

Muller, Charles (1972), *Einführung in die Sprachstatistik*. München: Hueber.

Nikitopoulos, Pantelis (1973), *Statistik für Linguisten*. Tübingen: Narr.

Nübold, Peter (1974), *Quantitative Methoden zur Stilanalyse literarischer Texte*. Braunschweig.

Piirainen, Ilpo Tapani (1972), Quantitative Vorgehensweisen in der automatischen Analyse der älteren deutschen Trivialliteratur. In: Schanze 1972, 88–106.

Sabourin, Conrad F. (1994), *Quantitative and statistical linguistics, bibliography*. Montréal: Infolingua.

Schanze, Helmut (Hrsg.), *Literatur und Datenverarbeitung*. Tübingen: Niemeyer, 1972.

Thumb, Albert (1911), Experimentelle Psychologie und Sprachwissenschaft. Ein Beitrag zur Methodenlehre der Philologie. Teil 1 und 2. In: *Germanisch-Romanische Monatsschrift* 3, 1–15; 65–74.

Zwirner, Eberhard/Zwirner, Kurt (Hrsg.), *Grundfragen der Phonometrie*. 3 Bände. 2., erw. u. verb. Aufl. Basel u. a.: Karger, 1966–1969. (1. Auflage Berlin: Metten, 1936).

Dieter Aichele, Worms (Deutschland)

3. Zur Vorgeschichte quantitativer Ansätze in der russischen Sprach- und Literaturwissenschaft

1. Einleitung
2. Anfänge
3. Mathematik in der Literatur
4. Resümee
5. Literatur (in Auswahl)

1. Einleitung

Wissenschaft erfordert Kontinuität, und nicht nur Hoffnung auf eine Perspektive, sondern auch Kenntnis der Retrospektive.

A. A. Reformatskij

(Frumkina 1987, 80)

Im vorliegenden Beitrag soll, aufbauend auf synoptischen Darstellungen wie denen von Papp (1966) oder Kempgen (1995; 1999), ein Überblick über die Vorgeschichte quantitativer Ansätze in der russischen Sprach- und Literaturwissenschaft in Russland geboten werden. Der zeitliche Rahmen, um den es dabei geht, erstreckt sich im Wesentlichen von der zweiten Hälfte des 19. Jh.s bis in die 30er Jahre des 20. Jh.s – ein Zeitraum also, der in Russland von weitreichenden kulturellen und politischen Veränderungen geprägt war.

Die Notwendigkeit, diesem Zeitraum besondere Aufmerksamkeit zu schenken, ergibt sich zum einen aus der Tatsache, dass gerade in Russland bereits zu dieser Zeit zahlreiche quantitative Fragestellungen in einem Maße verfolgt wurden, wie dies anderswo nicht der Fall war (bzw. nicht bekannt ist). Zum anderen weist gerade der russische Bereich – und das ist natürlich

nicht isoliert von den politischen Rahmenbedingungen zu sehen – eine sehr eigene Geschichte von Kontinuitäten und Kontinuitätsbrüchen auf, deren Auswirkungen sich nicht auf den genannten Zeitraum beschränken, sondern mitunter bis in unsere Gegenwart reichen.

Im Hinblick auf die genannte Fragestellung gilt es – vor allem auch im Vergleich zu anderweitigen Darstellungen – insbesondere darauf hinzuweisen, dass die folgenden Ausführungen sich nicht auf den engeren Bereich der quantitativen Linguistik (QL) beschränken, sondern allgemein den Bereich der Quantifizierung sprachlicher und textlicher Strukturen, also unter Einschluss auch literaturwissenschaftlicher (so etwa metrischer, versanalytischer, u. a.) Ansätze umfassen. Ebenso gilt es, einleitend auf die Notwendigkeit einer Präzisierung des Gegenstandsbereiches im Zusammenhang mit geschichtlichen und politischen Faktoren hinzuweisen: So wäre es im Hinblick auf die Anfänge der QL im Russland des 19. Jh.s aus historischer Sicht unzulässig, dem damaligen Russischen Zarenreich angehörige, heute jedoch selbständige Nationalstaaten (wie etwa Polen oder die Ukraine) aus dem Kreis der Betrachtung auszublenden. Ohne Frage ist jegliche diesbezügliche Ein-, Aus- oder Abgrenzung überaus problematisch; ungeachtet dessen sollen im vorliegenden Beitrag solche Arbeiten als dem Gegenstandsbereich der QL in Russland zugehörig angesehen werden, die in weiterem Sinne dem russischen Be-

reich zugerechnet werden können. Die zur Sprache kommenden Arbeiten können sich somit, müssen sich aber nicht auf die russische Sprache beziehen, und sie können, aber müssen nicht auf russisch publiziert worden sein. Der aufgezeigten Erweiterung des Gegenstandsbereichs steht somit eine Einschränkung gegenüber, die den vorliegenden Beitrag der Pflicht enthebt, weltweite quantitative Forschungen zur russischen Sprache zu berücksichtigen.

2. Anfänge

Sowohl ältere (vgl. Struve 1918) als auch neuere Darstellungen zur Geschichte der QL in Russland beziehen sich gern auf den Mathematiker *Viktor Ja. Bunjakovskij* (1804–1889), der – immerhin als damaliger Vizepräsident der Russischen Akademie der Wissenschaften – 1847 die Anwendung statistischer Verfahren auf grammatische, etymologische und sprachhistorisch-vergleichende Fragestellungen und somit die Etablierung einer *Linguostatistik* [статистика языка] forderte (vgl. Bunjakovskij 1847; Grzybek 2003). Zwar verwies Bunjakovskij auf eigene Forschungen in dem genannten Bereich – diese sind jedoch entweder nie durchgeführt worden, oder aber sie sind nie erschienen bzw. nicht erhalten.

Somit werden häufig die ersten sprachstatistischen Untersuchungen in Russland zeitlich gegen Ende des 19. Jh.s angesetzt – eine Zeit also, als der Linguistik im Wesentlichen noch die Entwicklung synchron ausgerichteteter Verfahren zur strukturellen Beschreibung von Sprache(n) bevorstand, und als auch die Statistik noch lange nicht ihre Methoden im heutigen Verständnis ausgearbeitet hatte. Bei diesen in Betracht gezogenen Arbeiten handelt es sich um Graphem- oder Laut-Statistiken. Eine erste solche Graphemstatistik findet sich z. B. bei *Anton S. Budilovič* (1846–1908), einem Schüler des in Petersburg lehrenden *Izmail I. Sreznnevskij* (1812–1880) – der übrigens 1839 eine von der Universität Charkov zunächst (1833) abgelehnte Dissertation über den Zusammenhang von Nationalökonomie und Statistik (*Opyt o predmete i elementach statistiki i političeskoj ekonomii sravnitel'no*) vorgelegt hatte. Die in seiner *Grammatik des Kirchenlawischen* enthaltene Graphemstatistik von *Budilovič* (1883) – der später (1892–1902) Rektor der Universität in

Tartu (Dorpat, bzw., nach der Russifizierung Anfang der 90er Jahre, Jur'ev) werden sollte – basierte allerdings nicht, wie mitunter angenommen wird, auf eigenen Auswertungen; vielmehr übernahm er sie aus der *Formenlehre des Kirchenlawischen* (Schleicher 1852) des damals in Prag lebenden und lehrenden August Schleicher, der seinerseits damit frühere Angaben von Förstemann (1846; 1852) zum Deutschen, Griechischen, Lateinischen und Gotischen ergänzen wollte; genauere Darstellungen dieses Sachverhalts finden sich bei Grzybek/Kelih (2003; 2004).

In Anbetracht dieses Umstandes sind es also die 90er Jahre des 19. Jh.s, aus denen die ersten originellen, quantitativ ausgerichteten Untersuchungen zum Russischen vorliegen. So legte 1891 der ukrainische Ethnologe *Mitrofan A. Dikarev* (1854–1902) eine erste Lautstatistik vor (vgl. Dikarev 1891); er hatte sich als Sekretär des statistischen Komitees im Gouvernement Voronež unter der Anleitung von *F. A. Ščerbina* (1849–1936), einem ausgewiesenen Statistiker und Begründer der Volkswirtschaftsstatistik (vgl. Ščerbina 1925), mit statistischen Methoden vertraut gemacht. Dikarevs Arbeit ist insofern bemerkenswert, als sie bewusst auf der Basis von Texten und nicht von Wörterbuchmaterial erstellt wurde. Offenbar wollte Dikarev so dem systemischen Charakter sprachlicher Erscheinungen Rechnung tragen: Individuelle Wörter betrachtete er in Analogie zu einzelnen Bauernhöfen – so wie letztere in Dorfgemeinschaften organisiert sind, sind es Wörter in Sätzen, und so wie die Dörfer den Gouvernements unterstehen, folgen Sätze den Regeln eines vollständigen Textes. Es ist also weniger der Faktor der Frequenz, den Dikarev als distinktiv zwischen Wörterbuch- und Textmaterial hervorhob, als vielmehr der intrinsisch-systemische Charakter, den er offenbar erkannte.

Ein erstes Frequenzwörterbuch wurde 1894 in Kiev von *V. N. Kunickij* vorgelegt, zeitlich also noch vor dem *Häufigkeitwörterbuch der deutschen Sprache* von Kaeding (1897/98), das üblicherweise als das erste seiner Art angesehen wird (vgl. Kunickij 1894). Allerdings bezog sich das Werk von Kunickij „nur“ auf den Text der Komödie *Gore ot uma* von A. S. Griboedov (unter Auslassung der sprechenden Personen, Text- und Regieanweisungen, u. ä.). Eingang fanden so 3370 verschiedene Wörter, basierend auf 13246 Wortverwendungen, die in alpha-

betischer Reihenfolge angeordnet wurden; den einzelnen Einträgen folgten neben der jeweiligen Vorkommenshäufigkeit die Nummern der Akte des Vorkommens (vgl. Ermolenko 1974).

Während es sich bei diesen ersten Arbeiten im Grunde genommen also um mehr oder weniger rein empirische Datenerhebungen handelte, wurden Anfang des 20. Jh.s quantitative Ansätze auch bereits in linguistische, vor allem sprachhistorische Argumentationszusammenhänge gestellt. Das gilt z. B. für *Dmitrij N. Kudrjavskij* (1867–1920); auch Kudrjavskij war, ebenso wie Budilovič, Absolvent der Petersburger Universität, und auch er lehrte später an der Universität Tartu [Jur'ev]. Kudrjavskij, der aufgrund seiner sprachpsychologischen Arbeiten mitunter als Nachfolger von *Aleksandr A. Potebnja* (1835–1891) angesehen wird, wandte in verschiedenen Arbeiten (vgl. Kudrjavskij 1909; 1911; 1912) extensiv statistische Verfahren an, mit dem Ziel, im Bereich der Linguistik zur objektivierenden Beschreibung sprachlicher Gesetzmäßigkeiten gelangen zu können. Dies formulierte Kudrjavskij (1909, 53) explizit, seiner Auffassung Ausdruck gebend, „dass die statistische Methode die Möglichkeit gibt, Erscheinungen zu bemerken, die sich normalerweise der Aufmerksamkeit des Forschenden entziehen.“ Für ihn lag die Bedeutung statistischer Methoden darin, „dass sie unserer Forschung eine Richtung verleiht. Und diese Richtung wird nicht durch unsere subjektiven Sichtweisen, sondern durch unerschrockene Ziffern bestimmt“ (Kudrjavskij 1909, 54).

Die Tatsache, dass sich im damaligen Dorpat/Jur'ev ein frühes Zentrum quantitativer Linguistik herausbildete, ist mitnichten zufällig. Die Querbezüge dieser historischen Grundlagen der QL in Tartu – wie sie sich mit dem 1976 erschienenen 1. Band der *Arbeiten zur Sprachstatistik* [Tõid keelestatistika alalt / Trudy po lingvostatistike], herausgegeben von der Gruppe für Sprachstatistik an der Philologischen Fakultät der Universität Tartu unter der Leitung von J. Tuldava, herausbildete – wären eigens zu untersuchen. Jedenfalls kamen zu Beginn des 20. Jh.s die entscheidenden Personen von der Petersburger Universität, die seinerseits zu einem erheblichen Maße von der Sprachwissenschaft der sog. Kazaner Schule geprägt war, welche nachhaltig den Boden für

die Herausbildung einer QL in Russland bereitete.

Die sog. Kazaner Schule hatte sich gegen Ende des 19. Jh.s unter der Leitung von *Ivan A. [Jan Ignacy Nicislaw] Baudouin de Courtenay* (1845–1929) herausgebildet. In seinem konstruktiv-kritischen Umgang mit den Prinzipien des Junggrammatikalismus legte Baudouin die Grundlagen zu dessen Überwindung und warf den Schatten auf Konzeptionen voraus, die später dann allerdings eher mit dem Namen von Ferdinand de Saussure verbunden werden sollten (mit dem er übrigens in engem persönlichen Kontakt stand).

Nach seinem Studium in Warschau und nach seiner Dissertation 1870 in Petersburg bei Sreznevskij war Baudouin 1875 nach Kazan, also an die östlichste Universität des damaligen russischen Reiches, berufen worden, wo er zunächst bis 1883 blieb.

Auch wenn Forderung und Anwendung mathematisch-statistischer Methoden in der Sprachwissenschaft nicht im Vordergrund seines Denkens standen, durchziehen doch entsprechende Hinweise das gesamte Œuvre Baudouins schon seit seinen ersten Kazaner Vorlesungen der Jahre 1877/78, in denen er die qualitative und quantitative Beschreibung sprachlicher Systeme forderte. Gerade in Kazan musste ein solches Postulat auf offene Ohren stoßen: Denn Kazan war im 19. Jh. vor allem im Bereich der Mathematik eine der führenden russischen Universitäten. So lehrte hier seit 1874 – also zeitgleich mit Baudouin – *Aleksandr V. Vasil'ev* (1853–1929), ein Mathematiker höchsten Ranges und Herausgeber der Werke des herausragenden Mathematikers *Nikolaj I. Lobačevskij* (1792–1856), der seinerseits selbst in Kazan studiert hatte, wo er dann auch 1822 eine ordentliche Professur erhielt und 1827–1846 Rektor der Universität war.

Zu den bedeutendsten Kazaner Studenten Baudouins zählten u. a. *Mikołai Kruszewski* (1851–1887) und *Vasilij A. Bogorodickij* (1857–1941), die später beide Baudouin auf dem Kazaner Lehrstuhl folgten: zunächst Kruszewski 1883 bis zu seinem frühen Tod 1887, dann Bogorodickij. Letzterer bezeichnete sein 1908 in Kazan erschienenes (stark von der Mathematik Lobačevskijs beeinflusstes) Buch *Osnovy arifmetičeskogo sčeta* [Grundlagen des arithmetischen Rechnens] (vgl. Bogorodickij 1908) als *induktive Philologie*; 1913 legte er umfangreiche Daten aus Studien zu Lautfrequenzen des Rus-

sischen vor, deren Ergebnisse u. a. in seinem *Obščij kurs russkoj grammatiki* [Allgemeinen Kurs der russischen Grammatik] (vgl. Bogorodickij 1935, 26 ff.) zu finden sind. Ebenfalls in Kazan erschienen um eben diese Zeit eine ganze Reihe von sprachstatistischen, vor allem lautstatistischen Arbeiten, so etwa von G. A. Aleksandrov (1911), V. Petrov (1911), oder V. Ermolaev (1915), der in seinen mathematisch-statistischen Bemühungen explizit auf die Anregungen seines Lehrers Bogorodickij verwies. De facto aber handelte es sich bei all diesen Untersuchungen um die Realisierung eines Postulats, das schon zuvor Baudouin in seinen frühen Kazaner Vorlesungen aufgestellt hatte, als er dazu aufforderte, die „Procentnyja otnošenija količestva otdel'nych zvukov v danom jazyke“ [die Prozentsätze der Laute in einer gegebenen Sprache] statistisch zu untersuchen und zu bewerten.

Baudouin selbst war 1883 zunächst an die Universität in Tartu übersiedelt, wo er neben seinen sprachwissenschaftlichen Studien auch höhere Mathematik studierte. Daran anschließend – nach der Russifizierung der Universität – ging er 1894 zunächst nach Krakau, bevor er 1900 nach Petersburg zurückkehrte, um dann schließlich vor Erlangen der politischen Selbständigkeit Polens (1918) bis zu seinem Tod nach Warschau zu gehen. Auch während seiner Petersburger Zeit wies er wiederholt dezidiert auf die Bedeutung der Mathematik für die Linguistik hin: Seiner Vorstellung nach sollte „die Linguistik sich mehr und mehr den exakten Wissenschaften annähern“ und „mehr und mehr quantitatives Denken nach dem Modell der Mathematik“ einführen (Baudouin 1909, 267). Vor diesem Hintergrund forderte er eine *neue Mathematik* [nowa matematyka], eine *Mathematik der Zukunft* [matematyka przyszłości]: Für Baudouin (1909, 268) ging es primär darum, dass die Untersuchung sprachlicher Fakten eine objektive Angelegenheit werden muss. Deswegen erwartete er von der Linguistik – ebenso wie von der Mathematik –, „unendliche Quantitäten auf endliche zu reduzieren, die dem analytischen Denken unterworfen werden können“ (Baudouin 1909, 267 f.).

Noch zwei Jahre vor seinem Tod setzte Baudouin (1927) sich explizit mit dem Problem der Quantität im sprachlichen Denken auseinander: „Die einzige Methode, die eine mathematische Grundlage hat und die bislang in der Sprachwissenschaft anwendbar

ist, ist die statistische Methode, d. h. die empirische Definition des wechselseitigen prozentualen Verhältnisses der verschiedenartigen Erscheinungen des sprachlichen Lebens (prozentuale Relationen in der beschreibenden und historischen Phonetik, in der Morphologie usw.). Andere mathematische Begriffe [...] werden bislang noch nicht angewendet“ (Baudouin 1927, 323). Er forderte deshalb abermals eine *Mathematik der Zukunft*, um im Sinne einer exakten Wissenschaft die „Gesetze der Welt der Sprache“ beschreiben zu können, welche die Abhängigkeit verschiedener Größen voneinander formal ausdrücken können (Baudouin 1927, 324).

Zu den Hörern der Petersburger Vorlesungen von Baudouin gehörten zahlreiche Sprach- und Literaturwissenschaftler, von denen sich ohne Übertreibung sagen lässt, dass sie alle in den folgenden Jahren bzw. Jahrzehnten nachhaltig das Profil (nicht nur) der russischen Sprach- und Literaturwissenschaft verändern und ausnahmslos zu führenden Personen des akademischen Betriebs werden sollten, so u. a. Sergej I. Bernštejn (1892–1970), Boris M. Ejchenbaum (1886–1959), Lev P. Jakubinskij (1892–1945), Evgenij D. Polivanov (1891–1937/38?), Viktor B. Šklovskij (1893–1984), oder Lev V. Ščerba (1880–1944). Sie alle stellten auch, gemeinsam mit Jurij N. Tynjanov (1894–1943) und Boris V. Tomaševskij (1890–1957), das personelle Gerüst der 1916 gegründeten Petersburger „Gesellschaft zur Erforschung der poetischen Sprache“ (*Opojaz*) dar. In dem Bestreben, die Sprache der Literatur mit Hilfe linguistischer Methoden zu untersuchen, hatte diese Gruppe einen gemeinsamen Interessenschwerpunkt mit den Angehörigen des Moskauer Linguistischen Kreises (MLK). Der 1915 auf die Initiative von Roman O. Jakobson (1896–1982) – damals noch Student – und Pëtr G. Bogatyřev (1893–1971) gegründete MLK bestand bis 1924; erster Präsident war Jakobson (1915–20), 1920 gefolgt von Michail N. Peterson (1885–1962), dann 1921 von A. A. Buslaev, einem Enkel von Fedor I. Buslaev, und schließlich 1922–24 von Grigorij O. Vinokur (1896–1947), der 1915 erster Sekretär des MLK gewesen war.

Im Vergleich zum Petersburger *Opojaz* war der MLK in methodologischer Hinsicht stärker linguistisch als literaturwissenschaftlich ausgerichtet, so dass bei der Erforschung der Literatur primär die Sprache in

ihrer ästhetischen Funktion im Vordergrund stand. Bei der Gründung hatte man sich aus Gründen der formalen Legitimation mit der Moskauer Dialektologischen Kommission an der Akademie der Wissenschaften assoziiert. Federführend waren dort *Aleksej A. Šachmatov* (1864–1920) und *Dmitrij N. Ušakov* (1873–1942), beide Schüler von *Filipp F. Fortunatov* (1848–1914), dem Begründer der sog. Moskauer Schule der Linguistik. Diese hatte sich nahezu zeitgleich mit der Kazaner Schule etwa ab den 70/80er Jahren des 19. Jh.s herausgebildet, und auch mit ihr hatte sich wesentlich der Junggrammatikalismus in Russland etabliert. Interessanterweise entstanden die Arbeiten sowohl der Moskauer als auch der Kazaner Schule vor dem Hintergrund eines allgemein-semiotischen Verständnisses von Sprache (vgl. Grzybek 1999): So war etwa in der Kazaner Schule für Bogorodickij Sprache ein „Zeichen für den Ausdruck von Begriffen und Gedanken“, und auch für Kruszewski war nicht nur das Wort allgemein ein „Zeichen des Dinges“, sondern er verstand Sprache als ein „System von Zeichen“, was ihn mit der (späteren) Linguistik Saussures vereinen sollte. Und auch Fortunatov (1956) deklarierte in seinen Vorlesungen Sprache als eine „Menge von Zeichen“, „Zeichen dessen, was absolut nicht unmittelbar in unserem Denken repräsentiert sein kann“ (Fortunatov 1956, 111 f.).

In der unmittelbaren Tradition Fortunatovs – der zufolge es Sprache vor allem in ihrem formalen Aspekt zu untersuchen galt – standen später dann führende russische Linguisten wie *Aleksandr M. Peškovskij* (1848–1933), *Michail N. Peterson* (1885–1962), *Viktor K. Poržežinskij* [*Porzeziński*] (1870–1929), oder eben die bereits erwähnten Šachmatov und Ušakov. Das Kazaner Bemühen um wissenschaftliche Exaktheit einerseits, und die streng formale Ausrichtung der Moskauer Schule andererseits stellten somit seitens der Linguistik essentielle Grundlagen für die Herausbildung der sog. Russischen Formalen Schule (Russischer Formalismus) dar, wie die Zusammenarbeit des Petersburger *Opojaz* und des MLK bezeichnet wurde. Genährt durch das gemeinsame Interesse an insbesondere auch synchron ausgerichteten Fragen, und in enger (personeller und sachbezogener) Verwobenheit mit der damaligen Avantgardekunst stand im Fokus dieser sprach-, literatur- und kunsttheoretischen Richtung die Untersu-

chung der künstlerischen Form. Form wurde hierbei nicht als Gegensatz, sondern als dialektisches Komplementat zum Inhalt verstanden, so dass ganz zwangsläufig die Frage nach der semantischen Funktion formaler Elemente in den Vordergrund rückte – was bei der formal-grammatischen Linie Fortunatovs so übrigens nicht der Fall war, so dass es den Begriff des *Formalismus* hier sehr differenziert zu sehen gilt.

Ungeachtet dessen liegt es auf der Hand, dass die strikte Trennung von Sprach- und Literaturwissenschaft im Russischen Formalismus aufgebrochen und de facto auch nicht aufrecht zu halten war. Im gemeinsamen Streben nach der Entwicklung und Etablierung einer synchronen Sprach- und Literaturwissenschaft als einer exakten Wissenschaft war der Rückgriff auf (damals verfügbare) mathematisch-statistische Verfahren zur Beschreibung sprachlicher und poetischer Strukturen mehr als nahe liegend und wurde in der Tat auch vollzogen. Während es sich im Bereich der Sprachwissenschaften allerdings überwiegend um die oben bereits genannten, in der Tradition der Kazaner und Moskauer Schule stehenden Linguisten handelte, war die Diskussion quantitativer Verfahren im Bereich der Literaturwissenschaft stärker durch ein weiteres, nicht unbedingt zum personellen Kernbestand der Formalen Schule zählendes Umfeld geprägt. Dieser Umstand erklärt auch die Tatsache, warum entsprechende historische Darstellungen des Russischen Formalismus (vgl. Erlich 1973; Hansen-Löve 1978; Steiner 1984) diesem Aspekt bestenfalls sporadisch gerecht werden; eine Aufarbeitung der Diskussion quantitativer Verfahren insbesondere im literaturwissenschaftlichen Bereich steht bis heute noch aus und kann im Folgenden auch nur ansatzweise geleistet werden.

3. Mathematik in der Literatur

Angeregt waren zahlreiche Arbeiten in dieser Richtung nicht zuletzt von Dichtern, die sich auch theoretisch mit den Grundlagen der poetischen Sprache und formalen Aspekte der Poesie beschäftigten. In erster Linie sind hier *Andrej Belyj* (1880–1934) und *Valerij Ja. Brjusov* (1873–1924) zu nennen.

Belyj, mit eigentlichem Namen Boris N. Bugaev, war ein Sohn von *Nikolaj Vasil'evič Bugaev* (1837–1903), einem Mathematiker,

der Dekan der physikalisch-mathematischen Fakultät der Moskauer Universität war. Belyj war nicht nur einer der wichtigsten symbolistischen Dichter seiner Zeit, sondern auch ein einflussreicher Literatur- und Kulturtheoretiker. In seinem 1910 erschienenen literatur- und kulturtheoretischen Sammelband *Simvolizm* (Belyj 1910) waren unter anderem eine Reihe statistischer Untersuchungen zur Rhythmik und Metrik poetischer Texte enthalten. Für diese Arbeiten erhielt Belyj später – ungeachtet der Tatsache, dass sie mitunter heftig kritisiert wurden – die ungeteilte Anerkennung, statistische Verfahren in die Analyse poetischer Texte eingeführt zu haben.

Ausgangsziel Belyjs war ohne Frage eine wissenschaftliche Beschäftigung mit Literatur bzw. Poesie. So ging er in seiner 1909 als Vortrag gehaltenen Studie „Лирика и эксперимент“ [Lyrik und Experiment] davon aus, dass die Grenzen exakten Wissens nicht durch das Objekt gegeben seien, sondern durch eine bestimmte Perspektive auf das Objekt (Belyj 1910, 231). Zwar sei es im Bereich der Ästhetik nicht zur Herausbildung einer solchen (eigenständigen) exakten Wissenschaft gekommen (Belyj 1910, 236) – diese sei auch nur als System von Wissenschaften denkbar – dennoch sei durchaus die wissenschaftliche Betrachtung eines poetischen Objekts möglich. So sei eine der Aufgaben der zeitgenössischen Poetik die exakte Bestimmung des Rhythmus, vor allem dessen Abgrenzung vom Metrum (Belyj 1910, 254). Als ein erstes Beispiel zur Veranschaulichung dessen, worauf er hinaus will, dienten Belyj (Belyj 1910, 259) die ersten beiden Verszeilen aus Puškins *Evgenij Onegin*, einem in vierfüßigem Jambus geschriebenen Versroman:

*Мой дядя самых честных правил,
Когда не в шутку занемог, [...]*

Mein Onkel, brav und hoch zu preisen,
scheint nun doch ernsthaft krank zu sein

Die erste Zeile näherte sich, so Belyj, aufgrund des Zusammenfallens der Betonungen in den Wörtern mit den langen Silben der jambischen Struktur an:

Мой дя | дя са | мых че | стных пра | вил

Die zweite Zeile hingegen weiche von der reinen jambischen Struktur ab, weil das Wort *занемог* eine zweite Betonung bzw. eine Nebenbetonung annehme, so dass der

Jambus (U–) durch eine Pyrrhichie (UU) ersetzt werde.

Когда | не в шут | ку за | немог
U– | U– | UU | U– |

Diese Pyrrhichien, deren Vorkommen Belyj als einen Beschleunigungsfaktor ansieht, können zwar in jedem der vier Takte vorkommen; die Art des Vorkommens aber ist von Dichter zu Dichter verschieden. So lautet jedenfalls das Ergebnis statistischer Berechnungen von Belyj, der Texte von 24 Dichtern im Hinblick auf diese Fragestellung verglichen hat und aufgrund der Ergebnisse die Poeten intuitiv in drei Gruppen einteilte (Belyj 1910, 262); versucht man mit modernen Methoden wie etwa Cluster-Analysen eine solche Gruppenbildung zu reproduzieren, so erweisen sich immerhin 20 der 24 Zuordnungen als haltbar. Abgesehen von seinen auf den genannten Berechnungen beruhenden Schlussfolgerungen hat Belyj jedoch noch einen anderen Weg der Bearbeitung eingeschlagen, nämlich die Erarbeitung spezifischer Graphiken, die ihm als Grundlage weiterer Berechnungen dienen. Verbindet man nämlich wie in Abb. 3.1a (Belyj 1910, 260) die jeweiligen Pyrrhichien miteinander, so ergeben sich bestimmte geometrische Figuren wie z. B. die in Abb. 3.1b dargestellten:

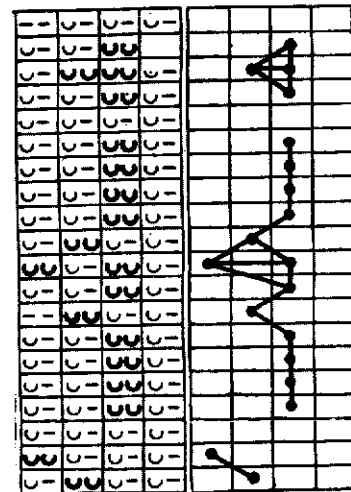


Abb. 3.1a: Visualisierung von Pyrrhichien nach Belyj (1910, 260)

Aufgrund der Art und Häufigkeit des Vorkommens dieser Figuren unterschied Belyj einen „armen“ und einen „reichen“ Rhyth-







Dach	umgekehrtes Dach	Quadrat	Leiter	Rhombus	Kreuz
					

Abb. 3.1b: „Figuren“ nach Belyj (1910)

mus, verglich abermals die Vorkommenshäufigkeit in den Texten der 24 Autoren je nach Vorkommensstelle in der Verszeile, usw.

Während Belyj (1910, 331) seine „graphischen und statistischen Tabellen“ allerdings noch als einen Beitrag zur „Anatomie des Stils“ (Belyj 1910, 286) ansah, und sie im Gegensatz zu subjektiven Wertungen als „unvoreingenommene Strukturbeschreibung“ betrachtete, betonte er später, seine Untersuchungen seien „experimentell, nicht theoretisch“ (Belyj 1929, 25). Ungeachtet dessen bezeichnete er sich selbst explizit als „Formalist vor den Formalisten in Russland“ (Belyj 1929, 28). Diese rückblickende Aussage lässt sich fraglos als Abgrenzung gegen die russischen Formalisten werten, insbesondere als Angriff gegen Viktor M. Žirmunskij (1891–1971), der seit 1920 Leiter der Abteilung für Literaturgeschichte am Petersburger Staatlichen Institut für Kunstgeschichte war, und der sich in seinem Buch *Vvedenie v metriku* (1925) u. a. kritisch mit Belyj auseinandergesetzt hatte. Ähnlich hatte auch Tomaševskij (1923b, 3) betont, dass es in Russland keine Wissenschaft vom Vers gebe: die Arbeiten von Belyj seien zwar durchaus wissenschaftlichen Charakters, ergäben aber kein wissenschaftliches System. Genau solche Stellungnahmen veranlassten Belyj dann offenbar, sich selbst insofern als „formalistischer als die Formalisten selbst“ (Belyj 1929, 25) zu nennen, da er die formale Untersuchung der Poesie bis an ihre Grenzen, nämlich bis zur Mathematik, geführt habe (Belyj 1929, 25). Damit weist Belyj allerdings genau auf den Punkt hin, welcher die methodologische Kritik auf sich zog – die Tatsache nämlich, dass er es nicht bei mathematisch basierten Strukturbeschreibungen beließ, sondern die auf den Berechnungen basierenden geometrischen Figuren spekulativ interpretierte und bewertete. So sah es auch Žirmunskij (1929, 204) in seiner Reaktion auf Belyjs (1929) *Ritm kak dialektika*: Berechnungen und Graphi-

ken, die eigentlich Hilfsmittel zur größeren Anschaulichkeit abstrakter Schlussfolgerungen sein sollten, würden bei Belyj zu einer Art statistischer Gesetze des Rhythmus, zu einer Art Zahlenmystik.

Ungeachtet aller (berechtigten und unberechtigten) Kritik, die Belyj entgegen gebracht wurde, folgten dem von ihm in seinem Buch *Simvolizm* (1910) eingeschlagenen Weg recht bald eine Reihe von Literaten und Vertheoretikern. Als erstes zu nennen sind hier *Sergej M. Luk'janov* (1855–1935) und *Valerian A. Čudovskij* (1891–1938?).

Luk'janov veröffentlichte 1914 eine Studie des Gedichts „Angel' Smerti“, das von dem zeitgenössischen, ein Jahr zuvor verstorbenen Poeten *Arsenij A. Goleniščev-Kutuzov* (1848–1913) stammte. Mit dieser Studie stellte *Luk'janov*, Medizin-Professor, Literat und Biograph des bedeutenden russischen Religionsphilosophen *Vladimir S. Solov'ev* (1853–1900), im Prinzip Fragen, die erst sehr viel später – dann freilich auf höherem Niveau – die quantitative Stilistik beschäftigen sollten. Für *Luk'janov* (1914) bestand der Wert seiner Untersuchungen insbesondere darin, „dass sie zur Erhellung der äußeren und technischen Mittel beitragen können, derer sich ein gegebener Dichter zum Ziel des künstlerischen Ausdrucks bedient“ (*Luk'janov* 1914, 351). Für ihn öffnete sich damit ein wissenschaftliches Feld, das bis dato noch kaum ausgearbeitet sei – nämlich die Physiologie und Psychologie des Unbewussten. *Luk'janov* bezog sich hierbei in erster Linie auf die damals aktuellen (und teilweise auch ins Russische übersetzten) Schriften von *Eduard von Hartmann* (*Die Philosophie des Unbewussten*, 1869), von *William James* (*The Varieties of Religious Experience*, 1902), oder *Joseph Grasset* (*Introduction physiologique à l'étude de la philosophie*, 1908). In seinen eigenen Analysen untersuchte *Luk'janov* u. a. die (in Silben berechnete) Vorkommenshäufigkeit von Wortlängen innerhalb der einzelnen Stro-

phen und im Gesamttext, und verglich sie mit einem Text von A. S. Puškin. Dabei stellte er fest, dass der Anteil ein- und zweisilbiger Wörter in dem untersuchten Text mit 73 % im Vergleich zu 70 % nur unbedeutend höher sei; auch den prozentualen Anteil von Verben und Substantiven untersuchte er und stellte keine wesentlichen Unterschiede fest; einer detaillierten Analyse des Konsonanten- und Vokalbestandes ließ er Untersuchungen zum lexikalischen Reichtum folgen und arbeitete heraus, dass im „Angel' Smerti“ von den 96 Wörtern 55 nur ein einziges Mal vorkommen, 41 hingegen zwei- oder mehrmals, so dass seiner Meinung nach von „lexikalischem Geiz“ keine Rede sein könne (Luk'janov 1914, 330); schließlich ordnete er den Text aufgrund eines Vergleichs der Textlänge in das Gesamtwerk des Autors ein.

Čudovskij, für den die Untersuchung des Verses als einer formalen Erscheinung in erster Linie eine Frage der Klassifikation ist (Čudovskij 1915, 57), ging es in einer Reihe von Aufsätzen (Čudovskij 1914; 1915; 1917) in erster Linie um „eine Theorie über den Zusammenhang von Form und Inhalt“ (Čudovskij 1914, 110), die er im Wesentlichen allerdings als eine „Angelegenheit der Zukunft“ betrachtete.

Bei der Vertiefung der von Belyj aufgeworfenen Frage nach einer adäquaten Differenzierung von Rhythmus und Metrum wirft Čudovskij (1914, 116) im ersten der drei Texte die Frage auf, wie viel Prozent der Silben im Russischen überhaupt eine Betonung aufweisen. Da seines Wissens diesbezüglich keinerlei sprachwissenschaftliche Untersuchungen vorliegen, führte er selbst Analysen an der Prosa von Turgenev durch, deren Unvollständigkeit er durchaus selbst einräumte. In der Tat sollte erst Jahre später Šengeli (1923) diese Frage noch einmal aufgreifen und in größerem Maßstab untersuchen (s. u.). Jedenfalls kam Čudovskij – ungeachtet der Problematik der Behandlung von Neben-Betonungen, für die bis dato noch keine Klassifikation erarbeitet worden sei – auf einen Prozentsatz von 35–42 %.

Der Schluss liege deshalb auf der Hand, dass in einem fünf Fußigen Jambus theoretisch nicht mehr als $4\frac{1}{2}$ Betonungen pro Verszeile zu erwarten seien. Die Bestätigung einer solchen Schlussfolgerung bedürfe jedoch, so Čudovskij (1914, 116), einer „kolossalen statistischen Arbeit“. So sei es zum Beispiel in Anbetracht des genannten Pro-

zentsatzes nicht angebracht, davon auszugehen, dass in einem Text mit fünf Fußigem Jambus $3\frac{1}{2}$ Betonungen pro Verszeile vorkämen. Vor dem Hintergrund dieses zu erwartenden Minimums einerseits und der ein Vorkommen von fünf Betonungen voraussetzenden Norm andererseits finde offensichtlich seitens des Dichters eine spezifische lexikalische Auswahl statt. Während Čudovskij (1914, 117 ff.) solche Schwankungen innerhalb eines Textes exemplarisch aufzeigte, forderte er in Weiterentwicklung dieser Überlegungen später die „vergleichende Berechnung der relativen Häufigkeit von Wörtern nach verschiedenen logometrischen Typen, d. h. nach Stelle der Betonung und Anzahl der Silben“ (Čudovskij 1917, 69). Als Ergebnis entsprechender Untersuchungen kam er zu dem Schluss, „dass Puškin zu einem gewissen Grad ein eigenes Lexikon für jedes Versmaß hatte [...], auf bestimmte Art vielleicht ein Art Determinante des Rhythmus ...“ (Čudovskij 1917, 69).

Čudovskij (1915, 61) entwickelte ein Verfahren zur Beschreibung der Versstruktur, das er selbst als „mathematisch“ bezeichnete, und dessen Basisschema wie folgt aussah:

Jambus	$2n$
Trochäus	$2n+1$
Anapäst	$3n$
Daktylus	$3n+1$
Amphybrachys	$3n+2$

Zur detaillierten Beschreibung eines Verses verwendete er das Plus- und Minuszeichen (+ bzw. –) zur Bezeichnung von Erweiterungen bzw. Kürzungen des Versschlusses (sog. hyperkatalektische bzw. katalektische Versschlüsse), gab in eckigen Klammern zusätzlich die Stelle des Ausfalls von Betonungen an, und führte die Anzahl der Wörter pro Verszeile inkl. Anzahl der Betonungen (in Form von hochgestellten Indizes) bei Wörtern mit zwei oder mehr Betonungen an. Eine Verszeile wie *Морозой пылю се-ребрится* erhielt demnach die folgende Form:

$$\frac{4(2n) + 1[3]}{3, 2, 4^2}$$

Auf dieser Grundlage berechnete Čudovskij (1915, 84) die Vorkommenshäufigkeiten der verschiedenen Möglichkeiten in verschiedenen Texten Puškins und nahm damit methodologisch Untersuchungen vorweg, die V. Nabokov erst Jahrzehnte später in seinem

Kommentar zu *Evgenij Onegin* zur Anwendung bringen sollte.

Ungeachtet seiner eigenen Berechnungen forderte Čudovskij (1917, 58) allerdings die genaue Berechnung der Abweichung von Durchschnittswerten; die Behandlung dieser Frage werde jedoch durch das Problem des Verhältnisses von Norm und spezifischem Text kompliziert: es stelle sich nämlich die grundsätzliche Frage, welche genauen Berechnungen eigentlich auf der Basis einer falsch gewählten Norm möglich seien.

Sowohl Luk'janov als auch Čudovskij entwickelten die von Belyj aufgeworfenen Fragen auf die eine oder andere Art weiter. Überaus positiv schätzte später deshalb auch der Statistiker Setnickij (s. u.) die Arbeiten von Luk'janov und Čudovskij ein. Die Arbeit von Luk'janov war für ihn ein Beispiel der Art von Daten, über die man für jeden einzelnen Dichter und für die russische Poesie insgesamt verfügen sollte (Setnickij 1922, 24). Und die Arbeiten von Čudovskij nannte er „den Moment, ab dem man von einer seitens der Verstheoretiker klar erkannten Notwendigkeit der ernsthaften, auf entsprechender statistischer Bearbeitung des Materials begründeter Erforschung des Verses“ sprechen könne (Setnickij 1922, 25). Insofern sei Čudovskij zweifellos als „Impuls für die breite, methodologisch fundierte und durchdachte Anwendung der statistischen Methode in der Verswissenschaft“ anzusehen (Setnickij 1922, 26).

Diese positive Entwicklung bemerkten auch zeitgenössische Rezensenten. So kritisierte etwa Bobrov in seiner 1916 erschienenen Aufsatzsammlung *Zapiski stichotvorca* (Bobrov 1916) an einem ebenfalls in Anlehnung an Belyj geschriebenen Beitrag von Nedobrovo (1912), dass dieser sich „in direkter Abhängigkeit von Belys *Simvolizm* befindliche Text“ faktisch „nicht der Rede wert“ (Bobrov 1916, 75) sei, da er nicht über das dort Dargestellte hinausgehe – hingegen seien die Beiträge von Luk'janov und Čudovskij immerhin darum bemüht, etwas Neues zu bieten, auch wenn sie im Prinzip nach dem Modell von Belyj arbeiteten.

Sergej Pavlovič Bobrov (1889–1971) trat wiederholt ebenso als Poet wie auch als Literaturtheoretiker in Erscheinung. Nachdem er eine Zeit lang Mitglied des sog. Rhythmischen Kreises von A. Belyj gewesen war, sagte Bobrov sich recht bald von dem zunächst von ihm bewunderten Belyj los und

wurde selbst Vordenker und führender Kopf einer bedeutenden literarischen Gruppe, nämlich der futuristisch ausgerichteten *Centrifuga* (aus der u. a. auch Boris Pasternak hervorging).

Neben seinen literarischen und literaturtheoretischen Beschäftigungen wurde Bobrov später ein anerkannter Spezialist für ökonomische Statistik, der in den 50er Jahren u. a. populäre Mathematiklehrbücher für Kinder schrieb, so z. B. *Das Zauberzweihorn* [Volšebnyj dvurog, 1949] und *Der Archimedische Sommer* [Archimedovo leto, 1959–62]. Bobrov war interessanterweise eine derjenigen Personen, die fast ein halbes Jahrhundert später, in den 60er Jahren, abermals mit Untersuchungen zur quantitativen Literaturwissenschaft prominent in Erscheinung traten (Bobrov 1964a; 1964b; 1965; 1967). So stellte er z. B. 1964 in der Zeitschrift *Teorija verojatnostej i ee primenenie* [Wahrscheinlichkeitstheorie und ihre Anwendungen] eine quantifizierende Versanalyse vor, in der er u. a. auf seine eigenen frühen Arbeiten (Bobrov 1915) ebenso wie auf die von B. V. Tomaševskij (1916a) verwies, mit dem er seit 1916 in engem Briefkontakt gestanden hatte (s. u.). Einen anderen Aufsatz leitete Bobrov (1965) in der Zeitschrift *Russkaja literatura* mit den programmatischen Worten ein: „Jeder Literaturwissenschaftler muss in unserer Zeit die Verdienste und Errungenschaften der statistischen Methode verstehen und anerkennen.“ (Bobrov 1965, 109). In diesem Text verwies er auch hier ein weiteres Mal auf Tomaševskij, der als Erster wahrscheinlichkeitstheoretische Verfahren in die Versforschung eingeführt habe.

Bobrov war seit 1916 in engem Kontakt und Briefwechsel mit *Boris V. Tomaševskij* (1890–1957), einem der wichtigsten Literaturtheoretiker des Russischen Formalismus, der später Leiter des Puškin-Instituts der Akademie der Wissenschaften der UdSSR wurde. Die Arbeiten Tomaševskijs nehmen im Kontext des Russischen Formalismus, vor allem im Hinblick auf die Anwendung mathematisch-statistischer Verfahren auf Sprache und Literatur, einen besonderen Stellenwert ein und vermitteln einen lebendigen Eindruck vom Status mathematischer Ansätze insbesondere in der Literaturwissenschaft der 10er und 20er Jahre.

Da Tomaševskij nach Abschluss des Gymnasiums aufgrund politischer Aktivitäten der Zugang zu den höheren Bildungsinstitutionen in Russland versperrt war, absolvierte er in Belgien (Liège) ein Studium der Elektrotechnik (1908–1912), in dem er sich grundlegende mathematische Kenntnisse eignete. Wie aus im Jahre 1990 publizierten Erinnerungen hervorgeht, ging es ihm schon

in seinen ersten wissenschaftlichen Arbeiten (ab 1915) „um die Anwendung von Methoden der mathematischen Statistik und der Wahrscheinlichkeitstheorie auf konkretes Material“ (Bobrov 1916, 143). Hierin deckten sich die Interessen von Tomaševskij und Bobrov, in dessen Tagebuchaufzeichnungen sich in ähnlicher Weise wie bei Tomaševskij bereits im Jahre 1912 die Absicht formuliert findet, „höhere Mathematik zu studieren, und mit ihr Fakten des Rhythmus zu bearbeiten“ (Bobrov 1916, 143) – in der Tat sollte Bobrov 1919 ein Studium der mathematischen Statistik aufnehmen.

Tomaševskij legte noch in den 10er Jahren eine Reihe statistischer Arbeiten zur Elektrotechnik mit praktischer Ausrichtung vor, so etwa 1918 eine „Statistik der Beschädigungen von Telegraphenleitungen“ (Tomaševskij 1918a) oder 1919 eine „Statistik der Post- und Telegraphenämter“ (Tomaševskij 1919a), die im Zusammenhang mit den „neuen Aufgaben des staatlichen Aufbaus“ erarbeitet wurden. Obwohl er in den 30er Jahren an einem Leningrader Schiffsbau-Institut höhere Mathematik lehrte, ist doch gerade der Bereich der poetischen Sprache als dominanter Anwendungsbereich seiner statistischen Arbeiten anzusehen. Eine der ersten Arbeiten, in denen Tomaševskij (1916a) dieses Programm in Angriff nahm, war seine Rezension zu dem Buch *Novoe o stichosloženii A. S. Puškina* (Moskva 1915) von Bobrov, dessen Untersuchung eine kritische Auseinandersetzung mit Analysen Brjusovs zur Verstechnik Puškins enthielt. In der Folge entwickelte sich zwischen Bobrov, der 1916 auch eine statistische Untersuchung zur Vorkommenshäufigkeit verschiedener Konsonantentypen in poetischen Texten vorgelegt hatte (vgl. Bobrov 1916), und Tomaševskij ein Briefwechsel. In diesem kam insbesondere die beiden gemeinsame Ablehnung des Herangehens von Belyj zum Ausdruck. So schrieb Tomaševskij 1916 in einem Brief an Bobrov: „Ich muss bei Belyj das völlige Fehlen von mathematischem Takt feststellen. Auch wenn er, vielleicht, über einen ausreichenden Vorrat an mathematischen Kenntnissen verfügt, so hat Belyj doch deutlich eine Unfähigkeit an den Tag gelegt, mathematische Untersuchungsverfahren zu beherrschen“ (vgl. Tomaševskij 1916b, 142 f.).

Wie Tomaševskij (1919b) in einem 1919 im MLK gehaltenen und 1923 erstmals publizierten Vortrag bemerkte, machte er

durchaus keinen Hehl daraus, dass er die prinzipielle Fruchtbarkeit von Belyjs Vorgehen „in der Anwendung der statistischen und deskriptiven Methode, im Unterschied zur nomenklatorischen und normativen Methode anderer Forscher“ (Tomaševskij 1919b, 139) schätzte; auch sah Tomaševskij Belyj objektiv als Begründer einer statistischen Theorie des russischen Verses an: „[...] тем не менее создал статистическую теорию русского стиха“ (Tomaševskij 1924, 324). Das versperrte ihm allerdings nicht den Blick, dessen Ansatz äußerst kritisch zu beurteilen. Was Tomaševskij an Belyj insbesondere kritisierte, war die seiner Meinung nach „voreilige Bewertung individueller Erscheinungen [...], welche die Strenge der Untersuchung zerstört“ (Tomaševskij 1919b, 139).

In einem 1922 als Vortrag gehaltenen Beitrag bezog Tomaševskij (1923a, 35) seine Auffassung, die statistische Methode werde ad absurdum geführt, wenn ungeübte Philologen sie dazu verwendeten, von ihnen berechnete Koeffizienten verschiedener Art wie Bewertungsnoten zu handhaben, explizit auf Belyj. In diesem Falle würden statistische Verfahren nur mit dem Ziel eingesetzt, evaluative Koeffizienten zu erhalten, damit ein Qualitätsurteil über den Untersuchungsgegenstand getroffen werden könne – so etwa ließe sich eine vermeintliche Überlegenheit von Puškin über andere Dichter feiern, weil in dessen Texten irgendein Koeffizient 20 und nicht etwa nur 5 betrage.

Es war insbesondere diese Vermengung quantitativer Verfahren mit dem Anspruch, aus ihnen unmittelbare qualitative Schlussfolgerungen zu ziehen, die Tomaševskij nachhaltig kritisierte: „Man darf nicht vergessen, dass man auch bei einer korrekt durchgeführten Berechnung als Ergebnis eine Zahl erhält, die lediglich die Häufigkeit einer Erscheinung charakterisiert, aber mitnichten ihre Qualität. Aber auch der Umgang mit einer solchen nackten Zahl verlangt, wenn man sie nicht nur als groben Index bezüglich der Vorkommenshäufigkeit oder -seltenheit verwendet, große Sorgfalt. Das Wichtigste ist es, ihren Grad an Exaktheit zu bestimmen. Hier überschreiten wir die Grenze der elementaren Aufgabenstellung der Berechnung und gehen zu speziell statistischen Fragen über (...)“ (Tomaševskij 1923a, 36).

Doch Tomaševskijs Skepsis richtete sich keineswegs nur gegen Belyj. Ein anderes

Objekt seiner Kritik waren die Arbeiten von *Georgij A. Šengeli* (1894–1956), die trotz aller Einwände, die sich ihnen gegenüber vorbringen lassen, Ecksteine der russischen Verstheorie darstellen: Sein zunächst 1921 erschienener *Traktat o russkom stiche* kam 1923 in zweiter, erweiterter Auflage heraus (Šengeli 1923), und auch sein später entstandenes Buch *Technika sticha* (Šengeli 1940) – welches vom Gegenstand her ausschließlich verstheoretisch ausgerichtet ist und daher im hier gegebenen Kontext nicht ausführlicher besprochen werden muss – erlebte 1960 eine zweite Auflage.

In seiner ersten verstheoretischen Arbeit, der 1916 fertig gestellten, 1918 eigenständig erschienenen Studie *Dva Pamjatnika* (Šengeli 1918), untersuchte Šengeli zwei gleich lautende Gedichte mit dem Titel „Pamjatnik“: das fünfstrophige Gedicht von Puškin sowie das sechsstrophige von Brjusov. Beide Texte weisen jeweils vierzeilige Strophen mit jambischem Metrum und Paarreim auf. Neben verschiedenen inhaltlichen stellte Šengeli auch eine Reihe quantitativer Vergleiche an: So berechnete er u. a. für die einzelnen Strophen und Texte die Anzahl der Worte und die Wortlängen in der Anzahl der Silben pro Wort. Auch die Häufigkeit der Konsonanten und Vokale berechnete er für beide Texte: Demnach kommen in dem Puškin-Text bei 306 Konsonanten und 230 Vokalen auf einen Vokal im Durchschnitt 1,33 Konsonanten, in dem Brjusov-Text bei 397 Konsonanten und 276 Vokalen sind es 1,44. Weiterhin berechnete Šengeli die mittlere Wortlänge in der Anzahl der Buchstaben pro Wort – dividiert man nun für beide Texte jeweils die Gesamtzahl der Worte durch die Summe der Buchstaben, so ergibt sich für den Puškin-Text bei 117 Wörtern und 536 Buchstaben eine mittlere Wortlänge von 4,58, für den Brjusov-Text bei 137 Wörtern und 673 Buchstaben eine mittlere Länge von 4,91 Buchstaben pro Wort. Aufgrund dieser Werte hat Šengeli – davon ausgehend, dass sich kürzere Wörter und Wörter mit einer (wie er meinte) günstigeren Konsonanten-Vokale-Relation leichter aussprechen lassen – so etwas wie eine *Transparenz* beider Texte berechnet: Dividiert man zunächst die Summe der Buchstaben des Brjusov-Textes durch die des Puškin-Textes ($673:536 = 1,26$), dividiert sodann die mittlere Wortlänge des Brjusov-Textes durch diejenige des Puškin-Textes ($4,91:4,58 = 1,07$), und multipliziert schließ-

lich die beiden Quotienten miteinander, so ergibt sich ein Wert von 1,35, den Šengeli (1918, 11) als Indiz einer um den Faktor 1,35 größeren Leseschwierigkeit interpretiert.

Als wesentliches Ergebnis seiner Formanalysen kam Šengeli (1918) zu der Schlussfolgerung, dass sich in fast allen Bereichen herausgestellt habe, dass Brjusov über „меньшее мастерство распрояжения словом“ [weniger Meisterschaft im Umgang mit dem Wort] verfüge als Puškin: „Der Rhythmus ist ärmer, die Wörter sind schwerer, die Armut an Alliterationen ist größer, die Menge von für das Gehör unangenehmen Lauten ist größer, Stauungen, Kakophonie“ (Šengeli 1918, 13).

Tomaševskij empfand das Vorgehen Šengelis als einen „zum Himmel schreienden Verstoß gegen das Grundgesetz der induktiven Statistik, das Gesetz der großen Zahlen“; die statistischen Untersuchungen seien aus diesem Grunde zwar „reich an Material“, aber „arm an Rückschlüssen“ und brächten „keinerlei unmittelbare Ergebnisse“ (Tomaševskij 1918b, 104 f.). Vor allem aber sahen Kritiker wie Bobrov oder Tomaševskij in solchen auf den Berechnungen fußenden Bewertungen eine nicht zulässige Vermengung von quantitativem und qualitativem Herangehen. Ähnlich kritisch äußerten sich in dieser Hinsicht auch andere Forscher. In dieselbe Richtung ging z. B. auch die Kritik von *Arkadij G. Gornfel'd* (1867–1941), einem Vertreter der sog. Charkover Linguistischen Schule, die ihre Ursprünge bei Sreznevskij hatte und dann vor allem von *A. A. Potebnja*, dem Lehrer Gornfel'ds, geprägt wurde. Gornfel'd (1922, 164) kommentierte die wesentliche Schlussfolgerung Šengelis über die geringere Meisterschaft Brjusovs wie folgt: „Es ist vollkommen klar, dass hier die Objektivität der Statistik durch reinste Subjektivität überdeckt ist“.

Gornfel'd (1922) ging in seiner Kritik noch weiter; für ihn zeigten die von Šengeli aufgrund seiner statistischen Untersuchungen erhaltenen Schlussfolgerungen vor allem, „dass die Statistik dort schädlich ist, wo sie übel angewendet wird“ (Gornfel'd 1922, 164), was ihn zu der zusammenfassenden (nicht mehr nur auf Šengeli) bezogenen Einschätzung führte: „Unterdessen ist eine gewisse Kenntnis der Literatur ausreichend, um zu sehen, dass bei unserer neuen Zusammenarbeit von Poetik und Statistik die Statistik sehr gelitten hat, und von daher hat bislang auch die Poetik ein wenig verloren“ (Gornfel'd 1922, 164).

Diese verallgemeinernde Kritik an den Studien von Šengeli gilt es allerdings in mehrerer Hinsicht zu relativieren: Zum einen hatte Šengeli seiner Studie die Bemerkung vorausgeschickt, dass diese 1916 abgeschlossen worden sei, und dass sich seither die theoretischen Grundlagen erheblich geändert, nämlich kompliziert hätten; zum zweiten fand die Studie von 1918 nicht, wie zunächst geplant, Eingang in Šengelis 1921 erschienene umfangreichere Monographie *Traktat o russkom stiche*, die zudem Gornfel'd bei seiner Pauschalverurteilung nicht vorlag. Und abgesehen davon finden sich im *Traktat* praktisch keinerlei wertende Schlussfolgerungen mehr – was freilich nicht bedeutet, dass das Buch nicht aus anderen Gründen Kritik hervorrufen sollte.

Zu Beginn des *Traktats* erklärt Šengeli (1923, 15) ohne Umschweife, dass mit den Arbeiten von Belyj eine neue Periode der russischen Verswissenschaft begonnen habe. Auch wenn dessen Arbeiten in keiner geschlossenen Theorie gemündet seien, hätten sie die Grundlage für eine vollkommen neue Methode, die statistische Methode, gelegt. Die von Belyj initiierte Suche nach „immanenten Gesetzen“ des Verses sei allerdings nur „durch die Anhäufung einer großen Anzahl statistischer Daten“ möglich. Šengeli (1923, 15) folgt Belyj keineswegs blind, der ohne Frage der Vater der Verswissenschaft [стиховедение] als einer Wissenschaft [наука] sei, selbst wenn bei ihm eine ganze Reihe von Schlussfolgerungen unbegründet bleibt, und auch wenn es bei ihm in der Datenerhebung und mathematischen Anwendung methodologische Fehler gebe.

In seinem *Traktat* konzentrierte sich Šengeli großteils auf die schon von Čudovskij angesprochene Frage nach dem Anteil betonter und unbetonter Silben in der russischen Sprache. Šengeli weist eingangs auf das Ergebnis von dessen Studie hin, der zufolge der Anteil betonter Silben bei etwa 35 % liege. Dies decke sich in etwa mit Daten von Kagarov, die dieser in einer nicht publizierten, 1918 in Charkov als Vortrag gehaltenen Studie vorgestellt, dabei aber auf Unterschiede in künstlerischer, wissenschaftlicher und journalistischer Prosa aufmerksam gemacht habe. Kagarovs Studie wurde vier Jahre später, also 1922 und somit nach dem Erscheinen der ersten, jedoch vor dem Erscheinen der zweiten Auflage von Šengelis *Traktat*, veröffentlicht (vgl. Kagarov 1922).

Šengeli selbst hat vor diesem Hintergrund umfangreiche Analysen durchgeführt; dabei kam er bei seiner Analyse von 135.983 Wörtern auf 50.000 betonte Silben, was einem Anteil von 36,77 % entspricht – mit anderen Worten: auf 1,72 unbetonte Silben kommt eine betonte. Den Umstand, dass Čudovskij im Vergleich dazu auf 1,86 unbetonte Silben gekommen war, erklärt Šengeli mit der unterschiedlichen Zählweise von Hilfswörtern (wie er sie bezeichnete): Während Čudovskij in dem Puškin'schen Vers *Когда не в шутку занемор* nur auf zwei Betonungen gekommen sei, rechnet Šengeli hier (korrekterweise) drei.

Šengeli lässt es jedoch nicht bei diesen simplen Berechnungen bewenden. Vielmehr analysiert er des Weiteren im Detail – unter Berücksichtigung verschiedener Textsorten – wie sich die Betonungen auf Wörter mit verschiedener Silbenanzahl verteilen, und an welchen Positionen innerhalb dieser Wörter die Betonungen vorkommen. Damit gelangt Šengeli sozusagen en passant auch erstmals zu einer umfangreichen Häufigkeitsverteilung eines russischen Textkorpus, auf die er jedoch nicht näher eingeht. Im Hinblick auf die für ihn im Vordergrund stehende Frage stellt Šengeli in einer Analyse von zehn Ausschnitten aus Prosatexten verschiedener russischer Schriftsteller des 19. und 20. Jh.s fest, dass bei Wörtern mit ungerader Silbenzahl eine Tendenz besteht, dass die Betonung genau auf der Wortmitte liegt, bei Wörtern mit gerader Silbenzahl auf dem mittleren Silbenpaar (Šengeli 1923, 21). Diese Regularität interpretiert Šengeli als „наличие в языке естественных слогоударных констант“ [Existenz natürlicher silbenbetonender Konstanten in der Sprache] (Šengeli 1923, 22). Deren Einfluss auf die Formgebung sei jedoch nicht unbegrenzt; vielmehr gebe es auch „автономные законы ритма“ [autonome Gesetze des Rhythmus] (Šengeli 1923, 22), so dass die Verskonstruktion in der Praxis dem Einfluss sowohl sprachlicher als auch rein rhythmischer Normen unterliege. Den Nachweis dieser Hypothese erbringt Šengeli durch analog angelegte Untersuchungen zum Vorkommen der Betonungen in Texten verschiedener Versmaße. So stellt er in Stichproben à 3.000 Wörter aus in Hexametern geschriebenen Texten von vier verschiedenen Autoren fest, dass die Ergebnisse sich von denen der Prosatexte nur unwesentlich unterscheiden. Vollkommen andere Bilder ergeben sich je-

doch aus den Analysen von Texten mit vier- oder fünffüßigem Jambus: Während in den Texten mit vierfüßigem Jambus insbesondere einsilbige und trochäische Wörter abnehmen, fehlen in den Texten mit fünffüßigem Jambus fast vollständig drei- und viersilbige Wörter mit Betonung auf der ersten Silbe. Diese Beobachtung interpretiert Šengeli so, dass sich einerseits einige Metren im Hinblick auf den Wortgebrauch den natürlichen Konstanten annähern und andere (interessanterweise häufiger verwendete) Metren davon entfernen, dass andererseits die natürlichen Konstanten auf die Gebräuchlichkeit der jeweiligen Formen einwirkten. Dies sei insgesamt so zu verstehen, „dass sich im Vers ein rein sprachliches und ein rein rhythmisches Element in ständiger Wechselbeziehung befinden“ (Šengeli 1923, 28).

Im Anschluss an weitere Detailanalysen zu einzelnen Versfüßen und weiterführende Überlegungen zur Fundierung der Strophe als einer wesentlichen Texteinheit überführt Šengeli (1923, 177 ff.) die dargestellten Überlegungen in die Berechnung eines von ihm so genannten Vorkommenskoeffizienten: Da das Vorkommen eines bestimmten rhythmischen Komplexes dem Einfluss zweier Faktoren unterliege (den sprachlichen und den rhythmischen Normen), lasse sich die Wahrscheinlichkeit dieses Vorkommens unter Rückgriff auf die Vorkommenshäufigkeit der dem Komplex zugrunde liegenden Komponenten bestimmen. Während sich allerdings üblicherweise die gemeinsame Vorkommenswahrscheinlichkeit zweier Elemente, von denen das eine die Wahrscheinlichkeit m/n , das andere die Wahrscheinlichkeit p/q hat, durch Multiplikation der einzelnen Wahrscheinlichkeiten ergibt, lässt sich der Vorkommenskoeffizient im gegebenen Fall nicht als mp/nq berechnen; vielmehr sei die Wahrscheinlichkeit des Vorkommens einer Verbindung von zwei oder mehr Elementen deswegen höher, weil ein bestimmtes metrisches Schema realisiert werden müsse, so dass eine Wahl vorliege. Deswegen sei die gemeinsame Vorkommenswahrscheinlichkeit zweier oder mehrerer Elemente durch die Berechnung des geometrischen Mittels der Koeffizienten der einzelnen Komponenten zu bestimmen. Habe man zum Beispiel ein Schema wie: $U - |U - |UUU - |$ und $U - U| - |UUU - |$, und seien die Vorkommenshäufigkeiten der Komponenten, die in dieses Schema Eingang finden,

0,165, 0,165, 0,110 sowie 0,135, 0,155, 0,0100, so berechne sich die Wahrscheinlichkeit des Vorkommens der beiden ersten Paare als $\sqrt{0,165 \cdot 0,165}$ bzw. $\sqrt{0,135 \cdot 0,155}$; man erhält somit ein Verhältnis von 33:29, d. h. von 1,14:1. De facto habe sich, so Šengeli (1923, 178), in seinen empirischen Untersuchungen bei den vierfüßigen Jamben ein Verhältnis von 304:236, d. h. von 1,29:1 ergeben (was er als recht gute Annäherung an die Erwartung interpretiert).

Überaus harsch äußerte sich Bobrov (1921) zu Šengelís *Traktat*, der seiner Meinung nach „mit Berechnungen operierte, denen er die laute Bezeichnung ‚statistisch‘ zueignete, ohne auch nur die geringste Vorstellung von der statistischen Methode zu haben“ (Bobrov 1921, 268). Weiterhin betonte Bobrov (1921), Šengeli sei „in der Statistik nicht über das geometrische Mittel hinausgekommen, das er mit unklaren Zielen und ebenso nebelhaften Resultaten anwende“ (Bobrov 1921, 268). Das einzig gute Haar, das Bobrov in seinem Verriss an der Arbeit von Šengeli ließ, war der Umstand, dass „möglicherweise ein gebildeter Mensch irgendetwas aus dessen Tabellen nicht ohne interessante Resultate verwenden könnte“ (Bobrov 1921, 268).

Auch Tomaševskij forderte in seiner Rezension des *Traktats* nachdrücklich „ostorožnost’ i kritičizma“ [Vorsicht und kritische Haltung] (Tomaševskij 1923b, 52) bei der Anwendung statistischer Verfahren – genau das sah er bei Šengeli ganz offensichtlich nicht gewährleistet: „Šengeli wendet auf breiter Ebene die statistische Methode an. Diese Methode, die in philologischen Kreisen auf Widerstand stößt, hat das Ziel, Fakten und Beziehungen zu erheben, aber sie liefert verständlicherweise niemals ‚Erklärungen‘. Wenn man sich ihr zuwenden will, so erfordert das eine vorsichtige und kritische Haltung, die nur solche Personen haben, die der mathematischen Kultur nahe sind. Bei dem Autor ist von dieser Vertrautheit mit der Mathematik nichts zu bemerken, und seine ‚Beweise‘ der ‚Koeffizienten von Vorkommensformen‘ sind ein bedauerliches Missverständnis.“ (Tomaševskij 1923b, 52).

In der Tat sind solche Koeffizienten oder Indizes so lange unbrauchbar, wie sie nicht standardisiert sind – so lange man nicht Erwartungswert und Varianz solcher Indizes (wie sie auch heute noch in verschiedenen Bereichen verwendet werden) kennt, sind

sie schlicht und einfach nicht zuverlässig interpretierbar. Vermutlich würde man heute die Vorkommenshäufigkeit benachbarter Ereignisse eher in Form von Übergangswahrscheinlichkeiten bestimmen, ein Verfahren, das ja gerade im damaligen Kontext von A. A. Markov entwickelt und an poetischem Material demonstriert worden war, und auf dessen Bedeutung schon Tomaševskij (1923a, 36) nachdrücklich hingewiesen hatte.

Andrej A. Markov (1856–1922), seit 1886 Professor an der Universität Petersburg und seit 1896 Mitglied der Petersburger Akademie der Wissenschaften, hatte die diesbezüglichen Überlegungen im Wesentlichen in zwei seiner Arbeiten dargelegt (Markov 1913; 1916). In der ersten Studie legte Markov (1913) eine statistische Untersuchung an Puškins Versepos *Evgenij Onegin* vor. Über erste Ergebnisse dieser Untersuchung hatte er schon im Januar 1913 in einem Brief an den bedeutenden Mathematiker *Aleksandr A. Čuprov* (1874–1922) berichtet, mit dem er seit 1910 in intensivem Briefkontakt stand, und der ebenfalls in Petersburg lebte und arbeitete (vgl. Ondar 1981). Aus dem genannten Brief vom 15. 1. 1913 geht hervor, dass Markov seine Untersuchung im Kontext von Arbeiten zum Gesetz der großen Zahlen durchführte, das Bernoulli ja genau 200 Jahre zuvor in seiner *Ars Coniectandi* aufgestellt hatte. Während die Gültigkeit dieses Gesetzes in der bis dahin akzeptierten Form allerdings auf unabhängige Variablen beschränkt gewesen war, lag es im Interesse Markovs, eine Erweiterung auch auf abhängige Variablen vorzunehmen.

Markov (1913) analysierte – unter Auslassung der Leerzeichen sowie der weichen und harten Zeichen (also ь und ъ) – die ersten 20.000 Buchstaben von Puškins *Evgenij Onegin*. Was ihn interessierte, war das Aufeinanderfolgen von Vokalen und Konsonanten (wobei er das ѝ allerdings als Vokal zählte). Markov untergliederte den von ihm untersuchten Textausschnitt in 200 Blöcke à 100 Buchstaben und ordnete diese zeilenweise fortlaufend in einer Matrix von 10×10 Buchstaben an.

Die auf diese Art erhaltenen 200 Blöcke betrachtete er als unabhängige Variablen, wobei jede Variable als das Ergebnis von 100 ab-

1	2	3	4	5	6	7	8	9	10
м	о	й	д	я	д	я	с	а	м
ы	х	ч	е	с	т	н	ы	х	п
р	а	в	и	л	к	о	г	д	а
н	е	в	ш	у	т	к	у	з	а
н	е	м	о	г	о	н	у	в	а
ж	а	т	с	е	б	я	з	а	с
т	а	в	и	л	н	л	у	ч	ш
е	в	ы	д	у	м	а	т	н	е
м	о	г	е	г	о	п	р	и	м
е	р	д	р	у	г	и	м	н	а
3	7	2	5	5	3	5	4	3	5

hängigen Versuchen anzusehen sei. Im nächsten Schritt addierte Markov die Anzahl der Vokale, zunächst in den einzelnen senkrechten Spalten, dann in den Spaltenpaaren $1 + 6$, $2 + 7$, $3 + 8$, $4 + 9$, $5 + 10$. Im obigen Beispiel (dem Textanfang des *Evgenij Onegin*) ergibt sich so $6 + 12 + 6 + 8 + 10 = 42$. Die Gesamtsumme aller Vokale beläuft sich auf 8638, so dass das arithmetische Mittel über alle 200 Blöcke 43,2 beträgt – die Wahrscheinlichkeit, dass ein Buchstabe ein Vokal ist, macht somit $p = 0,432$ aus. Die Summe aller Konsonanten beträgt entsprechend $20000 - 8638 = 11362$, was einer relativen Häufigkeit von $q = p - 1 = 0,568$ entspricht.

Die Wahrscheinlichkeiten für Sequenzen (Ketten) von Vokalen (V) bzw. Konsonanten (C) bezeichnete Markov des weiteren wie folgt.

Da die Summe der quadratischen Abweichungen insgesamt 1022,78 beträgt, beläuft sich das von Markov als *Dispersionskoeffizient* bezeichnete Abweichungsmaß auf $1022,78/20000 = 0,5114$. Außerdem berechnete Markov mit den relativen Häufigkeiten der Vokal- und Konsonantenvorkommnisse einen von ihm als *Dispersionsindex* bezeichneten Wert von $0,051/[0,432 \cdot (1 - 0,432)] = 0,208$.

Im Weiteren ging es Markov um die Art der Aufeinanderfolge von Konsonanten und Vokalen. Dazu stellte er folgende Überlegung an: Die Anzahl von Vokalen, die auf einen Vokal folgen beträgt 1104; dividiert man nun die Anzahl der Vokale, die auf einen Vokal folgen, durch die Summe aller Vokale, so erhält man $p_1 = 1104/8638 =$

p	$q = 1 - p$	p_1	p_0	$p_{1,1}$	$p_{1,0}$	$p_{0,1}$	$p_{0,0}$
V	C	V-V	C-V	V-V-V	V-C-V	C-V-V	C-C-V

0,128. Entsprechend berechnet sich die Anzahl der Konsonanten, die auf einen Vokal folgen, als $8638 - 1104 = 7534$. Dividiert man diese Anzahl durch die um 1 reduzierte Anzahl aller Konsonanten, erhält man $p_0 = 7534 / (20000 - 1 - 11362) = 0,663$.

Die Differenz der beiden erhaltenen Quotienten bezeichnet Markov als δ , so dass $\delta = 0,128 - 0,663 = -0,535$. Berechnet man mit dem Wert den theoretischen Dispersionskoeffizienten, so ergibt sich

$$\frac{1 + \delta}{1 - \delta} = \frac{0,465}{1,535} = 0,303$$

Markovs Schlussfolgerung aus dieser Berechnung ist, dass dieser Erwartungswert zwar nicht mit dem beobachteten von 0,208 übereinstimmt, diesem aber dennoch deutlich näher ist als dem Wert 1, der vollständige Unabhängigkeit der Stichproben bedeuten würde.

Ausgehend davon, dass sich die Vermutung der Abhängigkeit noch besser nachweisen lässt, wenn man nicht von einfachen, sondern von komplexen Ketten ausgeht (d. h. nicht nur Sequenzen aus zwei, sondern aus drei und mehr Buchstaben in Betracht zieht), berechnete Markov die Anzahl von drei aufeinander folgenden Vokalen bzw. Konsonanten. Im ersten Fall beläuft sich die Summe auf 115, im zweiten auf 505. Damit ergeben sich die Wahrscheinlichkeiten für $p_{1,1} = 115/1104 = 0,104$ bzw. für $q_{0,0} = 505/3827 = 0,132$.

p	$q = \frac{1-p}{1-p}$	p_1	p_0	$p_{1,1}$	$p_{0,0}$
0,432	0,568	0,128	0,663	0,104	0,132

Mit diesen Werten lassen sich in Analogie zur Berechnung von δ aufgrund der entsprechenden Differenzen ε und η berechnen, als deren Werte man

$$\varepsilon = -0,024/0,872 = -0,027 \text{ und} \\ \eta = -0,025/0,663 = -0,309 \text{ erhält.}$$

Setzt man diese Werte zur Berechnung des theoretischen Dispersionskoeffizienten in die Formel

$$\frac{\{q(1-3\varepsilon)(1-\eta) + p(1-3\eta)(1-\varepsilon) - 2(1-\varepsilon)(1-\eta)\}(1-\delta) + 2(1-\varepsilon\eta)}{(1-\delta)(1-\varepsilon)(1-\eta)} \\ = \frac{1+\delta}{1-\delta} \left\{ \frac{1+\varepsilon}{2(1-\varepsilon)} + \frac{1+\eta}{2(1-\eta)} \right\} + \frac{(q-p)(\eta-\varepsilon)}{(1-\varepsilon)(1-\eta)}$$

ein, so erhält man einen Wert von 0,195, der in der Tat dem empirisch beobachteten von 0,208 recht nahe kommt. Markov (1913) führt in der Folge noch weitere Berechnungen durch, doch kann man allein aufgrund dieser Tatsache ihm zufolge kaum in Zweifel ziehen, „dass die beobachtete Übereinstimmung der Zahlen nicht zufällig ist“ (Markov 1913, 160).

Damit war Markov einerseits der Nachweis gelungen, dass das Gesetz der großen Zahlen auch für abhängige Variablen gültig ist; abgesehen davon hatte er allerdings einen nachhaltigen Riesenschritt im Hinblick auf die Anwendung mathematischer Verfahren auf sprachliches Material geleistet.

Der zweite Aufsatz von Markov (1916) war im Wesentlichen eine Reaktion auf eine Untersuchung von *Nikolaj A. Morozov* (1854–1946), ein ehemaliger, dem Terror nahe stehender Volkstümpler [narodnik], der nach mehrjähriger Haft zu verschiedensten Themen arbeitete (von Chemie, Physik, Mathematik über Astronomie und Astrologie bis hin zu Geschichte und Theologie), und der 1932 zum Ehrenmitglied der Akademie der Wissenschaften der UdSSR wurde. In Morozovs Aufsatz aus dem Jahre 1915 lag sein Hauptinteresse zwar auf der Bestimmung individual-stilistischer Merkmale, dies aber – seinen Aussagen zufolge schon seit Mitte der 80er Jahre des 19. Jh.s – vor dem Hintergrund „allgemeiner stilometrischer Gesetze“ (Morozov 1915, 97). Wenn, so seine Überlegung, in der Natur und im gewöhnlichen Leben die vielfältigsten, scheinbar zufälligen, Erscheinungen zu einem ziemlichen Maße „gesetzmäßigen Charakter“ haben – warum soll das nicht auf den Bereich der Sprache zutreffen? Morozov (1915, 97) war fest davon überzeugt, dass „in unserer menschlichen Sprache all deren Elemente eine bestimmte Proportion haben“, und dass „statistische Gesetzmäßigkeiten nicht nur in sich häufig wiederholenden Erscheinungen der Natur und des öffentlichen Lebens, sondern auch in den Erscheinungen unserer Umgangs- und Schriftsprache existieren“ (Morozov 1915, 110). In dieser Überzeugung führte Morozov als eine erste

Veranschaulichung eine Häufigkeitsliste russischer Buchstaben an, allerdings ohne Angabe, woher die entsprechenden Zahlen stammen bzw. worauf sie basieren. Dennoch ist dies vor Proskurnin (1933) die früheste Graphemstatistik des Russischen, wenn man von der Untersuchung von Ol'chin (1907) absieht, die aufgrund ihrer Zielsetzung nicht das gesamte Graphemsystem in Betracht zieht (vgl. Grzybek/Kelih 2003; eine theoretische Modellierung russischer Graphemhäufigkeiten allgemein findet sich bei Grzybek/Kelih/Altmann 2004).

In expliziter Analogie zur Technologie der Spektralanalyse schwebten Morozov als Ergebnis seiner linguistischen Analysen von Texten individueller Autoren sog. linguistische Spektren vor; darunter verstand er Graphiken, welche die autoren spezifische Stilistik aufgrund unterschiedlicher Vorkommenshäufigkeiten von linguistischen Einheiten wie ausgewählten Präpositionen, Konjunktionen u. a.m. veranschaulichen sollten. Als ein Beispiel stellt Abb. 3.2 einen Vergleich der Vorkommenshäufigkeit der drei Präpositionen *в*, *на* und *с* in den ersten 1000 Wörtern aus verschiedenen Texten von Gogol' und Turgenjev dar.

Markov (1916) leitete seine Kritik an der Untersuchung von Morozov mit der Bemerkung ein, dass ein entsprechendes Vorgehen im Prinzip von großer Bedeutung sei – allerdings „nur unter der Bedingung, dass die Konstanz der Schlussfolgerungen, mit anderen Worten: ihre Stabilität, nicht auf Glauben basiert, sondern sich in der Untersuchung selbst herausstellt, wobei auch das Maß der Schwankungen geklärt werden muss“ (Markov 1916, 239). Worauf Markov mit diesen Bemerkungen abzielt, ist zweierlei: auf der einen Seite fordert er die Berücksichtigung nicht nur absoluter oder relativer Häufigkeiten, sondern auch die Angabe dazugehöriger Streuungsmaße, auf der anderen bemängelt er den zu geringen Stichprobenumfang.

Um bei dem obigen Beispiel von *в* und *на* zu bleiben: In den ersten 1000 Wörtern von Puškins *Kapitanskaja dočka*, *Baryšnja-krest'janka* und *Dubrovskij* ist das Verhältnis der beiden Präpositionen zueinander nach Morozov 32:11, 40:12, und 46:13. Markovs ergänzende Analyse der ersten 1000 Wörter des 2. Kapitels von *Kapitanskaja dočka* ergibt allerdings ein ganz und

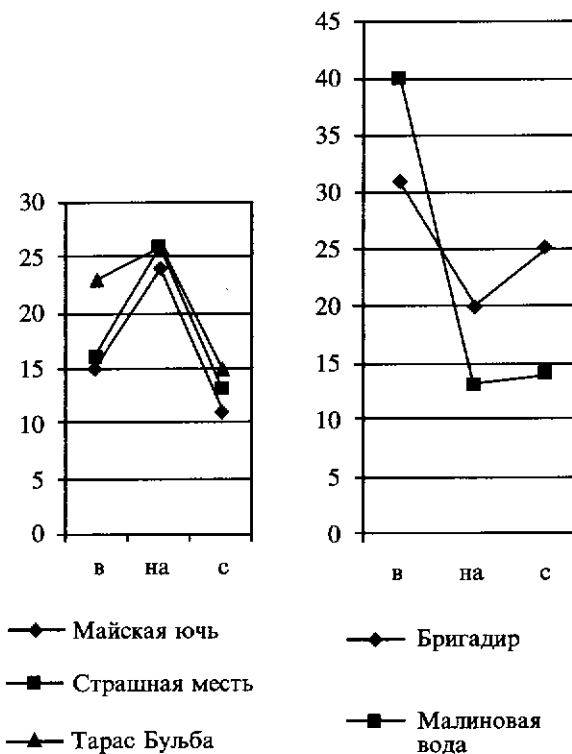


Abb. 3.2: Frequenz russischer Präpositionen nach Morozov (1915)

gar anderes Verhältnis von 15:21, welches gemäß den Analysen Morozovs eher dem Stil Gogol's entsprechen müsste, denn Morozov (1915, 111) ist das Verhältnis der beiden Präpositionen in *Majskaja noč'*, *Strašnaja mešt'* und *Taras Bulba* 15:24, 16:26, 23:20. Auch in diesem Fall führen Markovs Zusatzberechnungen allerdings dazu, Morozovs Schlussfolgerungen in Frage zu stellen, da seine Analyse der ersten 1000 Wörter aus dem 9. Kapitel der *Toten Seelen* mit einem Verhältnis von 37:12 de facto dem vermeintlichen Stil Puškins, nicht aber Gogol's entspricht. Markovs (1916, 241) Fazit lautete deshalb: „Nur eine bedeutende Erweiterung des Untersuchungsfeldes (nicht 5 Tausend, sondern hundert Tausende Wörter) kann den Schlussfolgerungen einen gewissen Grad an Grundlage verleihen.“

Das Postulat, linguo-statistische Untersuchungen jedweder Art insgesamt auf eine möglichst breite Basis zu stellen, war auch das Fazit des oben bereits erwähnten *Nikolaj A. Setnickij* (1888–1937). Setnickij war ein stark von den Religionsphilosophen *N. F. Fëdorov* und *V. S. Solov'ëv* beeinflusster Philosoph und Ökonom, der nach Abschluss seines Studiums an der juristischen Fakultät in Petersburg 1912 zunächst als Ökonom und Statistiker arbeitete, 1925 nach China emigrierte, und zwei Jahre nach seiner Rückkehr in die UdSSR 1935 verhaftet und 1937 erschossen wurde.

In seiner – zuvor auch in der Zeitschrift *Statističeskij Vestnik* abgedruckten – Studie *Statistik, Literatur und Poesie* geht Setnickij (1923, 18) davon aus, dass die Statistik „einer der Wege ist, das Wesen des Wortes zu erhellen und in außergewöhnliche Höhen zu erheben“. Vor dem Hintergrund der Annahme, seine Zeit sei durch eine „außergewöhnliche Entwicklung und Erweiterung der Anwendungsbreite der statistischen Methode“ (Setnickij 1923, 3) charakterisiert, wendet sich Setnickij gegen das Postulat, dass die Statistik primär „den Interessen des gegenwärtigen Moments entsprechen“ und „auf die Bedürfnisse des gegenwärtigen sozialen Aufbaus“ reagieren müsse – ein solches Postulat hatte nämlich kurz zuvor der Führer der staatlichen Statistik, *P. I. Popov*, auf der III. Statistischen Konferenz in Moskau (Januar 1921) vorgetragen. Gegen eine solche von ihm als pragmatizistisch bezeichnete Sichtweise wendet Setnickij (1923) ein, „dass die Statistik bzw. die statistische Methode an und für sich kein Werkzeug ist, das

ausschließlich dafür vorgesehen sei, praktische Ziele zu erreichen, die durch den heutigen Tag gesteckt würden“. Und weiter: „Wie in jeder Wissenschaft und im wissenschaftlichen Denken überhaupt, droht auch der Statistik durch das ausschließliche Stellen und Erfüllen von tagesaktuellen Aufgaben fast unausweichlich Blutleere und Schwächung des theoretischen Gedankenflugs, und in der weiteren Folge einer entsprechenden Wertminderung [...]“ (Setnickij 1923, 4). Die einzig sinnvolle Art und Weise, dem entgegenzuwirken, sei „die Konzentration der Aufmerksamkeit auf solche Bereiche und die Arbeit in solchen Bereichen, die nicht durch die offiziell berufene Charakteristik des Sozialen, Unaufschiebbareren und Unausweichlichen gekennzeichnet sind.“ In diesem Zusammenhang zählt Setnickij unter anderem explizit die Literatur im weiten Sinne des Wortes zu diesen Bereichen, „wo die Anwendung statistischer Methoden in der Gegenwart gereift ist und überaus fruchtbare Ergebnisse zeitigen kann“ (Setnickij 1923, 5).

Für die hauptsächlichen Objekte der statistischen Forschung im Bereich der Literatur hält Setnickij (1923, 6) den Autor und das Werk, die beide ihre statistischen Charakteristika haben; statistische Methoden versteht Setnickij dabei in breitem Sinne des Wortes. So meint er zum einen damit durchaus die Anlage statistischer Kartotheken zur Erfassung von autor- und werkbezogenen Daten. Zum anderen eröffnet sich für Setnickij (1923, 10) „eine Reihe neuer Arbeiten, wo ein speziell statistischer Zugang zu dem entsprechenden Material und dessen Erforschung mit Hilfe statistischer Methoden möglich ist“ – damit meint er die Erforschung sowohl von Poesie als auch von künstlerischer Prosa, deren Formen, Verfahren des Gebrauchs, grammatische und syntaktische Besonderheiten. Seiner Meinung nach ist „nur die in exakten Zahlen ausgedrückte Erforschung der Sprache und ihrer Tendenzen geeignet, eine Grundlage für seriöse und begründete Urteile in diesem Gebiet zu bereiten“ (Setnickij 1923, 11); allerdings muss er bedauernd einräumen, „dass die Theorie der künstlerischen Prosa noch relativ wenig ihre grundlegenden Elemente und Begriffe ausgearbeitet und geklärt habe“ (Setnickij 1923, 11). Wenn Letzteres allerdings der Fall wäre, „dann wäre die fruchtbare Grundlage für die Anwendung der statistischen Methode zur Erforschung

dieser Elemente und der mit ihnen verbundenen Ausdrucksverfahren der künstlerischen Prosa unzweifelhaft offenbar“ (Setnickij 1923, 11). Etwas besser sei die Lage im Bereich der Poesie, d. h. im Bereich der Theorie und Technik der Versfübung. Da es hier eine Reihe von Pionieren gebe, welche die statistische Methode auf Fragen der russischen Versfübung angewendet haben, sei auch bereits eine bedeutende Anzahl von herauskristallisierten Elementen bekannt, die als Grundlage für weitere Forschungen dienen können. Als Beispiele werden Metrum, Reim, Strophik genannt.

In diesem Kontext führt Setnickij (1922, 21–28) auch, ohne Anspruch auf Vollständigkeit, eine Reihe von Hinweisen auf entsprechende Vorarbeiten an. So erwähnt er u. a. Ja. Denisovs (1888) *Grundlagen der Metrik bei den Alten Griechen und Römern*. Die von Denisov (1888, 92; 107) angeführten Untersuchungen könnten zwar nicht im strengen Sinne des Wortes als eine Statistik der Versfübung bezeichnet werden; immerhin aber gehe es um Berechnungen der Vorkommenshäufigkeit verschiedener Versformen, die, auch wenn sie keinen eigenständigen Wert hätten, als Anstoß für weiterführende Untersuchungen dienen könnten.

Ebenfalls nennt Setnickij in diesem Zusammenhang *Fedor E. Korš* (1843–1915), den späteren Vorsitzenden der Moskauer Dialektologischen Kommission; dieser habe in seinen Untersuchungen zur Authentizität von Puškins *Rusalka* zwar keine eigenen Berechnungen angestellt, dennoch aber sei für die gesamte Arbeit „charakteristisch, dass sie unwillkürlich auf die Notwendigkeit und Vernünftigkeit aufmerksam mache, die gesamte Beweisführung nach dem Gesetz der großen Zahlen umzugestalten“ (Setnickij 1922, 22).

Während also Mathematiker bzw. Statistiker wie Markov oder Setnickij mit Recht auf das Problem der Stichprobengröße hinweisen und sich für die Orientierung am Gesetz der großen Zahlen aussprachen, sah ein mathematisch gebildeter Literaturwissenschaftler wie Tomaševskij dies differenzierter. Allerdings nahm Tomaševskij unter allen, die in den 10er und 20er Jahren auf irgendeine Art und Weise mathematisch-statistische Verfahren zur Analyse künstlerischer Texte zur Anwendung brachten, eine Sonderstellung ein. Und im Gegensatz zu den meisten anderen war es insbesondere Tomaševskij, der methodologisch verallgemeinernde Über-

legungen hinsichtlich der Anwendung solcher Verfahren anstellte.

So klassifizierte er in – im Grunde genommen heute noch – zutreffender Art und Weise die in der Analyse poetischer Texte angewandten Verfahren in zwei Grundprinzipien: „исследование на примерах и статистический метод“ [die Forschung anhand von Beispielen und die statistische Methode] (Tomaševskij 1923a, 34): Während ersterem Vorgehen zufolge eine bestimmte Erscheinung isoliert und dann an weiteren Beispielen verfolgt werde, werde bei der zweiten die Berechnung einer ganzen Reihe von Erscheinungen an mehr oder weniger umfangreichem, kompakten Material durchgeführt. Tomaševskij zufolge kann die isolierend-individualisierende Methode lediglich einen Weg für die vorbereitende Klassifikation von Erscheinungen darstellen (Tomaševskij 1923a, 34) – genau das war ja eines der Probleme, das auch Setnickij (1922, 11) angesprochen hatte, nämlich die operationale Bestimmung der Einheiten, die einer quantifizierenden Analyse unterworfen werden sollen. Die ausschließlich isolierend-individualisierende Methode aber sei – so Tomaševskij (1923a, 34) weiter – bei dem Versuch, Beziehungen zwischen den beschriebenen Erscheinungen herzustellen, zum Scheitern verurteilt, weil das nur an umfangreichem Material gelingen könne.

Insofern begrüßte Tomaševskij – bei aller Sympathie für mathematische Zugänge zu Sprache und Literatur – auch keineswegs jeglichen Versuch in dieser Richtung. So verurteilte er z. B. 1922 in einer Rezension von B. Ėjchenbaums *Melodika russkogo liričeskogo sticha* dessen „willkürliches Operieren mit mathematischen Relationen“ (vgl. Tomaševskij 1990, 143). Und in einer anderen Besprechung setzte Tomaševskij (1922) sich kritisch mit einem Versuch von A. A. Bobrik auseinander, in der damaligen Literaturwissenschaft verwendete Begriffe wie *Funktion*, *Relation*, *Vergleich* u. a. mit den entsprechenden mathematischen Termini gleichzusetzen und zu symbolisieren. Bobriks Abhandlung mit dem Titel *Nemnogo matematiki v teorii slovesnosti* (Bobrik 1921) kanzelte Tomaševskij schlichtweg als eine pathologische Erscheinung ab.

Im Gegensatz zu den oben dargestellten Versuchen, im Bereich der Poesie insbesondere Fragen von Rhythmus und Metrum mathematisch zu bearbeiten, geht auf Bobrik (1921) ein mitunter abstrus anmutender

Versuch zurück, zeitgenössische literaturwissenschaftliche Überlegungen zur Erzählforschung „in die Sprache der Mathematik zu übersetzen“ (Bobrik 1921, 4). Der in Gel'singfors publizierte Text steht mit Fragen der schon im 19. Jh. diskutierten, sog. historischen Poetik im Zusammenhang, wie sie von Potebnja und vor allem auch *Aleksandr N. Veselovskij* (1838–1906) entwickelt worden war. Insbesondere geht es dabei um die erzähltheoretischen Konzepte *Motiv* und *Sujet*, die später dann vor allem durch Vladimir Ja. Propps Untersuchungen an russischen Zaubermärchen präzisiert wurden, die als Vorläufer der strukturalistischen Narratologie angesehen werden können.

Im Kontext der genannten Arbeiten ist unter einem *Motiv* die einfachste narrative Einheit zu verstehen, deren Gesamtheit innerhalb eines Textes dessen *Sujet* verkörpert. Bobriks (1921) Abhandlung stellt nun einen Versuch dar, unter direkter Bezugnahme auf eine diesbezügliche Darstellung von Alfred L. Bem (1918/21) diese Überlegungen aus dem Bereich der Geschichte und Theorie der Literatur in die Sprache mathematischer und logischer Symbolik umzuformulieren.

Alfred L. Bem (1881–1945), dessen Text übrigens niemand anderem als Bobrik gewidmet war, war ein Schüler von Sreznevskij und Vengerov. Er emigrierte nach der Oktoberrevolution und lebte, nach mehrjährigem Aufenthalt in Warschau (1920–22), in Prag, wo sich seine Spuren verlieren, nachdem er 1945 unmittelbar nach Einmarsch der Roten Armee vom NKWD verhaftet wurde.

Die Annahme, dass sich das *Sujet* (*S*) als Summe der in einem Text vorkommenden *Motive* (m_k) verstehen lässt, erhält somit nach Bobrik (1921, 7) die folgende Form:

$$S = \sum_{k=1}^{k=n} m_k \quad (1)$$

Weiterhin greift Bobrik auch die Frage des von Bem diskutierten Verhältnisses zwischen dem Inhalt eines Werks und dessen *Sujet* (*S*) auf und folgt der Argumentation von Bem, der zufolge man bei einem einzelnen Werk (P_1), das nicht mit anderen Werken (P_2, P_3, \dots, P_n) in Relation steht, nur von dessen Inhalt (*I*) sprechen könne, während die Begriffe *Sujet* (*S*) und *Motiv* (*M*) nur als Ergebnis eines Vergleichs oder einer Verallgemeinerung mit anderen Werken anzusehen seien. Bobrik, der solche Begriffe wie *Vergleich*, *Relation*, *Verallgemeinerung* als

mathematische Begriffe identifiziert bzw. mit solchen gleichsetzt, formuliert entsprechend um: Demnach verhalte sich der Inhalt (*I*) zum *Sujet* (*S*) wie eine einzelne Vorstellung zu einer allgemeinen Vorstellung, was sich mathematisch als

$$P_1/P = 1/n \quad (2)$$

ausdrücken lasse.

Die ein *Sujet* konstituierenden *Motive* lassen sich nach Bobrik weiter differenzieren; dabei geht er von der Existenz eines Hauptmotivs (*iM*) und mehrerer diesem Hauptmotiv zu- bzw. untergeordneter Nebenmotivs verschiedener Ordnung (m_1, m_2, \dots, m_k) aus. Nimmt man etwa in Erzählungen wie Puškins oder Lermontovs *Kavkazskij Plennik* als Hauptmotiv (*iM*) „Liebe eines Mädchens zu einem Gefangenen“, so ergeben sich als mögliche *Motive* zweiter Ordnung (im_1) zwei Varianten: (a) auch der Gefangene liebt das Mädchen, (b) er liebt sie nicht. Als *Motive* dritter Ordnung (im_2) ergeben sich sodann, ausgehend von der Oberkategorie „Befreiung des Gefangenen“, die Optionen (a) es findet ein Befreiungsversuch statt, (b) es findet kein Befreiungsversuch statt. Auf der Basis einer solchen binären Erzähllogik ergibt sich somit bei einem gegebenen Hauptmotiv (*iM*) als Summe möglicher *Motive*

$$N_m = 1 + \sum_{k=1}^{k=n} 2^k, \quad (3)$$

was in entwickelter Form der Summe der Glieder einer geometrischen Reihe entspricht. Im Ergebnis erhält man unter diesen Voraussetzungen somit ein „Schema von Bifurkationen des Grundmotivs“ (Bobrik 1921, 18), welches in vielerlei Hinsicht spätere Konzepte der strukturalistischen Erzählforschung vorwegnimmt (vgl. Abb. 3.3).

Auf dieser Grundlage hat Bobrik (1921, 24 ff.) auch einen (konstruierten) Vergleich von Texten zweier Autoren vorgenommen. Gegeben sei das Schema A von Text A mit dem Schema $iS_1 = i(M + m_2^2 + m_3^3 + m_6^4)$ sowie Schema B von Text B mit dem Schema $iS_2 = i(M + m_4^4 + m_4^4 + m_6^3)$.

Dann lässt sich nach Bobrik zum Vergleich zweier Texte ein Koeffizient als Quotient aus der Anzahl aller in beiden Texten gemeinsam vorkommenden und der Anzahl aller vorkommenden *Motive* überhaupt berechnen; im obigen Beispiel wäre das konkret

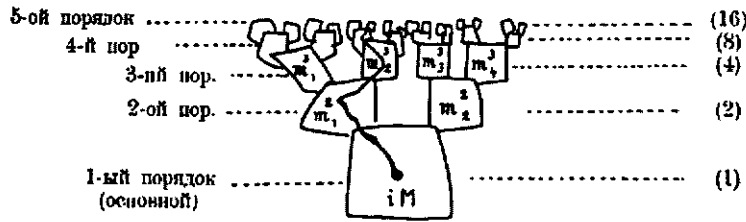


Abb. 3.3: Narrative Bifurkationen nach Bobrik (1921, 18)

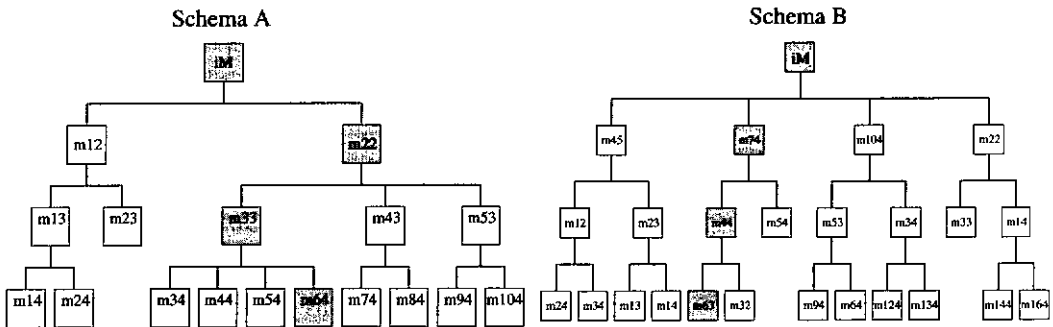


Abb. 3.4: Vergleichende Motivschemata nach Bobrik (1921, 24)

$$K = \frac{iM + iM}{2iM + m_2^2 + m_3^3 + m_6^4 + m_7^4 + m_4^4 + m_6^3} = \frac{2}{8} = 0,25.$$

Entsprechend hat Bobrik (1921, 25) auch einen Index zur Berechnung des Zusammenhangs zwischen den Schemata zweier Texte vorgeschlagen, der sich als Quotient aus der Anzahl der in beiden Schemata zusammenfallenden Motive und der Anzahl der insgesamt vorkommenden Motive berechnet, im gegebenen Fall also als

$$L = \frac{18 + 18}{18 + 25} \cdot \frac{36}{43} = 0,8372.$$

Der Versuch von Bobrik mutet mitunter recht kurios an – in der Tat muss ein solcher Versuch zum Scheitern verurteilt sein, solange die zur Disposition stehenden Motive bzw. die zur Beschreibung solcher Motive geeigneten semantischen Kategorien nicht in ein logisches kohärentes und konsistentes System gebracht werden können (was auch der strukturalistischen Erzählforschung in der zweiten Hälfte des 20. Jhd. nicht überzeugend gelang). Ungeachtet dessen ist Bobriks Intention und Art des Herangehens, eine literaturwissenschaftliche Meta-Sprache an der Exaktheit mathemati-

scher Konzepte auszurichten und zu messen, im Kontext der damaligen Zeit von großer Bedeutung, so dass die Abqualifikation von Tomaševskij (1922, 84), der Bobriks Versuch schlichtweg als pathologische Erscheinung bezeichnete, in gewissem Sinne zu relativieren ist.

Doch Tomaševskij sah sich auch selbst in seinen eigenen Versuchen, mathematische Verfahren im Bereich der Literatur anzuwenden, kritischen Reaktionen ausgesetzt. Einer dieser Kritiker war z. B. *Grigorij O. Vinokur* (1896–1947), ein Schüler von Ušakov und Gründungsmitglied des MLK. Er bemerkte in einer Rezension zu Tomaševskijs (1923 in Berlin publizierter, 1919 zuvor im Moskauer Linguistischen Kreis vorgestellter) Untersuchung zum 5-füßigen Jambus, es handle sich dabei um „keine philologische Untersuchung“, sondern um einen Versuch, „Fragen der poetischen Sprache mit Mitteln der Mathematik und Statistik zu lösen“ (Vinokur 1924, 263; zit. nach Tomaševskij 1929, 275).

Während sich also auf der einen Seite mathematisch ausgebildete (und kompetente) Wissenschaftler um die Qualität poetologischer Ergebnisse sorgten, die aufgrund mathematischer Inkompetenz in diesem Bereich zustande kamen, gab es andererseits die (auch heutzutage noch nicht minder ver-

breitete) Befürchtung, dass die Poetik als philologische Disziplin aufgrund der Anwendung mathematischer Methoden um ihren genuinen Gegenstandsbereich gebracht würde.

Tomaševskij (1929) entgegnete Einwänden wie dem von Vinokur in einer Art, die nicht nur an Klarheit nichts zu wünschen übrig lässt, sondern auch seine methodologische Weitsicht zum Ausdruck bringt: „Man darf die Philologie nicht der Statistik gegenüberstellen, da erstere sich durch das Forschungsobjekt definiert (die Sprache), und zweitere durch die Methode, Fakten zu bestimmen [...]. Eine Zahl, eine Formel, eine Kurve – all das sind im Wesentlichen eben solche Symbole wie ein Wort – verständlich nur für diejenigen, die das entsprechende System von Symbolen verstehen. Die Angst der Philologen vor der Zahl ist größtenteils durch die falsche Vorstellung von ihrer Rolle in der wissenschaftlichen Arbeit zu erklären. Denn die Zahl ‚entscheidet‘ nichts, d. h. sie interpretiert nicht, sondern dient nur als Verfahren, ein Faktum festzustellen und zu beschreiben. Man darf nur nicht glauben, dass die Zahlen in dieser Rolle nur im ersten Zehntel gesetzmäßig sind. Das, was mit Zahlen und Graphiken an Missbrauch getrieben worden ist, bringt nicht die Methode in Verruf: schuldig ist derjenige, der den Missbrauch treibt, nicht das, was er missbraucht.“ (Tomaševskij 1929, 276). Die hier angesprochene Frage des unreflektierten Einsatzes statistischer Verfahren thematisierte Tomaševskij (1923a) auch an anderer Stelle in verallgemeinerter Form: „Bei uns ist die Ansicht weit verbreitet, dass man zählen kann, was man will – in der Annahme, die Zahlen selbst würden schon zeigen, ob das Material zur Berechnung richtig ausgewählt ist. Und wenn die Berechnung irgendeine numerische Harmonie ergibt, irgendeine numerische Gesetzmäßigkeit, dann wird daraus entschieden der Schluss gezogen, dass die für die Berechnung ausgewählte Grundlage gerechtfertigt war. Aber die Zahl hat ihre eigene Gesetzmäßigkeit.“ (Tomaševskij 1923a, 35).

In seiner 1917 publizierten, seinen eigenen Angaben nach in den Jahren 1909–15 entstandenen Untersuchung, beschäftigte Tomaševskij (1917) sich mit der Rhythmik des vierfüßigen Jambus in Puškins *Evgenij Onegin* [EO]. Bezeichnenderweise ging es ihm dabei nicht darum, die eine oder andere Sichtweise auf den Jambus zu verteidigen,

sondern vielmehr objektives Material zur Beurteilung des vierfüßigen Jambus zur Verfügung zu stellen und eine exakte(re) Klassifikation rhythmischer Erscheinungen zu ermöglichen.

Tomaševskij unterschied – ähnlich wie das später Šengeli im Sinne einer sprachlichen und rhythmischen Norm tun sollte – zwei Faktoren der rhythmischen Gestaltung: einen statischen und einen organisierenden Faktor. Vor dem Hintergrund dieser Unterscheidung stellte Tomaševskij umfangreiche Berechnungen der lexikalischen und metrischen bzw. rhythmischen Struktur des EO an: In den von ihm ausgezählten 5320 Versen, die 17300 Wörter aufweisen, kommen seinen Angaben zufolge 2022 trochäische Wörter (–U) vor, was einer relativen Häufigkeit von 0,117 entspricht; der Anteil jambischer Wörter (U–) betrage 0,268, amphibrachischer Wörter (U–U) betrage 0,265, und einsilbige (betonte) Wörter kommen auf einen Anteil von 0,1085. Auf dieser Basis lasse sich die erwartete Häufigkeit eines Verses wie *В волненьи бурных дум своих* (U–U, –U, –, U–) als Produkt dieser Häufigkeiten berechnen: $0,265 \times 0,117 \times 0,1085 \times 0,268 = 0,0009$. Summiert man diese Erwartungswerte für alle möglichen Wortverbindungen im vierfüßigen Jambus auf, kommt man nach Tomaševskij auf eine Summe von 0,103. Bei 5230 Verszeilen kommt man somit für einen Vers der oben angeführten Art auf $0,00090 \times 5320/0,103 \approx 46$ erwartete Vorkommnisse. De facto komme der obige Verstyp im EO jedoch 90 mal, also fast doppelt so häufig vor. Dass dies nicht zufällig, sondern durch (bewusste oder unbewusste) Gestaltung bedingt ist, versuchte Tomaševskij durch quantitative Vergleiche mit anderen Texten zu belegen.

Bei einem Vergleich der Vorkommenshäufigkeit von Wörtern verschiedener metrischer Typen in den Oden Lomonosovs (1746–54) und dem *Demon* von Lermontov kommt Tomaševskij (1917) zu dem Schluss, „dass der rhythmische Reichtum der Sprache Puškins sich wenig von der Sprache Lomonosovs und Lermontovs unterscheidet“. (Tomaševskij 1917, 104). Dabei versteht Tomaševskij (1917) unter *rhythmischem Reichtum* die „relative Vorkommenshäufigkeit von Wörtern mit einheitlicher rhythmischer Struktur“, und ein *Wort* definiert er als eine „jede selbständige Gruppe von Silben, die unter eine Betonung fallen (d. h. Proklitika und Enklitika werden dem entsprechenden

Tab. 3.1: Relative Häufigkeiten verschiedener Versfüße

	Puškin <i>EO</i>	Lomonosov <i>Oden</i>	Lermontov <i>Demon</i>	Puškin <i>Pikovaja dama</i>
U –	26,8	23,3	27,3	17,7
U – U	26,5	29,4	28,6	14,6
– U	11,7	11,3	9,1	15,4
–	10,9	11,5	11,1	8,6
UU –	6,3	4,9	6,5	9,4
UU – U	4,5	7,3	4,0	8,3
– UU	3,6	2,8	3,9	7,2
U – UU	3,3	5,6	3,6	7,5

betonungstragenden Wort zugerechnet“ (Tomaševskij 1917, 102). Ganz anders stellen sich die Verhältnisse nach Tomaševskij (1917, 104) jedoch in einem prosaischen Text dar, was er an einem Vergleich mit dem Text von Puškins *Pikovaja dama* belegt, der sich vor allem durch eine höhere Anzahl einsilbiger Wörter und von Wörtern, die mit einer ungeraden Anzahl unbetonter Silben beginne, unterscheidet (vgl. Tab. 3.1).

Der hier angesprochenen Frage nach rhythmischen Strukturen in prosaischer Sprache ist Tomaševskij an anderer Stelle gerade unter Bezugnahme auf Puškins *Pikovaja dama* detaillierter nachgegangen, nämlich in seiner Untersuchung *Ritm prozy* (Tomaševskij 1920, 254 ff.), die auf einen Anfang 1920 im MLK und Mitte 1921 am Russischen Institut für Kunstgeschichte (RIII) gehaltenen Vortrag zurückgeht.

Tomaševskij, der davon überzeugt ist, dass der Prosarhythmus prinzipiell mit der statistischen Methode – sei es mit Hilfe von Zahlentabellen oder auf andere Art und Weise – untersucht werden muss (Tomaševskij 1920, 275), geht davon aus, dass die erste zu lösende Frage darin besteht, eine geeignete Untersuchungseinheit zu bestimmen, die er als „zvukovoj period“ bezeichnet (Tomaševskij 1920, 262). Als erstes untersucht Tomaševskij die Ebene des Satzes als eine solche möglicherweise geeignete Einheit, weil für ihn die Berechnung einer solchen *Lautperiode*, die nach der Wortzahl bestimmt würde, keine Basis sein kann: „Die Berechnung nach *Quantität* der Worte trägt in die Sphäre der lautlichen Erscheinungen Vorstellungen aus der semantischen Reihe hinein, d. h. es überlagern sich zwei Reihen der Wahrnehmung“ (Tomaševskij 1920, 272). Möglich hingegen sei es, dass ein Satz einer solchen Lautperiode entspreche, da es sich bei einem Satz um einen „syntak-

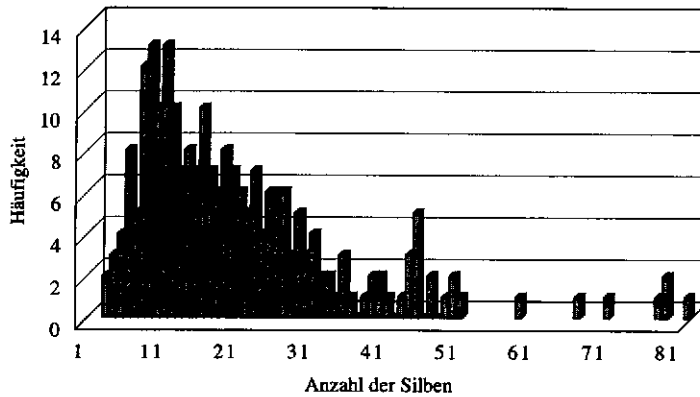
tisch abgeschlossenen Redeausschnitt“ (Tomaševskij 1920, 262) handle. Als konkretes Untersuchungsmaterial dienen ihm die beiden ersten Kapitel aus Puškins *Pikovaja dama*, Tab. 3.2 und Abb. 3.5 repräsentieren Vorkommenshäufigkeiten (f_i) der entsprechenden Silbenanzahl pro Satz (i).

Tab. 3.2: Satzlängen in *Pikovaja Dama*

i	f_i	i	f_i	i	f_i	i	f_i	i	f_i
3	2	14	8	25	6	36	1	51	1
4	3	15	7	26	6	38	1	59	1
5	4	16	10	27	6	39	2	67	1
6	8	17	7	28	3	40	2	71	1
7	5	18	6	29	5	41	1	78	1
8	12	19	8	30	3	43	1	79	2
9	13	20	7	31	4	44	3	82	1
10	10	21	6	32	2	45	5		
11	13	22	5	33	2	47	2		
12	10	23	7	34	1	49	1		
13	7	24	4	35	3	50	2		

Nach Tomaševskij (1920, 264) ergibt sich eine hinreichend gleichmäßige, schnell ansteigende und langsam fallende Kurve. Das Maximum von 11 Silben sei aber nicht charakteristisch, da der Mittelwert von 20 Silben zweimal so groß sei wie das Maximum. Tomaševskij hat sich in Anbetracht dieses Ergebnisses weniger an der Streuung (die sich gegebenenfalls durch Gruppeneinzelheiten reduzieren ließe) als vielmehr an der Abweichung von der offenbar erwarteten Symmetrie der Verteilung orientiert, da sich für ihn „jegliche Beziehung, jegliche Analogie zwischen der Länge der verschiedenen Perioden verliert“.

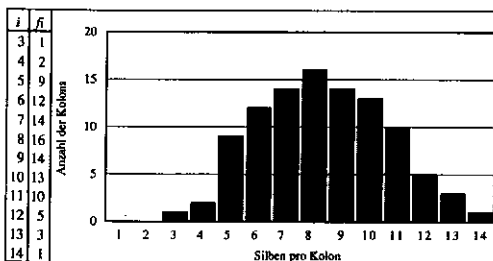
Unter Bezugnahme auf die sprachwissenschaftlichen Arbeiten zur Syntax – vor allem die in mehreren Auflagen erschienene *Grammatika russkogo jazyka* von Bogorodickij, Peškovskijs *Russkij Sintaksis v nauč-*

Abb. 3.5: Satzlängen in *Pikovaja Dama*

nom osveščenie (1914) und einschlägige Überlegungen von Ščerba – analysiert Tomaševskij (1920, 266 ff.) im weiteren Verlauf, inwiefern sich die Einheit des (von Ščerba als *Syntagma* bezeichneten) Kolons als lautliche Periode eignet.

Ein Kolon ist demnach definiert als ein „gesondertes, durch abgrenzende Intonationen abgegrenztes Satzglied“ (Tomaševskij 1920, 266). Dabei sieht sich Tomaševskij (1920) zwar gehalten einzuräumen, dass das Fehlen einer strikten Definition des Kolons es unmöglich mache, das Material mit strenger Objektivität zu untersuchen und den Einfluss eines gewissen Grades an Subjektivität bedinge; er ist sich jedoch sicher, dass dieser Umstand letztendlich keinen entscheidenden Einfluss auf die Schlussfolgerungen habe.

In der Tat erweist sich die Silbenlänge der Kola als wesentlich robuster, wie der Tab. 3.3 bzw. Abb. 3.6 zu entnehmen ist:

Tab. 3.3 / Abb. 3.6: Silbenanzahl pro Kolon in *Pikovaja dama*

Die offenbare Symmetrie der Verteilung – die sich darin äußert, dass sowohl Mittelwert als auch Maximum bei acht Silben liegen –

interpretiert Tomaševskij als Anzeichen einer korrekt gewählten Einheit zur Bestimmung dessen, was er ‚lautliche Periode‘ nannte. Dieser Befund deckte sich im Übrigen mit Beobachtungen von *Aleksandr Chr. Vostokov* (1781–1864), der bereits 1817 in seinem *Versuch über die Russische Versfü-gung* den Begriff der „prosodischen Periode“ ausführlich diskutiert und den Umfang einer solchen Periode bei sieben bis acht Silben veranschlagt hatte (Vostokov 1817, 100 ff.). Den Grund für eben diesen Umfang einer Periode – die Tomaševskij mit der Periode innerhalb einer Phrasenbetonung bei Peškovskij gleichsetzte – hatte Vostokov aus physiologischer Sicht in einer „natürlichen Begrenzung“ gesehen, die seiner Meinung nach darin bestand, „dass man bei einer erhöhten oder betonten Silbe nicht mehr erniedrigte oder unbetonte Silben aussprechen kann als die menschliche Brust aushält, die keine Atmung ausübt“ (Vostokov 1817, 100).

Tomaševskij (1920) stellt in diesem Zusammenhang klar, dass die von ihm getroffene Feststellung der Konstanz der rhythmischen Periode „nur das Faktum des normalen Umfangs ist und keinesfalls ein obligatorischer Umfang, d. h. Faktum einer normativen rhythmischen Verstärkung“ (Tomaševskij 1920, 273). Auch oder gerade deswegen stellt sich für ihn die Frage, ob und inwiefern sich ein (kausaler) Zusammenhang ergibt zwischen dem im Durchschnitt achtsilbigen Prosarhythmus und dem zur Entstehungszeit der *Pikovaja dama* gebräuchlichsten Metrum, dem vierfüßigen Jambus (mit eben dieser Silbenanzahl). Obwohl im Prinzip eine solche Beeinflussung in beide Richtun-

gen denkbar wäre, spricht Tomaševskij (1920) sich dafür aus, „die Konstanz der sylabischen Länge des rhythmischen Prosakolons außerhalb jeglicher prinzipiellen Verbindung mit der Struktur der dichterischen Rede anzusehen“ (Tomaševskij 1920, 274): Während sich im Vers der metrische Umfang als „Merkmal einer konstruktiven ästhetischen Norm“ erweise, seien die „Normen des Prosarhythmus kein konstruktives Gesetz“ (Tomaševskij 1920, 274).

In weiteren Schritten hat Tomaševskij an den ersten drei Kapiteln der *Pikovaja dama* die Verteilung der unbetonten Silben in bestimmten Positionen genauer untersucht und mit dem Gesamtvorkommen im Text verglichen.

Als erstes hat Tomaševskij den Auftakt (začin) untersucht, und zwar sowohl im Hinblick auf den Satzanfang als auch im Hinblick auf das Kolon innerhalb eines Satzes. Konkret stellt sich also die Frage, wie viele Silben der ersten betonten Silbe an der entsprechenden Position vorausgehen. Wie die Daten in Tab. 3.4 deutlich zeigen (vgl. Abb. 3.7 [1]), ist die Satzanfangsposition im Vergleich zum Gesamtvorkommen durch eine klare Dominanz trochäischer Auftakte und eine ebenso deutliche Abnahme jambischer Auftakte charakterisiert; im Durchschnitt sind es 0,81 Silben, die der ersten betonten Silbe vorangehen. Eine andere Tendenz ergibt sich im zweiten Fall, beim Kolon in der Satzmitte: hier nimmt insbesondere die Zahl kurzer Auftakte ab und die Zahl mehrsilbiger unbetonter Auftakte zu – im Durchschnitt gehen an dieser Position 1,35 Silben der betonten Silbe voraus.

Tab. 3.4: Positionsabhängige Verteilung von Betonungen

	gesamt (positions- unabhängig)	Satz- an- fang	Anfang mittl. Kolon
trochäisch	0 32,7	45,8	19,9
jambisch	1 41,3	29,8	39
anapästisch	2 20,8	22,6	29,9
paiononisch	3 4,5	1,3	8,2
hyperpaionisch	4 0,7	0,5	2,8
	5		0,2

Ein weiterführender Vergleich dieser Befunde mit der Verspraxis Puškins zeigt nach Tomaševskij (1920, 292) die – für ihn nicht unbedingt zu erwartende – Tendenz, dass die beobachteten Konstruktionen sich in

Vers und Prosa nicht wesentlich unterscheiden. Anders hingegen ist das bei einer Untersuchung der Satz- bzw. Kolaendungen. Denn bei einer entsprechenden Untersuchung der Endungen (koncovka) der rhythmischen Perioden am Satzende stellt sich eine klare „Tendenz zur Verlängerung“ (Tomaševskij 1920, 295) heraus: während an dieser Position der Anteil weiblicher Endungen im Vergleich mit dem gesamten Material abnimmt, steigt der Anteil weiblicher und vor allem daktylischer Endungen – hier gibt es keinen wesentlichen Unterschied im Vergleich zur Endung der satzinternen Kola.

Im Gegensatz zur Struktur der Auftakte sind somit die Endungen der Kola sowohl am Satzende als auch satzintern eher einförmig. Allerdings gibt es nach Tomaševskij (1920, 296) hierbei keine Analogie zur Struktur im Vers, insofern sich die im Vers ausdrückende Tendenz zur männlichen Beendigung einer Periode nicht in der Prosa spiegelt – für Tomaševskij (1920, 296) ein zentraler Hinweis darauf, „wie gefährlich es ist, im Allgemeinen eine Analogie zwischen Vers und Prosa anzuführen, ohne vorher das Verhältnis beider zueinander zu prüfen“ (Tomaševskij 1920, 296). Abb. 3.7 stellt die Befunde in anschaulicher Form dar: in der linken Graphik (Abb. 3.7 [1]) die Relationen der Auftakte, in der rechten die Kadenzen (Abb. 3.7 [2]). Jeweils punktiert sind die positionsunabhängigen Häufigkeiten, mit durchgezogener Linie die auf den Satz, mit gestrichelter Linie die auf die Kola bezogenen Relationen.

In dem selben Text ist Tomaševskij (1920, 302 ff.) der Frage der Intervalle, d. h. der Anzahl unbetonter Silben zwischen zwei betonten Silben, nachgegangen. Dabei ist er von den beobachteten Häufigkeiten ausgegangen, mit denen im untersuchten Text unbetonte Silben unmittelbar aufeinander Silben aufeinander folgen; nach Tomaševskij lassen sich (1920, 302) „nach dem Prinzip der Wahrscheinlichkeitstheorie die Vorkommenshäufigkeit der verschiedenen Intervalle berechnen“. Tomaševskij führt leider nicht an, wie er zu diesen theoretischen Werten gelangt ist; es ist jedoch am naheliegendsten, dass er die Binomialwahrscheinlichkeiten berechnet hat, die sich nach

$$P_x = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \dots, n \quad (4)$$

berechnen lassen, wobei sich p recht gut aus den beobachteten Daten als $\hat{p} = \bar{x}/n$ schät-

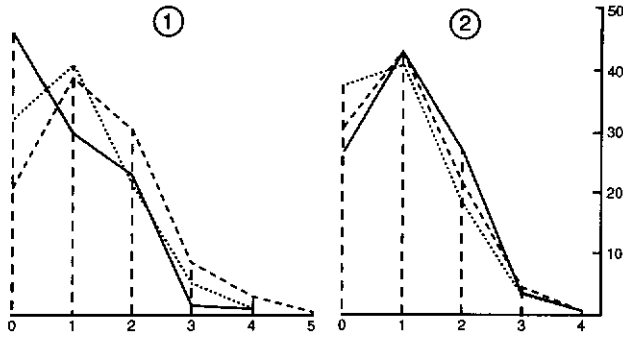


Abb. 3.7: Auftakte und Kadenzen nach Tomaševskij (1920, 289)

zen lässt und $\hat{q} = 1 - \hat{p}$ ist. Die entsprechenden Binomialwahrscheinlichkeiten sind in Prozenten den von Tomaševskij (1920, 302) angeführten Werten in der Tab. 3.5 hinzugefügt.

Die Werte der von Tomaševskij angegebenen Wahrscheinlichkeiten lassen sich, wie eine Re-Analyse seiner Daten zeigt, im Übrigen nur dann reproduzieren, wenn man bei der Berechnung der Binomialwahrscheinlichkeiten von $n = 9$ ausgeht (womit sich aufgrund von $\bar{x} = 1,87$ ein Wert von $p = 0,0208$ ergibt); eine Anpassung der Binomialverteilung an die Daten mit $n = 6$ (was ja dem tatsächlichen Klassenumfang entspricht) führt im Übrigen zu noch besseren Ergebnissen und bestätigt die Binomialverteilung als ein ausgezeichnetes Modell ($C = \chi^2/N = 0,0005$).

Tab. 3.5 enthält ebenfalls die Werte für die Länge derjenigen Intervalle, die unmittelbar auf ein einsilbiges Intervall folgen, welche sich vor allem durch einen deutlich erhöhten Anteil einsilbiger Intervalle auszeichnen (was im Anschluss an mehrsilbige Intervalle Tomaševskij zufolge nicht der Fall ist).

Abb. 3.8 veranschaulicht die Ergebnisse in graphischer Form: die punktierte Linie ist die relative Häufigkeit unabhängig von der

Position, die dünne Linie stellt die theoretischen Häufigkeiten dar, die dicke Linie symbolisiert die relativen Häufigkeiten nach einem einsilbigen Intervall.

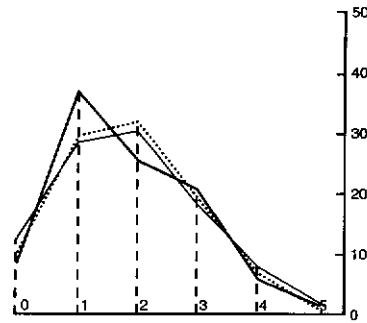


Abb. 3.8: Relative Häufigkeiten der Intervalle nach Tomaševskij (1920, 302)

Tomaševskij (1920, 305) weist abschließend einschränkend darauf hin, dass die von ihm beobachteten Erscheinungen noch keine Grundlage für weitreichende Verallgemeinerungen darstellen – lediglich (weitere) vergleichende Analysen könnten zeigen, was für das untersuchte Material spezifisch sei und was fehlerlos verallgemeinert werden könne.

Tab. 3.5: Beobachtete und theoretische Intervalle zwischen betonten Silben

	Länge des Intervalls (Anzahl unbetonter Silben zwischen betonten Silben)						
	0	1	2	3	4	5	6
beobachtet	10,3	29,8	32,1	19,7	6,8	0,8	0,5
theoretisch (nach Tomaševskij)	12,2	28,7	30,6	18,8	7,4	1,9	0,3
binomial (Re-Analyse)	10,6	28,8	32,7	19,8	6,7	1,2	0,1
n. 1-silbigem Intervall	9,4	37,5	25,1	20,7	5,9	1,3	0,3

Einen Versuch dieser Verallgemeinerung hat Tomaševskij selbst in seiner oben bereits erwähnten Untersuchung zum fünffüßigen Jambus bei Puškin unternommen, die auf einen Vortrag im MLK (Juni 1919) zurückgeht; diese Untersuchung wurde 1923 unter dem Titel *Očerki po poëtike Puškina* zusammen mit weiteren Analysen von P. G. Bogatyrev und V. B. Šklovskij in Berlin veröffentlicht und dann im Buch *O stiche* (Tomaševskij 1929, 138 ff.) wiederabgedruckt.

Tomaševskij skizziert eingangs kurz die Entstehungsgeschichte der Fragestellung nach dem Verhältnis von betonten und unbetonten Silben in Poesie und Prosa: Erstmals aufgeworfen wurde sie in der Zeitschrift *Sovremennik* 1855 von dem bekanntesten Literaturkritiker und Schriftsteller *Nikolaj G. Černyševskij* (1828–89). Aus Anlass der Besprechung einer Puškin-Ausgabe von Annenkov hinterfragte Černyševskij (1906a) die Dominanz der zweisilbigen jambischen und trochäischen Versfüße dahingehend, ob nicht der Jambus „das für die russische Sprache natürlichste Metrum“ (Černyševskij 1906a, 286) sei (vgl. Grzybek 2004a). Als erste Vermutung, dass dies nicht der Fall sei, führte er an, dass im Russischen (im Vergleich zum Deutschen) die Wörter eher mehrsilbig seien, und dass Präpositionen und Pronomina in der Regel keine, alle anderen Wörter nur jeweils eine Betonung tragen. Diese Beobachtung versuchte Černyševskij (1906a, 287) an ausgewählten Stichproben zu überprüfen. So zählte er in einem Textausschnitt von *A. F. Pisemskij* (1821–1881) 66 betonte von insgesamt 193 Silben, in zwei anderen Textausschnitten waren 25 von 75 bzw. 27 von 83 Silben unbetont. In allen drei Fällen (sowie in den drei aufsummierten Stichproben) lag somit die Relation betonter und unbetonter Silben in „erstaunlicher Nähe“ zueinander bei 1:3, was Černyševskij (1906a) veranlasste, „den unumgänglichen Schluss zu ziehen, dass Jambus und Trochäus, die bei 30 Silben 15 Betonungen verlangen, in der russischen Sprache bei weitem nicht so natürlich sind wie Daktylus, Amphibrachys, Anapäst, die bei 30 Silben 10 Betonungen verlangen.“ (Černyševskij 1906a, 287)

In einer anderen Besprechung aus dem selben Jahr wünschte Černyševskij (1906b, 359) sich, dass die mathematischen Wissenschaften mit dem von ihnen erreichten Stand in vielerlei Hinsicht als Beispiel dafür dienen sollten, wohin auch die übrigen Wissenschaften streben sollten. Er bezog das

zum einen auf die klare Strukturierung des Fachs und die damit verbundene Ausbildung auf verschiedenen Niveaus, lobte aber auch die Klarheit der Begriffsbildung, die er in einem konstruierten Beispiel veranschaulichte, indem er für die Kategorie des Dativs die folgende fiktive Darstellungsform notierte:

$$[(a\sqrt[5]{b^3} + c^4 - d^2/e^3)/(a + b^2)]^2$$

In seinen 1877/78 in der Zeitschrift *Russkij Vestnik* veröffentlichten, später wiederholt in Buchform erschienenen Ausführungen *O drame* setzte sich *Dmitrij V. Averkiev* (1836–1905) harsch mit den Überlegungen von Černyševskij auseinander.

Averkiev war ein bekannter Dramaturg, der 1859 die physikalisch-mathematische Fakultät der Universität Petersburg absolviert hatte; er stand seit jener Zeit in engem Kontakt mit zwei zentralen Figuren des literarischen Lebens, *Apollon A. Grigor'ev* (1822–64) und *Nikolaj N. Strachov* (1828–96), der übrigens ebenfalls in Sankt Petersburg Mathematik studiert und zunächst in Odessa, später in Sankt Petersburg Mathematik unterrichtet hatte.

Averkiev wies darauf hin, dass im vier- bzw. fünffüßigen Jambus oder Trochäus vier bzw. fünf realisierte Betonungen eine Seltenheit seien; vielmehr kämen auf 8–9 Silben in der Regel zwei oder drei, auf 10–11 Silben drei oder vier, auf 12–13 Silben vier oder fünf realisierte Betonungen – d. h. in etwa so viele, wie in Černyševskij's Prosa-Studien angegeben.

Ungeachtet aller Kritik, die sich in Anbetracht der Überlegungen von Černyševskij und Averkiev aufdrängt (vgl. Tomaševskij 1919b, 162 f.), war damit bereits Anfang der zweiten Hälfte des 19. Jh.s darauf hingewiesen worden, die Metrik der Verssprache nicht als isoliertes Problem, sondern im Lichte der praktischen Prosasprache zu sehen. Genau vor diesem Hintergrund lokalisiert Tomaševskij seine Untersuchungen. Bei einem einleitenden Vergleich der Vorkommenshäufigkeit unbetonter Silben in jambischen Versen verschiedener Länge kommt er dabei auf die in Tab. 3.6 angeführten Prozentzahlen.

Die Darstellung in Abb. 3.9 macht für Tomaševskij deutlich, „dass man eine gerade Linie erhält, die von einer Proportionalität der Zunahme des Verses und der Anzahl unbetonter Silben zeugt“ (Tomaševskij 1919b, 167).

Fast noch wichtiger aber ist die in diesem Zusammenhang von ihm gemachte Be-

Tab. 3.6: Verslänge und Anzahl unbetonter Silben nach Tomaševskij (1919b, 172)

Anzahl gerader Silben pro Vers		1	2	3	4	5	6
Prozentuale Häufigkeit	(beobachtet)	0	24	63	84	107	142
unbetonter Silben	(theoretisch)	0	28	56	84	112	140

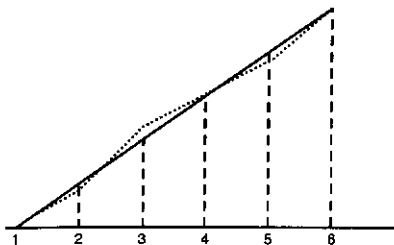


Abb. 3.9: Zusammenhang von Verslänge und Anzahl unbetonter Silben nach Tomaševskij (1919b, 172)

obachtung, „dass die Gesamtmenge der unbetonten geraden Silben nicht zur Gesamtzahl der geraden Silben eines Verses (der „Anzahl der Takte“) proportional ist, sondern nur zu der um 1 verminderten Zahl“ (Tomaševskij 1919b, 167). Dies drückt sich auch in der Formel für die theoretisch berechneten Werte aus, die auf der Gleichung

$$y = 0,28 \cdot (x - 1) \quad (5)$$

basieren, was in der Tat einen nahezu perfekten linearen Zusammenhang ergibt. Dies sei die unmittelbare empirische Widerlegung der Annahme von Averkiev, dass die nicht realisierte Betonung gerader Silben die Verssprache in die praktische Sprache überführe: Träfe diese Annahme nämlich zu, dann müsste eine exakte Proportionalität zwischen Nicht-Realisierung der Betonung und der vollständigen syllabischen Verslänge vorliegen. De facto aber liege eine solche Proportionalität zwischen Anzahl der Betonungen und Silbenzahl nur bis zur konstanten Betonung des Reims vor. Insofern komme gerade den letzten Silben eines Verses, da sie eine konstante Betonung aufweisen, besondere Bedeutung zu. Daraus wiederum folgt für Tomaševskij (1919b), dass man das „Gesetz der Anzahl von Betonun-

gen“ kenne, sobald man das „Gesetz der Verteilung der unbetonten Silben“ (Tomaševskij 1919b, 169), kenne. In der Tat lässt sich – da die Wahrscheinlichkeit dafür, dass eine der inneren geraden Silben betont ist, 0,72 beträgt – die Anzahl der zu erwartenden Betonungen (abgesehen von der letzten) in einem Vers berechnen, so etwa für einen 5-füßigen Jambus als $0,72 \times 4 = 2,88$.

Allerdings ist die Verteilung der Betonungen bei diesen inneren geraden Silben nicht gleichmäßig: Abgesehen vom zweifüßigen Jambus – der nur eine freie Silbe aufweist, so dass die durchschnittlichen 0,72 Betonungen nur auf die zweite Silbe fallen können – weist vor allem die letzte gerade Silbe vor dem Reim eine *Konstanz* auf, die darin besteht, dass diese Silbe unabhängig von der Verslänge in nur ca. 40 % der Fälle betont ist (vgl. Tab. 3.7).

Eine ähnliche Konstanz weist die erste gerade, also die zweite Silbe eines Verses auf – eine Ausnahme stellt hier der dreifüßige Jambus dar, von dessen 144 (für die inneren geraden Silben) zu erwartenden Betonungen 40 auf der vierten Silbe liegen, so dass alle übrigen auf die zweite fallen (vgl. Tab. 3.8).

Tab. 3.8 enthält die entsprechenden Daten, wobei Tomaševskij (1929, 174) die Prozentzahlen für den 6-füßigen Jambus von Šengeli übernommen hat, was gegebenenfalls aufgrund unterschiedlicher Zählweisen die Abweichungen erklärt. Mit Ausnahme des dreifüßigen Jambus weist also die zweite Silbe im Durchschnitt in 85 % der Fälle Betonungen auf, so dass an dieser Stelle statt der zu erwartenden 28 % nur 15 % unbetonter Silben zu erwarten sind.

Nach einer Reihe detaillierter Einzelberechnungen kommt Tomaševskij so zur theoretischen Verteilung der Betonungen in verschiedenen jambischen Verstyphen. Im

Tab. 3.7: Zusammenhang zwischen Verslänge und Betonungsanzahl

Anzahl der Takte (Verslänge)	3	4	5	6	\bar{x}
Anzahl der Betonungen der letzten geraden Silben vor dem Reim (in %)	39	43	40	38	40

Tab. 3.8: Zusammenhang zwischen Verslänge und Betonungsanzahl

Anzahl der Takte (Verslänge)	4	5	6	\bar{x}
Anzahl der Betonungen der letzten geraden Silben vor dem Reim (in %)	84	84	90	85

4-füßigen Jambus z. B. sollten bei 100 Versen im Durchschnitt 216 Betonungen ($400 - 100 - 56 = 216$) vorkommen; zu erwarten wäre, dass 85 Betonungen auf der zweiten und 40 auf der sechsten Silbe liegen, so dass für die vierte Silbe 91 Betonungen übrig bleiben; nach Tomaševskij (1919b, 177) liegen die im *Evgenij Onegin* zu beobachten-

den Häufigkeiten bei 84, 43 und 90. Da die Ergebnisse auch für die anderen aufgezeigten Typen gleich gut sind, ist Tomaševskij (1919b) überzeugt, „dass das gefundene Zahlengesetz von der Verteilung der Betonungen im Jambus richtig ist“ (Tomaševskij 1919b, 177).

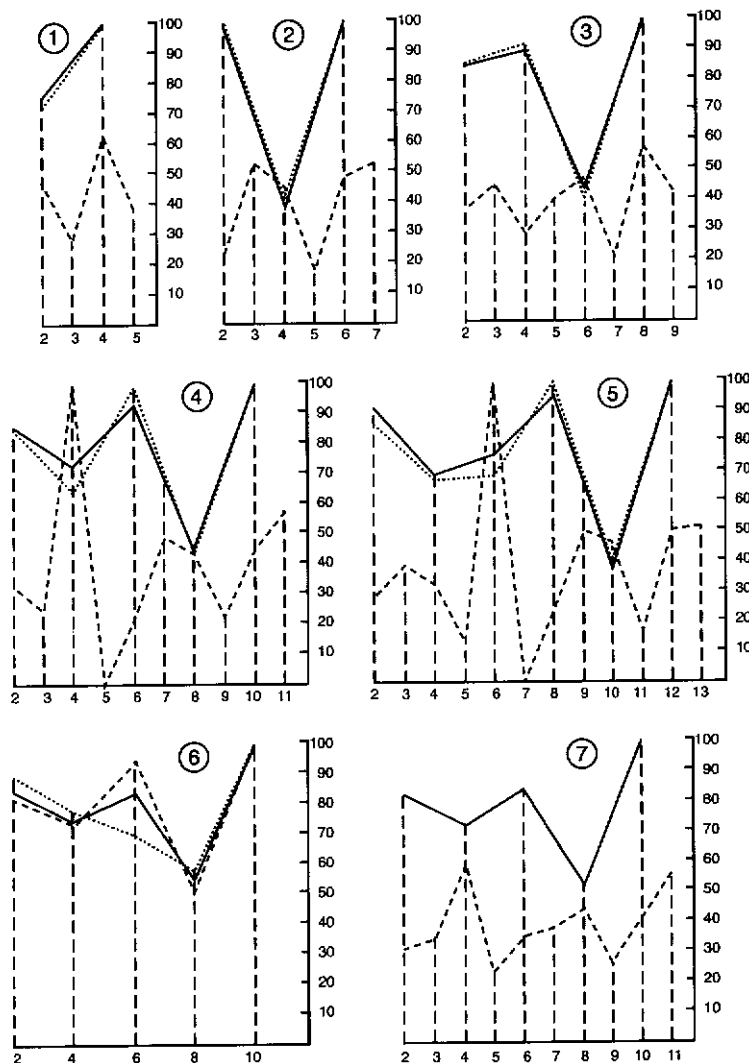


Abb. 3.10: Beobachtete und theoretische Häufigkeiten verschiedener jambischer Typen nach Tomaševskij (1919b, 179)

Tab. 3.9: Relative Frequenz von Wortlängen

Silben pro Wort	Prosa	Jambus				
		2-füßig	3-füßig	4-füßig	5-füßig	
					m. Zäsur	o. Zäsur
1	8,55	17,81	11,31	10,84	15,72	14,49
2	33,09	49,31	32,41	38,44	34,63	36,76
3	31,25	20,54	38,95	36,40	29,50	31,91
4	18,71	8,90	15,17	10,44	18,37	13,47
5	6,30	3,42	2,15	2,88	1,74	3,13
6	1,81			0,94	0,01	0,21
7	0,27			0,07	0,01	0,01
\bar{x}	2,88	2,31	2,64	2,59	2,56	2,55

Abb. 3.10 zeigt die beobachteten und theoretischen Häufigkeiten für die verschiedenen jambischen Typen: (1) 2-füßig, (2) 3-füßig, (3) 4-füßig, (4) 5-füßig mit Zäsur, (5) 6-füßig, (6) und (7) 5-füßig ohne Zäsur (m. unterschiedlichen Wortgrenzen); die beobachteten Betonungen sind mit durchgezogener, die theoretischen mit punktierter Linie dargestellt.

Abb. 3.11 stellt zusätzlich die Wortgrenzen für die einzelnen Verstypen dar, deren

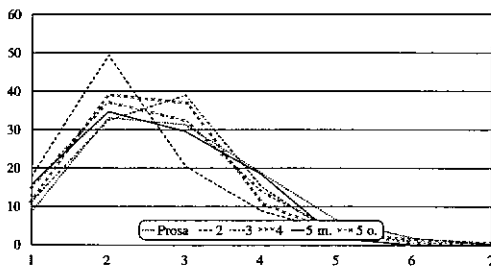


Abb. 3.11: Häufigkeiten der Wortgrenzen in verschiedenen Verstypen nach Tomaševskij (1919b, 197 ff.)

Verteilungen im Vers Tomaševskij (1919b, 197 ff.) im Detail untersucht. Die Ergebnisse einer akribischen Darstellung der Vorkommenshäufigkeiten von Betonungsstellen in Wörtern – differenziert nach der Wortlänge einerseits, nach Texttyp andererseits (in der Prosa und in fünf verschiedenen jambischen Verstypen) – hat Tomaševskij (1919b, 197 ff.) in drei resümierenden Tabellen zusammengefasst. Tab. 3.9 enthält die relativen Häufigkeiten aller Wortlängenvorkommnisse, getrennt nach Text- bzw. Verstyp.

Deutlich zu sehen ist zum einen, wie sich die mittlere Wortlänge, zum anderen das Verteilungsmodell in Abhängigkeit vom Texttyp ändert – besonders auffällig im Vergleich zur Prosa ist der hohe Anteil 2-silbiger und der niedrige Anteil 3-silbiger Wörter beim 2-füßigen Jambus, sowie der hohe Anteil 3-silbiger Wörter beim 3-füßigen Jambus.

Tab. 3.10 zeigt die Häufigkeit der Silbenanzahl, mit der unbetonte Silben der Betonung eines Wortes vorausgehen. Hier ist im Vergleich zur Prosa zum einen der Anstieg

Tab. 3.10: Relative Häufigkeit der Silbenanzahl vor der 1. Betonung

unbetonte Silben vor Wortbetonungen	Prosa	Jambus				
		2-füßig	3-füßig	4-füßig	5-füßig	
					m. Zäsur	o. Zäsur
0	32,67	26,71	20,72	26,29	24,67	30,01
1	41,27	60,27	63,07	56,83	61,91	54,31
2	20,82	3,42	13,40	10,99	8,24	9,73
3	4,54	9,58	2,80	5,88	5,11	5,83
4	0,68			0,01	0,01	0,12
5	0,01			0,01	0,01	0,01
6						
\bar{x}	0,99	0,96	0,98	0,97	0,94	0,92

Tab. 3.11: Relative Häufigkeit der Silbenanzahl nach der Wortbetonung

unbetonte Silben nach Wortbetonungen	Prosa	Jambus				
		2-füßig	3-füßig	4-füßig	5-füßig	
					m. Zäsur	o. Zäsur
0	37,45	65,06	42,87	46,36	50,93	47,56
1	40,61	34,94	46,68	45,14	36,58	42,73
2	18,38		7,85	8,00	12,05	8,93
3	3,28		0,59	0,51	0,42	0,69
4	0,28					
\bar{x}	0,88	0,35	0,64	0,63	0,62	0,63

von Wortformen, die mit einer ungeraden Zahl unbetonter Silben beginnen, auffällig, zum anderen die Tatsache, dass die mittlere Anzahl unbetonter Silben vor einer Wortbetonung insgesamt relativ konstant ist.

Anders ist das im Fall der Häufigkeit der Silbenanzahl, mit der unbetonte Silben der Betonung eines Wortes folgen (vgl. Tab. 3.11). Hier sind die deutlichsten Veränderungen im Vergleich zur Prosa zu beobachten, was sich insbesondere im extremen Abfall der Mittelwerte äußert. Tomaševskij (1919b) sieht deshalb aufgrund der Beobachtung, „dass die Zahlen der Wortanfänge und -endungen größeren Abweichungen unterliegen als die mittleren Wortlängen“, die Schlussfolgerung bestätigt: „Für den Vers ist nicht so sehr die Anzahl der Silben im Wort wichtig als die Position der Betonung.“ (Tomaševskij 1919b, 200).

Damit stellt sich als Nächstes die Frage, inwiefern ein Vers als eine in sich geschlossene Einheit anzusehen ist oder ob es zwischen den einzelnen Versen Beziehungen in Form von bestimmten Konfigurationen gibt. Solche Figuren hatte ja A. Belyj in den Vordergrund seines Interesses gestellt – Tomaševskij (1919b, 182 ff.) geht es nun darum, dieses Problem wahrscheinlichkeitstheoretisch zu betrachten. So stellt er als erstes die Frage nach der Wahrscheinlichkeit des Vorkommens eines bestimmten Typs von Betonungsverteilung innerhalb eines Verses in Isolation oder in Gruppen von n identischen Typen in unmittelbarer Aufeinanderfolge. Zur Berechnung geht er davon aus, dass ein solcher Verstyp die Vorkommenswahrscheinlichkeit α habe, so dass β die Summe der Wahrscheinlichkeiten aller anderen Typen sei, womit $\alpha + \beta = 1$. Die Wahrscheinlichkeit p_n , dass ein Verstyp in einer Gruppe von n identischen Verstypen in unmittelbarer

rer Aufeinanderfolge vorkommt, berechnet sich dann als:

$$p_n = \alpha^n \cdot \beta^2 \quad (7)$$

Nimmt man als Beispiel aus Puškins *Boris Godunov* mit seinen 1579 im fünffüßigen Jambus geschriebenen Verszeilen den Verstyp, in dem auf der 4. und 8. Silbe die Betonung nicht realisiert ist, so beläuft sich deren Frequenz auf 301, was einem Prozentsatz von 19,1 % entspricht. Die Wahrscheinlichkeit, dass dieser Verstyp in Isolation vorkommt, beträgt somit $p_1 = 0,19^1 \times 0,81^2$, was nach Multiplikation mit $N = 1579$ insgesamt 197 Vorkommnisse erwarten ließe – de facto sind es 182; für zwei bzw. drei direkt aufeinander folgende Verstypen sind die theoretischen und beobachteten Werte 38 und 43 bzw. 7 und 8. Auf ähnliche Weise berechnet Tomaševskij die Vorkommenswahrscheinlichkeit von Figuren à la Belyj. So berechnet er die theoretische Vorkommenshäufigkeit eines *Dachs* im *Boris Godunov* – einer Kombination also, bei der im ersten Vers nur die 6. und im zweiten Vers die 4. und 8. Silbe unbetont sind – als 10, in der Realität sind es 9; die Häufigkeit eines *Kreuzes* berechnet Tomaševskij mit 2, was genau der Beobachtung entspricht, usw.

Insgesamt sind die Ergebnisse für Tomaševskij (1919b, 187) dermaßen überzeugend, dass er die Notwendigkeit einer weiteren Behandlung der Frage für erledigt erklärt – und zwar aufgrund von objektiven Fakten, und nicht von willkürlichen Erklärungen. Belys Figuren sind für Tomaševskij (1919b) das mechanische Resultat einer zufälligen Verbindung unterschiedlicher Verstypen, deren „scharfsichtige Erläuterungen“ seines Erachtens „nicht mit den objektiven Vorzügen der Methode, sondern ausschließlich mit den herausragenden intuitiven Fähigkeiten

des Autors zu erklären sind“ (Tomaševskij 1919b, 188).

Der Ansatz Tomaševskijs, die Auslegungen von Belyj auf Wahrscheinlichkeitstheoretischer Ebene zu behandeln, kann historisch gesehen gar nicht hoch genug eingeschätzt werden. Das schützt Tomaševskij freilich nicht vor Kritik, die es im Hinblick auf die von ihm angestellten Berechnungen vorzubringen gilt: Zum ersten beträgt die Summe aller Wahrscheinlichkeiten seiner Formel nicht 1, was die Formel de facto für das von Tomaševskij angestrebte Ziel unbrauchbar macht; wenn man seine Formel auf $a^x \cdot b$ korrigiert, wären zwar die Wahrscheinlichkeiten korrigiert, aber jeweils nur für einen einzigen Typ von Vorkommnissen gleichzeitig.

Tomaševskij stellt weitere Berechnungen an, auf die hier nicht im Detail eingegangen werden kann. So berechnet er etwa im Hinblick auf den Reim zum einen die Anzahl der in einer sog. Reim-Kette verbundenen Reime, zum anderen die Größe und Häufigkeit der Intervalle zwischen den Gliedern eines Reims (Tomaševskij 1919b, 211 ff.). Auch stellt er Vergleiche von Puškin-Texten mit Texten anderer Autoren (in anderen Sprachen) an, um die Frage der Spezifik von Autor, Sprache und Metrum zu klären (Tomaševskij 1919b, 220); er vergleicht die Dichtung Puškins verschiedener zeitlicher Perioden miteinander (Tomaševskij 1919b, 231 ff.) und setzt die beschriebenen Verfahren zur Bestimmung strittiger Autorschaft ein (Tomaševskij 1919b, 249 f.).

Damit lässt sich zusammenfassend die Schlussfolgerung ziehen, dass es Tomaševskij nicht nur gelungen ist, erstmals Wahrscheinlichkeitstheoretische Überlegungen in die Versanalyse einzuführen, sondern die Grundlagen für quantitative Verfahren in Sprach- und Textwissenschaft schlechthin zu legen. Bezeichnend ist dabei, dass die disziplinären Grenzen zwischen philologischen und mathematischen Ansätzen dabei ebenso verschwimmen wie jene zwischen Sprach- und Literaturwissenschaft. Nicht zufällig bezieht er sich in seinen Versuntersuchungen immer wieder auf die zeitgenössischen linguistischen Arbeiten nicht nur von Baudouin-Schülern wie Bogorodickij oder Ščerba, sondern auch auf die Arbeiten des aus der Moskauer Fortunatov-Tradition stammenden Peškovskij.

Mit dem Verweis auf dessen Arbeiten schließt sich insofern zum einen der Kreis zwischen quantitativen Ansätzen in der Sprach- und Literaturwissenschaft; zum an-

deren zeigt sich deutlich, wie die von Baudouin de Courtenay und Fortunatov geschaffenen Grundlagen der Kazaner und der Moskauer Schule nunmehr ihre Früchte trugen. Tomaševskij, der selbst ja Baudouin-Schüler war, ist nur eines von vielen Beispielen dafür.

Die erste breite sprachstatistische Darstellung erschien Ende der 20er Jahre; es handelt sich um den *Versuch der Anwendung der statistischen Methode auf die Sprachwissenschaft* (Čistjakov/Kramarenko 1929) von V. F. Čistjakov und B. K. Kramarenko, der 1929 in 350 Exemplaren in Krasnodar erschien. Das Buch stellt erstmals in der Geschichte der Quantitativen Linguistik – nicht nur in Russland – ein breites Spektrum des sprachstatistischen Feldes dar. Dies wurde auch zur damaligen Zeit so gesehen: so wies etwa Kopeckij (1931, 169) in seiner Rezension darauf hin, dass sich durchaus auch schon zuvor Forscher wie (die im Übrigen auch im Anhang des Buches von Čistjakov/Kramarenko genannten) Bogorodickij, Peškovskij, u. a. mit sprachstatistischen Fragen beschäftigt hätten, dass aber niemand zuvor in einem derartigen Umfang statistische Methoden auf die Sprache angewendet habe. Ziel dieses Buches war es, wie im Vorwort dargelegt wird (Kopeckij 1931, 4):

- (1) die prinzipielle Anwendungsmöglichkeit der statistischen Methode auf sprachwissenschaftliche Fragen darzulegen;
- (2) eine Reihe von linguistischen Problemen aufzuweisen, deren Lösung entsprechender Verfahren bedarf;
- (3) die Anwendung der statistischen Methode an konkretem Material zu demonstrieren.

In einleitender Auseinandersetzung mit dem (junggrammatischen) Gesetzesbegriff der Sprachwissenschaft, und unter Bezugnahme auf diesbezügliche kritische Aussagen, wie sie etwa Aleksandr I. Tomson (1860–1935), ein Schüler Fortunatovs, in seinem Werk *Obščee jazykoznanie* (Tomson 1910) oder Ferdinand de Saussure getroffen hatten, sprechen die Autoren von „ideographischen Gesetzen“, zu deren Beschreibung die Anwendung statistischer Verfahren nahe liege.

Auf Saussures *Cours de linguistique générale*, der erst 1933 erstmals in russischer Übersetzung von A. M. Suhotin erschien, hatte in Russland in schriftlicher Form erstmals Peterson in den 20er

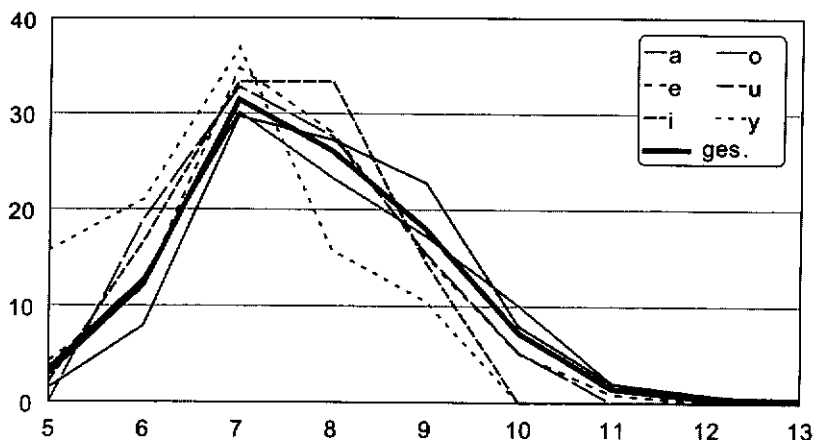


Abb. 3.12: Differenzierende Häufigkeitsverteilung der Wortlängen nach Čistjakov/Kramarenko (1929)

Jahren aufmerksam gemacht (vgl. Grzybek 1989, 21 ff.; 1999, 1290 ff.).

Dies schließt nicht die Existenz natur-historischer Gesetze auch im Bereich der Sprache aus; allerdings dürfe man die Rolle der Statistik im Bereich der Sprachwissenschaft nicht überschätzen, da man mit ihr nur statistische Gesetzmäßigkeiten aufzeigen könne – zu deren Erklärung seien hingegen spezifische Methoden der Linguistik notwendig (Kopeckij 1931, 7).

Im Anschluss an die einleitenden theoretischen Überlegungen werden zunächst „Formeln des Lautbestands“ (Kopeckij 1931, 9 ff.) analysiert, in denen es in erster Linie um statistische Analysen des Bestands der verschiedenen Laute und Lauttypen, deren Häufigkeiten und Kombinationen geht. Dabei ergeben sich jedoch durchaus auch über diesen engeren Bereich hinausgehende relevante Ergebnisse, die von den Autoren allerdings nicht weiter verfolgt werden. So analysieren sie z. B. alle Wörter aus dem Wörterbuch von V. I. Dal' mit der orthographischen Endung *-чик*, bzw. *-счик* und *-зчик* (was in phonetischer Hinsicht den Endungen *-цьк* bzw. *-щцьк* entspricht) im Hinblick auf die den Akzent tragenden Vokale. Es handelt sich dabei um insgesamt 694 Wörter, von denen die Autoren jeweils die Laute pro Wort zählen, und dies getrennt nach den einzelnen betonten Vokalen und insgesamt. Damit ergibt sich gleichzeitig auch eine Häufigkeitsverteilung der Wortlängen, die in Abb. 3.12 dargestellt ist.

Unter Hinweis darauf, dass die Kurven unabhängig vom jeweils betonten Vokal ei-

nen überaus ähnlichen Verlauf aufweisen, der aufgrund der linksipfligen Verschiebung eine deutliche Abweichung von einer Normalverteilung aufweist, haben Čistjakov/Kramarenko (1929, 43) zum einen Kenngrößen wie Schiefe und Kurtosis angegeben; darüber hinaus haben sie auch versucht, den Kurvenverlauf mit der folgenden Formel zu modellieren, die sie als Typ III des Pearson-Systems bezeichnen. Demnach würde die Formel wie folgt aussehen:

$$y = (1 + x/b)^c \cdot e^{-yx} \quad (8)$$

Allerdings lassen sich bei einer entsprechenden Re-Analyse weder mit dieser Formel noch mit der von Čistjakov/Kramarenko (1929, 43) dargestellten einigermaßen befriedigende Ergebnisse erzielen. Lediglich die Ergänzung um eine weitere (also vierte) Variable führt mit der sich dann ergebenden Formel

$$y = a \cdot (1 + x/b)^c \cdot e^{-yx} \quad (9)$$

zu einem guten Ergebnis ($R^2 = 0,98$). Ein ebenso gutes Ergebnis lässt sich allerdings auch mit einer sehr viel einfacheren Formel erzielen (vgl. Altmann 1980), nämlich

$$y = a \cdot x^b e^{-cx} \quad (10)$$

Das Ergebnis ist in der Abb. 3.13a dargestellt.

Abgesehen davon, dass das Ergebnis der theoretischen Anpassung nicht wirklich befriedigend ist und sich de facto nur durch eine andere Kurve erfassen lässt als die beschriebene, würde man heute eher eine diskrete, keine stetige Verteilung an die Daten

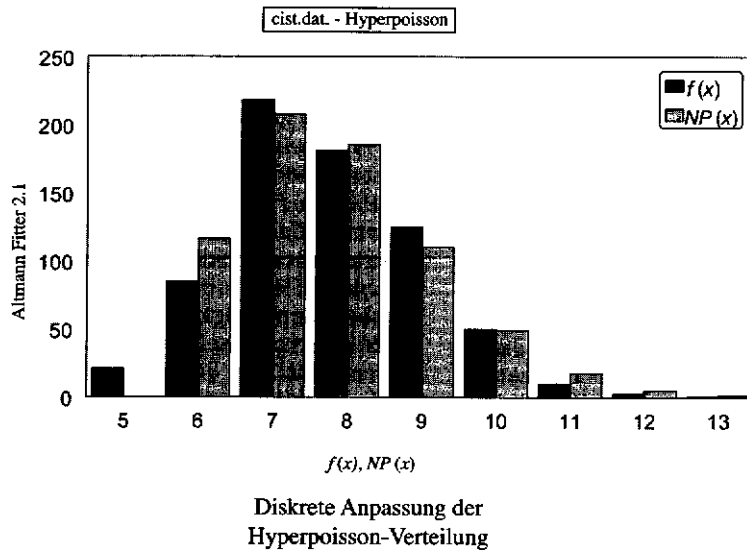
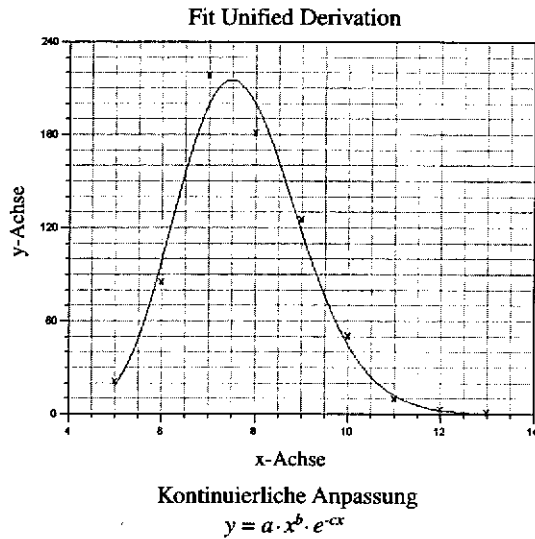


Abb. 3.13a/b: Kontinuierliche und diskrete Anpassung von Wortlängenhäufigkeiten

anzupassen versuchen. Führt man eine solche Re-Analyse durch, so stellt sich dabei interessanterweise heraus, dass die zu diesem Zweck für die Verteilung russischer Wortlängenhäufigkeiten in letzter Zeit wiederholt herangezogene Hyperpoisson-Verteilung überaus gut geeignet ist, die spezifischen Daten von Čistjakov/Kramarenko zu modellieren ($a = 1,7847$, $b = 0,0005$; $\chi^2_3 = 8,02$, $P = 0,16$).

Weiterhin werden auch quantitative Fragen der Morphologie diskutiert: im Detail

analysiert wird unter anderem die relative Vorkommenshäufigkeit russischer Präpositionen, Suffixe und Präfixe, wobei deren Vorkommenshäufigkeiten zur Produktivität von Wurzelmorphemen in Beziehung gesetzt werden. Ebenfalls enthalten sind statistische Untersuchungen zur Vorkommenshäufigkeit verschiedener Wortarten. So werden u. a. diesbezügliche Angaben auf der Basis von ca. 25.000 Wortvorkommnissen angeführt, differenziert nach vier verschiedenen (historischen) Stiltypen, die in Bezie-

hung zu entsprechenden (weniger umfangreichen) Untersuchungen zum Französischen, Deutschen, Englischen und (Alt-)Hebräischen gestellt werden.

Das Buch von Čistjakov/Kramarenko (1929) sollte eigentlich der erste von zwei geplanten Bänden sein – vorgesehen war ursprünglich auch ein zweiter, nie erschiener Band mit weiterführenden Untersuchungen zur Silbenstruktur, Syntax und Semasiologie. Doch auch das nur als erster Band geplante Buch von 1929 ist in der Folgezeit nahezu vergessen worden.

Einer der beiden Autoren, Čistjakov, der Ende der 30er Jahre mit dem *Slovar' komedii „Gore ot uma“ A. S. Griboedova* (Čistjakov 1939) das erste erklärende Wörterbuch dieser Art veröffentlichte, beschäftigte sich auch später noch bis in die 70er Jahre mit quantitativen Fragen wie etwa dem Verhältnis von Konsonanten und Vokalen in verschiedenen Sprachen (Čistjakov 1971; 1972a) oder dem Häufigkeitsvorkommen von Präpositionen (Čistjakov 1972b) – Fragen, die im Prinzip auch schon 1929 angesprochen und behandelt worden waren. Insofern wies Čistjakov (1964) vor dem Hintergrund der sich in den 60er Jahren in Russland zunehmend etablierenden Quantitativen Linguistik in einem Brief an die Redaktion der Zeitschrift *Filologičeskie nauki* mit ambivalent bedauernder Genugtuung auf die Ergebnisse seines 34 Jahre zuvor erschienenen Frühwerks hin. Deutlichere Töne ließ Čistjakov (1965) in einer akribischen Darstellung anklingen, in der er Passagen aus dem Werk von 1929 aktuellen Arbeiten von L. P. Zinder vergleichend gegenüberstellte, um zu belegen, dass Zinder sich in vielem auf die Arbeit dieses Werks stützte, ohne dies freilich angemessen kenntlich zu machen. Zinder hatte 1962 an der Universität Petersburg den ersten Lehrstuhl für Mathematik eingerichtet, und zwar an der 1958 ins Leben gerufenen Abteilung für Angewandte Linguistik.

Interessant ist allerdings, dass sich ab Mitte der 20er Jahre verstärkt auch Fortunatov-Schüler mit mathematischen Methoden in Anwendung auf Sprache und Literatur beschäftigten. So geht eine der ersten, immer wieder zitierten Lautstatistiken auf Peškovskij (1925) zurück, der sich schon zuvor (Peškovskij 1924) zur Frage von Prosa und Vers aus linguistischer Sicht geäußert hatte. Seine nachhaltig einflussreiche Studie „Zehntausend Laute“ entstand vor dem Hintergrund zahlreicher Arbeiten zum Problem der Lautinstrumentierung in poetischen Texten; die damit verbundenen Fragen zielten nach Peškovskij (1925, 167) entweder (a) auf die Dominanz oder (b) auf die Verteilung eines bestimmten Lautes bzw. einer Lautgruppe

in einem gegebenen Text. Peškovskij (1925, 167) verstand seine Studie als einen Beitrag für eine objektive Grundlage entsprechender Untersuchungen. Als Material dienten ihm Aufzeichnungen mündlicher Rede, aufgeteilt in 10 Stichproben à 1000 Laute; diese Stichprobengrößen sah Peškovskij als hinreichend für die anzustrebende Konstanz der Ergebnisse an.

Die Ergebnisse betrachtete Peškovskij (1925, 181) durchaus als über das Russische hinausgehend relevant, so den größeren Anteil von Vokalen im Vergleich zu den Konsonanten (bzw. der silbenbildenden Laute im Vergleich zu den nicht-silbenbildenden), der unbetonten Vokale im Vergleich zu den betonten, der kurzen und harten im Vergleich zu den langen und weichen, usw. Gerade in den konkreten Proportionen aber könne das Überwiegen der einen oder anderen Kategorie Unterschiede zwischen verschiedenen Sprachen, verschiedenen literarischen Sprachen oder verschiedenen literarischen Genres manifestieren. So habe z. B. eine Berechnung der ersten 1000 Laute aus Puškins *Evgenij Onegin*, Goethes *Faust* und Dantes *Divina Commedia* ergeben, dass der Anteil stimmloser Konsonanten im ersten Text ca. 20 %, im *Faust* ca. 24 %, und in der *Divina Commedia* ca. 18 % beträgt. Auf der einen Seite unterscheidet sich also der klangliche *Evgenij Onegin* deutlich von der russischen gehobenen Umgangssprache, die mit 25,5 % geräuschvoller sei, auf der anderen Seite unterscheidet sich die russische Sprache ebenfalls deutlich vom Deutschen und Italienischen – was Anlass zu analogen Berechnungen in weiteren Sprachen und Genres gebe.

Ein weiterer unmittelbarer Fortunatov-Schüler war *M. N. Peterson*, der dann nicht nur Lehrer u. a. von G. O. Vinokur war, sondern z. B. auch von *Vjačeslav Vs. Ivanov* und *Vladimir N. Toporov*, die in den späten 50er und 60er Jahren maßgeblich an der Herausbildung von Semiotik und Strukturalismus beteiligt waren.

Peterson war nach Absolvierung des Studiums an der Moskauer Universität dort seit 1919 als Professor tätig. Sein Schwerpunkt lag auf der vergleichend-historischen Grammatik der indoeuropäischen Sprachen, insbesondere Syntax und Morphologie. 1928 stellt er eine Studie zu „Konstruktionen mit der Präposition ‚iz‘ bei Lermontov“ (Peterson 1928) an. De facto handelt es sich hierbei um die erste quantitative Untersuchung

Tab. 3.12: Lautfrequenzen nach Peškovskij (1925)

		Vokale			Konsonanten			
		betont	unbetont	gesamt	hart	weich	gesamt	
silben-bildend	а	5,19	5,60	10,79	т	5,57	2,02	7,59
	ъ		8,28	8,28	н	3,9	2,27	6,17
	и	1,53	3,79	5,32	с	3,03	1,53	4,56
	э		4,67	4,67	л	2,02	1,73	3,75
	ы	1,41	2,51	3,92	р	2,24	1,46	3,7
	о	3,56	0,29	3,85	в	2,72	0,89	3,61
	э	3,23	0,20	3,43	к	2,88	0,38	3,26
	ы	0,73	1,36	2,09	д	1,59	1,36	2,95
	ges.	15,65	26,70	42,35	п	2,36	0,53	2,89
nicht silben-bildend	й			4,12	м	2,08	0,64	2,72
					ш	1,66	0,86	2,52
					б	1,2	0,53	1,73
					ч		1,69	1,69
					з	1,15	0,37	1,52
					х	1,01	0,02	1,03
					ц	1,03		1,03
					г	0,89	0,07	0,96
					ф	0,84	0,09	0,93
					ж	0,85	0,03	0,88
				γ	0,04		0,04	
	ges.	15,65	26,70	46,47	ges.	37,06	16,47	53,53

zur russischen Syntax. Peterson zufolge lässt sich ein umfassendes Bild der russischen Syntax nur durch die detaillierte Beschreibung spezieller syntaktischer Erscheinungen erhalten; dies wiederum sei nicht ohne Kenntnis der Verbreitung dieser syntaktischen Phänomene möglich, was seinerseits die Anwendung statistischer Verfahren notwendig mache (Peterson 1928, 410). Nachdem er in der erwähnten Studie 579 Vorkommnisse der Präposition *з* in prosaischen und poetischen Texten von Lermontov untersucht und deren syntaktische Einbettung quantitativ bestimmt hatte, stellte sich für ihn abschließend die Frage nach einem synchronen und diachronen Vergleich, die sich wiederum eben durch die Methode der statistischen Beschreibung lösen lasse.

Etwa zehn Jahre später erstellte Peterson (1936/37) eine zweite quantitative Studie, in der es um eine Analyse der russischen Satzstruktur ging. Peterson unterstellt dabei zunächst, dass ein Satz (a) aus einem einzigen Wort, (b) aus der Verbindung von zwei oder mehreren Wörtern, oder (c) aus der Verbindung von zwei oder mehreren Wortfügungen bestehen kann. Weiterhin ging er von der Annahme aus, dass der quantitative Bestand eines Satzes um so geringer sei, je einfacher, ungekünstelter die Rede sei, und ge-

langte so zu der Hypothese, dass die Satzstruktur in Hinsicht auf ihren quantitativen Bestand in Texten der Alltagssprache einfacher sei als in schriftlichen, literarischen Texten. Um diese Hypothese zu testen, verglich er den syntaktischen Bestand des Igorlieds [*Slovo o polku Igoreve*] mit demjenigen in Texten vorrevolutionärer Klassiker sowie in mündlichen Texten russischer Dorfbewohner. Im Ergebnis stellte sich heraus, dass 1-Wort-Sätze in der mündlichen Rede (8 %) deutlich häufiger vorkommen als in den schriftlich-literarischen Texten (1,5 % bzw. 0,6 %). In der mündlichen Rede hingegen kamen syntaktische Strukturen in Form von Wortverbindungen deutlich häufiger vor (79 % im Intervall von 2 bis 5 Wörtern) als in den literarischen Texten (42 % bzw. 44,2 % im Intervall von 2 bis 20 Wörtern), während komplexere Verbindungen von Wortfügungen wiederum in der mündlichen Rede deutlich seltener vorkamen (13 % im Umfang von zwei bis vier Wortfügungen) als in den literarischen Texten (56,5 % bzw. 55,2 % im Umfang von zwei bis 10 Wortfügungen). Zusammenfassend hält Peterson (1936/37, 592) fest, dass die Syntax des Igorlieds als Syntax der Literatursprache synthetischen Charakters sei und insofern „Züge von großer Künstlichkeit“ (Peterson 1936/37, 592) trage.

Ein weiterer unmittelbarer Fortunatov-Schüler war der bereits erwähnte Kaukasologe *Nikolaj F. Jakovlev* (1892–1974). Er veröffentlichte 1928 einen Aufsatz mit dem Titel „Eine mathematische Formel zur Konstruktion des Alphabets“ (Jakovlev 1928). Jakovlev, Gründungsmitglied des MLK, war aus der Fortunatov-Tradition hervorgegangen und hatte 1916 die Moskauer Universität abgeschlossen. Seine Arbeiten waren von großer Bedeutung für die Herausbildung der Moskauer Phonologischen Schule, als einer deren Mitbegründer er neben R. I. Afanesov, P. S. Kuznecov, V. N. Sidorov und A. M. Suchotin angesehen wird; gemeinsam mit N. S. Trubeckoj und R. O. Jakobson gilt er als einer der Begründer der strukturellen Phonologie.

In der genannten Studie ging es um die Konstruktion eines wissenschaftlich begründeten Alphabets, d. h. um die Bestimmung des optimalen Verhältnisses zwischen der Graphem- und der Phonemzahl einer Sprache. Eine solche Frage hatte auch schon früher Baudouin de Courtenay (1912, 67) aufgeworfen, als er die Bestimmung des Zusammenhangs von Graphemen und Phonemen als Grundfrage einer Theorie der Graphik bezeichnet und erklärt hatte, dass der Zusammenhang in einer mathematischen Formel auszudrücken sei.

Jakovlev (1928) ging in seiner Darstellung davon aus, dass es nicht ökonomisch sei, wenn eine Sprache ein Alphabet mit genau so vielen Graphemen wie vokalischen und konsonantischen Phonemen habe. Statt dessen sei es möglich und sinnvoll, das Grapheminventar zu reduzieren; dazu ging er von den folgenden Grundgrößen aus:

- A Anzahl der Zeichen des Alphabets (Inventarumfang)
- C Anzahl der (selbständigen) Konsonantenphoneme
- V Anzahl der (selbständigen) Vokalphoneme
- C' Anzahl der in einem bestimmten Merkmal voneinander verschiedenen konsonantischen Phonempaare
- V' Anzahl der in einem bestimmten Merkmal voneinander verschiedenen vokalischen Phonempaare

Die quantitative Optimierung eines Grapheminventars (im Vergleich zum Phoneminventar) lässt sich nach Jakovlev dadurch erreichen, dass die Anzahl der Zeichen für die selbständigen Phoneme um die Diffe-

renz zwischen den vokalischen und konsonantischen Phonempaaren reduziert wird. Das notwendige Grapheminventar lässt sich nach Jakovlev somit nach der folgenden Formel berechnen, die er als „maximal ökonomische Konstruktionsformel“ bezeichnete:

$$A = (C + V) - (\pm C' \pm V') + 1 \quad (11)$$

In Anwendung auf das Russische erhält man nach Jakovlev die folgenden Daten:

- (1) Die selbständigen Konsonanten und Vokale belaufen sich auf $C = 33$ bzw. $V = 5$.
- (2) In die Berechnung von C' gehen die harten und weichen Konsonanten ein, die sich durch das Merkmal der Palatalisierung voneinander unterscheiden; da Jakovlev sich bei der Behandlung von $g-g'$, $k-k'$ sowie $x-x'$ nicht sicher war, bezifferte er $C' = 12$ ($b-b'$, $v-v'$, $d-d'$, $z-z'$, $f-f'$, $p-p'$, $r-r'$, $s-s'$, $t-t'$, $l-l'$, $m-m'$, $n-n'$).
- (3) Unter der Annahme, dass die Buchstaben я und ы Varianten ein und desselben Phonems bezeichnen und dass das Phonem $/e/$ nur jeweils einem der beiden Glieder der 12 C' -Paare folgen kann, ergibt sich $V' = 4$ (a , i , o , u).
- (4) Die zusätzliche $+1$ am Ende der Formel soll für ein graphematisches Zeichen stehen, das dann ins Spiel kommt, wenn der zu bezeichnende Konsonant nicht von einem Vokal gefolgt wird, d. h. also, wenn dieser Konsonant vor einem anderen Konsonanten innerhalb eines Wortes oder aber am Ende eines Wortes steht; konkreter geht es um den Ausdruck der Palatalität von Konsonanten vor anderen Konsonanten bzw. am Wortende, d. h. um das Weichheitszeichen ь .

Somit ergibt sich für das Russische: $A_{\text{russ}} = (33 + 5) - (12 - 4) + 1 = 31$. Nach Jakovlev sollte also ein phonemisches Alphabet für das Russische mit 31 Zeichen, d. h. mit einem im Vergleich zum Phoneminventar von 38 Phonemen um sieben Zeichen geringeren Grapheminventar auskommen. Interessanterweise ergibt sich dasselbe Ergebnis, wenn man auch $g-g'$, $k-k'$, $x-x'$ zu den paarigen Konsonantenphonemen zählt, allerdings auf Basis einer anderen Berechnung: $A = (36 + 5) - (15 + 4) + 1 = 31$.

Diese Anzahl darf allerdings, wie auch Kempgen (1995, 18) zu Recht bemerkt, nicht als verbindlich angesehen werden, da sie aus

verschiedenen Gründen in Frage zu stellen ist; die wichtigsten Einwände dürften die beiden folgenden sein:

- (1) Berücksichtigt wird von Jakovlev nur eine einzige Sprachebene, nämlich der Inventarumfang des Alphabets – dass dieser Inventarumfang sich innerhalb einer strukturell bzw. synergetisch gedachten Konzeption auf andere sprachliche Einheiten (Wortlänge) mit ihren eigenen Systembedürfnissen auswirkt, bleibt also unberücksichtigt;
- (2) Graphem- und Lautebene werden nicht konsequent unterschieden – sonst wäre neben der Palatalitätskorrelation auch diejenige von Stimmhaftigkeit/Stimmlosigkeit zu berücksichtigen (gewesen).

Die Frage nach der Konstruktion von Alphabeten stand in den 20er Jahren in einem ganz konkreten kulturpolitischen Kontext: In Folge der Oktoberrevolution von 1917 und den anschließenden (nicht zuletzt ideologisch motivierten) Alphabetisierungsbestrebungen standen Fragen der Schrift wiederholt im Vordergrund – sei es im Hinblick auf erstmalige Verschriftlichung von Sprachen der expandierenden UdSSR, die modernisierende Modifikation schon vorhandener Schriftsprachen, oder deren akkultivierende Unifizierung. Nachdem 1922 auf Initiative der aserbajdschanischen Sowjetrepublik ein Komitee des Neuen Türkischen Alphabets [Komitet novogo tjurskogo alfavita / NTA] gegründet worden war, und nachdem 1923 das zuvor geschaffene lateinische Alphabet als „staatlich und obligatorisch“ deklariert worden war, fand 1926 in Baku der 1. Turkologische Allunionskongress statt. Auf diesem Kongress hatte Jakovlev seinen zwei Jahre später veröffentlichten Beitrag als zentralen Vortrag gehalten, den er im übrigen explizit als Beitrag zur Angewandten Linguistik bzw. zu einer Theorie der Angewandten Linguistik bezeichnet hatte. Auf dem Kongress in Baku wurde auch das Zentralkomitee des Neuen Türkischen Alphabets [Central'nyj komitet novogo tjurskogo alfavita / CK NTA] gegründet, das 1929 in Allunions-Zentralkomitee des Neuen Alphabets (VCK NA) beim Präsidium des Sowjets der Nationalitäten des Zentralen Exekutivkomitees der UdSSR umbenannt wurde.

Auch beim (damals noch von A. A. Lunačarskij geleiteten) Volkskommissariat für

Bildung war eine Subkommission unter dem Vorsitz von Jakovlev eingesetzt worden, deren Ziel die Latinisierung des russischen Alphabets war. Ein ähnliches Schicksal erlebte auch die Redaktionskommission zur Erarbeitung von Vorschlägen bezüglich einer Reform der russischen Orthographie [Redakcionnaja kommissija po predloženijam o reforme ruskoj orfografii] beim Wissenschaftlichen Forschungsinstitut für Sprachwissenschaften [NIJaz]. Die Arbeit dieser Kommission, der u. a. A. A. Reformatskij, V. N. Sidorov, A. M. Suchotin angehörten, wurde wegen der vermeintlichen Radikalität der unterbreiteten Vorschläge (wie etwa der Einführung des Buchstaben *j* ins russische Alphabet) gestoppt. Hintergrund dieser Eingriffe war letztendlich der Stalin'sche Schwenk weg von der internationalistischen Idee hin zur russisch dominierten Revolution in einem Lande. Reformatskij war im Übrigen auch Vorsitzender einer erfolgreicherer Kommission, der sog. UFT-Kommission (Kommission zur Herausarbeitung einer unifizierten phonetischen Transkription), die von 1929 bis 1931 am Wissenschaftlichen Forschungsinstitut der Vereinigung der staatlichen Verlage [NII Ogiz] unter Mitarbeit von Peterson, Vinokur u. a. wirkte.

Aleksandr A. Reformatskij (1900–1978), der u. a. bei Peterson und Ušakov studiert hatte, war zu Beginn der 20er Jahre aktives Mitglied bei *Opojaz* und MLK gewesen und hatte in diesem Zusammenhang u. a. strukturelle Analysen von Literatur und Musik vorgelegt. Die Arbeit an der Orthographie reform bezeichnete er, ähnlich wie Jakovlev, explizit als „angewandte Tätigkeit“. Am NIJaz arbeitete übrigens auch *Pëtr S. Kuznecov* (1899–1968), ein Jugendfreund des bedeutenden Mathematikers *Andrej N. Kolmogorov* (1903–1987). Zusammen mit dessen Schüler V. A. Uspenskij und Vjač. Vs. Ivanov sollte Reformatskij 1956 dann an der Moskauer Universität das erste Seminar zur mathematischen Linguistik in der UdSSR anbieten (Некоторые применения математических методов исследования в языкознании), und gemeinsam mit Kuznecov und A. A. *Ljapunov* (1911–73), ebenfalls Kolmogorov-Schüler (und Leiter des sich seit ca. 1954 mit Fragen der Sprachstatistik beschäftigenden Steklov-Instituts für Mathematik an der AN SSSR), sollte er 1956 in *Voprosy jazykoznanija* den programmatischen Artikel „Grundprobleme der maschinellen Übersetzung“ zur Diskussion bringen. Reformatskij, zu dessen Schülern u. a. T. M. Nikolaeva und R. M. Frumkina gehörten, war von 1958–70 dann Leiter des Sektors für strukturelle und angewandte Linguistik an der AN der UdSSR.

Von dieser Gruppe wurde wesentlich der *Vserossijskij s'ezd po reforme orfografii* 1931 vorbereitet, an dem noch Marristen und Etymologen gemeinsam teilnahmen und diskutierten.

Ebenfalls in diesem Kontext sind zahlreiche Arbeiten des bereits erwähnten E. D. Polivanov anzusiedeln. Polivanov, Schüler von Baudouin und Ščerba (selbst Schüler von Baudouin), studierte in Petersburg (1908–1912) und war Teilnehmer an den frühen Aktivitäten des *Opojaz*. Während und nach der Revolution engagierte er sich politisch in der russischen KP und wurde in den 20er Jahren zu einem der wichtigsten Sprachwandeltheoretiker. In Anlehnung an Baudouins Unterscheidung von äußeren und inneren Kräften als Einflussfaktoren auf den Sprachwandel trennte Polivanov strikt zwischen äußeren und inneren Gesetzen der Sprachentwicklung und dem unmittelbaren Einfluss sozialer Faktoren auf die Sprache. Gerade letzterer wurde jedoch zunehmend von dem immer mehr Einfluss nehmenden sog. Marrismus verabsolutiert: Der Schöpfer dieser sog. Neuen Lehre des Marrismus, *Nikolaj Ja. Marr* (1864–1934), vertrat, ebenso wie nach seinem Tode sein wichtigster Nachfolger (*Ivan I. Meščaninov*, 1883–1967), die Auffassung, dass Entstehung und Entwicklung von Sprache in unmittelbarer Abhängigkeit von den (jeweiligen) sozio-ökonomischen Verhältnissen zu sehen sei, so dass sich je nach sozio-ökonomischer Entwicklungsstufe einer Gesellschaft auch die typologische Struktur einer Sprache ändere – der klassenhafte Charakter der Sprache wurde so mit einer Stadialtheorie der Sprache (in Anlehnung an die Stadialtheorie der Gesellschaftsentwicklung) verknüpft. Es liegt auf der Hand, dass diese Konzeption weniger durch wissenschaftliche Überzeugungskraft als vielmehr aufgrund der politischen Aktivität ihrer Anhänger zur herrschenden Doktrin und zu einem unantastbaren Dogma der 30er und 40er Jahre werden konnte. Dabei war der Marrismus – natürlich – extrem empiriefrem bzw. empiriefeindlich; entsprechende Forschungen wurden vollständig unterdrückt und bekämpft. Der Streit um die vermeintlich wahre marxistische Sprachwissenschaft wurde zunehmend politisiert, bis hin zur persönlichen Denunzierung der ideologischen (!) Gegner.

Die Gefahr dieser Tendenz wurde durchaus schon Mitte der 20er Jahre gesehen; so hatte z. B. der

seit 1922 in Wien lebende und lehrende, als Porzeziński-Schüler ursprünglich in der Tradition der Moskauer Schule stehende *Nikolaj S. Trubeckoj* (1890–1938) unmittelbar nach der Lektüre von Marrs Aufsatz „O jafetičeskoj teorii“ (1924) am 6.11.1924 an Jakobson geschrieben: „Der Aufsatz von Marr übersteigt das Maß dessen, was er bis jetzt geschrieben hat ... Ich bin fest davon überzeugt, daß eine kritische Besprechung dieses Aufsatzes nicht von einem Linguisten, sondern von einem Psychiater gemacht werden müßte. Zum Schaden für die Wissenschaft ist Marr noch nicht verrückt genug, um in eine Anstalt gesperrt zu werden, aber mir ist klar, daß er verrückt ist.“ (Jakobson 1977, 281).

In Trubeckoj's 1939 erscheinener *Einführung in die Phonologie* (Trubeckoy 1939) ist im übrigen ein ganzes Kapitel mit dem Titel „Zur phonologischen Statistik enthalten“, in der er nicht nur zwischen der Statistik der Buchstaben, der Laute und der Phoneme sowie Phonemverbindungen differenzierte, sondern auch Angaben zu Vorkommenshäufigkeiten von (in der Anzahl der Silben pro Wort berechneten) Wortlängen machte. – Im hier gegebenen Zusammenhang würde es allerdings zu weit führen, auch quantitativen Arbeiten von aus Russland bzw. aus der UdSSR Emigrierten nachzugehen, was in gleicher Weise für Jakobson und andere gilt.

Das betraf – neben anderen Sprachforschern wie Peškovskij, Peterson, Ščerba, Ušakov, u. a. – auch und gerade Polivanov, der mehr als 20 Jahre aktiv in die Frage der Latinisierung des Alphabets involviert war. 1927 zum Vorsitzenden der Linguistischen Sektion des Instituts für Sprache und Literatur (RANION) ernannt, nahm er u. a. auch an dem oben genannten Kongress in Baku teil und wurde 1928 in den Wissenschaftlichen Rat des CK NTA aufgenommen. Nachdem Polivanov sich schon 1929 in einem Vortrag mit dem Titel „Das Problem einer marxistischen Sprachwissenschaft und die japhetische Theorie“ kritisch mit dem Marrismus auseinandergesetzt hatte, enthielt auch sein – in einem Sammelband mit dem Titel *Für eine marxistische Sprachwissenschaft* 1931 veröffentlichter – Aufsatz „Auch die Mathematik kann nützlich sein“ (Polivanov 1931) entsprechende methodologische Kritik. Durchaus eingestehend, dass die Mathematik nicht gerade seine Disziplin sei, räumte Polivanov statistischen Methoden wichtige Funktionen in drei Bereichen der Sprachwissenschaft ein:

- (1) in der Experimentalphonetik
- (2) in der dialektologischen Statistik
- (3) in der Etymologie.

Insbesondere im Bereich der Etymologie war der Konflikt mit dem Marrismus vorprogrammiert, ging dieser doch im Rahmen seiner sog. Vier-Elemente-Theorie davon aus, dass alle Wörter sämtlicher Sprachen der Erde auf vier Wurzeln (*sal, *ber, *jonb, *roš) zurückgeführt werden können: Einen wahrscheinlichkeitstheoretischen Ansatz sowohl auf die centum-çatám-Frage als auch allgemein auf das Prinzip des japhetidologischen Ansatzes bei der etymologischen Rekonstruktion anwendend, kam Polivanov (1931) zu der Schlussfolgerung, „dass die gegebene marr'sche Etymologie offenkundig unrichtig ist“ (Polivanov 1931, 294). Seither hatte Polivanov zunehmend Schwierigkeiten, seine Arbeiten zu publizieren; 1937 wurde er verhaftet, 1938 zum Tode verurteilt und erschossen. Erst 1963 wurde er vollständig rehabilitiert.

Die sowjetische Sprachwissenschaft jedoch war seit den 30er Jahren großteils von der internationalen Linguistik isoliert. Als Stalin selbst 1950 mit seiner Abhandlung *Marxismus und Fragen der Sprachwissenschaft* in die Diskussion eingriff und den Marrismus als falsch verurteilte, befreite er zwar die sowjetische Sprachwissenschaft von dem einen Dogma, schuf aber gleichzeitig ein neues, indem er jegliche soziologisch bedingte Heterogenität der Sprache kategorisch ablehnte. So wurde erst nach seinem Tod (1954) bzw. nach dem XX. Parteitag der KPdSU (1956) prinzipiell eine Neu-Diskussion der „Aktuellen Aufgaben der sowjetischen Sprachwissenschaft“ möglich, bei der insbesondere im Zusammenhang mit der sog. Strukturalismus-Diskussion von 1957 auch quantitativen Methoden eine neue Entfaltungsmöglichkeit eingeräumt wurde (vgl. Art. Nr. 4).

4. Resümee

Aus der obigen Darstellung ergeben sich eine Reihe von Befunden, Schlussfolgerungen und Perspektiven, die sich im Wesentlichen wie folgt resümieren lassen:

- (1) Die Anwendung quantitativer Verfahren bezieht sich in Russland seit den ersten, ab Mitte des 19. Jh.s entstandenen Arbeiten nie nur auf den engeren Bereich der Sprache bzw. Sprachwissenschaft, sondern von jeher gleichermaßen auch auf den Bereich der Literatur(wissenschaft);

- (2) die (Vor-)Geschichte der Anwendung quantitativer Verfahren auf Sprache und Literatur weist in der russischen Tradition einen Reichtum an Arbeiten auf, wie er andernorts in dieser Vielfalt kaum zu beobachten ist;
- (3) nicht zuletzt aufgrund der Vielfalt der seit Ende des 19., insbesondere Anfang des 20. Jh.s geleisteten Arbeiten sind in der russischen Sprach- und Literaturwissenschaft Grundlagen für quantitative Untersuchungen gelegt worden, die einen Bezugs- und Anknüpfungspunkt für spätere Untersuchungen darstellen, wie sie auf internationaler Ebene ab Mitte der 50er Jahre die sich dann herausbildende quantitative Linguistik und quantitative Stilistik charakterisieren;
- (4) aufgrund der äußeren Umstände stellt die russische Tradition keine gradlinige Entwicklung dar; sie weist zwar inhaltliche und personelle Kontinuitäten auf, ist aber ebenso von Schwankungen und Kontinuitätsbrüchen gekennzeichnet, im Zusammenhang mit denen verlorenes Terrain immer wieder neu erarbeitet werden musste.

Die sich aus der historischen Darstellung ergebenden Fakten und Zusammenhänge lassen bestimmte Entwicklungen und gegenwärtige Trends in einem deutlicheren Licht erscheinen — sie weisen aber auch auf die Notwendigkeit hin, die historiographische Archäologie quantitativer Verfahren in Sprach- und Literaturwissenschaft weiterhin zu vertiefen.

5. Literatur (in Auswahl)

Altmann, Gabriel (1980), Prolegomena to Menzerrath's Law. In: *Glottometrika* 2. (Hrsg. R. Grotjahn). Bochum: Brockmeyer, 1–10.

Baudouin de Courtenay, Jan (1877), *Otčety komandirovannago Ministerstvom Narodnago Prosvěščenija za granicu s učenoju celju I. A. Boduënade-Kurtené o zanjatjach po jazykovedeniju v tečenie 1872 i 1873 gg.* Kazan'.

Baudouin de Courtenay, Jan (1909), *Zarys historii językoznawstwa czyli lingwistyki (glottologii)*. Warszawa.

Baudouin de Courtenay, Jan (1912), *Ob omošenii russkago pis'ma k russkomu jazyku*. S.-Peterburg.

Baudouin de Courtenay, Jan (1927), *Ilościowość w myśleniu językowym*. In: *Symbolae Grammaticae in honorem Joannis Rozwadowski, t. 1*. Kra-

- ków, 1929, 3–18. [Zit. n.d. russ. Übersetzung: Kolichestvennost' v jazykovom myšlenii. In: Boduën de Kurtene, I. A. (1963), *Izbrannye trudy po obščemu jazykoznaniju. Tom II*. Moskva, 311–324.]
- Belyj, Andrej (1910), *Simvolizm*. Moskva. [Repr.: München, 1969]
- Belyj, Andrej (1929), *Ritm kak dialektika*, Moskva. [Repr.: Chicago 1968]
- Bem, Al'fred L. (1918/21), K ujasneniju istoriko-literaturnych ponjatij. In: *Izvestija otdelenija ruskogo jazyka i slovesnosti Rossijskoj akademii nauk, XXIII/1*, 225–245.
- Bobrik, A. A. (1921), *Nemnogo matematiki v teorii slovesnosti*. Gel'singfors.
- Bobrov, Sergej P. (1915), *Novoe o sticholoženii Puskina*. Moskva.
- Bobrov, Sergej P. (1916), *Zapiski stichotvorca*. Moskva. [Repr.: Letchworth (GB), 1973]
- Bobrov, Sergej P. (1921), [Rez.:] G. Šengeli, *Traktat o stiche*. Odessa, 1921. In: *Pečat' i revoljucija* 3, 268–269.
- Bobrov, Sergej P. (1964a), Opyt izucenija vol'nogo sticha Puškinskich „Pesen zapadnych slavjan“. In: *Teorija verojatnostej i ee primenenija, IX/2*, 262–272.
- Bobrov, Sergej P. (1964b), K voprosu o podlinnom stichotvornom razmere puškinskich „Pesen zapadnych Slavjan“. In: *Russkaja literatura* 3, 119–137.
- Bobrov, Sergej P. (1965), Tesnota stichovogo rjada (Opyt statističeskogo analiza literaturovedčeskogo ponjatija, vvedennogo Ju. N. Tynjanovym). In: *Russkaja literatura* 3, 109–124.
- Bobrov, Sergej P. (1967), Russkij toničeskij stich s ritmom neopredelennoj četnosti i var'irujuščej sillabikoj. In: *Russkaja literatura* 3, 42–64.
- Bogorodickij, Vasilij A. (1908), *Osnovy arifmetičeskogo sčeta*. Kazan'.
- Bogorodickij, Vasilij A. (1935), *Obščij kurs ruskoj grammatiki (iz universiteskich čtenij)*. Moskva.
- Budilovič, Anton S. (1883), *Načertanie cerkovno-slavjanskoj grammatiki primenitel'no k obščej teorii ruskago i drugih rodstvennych jazykov*. Varšava.
- Bunjakovskij, V. Ja. (1847), O vozmožnosti vvedenija opredelitel'nych mer doverija k rezul'tatam nekotorych nauk nabljudatel'nych i preimuščestvenno statistiki. In: *Sovremennik III/2*, 36–49.
- Černyševskij, Nikolaj G. (1906a), [Rez.:] Sočinenija Puškina. In: *Polnoe sobranie sočinenij. Tom I: Kritika i bibliografija*. (Hrsg. N. G. Černyševskij). Sankt Peterburg, 1906, 245–330. [1. Aufl. 1855]
- Černyševskij, Nikolaj G. (1906b), [Rez.:] Grammatičeskija zametki. In: *Polnoe sobranie sočinenij. Tom I: Kritika i bibliografija*. (Hrsg. N. G. Černyševskij). Sankt Peterburg, 1906, 359–364. [1. Aufl. 1855]
- Čistjakov, Vasilij F. (1939), *Slovar' komedii „Gore ot uma“ A. S. Griboedova*. Smolensk.
- Čistjakov, Vasilij F. (1964), Pis'mo v redakciju. In: *Filologičeskie nauki* 2, 208–209.
- Čistjakov, Vasilij F. (1965), *K istorii primenenija statističeskich metodov v jazykoznanii*. Voronež.
- Čistjakov, Vasilij F. (1971), Častoti glasnych i soglasnych v jazykach raznych sistem. In: *Slavia. Časopis pro slovanskou filologii* 40, 204–205.
- Čistjakov, Vasilij F. (1972a), Častotnosti glasnych i soglasnych v 50 jazykach raznogo grammatičeskogo stroja. In: *Lingua Posnaniensis* 16, 45–48.
- Čistjakov, Vasilij F. (1972b), Predlogi v ruskom jazyke. In: *Materialy po rusko-slavjanskomu jazykoznaniju*. Voronež, 171–179.
- Čistjakov, Vasilij F./Kramarenko, B. K. (1929), *Opyt priloženija statističeskogo metoda k jazykoznaniju. Vyp. I*. Krasnodar.
- Čudovskij, Valerian (1914), O ritme puškinskoi „Rusalki“ (Otryvok). In: *Apollon* 1–2, 108–121.
- Čudovskij, Valerian (1915), Neskol'ko myslej k vozmožnomu učeniju o stiche (s primernym, razborom stichosloženija v I glave „Evgenija Onegina“). In: *Apollon* 7–8, 55–95.
- Čudovskij, Valerian (1917), Neskol'ko utverždenij o ruskom stiche. In: *Apollon* 4–5, 58–69.
- Denisov, Ja. (1888), *Osnovy metriki u drevnich grekov i rimljan*. Moskva.
- Dikarev, Mitrofan A. (1891), *Voronežski etnografičeskij sbornik*. Voronež.
- Erlich, Viktor (1973), *Russischer Formalismus*. Frankfurt/M.: Suhrkamp. [1. Aufl. 1955]
- Ermolenko, Georgij V. (1974), O prologe k komedii A. S. Griboedova „Gore ot uma“. In: *Voprosy statističeskoj stilistiki*. Kiev, 251–262.
- Förstemann, Ernst (1846), Über die numerischen Lautverhältnisse im Deutschen. In: *Germania. Herausgegeben von der Berlinischen Gesellschaft für deutsche Sprache und Altherthumskunde* 7, 83–90.
- Förstemann, Ernst (1852), Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen. In: *Germanische Zeitschrift für Vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 1, 163–179.
- Fortunatov, Filipp F. (1956), *Sravnitel'noe jazykovedenie Obščij kurs*. In: *Izbrannye trudy. Tom I*. (Hrsg. F. F. Fortunatov). Moskva, 21–197.
- Frumkina, Revekka M. (1987), Moj učitel' A. A. Reformatskij. In: *Znanie – sila* 7, 78–82.
- Gornfel'd, Arkadij (1922), Chudožestvennoe slovo i naučnaja cifra. In: *Literaturnaja mysl* 1, 163–170.
- Grzybek, Peter (1989), *Studien zum Begriff der sowjetischen Semiotik (Moskauer und Tartuer Schule)*. Bochum: Brockmeyer.

- Grzybek, Peter (1999), Sowjetische und russische Konzepte der Semiotik. In: *Handbuch der sprachwissenschaftlichen Russistik und ihrer Grenzdisziplinen*. (Hrsg. H. Jachnow). Wiesbaden: Harrassowitz, 1274–1305.
- Grzybek, Peter (2003), Viktor Ja. Bunjakovskij (1804–1889) – A Forerunner of Quantitative Linguistics in Russia. In: *Glottometrics* 6, 103–106.
- Grzybek, Peter (2004), Nikolaj Gavrilovič Černyševskij (1828–1889) – A Forerunner of Quantitative Stylistics in Russia. In: *Glottometrics* 7, 91–93.
- Grzybek, Peter/Kelih, Emmerich (2003), Graphemhäufigkeiten (Am Beispiel des Russischen). Teil I: Methodologische Vor-Bemerkungen und Anmerkungen zur Geschichte der Erforschung von Graphemhäufigkeiten im Russischen. In: *Anzeiger für Slavische Philologie* 31; 131–162.
- Grzybek, Peter/Kelih, Emmerich (2004), Anton Seměnovič Budilovič (1846–1908) – A Forerunner of Quantitative Linguistics in Russia? In: *Glottometrics* 7, 94–100.
- Grzybek, Peter/Kelih, Emmerich/Altmann, Gabriel (2004), Graphemhäufigkeiten. Teil II: Theoretische Modelle der Häufigkeitsverteilung (mit einer empirischen Untersuchung russischer Graphemhäufigkeiten). In: *Anzeiger für Slavische Philologie* 32, 25–54.
- Hansen-Löve, Aage (1978), *Der russische Formalismus: Methodologische Rekonstruktion seiner Entwicklung aus dem Prinzip der Verfremdung*. Wien: Verlag der Österr. Akademie der Wiss.
- Jakobson, Roman O. (1977), Autobiographische Notizen von N. S. Trubetzkoy. In: *Grundzüge der Phonologie*. (Hrsg. N. S. Trubetzkoy) 6. Aufl. Göttingen: Vandenhoeck & Ruprecht, 273–288.
- Jakovlev, Nikolaj F. (1928), Matematičeskaja formula postroenija alfavita. In: *Kul'tura i pis'mennost' Vostoka, kn. I*. Moskva, 41–64. [Repr.: Reformatskij, A. A. (Hrsg.), *Iz istorii otečestvennoj fonologii*. Moskva, 1970, 123–148].
- Kaeding, Friedrich W. (1897/98), *Häufigkeitwörterbuch der deutschen Sprache*. Berlin-Steglitz: Selbstverlag.
- Kagarov, Evgenij (1922), O ritme russkoj prozaičeskoj reči. In: *Nauka na Ukraine* 4, 324–332.
- Kempgen, Sebastian (1995), *Russische Sprachstatistik. Systematischer Überblick und Bibliographie*. München: Sagner.
- Kempgen, Sebastian (1999), Quantitative Aspekte. In: *Handbuch der sprachwissenschaftlichen Russistik und ihrer Grenzdisziplinen*. (Hrsg. H. Jachnow). Wiesbaden: Harrassowitz, 525–550.
- Kopeckij, Leontij (1931), [Rez.:] Čistjakov, V. F./Kramarenko, B. K. (1929): Opyt priloženija statističeskogo metoda k jazykoznaniju. Krasnodar. In: *Slavia* 10, 169–171.
- Kudrjavskij, Dmitrij (1909), K statistike glagol'nych form v Lavrent'evskoj letopisi. In: *Izvestija otdelenija russkago jazyka i slovesnosti Imperatorskoj Akademii Nauk, t. XIV, č. 2*, 49–56.
- Kudrjavskij, Dmitrij (1911), K istorii russkago proeđščago vremeni. In: *Russkij filologičeskij vestnik LXV*, 119–139.
- Kudrjavskij, Dmitrij (1912), Drevne-russkija pričastija nastojaščago vremeni dejstvitel'nago zaloga na -a. In: *Russkij filologičeskij vestnik LXVIII*, 119–139.
- Kunickij, V. N. (1894), *Jazyk i slog komedii „Gore ot uma“*. Kiev.
- Luk'janov, S. M. (1914), „Angel' smerti“ gr. A. A. Goleniščeva-Kutuzova. In: *Žurnal ministerstva narodnago prosvěščenija XLIX*, 316–352.
- Markov, Andrej A. (1913), Primer statističeskogo izsledovanija nad tekstem „Evgenija Onegina“ illjustrirujuščij svjaz' ispytanij v cep'. In: *Izvestija Imperatorskoj Akademii Nauk / Bulletin de l'Académie Impériale des Sciences de St.-Petersbourg, ser. VI, t. 7, no. 3*, 153–162.
- Markov, Andrej A. (1916), Ob odnom primenenii statističeskogo metoda. In: *Izvestija Imperatorskoj Akademii Nauk / Bulletin de l'Académie Impériale des Sciences, ser. VI, t. 10, no. 4*, 239–242.
- Morozov, Nikolaj A. (1915), Lingvističeskie spektry. In: *Izvestija otdelenija russkago jazyka i slovesnosti imperatorskoj akademii nauk XX(1–4)*, 95–127.
- Nedobrovo, N. V. (1912), Ritm, metr i ich vzaimootnošenie. In: *Trudy i dni* 2, 14–23.
- Ol'chin, P. (1907), Pervaja opora pri postroenii racional'noj stenografii. In: *Stenograf. Ežemesjačnyj žurnal, posvjaščennyj voprosam naučnoj i praktičeskoj stenografii, no. 4–5*, 114–118.
- Ondar, Kh. O. (Hrsg.), *The Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. New York u. a.: Springer, 1981.
- Papp, Ferenc (1966), *Mathematical Linguistics in the Soviet Union*. London u. a.: Mouton.
- Peškovskij, Aleksandr M. (1914), *Russkij Sintaksis v naučnom osveščeni*. Moskva, 51935.
- Peškovskij, Aleksandr M. (1924), Stichi i proza [s lingvističeskoj točki zrenija]. In: *Svitok. Al'manach* 3. Moskva/Leningrad, 197–223.
- Peškovskij, Aleksandr M. (1925), Desjat' tysjač zvukov. (Opyt zvukovoj charakteristiki russkogo jazyka, kak osnovy dlja eufoničeskich issledovanij). In: *Metodika rodnogo jazyka, lingvistika, stilistika poetika. Sbornik statej*. (Hrsg. A. M. Peškovskij). Leningrad/Moskva, 167–191.
- Peterson, Michail N. (1928), Konstrukcii s predlogom „iz“ u Lermontova. In: *Sbornik statej v čest' akademika Alekseja Ivanoviča Sobolevskogo*. Leningrad, 410–412. [= *Sbornik otdelenija russkago jazyka i slovesnosti* 101, 3]

Peterson, Michail N. (1936/37), Sintaksis „Slova o polku Igoreve“. In: *Slavia. Casopis pro slovan-skou filologii* 14, 547–592.

Polivanov, Evgenij D. (1931), I matematika možet byt' poleznoj. In: *Za marksistkoe jazykoznanie*. Moskva, 173–181. [Repr.: *Stat'i po obščemu jazykoznaniju*. Moskva, 287–294.]

Proskurnin, N. (1933), Podščety častoty liter i komplektovka šrifta. In: *Revoljucija i pis'mennost'*. Sbornik I. Moskva-Leningrad, 72–82.

Schleicher, August (1852), *Die Formenlehre der kirchenslawischen Sprache, erklärend und vergleichend dargestellt*. Bonn: König.

Setnickij, Nikolaj (1922), *Statistika, literatura i po-čezija. K voprosu o plane issledovanija*. Odessa.

Setnickij, Nikolaj (1923), *Statistika, literatura i po-čezija. K voprosu o plane issledovanija*. Odessa.

Steiner, Peter (1984), *Russian Formalism. A Meta-poetics*. Ithaca/London: Cornell University Press.

Struve, Petr B. (1918), Kto pervyj ukazal na primenenie statistiki k filologičeskim izsledovanijam? In: *Izvestija Rossijskoj Akademii Nauk [Bulletin de l'Académie des Sciences de Russie]*, t. XII, VI ser., no. 12–18, čast' 2, 1317–1318.

Ščerbina, Fedor A. (1925), *Statistika. Istorija statistiki i statističnich ustanov*. Praga.

Šengeli, Georgij (1918), *Dva „Pamjatnika“. Sravnitel'nyj razbor ozaglavlennyh etim imenem stichtvorenij Puškina i Brjusova*. Petrograd.

Šengeli, Georgij (1923), *Traktat o russkom stiche*. Moskva, 1921.

Šengeli, Georgij (1940), *Technika sticha*. Moskva, 1960.

Tomaševskij, Boris V. (1916a), [Rez.:] S. Bobrov (1915), *Novoe o stichosloženii A. S. Puškina*. Moskva. In: *Apollon* 10, 72–74.

Tomaševskij, Boris V. [1916b](1990), Pis'ma B. V. Tomaševskogo k S. P. Boborvu. [Publikacija, vstupil'naja stat'ja i primečanja K. Ju. Postoutenko]. In: *Pjatye Tynjanovskie čtenija*. Riga, 133–148.

Tomaševskij, Boris V. (1917), Ritmika četyrechstopnogo jamba po nabljudenijam nad stichom „Evgenija Onegina“. In: *Tomaševskij* 1929, 94–137.

Tomaševskij, Boris V. (1918a), Statistika kratkosti povreždenij telegrafnyh provodov. In: *Počtovo-*

telegrafnyj žurnal 5–8 [čast' neoficial'naja], 33–44.

Tomaševskij, Boris V. [1918b](1977), O šestistopnom jambe u Puškina. In: *Trudy po znakovym sistemam* 9. Tartu, 103–112.

Tomaševskij, Boris V. (1919a), Statistika počt i telegrafov v svjazi s novymi zadačami gosudarstvennogo stroitel'stva. In: *Proletarij svjazi* 1, 23–24.

Tomaševskij, Boris V. (1919b), Pjatisťopnyj jamb Puškina. In: *Tomaševskij* 1929, 138–253.

Tomaševskij, Boris V. (1920), Ritm prozy. In: *Tomaševskij* 1929, 254–318.

Tomaševskij, Boris V. (1922), [Rez.:] Bobrik, A. A. (1921), *Nemnogo matematiki v teorii slovesnosti*. Gel'sinfgors. In: *Kniga i revoljucija* 3/15, 84.

Tomaševskij, Boris V. (1923a), Problema stichotvornogo ritma. In: *Literaturnaja mysl* 2, 124–140. [Repr. in.: *Tomaševskij, Boris V. (1929), O stiche*. Leningrad, 3–36.]

Tomaševskij, Boris V. (1923b), *Russkoe stichosloženie*. Petrograd. [Repr. München: Fink, 1971]

Tomaševskij, Boris V. (1923b), [Rez.:] Šengeli, G. A. (1921), *Traktat o rusckome stiche*. Odessa. [2. Aufl. 1923]. In: *Kniga i revoljucija*, 1/25, 51–52.

Tomaševskij, Boris V. (1924), Valerij Brjusov kak stichoved. In: *Tomaševskij* 1929, 319–325.

Tomaševskij, Boris V. (1929), *O stiche*. Leningrad. [Repr.: München: Fink, 1970]

Tomson, Aleksandr I. (1910), *Obščee jazykoznanie*. Odessa.

Trubetzkoy, Nikolaj S. (1939), *Einführung in die Phonologie*. Prague. [Grundzüge der Phonologie. Göttingen: Vandenhoeck & Ruprecht, 1977].

Vinokur, Grigorij O. (1924), [Rez.:] (1923) *Očerki po poëtike Puškina*. Berlin. In: *Russkij Sovremennik* 3, 263–264.

Vostokov, Aleksandr Ch. (1817), *Opyt o rusckom stichosloženii*. Sankt Peterburg.

Žirmunskij, Viktor M. (1929), Po povodu knigi „Ritm kak dialektika“. In: *Zvezda* 8, 203–208.

Peter Grzybek/Emmerich Kelih,
Graz (Österreich)