

Sebastian Kempgen

## **Phonemcluster und Phonemdistanzen (im Russischen)**

### **1. Einleitung**

Die Beschreibung der phonologischen Struktur einer Sprache besteht u.a. aus einer paradigmatischen und einer syntagmatischen Komponente, d.h. aus einer Beschreibung der phonologischen oder distinktiven Merkmale eines jeden Segmentes und aus einer Beschreibung der Kombinationen der Segmente. Dies kann man auch die internen und die externen Eigenschaften der Phoneme nennen. Die Untersuchung dieser beiden Bereiche gehört seit der Entwicklung der Phonologie als eigenständiger Disziplin zu ihrem Programm.<sup>1</sup>

Allgemein bekannt ist die Tatsache, daß Sprachen keineswegs alle Zweierkombinationen ihrer Konsonanten ausnutzen, bevor sie zu Dreier- oder Viererkombinationen greifen. Es sollte deshalb, so vermutet man, einen Grund geben, warum bestimmte Zweierkombinationen vorkommen, andere nicht. Schon lange ist man auf den naheliegenden Gedanken gekommen, die externen Fähigkeiten der Phoneme, also ihre Kombinationseigenschaften, in Beziehung zu ihren internen Eigenschaften, also ihren Merkmalen, zu setzen. Etwas genauer: die Frage, ob und wie häufig ein Konsonantencluster auftritt, wird in Zusammenhang mit der Ähnlichkeit oder Unähnlichkeit der beteiligten Phoneme hinsichtlich ihrer phonologischen Merkmale gebracht.

Daß die phonologischen Merkmale der beteiligten Phoneme nicht alle Phänomene der Phonemkombinatorik erklären können, zeigt freilich eine einfache Überlegung vorab schon: wie immer man sie auch mißt – die phonologische Ähnlichkeit ist eine reziproke Relation, d.h. die Ähnlichkeit von beispielsweise /p/ zu /r/ ist die gleiche wie zwischen /r/ und /p/. Wenn nun die Folge /pr-/ am Wortanfang vorkommt, /rp-/ aber nicht, so ist klar, daß noch andere Faktoren im Spiel sein müssen. Offensichtlich wird die Silbe als ganze noch nach weiteren Prinzipien organisiert, von denen etwa die "steigende Sonorität" ein Versuch ist, ein solches Prinzip zu erkennen. Was für die phonologische Ebene gilt, muß übrigens auf der phonetischen Ebene nicht gelten: Es läge nahe, hier eine Art von Übergangsschwierigkeit vom ersten zum zweiten Konsonanten zugrunde zu legen, die man auch als gerichtete Relation konstruieren könnte (hierzu s.u.).

Werfen wir einen kurzen Blick auf die im gegebenen Zusammenhang einschlägige Literatur.<sup>2</sup>

---

<sup>1</sup> Vgl. als ausführliche Übersicht zu diesem Thema BIRNBAUM (1967).

<sup>2</sup> Weitere Literatur vgl. in der Auflistung bei ALTMANN/LEHFELDT (1980, 82).

**1.1.** Der erste Versuch einer Erklärung und eines mathematischen Modells für die Analyse von Phonemverbindungen stammt von **SAPORTA** (1955). Er versucht nach eigenen Worten, psycholinguistische Ansätze und solche der strukturalistischen Linguistik miteinander zu verbinden. Insbesondere geht er in seinem kurzen Aufsatz davon aus, daß die relative Häufigkeit sprachlicher Erscheinungen ein wichtiges Merkmal sein könne, um ihre Funktion zu bestimmen. Ohne ZIPF zu nennen, der hierfür die Grundlagen gelegt hat, konstruiert er Situationen, die für Sprecher bzw. für Hörer vom “principle of least effort” bestimmt wären, wenn nämlich die Elemente von Konsonantengruppen einander möglichst ähnlich wären – optimal für den Sprecher, oder wenn sie möglichst unähnlich wären – optimal für den Hörer. Minimierung des Enkodierungs- und des Dekodierungsaufwandes sind also einander widerstrebende Anforderungen, zwischen denen die Sprache einen Ausgleich suche, woraus er folgende Hypothese ableitet: *“the average frequency of a consonant cluster is a function of the difference between the phonemes in the cluster: low frequencies are expected for clusters which are either extremely similar or extremely dissimilar; high frequencies are expected for clusters which are at neither extreme”* (1955, 25).

Für die Berechnung des Unterschiedes zwischen Phonemen benutzt SAPORTA das, was man heute als die “doppelte Phonemdistanz” bezeichnet (s.u.). Auf dieser Grundlage berechnet er am Beispiel des Englischen die Zahl der theoretisch denkbaren Cluster einer jeden Distanz und summiert die Häufigkeiten aller realisierten Cluster einer Distanz, teilt dann jedoch den letzten Wert durch den ersten, um die durchschnittliche Frequenz der Cluster einer gegebenen Frequenz zu ermitteln. Der Ansatz ist zwar im Prinzip richtig, weist jedoch noch methodische Mängel auf – die durchschnittliche Häufigkeit kann sinnvollerweise nur auf die existierenden Cluster bezogen werden, nicht auf alle denkbaren Cluster.<sup>3</sup> Jedenfalls behauptet SAPORTA, die sich ergebenden Kurven ähnelten der Glockenkurve der Normalverteilung. Ein statistischer Test hierfür erfolgt jedoch nicht.<sup>4</sup>

**1.2. CARROLL** (1958) benennt die Unzulänglichkeiten bei SAPORTA und entwickelt ausführlich einen mathematisch korrekten Ansatz. Er besteht in dem allgemeinen Grundgedanken, die beobachtete Frequenzverteilung mit einer theoretischen Verteilung zu vergleichen, und zwar derjenigen, die zu erwarten wäre, wenn *keine* Abhängigkeit zwischen den Phonemdifferenzen und den

---

<sup>3</sup> Der gleiche Fehler wird in einem späteren Aufsatz wiederholt (KELLER/SAPORTA 1957), obwohl dort auch die Anzahl der Cluster-Types tabelliert wird.

<sup>4</sup> Fragwürdig ist auch SAPORTAs Empfehlung, dort die Daten an das Modell anzupassen, wo sie ihm nicht recht entsprechen wollen, also z.B. die Phonemisierungsprinzipien zu ändern.

Clusterfrequenzen bestünde – dies ist die sog. Nullhypothese. Beide Zahlenwerte sollen daraufhin mittels eines statistischen Testes miteinander verglichen und auf signifikante Abweichung bewertet werden, woraufhin die Annahme der Unabhängigkeit akzeptiert oder verworfen werden kann. Eine von ihm vorgenommene erneute Prüfung der von SAPORTA vorgelegten Daten ergibt tatsächlich eine Abhängigkeit der Phonemcluster-Frequenzen von der Phonemdistanz. CARROLL korrigiert aber nicht nur die Methodik zur Untersuchung der genannten Fragestellung, sondern erweitert auch die linguistische Fragestellung, indem er auf die Möglichkeit hinweist, nicht nur das Auftreten der *tokens*, sondern auch das der *types* zu untersuchen. Das heißt: er vermutet nicht nur einen Zusammenhang zwischen den Phonemdistanzen und der Frequenz der Kombinationen, sondern auch zwischen den Phonemdistanzen und der Frage, wieviele verschiedene Kombinationen von jeder Distanz verwirklicht werden.

Seine Tests, die auf Textauszählungen beruhen, zeigen jedoch, daß in bezug auf diese zweite Fragestellung die Nullhypothese anzunehmen ist, d.h. die Frage, welche Cluster existieren und welche nicht, kann *nicht* mit deren interner Struktur beantwortet werden.

**1.3. MULJAČIĆ** (1967) beschäftigte sich auf der großen Phonologie-Tagung von 1966 in Wien ebenfalls mit der Abhängigkeit der Kombinatorik von den internen Merkmalen und untersuchte sie konkret an italienischem und serbokroatischem Material, offensichtlich jedoch ohne Kenntnis der älteren Arbeiten. Zur Bestimmung der doppelten Phonemdistanz beispielsweise greift er auf eine Arbeit von G. HEIKE zum Deutschen zurück. Tatsächlich ist die Arbeit von MULJAČIĆ denn auch in methodischer Hinsicht ein Rückschritt gegenüber CARROLL: er ermittelt nur die existierenden Cluster und ordnet sie nach den (doppelten) Phonemdistanzen, um anschließend die Zusammensetzung der einzelnen Klassen zu analysieren.

Resümierend wird die Ansicht vertreten, in den untersuchten Sprachen schwanke die Phonemdistanz in den realisierten Paaren um einen mittleren Wert und dürfe weder zu klein noch zu groß sein.

**1.4. TOLSTAJA** (1968) ist unmittelbar von MULJAČIĆ inspiriert. Sein Beitrag gilt dieser Autorin als “erster – und erfolgreicher – Versuch einer syntagmatischen Interpretation der phonologischen Distanz, der die Abhängigkeit der distributionellen Besonderheiten eines Phonems von seinen paradigmatischen Charakteristika bekräftige” (1968, 67) – aber nur, weil TOLSTAJA selbst die älteren Ansätze auch nicht kennt.

Die Autorin beschäftigt sich zunächst mit dem Begriff der Phonemdistanz, um dann den Zusammenhang mit der Phonemkombinatorik an polnischem, tschechischem, russischem, serbokroatischem und bulgarischem Material zu untersuchen. Für jede Sprache gibt sie eine Matrix mit den phonologi-

schen Merkmalen, eine Matrix mit der paarweisen (doppelten) Phonemdistanz und eine Tabelle mit der Verteilung der Konsonantencluster des Wortanfangs bzw. des Wortendes auf die Phonemdistancen an, wobei die Zahl der verwirklichten Cluster der Maximalzahl numerisch und in Form einer Graphik gegenübergestellt wird.

Neu ist hieran nur die graphische Darstellung, die Form der Tabellen wird von MULJAČIĆ übernommen. Irgendeine Art statistischer Analyse fehlt; stattdessen werden nur die absoluten Zahlen interpretiert, was unsinnig ist, wie vielleicht schon andeutungsweise klar geworden ist. Außerdem sind die abgedruckten Zahlen in sich inkonsistent und in der vorgelegten Form nicht zu benutzen.<sup>5</sup> Nach Meinung der Autorin jedenfalls bekräftigen ihre Befunde die Resultate von MULJAČIĆ, d.h. sollen Vermeidung sehr kleiner und sehr großer Distanzen sowie eine Bevorzugung mittlerer Distanzen zeigen.

**1.5.** Die nächste Etappe bringt theoretisch-methodische Fortschritte: **ALTMANN** (1969) systematisiert alle Aspekte, hinsichtlich derer sich Phoneme unterscheiden können, darunter auch die Distanz nach den phonologischen Merkmalen.<sup>6</sup> **ALTMANN/LEHFELDT** beschäftigen sich in ihrer "Einführung in die Quantitative Phonologie" erstmals ausführlich mit den theoretischen Grundlagen einer Messung der Phonemähnlichkeit (1980, 80-87). Aus ihren Ausführungen ergibt sich eine Empfehlung für eine Berücksichtigung *aller* phonologischen Merkmale, also positiv wie negativ spezifizierter, und gegen eine Gewichtung der einzelnen Merkmale.

**1.6.** Resümieren wir die Situation, so müssen wir konstatieren, daß bisher für keine einzige slawische Sprache der Zusammenhang zwischen der Phonemparadigmatik und der Phonemkombinatorik in methodisch befriedigender Weise untersucht worden ist. Der vorliegende Beitrag versucht nun, diesen Zusam-

---

<sup>5</sup> Die Inkonsistenz betrifft insbesondere die theoretische Maximalzahl der Cluster einer gegebenen Distanz. Die Summe dieser Zahlen muß gleich dem Quadrat der Phonemzahl sein. TOLSTAJA berechnet nur die Werte für die Hälfte der – natürlich spiegelbildlichen – Matrix inklusive der Diagonale. Die Maximalwerte müssen aber die ganze Tabelle einbeziehen, denn theoretisch existieren kann natürlich sowohl der Cluster /ab/ als auch der Cluster /ba/, auch wenn die Distanzwerte zwischen ihnen gleich sind. Bei TOLSTAJA müssen deshalb alle Maximalwerte bis auf Distanz 0 verdoppelt werden. Bei einigen Sprachen (Tschechisch, Russisch) ergibt die Summe der veränderten Werte tatsächlich das geforderte  $K^2$ , bei anderen Sprachen (Polnisch, Serbokroatisch, Bulgarisch) bleiben Abweichungen, so daß weitere Fehler in den Ausgangszahlen enthalten sein müssen. – In einer späteren Arbeit (TOLSTAJA 1974) werden die am Wortanfang bzw. Wortende existierenden Cluster für einen typologischen Vergleich der slawischen Sprachen benutzt.

<sup>6</sup> Die Aufstellung umfaßt phonetische, phonologische, funktionale und distributionelle Unterschiede sowie Unterschiede hinsichtlich der Frequenz und des Ranges.

menhang für das Russische zu überprüfen. Dabei beschränke ich mich – wie alle anderen Autoren auch – auf zweigliedrige Verbindungen, da die theoretischen Probleme im Zusammenhang mit drei- und mehrgliedrigen Verbindungen noch ungelöst sind.<sup>7</sup>

## 2. Daten zum Russischen

### 2.1. Das Phoneminventar des Russischen

Das Aufstellen eines Phoneminventars ist ein Punkt der Untersuchung, an dem der Eingriff der Theorie in die zugrundeliegenden Daten besonders deutlich wird. Es gibt keinen induktiven Weg von phonetischen Daten zu “dem richtigen” Phoneminventar: das Phoneminventar wird nicht am Korpus aufgefunden, sondern vom Linguisten postuliert. Wesentliches Kriterium bei den zu treffenden Entscheidungen ist u.a. die Frage, welche Beschreibung sich als zweckmäßig, widerspruchsfrei und beschreibungsadäquat erweist, d.h. welche Beschreibung unter der gewählten Fragestellung die Daten vollständig zu erfassen und zu explizieren in der Lage ist (vgl. PILCH 1974, 104f.). Ein Kernbestand von Phonemen ist für das Russische natürlich unbestritten (vgl. PANZER 1975, 32), und die Zahl der Phoneme schwankt meist um die 40. Alle Beschreibungen mit weniger als 48 Phonemen – und das sind mit einer Ausnahme (LEKOMCEVA 1968, 138f.) alle – weisen das Russische als Sprache mit normal hoher Phonemzahl aus (vgl. LEHFELDT 1975, 285).

Das Phonemsystem, von dem ich hier ausgehe, umfaßt die folgenden Elemente:

*Vokale:* /a e i o u/

*Konsonanten:* /b b' v v' g g' d d' ž z z' j k k' l l' m m' n n'  
p p' r r' s s' t t' f f' x c č š/

Dieses Inventar entspricht weitgehend den Darstellungen von KUČERA/MONROE (1968, 22ff.) und OLIVERIUS (1974, 95).<sup>8</sup>

### 2.2. Die distinktiven Merkmale der russischen Phoneme

Für die Beschreibung der phonologischen Merkmale greife ich auf PANZER (1975, 33f.) zurück, der eine Mischung aus artikulatorisch interpretierbaren und akustischen Merkmalen nach JAKOBSON/HALLE verwendet.

<sup>7</sup> TOLSTAJA (1968, 71) spricht diese Probleme wenigstens kurz an, ohne sie jedoch vollständig abzuhandeln. Die Schwierigkeiten ergeben sich u.a. daraus, daß die Phonemdistanz keine transitive Relation ist – es gilt also nicht logisch, daß  $D(a, b) + D(b, c) = D(a, c)$ .

<sup>8</sup> Bei KUČERA/MONROE fehlt /g'/, weil sich in der Standardsprache kein Minimalpaar finden lasse; bei OLIVERIUS findet sich zusätzlich /x'/.

Die Beschreibung benutzt insgesamt 14 Merkmale, wobei jedes Phonem im Hinblick auf alle Merkmale vollständig – positiv oder negativ – spezifiziert wird, d.h. es liegt keine möglichst redundanzfreie Transkription vor, was für unsere Zwecke im Augenblick auch nicht erforderlich ist. Die im weiteren erhaltenen Ergebnisse sind natürlich von den Entscheidungen, die dieser Beschreibung zugrundeliegen, abhängig – eine andere Beschreibung könnte folglich eventuell zu anderen Ergebnissen führen.

### 2.3. Einfache Phonemdistanz

Da die Hypothese besagt, daß das Auftreten und die Häufigkeit der einzelnen Phonemcluster von der Ähnlichkeit bzw. Differenz der beteiligten Elemente abhängt, muß im nächsten Schritt die Differenz zwischen den Phonemen quantifiziert werden.

Hierfür gibt es in der Literatur mehrere Vorschläge. Ich benutze hier die sogenannte „*einfache Phonemdistanz*“, bei der alle Phoneme paarweise in bezug auf die Ausprägungen ihrer paradigmatischen phonologischen Merkmale miteinander verglichen werden.<sup>9</sup> Dabei werden die theoretisch möglichen Resultate folgendermaßen bewertet:

Phonem A	Phonem B	Wertung
+	+	0
-	-	0
0	0	0
+	-	1
-	+	1
0	+/-	0,5
+/-	0	0,5

Die Distanz zwischen zwei Phonemen wird als Summe aller Einzelwertungen berechnet, alle Merkmale werden also gleich gewichtet:

$$D(a,b) = \sum_{i=1}^n |a_i - b_i|$$

Diese Phonemdistanz ist im übrigen eine symmetrische Relation: die Distanz zwischen /a/ und /b/ ist gleich der zwischen /b/ und /a/.

Bei 14 Merkmalen liegt der theoretische Höchstwert der Differenz folglich bei 14. Da bei PANZER nur eine positive oder eine negative Angabe vor-

<sup>9</sup> TOLSTAJA (1968, 66) nennt weitere Möglichkeiten, z.B. die Anzahl der Schritte, um in einer Baum-Darstellung von einem Phonem zum anderen zu kommen. Sie spricht auch die Möglichkeit (und Schwierigkeit) an, eine (angenommene) Hierarchie der phonologischen Merkmale in eine Gewichtung derselben umzusetzen. Vgl. auch die Ausführungen von LEHFELDT (1978), BELOOZEROV (1964, 55), LEVIN (1964, 114).

kommt, können als Distanzwert selbstverständlich nur ganzzahlige Werte auftreten. Es zeigt sich, daß der höchste tatsächlich vorkommende Werte 10 ist (z.B. zwischen /t' – a, t' – o, t' – u/ usw.). Der niedrigste Wert, also Null, kann natürlich nur zwischen identischen Phonemen auftreten, also bei Doppelkonsonanz. **Tab. 1** zeigt die einfachen Phonemdistanzen für das Russische.

Sinnvoller als die Verwendung der einfachen oder doppelten Phonemdistanz ist im übrigen die *relative Phonemdistanz*, wie auch von ALTMANN/LEHFELDT (1980) verwendet. Bei ihr werden die absoluten Werte auf das jeweilige theoretische Maximum, in unserem Falle also 14, bezogen und so in das Einheitsintervall <0;1> transformiert. Damit ist eine Vergleichbarkeit zwischen verschiedenen Phonemmodellen, verschiedenen Sprachen und verschiedenen Studien gegeben, was bei den absoluten Werten natürlich nicht der Fall ist. Wir ersparen es uns an dieser Stelle aus Platzgründen, diese transformierten Werte abzudrucken.

Wenn die phonologischen Merkmale artikulatorisch definiert sind, liegt es nahe, die Distanz zwischen den Elementen eines Clusters mit der Schwierigkeit, ihn auszusprechen, in Verbindung zu bringen. Schon das vorwissenschaftliche Sprecherbewußtsein spricht ja bekanntlich von den "schwierigen" Lautfolgen einer fremden Sprache. Es gibt jedoch bisher kein allgemein akzeptiertes Verfahren, die Schwierigkeit des Sprechbewegungsablaufes objektiv zu quantifizieren.<sup>10</sup> Die Distanzmessungen auf der Phonemebene dürfen deshalb nicht im Sinne der Sprechschwierigkeit mißverstanden werden, auch wenn sie vielleicht in einem abstrakteren Sinne den gleichen Objektbereich modellieren.

---

<sup>10</sup> Beachtenswerte Vorschläge und kritische Diskussion hierzu stammen von LEHFELDT (1980) sowie GROTHJAHN (1980a, 1980b). Vgl. auch FRUMKINA/VASIL'EVIC (1968), LADEFOGED (1970).





## 2.4. Initiale und finale Konsonantencluster des Russischen

Die Frequenz aller initialen und finalen zweigliedrigen Konsonantencluster des Russischen wurde anhand zweier Wörterbücher erhoben, und zwar dem ORFOGRAFIČESKIJ SLOVAR' (1967) und dem OBRATNYJ SLOVAR' (1974). Bei unserem Material handelt es sich folglich um eine sog. systemische Stichprobe, nicht um eine pragmatische. Für sie gilt zudem die Einschränkung, daß natürlich nur die sogenannten Nennformen der flektierenden Lexeme erfaßt werden. Die zahlreichen Konsonantenkombinationen, die sich beispielsweise in den endungslosen Genitiv-Plural-Formen ergeben, können so nicht berücksichtigt werden. Die so erhobenen Daten lassen sich natürlich in zweierlei Hinsicht auswerten: Sie geben Auskunft über die *types*, also darüber, welche Cluster es überhaupt gibt. Und sie lassen die *tokens* erkennen, d.h. besagen, wie oft die *types* ausgenutzt werden. Zunächst aber einmal gilt: Dies sind graphemische Daten.

## 2.5. Prinzipien der Phonemisierung

Um das untersuchte Korpus *phonetisch* zu transkribieren, wurde als Standardwerk AVANESOV (1972) herangezogen. In diesem Werk wird der sogenannte "neue Standard" gegenüber der "alten Moskauer Aussprache" dargestellt; er zeichnet sich u.a. durch größere Nähe zur graphematischen Ebene aus (vgl. auch KUČERA/MONROE 1968, 22f.), insbesondere bei den vielfältigen Palatalitätsassimilationen.

Bei der *phonologischen* Interpretation der phonetischen Daten erfolgte die Entscheidung für ein Modell, das der Ebene der Phonemrealisierungen sehr nahesteht, weil dies für unsere Fragestellung einzig sinnvoll ist. Ähnlich gehen auch KUČERA/MONROE (1968) vor. Das heißt also: die bekannten Assimilationen nach Stimmbeteiligung, Palatalität, Artikulationsort und -modus werden nicht auf die allophonische Ebene abgeschoben, sondern als Regularitäten der Phonemebene betrachtet. Einen Konsonantencluster \*/st'-/ beispielsweise kann es in unserem Modell nicht geben, weil dentale Frikative vor palatalisierten dentalen Plosiven typischerweise regressiv palatalisiert werden, so daß wir in diesen Fällen immer /s't'-/ schreiben.

Die **Tab. 2** und **Tab. 3** zeigen die auf die Weise phonologisch interpretierten graphematischen Daten: die wortinitialen bzw. wortfinalen zweigliedrigen Clusterfrequenzen des Russischen.

	b	b'	v	v'	g	g'	d	d'	ž	z	z'	j	k	k'	l	l'	m	m'	n	n'	p	p'	r	r'	s	s'	t	t'	f	f'	x	c	c'	š	Σ	
b								3							201	112							279	150											745	
b'												2																								2
v	4	11	27		9	4	98	28	9	72	18	15			61	45	25	53	73	73			111	36												772
v'			58							22																										80
g			23	5			6	1							173	65	2	46	39				366	161												887
g'																																			0	
d			312	32					27	1	5				1	33	4	12					182	98												707
d'												14																							14	
ž	2		3		9	1		41									9	1	7			4	15												92	
z	58		78		43	8	55		4					102			55					3	19												425	
z'			34	62			36								6	32																			170	
j																																			0	
k			102	26										125	217	1	11	52				1	542	341	1	17	13	1							1451	
k'												1																							1	
l	2				5			30							1																				38	
l'			1	4	3		4	3				4							32	2															53	
m															12	6		15				19	1												238	
m'																																			0	
n																																				10
n'																																				10
p															470	207							2956	2933	16	76	23	26	1	2	1				1	6768
p'										27																										27
r			13	1			5	22																											45	
r'																																			2	
s	149												36	560	69	338	250	104	227			103	99	62	712	33			65	43					2850	
s'			467												196	181	102	207							12	428	22								1615	
t	19		25											15	11	4						511	452												1037	
t'																	3										1								4	
f														71	6	59	68					66	79	117	60	55	111	99	44						899	
f'																																			1	
x			76												125	104	6	14	5			131	77												538	
c																																			45	
č			5												17	7																			48	
š	32		20											48	4	18	79	6	7	13	4	63	30	4	8										808	
Σ	64	47	840	745	69	12	158	81	130	77	23	125	694	79	1686	1166	306	295	520	335	356	317	5338	4461	134	216	998	532	35	24	79	53	314	64	20373	

Tab. 2: Frequenzen zweigliedriger Anlautgruppen (systemisch)

	b	b'	v	v'	g	g'	d	d'	ž	z	z'	j	k	k'	l	l'	m	m'	n	n'	p	p'	r	r'	s	s'	t	t'	f	f'	x	c	c'	š	š'	Σ
b																9							9	6												24
b'																3				1			25													0
v																																				29
v'																																				0
g																																				9
g'																																				0
d																																				14
d'																																				0
ž																																				0
z																																				574
z'																																				17
j																																				53
k																																				132
k'																																				0
l																																				32
l'																																				56
m																																				26
m'																																				0
n																																				464
n'																																				4
p																																				35
p'																																				0
r																																				374
r'																																				0
s																																				841
s'																																				4615
t																																				163
t'																																				0
f																																				31
f'																																				0
x																																				9
c																																				0
č																																				1
š																																				30
Σ	0	0	0	0	0	0	0	0	0	0	0	0	0	254	0	18	35	622	0	39	24	32	1	213	12	172	7	1338	4633	23	7	25	14	37	27	7533

Tab. 3: Frequenzen zweigliedriger Auslautgruppen (systemisch)

### 3. Test der Hypothese

Die Hypothese zum Zusammenhang zwischen distinktiven Merkmalen und Clusterfrequenzen wollen wir im folgenden dreifach prüfen. Die ersten beiden Tests übernehmen wir dabei von SAPORTA bzw. CARROLL, mit dem dritten Test entwickeln wir die bisherigen Ansätze um einen neuen Gedanken weiter.

#### 3.1. Existenz der Clustertypes

Zunächst wollen wir die Frage prüfen, ob die Zahl der *verschiedenen* Cluster, also der *types*, die es im Russischen am Wortanfang wie am Wortende gibt, von der internen Struktur dieser Cluster abhängt. In **Tab. 4** und **Tab. 5** sind die Werte zusammengestellt, die für eine Überprüfung der Hypothese benötigt werden, und zwar getrennt für Wortanlaut wie für Wortauslaut.

$x_i$	$\max_i$	$n_i$	$n_i / \max_i$	$E(n_i)$	$\chi^2$
0	34	8	0.2353	6.44	
1	84	9	0.1071	15.91	1.28
2	118	15	0.1271	22.36	2.42
3	168	26	0.1548	31.83	1.07
4	256	56	0.2188	48.50	1.16
5	264	53	0.2008	50.01	0.18
6	168	33	0.2063	30.31	0.24
7	58	14	0.2414	10.99	
8	12	4	0.3333	2.27	2.11
9	2	1	0.5000	0.38	
10	0	0	–		
$\Sigma$	1156 = $N_{\max}$	219 = $N_n$	rel. Anteil	Erwartungswert	8.46

Tab. 4: Anlaut und Phonemdistanzen im Russischen

$x_i$	$\max_i$	$n_i$	$n_i / \max_i$	$E(n_i)$	$\chi^2$
0	34	0	–	3.38	
1	84	4	0.0476	8.36	<b>5.10</b>
2	118	2	0.0170	11.74	<b>8.08</b>
3	168	14	0.0833	16.71	0.44
4	256	30	0.1172	25.47	0.81
5	264	26	0.0985	26.26	0.0003
6	168	22	0.1375	15.92	
7	58	11	0.1897	5.77	
8	12	4	0.3333	1.19	<b>10.98</b>
9	2	2	1.0000	0.20	
10	0	0	–		
$\Sigma$	1156 = $N_{\max}$	115 = $N_n$	rel. Anteil	Erwartungswert	<b>25.41</b>

Tab. 5: Auslaut und Phonemdistanzen im Russischen

In Spalte 1 steht die Phonemdistanz  $x_j$  vom Minimalwert 0 bis zum beobachteten höchsten Wert 10 (der allerdings nur zwischen Vokalen und Konsonanten vorkommt). Wertet man die Tabelle mit den Phonemdistanzen aus, so kommt man auf die Werte  $max_j$ , die in der zweiten Spalte stehen. Sie geben an, wieviele verschiedene Cluster der gegebenen Distanz es maximal geben könnte, dann nämlich, wenn sich jeder Konsonant mit jedem anderen verbände. Deshalb ist die Gesamtzahl  $N_{max}$  gleich  $K^2 = 34^2 = 1156$ . Jeder existierende Cluster (zu entnehmen den Tabellen 2 und 3) läßt sich nun daraufhin überprüfen, welche Distanz zwischen seinen Elementen besteht. Dies ergibt dann die Zahl der tatsächlich verwirklichten Cluster einer gegebenen Distanz, die in Spalte drei unter  $n_j$  steht. Das heißt also z.B.: von 34 möglichen Doppelkonsonanten – Distanz natürlich 0 – kommen am Wortanfang 8 Paare tatsächlich vor, von 84 Clustern mit der Distanz 1 ganze 9 usw. Wie hoch der Anteil der tatsächlich vorhandenen Paare an den theoretisch möglichen ist, zeigt die vierte Spalte, also  $n_j/max_j$ .

Werfen wir einen Blick auf die Ergebnisse.

Betrachtet man die absoluten Zahlen ( $max_i$  und  $n_i$ , Spalten 2 und 3) der Anlaut- und Auslaut-Werte, so ergibt sich bei graphischer Umsetzung der Zahlen die Andeutung einer Glockenkurve mit einem Gipfel bei Distanz 4 bzw. 5 (vgl. **Fig. 1**). Dies hat die früheren Autoren dazu gebracht, anzunehmen, geringe und hohe Distanzen würden vermieden, mittlere bevorzugt. Selbstverständlich ist es richtig, daß von den Clustern mit mittlerer Distanz besonders viele verschiedene existieren. Das kann aber gar nicht verwundern, weil es für sie ja auch besonders viele verschiedene Möglichkeiten gibt, wie die Maximalwerte zeigen. Es wäre also im Gegenteil verwunderlich, wenn dies nicht der Fall wäre!

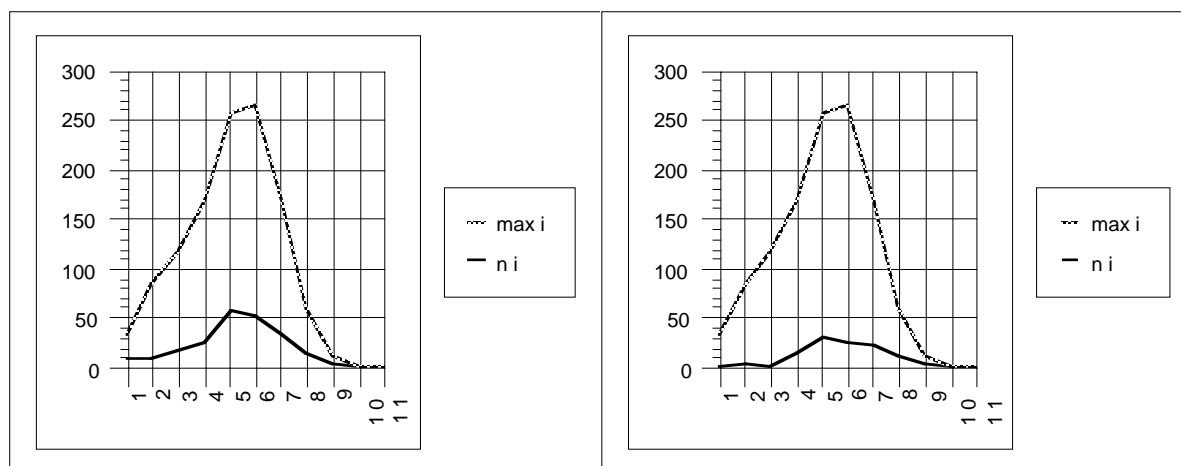


Fig. 1: Clustertypes nach Distanzen (Wortanfang/Wortende)

Alle Autoren, die nur diese absoluten Zahlen mit ihrer Glockenform betrachten und daraus entsprechende Schlüsse ableiten, und das sind SAPORTA, MULJAČIĆ und TOLSTAJA, erliegen also einem Trugbild.

Die beobachteten Werte müssen unbedingt auf die Maximalwerte bezogen werden, um die wirklichen Verhältnisse zu ermitteln. Und bei diesen relativen Anteilen ergibt sich nun ein doch etwas anderes Bild:

Bei den initialen Verbindungen steigt der Anteil der realisierten Verbindungen von der Distanz 1 an kontinuierlich an: von 10 auf 50 Prozent. Nur die Doppelkonsonanten fallen aus dieser Abfolge etwas heraus. Bei den finalen Clustern zeigt die Kurve ein kleines Auf und Ab, die Tendenz ist aber die gleiche. Es scheint also zu gelten: Je unterschiedlicher die Bestandteile einer Konsonantenverbindung sind, desto mehr Verbindungen von allen theoretisch möglichen gibt es. Dies könnte darauf hindeuten, daß – jedenfalls im Russischen – eher folgendes gilt: *zu ähnliche Cluster werden vermieden und solche mit stärker differenzierter Struktur bevorzugt.*

Die Frage, ob eine derartige Abhängigkeit zwischen der Phonemdistanz und der Zahl der verschiedenen Cluster dieser Distanz besteht, kann mit einem einfachen statistischen Test überprüft werden: Man berechnet einen Erwartungswert und vergleicht diesen mit dem beobachteten Wert. Sind die Unterschiede signifikant, so liegt ein linguistisch interpretierbares Ergebnis vor.

In Analogie zu den Vorschlägen von CARROLL ergibt sich der Erwartungswert für die Zahl der Cluster pro Distanz folgendermaßen: Die Anzahl der insgesamt verwirklichten verschiedenen Cluster wird in Abhängigkeit von dem theoretischen Maximum von Clustern einer bestimmten Distanz prozentual auf die einzelnen Distanzen aufgeteilt, also:

$$E(n_i) = \max_j \frac{N_n}{N_{\max}}$$

An einem Beispiel sei die Berechnung demonstriert: Von den insgesamt 1156 theoretisch denkbaren Konsonantenverbindungen werden 118 der Phonemdistanz 2 zugerechnet. Tatsächlich existieren jedoch nicht 1156, sondern nur 219 verschiedene Kombinationen am Wortanfang. Es sind deshalb für die Distanz 2 auch nicht 118 verschiedene Cluster zu erwarten, sondern nur ein entsprechender Prozentsatz, der dem Verhältnis von 219 zu 1156 entspricht, also  $219/1156 = 0.1894$  oder knapp 19%. Multipliziert man 118 mit diesem Faktor, so ergibt dies 22.3546. Das heißt also: es wären gut 22 verschiedene Cluster der Distanz 2 zu erwarten, wenn sich die tatsächlich existierenden Cluster gleichmäßig auf alle Distanzen verteilten – und damit unabhängig von ihnen wären.

$$E(n_{i=2}) = 118 \frac{219}{1156} = 22.3546$$

Die Erwartungswerte sind in den beiden Tabellen ebenfalls aufgeführt.

Der Unterschied zwischen dem beobachteten Wert (in unserem Beispiel 15 für die Distanz 2) und dem Erwartungswert (22.36) wird sodann mit dem sog. Chi-Quadrat-Test auf Signifikanz getestet.<sup>11</sup> Die Ergebnisse stehen ebenfalls in den Tabellen. Hier gilt: Je größer der Wert, desto größer die Abweichung. Ein Einzelergebnis ist signifikant, wenn  $\chi^2 > 3.84$ . Signifikante Werte sind fett gedruckt. Die Summe der Chi-Quadrat-Werte steht für die Wahrscheinlichkeit, mit der die beobachtete Verteilung *insgesamt* zufällig zustande kommen könnte. Ist die Wahrscheinlichkeit *groß*, wird man sagen, daß *keine* Abhängigkeit zwischen Phonemdistanz und Clusterzahl besteht. Ist die Wahrscheinlichkeit klein, wird die Hypothese von der Unabhängigkeit verworfen.<sup>12</sup> In unserem Falle ergibt sich:

Wortanfang	FG = 6	P << 0.2	⇒	Unabhängigkeit
Wortende	FG = 5	P << 0.0001	⇒	Abhängigkeit

Das heißt also: Für den *Wortanfang* wird die Annahme, daß die Zahl der verwirklichten Konsonantenverbindungen von der Distanz zwischen den beteiligten Phonemen abhängt, in dieser Pauschalität *nicht* bestätigt, für das *Wortende* hingegen sehr wohl. Dies entspricht im übrigen CARROLLS Ergebnissen zum Englischen und zeigt, daß die phonotaktischen Restriktionen am Wortende größer sind als am Wortanfang. Eine Tendenz ist zugleich deutlich: Bei Clustern geringer Distanz (0 – 3) gibt es ausnahmslos weniger Verbindungen als erwartet, bei größeren Distanzen (4 – 9) ist es genau andersherum. Das heißt aber: damit ist SAPORTAS ursprüngliche Hypothese in bezug auf die bloße Cluster-Existenz nur teilweise bestätigt, und statistisch signifikant nur für das Wortende.

### 3.2 Clusterhäufigkeit – Tokens

Der zweite Test der Hypothese soll nun darin bestehen, auch die Frequenzen, mit denen die einzelnen Cluster auftreten, zu berücksichtigen. Diese Frequenzen finden sich ja in den Tab. 2 und Tab. 3.

Für den Test summiert man alle Häufigkeiten der Cluster einer bestimmten Distanz. Diese Werte stehen in der **Tab. 6** jeweils unter  $f_j$ . Ein Beispiel: Die acht Doppelkonsonanten treten zusammen 227 mal am Wortanfang auf, die neun Cluster der Distanz 1 hingegen werden 564 mal benutzt usw. Auch hier zeigen die beobachteten Werte und die Erwartungswerte wieder die schon

<sup>11</sup> Bei diesem Test sollen erwartete Werte nicht kleiner als 10 sein, weshalb gegebenenfalls benachbarte Gruppen zusammengefaßt werden.

<sup>12</sup> Der Schwellenwert für die Entscheidung zwischen beiden Möglichkeiten wird konventionell gewählt, gewöhnlich als  $\alpha = 0.05$ , d.h. bei weniger als 5% Wahrscheinlichkeit spricht man von einem interpretierbaren Zusammenhang. "FG" steht für die Anzahl der sog. Freiheitsgrade, d.h. für die Zahl der unabhängigen Entscheidungen, ist also um einen Punkt kleiner als die Zahl der getesteten Klassen.

bekannte Glockenform, das Wortende dabei jedoch viel extremer als der Wortanfang: wirklich häufig sind am Wortende nur Cluster mit der Distanz 3 und 4.

$x_i$	Wortanfang			Wortende		
	$f_i$	$E(f_i)$	$\chi^2$	$f_i$	$E(f_i)$	$\chi^2$
0	227	332.62	34.58	0	72.75	72.75
1	564	866.93	105.85	18	284.99	250.12
2	1115	1023.02	8.27	28	127.33	77.40
3	2035	2923.37	269.97	5795	3250.55	1991.74
4	3664	4687.81	233.60	1124	2182.52	513.38
5	5273	5049.14	9.93	100	859.62	671.25
6	4596	3421.19	403.42	126	653.73	426.01
7	1837	1382.46	149.45	251	61.68	581.11
8	957	583.18	239.62	72	38.43	
9	125	122.91	0.04	19	1.39	78.65
$\Sigma$	20393		1443.67	7533		4662.41

Tab. 6: Clusterfrequenzen und Phonemdistancen im Russischen

Das Testverfahren ist den oben dargestellten Gedankengängen ganz analog und braucht hier nicht noch einmal erläutert zu werden.<sup>13</sup> In unserem Falle ergibt sich:

Wortanfang	FG = 9	$P \ll 0.0001$	⇒	Abhängigkeit
Wortende	FG = 8	$P \ll 0.0001$	⇒	Abhängigkeit

Das heißt also: sowohl für den Wortanfang wie für das Wortende wird mit fast 100% Sicherheit die Hypothese, daß *kein* Zusammenhang zwischen der Phonemdistanz und der Frequenzverteilung besteht, abgelehnt. Oder andersherum: die Ausgangsfrage nach einem möglichen Zusammenhang zwischen Phonemdistanz und Häufigkeit der entsprechenden Cluster findet in dieser allgemeinen Form und Formulierung bei Berücksichtigung der Clusterfrequenzen eine überzeugende Bestätigung, und zwar für den Wortanfang wie für das Wortende.

Trotzdem müssen die Werte noch etwas genauer betrachtet werden, um SAPORTAs Hypothese in ihrer speziellen Form zu prüfen. Sie würde, in statistische Terminologie übersetzt, ja lauten, daß bei sehr kleinen und bei sehr großen Distanzen die tatsächlichen Frequenzen unter den Erwartungswerten bleiben müßten, während mittlere Distanzen häufiger ausgenutzt werden soll-

<sup>13</sup> Die Anwendung des Chiquadrattests erfolgt auch hier nach CARROLL.



ten. Diese Annahme stimmt in dieser Form offenbar nicht ganz: Am Wortanfang beobachten wir vielmehr wieder, wie oben, daß Cluster geringer Distanz (0 – 1 und 3 – 4) weniger häufig realisiert werden als erwartet, Cluster größerer Distanz (5 – 9) dagegen durchweg häufiger. (Bei den Clustern mit der Distanz 2 liegen tatsächlicher und erwarteter Wert ziemlich nahe beieinander.) Das heißt also: am Wortanfang werden zu ähnliche Cluster eher vermieden und unähnliche bevorzugt. Am Wortende hingegen sieht das Bild schon eher so aus, wie SAPORTA es gemalt hat: Cluster der Distanz 3 werden eindeutig bevorzugt, Cluster geringer Distanz (0 – 2) und größerer Distanz (4 – 6) vermieden. Erst bei sehr großer Distanz (7 – 9) stimmt das Wortende wieder mit dem Wortanfang überein.

### 3.3. Clusterstatus

Wenn die Frequenzen der einzelnen Phonemverbindungen berücksichtigt werden, können über den Status der Verbindungen genauere Aussagen gemacht werden als nur die, daß sie existieren oder nicht. Jede Phonemverbindung erfährt aufgrund ihrer Frequenz eine bestimmte “Gewichtung”. Diese Frequenz kann auch benutzt werden, um den *Status* des betreffenden Clusters in der Sprache klassifikatorisch zu charakterisieren. Hier hat ALTMANN (1973) ein mathematisches Modell entwickelt, das von ALTMANN/LEHFELDT (1980) weiter präzisiert worden ist.<sup>14</sup>

In diesem Modell wird jede Phonemverbindung daraufhin überprüft, ob sich ihre Elemente in einer assoziativen, neutralen oder dissoziativen Tendenz zueinander befinden, je nachdem, ob die tatsächliche Frequenz  $n_{ij}$  signifikant größer, gleich oder kleiner als der theoretische Erwartungswert  $E_{ij}$  ist. Durch die Berücksichtigung der Tatsache, ob die betreffende Verbindung existiert, kommt man zu einer sechsfachen Gruppierung, von denen fünf Gruppen Sinn machen (vgl. ALTMANN 1973, 104):

	$n_{ij} > E_{ij}$	$n_{ij} = E_{ij}$	$n_{ij} < E_{ij}$
$n_{ij} > 0$	P = bevorzugt	A = aktuell	M = marginal
$n_{ij} = 0$	–	V = virtuell	I = unzulässig

Als “bevorzugt” werden solche Verbindungen gekennzeichnet, die eine starke Assoziationstendenz aufweisen, also besonders häufig auftreten. Sie charakterisieren eine Sprache in besonderem Maße. Die Verbindungen, die mit ihrer Frequenz von dem zu erwartenden Wert nicht signifikant abweichen, werden als “aktuell” bezeichnet. Es sind dies die normalen Cluster einer Sprache. Wenn die tatsächliche Frequenz signifikant unter der zu erwartenden liegt, wird dies als dissoziative Tendenz gewertet und der Cluster als “marginal” be-

<sup>14</sup> Zu einem anderen Ansatz vgl. BARINOVA/ŠEVOROŠKIN (1972).

zeichnet. Solche Cluster können z.B. für den peripheren Bereich des phonotaktischen Systems einer Sprache charakteristisch sein. Existiert eine denkbare Kombination nicht, so wird sie als "virtuell" bezeichnet, wenn die beobachtete Frequenz damit dem Erwartungswert entspricht. Dies könnte man die "zufälligen Lücken" in der Phonemkombinatorik nennen. Eine nicht existierende Kombination wird hingegen als "unzulässig" bezeichnet, wenn sich die beobachtete Frequenz (also Null) signifikant unter dem Erwartungswert befindet. Solche Kombinationen könnte man als "strukturelle Lücken" in der Phonemkombinatorik bezeichnen.

Die Namen dieser fünf Klassen sind natürlich reine Etiketten, genauso gut könnte man andere Bezeichnungen verwenden, falls die vorliegenden als nicht glücklich empfunden würden. Sie sind jedoch in jedem Fall prägnanter und griffiger als die präzisen Testergebnisse, für die sie stehen.

Der eigentliche statistische Test soll hier nicht im einzelnen dargestellt werden, da er sich in der genannten Literatur nachlesen läßt. Ergebnis der Anwendung ist eine Tabelle, in der die Frequenzwerte durch die Klassensymbole P – A – M – V – I ersetzt werden. Die **Tab. 7** und **Tab. 8** (im Anhang) zeigen die Ergebnisse für das Russische. Von diesen fünf Klassen interessieren uns im folgenden allerdings nur die ersten drei, das heißt die Bewertung derjenigen Cluster, die tatsächlich existieren.

Bei der Frage, ob Phonemdistanz und Clusterexistenz bzw. Clusterfrequenz zusammenhängen, könnte es naheliegen zu vermuten, daß von den Clustern, die im Russischen als marginal eingestuft werden, besonders viele unter geringen bzw. hohen Phonemdistanzen zu finden sind. Umgekehrt könnte es sein, daß sich die präferierten Konsonantencluster vorzugsweise unter denjenigen mit mittlerer Distanz befinden. Ob dies so ist, soll abschließend überprüft werden. Die **Tab. 9** und **Tab. 10** schlüsseln die Anteile der präferierten, aktuellen und marginalen Cluster an den existierenden Verbindungen im Anlaut bzw. Auslaut auf.

$x_i$	$n_i$	$n_P$	$n_A$	$n_M$	$n_P/n_i$	$n_A/n_i$	$n_M/n_i$
0	8	4	4	0	0,500	0,500	0,000
1	9	8	0	1	0,889	0,000	0,111
2	15	8	3	4	0,533	0,200	0,267
3	26	13	7	6	0,500	0,269	0,231
4	56	32	10	14	0,571	0,178	0,250
5	53	26	15	12	0,490	0,283	0,226
6	33	19	6	8	0,576	0,182	0,242
7	14	7	3	4	0,500	0,214	0,286
8	4	4	0	0	1,000	0,000	0,000
9	1	0	1	0	0,000	1,000	0,000
10	0	–	–	–	–	–	–
$\Sigma$	219 $N_n$	121 $N_P$	49 $N_A$	49 $N_M$			

Tab. 9: Wortanfang und Clusterstatus im Russischen

$x_i$	$n_i$	$n_P$	$n_A$	$n_M$	$n_P/n_i$	$n_A/n_i$	$n_M/n_i$
0	0	–	–	–	–	–	–
1	4	3	0	1	0,750	0,000	0,250
2	2	2	0	0	1,000	0,000	0,000
3	14	11	2	1	0,786	0,143	0,071
4	30	23	3	4	0,767	0,100	0,133
5	26	15	6	5	0,577	0,231	0,192
6	22	13	7	2	0,591	0,318	0,909
7	11	8	3	0	0,727	0,273	0,000
8	4	2	2	0	0,500	0,500	0,000
9	2	2	0	0	1,000	0,000	0,000
10	0	–	–	–	–	–	–
$\Sigma$	115 $N_n$	79 $N_P$	23 $N_A$	13 $N_M$			

Tab. 10: Wortende und Clusterstatus im Russischen

Diese Aufschlüsselung nach dem Status der Verbindung zeigt ein überraschendes Ergebnis, wenn wir die relativen Werte betrachten: Die präferierten (P) und die aktuellen Verbindungen (A) haben einen besonders hohen Anteil an den Clustern mit geringer oder hoher Differenz; “in der Mitte” liegt ihr Anteil bei ziemlich konstanten 50% bzw. 20-25%. Der Anteil der marginalen Verbindungen (M) ist hingegen bei den Clustern mit mittlerer Distanz deutlich höher als bei den Extrema. Also scheint zu gelten: Wenn es schon Cluster mit geringer oder hoher interner Distanz gibt, dann haben diese Cluster aber in der Sprache mindestens einen “normalen” Status.

#### 4. Ergebnisse

Fassen wir die verschiedenen Ergebnisse noch einmal zusammen:

1) Die Annahme, es gebe eine Abhängigkeit zwischen den Phonemdistanzen und der Phonemkombinatorik, findet keine sehr gute Bestätigung, wenn man nur die bloße Existenz der Cluster, also die *types*, berücksichtigt.

2) Die Annahme einer Abhängigkeit ergibt sich eindeutig bei Berücksichtigung der Clusterfrequenzen, also bei den *tokens*.

3) Wortanfang und Wortende weisen im Russischen eine so unterschiedliche Struktur auf, daß sie nicht durchweg über einen Kamm geschoren werden dürfen.

4) Die Vorstellung SAPORTAs, wonach Cluster geringer und hoher Distanz vermieden und solche mittlerer Distanz bevorzugt werden sollen, ist zu revidieren. Mindestens im Russischen ist es so, daß es eher eine Zweiteilung gibt: zu ähnliche Cluster werden vermieden, es kombinieren sich bevorzugt solche Konsonanten miteinander, die zueinander einen mittleren oder größeren Unterschied aufweisen. Das heißt aber auch: die theoretische Herleitung, warum es viele Cluster mittlerer Distanz geben soll, die ja auf der Enkodierungs- bzw. Dekodierungsleistung von Sprechern und Hörern fußt, ist in dieser einfachen Form möglicherweise als Erklärung nicht für alle Sprachen ausreichend. Man könnte allerdings auch folgenden Schluß ziehen: Die genannte Hypothese gilt weiterhin, und zwar als Tendenz. Dann wiche das Russische aus irgendeinem Grunde hiervon ab, und es müßte nach dem Grund für dieses abweichende Verhalten gesucht werden. Welcher Fall vorliegt, kann nur durch die Untersuchung weiterer Sprachen entschieden werden.

Es sei bei der Würdigung dieser Ergebnisse jedoch noch einmal daran erinnert, daß unsere statistischen Tests die Ergebnisse zweier linguistischer Modelle benutzen und somit von ihnen abhängen. Diese beiden Modelle sind die analytische Beschreibung der Phoneme hinsichtlich ihrer phonologischen Merkmale und die Messung der Phonemdistanz. Bei einer anderen komponentalen Phonemanalyse und einem anderen Distanzbegriff käme man möglicherweise zu anderen Ergebnissen. Die oben gemachten Aussagen über Abhängigkeit oder Unabhängigkeit der Konsonantencluster von ihrer internen Struktur dürfen also nicht von ihren Voraussetzungen losgelöst und verabsolutiert werden.

## LITERATUR

Altmann, G.:

1969 Differences between phonemes. *Phonetica* 19, 118-132.

1973 Probabilistische Klassifikation von Konsonantenverbindungen des Indonesischen. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 123, 98–116.

Altmann, G., Lehfeldt, W.:

1973 *Allgemeine Sprachtypologie. Prinzipien und Meßverfahren*. München: W. Fink.

1980 *Einführung in die Quantitative Phonologie* (Quantitative Linguistics 7). Bochum: Brockmeyer.

Avanesov, R.I.:

1972 *Russkoe literaturnoe proiznošenie. Izd. 5-oe, pererab. i dop.* Moskva.

Barinova, G.A., Ševoroškin, V.V.:

1972 Reguljarnye i nereguljarnye sočetanija soglasnych v russkom jazyke. In: *Problemy strukturnoj lingvistiki 1971*, Moskva, 334–341.

Beloozerov, V.N.:

1964 Formal'noe opredelenie fonemy. *Voprosy jazykoznanija* 5, 54–60.

Birnbaum, H.:

1967 Syntagmatische und paradigmatische Phonologie. In: J. Hamm (Hg.), *Phonologie der Gegenwart*, Graz–Wien–Köln, 307–352.

Carroll, J.B.:

1958 The Assessment of Phoneme Cluster Frequencies. *Language* 34, 267–278.

Frumkina, R.M., Vasilevič, A.P.:

1968 Izučenie "proiznositel'noj trudnosti" russkich trechbukvennych sočetanij metodom škalirovanija. In: *Materialy vtorogo simpoziuma po psiholingvistike*, Moskva.

Grotjahn, R.:

1980a Einige kritische Bemerkungen zu W. Lehfeldt "Zur numerischen Erfassung der Schwierigkeit des Sprechbewegungsablaufes". In: W. Matthäus (ed.), *Glottometrika* 3, Bochum, 34-35.

1980b Zur Quantifizierung der Schwierigkeit des Sprechbewegungsablaufes. In: R. Grotjahn, E. Hopkins (eds.), *Empirical Research on Language Teaching and Language Acquisition*, Bochum, 199-231.

Keller, K.C., Saporta, S.:

1957 The Frequency of Consonant Clusters in Chontal. *IJAL* XXIII, 28–35.

Kučera, H., Monroe, G.K.:

1968 *A Comparative Quantitative Phonology of Russian, Czech, and German*. New York.

Ladefoged, P.:

1970 The Measurement of Phonetic Similarity. *Statistical Method in Linguistics* 6, 23–32.

Lehfeldt, W.:

1975 Die Verteilung der Phonemzahl in den natürlichen Sprachen. *Phonetica* 31, 274–287.

1978 Zur Messung der phonetischen Lautdifferenz. Eine begriffskritische Untersuchung. In: G. Altmann (ed.), *Glottometrika* 1, Bochum, 26–45.

1980 Zur numerischen Erfassung der Schwierigkeit des Sprechbewegungsablaufes. In: R. Grotjahn (ed.), *Glottometrika* 2, Bochum, 44–61.

Lekomceva, M.I.:

- 1968 *Tipologija struktur sloga v slavjanskih jazykach*. Moskva: Nauka.
- Levin, Ju.I.:  
 1964 Ob opisanií sistemy lingvističeskich ob'ektov, obladajuščich obščimi svojstvami. *Voprosy jazykoznanija* 4, 112–119.
- Muljačić, Ž.:  
 1967 La combinabilité des phonèmes sur l'axe syntagmatique dépend-elle de leurs traits distinctifs? In: J. Hamm (ed.), *Phonologie der Gegenwart*, Graz usw., 273–286.
- Obratnyj slovar'  
 1974 *Obratnyj slovar' russkogo jazyka*. Moskva: Sovetskaja Ėnciklopedija.
- Oliverius, Zd.F.:  
 1974 *Vvedenije [sic] v fonetiku russkogo jazyka. Lekcii i upražnenija*. Praha.
- Orfografičeskij slovar'  
 1967 *Orfografičeskij slovar' russkogo jazyka*. Pod red. S.G. Barchudarova, S.I. Ožegova i A.B. Šapiro. Izd. 7-oe. Moskva.
- Panzer, H.:  
 1975 *Strukturen des Russischen. Eine Einführung in die Methoden und Ergebnisse der deskriptiven Grammatik*. München.
- Pilch, H.:  
 1974 *Phonemtheorie*. 1. Teil. 3., vollständig neu bearb. Auflage. Basel usw.: S. Karger.
- Saporta, S.:  
 1955 Frequency of Consonant Clusters. *Language* 31, 25–30.
- Tolstaja, S.M.:  
 1968 Fonologičeskoe rasstožanie i sočetaemost' soglasnych v slavjanskih jazykach. *Voprosy jazykoznanija* 3, 66–81.  
 1974 K charakteristike konsonantnyh sočetanij v slavjanskih jazykach (načalo i konec slova). *Slavia* 43, 113–133.
- Toporov, V.N.:  
 1966 Materialy dlja distribucii grafem v pis'mennoj forme russkogo jazyka. In: *Strukturalnaja tipologija jazykov*, Moskva, 65–143.

### Summary

The article investigates the relationship between the phonemic distance of Russian consonant clusters and their existence resp. frequency at the beginning and at the end of words in the lexicon. Using statistical tests, it is shown that the existence of cluster types does not depend in a significant way on the phonemic distance between the consonants involved, whereas there is a clear dependency when cluster frequency is taken into account. The results obtained here for Russian support the findings published by CARROLL for English. In Russian, though, consonant clusters tend to prefer greater distances while avoiding small distances. This makes it necessary to review once more the general assumptions proposed first by SAPORTA. It is also shown that previous treatments of the dependency in question (by MULJAČIĆ and TOLSTAJA) suffer from methodological problems which make their findings practically useless.

	b	b'	v	v'	g	g'	d	d'	ž	z	z'	j	k	k'	l	l'	m	m'	n	n'	p	p'	r	r'	s	s'	t	t'	f	f'	x	c	č	š	
b	V	V	A	V	V	V	V	V	A	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
b'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
v	V	V	A	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
v'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
g	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
g'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
d	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
d'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
ž	V	V	A	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
z	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
z'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
j	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
k	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
k'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
l	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
l'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
m	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
m'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
n	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
n'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
p	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
p'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
r	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
r'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
s	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
s'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
t	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
t'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
f	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
f'	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
x	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
c	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
č	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
š	V	V	P	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V

Tab. 7: Status zweigliedriger Anlautgruppen (systemisch)

