

Sebastian Kempgen, Bamberg

Quantitative Aspekte des Russischen

0. Vorbemerkungen

Die russische Sprache ist – wie jede menschliche Sprache – auch ein Massenphänomen, dessen Eigenschaften und Erscheinungen sich folglich mit den speziellen Methoden untersuchen lassen, die für derartige Aspekte eines Objektes in den empirischen Wissenschaften entwickelt worden sind. Im Gegensatz zu den meisten anderen Beiträgen dieses Handbuches ist Gegenstand des vorliegenden Beitrages also nicht eine bestimmte sprachliche Ebene oder Erscheinung, sondern ein besonderer methodischer Zugang zur Beschreibung der Strukturen des Russischen. Hierfür einen eigenen Beitrag vorzusehen, scheint u.a. auch deshalb gerechtfertigt, weil die statistische Beschreibung einer Sprache sich zu einer eigenen linguistischen Disziplin entwickelt hat, deren Vorgehen und Ergebnisse in die üblichen Darstellungen gewöhnlich nicht ohne weiteres, im vorliegenden Falle jedenfalls nicht ohne häufige Wiederholungen eingebaut werden könnten.

Der Beitrag versucht deshalb, eine Vorstellung davon zu vermitteln, von welchen Grundgedanken die quantitative Linguistik (russ: *квантитативная лингвистика, вычислительная лингвистика, лингвостатистика, статистика речи*) ausgeht, welche typischen Fragestellungen sie verfolgt, und zu welchen Ergebnissen sie in Anwendung auf das moderne Russisch bisher gekommen ist. Daß diese Skizze sich dabei aus Platzgründen des öfteren auf allgemeine Angaben beschränken muß und nicht immer in die Details gehen kann, versteht sich von selbst. Eine vertiefte und wesentlich umfangreichere Darstellung, die auch sprachgeschichtliche Untersuchungen berücksichtigt und eine umfangreiche Bibliographie enthält, ist bereits an anderer Stelle separat veröffentlicht worden (KEMPGEN 1995).

Der vorliegende Beitrag setzt keinerlei spezielle Vorkenntnisse voraus, sondern versucht vielmehr, den „normalen“ Russisten anzusprechen, der sich über einige wesentliche quantitative Aspekte des Russischen informieren will. Der Aufbau folgt dabei zunächst dem strukturalistischen Schichtenmodell der Sprache, woran sich weitere, speziell angewandte, Bereiche anschließen.

1. Grundannahmen und Methoden der Sprachstatistik

Die quantitative Linguistik (vgl. etwa ALTMANN 1972, 1973 oder HOFFMANN/PIOTROWSKI 1979) interessiert sich, allgemein gesprochen, für den großen Bereich der statistischen Zusammenhänge, der Tendenzen in der Sprache, denen die ausnahmslos gültigen oder kategorischen Zusammenhänge als kleinere Gruppe der sprachlichen Regeln gegenüberstehen. Da sich Elemente und deren Anordnung besonders gut zählen lassen, sind sprachstatistische Arbeiten explizit oder unreflektiert oft den Grundannahmen des sprachwissenschaftlichen Strukturalismus verpflichtet. Im Prinzip kann aber jeder Beschreibungsansatz quantitative Methoden verwenden. Ihre Fragestellungen untersucht die Sprachstatistik entweder am Lexikon („systemische Stichprobe“) oder an Texten („pragmatische Strichprobe“), und damit entweder die Elemente des Inventars (*langue*) oder deren Häufigkeit im Gebrauch (*parole*), wobei selbstverständlich die synchrone wie die diachrone Perspektive benutzt werden kann. Dabei geht es ihr um innere wie äußere Eigenschaften der Objekte; zu den ersteren gehört die Frage, wie sich eine sprachliche Einheit aus Elementen der nächstniedrigeren Ebene zusammensetzt, zu den letzteren die Frage, wie sich Elemente miteinander zu höheren Einheiten kombinieren. Neben das Zählen absoluter Frequenzen tritt die Angabe der relativen Häufigkeiten, die die Werte verschiedener Zählungen vergleichbar macht. Um die Ergebnisse linguistisch sinnvoll interpretieren zu können, bedarf es meistens eines Testes, der prüft, ob die untersuchte Variable nur zufällig verschiedene Werte angenommen hat oder ob sich hinter den Differenzen strukturelle, „signifikante“, Unterschiede verbergen. Neben diesem Bereich der deskriptiven Statistik beschäftigt sich die quantitative Linguistik aber auch damit, theoretische Verteilungen für empirische Häufigkeitsverteilungen zu finden, Sprachgesetze aus Grundannahmen abzuleiten, Zusammenhänge zwischen solchen Gesetzen zu formulieren usw., um so letzten Endes zu einer Sprachtheorie (bzw. zu Teilbereichen aus ihr) zu kommen. Der vorliegende Artikel beschränkt sich auf die sprachstatistischen, deskriptiven Anteile der einschlägigen Literatur. Als Beispiel eines Sprachgesetzes sei das sog. „Menzerathsche Gesetz“ genannt, das allgemein lautet: „Je größer (komplexer) ein sprachliches Konstrukt, desto kleiner (einfacher) seine Komponenten“. Die Paare von Konstrukt und Komponenten, die sich hier einsetzen lassen, sind z.B. Text/Textabschnitt, Satz/Teilsatz, Syntagma/Wort, Wort/Morphem oder Wort/Silbe.

Großen Wert legt die quantitative Linguistik ferner auf eine Klärung der jeweils verwendeten Begriffe: mit kategorischen Begriffen ordnet man in eine Klasse ein oder schreibt man eine Eigenschaft zu („*k* ist ein Konsonant“), mit komparativen Begriffen vergleicht man bzw. bringt man in

eine Rangfolge („*i* ist weniger offen als *e*“), mit quantitativen Begriffen schließlich mißt man („*стамучука* ist 10 Buchstaben lang“, „ist doppelt so lang wie *книга*“). Für jede dieser drei Begriffsformen gibt es bestimmte Bedingungen, die erfüllt sein müssen, damit von einer korrekten Verwendung des betreffenden Begriffes gesprochen werden kann. Ein kategorischer oder nominativer Begriff etwa muß eindeutig und erschöpfend auf seine Objektmenge angewendet werden können, wobei als Beispiel die Wortartenklassifikation des Russischen dienen mag: jedes Wort soll genau einer Wortart angehören und kein Wort darf „übrigbleiben“. Über weitere Details informiert die einschlägige wissenschaftstheoretische Literatur.

2. Wissenschaftsgeschichtliche Anmerkungen

Es ist ebenso interessant wie nützlich, sich die Entwicklung der sprachstatistischen Forschung in Rußland wenigstens in ihren Grundzügen zu vergegenwärtigen. Erste einschlägige Arbeiten finden sich bereits am Ende des 19. Jahrhunderts (BUDILOVIĆ 1883, DIKAREV 1891) und in den ersten Jahren des 20. Jahrhunderts (KUDRJAVSKIJ 1911, MOROZOV 1915). Wenn sie aus heutiger Sicht nicht alle gleichermaßen methodisch befriedigen, so muß man sich bei diesem Urteil vor Augen halten, daß zu diesem Zeitpunkt weder die strukturalistische Linguistik noch die Statistik als eigene Disziplin mit ihren spezifischen Methoden existierten. Dennoch gab es bahnbrechende Entwicklungen, wie z.B. erste Ansätze zur späteren Informationstheorie bei A. A. MARKOV 1913, der die von ihm untersuchten und später nach ihm benannten Markov-Ketten am Text des „Evgenij Onegin“ demonstrierte.

In den ersten Jahrzehnten der Sowjetzeit ist die Entwicklung und das weitere Schicksal der Sprachstatistik unmittelbar mit demjenigen der Linguistik als ganzer verknüpft. Folgerichtig finden wir zunächst interessante Arbeiten wie die große Lautstatistik von PEŠKOVSKIJ 1925, das schon seinerzeit leider wenig beachtete Büchlein von ČISTJAKOV/KRAMARENKO 1928 oder die syntaktischen Studien von PETERSON 1937. Diese Ansätze konnten jedoch nicht konsequent weiterentwickelt werden, als in der Zeit des sog. Marrismus (s. dazu an anderer Stelle dieses Handbuches) alle „formalen“ Ansätze, darunter auch die sog. Moskauer oder formale Schule ihres Hauptvertreters F. F. Fortunatov, in Bedrängnis gerieten.

Erst mit dem Zweiten Weltkrieg wendete sich das Blatt wieder, und hier konnte sich die Sprachstatistik noch vor der übrigen sprachwissenschaftlichen Forschung, die sich erst im Zuge der sog. „Strukturalismus-Diskussion“ in den 50er Jahren ihrer dogmatischen Fesseln zu entledigen vermochte, wieder betätigen, wenn auch zunächst nur im geheimen, nämlich zu militä-

rischen Zwecken: Zur Entwicklung der Nachrichtentechnik ließ Stalin gegen Ende des Zweiten Weltkrieges eine Gruppe von Philologen und Mathematikern in einem Speziallager an Lösungen zur Sprachverschlüsselung arbeiten. In den USA entstand gleichzeitig die klassische Ausarbeitung der Informationstheorie in den Arbeiten von C. E. Shannon. Einen entscheidenden „Schub“ erfuhren diese Arbeiten durch die Erfindung und Weiterentwicklung des Computers, zunächst in Form von Großrechnern, die erstens empirische – und damit auch sprachstatistische – Arbeiten erleichterten, zweitens aber auch zur Entwicklung der Maschinellen Übersetzung als eigener Disziplin führten, ebenfalls wesentlich durch den Kalten Krieg motiviert.

In der Sowjetunion konnten sich in den 50er Jahren sowohl in Moskau wie in Leningrad Zentren der Sprachstatistik etablieren, wobei hier besonders die Leningrader Gruppe „Statistika reči“ zu nennen ist, deren Tätigkeit maßgeblich von R. G. Piotrovskij und N. D. Andreev geformt wurde. Alle organisatorischen und inhaltlichen Neuorientierungen dieser Zeit kulminierten in den 60er Jahren in einer Flut von Publikationen zur quantitativen Linguistik, deren Vorläufer bereits Ende der 50er Jahre erschienen, und die andererseits bis heute nachwirken: Die 60er Jahre waren ohne Zweifel die Blütezeit der sowjetischen Sprachstatistik; auf den methodischen und empirischen Ergebnissen, die damals erarbeitet wurden, basierte die Forschung und die publikatorische Tätigkeit der nächsten Jahrzehnte, die z.B. die ersten russischen Häufigkeitswörterbücher hervorbrachten. Als Publikationsorgan sei auch die 1952 gegründete Zeitschrift „Voprosy jazykoznanija“ erwähnt, die der Sprachstatistik eine ganze Reihe von Jahren mit der Rubrik „Prikladnoe i matematičeskoe jazykoznanie“ ein Forum bot, bevor sie wieder auf einen Strukturalismus-feindlichen und anti-formalen Kurs einschwenkte.

Einige programmatische und/oder klassische sprachstatistische Arbeiten aus den 60er Jahren sind: JAGLOM et al. 1960, FRUMKINA (1960b, 1964), ACHMANOVA et al. 1961, PIOTROVSKAJA et al. 1962, PIOTROVSKIJ (1966, 1969), ALEKSEEV et al. 1968. Zur Wissenschaftsgeschichte vgl. auch PAPP (1964, 1966).

3. Graphem- und Phonemebene

Da Buchstaben bzw. Phoneme klar abgegrenzte und wegen des geringen Inventarumfanges gut zählbare Einheiten sind, gehört der Bereich der Graphem- bzw. Phonemstatistik seit Anfang an zu den quantitativ gut bearbeiteten Aspekten des Russischen. Die Häufigkeit dieser Segmente, eine ihrer externen Eigenschaften, kann an Texten oder im Lexikon gezählt werden,

und zwar entweder das Vorkommen in bestimmten Positionen (initial, medial, final) oder ohne Berücksichtigung dieses Aspektes. Gezählt wird ferner die Häufigkeit der einzelnen Segmente für sich oder die Frequenz zwei-, drei- oder viergliedriger Kombinationen. Als Rahmeneinheit, innerhalb deren die Kombinatorik der Phoneme untersucht wird, dient die Silbe, das Morphem oder das Wort. Nur die erste Einheit ist dabei ebenfalls ausschließlich phonologisch definiert. Diese Häufigkeitszählungen basieren notwendigerweise immer auf einer begrenzten Stichprobe aus der Grundgesamtheit. Als Ergebnis erhält man Informationen über die Phonemstruktur von Silben, Morphemen und Wortformen, wobei man z.B. auch Klassen bilden kann (Vokale, Konsonanten oder dergl.), es lassen sich die für das Russische typischen Phonemfolgen ermitteln usw.

Ältere Schrift- und Lautstatistiken finden sich bei PEŠKOVSKIJ 1925 bzw. PROSKURNIN 1933. Eine ausführliche neuere Graphemstatistik stammt von BELONOGOV/FROLOV 1963, die die Frequenzen ein- und zweigliedriger Verbindungen mit und ohne Berücksichtigung der Position im Wort an Texten ermittelt haben. Eine systemische Stichprobe läßt sich dem OBRATNYJ SLOVAR' 1974 entnehmen. Nach Rängen geordnet, erhält man die Ergebnisse, die in Tabelle 1 zusammengestellt sind.

Am Wortanfang fehlen natürlich aus systematischen Gründen *й-, б-, в-* und *г-*, letzteres seit der Orthographiereform von 1917 ja auch am Wortende. Zwischen den verschiedenen Zählungen gibt es einige bedeutende Differenzen, die sich jedoch oftmals ohne weiteres erklären lassen. Allgemein ist z.B. *е* einer der häufigsten Buchstaben, jedoch nicht am Wortanfang, was sprachgeschichtliche Gründe hat: In dieser Position ist es fast immer zu *о* übergegangen (*единъ > один*). Umgekehrt ist *-и* am Wortende im Lexikon extrem selten (es gibt im Grunde nur einige wenige indeklinable Fremdwörter, die so auslauten), in Texten jedoch wegen der vielen Flexionsendungen äußerst häufig. Systemisch sind hingegen *-б, -й, -я* und *-а* besonders häufig: 70% aller russischen Lexeme enden auf einen dieser vier Auslaute.

Unter den Zweierkombinationen dominieren am Wortende bei systemischen Zählungen (s. OBRATNYJ SLOVAR' 1974) *-ть, -ий* und *-ся*, die 50% aller Lexeme charakterisieren; sie sind zugleich natürlich markante Allomorphe derjenigen Wortarten, die in der Nennform nicht ein Nullallomorph aufweisen (also Verben, Adjektive). LEBEDEV/GARMAŠ 1959 haben Dreierkombinationen anhand einer pragmatischen Stichprobe ausgewertet. Da das Leerzeichen auch als Zeichen gezählt wurde, ergaben sich als häufigste Ketten *_и_, _ть_, _ла_, _он_, _в_, _не_, _что_, _его_, _ся_*. In ihnen erkennt man die häufigsten Präpositionen, Konjunktionen, Pronomen und Suffixe des Russischen. Die Häufigkeit der zugrundeliegenden Inhalte bestimmt also hier bereits die Frequenzen der Buchstabenkombinationen.

<i>Graphem</i>	<i>allg., pragm. (Peškov.)</i>	<i>allgem., pragmat. (Bel./Frol.)</i>	<i>initial, pragmat. (Bel./Frol.)</i>	<i>medial, pragmat. (Bel./Frol.)</i>	<i>final, pragmat. (Bel./Frol.)</i>	<i>final, system. (Obr. Slov.)</i>
О	1	1	2	1	6	7
Е	3	2	18	2	2	5
А	2	3	15	3	4	4
Н	5	4	6	4	16	10
И	6	5	5	7	1	11
Т	4	6	12	6	7	8
Р	10	7	8	5	20	9
В	8,5	8	3	8	12	18
С	7	9	4	9	22	15
Д	11	10	7	10	17	16
П	15	11	1	16	23	21
М	12	12	13	14	5	12
К	13	13	9	12	14	6
Я	16,5	14	26	21	3	3
Л	8,5	15	21	11	19	17
У	14	16	11	18	13	22
Ч	18	17	17	13	25	28
Ы	20	18	–	17	10	14
З	21,5	19	14	15	18	20
Б	19	20	10	20	26	26
Ь	16,5	21	–	19	11	1
Й	23	22	–	24	8	2
Х	24,5	23	22	26	9	23
Г	21,5	24	16	23	21	19
Ж	27	25	25	22	24	24
Ю	26	26	27	27	15	29
Ш	24,5	27	24	25	28	27
Щ	29,5	28	28	28	29	30
Ц	29,5	29	23	29	27	13
Ф	31	30,5	20	32	31	25
Э	28	30,5	19	30,5	31	31
Ъ	32	32	–	30,5	31	–

Tab. 1: Rangverteilung der russischen Grapheme nach ihrer Häufigkeit

Die *Informationstheorie*, deren Vorgehen und Fragestellungen sich mit denjenigen der quantitativen Linguistik überschneidet, untersucht u.a. die Frage, ob und wie gut sich vorhersagen läßt, welches Graphem als nächstes folgen wird, wenn bereits eine Kette von n Graphemen bekannt ist. So sind z.B. die möglichen Fortsetzungen der Kette *um-* relativ begrenzt (*o, e, u*). Läßt sich das nächste Element erraten, so trägt es in diesem speziellen Verständnis des Begriffes keine Information; je unwahrscheinlicher es ist, desto mehr Information würde sein Auftreten haben. Solche informationstheoretischen Untersuchungen (vgl. JAGLOM et al. 1960) können anschaulich zei-

gen, daß allein die Berücksichtigung der Frequenzen immer längerer Graphemketten selbst bei rein zufälligem Kombinieren allmählich zu „echten“ Wörtern führt:

- F₀: оухеррохьдбщ яыхвшихжтифвнарфенвштфрпхгпчькизряс
 F₁: т цыяь серв однг збя енвтша ъуемлолйк
 F₂: кая всванный рося ных ковкров
 F₃: покак поститвленный пот дурноскака наконецпо зно стволовил
 F₄: весел вратсья не сухом и непо и добре

Neben diesen externen Eigenschaften sind die internen Merkmale von Phonemen bzw. Graphemen seltener Gegenstand quantitativer Untersuchungen gewesen. TOLSTAJA 1968 etwa berechnet die Ähnlichkeit der russischen Konsonanten hinsichtlich ihrer Spezifikationen nach distinktiven Merkmalen; KEMPGEN 1993 quantifiziert konstruktive Elemente von Graphemen und Graphemsystemen. Zur Phonemebene gehört als Einheit auch die Silbe, zu deren Struktur (Länge, Struktur als Abfolge von V und K) es mehrere Untersuchungen gibt, vgl. etwa ELKINA/JUDINA 1964 oder MOISEEV 1975. Zahlreiche Aspekte der Phonemebene untersuchen auch KUČERA/MONROE in ihrer Monographie von 1968.

4. Morphemebene

4.1. Flexion: Ausdrucksseite

Auf der Ausdrucksebene betrachtet, läßt sich an Morphemen vor allem ihre Frequenz untersuchen, sodann die Frage, wie sie sich aus Einheiten der nächstniedrigeren Ebene zusammensetzen (Morphemlänge in Buchstaben/Phonemen/Silben). Auch Häufigkeitsuntersuchungen zu grammatischen Phänomenen, denen kein inhaltlicher Unterschied entspricht (Deklinations- und Konjugationsklassen), gehören hierher.

Als wichtigste einschlägige Untersuchung sei hier OLIVERIUS 1976 genannt. Dieser Autor hat das in ŠTEJNFEL'DT 1963 enthaltene Material benutzt, um daraus eine Morphemstatistik zu gewinnen. Eine zusammenfassende Übersicht über Anzahl und Häufigkeit verschiedener grammatisch definierter Morphemtypen (Flexionsendungen sind ausgeschlossen) enthält Tabelle 2. Unter den isolierten Wurzeln sind hier Präfixe, Präpositionen, Konjunktionen, Interjektionen und Partikel zu verstehen. Das Type-Token-Verhältnis erlaubt Aufschlüsse über die Ausnutzung der betreffenden Wortstämme in der Rede: Präpositionen und Konjunktionen beispielsweise machen nur eine kleine Klasse aus, werden aber häufig benutzt.

<i>Morphemtyp</i>	<i>Types</i>	<i>Tokens</i>
Isolierte Wurzeln	81	151371
Verbale Wurzeln	273	82145
Substant. Wurzeln	425	51341
Nominale Wurzeln	142	26782
Pronom. Wurzeln	28	72342
Relationale Verbalsuffixe	6	44380
Relationale Subst.suffixe	40	40439
Relationale Nominalsuffixe	26	91218
Relationale Pronom.suffixe	13	55379
Gemischte Substant.suffixe	7	3884
Gemischte Nominalsuffixe	3	181
Modifizierende subst. Suffixe	6	3762
Modifiz. nominale Suffixe	4	3517
<i>Summe</i>	1054	626831

Tab. 2: *Frequenzen verschiedener Morphemtypen*

Von den beiden Konjugationsklassen dominiert nach KEIL (1965/66) mit einem Anteil von 61% die e-Konjugation, während auf die i-Konjugation 36% entfallen; den Rest machen die Sonderfälle aus (athematische Verben etc.). Angaben zur Häufigkeit der verschiedenen Varianten der Imperativbildung finden sich bei BERGER 1981.

Praktisch alle Aspekte der Flexionsmorphologie, die in A. A. Zaliznjaks ‚Grammatischem Wörterbuch‘ markiert sind, wurden von ILOLA/ MUSTAJOKI 1989 mit Methoden der deskriptiven Statistik ausgewertet. Das Buch enthält ausschließlich Tabellen mit Frequenzangaben und sprachlichen Beispielen. Statistische Tests zur Signifikanz und linguistische Auswertungen des umfangreichen Materials wären jedoch erst noch zu vorzunehmen. Zu allen Phänomenen der grammatischen Variation siehe auch das Wörterbuch von GRAUDINA et al. 1976, zur Verteilung der Akzentschemata in den Flexionsklassen des russischen Grundwortschatzes TORNOW 1984.

Experimente (PIOTROVSKIJ 1965, PETROVA/PIOTROVSKIJ 1966) haben gezeigt, daß die Information in Morphemen (ebenso in Wortformen und Sätzen) ungleichmäßig verteilt ist: Am Anfang ist die Unsicherheit über die Fortsetzung relativ hoch, um danach abzusinken, beispielsweise zum Stammende hin, bevor sie dann am Anfang der Flexionsendung wieder ansteigt. Als Kurve dargestellt, ergibt sich ein wellenförmiges Auf und Ab, wobei das Ansteigen plötzlich und steil erfolgt, der Abfall hingegen mehr oder weniger allmählich. Solche Untersuchungen liefern eine Begründung

für die bekannte Beobachtung, daß der Wortauslaut gewöhnlich stärkeren Abschleifungen, Verkürzungen etc. unterworfen ist als der Anlaut: er trägt weniger Information.

4.2. Flexion: Inhaltsseite

Als Untersuchung der Inhaltsseite können auf morphologischer Ebene am ehesten Häufigkeitszählungen grammatischer Phänomene verstanden werden, also etwa die Häufigkeit der beiden Aspekte, der Tempora, der Personen, Numeri, Kasus usw. Praktisch alle grammatischen Kategorien des Russischen sind bereits auf diese Weise untersucht worden.

Angaben zur Frequenz der russischen Kasus lassen sich mehreren Quellen entnehmen; erwähnt sei hier NIKOLAEV 1960, der für fremdsprachendidaktische Zwecke Zeitungstexte ausgewertet hat. Am häufigsten ist danach der Nominativ (33%), gefolgt von Genitiv (25%) und Akkusativ (20%); danach folgen die eher peripheren Kasus Präpositiv, Instrumentalis und Dativ, die nur einen Anteil von je 5–10% haben. GREENBERG (1974) hat den Kasusgebrauch weiter nach semantischen Gruppen von Substantiven aufgeschlüsselt: bei Körperteilen beispielsweise ist der Instrumentalis überproportional häufig, bei Ortsangaben nicht unerwartet der Präpositiv.

Die relativen Häufigkeiten unter den wichtigsten grammatischen Kategorien des Verbs zeigt Tabelle 3, und zwar in der Umgangssprache (nach KOŠAČEVA 1976) und in schriftlichen Texten (nach ŠTEJNFEL'DT 1963).

Beim Tempus zeigen sich deutliche Unterschiede zwischen beiden Funktionalstilen, ebenso wohl auch beim Aspektgebrauch (hierzu vgl. insbesondere KOUBOURLIS 1967). Reflexive Verbformen weisen im Durchschnitt einen Anteil von 20% auf (BELOUSOVA 1969), unter den Verben der Fortbewegung dominieren die determinierten mit 3/4 aller Fälle (BOBERTZ 1966).

<i>Kategorie</i>	<i>Bedeutung</i>	<i>Mündlich</i>	<i>Schriftlich</i>
Modus	Indikativ	0,81	0,79
	Imperativ	0,05	0,04
	Konjunktiv	0,01	0,02
	Infinitiv	0,13	0,15
Tempus	Präsens	0,49	0,32
	Präteritum	0,41	0,57
	Futur	0,10	0,11
Aspekt	unvollendet	0,43	0,53
	vollendet	0,57	0,47

Tab. 3: *Relative Häufigkeit verbaler Kategorien*

4.3. Derivation

Phänomene der Wortbildung sind bislang sehr viel seltener mit quantitativen Mitteln untersucht worden. Einer systematischen Untersuchung des gesamten Bereiches (vgl. zum Thema auch TICHONOV 1983) steht sicher auch entgegen, daß die zu untersuchenden Aspekte zu vielschichtig und vielgestaltig sind, als daß sie alle auf einmal erfaßt werden könnten. So gibt es bisher hauptsächlich Arbeiten zu einzelnen konkreten Wortbildungsverfahren, Affixen etc.

Beispielsweise untersucht RAKUŠAN 1981 Affixe mit expressiver Funktion, indem er dem großen Akademiewörterbuch eine umfangreiche Stichprobe von Basiswörtern und Ableitungen entnimmt und nach den verwendeten Suffixen – insgesamt 66 – aufschlüsselt. Zu jedem Suffix findet sich eine Auswertung, ob seine Funktion vorwiegend „neutral“, „meliorativ“, „pejorativ“ oder „meliorativ wie pejorativ“ ist. Das Suffix *-ушк-* z.B. ist primär pejorativ, während als Normalfall aller Suffixe eine meliorative Funktion festzustellen ist. Die Derivationsmuster russischer Neologismen der Gegenwart untersucht KEMPGEN 1988 an empirischem Material und schlüsselt die wichtigsten Verfahren und Lehnbeziehungen getrennt nach Wortarten auf. Hierbei zeigt sich, daß eine Derivation innerhalb der gleichen Wortart dominiert, während der Wortartwechsel in den Hintergrund tritt. Eher peripher – aber mit einigen typischen Konstanten versehen – sind in der Sowjetzeit auch die Entlehnungen, was ihre Anzahl und die bevorzugten Quellsprachen betrifft.

5. Wortebene

Beginnen wir mit der internen Struktur von Wortformen, so läßt sich hier in erster Linie die Wortlänge in Einheiten verschiedener Ebenen untersuchen. Für die *Wortlänge in Buchstaben* gibt KURBAKOV 1965 eine aus Texten gewonnene Verteilung an, aus der u.a. hervorgeht, daß 73% aller *types* zwischen 5 und 11 Buchstaben lang sind. Die durchschnittliche *Wortlänge in Phonemen* beträgt nach NIKONOV 1978 in Texten 5,8. Lexeme können nach den Daten des OBRATNYJ SLOVAR' 1974 bis zu 12 Silben umfassen, einen hohen Anteil am Lexikon haben jedoch nur die Drei-, Vier- und Fünfsilbler.

Den Anteil der üblichen *Wortarten* des Russischen in einer pragmatischen Stichprobe (wobei MARKOV 1960 Autorenrede und Personenrede getrennt ausgewertet hat) und in einer systemischen Stichprobe (nach dem OBRATNYJ SLOVAR' 1974) zeigt Tabelle 4. Daß die Wortarten am Lexikon und in Texten unterschiedliche Anteile haben, ist offensichtlich und leicht einsichtig: ein Wörterbuch verzeichnet ja *types*, im Text finden wir *tokens*. Alle geschlossenen Klassen – Pronomina, Präpositionen und Konjunktionen – umfassen nur wenige Dutzend Elemente, die aber häufig gebraucht werden: in jedem Häufigkeitswörterbuch sind sie in der Rang-Frequenzliste ganz vorne zu finden. Aber auch der mündliche Sprachgebrauch (hier durch die Personenrede vertreten) unterscheidet sich vom schriftlichen: Mündlich dominieren die Pronomina noch stärker, Partizipien fehlen dagegen fast völlig, Partikel haben hier ihre Domäne.

<i>Wortart</i>	<i>Autorenrede (pragm.)</i>	<i>Personenrede (pragm.)</i>	<i>Lexikon (system.)</i>
Substantiv	28,0	19,4	46,35
Adjektiv	9,6	3,9	20,40
Numerale	1,4	0,8	0,10
Pronomen	10,6	20,2	0,08
Verb	15,8	20,0	30,71
Partizip	2,5	0,3	–,–
Adverbialpart.	1,2	0,0	–,–
Adverb	7,6	10,5	1,58
Präposition	12,4	7,5	0,12
Konjunktion	7,8	6,3	0,09
Partikel	3,1	10,1	0,12
Interjektion	0,0	1,0	0,28

Tab. 4: Prozentualer Anteil der Wortarten des Russischen

Frequenzwörterbücher sind das vielleicht bekannteste Ergebnis der Anwendung sprachstatistischer Methoden, die freilich selbst erst erarbeitet werden mußten. Inzwischen gibt es bereits etliche Häufigkeitswörterbücher zum Russischen, eine Ironie der Geschichte – vor dem Hintergrund der wissenschaftsgeschichtlichen Anmerkungen (s.o.) jedoch verständlich – ist es jedoch, daß das erste große Frequenzwörterbuch zum Russischen Anfang der 50er Jahre in den USA erschien, das zweite, didaktisch orientiert, ein Jahrzehnt später im Baltikum erarbeitet wurde, und erst Ende der 70er Jahre das erste „russische“ Häufigkeitswörterbuch erschien. Seitdem sind außerhalb Rußlands weitere Frequenzlisten publiziert worden.

Das erste große Häufigkeitswörterbuch des Russischen, JOSSELSON 1953, wurde – erstmals überhaupt – mit Computerhilfe erstellt und basiert auf Texten von insgesamt mehr als 1 Mio. laufenden Wörtern aus dem 19. und dem 20. Jh. (Publizistik, Belletristik, Wissenschaft, Sachtexte; die klassische russische Literatur vor dem 19. Jh. kam bei diesen Einschränkungen etwas zu kurz). Von den mehr als 40.000 verschiedenen Lexemen, die in den Stichproben vorkamen, fanden die 5320 häufigsten in die alphabetisch geordneten Listen dieses Wörterbuches, dem merkwürdigerweise eine einfache Rang-Frequenzliste fehlt, Aufnahme. Das Wörterbuch enthält darüber hinaus Informationen über die Häufigkeit einzelner grammatischer Phänomene.

Immer noch aktuell wegen seiner didaktischen Orientierung, die auch dazu geführt hat, daß das Wörterbuch in mehrere westliche Sprachen übersetzt wurde, ist ŠTEJNFEL'DT 1963. Es sollte insbesondere der Ausarbeitung eines Grundwortschatzes für den fremdsprachigen Russischunterricht dienen. Zusammengestellt wurden hier die 2.500 häufigsten Wörter, also ein erweiterter Grundwortschatz. Von besonderem Nutzen sind die zusätzlichen Auswertungen, etwa die nach Wortarten getrennten Listen, die auch Angaben zur Frequenz der jeweiligen Teilparadigmen bzw. einzelnen Wortformen enthalten, die Liste zur Rektion der Verben usw. Das in diesem gar nicht einmal sehr umfangreichen Büchlein enthaltene Material hat vielen sprachstatistischen Untersuchungen als Grundlage gedient. Einschränkend gilt es anzumerken, daß – der Zielstellung entsprechend – die Auswahl der Texte nicht ganz repräsentativ für das Russische insgesamt ist.

Das russische Standardwerk liegt mit ZASORINA 1977 vor. Es bietet knapp 40.000 Wörter in einer alphabetischen Liste und die 9.000 häufigsten von ihnen in der Rang-Frequenz-Liste. Hinweise zur Gebräuchlichkeit allgemein (verstanden als die Zahl der Texte, in denen das Lexem angetroffen wurde) und die Frequenz in den vier unterschiedenen Funktionalstilen runden die Angaben ab. Was als Desiderat selbst nach diesem umfangreichen Werk bleibt, ist ein Häufigkeitswörterbuch, das genaue

Angaben zur Frequenz jeder einzelnen grammatischen Form jedes Lexems enthält – zusammenfassen kann man solche Einzeldaten immer noch, wenn man nur an globaleren Werten interessiert ist, aber nicht umgekehrt.

Als neuestes Häufigkeitswörterbuch ist vor einigen Jahren LÖNNGREN 1993 erschienen, das aber die älteren Arbeiten ergänzt, nicht ersetzt. Als Besonderheit sei hier erwähnt, daß das zugrundeliegende Textkorpus der Forschung in computerlesbarer Form für weitere Auswertungen zur Verfügung steht.

Neben diesen großen, allgemeinen Häufigkeitswörterbüchern gibt es zum Russischen auch etliche Frequenzlisten zu speziellen Funktionalstilen oder sogar einzelnen Autoren bzw. Werken: Kindersprache (CHARAKOZ 1971), Umgangssprache (MARKOV/VIŠNJAKOVA 1968, TURKO 1968), Elektronik (KALININA 1968), Zeitung (POLJAKOVA/SOLGANIK 1971), Wissenschaft/Technik (DENISOV et al. 1978), Militär (KOLGUŠKIN 1970), verschiedene Fachwortschatzbereiche (vgl. als Beispiel HOFFMANN 1973) u.a.m. Eine umfangreiche Sekundärliteratur hat sich mit einem Vergleich der Ergebnisse, den Konsequenzen für die Erstellung von Lehrbüchern, Grundwortschätzen usw. beschäftigt.

Von den spezielleren Frequenzwörterbüchern sei an dieser Stelle stellvertretend dasjenige zu Puškin genannt, dessen Werk (mehr als eine halbe Million laufende Wörter) vollständig ausgewertet wurde (FRUMKINA 1960a, 1963). Puškin hat danach 21.197 verschiedene Lexeme benutzt, was einen sehr hohen aktiven Wortschatz darstellt. Von ihnen hat er 6.388 Wörter nur ein einziges Mal verwendet, aber genau diese Wörter sind es, die den Reichtum seines Wortschatzes ausmachen. Von Puškins 500 häufigsten Wörtern – mit denen man im Durchschnitt bereits knapp 60% eines Textes erfaßt – brauchen heute nur 10% als veraltet zu gelten; deshalb gehören seine Texte durchaus noch zum modernen Russisch.

Tabelle 5 zeigt die 25 häufigsten Wörter des Russischen nach einigen dieser Frequenzlisten: je spezieller die Textsorte, desto eher sind typische Autosemantika bereits unter diesen allerfrequentesten Einheiten zu finden, vgl. etwa *напряжение, электронов, поля* im Bereich der Elektronik oder *я, ты, идти, дать, играть* in der Kindersprache.

<i>Rang</i>	<i>Zasor.</i> <i>Allg.</i>	<i>Štejn.</i> <i>Allg.</i>	<i>Lönn.</i> <i>Allg.</i>	<i>Turko</i> <i>Uggs.</i>	<i>Pl./Sg.</i> <i>Zeitung</i>	<i>Denis.</i> <i>Wiss.</i>	<i>Kalin.</i> <i>Elektr.</i>	<i>Xarak.</i> <i>Kinder.</i>	<i>Puškin</i> <i>Belletr.</i>
1	в	и	и ¹	не	в	в	в	я	и
2	и	в	в ¹	я	и	и	и	ты	в
3	не	на	на ¹	а ¹	на	на	при	а	я
4	на	не	не	в	он	с	на	не	он
5	я	он	с	у ²	с	при	для	он	не
6	быть	я	этот	и ¹	не	этот	с	в	быть
7	что	что	быть	ну	что	для	от	и	на
8	он	с	он	так	этот	по	по	что	с
9	с	этот	а ¹	нет	быть	тот	что	у	что
10	а	быть	весь ¹	он	по	из	рис.	быть	ты
11	как	а	что ²	что ²	весь	к-рый	из	мы	мой
12	это	весь	тот	да ¹	к	что	как	вот	она
13	вы	они	как	на ¹	о	к	не	на	свой
14	ты	она	к	ты	год	быть	к	пойти	но
15	к	как	по	вы	за	не	а	ну	к
16	мы	мы	но ¹	как	к-рый	или	тока	нет	весь
17	этот	к	они	это	тот	от	можно	идти	они
18	она	у	и ²	она	а	как	ток	дать	как
19	они	ты	что ¹	с	свой	а	напряж.	с	тот
20	но	за	она	что <i>и</i>	из	весь	так	как	мы
21	по	тот	из	вот	но	иметь	э.трон.	этот	о
22	весь	но	к-рый	мне <i>д</i>	для	мочь	случае	сейча с	вы
23	за	вы	от	там	как	так	до	да	а
24	то	по	за	меня <i>р</i>	это	такой	или	весь	за
25	все	из	я	очень	мы	число	поля	играть	это

Tab. 5: Die 25 häufigsten Wörter des Russischen im Vergleich

5.3. Merkmalszusammenhänge

Auch der wechselseitige Zusammenhang zweier Eigenschaften läßt sich sinnvoll untersuchen. So beschäftigt sich MOISEEV 1976 z.B. mit der Position der Betonung in Abhängigkeit von der Wortlänge (in Silben); es ergibt sich, daß die Betonungsstelle zur zweiten Worthälfte hin tendiert. Zum gleichen Thema vgl. zuerst NIKONOV 1963. Auch die Zuordnung der russischen Lexeme zum festen bzw. zum beweglichen Akzent ist eine Tendenz: bewegliche Betonung ist praktisch nur unter Wörtern mit bis zu drei Silben relevant. MUSTAJOKI 1981 untersucht die Zugehörigkeit zu diesen beiden Typen in Abhängigkeit von der Wortfrequenz: bewegliche Betonung ist unter häufigen Wörtern üblicher. HENTSCHEL 1990 kommt zu dem Schluß, daß hohe Frequenz historisch auch den Übergang zur beweglichen Betonung begünstigt habe. Daten zu den Wortarten in Abhängigkeit vom Wortauslaut liefert das OBRATNYJ SLOVAR' 1974: natürlich gibt es hier signifikante und offensichtliche Abhängigkeiten.

6. Satzebene

Für die Satzebene gilt natürlich per definitionem, daß alle Untersuchungen auf einer pragmatischen Stichprobe beruhen müssen, da sich Sätze bzw. Äußerungen nicht inventarisieren lassen, höchstens Satztypen, Strukturschemata usw., also abstrakte und dabei in hohem Maße beschreibungsabhängige Größen.

Die klassische quantitative Untersuchung auf Satzebene betrifft die *Satzlänge*, genauer die Frage, wie häufig ein Satz gegebener Länge auftritt. Für das Russische ist die Bearbeitung dieser Frage vor allem LESSKIS (u.a. 1962, 1964) zu verdanken, der im Laufe der Jahre ein außerordentlich großes Textkorpus untersucht hat. Seine Daten sind jedoch nicht für das Russische insgesamt repräsentativ, sondern betreffen einzelne Funktionalstile, Genres oder Zeiträume, so z.B. den psychologischen Roman der 2. Hälfte des 19. Jahrhunderts (Dostoevskij, Leskov, Tolstoj, Turgenjev) oder die wissenschaftliche Prosa der gleichen Zeit. Seine Daten, die gewöhnlich auch graphisch dargestellt werden, zeigen u.a., daß die wissenschaftliche Prosa zu sehr viel längeren Sätzen tendiert (Gipfel der Verteilung bei 20 Wörtern) als die belletristische Autorenrede (Gipfel bei Satzlänge 7), außerdem durchweg komplexe Sätze bevorzugt.

Weitere Untersuchungen betreffen etwa den Informationsverlauf in Sätzen, den Anteil von Objektsätzen an den komplexen Sätzen als Stilmerkmal, den Anteil der subordinativen Sätze in Texten verschiedener Funktionalstile, das Verhältnis der Satztypen Deklarativsatz, Interrogativsatz und Ausruf zueinander, das Auftreten einfacher und komplexer Sätze in der Zei-

tungssprache (Überschrift, Leitartikel, Kommentar), die Wortstellung am Satzanfang bzw. die Stellung bestimmter Wortarten zueinander (BUTTKE 1965) u.a.m. Die Zahl derartiger Fragestellungen läßt sich leicht vervielfachen.

Einige wenige Arbeiten widmen sich auch syntaktischen Relationen innerhalb von Sätzen, etwa speziellen Syntagmen, Rektionsmustern oder Kongruenzregeln. Den Versuch, einen morphosyntaktischen Begriff quantitativ zu fassen, finden wir bei APRESJAN 1964, der auf diese Weise zwischen starker und schwacher Rektion unterscheiden will.

7. Textebene

An Texten ist bisher vor allem das sog. *Type-Token-Verhältnis* untersucht worden, d.h. das Verhältnis der verschiedenen Wörter zur Zahl der laufenden Wörter (genauer: Wortformen). Damit ist impliziert, daß sich Wörter im Text wiederholen, was zur Definition der sog. *Wiederholungsrate* führt. Am Anfang eines Textes ist praktisch jedes Wort neu; je weiter ein Text fortschreitet, desto häufiger wiederholen sich die Wörter jedoch und ein Textthema etabliert sich. PAPP 1961 unterteilt für seine Untersuchung Texte in Stücke von je 100 Wörtern und zählt die jeweils neuen Lexeme aus. Deren Anteil liegt zunächst bei 67–77%, sinkt bis zur vierten Gruppe stark ab, um sich dann bei etwa 30% bei einer Textlänge von 2000 Wörtern einzupendeln. Das heißt: in längeren Texten ist regelmäßig immerhin noch ein Drittel der Wörter neu: neue Personen werden eingeführt und charakterisiert, neue Handlungsstränge beginnen, neue Beschreibungen setzen ein usw.

Unter der *Textdeckung* versteht man die Frage, wieviel Prozent eines gegebenen Textes durch die x häufigsten Wörter einer Sprache abgedeckt werden. Solche Angaben lassen sich aus den Frequenzlisten von Häufigkeitswörterbüchern gewinnen bzw. entnehmen. ZASORINA 1977 etwa läßt erkennen, daß mit den 10 häufigsten Wörtern des Russischen durchschnittlich bereits 18% eines beliebigen Textes erfaßt sind (was aber nicht viel besagt, da die 10 häufigsten Wörter nichtssagende Funktionswörter sind), mit den 230 häufigsten etwa 50% eines Textes. Mit 2.000 Wörtern, dem normalen Umfang eines Grundwortschatzes, erfaßt man im Durchschnitt 76%. Trägt man solche Wertepaare in ein Koordinatensystem ein, so ergibt sich eine charakteristische Kurve, die zunächst steil ansteigt, dann aber in einem weiten Bogen nach rechts geht, ohne die Marke der hundertprozentigen Deckung jemals zu erreichen. Anschaulich sieht man hieran die bekannte Tatsache, daß die Kosten-Nutzen-Relation beim Wortschatzlernen leider desto schlechter wird, je mehr man schon be-

herrscht. Bei ŠTEJNFEL'DT 1966 finden sich Beispielstexte, an denen die Textdeckung anschaulich demonstriert wird. Auf diese Weise kann man sich leicht von der Effektivität von Frequenzlisten für die Sprachpraxis überzeugen.

8. Semantische Ebene

Die Inhaltsebene ist direkter Beobachtung naturgemäß nicht zugänglich, so daß es hier darauf ankommt, zunächst sinnvolle Vorgehensweisen zu formulieren, bevor irgend etwas gezählt werden kann.

Die Häufigkeit von *Semen* ist von KARAULOV 1980 anhand der bekannten Lexika von Ožegov und Ušakov untersucht worden, und zwar aufgrund folgender Überlegung: In einem einsprachigen Wörterbuch wird die Bedeutung eines gegebenen Eintrages durch andere Lexeme paraphrasiert. Wie oft bestimmte Lexeme zur Bedeutungsbeschreibung herangezogen werden, läßt sich auszählen. In der genannten Arbeit werden mehr als 5.500 Wortstämme mit ihren Häufigkeiten in zwei verschiedenen Listen aufgeführt.

Eine andere Eigenschaft, die sich systemisch am Lexikon an der Zahl der Bedeutungsangaben je Eintrag untersuchen läßt, ist die *Polysemie*. KRYLOV 1982 hat diese Frage anhand des Wörterbuches von Ožegov untersucht. Danach sind dort mehr als 23.400 Lexeme eindeutig, mehr als 6.200 bzw. 1.600 zwei- bzw. dreideutig; vier bis sechs Bedeutungen weisen jeweils mehrere Hundert Lexeme auf, sieben bis zehn Bedeutungen jeweils mehrere Dutzend Wörter, und elf bis sechzehn Bedeutungen jeweils weniger als zehn Einträge. Die Zahl der Bedeutungen, über die ein Wort verfügt, korreliert dabei mit seiner Häufigkeit: je allgemeiner und vielfältiger die Bedeutung eines Wortes ist, desto häufiger kann es benutzt werden.

Für eine pragmatische Stichprobe läßt sich die Zahl der *Bedeutungen* folgendermaßen operationalisieren: Je allgemeiner die Bedeutung eines Wortes, desto häufiger kann es sich mit anderen Wörtern in einem Syntagma kombinieren. PLOTNIKOV 1979 ermittelt deshalb die Zahl der Kombinationen, in die ein gegebenes Wort eingeht, und findet in fachsprachlichen Texten auf 1.000 Kombinationen nur 200 bis 250 verschiedene Lexeme, in belletristischen Texten hingegen 500 bis 600. In ähnlicher Weise untersucht er in der Arbeit von 1986 den Grad an Synonymie innerhalb des kleinen Wortfeldes ‚Denken‘, indem er die Zahl der gleichen Kombinationspartner ermittelt. Die alte Behauptung, daß es vollständige Synonymie gar nicht gebe, findet hier ihre Bestätigung, wenn etwa *мысль* und *идея* einen Synonymiegrad von 0,93 aufweisen: beide Wörter können sich oft, aber eben nicht immer gegenseitig ersetzen.

Die Beliebtheit der russischen *Vornamen* unterscheidet sich deutlich nach städtischen und ländlichen Gebieten: *Елена, Марина, Ирина* sowie *Олег* und *Игорь* werden in Städten bevorzugt, *Галина, Татьяна* und *Иван* hingegen sind auf dem Lande beliebt (NIKONOV 1967). Daß nach der Revolution Vornamen wie *Владимир* und *Надежда* einen Aufschwung erlebten, wird kaum verwundern, vielleicht aber die Tatsache, daß bis dahin mit den zehn häufigsten Vornamen mehr als die Hälfte aller Kinder benannt wurde (BONDALETOV 1972).

Weitere Untersuchungen betreffen etwa Strukturtypen von Flurnamen (PLESKALOVÁ 1984), die stilistische Bewertung des russischen Wortschatzes im Lexikon (DENISOV/KOSTOMAROV 1969), die Struktur des Farbwortfeldes (MOSKOVIČ 1969), die Konkurrenz zwischen englischen Fremdwörtern und ihren russischen Äquivalenten im Bereich des Sportes (REINTON 1978) u.a.m.

9. Soziolinguistik, Sprache und Norm

Aus sowjetischer Zeit existiert eine (einzige) groß angelegte empirische soziolinguistische Untersuchung, deren Ergebnisse vor allem in den beiden Bänden KRYSIN 1974 sowie KRYSIN/ŠMELEV 1976 niedergelegt sind, ferner in einigen ergänzenden Aufsätzen. Die empirischen Daten wurden Anfang der 60er Jahre per Fragebogen erhoben. Die 3.000 bis 4.000 Probanden waren Städter mit höherer Schulbildung, d.h. (vermutliche) Träger der Literatursprache. Abgeprüft wurden praktisch sämtliche wesentlichen Variationen des Russischen auf phonologischer, morphologischer und morphosyntaktischer Ebene, die assimilatorische Erweichung in Konsonantengruppen und die Vokalreduktion ebenso wie der Gebrauch des zweiten Genitivs bzw. Lokativs, der Nom.Pl. auf *-a* und die Flexion von Ortsnamen ebenso wie die Konjugation bestimmter Verbklassen, konkurrierende Derivationsmuster und Kongruenzphänomene, speziell in Verbindung mit Bezeichnungen für Frauen. Die Auswertung der Ergebnisse folgt fast immer einem festen Schema: nach historischen Erläuterungen zur betreffenden Variationsreihe werden die relativen Anteile der einzelnen ‚Konkurrenten‘ genannt und weiter nach den Merkmalen Alter, Beruf, Bildungsstand und Herkunft aufgeschlüsselt. Die Daten werden tabellarisch und in Form von Diagrammen dargestellt, aber nur selten auf ihre Signifikanz hin geprüft; interpretiert werden die augenfälligsten Besonderheiten. Dabei finden sich Erklärungsansätze wie die „Nähe zu den hochsprachlichen Zentren“, die Zugehörigkeit zu den dialektalen Großgruppen des Russischen (Norden, Mitte, Süden), die bewußtere Beherrschung der hochsprachlichen Normen bei „Philologen“ u.a. Auffällig ist, daß die Al-

tersstufe, die natürlich gerade kein soziales Merkmal ist, besonders häufig einen gewichtigen Einfluß zu spielen scheint.

Quantitative Untersuchungen zu Norm und Usus des Russischen unterscheiden sich von den soziolinguistischen vor allem dadurch, daß in ihnen versucht wird, den Gebrauch der unterschiedlich häufigen Varianten in erster Linie mit sprachlichen Randbedingungen zu korrelieren; empirische Erhebung und Datenauswertung an sich sind in beiden Bereichen sehr ähnlich. Die wichtigste Autorin im Problembereich sprachliche Norm und Statistik ist ohne Zweifel L. K. Graudina mit zahlreichen Arbeiten, wichtigstes „Produkt“ neben dem Grundsatzartikel von 1970 das Nachschlagewerk GRAUDINA et al. 1976, das eine Fülle von Material zu den Bereichen Flexion und Morphosyntax enthält.

Daß in beiden hier genannten Arbeitsgebieten Syntax und Semantik wesentlich schlechter bearbeitet sind als Phonologie und Morphologie, läßt sich leicht – neben dem Pilotcharakter der Studien – dadurch erklären, daß der Begriff der Variationsreihe auf den niedrigeren Ebenen der Sprache sehr viel besser zu handhaben ist. Auf lexikalischer Ebene könnte man sich aber eine Abrundung der vorliegenden Untersuchung durch weitere Studien gut vorstellen.

10. Psycholinguistik

Quantitative psycholinguistische Untersuchungen zum Russischen sind bislang eher selten; eine Ausnahme stellt das Wörterbuch von LEONT'EV 1977 zu den „assoziativen Normen“ dar, das die häufigsten Syntagmen und typische semantische Felder verdeutlichen soll, in die ein gegebenes Wort eingeht. Einträge des Wörterbuches sind 196 Stimuli, zu denen mehrere Hundert erwachsene Muttersprachler als Probanden den jeweils ersten Response beisteuerten. Zu jedem Stimulus wird sein Rang und seine Frequenz in den drei wichtigsten Häufigkeitswörterbüchern angegeben, dazu, ihrer Häufigkeit nach geordnet, die Responses. Die häufigsten Reaktionen etwa auf *русский* waren *язык* (40), *человек* (36), *народ* (12), *казах* (10), *немец* (8), *лес* (7). Das Wörterbuch zielt in erster Linie auf praktische Anwendungen (Erstellung von landeskundlichen Materialien für den Fremdsprachenunterricht), bietet selbst aber keine derartigen Auswertungen. Neben rein linguistischen Fragestellungen eröffnen sich hierfür auch interessante kulturwissenschaftliche Perspektiven.

Als psycholinguistisch kann man auch die Untersuchung von MORGAN/CORBETT 1989 zu den Farbwörtern des Russischen verstehen. Sie übertrugen ein bekanntes Experiment auf das Russische und ließen eine Gruppe von Moskauern alle ihnen einfallenden Farbwörter aufschreiben, wobei

nach jeder Minute eine Markierung in der Liste zu machen war. Bei der Auswertung der Ergebnisse kam es darauf an, wie oft ein konkretes Farbwort in jedem der fünf Zeitabschnitte genannt wurde. Erwartungsgemäß stehen die Grundfarben an der Spitze der Nennungen innerhalb der ersten Minute.

11. Dialektologie

Einer der Vorläufer der quantitativen Linguistik in Rußland hat bereits dialektologische Daten bearbeitet, nämlich DIKAREV 1891. Er untersuchte an den Dialekten von Priluck und Voronež die Häufigkeit von Einzellauten sowie zweigliedriger Lautverbindungen und verglich diese Daten mit denjenigen anderer slawischer Idiome. Seit dieser Zeit ist dialektologisch wenig mit quantitativen Methoden gearbeitet worden, obwohl alle modernen Sprachatlanten hierfür als Basis dienen könnten. Auch die in Westeuropa bereits in Bezug auf mehrere Sprachen angewendete „Dialektometrie“, das heißt die quantitative Bearbeitung, Auswertung und Darstellung von Dialektdaten, hat in die Russistik praktisch noch keinen Eingang gefunden. Lediglich verschiedene Arbeiten von PŠENIČNOVA (1973, 1977) vertreten diesen Ansatz und stellen zugleich eine frühe Übernahme der Prinzipien der sog. „Numerischen Taxonomie“, d.h. der algorithmischen Klassenbildung auf der Basis quantifizierter Merkmale, auf sprachwissenschaftliche Probleme dar. Z.B. werden von der genannten Autorin die Dialekte im Gebiet Ladoga – Novgorod – Toržok – Pskov in Form einer Merkmalsmatrix beschrieben, auf dieser Grundlage ihre gegenseitige Ähnlichkeit berechnet und schließlich eine Klassifikation aus der Ähnlichkeitsmatrix abgeleitet, die der traditionellen Dialekteinteilung für dieses Gebiet gegenübergestellt wird.

12. Poetische Sprache

Der Bereich „Statistik und Stil“ ist besonders von den Vertretern der Prager Schule theoretisch wie praktisch ausgearbeitet worden, vgl. etwa den Sammelband DOLEŽEL/BAILEY 1969, dem sich viele Informationen entnehmen lassen. Aus der russischen Sekundärliteratur sei KAUFMAN 1970 genannt. Zum Thema Satzlänge vgl. die schon genannten Arbeiten von LESSKIS, der dieses Merkmal ja u.a. an belletristischen Werken untersucht hat.

Ein Bereich des Russischen, der seit Anfang des 20. Jhs. ganz besondere Aufmerksamkeit gefunden hat, ist die quantitative Versstatistik, etwa in frühen Arbeiten Andrej Belyjs. Nach dem Standardwerk von TARANOVSKI

1953 erschienen Anfang der 60er Jahre in schneller Abfolge zahlreiche weitere Arbeiten (z.B. GASPAROV 1963, 1965, KOLMOGOROV 1963, KOLMOGOROV/KONDRATOV 1962, KOLMOGOROV/PROCHOROV 1963, KONDRATOV 1962, 1963), die sich u.a. intensiv mit Majakovskij beschäftigten. Die umfangreichste systematische Untersuchung stellt der Sammelband RUSSKOE STICHOSLOŽENIE 1979 dar, der als Beginn einer Serie geplant war.

Ein geradezu klassisches Problem ist der Versuch, mit quantitativen Methoden einen Autorschaftsnachweis zu führen. In der russischen Literatur ist es der *Tichij Don*, für den Michail Šolochov den Nobelpreis bekommen hat, dessen Autor aber in Wirklichkeit der Kosakenschriftsteller Fëdor Krjukov gewesen sein soll, wie seinerzeit vor allem im Westen behauptet wurde. KJETSAA nimmt sich in mehreren Arbeiten (1976, 1978) der Frage an, indem er unumstrittene Werke beider Autoren mit dem Text des *Tichij Don* vergleicht und zu beantworten versucht, wem von beiden der umstrittene Text ähnlicher ist. Konkret wird – jeweils an Stichproben – die Satzlänge, das Auftreten der Wortarten an Satzanfang und Satzende, die Wortlänge und der lexikalische Reichtum der Texte untersucht. Jeweils ergibt sich mit mehr oder minder großer Signifikanz, daß der *Stille Don* dem Stil Šolochovs näher steht als demjenigen Krjukovs. Das bedeutet aber nur folgendes: Die Wahrscheinlichkeit dafür, daß Krjukov den veröffentlichten Text wortwörtlich so geschrieben hat, ist relativ gering, wohingegen aber nach wie vor die Behauptung nicht widerlegt ist – und darum ging es ja hauptsächlich –, daß sich Šolochov von Krjukov gesammeltes Material, Skizzen und weit fortgeschrittene Entwürfe angeeignet und stilistisch zu dem publizierten Text überarbeitet hat. Anderen Fällen fraglicher oder unbekannter Autorschaft sind ERMOLENKO 1988 und MILOVA 1994 gewidmet.

13. Schluß

Wie diese kurze Übersicht angedeutet hat, sind bereits zahlreiche Bereiche des Russischen quantitativ bearbeitet worden; das Problem liegt jedoch oft darin, die einschlägigen Arbeiten zu ermitteln und die konkreten Ergebnisse herauszudestillieren. Während für den vorliegenden Beitrag keinerlei Vorkenntnisse vorausgesetzt wurden, gilt dies natürlich nicht für sämtliche hier zitierte Literatur; mindestens bei einem Teil von ihr muß man bereit sein, ein Mindestmaß an mathematischer Denkweise und Formeln nachzuvollziehen. Dies sollte jedoch niemanden davon abhalten, die Resultate, die von der quantitativen Linguistik bereits erzielt worden sind,

wenigstens zu rezipieren und einfache deskriptive Verfahren selbst anzuwenden.

Literatur

- Altmann, G. (1972), Status und Ziele der quantitativen Linguistik. In: S. Jäger (ed.), *Statistik und Linguistik*, Braunschweig, 1–9.
- Altmann, G. (1973), Mathematische Linguistik. In: W. A. Koch (ed.), *Perspektiven der Linguistik I*, Stuttgart, 208–232.
- Berger, T. (1981), The Formation of the Imperative in Modern Russian. *Russian Linguistics* 6, 65–80.
- Bobertz, J. (1966), Die Doppelzeitwörter der Fortbewegung. Untersuchungen zu Fragen des Terminus, des Bestandes, der Klassifizierung und der Häufigkeit. *Zeitschrift für Slawistik* 11, 535–549.
- Bondaletov, V. D. (1972), Der russische Vornamensschatz, seine Zusammensetzung, statistische Struktur und die Besonderheiten seiner Veränderung. *Zeitschrift für Slawistik* 17, 3–18.
- Buttke, K. (1965), Zur quantitativen Charakteristik der Wortfolge im Russischen. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig* 14, 53–68.
- Doležel, L./ Bailey, R. W. (eds.) (1969), *Statistics and Style* (Mathematical Linguistics and Automatic Language Processing, vol. 6). New York.
- Greenberg, J. (1974), The relation of frequency to semantic feature in a case language (Russian). *Working Papers on Language Universals* 16, 21–45.
- Hentschel, G. (1990), Natürlichkeit vs. Frequenz. Überlegungen zur betonungsabhängigen Allomorphie in der Deklination russischer Substantive. In: Dosuna, J. M./ Pensado, C. (eds.), *Naturalists at Krems*, Salamanca, 65–76.
- Hoffmann, L. (Hg.) (1973), Fachwortschatz Chemie. Häufigkeitswörterbuch Russisch – Englisch – Französisch. Leipzig.
- Hoffmann, L./ Piotrowski, R. G. (1979), *Beiträge zur Sprachstatistik*. Leipzig.
- Ilola, E./ Mustajoki, A. (1989), *Report on Russian Morphology as it Appears in Zaliznyak's Grammatical Dictionary* (Slavica Helsingiensia, 7). Helsinki.
- Josselson, L. G. (1953), *The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian*. Detroit.
- Keil, R.-D. (1965/66), Zur Häufigkeit russischer Verben und Verbformen. *Zeitschrift für den Russischunterricht* 1/2, 35–50.

- Kempgen, S. (1988), Zur lexikalischen Entwicklung der „Weltsprache Russisch“. In: J. Raecke (ed.), *Slavistische Linguistik 1987*, München, 184–208.
- Kempgen, S. (1993), Spezifika slawischer Schriften. In: ders. (ed.), *Slavistische Linguistik 1992*, München, 111–143.
- Kempgen, S. (1995), *Russische Sprachstatistik. Systematischer Überblick und Bibliographie* (Vorträge und Abhandlungen zur Slavistik, Bd. 26). München.
- Kjetsaa, G. (1976), Storms on the Quiet Don. A Pilot Study. *Scando-Slavica* XXII, 5–24.
- Kjetsaa, G. (1978), Проблема авторства в романе ‘Тихий Дон’. *Scando-Slavica* XXIV, 91–105.
- Koubourlis, D. J. (1967), *A Statistical Analysis of the Russian Verbal Aspect*. The University of Washington: Ph.D. diss.
- Kučera, H./Monroe, G. K. (1968), *A Comparative Quantitative Phonology of Russian, Czech, and German* (Mathematical Linguistics and Automatic Language Processing, vol. 4). New York.
- Lönngren, L. (ed.) (1993), *Частотный словарь современного русского языка. With a Summary in English. A Frequency Dictionary of Modern Russian* (Acta Universitatis Upsaliensis, Studia Slavica Upsaliensia, 32). Uppsala.
- Morgan, G./ Corbett, G. (1989), Russian Colour Term Salience. *Russian Linguistics* 13, 125–141.
- Mustajoki, A. (1981), Russian stress: Word frequency as an indicator of mobile stress. *Scando-Slavica* XXVII, 199–211.
- Nowikowa, I. (ed.) (1972), *2380 slov. Die gebräuchlichsten Wörter der russischen Umgangssprache*. Hamburg.
- Oliverius, Zd. F. (1976), *Морфемы русского языка. Частотный словарь*. Praha.
- Papp, F. (1964), Mathematische und strukturelle Methoden in der sowjetischen Sprachwissenschaft. *Acta Linguistica Acdemiae Scientiarum Hungaricae* XIV, 119–137.
- Papp, F. (1966), *Mathematical Linguistics in the Soviet Union* (Janua Linguarum, Series Minor No. XL). The Hague.
- Pleskalová, J. (1984), Zur Klassifizierung der slawischen Flurnamen nach Modellen. *Zeitschrift für Slawistik* 29, 80–89.
- Plotnikov, B. A. (1979), Meaning, Word Combination, and Statistics. *Wiener Slawistischer Almanach* 18, 251–260.
- Rakušan, J. (1981), On the Expressivity of Nominal Suffixes in Modern Russian. An Attempt at Quantitative Analysis. *Russian Linguistics* 6, 41–55.

- Reinton, J. E. (1978), The Relationship between English Loanwords and Their Synonyms in Russian Sport Terminology. *Scando-Slavica* XXIV, 217–237.
- Tornow, S. (1984), *Die häufigsten Akzenttypen in der russischen Flexion*. Berlin–Wiesbaden.
- Алексеев, П. М./ Калинин, В. М./ Пиотровский, Р. Г. (ред.), *Статистика речи*. Ленинград (deutsch u.d.T. „Sprachstatistik“, München–Salzburg 1973).
- Апресян, Ю. Д. (1964), О сильном и слабом управлении. *Вопросы языкознания* 3, 32–49.
- Ахманова, О. С./ Мельчук, И. А./ Падучева, Е. В., Фрумкина, Р. М. (1961), *О точных методах исследования языка*. Москва.
- Белоногов, Г. Г./ Фролов, Г. Д. (1963), Эмпирические данные о распределении букв в русской письменной речи. *Проблемы кибернетики* 9, 287–305.
- Белоусова, Е. А. (1969), Статистический анализ некоторых грамматических явлений в русском литературном языке (На материале употребления глагольных форм на -ся). В: *Вопросы теории и истории языка*, Ленинград, 284–291.
- Будилович, А. (1883), *Начертание церковнославянской грамматики, применительно к общей теории русского и других родственных языков*. Warszawa.
- Гаспаров, М. Л. (1963) Статистическое исследование русского трехударного дольника. *Теория вероятностей и ее применения* VIII, 1, 102–108.
- Гаспаров, М. Л. (1965) Вольный хорей и вольный ямб Маяковского. *Вопросы языкознания* 3, 76–88.
- Головин, Б. Н. (1964), *О вероятностно-статистическом изучении стилевой дифференциации языка*. Киев.
- Граудина, Л. К. (1970), Норма и статистика. В: *Актуальные проблемы культуры речи*, Москва, 326–368.
- Граудина, Л. К./ Ицкович, В. А./ Катлинская, Л. П. (1976), *Грамматическая правильность русской речи. Опыт частотно-стилистического словаря вариантов*. Москва.
- Денисов, П. Н./ Костомаров, В. Г. (1969), Стилистическая дифференциация лексики и проблема разговорной речи (по данным „Словаря русского языка“ С. И. Ожегова, изд. 3, М., 1953). В: *Вопросы учебной лексикографии*, Москва, 105–113.
- Денисов, П. Н./ Морковкин, В. В./ Сафьян, Ю. А. (1978), *Комплексный частотный словарь русской научной и технической лексики. 3047 слов*. Москва.

- Дикарев, М. А. (1891), Очерк воронежского мещанского говора сравнительно с украино-русским наречием (опыт статистического исследования звуковых явлений). В: *Воронежский этнографический сборник*, Воронеж, 1–68.
- Елкина, В. Н./ Юдина, Л. С. (1964), Статистика слогов русской речи. В: *Вычислительные системы. Сборник трудов ин-та математики СО АН СССР*, вып. 10, 58–78.
- Ермоленко, Г. В. (1988), *Анонимные произведения и их авторы. На материале русских текстов второй половины XIX–начала XX в.* Минск.
- Засорина, Л. Н. (1977), *Частотный словарь современного русского языка*. Москва.
- Калинина, Е. А. (1968), Частотный словарь русского подъязыка электроники. В: *Статистика речи*, Ленинград, 144–150.
- Караулов, Ю. Н. (1980), *Частотный словарь семантических множителей русского языка*. Москва.
- Кауфман, С. И. (1970), *Из курса по статистической стилистике*. Москва.
- Колгушкин, А. Н. (1970), *Лингвистика в военном деле (Разработка и использование частотных словарей военной лексики)*. Москва.
- Колмогоров, А. Н. (1963), О дольнике современной русской поэзии. Общая характеристика. *Вопросы языкознания* 6, 84–95.
- Колмогоров, А. Н./ Кондратов, А. М. (1962), Ритмика поэм Маяковского. *Вопросы языкознания* 3, 62–74.
- Колмогоров, А. Н./ Прохоров, А. В. (1963), К изучению ритмики Маяковского. *Вопросы языкознания* 4, 64–71.
- Кондратов, А. М. (1962), *Математика и поэтика*. Москва.
- Кондратов, А. М. (1963), Статистика типов русской рифмы. *Вопросы языкознания* 6, 96–106.
- Косачева, Л. Г. (1976), Лексико-статистическое описание морфологических особенностей разговорной речи (глагол). В: *Вопросы лингвостатистического анализа русской разговорной речи, Сб. статей*, Москва, 77–86.
- Кудрявский, Д. Н. (1911), К истории русского прошедшего времени. *Русский филологический вестник* LXV, 1, 119–139.
- Крылов, Ю. К. (1982), Об одной парадигме лингвистических распределений. В: *Tartu Riikliku Ülikooli Toimetised* 628, Tartu, 80–97.
- Крысин, Л. П. (ред.) (1974), *Русский язык по данным массового обследования. Опыт социально-лингвистического изучения*. Москва.
- Крысин, Л. П./ Шмелев, Д. Н. (1976), *Социально-лингвистические исследования*. Москва.

- Курбаков, К. И. (1965), Некоторые характеристики распределения символов русского алфавита. *Научно-техническая информация* 9, 41–46.
- Лебедев, Д. С./ Гармаш, В. А. (1959), Статистический анализ трехбуквенных сочетаний русского текста. *Проблемы передачи информации* II, 78–80.
- Леонтьев, А. А. (ред.) (1977), *Словарь ассоциативных норм русского языка*. Москва.
- Лесскис, Г. А. (1962), О размерах предложений в русской научной и художественной прозе 60-х годов XIX в. *Вопросы языкознания* 2, 78–95.
- Лесскис, Г. А. (1964), О зависимости между размером предложения и его структурой в разных видах текста. *Вопросы языкознания* 4, 99–123.
- Марков, А. А. (1913), Пример статистического исследования над текстом ‚Евгения Онегина‘, иллюстрирующий связь испытаний в цепь. *Изв. Росс. АН*, серия VI, т. 7, 3, 153–162.
- Марков, Ю. (1960), К вопросу о частотности грамматических категорий. *Русский язык в национальной школе* 4, 19–20.
- Марков, Ю./ Вишнякова, Т. (ред.) (1968), *2380 слов, наиболее употребительных в русской разговорной речи*. Москва.
- Милова, Л. В. (ред.) (1994), *От Нестора до Фонвизина. Новые методы определения авторства*. Москва.
- Моисеев, А. (1975), Типология слогов в современном русском литературном языке. *Вопросы языкознания* 6, 109–115.
- Моисеев, А. (1976), Место словесного ударения в современном русском литературном языке. *Studia Rossica Posnaniensia* VII, 77–87.
- Морозов, Н. А. (1915), Лингвистические спектры. Средство для отличения плагиатов от истинных произведений того или другого известного автора. Стилеметрический этюд. *Изв. ОРЯС Имп. АН*, т. XX, кн. 4, 93–134.
- Москович, В. А. (1969), *Статистика и семантика. Опыт статистического анализа семантического поля*. Москва.
- Николаев, В. (1960), Некоторые данные о частотности употребления падежных форм в современном русском литературном языке. *Русский язык в национальной школе* 5, 19–26.
- Никонов, В. А. (1963), Место ударения в русском языке. *IJSLP* VI, 1–8.
- Никонов, В. А. (1967), Личные имена в современной России. *Вопросы языкознания* 6, 102–111.
- Никонов, В. А. (1978), Длина слова. *Вопросы языкознания* 6, 104–111.

- Обратный словарь (1974), *Обратный словарь русского языка. Около 125000 слов*. Москва.
- Пап (= Papp), Ф. (1961), Количественный анализ словарной структуры некоторых русских текстов. *Вопросы языкознания* 6, 93–100.
- Петерсон, М. Н. (1937), Синтаксис слова. *Slavia* 14, 347–392.
- Петрова, Н. В./ Пиотровский, Г. Г. (1966), Слово, контекст, морфология. *Вопросы языкознания* 2, 111–124.
- Пешковский, А. М. (1925/26), Десять тысяч звуков русского языка. В: *Методика родного языка, лингвистика, стилистика, поэтика*. Ленинград–Москва, 167–191.
- Пиотровская, А. А./ Пиотровский, Р. Г./ Разживин, К. А. (1962), Энтропия русского языка. *Вопросы языкознания* 6, 115–130.
- Пиотровский, Р. Г. (1965), Теоретико-информационная структура русского слова. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 18, 149–172.
- Пиотровский, Р. Г. (ред.) (1966), *Энтропия языка и статистика речи*. Минск.
- Пиотровский, Р. Г. и др. (ред.) (1969), *Статистика текста. Сб. статей. Т. 1: Лингвостатистические исследования*. Минск.
- Плотников, Б. А. (1979), *Дистрибутивно-статистический анализ лексических значений*. Минск.
- Полякова, Г. П./ Солганик, Г. Я. (1971), *Частотный словарь языка газеты*. Москва.
- Проскурнин, Н. (1933), Подсчеты частоты литер и комплектовка шрифта. В: *Письменность и революция, сб. I*, Москва–Ленинград, 72–82.
- Пшеничнова, Н. Н. (1973), Статистический анализ многочленного диалектного соответствия (глухие долгие шипящие в русских говорах). *Вопросы языкознания* 6, 80–91.
- Пшеничнова, Н. Н. (1977), Применение метода таксономического анализа к классификации говоров. *Диалектологические исследования по русскому языку*, Москва, 3–14.
- Русское стихосложение (1979), *Русское стихосложение XIX в. Материалы по метрике и строфике русских поэтов*. Ред. коллегия М. Л. Гаспаров, М. М. Гиршман, Л. И. Тимофеев. Москва.
- Тарановски, К. (1953), *Руски дводелни ритмови. I–II*. Београд.
- Тихонов, А. Н. (1983), Система русского словообразования с точки зрения количественных данных. В: *Исследование деривационной подсистемы количественным методом*, Владивосток, 61–73.
- Толстая, С. В. (1968), Фонологическое расстояние и сочетаемость согласных в славянских языках. *Вопросы языкознания* 3, 66–81.

- Турко, Л. А. (1968), Частотный словарь русской разговорной речи. В: *Статистика речи*, Ленинград, 191–199.
- Фрумкина, Р. М. (1960а), Статистическая структура лексики Пушкина. *Вопросы языкознания* 3, 78–81.
- Фрумкина, Р. М. (1960б), Применение статистических методов в языкознании. *Вопросы языкознания* 4, 129–133.
- Фрумкина, Р. М. (ред.) (1963), *Материалы к частотному словарю языка Пушкина (проспект)*. Москва.
- Фрумкина, Р. М. (1964), *Статистические методы изучения лексики*. Москва.
- Харакоз, П. И. (1971), *Частотный словарь современного русского языка*. Фрунзе.
- Чистяков, В. А./ Крамаренко, Б. К. (1929), *Опыт приложения статистического метода к языкознанию*. Вып. I. Краснодар.
- Штейнфельдт, Э. А. (1963), *Частотный словарь современного русского языка. 2500 наиболее употребительных слов*. Таллин.
- Штейнфельдт, Э. А. (1966), *Häufigkeitwörterbuch der russischen Sprache. 2500 meistgebrauchte Wörter der modernen russischen Schriftsprache. Handbuch für Russischlehrer*. Moskau.
- Яглом, И. М./ Добрушин, Р. Л./ Яглом, А. М. (1960), Теория информации и лингвистика. *Вопросы языкознания* 1, 100–110.