgemeine Grundansatz ergibt:

$$(3) \quad P_x = g(x) P_{x-1}.$$

Je nach Art der Funktion ergeben sich so unterschiedliche Verteilungsmodelle (vgl. Wimmer et al. 1994, Wimmer/Altmann 1996). Setzt man in (3) beispielsweise für  $g(x) = (a + bx)/(c + dx)$ , so ergibt sich aufgrund der Differenzgleichung

$$(4) \quad P_x = \frac{a + bx}{c + dx} P_{x-1}$$

---

<sup>1</sup> Üblicherweise wird die Güte der Anpassung mit dem Chi-Quadrat-Anpassungstest berechnet; da der Chi-Quadrat-Test jedoch enorm anfällig für die Stichprobengröße ist, hat es sich in der quantitativen Linguistik durchgesetzt, statt dessen den Kontingenzkoeffizienten  $C$  als Prüfgröße heranzuziehen, der sich als  $X^2/N$  definiert; bei einem Kontingenzkoeffizienten von  $C < 0.02$  bzw.  $C < 0.01$  spricht man von einer guten bzw. sehr guten Anpassung (vgl. Grotjahn/Altmann 1993).

<sup>2</sup> Diese Beobachtung ist insofern bemerkenswert, als die Hyperpascal-Verteilung ansonsten als allgemeines Modell für die Häufigkeitsverteilung von Satzlängen postuliert wird (Altmann 1989).

nach Reparametrisierung die genannte Hyperpascal-Verteilung:

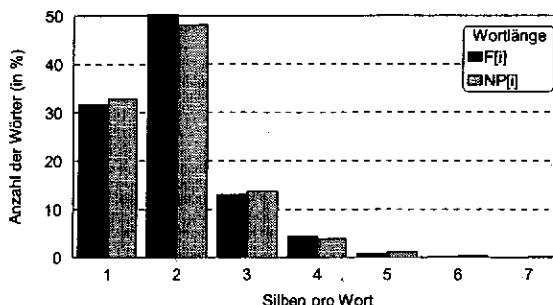
$$(5) P_x = \frac{\binom{k+x-1}{x}}{\binom{m+x-1}{x}} q^x P_0 \quad x = 0, 1, 2, \dots$$

Die Ergebnisse der Anpassung dieser Verteilung an die Daten der estnischen Sprichwörter (vgl. Tab. 1) finden sich in Tab. 2; deutlich zu sehen ist die Güte der Anpassung bei einem Kontingenzkoeffizienten von  $C < 0.01$ ; Abb. 2 enthält eine graphische Darstellung der Resultate:

Tab. 2: Re-Analyse der Wortlängen-häufigkeiten nach Krikmann (1967)

X[i]	F[i]	NP[i]	F [i] (%)	NP [i] (%)
1	6648	6916,16	31,60	32,87
2	10573	10122,59	50,26	48,12
3	2730	2896,25	12,98	13,77
4	920	801,71	4,37	3,81
5	149	219,41	0,71	1,04
6	16	59,70	0,08	0,28
7	2	22,19	0,01	0,11
<i>n</i>	0,0872	$\rightarrow$	130,39	
<i>m</i>	0,0159	<i>FG</i>	3	
<i>q</i>	0,2674	<i>P</i>	< 0,0001	
		<i>C</i>	0,0062	

Abb. 2: Graphische Darstellung der Re-Analyse von Wortlängen-häufigkeiten nach Krikmann (1967)



Dieser Umstand ist in mehrerer Hinsicht von Bedeutung.

1. Von eher mathematisch-theoretischem Interesse ist die Beobachtung, dass sich sowohl die Hyperpascal-Verteilung als auch die Poisson-Verteilung aus einer gemeinsamen Basisfunktion für die Verteilung von Wortlängen-häufigkeiten ableiten lassen. Setzt man diese Basisfunktion mit Wimmer/Altmann (1996) als  $g(x) = a/x^m$  bzw.  $g(x) = a \cdot x^{-m}$  an, so erhält man unter der Voraussetzung, dass in dieser Formel  $m = 1$ , die "normale" Poisson-Verteilung bzw. für den Fall, dass  $P_0 = 0$ , die in (2) beschriebene, von Krikmann verwendete, positive Poisson-Verteilung (bzw. Fucks-Verteilung). Und ebenfalls unter der Voraussetzung  $m = 1$  führt die Erweiterung  $g(x) = (a + bx)/(c + dx)$  zu der in (4) bzw. (5) dargestellten Hyperpascal-Verteilung.
2. Von besonderem Interesse für die Linguistik und für die Parömiologie hingegen ist eine andere Beobachtung, nämlich dass es gerade die Hyperpascal-Verteilung ist, die sich zur Anpassung an die sprichwörtlichen Daten eignet. Wie nämlich Bartens/Best (1996) argumentieren und zeigen, eignet sich zur Modellierung der Wortlängen-häufigkeit in estnischen Fließtexten eine ganz andere als die Hyperpascal-Verteilung am besten, nämlich die sog. Hyperpoisson-Verteilung.<sup>3</sup> Sollten sich also diese Ergebnisse auch an einem größeren Korpus estnischer Sprichwörter bestätigen, würde das bedeuten, dass sich die Wortlängen-häufigkeit in Sprichwörtern und Fließtexten voneinander unterscheidet; überraschender-

<sup>3</sup> Auch diese lässt sich übrigens als Modifikation der Basisfunktion nach Altmann/Wimmer (1996) verstehen, insofern die Funktion  $g(x) = a/(b + x)$  unmittelbar zu ihr führt.

weise scheint dabei das Modell für die Wortlängenhäufigkeit in Sprichwörtern "komplexer" zu sein als das für die Wortlängenhäufigkeit in Fließtexten, insofern hier mehr Parameter als bei Fließtexten ins Spiel kommen. Dieser Frage wird jedoch an anderer Stelle noch einmal ausführlicher nachzugehen sein.

Im hier gegebenen Zusammenhang wäre vielmehr als ein wichtiges Ergebnis zunächst einmal festzuhalten, dass die Häufigkeitsverteilung von Wortlängen in den estnischen Sprichwörtern nicht zufällig bzw. chaotisch ist, sondern bestimmten Gesetzmäßigkeiten folgt, die sich formalisieren lassen. Natürlich ist zu vermuten, dass auch sprachspezifische Faktoren auf die Modellierung der Wortlängenhäufigkeit Einfluss haben. Insofern sollen im folgenden die Untersuchungen auch an einem Korpus slowenischer Sprichwörter durchgeführt werden, an denen bereits Untersuchungen zur Verteilung von Satzlängenhäufigkeiten durchgeführt wurden (Grzybek 2000b).

## **2. Zur Wortlänge slowenischer Sprichwörter**

Es handelt sich um die Sammlung *Pregovori, prilike in reki* von Kocbek (1887), die eigentlich erste umfassende, eigenständige Sprichwortsammlung des Slowenischen, die nicht nur allgemein als ein Meilenstein, sondern als Fundament der slowenischen Parömiographie schlechthin anzusehen ist. Natürlich beruht auch diese Kompilation auf vorbereitenden Arbeiten und kleineren zuvor an unterschiedlichen Stellen veröffentlichten Sammlungen; doch nimmt das Werk von Kocbek ungeachtet dessen einen besonderen Stellenwert ein, weil es – zumindest zu einem großen Teil – auch die Grundlage für spätere Sammlungen darstellt, und zwar nicht nur für die Erweiterung von Kocbek/Šašelj (1934), sondern auch für solche Sammlungen wie diejenigen von Bojc (1974 u.a.) oder Prek (1970 u.a.).<sup>4</sup>

Die Sammlung von Kocbek weist insgesamt 2.429 sprichwörtliche Sätze auf. In diesem Korpus beträgt die in Worten gemessene Satzlänge im Durchschnitt  $\bar{x} = 6.37$  Wörter bei einer Standardabweichung von  $s = 2.48$ ; die in Silben berechnete Durchschnittslänge der 2.429 Sätze beträgt  $\bar{x} = 11.68$  bei einer Standardabweichung von  $s = 4.31$ . Diese beiden Ergebnisse zusammengenommen decken sich mit der Tatsache, daß das Sprichwortkorpus einen Umfang von 15.467 Wörtern bzw. von 28.363 Silben aufweist. Dies wiederum entspricht einer durch-

---

<sup>4</sup> Die mehrfachen Auflagen der Sammlung von Bojc und Prek sprechen für ihre Popularität; doch gilt es zu beachten, dass in den späteren Auflagen zum Teil in erhöhtem Maße nicht nur volkstümliche Sprichwörter, sondern auch Geflügelte Worte und Aphorismen (durchaus nicht nur slowenischer Herkunft) verstärkt Niederschlag finden.



schnittlichen in Silben gemessenen Wortlänge von  $\bar{x} = 1.83$  Silben bei einer Standardabweichung von  $s = 0.84$ .<sup>5</sup> Die absolute Häufigkeit ( $F[i]$ ) bzw. die relative Häufigkeit ( $F[i]\%$ ) der  $x$ -silbigen Wörter ist der zweiten bzw. vierten Spalte der Tabelle 3 zu entnehmen.

Als nächstes gilt es nun also herauszufinden, ob sich auch für diese Häufigkeitsverteilung ein geeignetes theoretisches Modell finden lässt. Diese Frage lässt sich im Endergebnis positiv beantworten, insofern sich herausstellt, dass sich die sog. **Hyperbinomial-Verteilung** als ein bestens geeignetes Modell für die Verteilung der Wortlängenhäufigkeiten erweist ( $C < 0.002$ ). Diese leitet sich aus der Differenzgleichung

$$(6) \quad g(x) = \frac{a - bx}{c + dx}$$

ab, was nach Reparametrisierung in der folgenden Wahrscheinlichkeitsfunktion resultiert:

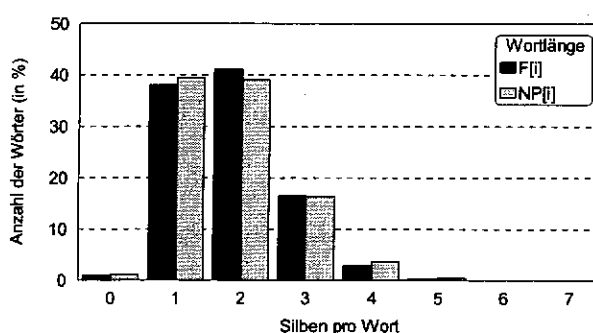
$$(7) \quad P_x = \frac{\binom{n}{x}}{\binom{m+x-1}{x}} q^x P_0, \quad x = 0, 1, \dots, n$$

In den Spalten 3 und 5 der Tab. 3 finden sich die Werte der theoretischen Anpassung an die beobachteten Häufigkeiten; Abb. 3 stellt die Ergebnisse in graphischer Form dar.

**Tab. 3: Wortlängenhäufigkeiten slowenischer Sprichwörter**

X[i]	F[i]	NP[i]	F[i] (%)	NP[i] (%)
0	163	169,15	1,05	1,09
1	5880	6102,01	38,02	39,45
2	6363	6029,95	41,14	38,99
3	2556	2523,12	16,53	16,31
4	442	566,13	2,86	3,66
5	58	71,65	0,37	0,46
6	4	4,84	0,03	0,03
7	1	0,14	0,01	0,00
$n$	7			56,94
$m$	0,033		FG	3
$q$	0,17		P	< 0,0001
			C	0,0037

**Abb. 3: Graphische Darstellung der Wortlängenhäufigkeiten slowenischer Sprichwörter**



Auch dieser Befund ist äußerst bemerkenswert, und zwar in dreierlei Hinsicht:

<sup>5</sup> Die in Grzybek (2000b) angegebenen Werte sind teilweise von den hier angeführten geringfügig verschieden; dies hängt damit zusammen, dass in der früheren Publikation die einsilbigen Wörter mit silbenbildendem Liquid (z.B. 'smrt') nicht richtig berechnet wurden.

- (1) zum einen kann nämlich damit auch an slowenischem Sprichwortmaterial die gesetzmäßige Organisation der Häufigkeitsverteilung von Wortlängen als nachgewiesen gelten;
- (2) zum anderen handelt sich bei den slowenischen Sprichwörtern um ein anderes Verteilungsmodell als bei den estnischen Sprichwörtern;
- (3) offenbar eignet sich für die Verteilung der Wortlängenhäufigkeiten auch bei den slowenischen Sprichwörtern ein anderes Modell besser als für die Verteilung der Wortlängenhäufigkeiten (a) in slowenischen Wortlisten<sup>6</sup> und (b) in slowenischen Fließtexten<sup>7</sup>.

### ***3. Ergebnisse und Perspektiven***

Mit den vorangegangenen Überlegungen und Analysen lassen sich eine Reihe von Ergebnissen festhalten, die zu Hypothesen für in der Zukunft durchzuführende Folgestudien führen:

1. die Häufigkeit, mit der Wörter einer bestimmten Länge in einem Sprichwortkorpus enthalten sind, ist nicht zufällig (chaotisch), sondern folgt einer bestimmten Regel bzw. Gesetzmäßigkeit;
2. diese Regelmäßigkeit lässt sich nicht nur beschreiben, sondern auch formalisieren;
3. diese Regelmäßigkeit lässt sich so formalisieren, dass sich Querverbindungen zu allgemeinen (nicht nur auf Sprichwörter bezogenen) Untersuchungen und theoretischen Annahmen zur Wortlängenverteilung herstellen lassen;
4. die sich in einem Sprichwortkorpus manifestierende Häufigkeitsverteilung von Wortlängen ist weder mit derjenigen von Wortlisten noch mit derjenigen von Fließtexten identisch;
5. diese Querverbindungen erlauben Aussagen über die Spezifik von Sprichwörtern und ihre sprachlichen Konstituenten;

---

<sup>6</sup> Wie Grzybek (2000c) gezeigt hat, erweist sich bei slowenischen Wortlisten die sog. Conway-Maxwell-Poisson-Verteilung als geeignet, die Häufigkeitsverteilung von Wortlängen zu modellieren.

<sup>7</sup> Für Fließtexte liegen im Hinblick auf das Slowenische bislang noch nicht in ausreichendem Maße Daten vor – die bisherigen Ergebnisse weisen aber darauf hin, dass die erweiterte positive Binomialverteilung ein geeignetes Modell für die Modellierung von Wortlängenhäufigkeiten slowenischer Fließtexte darzustellen scheint (Grzybek 2001b).

6. die Wortlängenverteilung in Sprichwörtern folgt keinem einheitlichen (sprachübergreifend-parömiologischen) Gesetz, sondern unterliegt (auch) dem Einfluss anderer, gegebenenfalls auch sprachspezifischer, Faktoren.

#### 4. Literatur

- Altmann, Gabriel (1988), "Verteilung der Satztlängen", in: Schultz, K.-P. (Hg.), *Glottometrika* 9, Bochum, 147-169.
- Bartens, Hans-Hermann; Best, Karl-Heinz (1996), "Wortlängen in estnischen Texten", *Uraltaische Jahrbücher*, N.F. 14, 112-128.
- Bojc, Etbin (1974), *Pregovori in reki na Slovenskem*. Ljubljana [<sup>2</sup>1980, <sup>3</sup>1987].
- Grotjahn, Rüdiger; Altmann, Gabriel (1993), "Modelling the Distribution of Word Length. Some Methodological Problems", in: Köhler, Reinhard; Rieger, Burkhard (Hg.), *Contributions to Quantitative Linguistics*, Dordrecht, NL, 141-153.
- Grzybek, Peter (2000a), "Zum Status der Untersuchung von Satztlängen in der Sprichwortforschung – Methodologische Vor-Bemerkungen", in: *Слово во времени и пространстве. К 60-летию профессора В.М. Мокшенико*, Sankt Petersburg, 430-457.
- Grzybek, Peter (2000b), "Wie lang sind slowenische Sprichwörter? Zur Häufigkeitsverteilung von (in Worten berechneten) Satztlängen slowenischer Sprichwörter", in: *Anzeiger für slavische Philologie*, 27 (1999) [2000], 87-108.
- Grzybek, Peter (2000c), "Pogostnostna analiza besed iz elektronskega korpusa slovenskih besedil" [= Lexikalische Frequenzanalyse aus dem elektronischen Korpus slowenischer Wörter], *Slavistična Revija*, 48(2); 141-157.
- Grzybek, Peter (2001a), "Zur Satz- und Teilsatzlänge zweigliedriger Sprichwörter", in: Köhler, Reinhard; Uhliřová, Lidia (Hg.), *Text as a Linguistic Paradigm: Levels, Constituents, Constructs*, Trier. [Im Druck]
- Grzybek, Peter (2001b), "Zur Wortlängenverteilung in slowenischen Fließtexten." [In Vorb.]
- Grzybek, Peter; Krikmann, Arvo (2001): "On Word Length Distribution in Estonian Proverbs." [In Vorb.]
- Kocbek, Fran (1887), *Pregovori, prilike in reki*. Celje.
- Kocbek, Fran; Šašelj, Ivan (1934), *Slovenski pregovori, reki in prilike*. Ljubljana.
- Krikmann, Arvo (1967), "Keelestatikat Eesti vanasõnadest" [= Sprachstatistik estnischer Sprichwörter], *Emakeele Seltsi Aastaraamat*, 13; 127-154.

Prek, Stanko (1972), *Ljudska modrost. Pregovori, domislice in reki*. Maribor. [Ljubljana, <sup>2</sup>1974, <sup>3</sup>1982, <sup>4</sup>1986, <sup>5</sup>1996]

Wimmer, Gejza; Grotjahn, Rüdiger; Altmann, Gabriel (1994), "Towards a Theory of Word Length Distribution", *Journal of Quantitative Linguistics*, 1; 98-106.

Wimmer, Gejza; Altmann, Gabriel (1996), "The Theory of Word Length: Some Results and Generalizations", in: Schmidt, Peter (Hg.), *Issues in General Linguistic Theory and the Theory of Word Length* [= *Glottometrika*; 15], Trier, 112-133.