PETER GRZYBEK (GRAZ)

# Quantitative Aspekte slawischer Texte (am Beispiel von Puškins *Evgenij Onegin*) [1]

Ohne Frage ist Puškins *Evgenij Onegin* (*EO*) eines der am meisten und wohl auch am besten untersuchten Werke der russischen Literatur – und zwar in verschiedenster Hinsicht, angefangen von konzeptuellen Fragen der Thematik über strukturelle Fragen der Komposition bis hin zu Spezifika von Metrik, Strophik, Reimschema u. v. a. m. Dabei sind – vor dem Hintergrund nicht nur russischer, sondern auch gesamteuropäischer Traditionen – sowohl intertextuelle als auch extratextuelle Bezüge Gegenstand vielfältigster Forschungen gewesen; auch Fragen der sprachlichen Gestaltung wie z. B. der lexikalischen oder syntaktischen Struktur haben die Aufmerksamkeit der Forschung auf sich gelenkt.

Dennoch sind ungeachtet dieses großen Interesses und trotz der mannigfaltigsten Herangehensweisen einige wichtige Aspekte der sprachlichen Gestaltung des *EO* bislang nicht in ausreichendem Maße von der Forschung berücksichtigt worden: die Frage nämlich, wie sich die poetische Spezifik des Textes zu allgemeinen sprachlichen Regularitäten verhält bzw. inwiefern allgemeine Gesetzmäßigkeiten der Textkonstruktion bei aller dichterischen Individualität und innovatorischen Genialität auch die Dynamik dieses Textes regulieren.

Der hier ins Spiel kommende Begriff der dynamischen (Selbst-)Regulation nimmt dabei Bezug auf neuere systemtheoretische Konzeptionen, die sich vom ‚klassischen' Strukturalismus der Saussureschen Tradition wesentlich unterscheiden. Denn dem von de Saussure ausgehenden strukturalistischen Paradigma zufolge, das Sprache (und in der Folge dann auch ein jegliches nichtsprachliches System) als ein aus Elementen, Relationen und Strukturen bestehendes geschlossenes System [„un système serré"] verstand, war Sprache als ein abstraktes Netz von Relationen (Hjelmslev) zu begreifen.

Innerhalb dieses Paradigmas stellte sich ein System somit als ein mehr oder weniger stabiler Gleichgewichtszustand zwischen den Elementen und Relationen auf verschiedenen Ebenen dar. Zwar argumentierten Forscher wie S. I.

[1] Die vorliegende Untersuchung entstand im Zusammenhang mit dem Grazer FWF-Projekt # 15485 »Wortlängen(häufigkeiten) in Texten slawischer Sprachen«; vgl.: http://www-gewi.uni-graz.at/quanta

Karcevskij oder R. O. Jakobson früh (vor allem in Anbetracht von Fragen der Sprachentwicklung und der dichterischen Sprache) schon recht für dynamische(re) Konzeptionen; doch eigentlich hat sich erst in jüngerer Zeit, vor allem auch unter dem Einfluss chaostheoretischer Ansätze, die Ansicht durchgesetzt, dass wir es bei jeglicher Art von System (also auch in der Sprache allgemein bzw. in sprachlichen und nicht-sprachlichen Texten im weitesten, semiotischen Sinne) nicht mit stabilen Gleichgewichtszuständen zu tun haben, sondern mit einer mehr oder weniger stabilen Abfolge von Zuständen (einem „Fließ-Gleichgewicht"). Diese dynamische Sichtweise führte auch zu einem veränderten Verständnis von *Regelabweichungen*: Während diese im Rahmen einer Konzeption von Sprache als einem stabilen System als Verstoß gegen die Norm und infolge dessen als System*störung* ausgelegt werden mussten, verstand man sie im Rahmen von Konzeptionen der Sprache als einem dynamischen, evoluierenden System in gewissem Sinne als systemimmanent und erachtete sie als notwendig für die ständige Entwicklung. Vor allem auch vor dem Hintergrund chaostheoretischer Ansätze verschob sich durch den Rückgriff auf das Konzept der Selbst-Regulation der Fokus stärker auf sog. Attraktoren, postulierte (End)-Zustände, auf die hin evoluierende Systeme zusteuern und die letztendlich für die (Wieder)-Herstellung des Fließ-Gleichgewichts eines (sich ständig in Veränderung befindlichen) Systems verantwortlich sind. Vor diesem allgemein-theoretischen Hintergrund präsentieren sich natürlich Fragen der Selbst-Organisation nicht nur in der Natur, sondern auch in der Literatur und Kultur auf ganz neue Art und Weise (vgl. HAYLES 1990, HAYLES ed. 1991, WERNER 1999). Genau dies soll in der vorliegenden Studie an einem abgegrenzten Bereich diskutiert werden.

Bei diesem abgegrenzten Bereich soll es sich um nicht mehr und nicht weniger als um Fragen von Vers- und Wortlänge im *EO* handeln. Während die Wortlänge durchgehend in der Anzahl der Silben pro Wort berechnet wird, muss die Bestimmung der Verslänge alternativ in der Anzahl der Silben pro Vers und in der Anzahl der Wörter pro Vers geleistet werden.

## VERSLÄNGE

Die in der Anzahl der S i l b e n p r o V e r s berechnete Verslänge wird natürlich stark durch die spezifische Strophik des *EO* bestimmt, die in engem Zusammenhang mit der Metrik steht und insofern nicht ohne Auswirkung auf die angesprochene Frage bleibt: Denn die Oneginsche Strophe mit ihren 14 vierfüßig-jambischen Verszeilen weist aufgrund ihrer Reimstruktur *AbAbCCdd EffEff* in jeder Strophe acht 8-silbige Verszeilen mit männlicher Klausel (also: ⏑–⏑–⏑–⏑–) und sechs 9-silbige Verszeilen mit weiblicher Klausel (also: ⏑–⏑–⏑–⏑–⏑) auf. Insgesamt sollten sich also, wenn das Schema ausnahmslos realisiert ist, in jeder Strophe 118 Silben auf jeweils 14 Verszeilen verteilen. Im „Regelfall" würde dies also einer mittleren Verslänge von $\bar{x} = 8{,}43$ Silben pro Vers bei einer Standardabweichung von $s = 0{,}50$ entsprechen.

Allerdings weisen von den acht Kapiteln des *EO* nur die Kapitel I, II, V, VI, und VII die Oneginsche Strophik voll ausgebildet auf:

1. im Kapitel III sind in der 3. Strophe nur die ersten acht Zeilen realisiert, die ausgelassenen sechs Verse sind durch Punkte markiert; außerdem weist das „Песня девушек“ insgesamt 18 Verszeilen mit jeweils 7 Silben auf; auch die 79 Verse von Tatjanas Brief an Evgenij entsprechen nicht der Oneginschen Strophe;
2. im Kapitel IV lautet der letzte Vers der 37. Strophe „И одевался …“, so dass sich hier ein 5-silbiger Vers ergibt;
3. im Kapitel VIII weist die 2. Strophe nur die ersten vier, die 25. Strophe nur die ersten acht Zeilen auf, und auch der Brief von Evgenij an Tat'jana entspricht mit seinen jeweils 30 acht- und neunsilbigen Versen nicht der Oneginschen Strophe.

Aufgrund dieser Umstände ergeben sich in den einzelnen Kapiteln geringfügige Abweichungen. Diese sind vom Umfang her insgesamt jedoch so gering, dass sie den o. a. theoretischen Durchschnittswert von $\bar{x} = 8{,}43$ nicht beeinträchtigen ($s = 0{,}50$).

Interessanter als die Frage nach der in der Anzahl der Silben pro Vers berechneten Verslänge (die ja unmittelbar von der syllabotonischen Metrik abhängt) ist jedoch die Berechnung der Verslänge in der Anzahl der W ö r t e r p r o V e r s , da es hier ja einen deutlich größeren gestalterischen Spielraum gibt. Allerdings kommen hierbei Fragen der Wortdefinition zum Tragen, die sich im Hinblick auf die Berechnungen insofern auswirken, als man bei einer graphematischen Wortdefinition im Russischen (wie in anderen slawischen Sprachen auch) von Wörtern auszugehen hat, welche kein silbenbildendes Element (im Russischen also kein Element mit Vokal-Qualität) aufweisen. Diese Wörter – in den meisten dieser Fälle handelt es sich hierbei um Präpositionen – werden zum Teil bei entsprechenden Berechnungen ignoriert, zum Teil aber auch einer eigenständigen Wortklasse von sog. 0-silbigen Wörtern zugeordnet, so dass sich die Gesamtanzahl der Wörter im Text (und entsprechend die mittlere Anzahl der Wörter pro Vers) unter dieser Bedingung erhöht.

Um die Auswirkung einer solchen Entscheidung unter Kontrolle zu halten, scheint es sinnvoll, die Berechnungen sowohl unter Berücksichtigung als auch unter Nicht-Berücksichtigung der sog. 0-silbigen Wörter (vgl. ANTIĆ – KELIH 2003) als einer eigenständigen Wort(längen)klasse anzustellen. Im Ergebnis zeigt sich, dass bei Berücksichtigung der 0-silbigen Wörter als einer eigenen Wort(längen)klasse die Verslänge in *EO* im Durchschnitt $\bar{x} = 4{,}29$ Wörter pro Vers bei einer Standardabweichung von $s = 1{,}06$ beträgt; werden die 0-silbigen Wörter nicht als eigenständige Wort(längen)klasse gerechnet, ist die mittlere Verslänge niedriger und beträgt $\bar{x} = 4{,}08$ ($s = 0{,}96$). Wie nicht anders zu erwarten, ist der Unterschied der mittleren Verslänge in Abhängigkeit

von der Art der Berücksichtigung der 0-silbigen Wörter als eigener Wortklasse hoch signifikant ($z = 10{,}62$; $p < 0{,}001$).[2]

Noch interessanter als die Frage nach der durchschnittlichen Anzahl der Wörter pro Vers im gesamten Text ist jedoch die Frage, ob es einen Unterschied in der Wortanzahl pro Vers zwischen den 8- und 9-silbigen Versen gibt.

Tab. 1 enthält die entsprechenden Werte, und zwar für beide Arten der Berücksichtigung 0-silbiger Wörter. Aufgrund der geringen Anzahl der 5- und 7-silbigen Verse (s. o.) beschränkt sich die Darstellung auf die Angaben für die 8- und 9-silbigen Verse.

T a b . 1 : Durchschnittliche Wortzahl pro Vers in *EO*

| | | **Silben pro Vers** | | |
|---|---|---|---|---|
| | | **8-silbige** | **9-silbige** | **gesamt** |
| **mit 0-** | $\bar{x}$ | 4,26 | 4,34 | 4,29 |
| **Silbern** | $s$ | 1,04 | 1,07 | 1,06 |
| **ohne** | $\bar{x}$ | 4,06 | 4,12 | 4,08 |
| **0-Silber** | $s$ | 0,95 | 0,96 | 0,96 |

Wie zu sehen ist, beträgt die durchschnittliche Verslänge in den 8-silbigen Versen $\bar{x} = 4{,}26$ ($s = 1{,}04$), in den 9-silbigen Versen $\bar{x} = 4{,}34$ ($s = 1{,}07$) Wörter pro Vers. Werden die 0-silbigen Wörter nicht als eigenständige Wortklasse gerechnet, ist die durchschnittliche Verslänge natürlich niedriger: Unter dieser Voraussetzung beträgt sie in den 8-silbigen Versen $\bar{x} = 4{,}06$ ($s = 0{,}95$), in den 9-silbigen Versen $\bar{x} = 4{,}12$ ($s = 0{,}96$). Der Befund, dass in den 8-silbigen Versen im Durchschnitt weniger Wörter vorkommen als in den 9-silbigen Versen, entspricht durchaus der Erwartung; der Unterschied der durchschnittlichen Wortanzahl pro Vers ist zwar insgesamt gering, dennoch aber unabhängig von der Behandlung der 0-silbigen Wörter auf dem 5%-Niveau signifikant (jeweils $p < 0{,}05$). Diese Beobachtung führt direkt zur Frage der Wortlänge, die es ebenfalls nicht nur im Hinblick auf den Gesamttext, sondern auch spezifisch in Abhängigkeit von der Verslänge zu untersuchen gilt, und zwar unter getrennter Behandlung der 0-silbigen Wörter.

[2] Zur Überprüfung von Mittelwertunterschieden bei kleineren Stichproben wird üblicherweise der sog. *t*-Test für unabhängige Stichproben verwendet. Da die dem *t*-Test zugrundeliegende Student-Verteilung gegen die Normalverteilung konvergiert und bei ca. $N > 100$ hinreichend genau die Normalverteilung approximiert, führt in unseren Analysen ein *t*-Test zu demselben Ergebnis wie ein Normalverteilungstest; aus diesem Grunde wird im vorliegenden Aufsatz nur auf den *z*-Wert der Normalverteilungstests verwiesen, was die Angabe von Freiheitsgraden erspart.

Zunächst aber soll die Untersuchung der Verslänge noch von einer anderen Richtung her angegangen werden: Wenn in der bisherigen Darstellung bei der Frage nach der Verslängenhäufigkeit die in der durchschnittlichen Anzahl der Wörter pro Vers gemessene Verslänge fokussiert wurde, so gilt es dabei zu berücksichtigen, dass ein und derselbe Mittelwert aufgrund von sehr unterschiedlichen Häufigkeitsverteilungen zustande kommen kann (wobei das Maß der Abweichungen von diesem Mittelwert durch den Wert der Standardabweichung angegeben wird). Insofern lässt sich die Frage nach der Verslänge auch anders stellen, indem man nämlich nicht nach dem Mittelwert als einem Maß der zentralen Tendenz fragt, sondern das Profil der gesamten Häufigkeitsverteilung betrachtet. Die Frage lautet dann: Wie viele Verse mit zwei, drei, vier usw. Wörtern gibt es eigentlich, und lässt sich die konkrete Häufigkeitsverteilung theoretisch modellieren oder gar im Sinne einer bestimmten Regularität verstehen?

Natürlich unterliegt auch eine Antwort auf diese Frage allgemeinen Prinzipien der Wortdefinition, insofern sich die Häufigkeit der Wort(längen)klassen je nach Berücksichtigung der 0-silbigen Wörter verändert. Aus diesem Grunde enthält die Tab. 2 neben den jeweiligen (in der Anzahl der Wörter pro Vers berechneten) Verslängen in der zweiten bzw. fünften Spalte die entsprechenden Häufigkeitswerte *f[i]* für den gesamten Text, und zwar für beide Varianten der Behandlung der 0-silbigen Wörter als einer eigenständigen Wortklasse; die Spalten 3 und 4 bzw. 5 und 6 enthalten die entsprechenden Werte jeweils für die 8- bzw. 9-silbigen Verse.

T a b . 2 : Verslängenhäufigkeiten in *EO*

| | **mit 0-silbigen Wörtern** | | | **ohne 0-silbige Wörter** | | |
|---|---|---|---|---|---|---|
| **Wörter pro** | | **Verslänge** | | | **Verslänge** | |
| **Vers** | **gesamt** | **8 Silben** | **9 Silben** | **gesamt** | **8 Silben** | **9 Silben** |
| **2** | 42 | 27 | 82 | 50 | 31 | 95 |
| **3** | 670 | 473 | 1149 | 813 | 599 | 1417 |
| **4** | 1160 | 849 | 2009 | 1238 | 887 | 2125 |
| **5** | 736 | 582 | 1318 | 654 | 546 | 1200 |
| **6** | 284 | 244 | 528 | 182 | 163 | 345 |
| **7** | 63 | 66 | 129 | 26 | 20 | 46 |
| **8** | 10 | 5 | 15 | 2 | 0 | 2 |

Auf der Basis dieser Häufigkeitswerte stellt sich als nächstes die Frage, ob die Häufigkeitsverteilungen einer bestimmten Regularität folgen, und – wenn ja – ob und wie diese sich theoretisch modellieren lässt. Als Basis für einschlägige Überlegungen können die Arbeiten von WIMMER ET AL. (1994)

bzw. WIMMER – ALTMANN (1996) dienen, die zunächst im Hinblick auf eine Theorie der Wortlängenhäufigkeit vorgelegt wurden, die sich jedoch ohne weiteres auf andere sprachliche Ebenen übertragen lassen. Eine Grundannahme dieser Arbeiten lautet, dass die Verteilung der Vorkommenshäufigkeiten der jeweiligen Längenklassen nicht chaotisch organisiert ist, sondern dass die einzelnen Klassen in einem Proportionalitätsverhältnis zueinander stehen. Übertragen auf die Verslänge lautet die Frage also, ob die Häufigkeit, mit der 2-Wort-Verse vorkommen, in einer spezifischen Beziehung steht zu der Häufigkeit, mit der 3-Wort-Verse vorkommen, und ob diese ihrerseits in einer spezifischen Beziehung zur Häufigkeit von 4-Wort-Versen steht, usw. Mathematisch umformuliert: Lässt sich die Häufigkeit einer gegebenen Klasse zur Häufigkeit der jeweiligen Nachbarklasse in Form einer Relation $P_x \sim P_{x-1}$ (d. h. als dynamisches System) verstehen?

Das zugrundeliegende Proportionalitätsverhältnis ist dabei nicht als konstant aufzufassen, sondern lässt sich mit der Funktion *g(x)* beschreiben, so dass sich der folgende allgemeine Grundansatz ergibt:

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

Je nach Art der Funktion ergeben sich so unterschiedliche Verteilungsmodelle. Als Basisform setzen WIMMER – ALTMANN (1994: 115) für $g(x) = a \cdot x^{-b}$ an; diese führt aufgrund der Differenzengleichung (1) zu

(2) $$P_x = \frac{a^x}{(x!)^b} \cdot P_0 \qquad x = 0,1,2,\ldots \;\; a, b > 0, \;\; P_0^{-1} = \sum_{i=0}^{\infty} \frac{a^i}{(i!)^b}$$

d. h. zur sog. Conway-Maxwell-Poisson-Verteilung. Interessanterweise lassen sich die in Tab. 2 dargestellten Häufigkeiten alle mit eben diesem Verteilungsmodell theoretisch beschreiben. Mit anderen Worten: Die Häufigkeit, mit der Verse mit einer bestimmten Anzahl von Wörtern vorkommen, ist stochastisch reguliert. Da in unseren Daten die Verslängenklassen $x = 0$ und $x = 1$ beide nicht besetzt sind, lautet die Formel für die entsprechend 2-verschobene Verteilung – in Modifikation der obigen Formel – wie folgt:

(3) $$P_x = \frac{a^{x-2}}{(x-2!)^b} \cdot P_0, \qquad x = 2,3,\ldots$$

Bei der konkreten Anpassung dieser Verteilung an empirische Daten geht es nun darum, die Parameter (in unserem Fall: *a* und *b*) so zu bestimmen, dass die empirischen und theoretischen Werte möglichst gut übereinstimmen. Dazu gibt es entsprechende Spezialsoftware, die in iterativen Prozessen die Werte optimiert (ALTMANN – FITTER 1997). Für die in Tab. 2 dargestellten Verslängen-

häufigkeiten des *EO* wären demnach die in Tab. 3a/b aufgeführten Werte einzusetzen, aufgrund derer sich die in der dritten Spalte der Tab. 3a/b dargestellten theoretischen Werte *NP[i]* ergeben.

Die Güte der Anpassung wird üblicherweise durch den $\chi^2$-Wert des $\chi^2$-Anpassungstests bewertet; allerdings hat der $\chi^2$-Test die Eigenschaft, dass der $\chi^2$-Wert linear mit der Stichprobengröße zunimmt. Aus diesem Grunde verwendet man in diesem Fall auch den sog. Diskrepanzkoeffizienten *C*, der sich als $C = \chi^2 / N$ berechnet und der für $C < 0{,}02$ auf gute, für $C < 0{,}01$ auf sehr gute Anpassungsergebnisse hinweist. Da wir es bei unseren Analysen ausnahmslos mit großen Stichproben zu tun haben, werden wir die Anpassungsgüte im Folgenden ausschließlich mit Bezugnahme auf den Diskrepanzkoeffizienten *C* überprüfen.

Wie der Tab. 3a/b zu entnehmen ist, stellt die Conway-Maxwell-Poisson-Verteilung unter beiden Bedingungen ein ausgezeichnetes Modell dar[3], die Häufigkeit der in der Anzahl der Wörter pro Vers berechneten Verslängen theoretisch zu beschreiben, insofern $C < 0{,}01$.

T a b . 3 a / b : Verslängenhäufigkeiten in *EO* (empirisch und theoretisch)

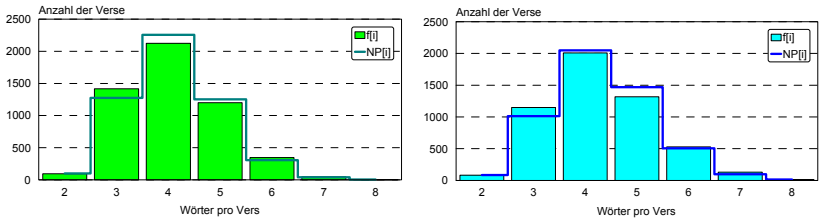
| x[i] | f[i] | NP[i] |
|---|---|---|
| **2** | 82 | 84,64 |
| **3** | 1149 | 1011,31 |
| **4** | 2009 | 2048,69 |
| **5** | 1318 | 1469,71 |
| **6** | 528 | 504,79 |
| **7** | 129 | 97,92 |
| **8** | 15 | 12,95 |
| | *a* | 11,95 |
| | *b* | 2,56 |
| | *C* | 0,0089 |

| x[i] | f[i] | NP[i] |
|---|---|---|
| **2** | 95 | 98,90 |
| **3** | 1417 | 1274,32 |
| **4** | 2125 | 2257,14 |
| **5** | 1200 | 1252,35 |
| **6** | 345 | 304,95 |
| **7** | 46 | 39,20 |
| **8** | 2 | 3,14 |
| | *a* | 12,88 |
| | *b* | 2,86 |
| | *C* | 0,0063 |

Abb. 1a/b stellt die beobachteten und theoretischen Häufigkeiten dar, Abb. 3a mit, Abb. 3b ohne Berücksichtigung 0-silbiger Wörter als eigenständiger Wort(längen)klasse.

[3] Es ist hier nicht der Ort, auf detailliertere Analysen einzugehen; dennoch gilt es festzuhalten, dass sich interessanterweise die Conway-Maxwell-Poisson-Verteilung auch dann als ein geeignetes Modell darstellt, wenn man die (in der Anzahl der Wörter pro Vers berechneten) Verslängenhäufigkeiten jeweils gesondert für die 8- und 9-silbigen Verse betrachtet; allerdings erweist sich hierbei ein anderes Verteilungsmodell, die sog. Hyperbinomial-Verteilung, unter allen Bedingungen als noch besser.

A b b . 3 a / b : Verslängenhäufigkeiten in *EO*




Indem wir somit festhalten können, dass die Häufigkeiten der (in der Wortanzahl pro Vers berechneten) Verslängen nicht chaotisch organisiert sind, sondern einer bestimmten Regularität folgen, soll es als nächstes um die Frage der Wortlänge gehen.

### WORTLÄNGE IN DER ANZAHL DER SILBEN PRO WORT BERECHNET

Die durchschnittliche Wortlänge berechnet sich im Prinzip als Quotient aus der Anzahl der Wörter und der Silben im Text; damit unterliegt auch die Berechnung der Wortlänge Fragen der Wort- bzw. Silbendefinition. Dies ist im Hinblick auf die mittlere Wortlänge insofern von Bedeutung, als sich unter Berücksichtigung der 0-Silber als einer eigenständigen Wort(längen)klasse die Gesamtzahl der Wörter erhöht (und folglich die durchschnittliche Wortlänge abnimmt), auch wenn der Umfang dieser Wortklasse relativ gering ist: im *EO* macht er insgesamt 4,81% aus.

Unter Berücksichtigung der 0-silbigen Wörter als eigener Wort(längen)-klasse beträgt die mittlere Wortlänge im *EO* $\bar{x}$ = 1,96 ($s$ = 1,06). Im Hinblick auf die syllabotonische Struktur der beiden Texte scheint jedoch die Frage nach der durchschnittlichen (in Silben berechneten) Länge auf der Basis nur der zur Silbenstruktur beitragenden Wörter (also ohne Berücksichtigung der 0-silbigen Wörter als eigenständige Wortlängenklasse) die der Fragestellung angemessenere Vorgangsweise zu sein. Unter dieser Voraussetzung beträgt im *EO* die mittlere Wortlänge $\bar{x}$ = 2,06 Silben ($s$ = 0,98).

Vor dem Hintergrund der obigen Überlegungen zur Verslänge liegt es jedoch nahe, die Wortlänge gesondert für die jeweiligen Verslängen zu berechnen. Anlass zu dieser Überprüfung geben zwei Annahmen:

1. Nach der oben getroffenen Feststellung zur unterschiedlichen Wortanzahl pro Vers in den verschiedenen Verslängen gilt es, die Wortlänge in den einzelnen Verslängen zu prüfen, und zwar in Abhängigkeit von der Berechnungsbasis der Verslänge (d. h. zum einen auf der Grundlage der Silbenanzahl pro Vers, zum anderen auf Basis der Wörter pro Vers).

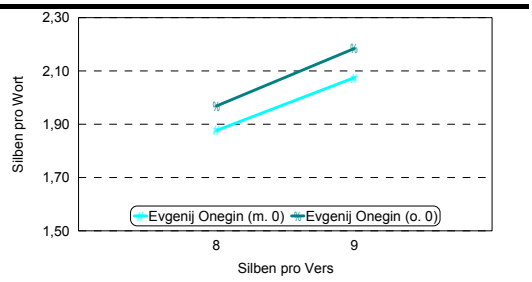
2. Aus der Quantitativen Linguistik sind allgemeine Annahmen über den Zusammenhang von Konstrukt und (unmittelbaren) Konstituenten, z. B. zwischen Satz- bzw. Teilsatzlänge und Wortlänge, bekannt. So besagt das sog. Menzerath-Altmann-Gesetz in allgemeiner Form: „Je größer bzw. komplexer ein sprachliches Konstrukt ist, um so kleiner bzw. einfacher sind seine Konstituenten“ (ALTMANN 1980). Diesen Annahmen zufolge wäre zu erwarten, dass die Wortlänge bei zunehmender Verslänge abnimmt, wenn der Vers als Konstrukt verstanden und in Relation zum Wort gesetzt wird. Dabei wäre allerdings zu testen, welche Einheiten (also etwa Silben pro Vers oder Wörter pro Vers) dem Konstrukt zugrunde zu legen sind. Hierzu fehlen bislang jegliche systematische Untersuchungen.[4]

WORTLÄNGE IN ABHÄNGIGKEIT VON DER VERSLÄNGE (IN SILBEN PRO VERS)

Tab. 4 enthält die in Abb. 2 anschaulich dargestellte durchschnittliche Wortlänge (mit dazugehöriger Standardabweichung) für beide Varianten der Berücksichtigung der 0-silbigen Wörter, und zwar sowohl für den gesamten Text (s. o.) als auch differenziert nach den jeweiligen Verslängen.

T a b . 4 / A b b . 2 : Wortlänge in *EO*

| | **m. 0** | **o. 0** |
|---|---|---|
| $\bar{x}$ | 1,96 | 2,06 |
| $s$ | 1,06 | 0,98 |
| $\bar{x}_8$ | 1,88 | 1,97 |
| $s_8$ | 1,00 | 0,94 |
| $\bar{x}_9$ | 2,07 | 2,18 |
| $s_9$ | 1,11 | 1,03 |




Es zeigt sich deutlich, dass die mittlere Wortlänge in den längeren Versen länger als in den kürzeren ist – ein Unterschied, der unabhängig von der Art der Behandlung der 0-silbigen Wörter hoch signifikant ist ($p < 0{,}0001$).

Im Hinblick auf die Erwartung abnehmender Wortlänge bei zunehmender Verslänge stellt sich also genau die entgegengesetzte Tendenz heraus. Es liegt nahe, dass die erwartete Tendenz deshalb nicht zum Ausdruck kommt, weil das Konstrukt ‚Vers‘ im Hinblick auf das Menzerath-Altmannsche Gesetz nicht zu

[4] Immerhin konnte an anderer Stelle eine solche Tendenz auch bereits an dem avantgardistischen Poem («Žuravl’») von V. Chlebnikov aufgezeigt werden; demnach ergab sich eine Verallgemeinerung folgender Art: »Je länger die Verse, desto kürzer die Wörter« (GRZYBEK 2001).

seinen unmittelbaren (!) Konstituenten in Beziehung gesetzt wird, so dass der obige Befund eher als Indiz für das sog. Arenssche Gesetz (ALTMANN 1983) zu werten wäre. Insofern stellt sich die Frage, ob sich die erwartete Tendenz unter der Bedingung bestätigt, dass das Konstrukt ‚Vers' nicht in der Anzahl der Silben, sondern in der Anzahl der Wörter pro Vers berechnet wird.

WORTLÄNGE IN ABHÄNGIGKEIT VON DER VERSLÄNGE (IN WÖRTERN PRO VERS)

Vor dem Hintergrund der obigen Befunde stellt sich als nächstes die Frage nach einem Zusammenhang zwischen der in der Anzahl der Wörter pro Vers berechneten Verslänge und der in der Anzahl der Silben pro Wort berechneten Wortlänge. Auch bei dieser Betrachtung wirkt sich natürlich die Behandlung der 0-silbigen Wörter wesentlich aus, weshalb es abermals nahe liegt, die Berechnungen doppelt anzustellen.

Die Tab. 5 enthält die durchschnittliche, in Silben gerechnete Wortlänge in Abhängigkeit von der jeweiligen in der Anzahl der Wörter gerechneten Verslänge.

T a b 5 : Wortlänge in *EO* Abhängigkeit von der Verslänge

| **Wörter pro Vers** | **Silben pro Wort** | |
|---|---|---|
| | m. 0-Silber | o. 0-Silber |
| 2 | 4,0732 | 4,0789 |
| 3 | 2,8021 | 2,8064 |
| 4 | 2,1056 | 2,1044 |
| 5 | 1,6883 | 1,6910 |
| 6 | 1,4104 | 1,4121 |
| 7 | 1,2159 | 1,2050 |
| 8 | 1,0417 | 1,0000 |

Deutlich erkennbar ist, dass bei zunehmender Wortanzahl pro Vers die durchschnittliche, in der Anzahl der Silben pro Wort berechnete Wortlänge abnimmt – unabhängig von der Art der Behandlung der 0-silbigen Wörter. Die Frage ist, ob und wie sich auch dieser Trend theoretisch modellieren lässt.

Wie oben bereits angesprochen, gibt es im Bereich der Quantitativen Linguistik allgemeine Grundsatzüberlegungen zur Lösung dieser Frage, die in dieser Form bislang kaum jemals auf Fragen der poetischen Textstruktur angewendet worden sind. In der Regel ist das Menzerath-Altmannsche Gesetz nur auf den Zusammenhang von Satz- und/oder Teilsatz- und Wortlänge bezogen worden, und es wäre im hier gegebenen Zusammenhang zu fragen, ob es sich auch im Hinblick auf den Zusammenhang von Vers- und Wortlänge als operational erweist.

Altmann hat eine mathematische Formalisierung dieser Tendenz in Form einer nicht-linearen Regressionsgleichung vorgeschlagen; in ihrer allgemeinsten Form (III) lautet diese

(4) $y = ax^b e^{-cx}$

mit den beiden Spezialfällen (I) $b \neq 0$, $c \neq 0$ und (II) für $b \neq 0$, $c = 0$.

T a b . 6 : Das Menzerathsche Gesetz in der Formalisierung von ALTMANN (1980)[5]

| | | | |
|---|---|---|---|
| I | $b = 0$ | | $y = Ae^{-cx}$ |
| II | $b \neq 0$ | $c = 0$ | $y = Ax^b$ |
| III | | $c \neq 0$ | $y = Ax^b e^{-cx}$ |

Ebenso wie im Fall der diskreten Häufigkeitsverteilungen geht es auch bei der nicht-lineraren Regression darum, in den gegebenen Formeln die Parameter so zu optimieren, dass die sich aufgrund der Anpassung ergebenden theoretischen Werte möglichst gut mit den beobachteten übereinstimmen. Die Güte der Anpassung muss natürlich getestet werden; sie wird durch den so genannten Determinationskoeffizienten $R^2$ überprüft, ein Maß für den Zusammenhang zwischen den beobachteten und den theoretischen Werten im Intervall zwischen 0 und 1. Dabei betrachtet man in der Regel solche Anpassungen als gelungen, bei denen die theoretische Schätzung auf eine möglichst geringe Anzahl von Parametern reduziert werden kann. In diesem Sinne wird Formel (II) üblicherweise als „Standardfall" für das Menzerath-Altmannsche Gesetz angesehen.

Für die in der Tab. 5 aufgeführten Daten erweist sich in der Tat der „einfache" Fall (II) als geeignet, die beschriebene Tendenz zu formalisieren; da sich die Werte je nach Art der Berücksichtigung der 0-Silber kaum unterscheiden, sind auch die Parameter $a$ und b für beide Bedingungen fast identisch ($a = 7.9989$, $b = 0.9666$). Es stellt sich bei einem $R^2$ von 0.9996 eine nahezu perfekte Anpassung an die beobachteten Daten heraus – mit anderen Worten: die durchschnittliche Wortlänge ist zu mehr als 99,9% durch die Verslänge determiniert! Abb. 3 veranschaulicht diesen Befund graphisch; aufgrund der

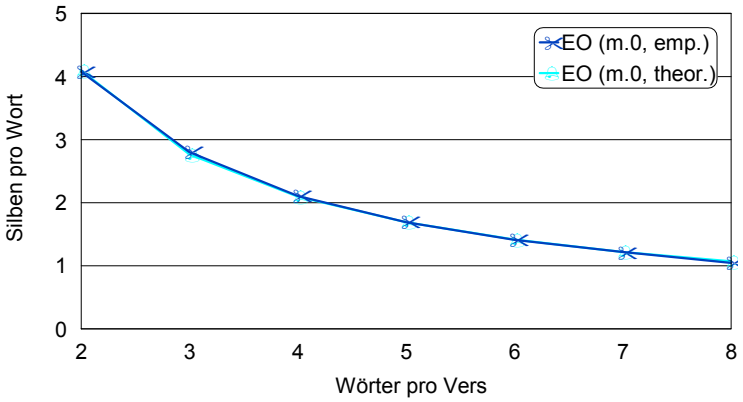
[5] In einem erweiterten Ansatz gehen WIMMER – ALTMANN (2002, 2003) von einer verallgemeinerten Formel aus:

$y = Ce^{a_0 x}\, x^{a_1} e^{-a_2/x - a_3/(2x^2) - a_4/(3x^3) - \ldots}$

Hier erhält man die oben erwähnte Form des Menzerath-Altmannschen Gesetzes für den Fall, dass $a_1 < 0$, $a_0 = a_2 = a_3 = \ldots 0$.

hochgradigen Ähnlichkeit des Kurvenverlaufs beider Arten der Berücksichtigung 0-silbiger bezieht sie sich lediglich auf die erste Bedingung.

A b b . 3 : Abhängigkeit von Wortlänge (in Silbenanzahl pro Wort) und Verslänge (in Wortanzahl pro Vers)




WORTLÄNGENHÄUFIGKEITEN

Aus dieser Beobachtung heraus stellt sich abschließend die nochmals weiterführende Frage nach der silbischen Struktur des lexikalischen Bestandes: Mit welcher Häufigkeit kommen Wörter einer bestimmten Länge im Text vor, und folgen (auch) diese Vorkommenshäufigkeiten ebenfalls einer bestimmten Regularität? Tab. 7 enthält die absoluten Werte ($f_i$) und die entsprechenden Prozentwerte, die sich unter Berücksichtigung der 0-silbigen Wörter als eigener Wort(längen)klasse ergeben.

T a b . 7 : Vorkommenshäufigkeiten der Wortlängen

| **Silben pro Wort** | ***Evgenij Onegin (m. 0)*** | |
|---|---|---|
| | *f(i)* | (in %) |
| **0** | 1079 | 4,81 |
| **1** | 7210 | 32,15 |
| **2** | 7599 | 33,88 |
| **3** | 4894 | 21,82 |
| **4** | 1314 | 5,86 |
| **5** | 277 | 1,24 |
| **6** | 51 | 0,23 |
| **7** | 4 | 0,02 |

| | | |
|---|---|---|
| **Σ** | 22428 | |

Als nächstes stellt sich nun wiederum die Frage, ob und wie sich die beiden Häufigkeitsverteilungen theoretisch beschreiben lassen. Wie oben bereits im Zusammenhang mit der Modellierung der Verslängenhäufigkeiten bemerkt wurde, lautet nach WIMMER ET AL. (1994) bzw. WIMMER – ALTMANN (1996) eine Grundannahme, dass die Verteilung der Vorkommenshäufigkeiten der jeweiligen Längenklassen nicht chaotisch organisiert ist, sondern dass die einzelnen Längenklassen in einem Proportionalitätsverhältnis zueinander stehen. Setzt man, ausgehend von dem oben dargestellten allgemeinen Grundansatz

(1) $P_x = g(x)P_{x-1}$,

für $g(x) = \frac{a-bx}{x}$, so ergibt sich aufgrund der Differenzengleichung

(5) $P_x = \frac{a-bx}{x} P_{x-1}$

nach entsprechender Reparametrisierung die Binomialverteilung, und zwar für $P_0 \neq 0$ deren Standardform (7), für $P_0 = 0$ die sog. positive Binomialverteilung (8):

(6) $P_x = \binom{n}{x} p^x q^{n-x} \qquad x = 0,1,...,n$

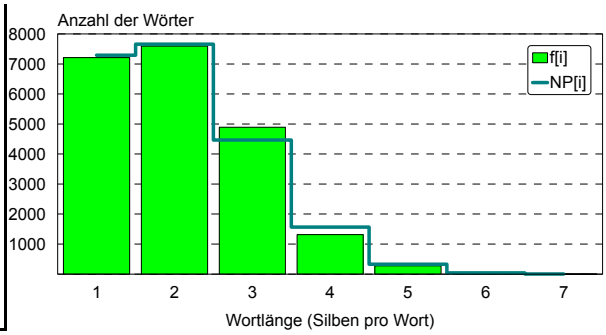
(7) $P_x = \frac{\binom{n}{x} p^x q^{n-x}}{1-q^n} \qquad x = 1,2,...,n$

Genau dieses Verteilungsmodell erweist sich im Hinblick auf die Wortlängenhäufigkeiten als hervorragend geeignet, wenn die 0-silbigen Wörter nicht als eigene Wort(längen)klasse berücksichtigt werden (C < 0,005); die in Abb. 4 veranschaulichten Werte sind der Tab. 8 zu entnehmen.

T a b . 8 / A b b . 4 : Wortlängenhäufigkeiten in *EO*

| x[i] | f[i] | NP[i] |
|---|---|---|
| **1** | 7210 | 7292,41 |
| **2** | 7599 | 7657,41 |
| **3** | 4894 | 4467,04 |
| **4** | 1314 | 1563,54 |
| **5** | 277 | 328,36 |
| **6** | 51 | 38,31 |
| **7** | 4 | 1,92 |

| | |
|---|---|
| $n$ | 7 |
| $p$ | 0,2593 |
| $C$ | 0,0045 |



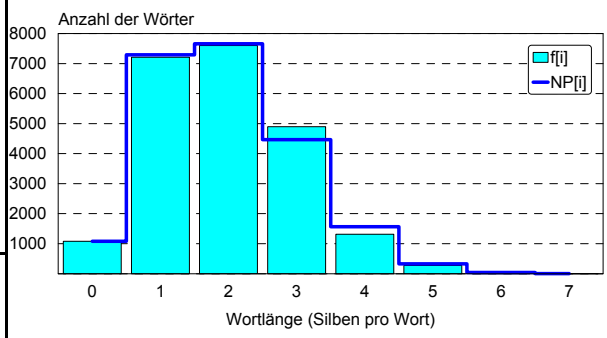

Interessanterweise stellt sich im Falle der Berücksichtigung der 0-silbigen Wörter als eigener Wort(längen)klasse eine Modifikation der (positiven) Binomialverteilung als am besten geeignetes Modell heraus, nämlich die sog. erweiterte positive Binomialverteilung. Wie aus Formel (9) ersichtlich ist, beinhaltet diese die Erweiterung der gestutzten Verteilung um die Klasse $x = 0$ und die anschließende Normierung aller Klassen:

$$(8)\qquad Px = \begin{cases} 1-\alpha & x = 0 \\ \dfrac{\alpha \binom{n}{x} p^x q^{n-x}}{1-q^n} & x = 1,2,\ldots,n \end{cases} \qquad \begin{array}{l} n \in \mathrm{N} \\ 0 \le p \le 1 \\ q = 1-p \\ 0 \le \alpha \le 1 \end{array}$$

Bei der Anpassung dieses Verteilungsmodells an die Daten des *EO* ergeben sich die in Tab. 9 aufgeführten und in Abb. 5 veranschaulichten Werte:

T a b . 9 / A b b . 8 : Beobachtete und theoretische Vorkommenshäufigkeiten der Wortlängen (mit 0-silbigen Wörtern als eigenständiger Wortlängenklasse)

| x[i] | f[i] | NP[i] |
|---|---|---|
| **0** | 1079 | 1079 |
| **1** | 7210 | 7292,70 |
| **2** | 7599 | 7657,43 |
| **3** | 4894 | 4466,89 |
| **4** | 1314 | 1563,43 |
| **5** | 277 | 328,32 |
| **6** | 51 | 38,31 |
| **7** | 4 | 1,92 |
| | $n$ | 7 |
| | $p$ | 0,2593 |
| | $\alpha$ | 0,9519 |
| | $C$ | 0,0045 |



## RESÜMEE UND AUSBLICK

Als ein Ergebnis der vorliegenden Untersuchungen lässt sich die streng gesetzmäßige Organisation des *EO* im Hinblick auf die folgenden sprachlichen Fragen festhalten: (a) Häufigkeit der Verslängen, (b) Häufigkeit der Wortlängen, (c) Abhängigkeit von Wort- und Verslänge.

Aufgrund des völligen Fehlens vergleichbarer Untersuchungen im Bereich der Analyse poetischer Texte führen die oben dargestellten Ergebnisse zu einer ganzen Reihe weiterführender Fragen, deren Lösung zu einer Einschätzung der obigen Befunde (vor allem auch im Hinblick auf eine Bewertung der Spezifik des Textes, der Poetik Puškins oder poetischer Texte allgemein) notwendig ist:

1. es gilt zunächst – vor allem in Anbetracht der mehrjährigen Entstehung des *EO* – die Homogenität des Gesamttextes bzw. dessen mögliche Heterogenität im Hinblick auf die einzelnen den Text konstituierenden Kapitel zu prüfen, d. h. der Frage nachzugehen, wie sich die beschriebenen Regularitäten in den einzelnen Kapiteln im Vergleich zum Gesamttext verhalten (GRZYBEK 2003b);
2. als nächstes gilt es, die sich hierbei einstellenden Befunde auf andere, vor allem auch in anderem Versmaß geschriebene Texte Puškins zu beziehen; einen ersten Ansatz kann hierzu ein Vergleich des *EO* mit dem 1830 entstandenen *Domik v Kolomne* bieten (GRZYBEK 2002);
3. eine weitere Relativierung der obigen Befunde wäre durch eine Analyse von Übersetzungen des *EO* in andere Sprachen zu erwarten; hier versprechen Analysen kroatischer, serbischer und slowenischer Übersetzungen erste wichtige Anhaltspunkte (GRZYBEK 2003a);
4. schließlich wären entsprechende Analysen sowohl für andere Texte der russischen Poesie (vgl. GRZYBEK 2001) als auch poetische Texte anderer Sprachen durchzuführen; auch hier bieten Analysen kroatischer und slowenischer Texte von Mažuranić und Prešeren erste Einsichten (GRZYBEK 2003 c,d).

## A b k ü r z u n g e n


ALTMANN 1980: G. Altmann, Prolegomena to Menzerath's law, Glottometrika 2, 1-10

ALTMANN 1983: G. Altmann, H. Arens' »Verborgene Ordnung« und das Menzerathsche Gesetz, in: M. Faust (Hrsg.), Allgemeine Sprachwissenschaft, Sprachtypologie und Textlinguistik. Festschrift für Peter Hartmann, Tübingen, 31-39

ANTIĆ – KELIH 2003: G. Antić – E. Kelih, On the question of so-called 0-syllable words in determining word length, in: P. Grzybek (ed.), Word Length Studies [im Druck]

GRZYBEK 2001: P. Grzybek, Zur Mikropoetik bei V. Chlebnikov. Vortrag am Institut für Slawistik der Universität Salzburg

GRZYBEK 2002: P. Grzybek, Versuchen wir einmal, die Kräfte aus dem Gleichgewicht zu bringen … Quantitative Aspekte von Puškins Evgenij Onegin und Domik v Kolomne, in: J. Bernard – P. Grzybek – G. Withalm (Hrsg.), Form – Struktur – Komposition. Akten des III. Internationalen Symposiums »Offene Grenzen«, Graz [im Druck]

GRZYBEK 2003a: P. Grzybek, Evgenij Onegin in kroatischer, serbischer und slowenischer Übersetzung – Quantitative Aspekte. Vortrag beim XIII. Internationalen Slawistenkongress 2003, Ljubljana

GRZYBEK 2003b: P. Grzybek, Puškins Evgenij Onegin – Zur Frage der Homogenität des Textes und seiner Kapitel, Glottometrics 5 [in Vorb.]

GRZYBEK 2003c: P. Grzybek, Anmerkungen zur Wort- und Versstruktur in France Prešerens Krst pri Savici [in Vorb.]

GRZYBEK 2003d: P. Grzybek, Bemerkungen zur Wort- und Versstruktur in Smrt Smail-age Čengića von Ivan Mažuranić [in Vorb.]

HAYLES 1990: K. N. Hayles, Chaos Bound. Orderly Disorder in Contemporary Literature and Science, Ithaca/London

HAYLES 1991: K. N. Hayles (ed.), Chaos and Order. Complex Dynamics in Literature and Science, Chicago/London

WERNER 1999: H. C. Werner, Literary Texts as Nonlinear Patterns: A Chaotic Reading of »Rainforest«, »Transparent Things«, »Travesty«, and »Tristram Shandy«, Acta Universitatis Gothoburgensis, Goeteborg

WIMMER – ALTMANN 1996: G. Wimmer – G. Altmann, The Theory of Word Length Distribution: Some Results and Generalizations, in: P. Schmidt (Hrsg.), Glottometrika 15, 112-133

WIMMER ET AL. (1994): G. Wimmer – R. Köhler – R. Grotjahn – G. Altmann, Towards a Theory of Word Length Distribution, Journal of Quantitative Linguistics 1, 98-106

WIMMER – ALTMANN 2002a: G. Wimmer – G. Altmann, Unified derivation of some linguistic laws, in: Handbook of Quantitative Linguistics [im Druck]

WIMMER – ALTMANN 2002b: G. Wimmer – G. Altmann, Unified derivation of some linguistic laws, in: P. Grzybek (ed.), Word Length Studies [im Druck]