

Phonemhäufigkeiten in slawischen Sprachen: Empirisches Verhalten und Modellierung

Emmerich Kelih

0. Einleitung

Die Häufigkeiten von Phonemen in slawischen Sprachen sind mehrfach Gegenstand linguistischer Abhandlungen, in denen unterschiedlichste Aspekte untersucht werden. Häufigkeiten von Subklassen von Phonemen (wie z.B. Vokale, Konsonanten usw.) werden als sprachtypologische Eigenschaft interpretiert. Vgl. dazu die älteren Arbeiten von Budilovič (1883), Dikarev (1891/1892), die späteren strukturalistischen Arbeiten wie z.B. Burlakova et al. (1962), Segal (1972), bis hin zu erst unlängst erschienenen Arbeiten von Majewicz (1989), Sawicka/Holvoet (1991) und Sawicka (2008).

In den angeführten Arbeiten stehen vor allem deskriptive Aspekte im Vordergrund und im Grunde ist man bislang nicht darüber hinausgekommen. In den letzten Jahrzehnten setzte sich innerhalb der quantitativen Linguistik (vgl. Altmann 1972, 1973, 1978, 1997) und insbesondere in der synergetischen Linguistik (vgl. Köhler 1986, 2005) die Erkenntnis durch, dass es nicht nur um die deskriptive Erfassung von quantitativ erfassbaren Sprachmerkmalen gehen sollte, sondern vielmehr um die Frage von Interrelationen und Wechselbeziehungen zwischen Sprachebenen und Spracheigenschaften.

Insbesondere stellte sich in den letzten Jahren heraus, dass die quantitative Ausprägung des Phoneminventars einen entscheidenden Einfluss auf die Silben- und Wortstruktur (vgl. Fenk-Oczlon/Fenk 2008, Kelih 2015, Köhler 1986) und die Anzahl von Phonemverbindungen (vgl. Kelih 2009a) hat. Ähnliche Tendenzen lassen sich bei Graphemsystemen finden. In einer Reihe von Untersuchungen zu Graphemhäufigkeiten in slawischen Sprachen konnte gezeigt werden, dass der Inventarumfang einen direkten Einfluss auf die Form der Ranghäufigkeitsverteilung von Graphemen hat (vgl. u.a. Grzybek/Kelih/Stadlober 2006, Grzybek/Kelih 2004). Damit erweist sich, dass eine paradigmatische Eigenschaft (wie eben der Inventarumfang eines Phonem- bzw. Graphemsystems) einen empirisch nachweisbaren Einfluss auf syntagmatische Eigenschaften wie die Häufigkeit von Phonemen bzw. Graphemen, die Anzahl von Phonemverbindungen, die Silbenstruktur usw. hat.

Phonemhäufigkeiten in slawischen Sprachen: Empirisches Verhalten und Modellierung

Emmerich Kelih

0. Einleitung

Die Häufigkeiten von Phonemen in slawischen Sprachen sind mehrfach Gegenstand linguistischer Abhandlungen, in denen unterschiedlichste Aspekte untersucht werden. Häufigkeiten von Subklassen von Phonemen (wie z.B. Vokale, Konsonanten usw.) werden als sprachtypologische Eigenschaft interpretiert. Vgl. dazu die älteren Arbeiten von Budilovič (1883), Dikarev (1891/1892), die späteren strukturalistischen Arbeiten wie z.B. Burlakova et al. (1962), Segal (1972), bis hin zu erst unlängst erschienenen Arbeiten von Majewicz (1989), Sawicka/Holvoet (1991) und Sawicka (2008).

In den angeführten Arbeiten stehen vor allem deskriptive Aspekte im Vordergrund und im Grunde ist man bislang nicht darüber hinausgekommen. In den letzten Jahrzehnten setzte sich innerhalb der quantitativen Linguistik (vgl. Altmann 1972, 1973, 1978, 1997) und insbesondere in der synergetischen Linguistik (vgl. Köhler 1986, 2005) die Erkenntnis durch, dass es nicht nur um die deskriptive Erfassung von quantitativ erfassbaren Sprachmerkmalen gehen sollte, sondern vielmehr um die Frage von Interrelationen und Wechselbeziehungen zwischen Sprachebenen und Spracheigenschaften.

Insbesondere stellte sich in den letzten Jahren heraus, dass die quantitative Ausprägung des Phoneminventars einen entscheidenden Einfluss auf die Silben- und Wortstruktur (vgl. Fenk-Oczlon/Fenk 2008, Kelih 2015, Köhler 1986) und die Anzahl von Phonemverbindungen (vgl. Kelih 2009a) hat. Ähnliche Tendenzen lassen sich bei Graphemsystemen finden. In einer Reihe von Untersuchungen zu Graphemhäufigkeiten in slawischen Sprachen konnte gezeigt werden, dass der Inventarumfang einen direkten Einfluss auf die Form der Ranghäufigkeitsverteilung von Graphemen hat (vgl. u.a. Grzybek/Kelih/Stadlober 2006, Grzybek/Kelih 2004). Damit erweist sich, dass eine paradigmatische Eigenschaft (wie eben der Inventarumfang eines Phonem- bzw. Graphemsystems) einen empirisch nachweisbaren Einfluss auf syntagmatische Eigenschaften wie die Häufigkeit von Phonemen bzw. Graphemen, die Anzahl von Phonemverbindungen, die Silbenstruktur usw. hat.

Vor dem Hintergrund einer sich abzeichnenden Wechselbeziehung zwischen dem Phoneminventar, der Häufigkeit von Phonemen und Phonemverbindungen, der Silben- und der Wortstruktur, ist die vorliegende Arbeit ein kleiner Baustein zur Analyse slawischer Phonemsysteme und deren Interrelationen zu anderen sprachlichen Ebenen. Folgende Punkte hinsichtlich der Phonemhäufigkeiten in slawischen Sprachen werden diskutiert:

(1) die Häufigkeit von Vokalen und Konsonanten im Text und deren Relation zum Anteil von Vokalen und Konsonanten im Phoneminventar als paradigmatische Spracheigenschaft.

(2) eine linguistische Interpretation von an unterschiedlicher Stelle publizierten Daten zu Phonemhäufigkeiten in zehn slawischen Standardsprachen.

(3) ein in der quantitativen Linguistik intensiv diskutiertes theoretisches Modell (stetige Funktion) wird empirisch getestet, welches sich bislang u.a. für slawische Graphemhäufigkeiten als geeignet erwiesen hat. Nunmehr soll überprüft werden, ob dieses Modell, konzipiert als ein sogenanntes Schichten-Modell, auch für Phonemhäufigkeiten geeignet ist.

1. Vorstellung des Datenmaterials

Quantitative Eigenschaften von Phonemsystemen slawischer Sprachen stehen immer wieder im Fokus von sprachwissenschaftlichen Untersuchungen. Zumindest sind für eine Reihe von slawischen Standardsprachen Daten zu Phonemhäufigkeiten (Slowenisch, Serbisch, Makedonisch, Bulgarisch, Russisch, Weißrussisch, Polnisch, Tschechisch, Slowakisch, Niedersorbisch) publiziert, die im Folgenden als Datenbasis für die oben angeführten Fragestellungen herangezogen werden. Dieses Vorgehen, d.h. das Verwenden von Daten aus der Sekundärliteratur, muss zumindest kurz kommentiert werden: Die Gewinnung von Phonemhäufigkeiten setzt eine stringente Definition der zu untersuchenden Einheiten voraus. In Anbetracht der Tatsache, dass bislang – unseres Wissens nach – keine systematischen und vergleichenden Phonemstudien zu slawischen Sprachen vorliegen, ist ein gangbarer Weg darin zu sehen, bestehende und publizierte Phonemdaten einer Re-Analyse zu unterziehen. Dieses Vorgehen impliziert aber eine ganze Reihe von Problemen, die es zu beachten gilt.

In erster Linie setzt man sich der Gefahr aus, dass das Phonem als Einheit nicht einheitlich definiert wird (unterschiedlicher Fokus auf phonetische, allophonische, morphologische und morphologische Kriterien bei der Bestimmung des Phoneminventars). Somit ist bereits a priori eine uneingeschränkte Vergleichbarkeit der Untersuchungen nicht gewährleistet. Zu denken ist auch an die unterschiedlichen Möglichkeiten der Interpretation segmentaler und su-

prasegmentaler Eigenschaften, die einen nachhaltigen Einfluss auf den Umfang eines Phoneminventars und damit auf die Häufigkeitsverteilung haben. Man denke hier z.B. an die komplexe Interrelation von Vokalquantitäten, Öffnungsgrad und Betonung im Slowenischen bzw. Serbischen und Ähnliches.

Beim Problem einer fruchtbaren und operational anwendbaren Phonemdefinition spielt auch die Frage, in welchem sprachlichen Material Phonemhäufigkeiten bestimmt werden, eine sehr große Rolle. Man erinnere sich an dieser Stelle an den bekannten Einwand von N.S. Trubetzkoy, der sich bereits 1939 kritisch mit Phonemzählungen auseinandergesetzt und dabei auf die Notwendigkeit einer Untersuchung sowohl im Text als auch im Wörterbuch verwiesen hat (vgl. Trubetzkoy 1939: 231ff). Darüber hinaus ist zu beachten, dass sich die Häufigkeit von Phonemen in unterschiedlichen Textsorten bzw. Funktionalstilen beträchtlich unterscheiden kann.¹

In Tab. 1 sind für zehn slawische Standardsprachen die jeweiligen Bezugsquellen der Daten und der in diesen Arbeiten verwendete Inventarumfang von Phonemen angeführt. Ergänzt werden diese Angaben durch eine stichwortartige Präzisierung der ausgewerteten „Textbasis“, d.h. es wird spezifiziert, in welchen Textsorten die Häufigkeiten bestimmt wurden. Hervorhebenswert – da insbesondere bei der statistischen Modellierung von großer Bedeutung – ist die Frage, ob es sich dabei um Stichproben (z.B. Textfragmente) handelt, oder ob jeweils abgeschlossene Texte verwendet werden. Die aus der Sekundärliteratur übernommenen absoluten Phonemhäufigkeiten² sind im Anhang 1, 2 und 3 tabelliert. Alle Datensätze beziehen sich auf die jeweilige Standardsprache, außer im Fall des Niedersorbischen, für welches nur Daten aus Dialektuntersuchungen eruiert werden konnten. Insofern ist ein Vergleich des Niedersorbischen mit den anderen slawischen Sprachen aufgrund der Datenlage nur bedingt möglich. Die niedersorbischen Daten werden der Vollständigkeit halber dennoch analysiert und interpretiert.

¹ Aus diesen Gründen ist die hier vorgelegte Untersuchung von „fremden“ Daten und deren Interpretation als vorläufig und tentativ zu verstehen. Die vorliegende Untersuchung ist zumindest als ein erster Schritt in Richtung einer systematischen Untersuchung von slawischen Phonemhäufigkeiten zu begreifen.

² Um die Datenbasis nicht unnötigerweise zu vergrößern, wurde pro Sprache nur ein einziger Datensatz ausgewählt, obwohl für viele slawische Sprachen wie beispielsweise Polnisch, Tschechisch, Russisch u.a. durchaus mehrere Datensätze zur Verfügung stehen. Für das Ukrainische konnten keine passenden Daten gefunden werden. Die in Perebyjnis (1970) veröffentlichten Daten sind Phonemhäufigkeiten, die jeweils aus bestimmten Wortlängen (Wörter mit 1, 2, 3 ... x Morphemen) gewonnen wurden. Damit sind sie nicht mit den in dieser Arbeit verwendeten Daten vergleichbar.

Tab. 1: Analyierte slawische Sprachen: Phoneminventarumfang, Textbasis und Quelle

Sprache	Anzahl von Phonemen	Textbasis	Quelle
Slowenisch	29	literarischer Text	Kolter (1994)
Serbisch (ekavisch)	30	Prosa-Texte, journalistische Texte	Ković (1982)
Makedonisch	32	Prosa-Text (Stichprobe)	Džukeski (1955)
Bulgarisch	42	Prosa-Texte	Marinova/Marinov (1964)
Russisch	38	unterschiedliche Funktionalstile	Kučera/Monroe (1968)
Weißrussisch	43	nicht angeführt	Podlužnyj (1966)
Polnisch	39	mündliche Rede (Hörspiele, Radio-diskussionen, Dialoge)	Jassem/Łobacz (1976) ³
Tschechisch	38	unterschiedliche Funktionalstile	Ludviková/Königová (1967)
Slowakisch	46	Volksmärchen	Bosák (1965)
Niedersorbisch (Dialekt) ⁴	33	mündliche Dialoge	Faßke (1964)

³ Diese Untersuchung geht von einer explizit allophonischen Transkription aus und berücksichtigt die Palatalisierungskorrelation im Polnischen, die ebenfalls in der polnischen Sprachwissenschaft umstritten ist, nicht.

⁴ Die untersuchte Vetschauer Mundart verfügt laut der Analyse von Faßke (1964: 25 und 39-40) über 7 Vokalphoneme und 29 Konsonantenphoneme. Damit ist von einem Inventar von 36 Phonemen auszugehen. In den von Faßke (1964) durchgeführten Zählungen sind drei Phoneme (/ś, ć, ź/) nicht realisiert, sodass für die vorliegende Untersuchung von einem Inventar von 33 Phonemen ausgegangen wird.

Gemeinsam ist allen Untersuchungen, dass pro Sprache von einem Phoneminventar ausgegangen wird, welches in der Literatur mehr oder weniger unbestritten ist. Darüber hinaus sind auch keine Daten, die auf Zählungen aus Wörterbüchern beruhen, aufgenommen worden.⁵ Die Stichprobengröße, d.h. die absolute Gesamtanzahl von Phonemen in Texten findet sich im Anhang 1, 2 und 3. Diese variiert dabei beträchtlich (zwischen 1000 Phonemen und über 1000000 Phonemen), wobei die komplexe Frage eines adäquaten Stichprobenumfangs an dieser Stelle nicht thematisiert werden kann (vgl. dazu u.a. Grzybek/Kelih 2003, Kubáček 1994).

2.1. Vokale und Konsonanten: Anzahl im Phoneminventar vs. Häufigkeit im Text

Die bekannteste Untersuchung des Anteils von Vokalen und Konsonanten in einem Phonemsystem geht auf Isačenko (1939/1940) zurück. Dieser hatte versucht, aufgrund dieses paradigmatischen Merkmals innerhalb der slawischen Sprachen zwischen vokalischen und konsonantischen Sprachen zu unterscheiden. Trotz einer Reihe von methodologischen Problemen dieser Studie (vgl. Kempgen 1991, Kelih 2010a), die darin bestehen, dass Isačenko (1939/1940) segmentale und suprasegmentale Eigenschaften bei der Bestimmung des Phoneminventars nicht trennt und die gezogene Grenze zwischen vokalischen und konsonantischen Sprachen nicht begründet, kann die jeweilige Anzahl von Vokalen und Konsonanten in einem Phonemsystem ohne Zweifel als ein sprachtypologisches Merkmal slawischer Sprachen verstanden werden. Als eine Begründung für die beobachtbaren Unterschiede bezüglich der Anzahl von Vokalen und Konsonanten in slawischen Sprachen gilt das unterschiedliche Ausmaß der Palatalisierungskorrelation und der suprasegmentalen Eigenschaften. Es wird in diesem Zusammenhang explizit von einer gegenseitigen Kompensation gesprochen (vgl. Andersen 1978; Trummer 1993), d.h. die Sprachen bauten entweder den konsonantischen Bereich aus (z.B. das Russische, das Ukrainische, das Weißrussische und das Polnische mit einer relativ hohen Anzahl von Konsonanten) oder aber der vokalische Bereich zeichnet sich durch zusätzliche suprasegmentale Eigenschaften aus (Slowenisch, Serbisch, Tschechisch, Slowakisch). Offen bleibt in dieser Diskussion, ob die Anzahl von Vokalen und Konsonanten in einem Phonemsystem in irgendeiner Weise mit der syntagmatisch realisierten Häufigkeit von Phonemen in einem

⁵ Auf die potentiellen Unterschiede von Phonemhäufigkeiten, die aus Texten gewonnen werden und denjenigen aus Wörterbüchern hat bereits Trubetzkoy (1939: 232) verwiesen. Linguistisch erklärbar ist dies durch die hohe Frequenz von Synsemantika in Texten, die in der Regel kürzer sind als Autosemantika.

Zusammenhang steht. Um diese potentielle Interrelation näher zu untersuchen, wird für die in Tab. 1 angeführten slawischen Sprachen (1) die Anzahl von vokalischen und konsonantischen Phonemen bestimmt und (2) die jeweilige relative Häufigkeit von Vokalen und Konsonanten berechnet.

Die Bestimmung der als vokalisch bzw. konsonantisch geltenden Phoneme bedingt eine Reihe von linguistischen Entscheidungen, die allerdings im vorliegenden Fall bereits durch die zur Verfügung stehenden Datensätze zum größten Teil vorgegeben sind: Suprasegmentale Eigenschaften von Vokalen im Serbischen⁶ können nicht berücksichtigt werden. Analytierte Diphthonge (vor allem im Tschechischen und Slowakischen) werden monophonematisch, d.h. als ein einziges Phonem interpretiert, wenn auch eine linguistisch begründbare biphonematische Interpretation möglich wäre (und stellenweise auch nötig ist, wie z.B. in Fremdwörtern).

Hinsichtlich der silbenbildenden Phoneme /l/ und /r/ konnte kein einheitliches Vorgehen gewählt werden, da diese in einigen Fällen bei den Zählungen nicht gesondert berücksichtigt wurden. Es stehen für das Makedonische, das Tschechische und das Slowakische die Häufigkeiten der silbenbildenden Konsonanten zur Verfügung, während dies für das Serbische und Slowenische nicht der Fall ist.

Ein weiteres Problem ist, dass z.B. innerhalb der ostslawischen Sprachen vor allem der phonemische Status von /i/ vs. /y/ umstritten ist bzw. diese Phoneme, je nach linguistischer Theorie, unterschiedlich gehandhabt werden: Für das Russische wird von einem Vokalsystem von fünf – ohne Unterscheidung von /i/ und /y/ Phonemen ausgegangen (vgl. Tab. 2 unten), während für das Weißrussische Podlužnyj (1966) diese Phoneme als zwei phonologisch distinktive Vokale interpretiert und demnach für das Weißrussische ein Inventar von sechs Vokalphonemen ansetzt. Weitere Inkonsistenzen betreffen auch die phonologische Interpretation der Nasalität und Palatalisierungskorrelation im Polnischen und die Interpretation von Quantitätsunterschieden und von Öffnungsgraden im Slowenischen. Die Affrikaten werden in der Regel in allen Untersuchungen als monophonematisch interpretiert. Die Kategorien von Halbvokalen/Glides bzw. Liquida werden in den Daten nicht gesondert berücksichtigt, sie werden durchgehend als konsonantisch interpretiert.

Bei den vorliegenden Daten wird von einem empirisch bestimmbar, allophonisch interpretierten Phonembegriff ausgegangen. Die Resultate der Berechnungen zum Anteil von Vokalen (Vsys.) und Anteil von Konsonanten

⁶ Für das Serbische bzw. Kroatische sind dem Verfasser keine systematischen Studien zur Phonemhäufigkeit bekannt, in denen bei den Zählungen die komplexen Vokalverhältnisse hinsichtlich Länge, Akzent und Tonhöhe berücksichtigt wurden.

(Csys.) am Gesamtinventar und die relative Häufigkeit von Vokalen (Vfreq.) und relative Häufigkeiten von Konsonanten (Cfreq.) sind in Tab. 2 zusammengefasst.

Tab. 2: Vokale und Konsonanten: Anzahl im Phoneminventar vs. Häufigkeit

Sprache	Inventar	V	C	Vsys	Csys	Vfreq	Cfreq
Slowenisch	29	8	21	0,27	0,72	0,41	0,58
Serbisch (ekavisch)	30	5	25	0,16	0,83	0,44	0,55
Makedonisch	32	6	26	0,18	0,81	0,45	0,54
Bulgarisch	42	6	36	0,14	0,85	0,43	0,56
Weißrussisch	43	6	37	0,13	0,86	0,40	0,59
Russisch	38	5	33	0,13	0,86	0,41	0,58
Polnisch	39	8	31	0,20	0,79	0,40	0,59
Tschechisch	32	7	25	0,21	0,78	0,41	0,58
Slowakisch	39	12	27	0,30	0,69	0,43	0,56
Niedersorbisch	33	7	26	0,21	0,78	0,43	0,56

Der jeweilige Anteil von Vokalen (Vsys.) und Konsonanten (Csys.) in den einzelnen Phonemsystemen kann folgendermaßen knapp umrissen werden. Als „Extrempunkte“ treten das Slowakische (Vsys = 0,30) und das Slowenische (0,27) auf, die aufgrund ihres hohen paradigmatischen Anteils an Vokalen am Gesamtinventar durchaus als „vokalische“ Sprachen bezeichnet werden können. Den anderen Extrempunkt stellen die ostslawischen Sprachen (Russisch, Weißrussisch) dar, die in Bezug auf alle anderen slawischen Sprachen den geringsten Vokalanteil⁷ im Phonemsystem, d.h. die geringste Anzahl von Vokalphonemen in Bezug auf das Gesamtinventar, haben.⁸

Aufgrund der linguistisch umstrittenen und methodologisch bislang nicht zufriedenstellend aufgearbeiteten Frage einer intersubjektiv argumentierbaren Trenngrenze (vgl. dazu auch Krámský 1946/1948: 43; 1959) zwischen vokalischen und konsonantischen Sprachen muss man sich einstweilen mit diesem einfachen Befund abfinden. Allgemein kann gesagt werden, dass auf paradigmatischer Ebene alle slawischen Sprachen als „konsonantische“ Sprachen be-

⁷ Dieser Unterschied wird bezogen auf die slawischen Sprachen dahingehend interpretiert, dass mit dem Ausbau der Palatalitätskorrelation, die, so gemeinhin angenommen, erst nach dem Jer-Wandel einsetzte, eben das Konsonanteninventar erhöht wurde.

⁸ Die angeführten Verhältnisse würden sich selbst unter Berücksichtigung von suprasegmentalen Eigenschaften nicht grundlegend ändern.

zeichnet werden können, da in keinem einzigen Fall der Vokalanteil höher ist als der Anteil an Konsonanten.

Damit kann zu einer weiteren – und bislang nicht untersuchten – Fragestellung übergeleitet werden. Es geht darum, zu untersuchen, ob und in welcher Wechselbeziehung der Vokal-Konsonantenanteil im Phonemsystem mit der Häufigkeit von Vokalen und Konsonanten im Text steht. Dieses Verhältnis ist eine wichtige systemlinguistische Eigenschaft,⁹ die Auskunft über den jeweiligen Ausnutzungsgrad einzelner Phoneme im Text gibt.

Zu beginnen ist mit einer Vorstellung der berechneten Häufigkeit von Vokalen und Konsonanten. Interpretiert man die in Tab. 2 berechneten relativen Vokal- und Konsonantenhäufigkeiten, so fällt Folgendes auf: Die höchste Vokalhäufigkeit hat das Makedonische (0,4596), während das Polnische – zumindest in dem von uns analysierten Datensatz – die geringste Frequenz an Vokalen (0,4069) aufweist. Weitere Sprachen mit einem relativ hohen Vokalanteil sind das Serbische, das Bulgarische, das Niedersorbische und das Slowakische, die allesamt eine Vokalhäufigkeit größer als 0,43 haben. Hinsichtlich der ostslawischen Sprachen – diese haben auf paradigmatischer Ebene den niedrigsten Vokalanteil – ergibt sich für das Russische ein Vokalanteil in der Höhe von $\approx 0,41$, während das Weißrussische ($V_{\text{freq.}} = 0,4085$) knapp darunter liegt. Dieser Unterschied zwischen Russisch und Weißrussisch ist statistisch (unter Anwendung eines asymptotischen Tests) signifikant ($z > 1,96$).

Es zeigt sich insgesamt, dass man in Bezug auf die slawischen Sprachen offensichtlich kaum von irgendwelchen „konstanten Häufigkeitsverhältnissen“ ausgehen kann. Auch gibt es zwischen dem Anteil von Vokalen/Konsonanten im Phonemsystem und ihrer Häufigkeit im Text offensichtlich keine Zusammenhänge im Sinne einer statistischen Korrelation. D.h. ein hohes Inventar an Vokalen zieht nicht automatisch eine hohe bzw. niedrige Vokalhäufigkeit nach sich. In Abb. 1 ist diese Wechselbeziehung für alle slawischen Sprachen graphisch dargestellt.

⁹ Weitgehend nicht untersucht ist die Frage, welche Faktoren die Vokal- bzw. Konsonantenhäufigkeit innerhalb einer Sprache beeinflussen. In Kelih (2010b) wird gezeigt, dass selbst innerhalb einer Sprache mit nicht unerheblichen Schwankungen der Vokal- und Konsonantenhäufigkeit zu rechnen ist, da die Häufigkeit von Vokalen in indirekter Relation (aufgrund eines Zusammenhangs zwischen der Wortlänge, der Silbenlänge und der Vokalhäufigkeit) zur verwendeten Textsorte bzw. zum Funktionalstil steht. Diese Einflussfaktoren sind bei einer sprachtypologischen Interpretation von Vokal- und Konsonantenhäufigkeiten zu berücksichtigen.

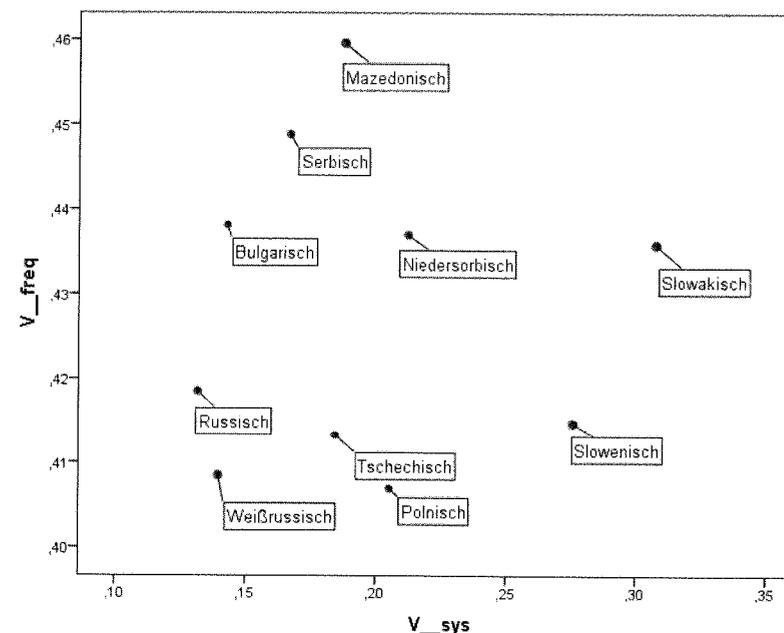


Abb. 1: Vokalanteil im System vs. Vokalhäufigkeit im Text

Deutlich zu sehen ist aus Abb. 1, dass bestimmte Sprachen wie das Russische, das Weißrussische, das Bulgarische, das Serbische und das Makedonische jeweils abgegrenzte Bereiche einnehmen. Dies gilt auch für das Slowenische, das Slowakische und das Tschechische, das Polnische und das Niedersorbische, die in der Tat „Subgruppen“ zu bilden scheinen.

Auch wenn die Frage einer Gruppenbildung methodologisch durch elaborierte statistische Verfahren abgesichert werden müsste (multivariate Diskriminanzanalysen, post-hoc Analysen und ähnliches), kann nunmehr zumindest für die slawischen Sprachen festgestellt werden, dass Eigenschaften von phonologischen Systemen und der Ausnutzungsgrad von Phonemen in Texten in keinerlei statistischer Wechselbeziehung stehen. Mit diesem Befund können beide Merkmale als autonome Merkmale für sprachtypologische Fragestellungen und Interpretationen herangezogen werden. Wichtiges gemeinsames Merkmal der slawischen Sprachen ist, dass eine im Grunde genommen niedrige Anzahl von Vokalen in einem Phoneminventar in einem Text überdurchschnittlich häufig gebraucht wird (die relative Frequenz ist in allen sla-

wischen Sprachen $> 0,40$). Linguistisch ist dies vermutlich dadurch zu erklären, dass die Vokale die zentralen Bausteine von Silben darstellen und somit eine hohe Funktionalität bei der Bildung von Morphemen und Morphemketten (Wortformen, Lexeme usw.) haben. Darüber hinaus gibt es Grund zur Annahme, dass Vokale über eine „grammatische“ Funktion verfügen, zumal diese – und dies wird erst in Zukunft zu überprüfen sein – bei Affixen und in der Regel kurzen Synsemantika überdurchschnittlich häufig auftreten.

2.2. Häufigkeit von Vokalen im Text

Nachdem nun zwei Merkmale, der Vokal-Konsonantenanteil im Phonemsystem und die Vokal-Konsonantenhäufigkeit im Text in den slawischen Sprachen, näher besprochen wurden, ist im Folgenden auf die Häufigkeit der einzelnen Phoneme einzugehen. Um einen detaillierten Einblick in das Häufigkeitsverhalten der Phoneme zu bekommen, wird für jedes Phonem der absolute Rang bestimmt, d.h. das häufigste Phonem innerhalb einer Sprache erhält den Rangplatz 1, das zweithäufigste den Rang 2 usw. Um es in der Sprache der Statistik auszudrücken: die Phonemhäufigkeiten werden auf einer Ordinalskala abgebildet, in der der Rang als Hilfsvariable auftritt. Um eine direkte Vergleichbarkeit der Daten zu gewährleisten, werden in einem nächsten Schritt die absoluten Häufigkeiten in relative Häufigkeiten überführt. Die entsprechenden Rohdaten sind im Anhang 1 zu finden. An dieser Stelle erfolgt eine selektive Besprechung von einigen auffälligen Tendenzen hinsichtlich der Häufigkeit von Vokalen:

1. In allen untersuchten slawischen Sprachen besetzen Vokale die ersten beiden Rangplätze, d.h. die Vokale kommen am häufigsten vor. Es ist dies aber nicht immer der gleiche Vokal. In fast allen Sprachen besetzt das Phonem /a/ Platz eins oder Platz zwei, nur im Tschechischen tritt der Vokal /e/ als häufigstes Phonem (= Rangplatz 1) auf. Der 3. Rangplatz ist – außer im Russischen, das an dieser Stelle bereits den Konsonanten /t/ aufweist – ebenfalls mit vokalischen Phonemen besetzt.

2. Bereits ab dem 4. Rangplatz lassen sich hinsichtlich der Vokalhäufigkeit kaum noch allgemeine Tendenzen feststellen: Während für das Slowenische, das Serbische, das Makedonische, das Bulgarische, das Weißrussische, das Tschechische und das Slowakische in dieser Position ebenfalls noch vokalische Phoneme auftreten, überlassen alle anderen Sprachen diesen Rang bereits Konsonanten. Weitere Details dazu finden sich in Kapitel 2.3.

3. Der vordere Rangbereich wird in der Regel von Vokalen besetzt. Bestimmte Sprachen verfügen aber auch über Vokale mit einer sehr niedrigen Häufigkeit, die entsprechend die hinteren Rangplätze belegen. Dies gilt beispielsweise für den Vokal /u/, der im Slowenischen den Rang 23 und im Bulgarischen den Rang 17 einnimmt. Einzig und allein im Serbischen nimmt das /u/ einen vorderen Rangplatz (8) ein. In jedem Fall kann daraus geschlossen werden, dass man innerhalb der Vokale durchaus zwischen hochfrequenten und weniger häufigen Vokalen zu unterscheiden hat. Unterschiede in den Häufigkeiten von Vokalen zeigen sich auch, sofern man ausgewählte „markierte“ (= niedrigfrequente) Phoneme näher betrachtet. So hat beispielsweise das offene /e:/ im Slowenischen eine sehr geringe relative Häufigkeit von 0,02594, im Polnischen hat das nasale /ɛ/ eine verschwindend geringe Frequenz von 0,001 und im Tschechischen hat beispielsweise der Diphthong /au/ eine Frequenz von nur 0,00065781. Dies bedeutet, dass man aus quantitativer Sicht durchaus von sogenannten peripheren Phonemen sprechen kann: Es sind dies Vokale, die ein konstitutiver Bestandteil des Phonemsystems sind, aber deren Funktionalität im Sinne der Gebrauchshäufigkeit sehr niedrig ist.

Zusammengefasst zeigt sich, dass die Häufigkeit von Vokalen bei einem durchschnittlichen Gesamtanteil von ca. 40% in allen slawischen Sprachen deutlich geringer ist als die Konsonantenhäufigkeit (vgl. die entsprechenden Häufigkeiten in Tab. 2). Damit erklärt sich auch die insgesamt niedrigere Häufigkeit von Vokalen in Relation zu Konsonanten.

Dieses Häufigkeitsverhalten deutet darauf hin, dass die Funktionalität der Vokalphoneme unterschiedlich verteilt ist und offensichtlich die Tendenz dazu besteht, einzelne Vokale überdurchschnittlich zu nützen. Dies führt zu dem systemlinguistisch interessanten Befund, dass für ein geschlossenes System von Phonemen (in diesem Fall die vokalischen Phoneme) unterschiedliche Häufigkeitsgrade und damit unterschiedliche Redundanzgrade vonnöten sind. Da dieser Befund eine entscheidende Rolle bei der Diskussion um ein adäquates statistisches Modell spielen wird (vgl. Kap. 2.6), sei dieses auffällige Verhalten – die Über- und Unterbelastung von Vokalen – abschließend anhand des Slowenischen graphisch demonstriert. In Abb. 2 sind die relativen Häufigkeiten aller slowenischen Vokale in absteigender Reihenfolge dargestellt.

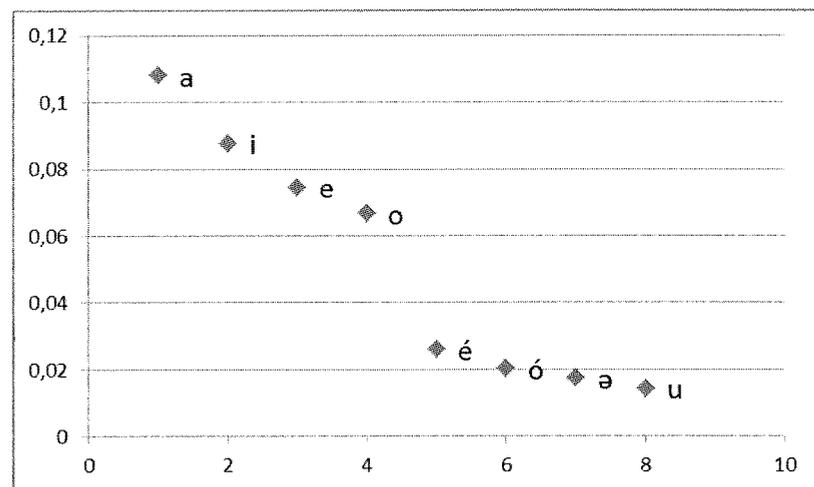


Abb. 2: Vokalhäufigkeiten (rel.) im Slowenischen¹⁰

Abb. 2 zeigt deutlich, dass man es mit einer überdurchschnittlichen Nutzung von /a/, /i/, /e/ und /o/ (dies sind die häufigsten Phoneme im Slowenischen) zu tun hat, während die Vokale /ɛ/, /ɔ/ /u/ und der Halbvokal /ə/ im Vergleich dazu „abgeschlagen“ sind und eine weitaus geringere Frequenz haben. Ein analoges Bild ergibt sich auch für alle anderen slawischen Sprachen, welches aber an dieser Stelle nicht im Detail besprochen werden muss. Zusammengefasst lässt sich festhalten, dass keine Gesamttendenzen hinsichtlich der Häufigkeit von Vokalen in Texten ausgemacht werden können, sondern hinsichtlich des Häufigkeitsverhaltens durchaus eine immanente Heterogenität zu beobachten ist.

2.3. Häufigkeit von Konsonanten-Subgruppen

Hinsichtlich der Häufigkeit von Konsonanten in den analysierten Texten lässt sich auf der Basis der analysierten Daten festhalten, dass diese in jeder slawischen Sprache – und dies lässt sich leicht aus den Daten zur Vokalhäufigkeit ableiten – höher ausfällt (im Durchschnitt ca. 0,60) als die entsprechende Vokalhäufigkeit. Da eine vergleichende Untersuchung der Häufigkeit von Konsonanten keine auffälligen Gemeinsamkeiten ergeben hat, wurde die Häufigkeit in vier Subgruppen von Konsonanten zusammengefasst. Aus phonetisch-phonologischer Sicht wird zwischen Sonanten, Verschlusslauten, Frikativen und

¹⁰ Aus technischen Gründen ist in der Graphik das é als /ɛ/ und ó als /ɔ/ zu lesen.

Affrikaten unterschieden. D.h. eine Zusammenfassung der konsonantischen Phoneme orientiert sich an phonetisch-akustischen Merkmalen. Die Zugehörigkeit der einzelnen Phoneme zu einer der erwähnten Gruppen orientiert sich an den Einzelbeschreibungen der slawischen Sprachen in Rehder (1998) bzw. Comrie/Corbett (1993). Die Grobheit dieser Zusammenfassungen wird bewusst in Kauf genommen, da es an dieser Stelle hauptsächlich darum geht, einige auffällige Tendenzen aufzuzeigen.

Die Ergebnisse dieser Zusammenfassungen sind in Tab. 3 zu finden, in der der jeweilige relative Anteil dieser Subgruppen angeführt ist. Diese Gruppierung ermöglicht es, einen detaillierten Einblick in das Verhalten der Konsonantenhäufigkeiten der slawischen Sprachen zu bekommen.

Die sonoren Laute haben innerhalb der slawischen Sprachen einen sehr hohen durchschnittlichen Anteil. Dieser beträgt ca. 0,22. Dieser Mittelwert soll aber nicht darüber hinwegtäuschen, dass die Bandbreite der Häufigkeit von Sonanten innerhalb der einzelnen Sprachen doch recht hoch ist. Das obere Ende bilden das Slowenische und das Niedersorbische mit einem Anteil von 0,28. Diese Auffälligkeit ist vermutlich durch die hohe Frequenz des als sonantisch interpretierten Phonems /v/ in beiden Sprachen erklärbar.

Tab 3: Häufigkeiten von Konsonanten-Subgruppen

Sprache	Inventar	Sonanten	Verschlusslaute	Frikative	Affrikaten
Slowenisch	29	0,28	0,17	0,10	0,01
Serbisch	30	0,21	0,17	0,13	0,02
Makedonisch	32	0,17	0,21	0,13	0,01
Bulgarisch	42	0,18	0,20	0,15	0,01
Weißrussisch	43	0,23	0,14	0,15	0,05
Russisch	38	0,22	0,18	0,14	0,03
Polnisch	39	0,22	0,16	0,15	0,05
Tschechisch	32	0,21	0,18	0,16	0,02
Slowakisch	39	0,20	0,18	0,15	0,02
Niedersorbisch	33	0,28	0,17	0,09	0,01

Das untere Ende hinsichtlich des Anteils von sonoren Lauten bilden das Makedonische und das Bulgarische mit einem Anteil von 0,17 bzw. 0,18. Diese – im Vergleich zu anderen slawischen Sprachen – geringe Häufigkeit lässt sich dadurch erklären, dass im Bulgarischen und Makedonischen gleichsam kompensatorisch eine andere Gruppe von Konsonanten, nämlich einzelne Verschlusslaute, sehr häufig vorkommen. Konkret ist dies in beiden Sprachen das Phonem /t/. Die relative Häufigkeit von /t/ beträgt im Makedonischen 0,0705,

während im Bulgarischen eine relative Frequenz von 0,0796 festgestellt werden kann. In beiden Fällen gilt, dass das Phonem /t/ in der Ranghäufigkeit von Konsonanten das häufigste konsonantische Phonem ist und somit auch häufiger als Sonanten vorkommt, die üblicherweise in slawischen Sprachen die häufigsten Konsonanten sind. Die hohe Frequenz dieses Phonems lässt sich unter Umständen durch die Verwendung des postpositiven Artikels (-to, -ta usw.) erklären. Demnach zeigt sich, dass die morphologische Struktur einen direkten Einfluss auf die Form der Häufigkeitsverteilung von Phonemen hat. Auch dazu müssen detaillierte Studien noch erfolgen.

Die zweithäufigste Gruppe innerhalb der Konsonanten – außer im Bulgarischen und Makedonischen – bilden die Verschlusslaute. Im Durchschnitt ergibt sich ein Anteil von 0,1760. Auffällig ist in dieser Gruppe nur das Weißrussische, welches innerhalb der slawischen Sprachen den geringsten Anteil (0,1454) an Verschlusslauten hat.

Hinsichtlich der dritthäufigsten Gruppe innerhalb der Konsonanten, den Frikativen, ergibt sich für die slawischen Sprachen ein durchschnittlicher Anteil von ca. 0,15. Auffällig sind in dieser Hinsicht nur das Slowenische und das Niedersorbische, deren Frikativanteil in etwa bei 0,10 liegt.

Die Gruppe mit der geringsten Häufigkeit bilden die Affrikaten, die nur im Polnischen und im Weißrussischen einen überdurchschnittlich hohen Anteil von 0,05 haben, während der Durchschnitt für alle anderen Sprachen in einem Intervall zwischen 0,0318 (Russisch) und 0,0170 (Makedonisch) liegt.

Damit können abschließend die Konsonantenhäufigkeiten in slawischen Sprachen (ohne das Bulgarische und das Makedonische, die aus erklärbaren Gründen ein abweichendes Verhalten zeigen) folgendermaßen charakterisiert werden: Es ergibt sich für alle slawischen Sprachen folgende Rangreihenfolge: Die häufigste Gruppe sind die sonoren Laute, gefolgt von den Verschlusslauten (außer im Weißrussischen, dort ist der Anteil an Frikativen höher), Frikativen und Affrikaten. In Anbetracht der Tatsache, dass immer wieder Ausnahmen auftreten, können die hier gemachten Beobachtungen wohl insgesamt nur als eine generelle Tendenz verstanden werden. In Zukunft müssen zu diesem Themenkomplex sicherlich noch systematische Analysen durchgeführt werden.

2.3.1. Ranghäufigkeiten von Konsonanten: Fallbeispiel Slowenisch

Die in Kapitel 2.3. gefundene Tendenz hinsichtlich der Ranghäufigkeit von Konsonantengruppen mit der Rangreihenfolge Sonanten – Verschlusslaute – Frikative – Affrikaten darf allerdings nicht zu der Fehlinterpretation führen, dass sich hinsichtlich der Ranghäufigkeit der einzelnen Phoneme in den jeweiligen Gruppen ebenfalls eine derart „harmonische“ Reihenfolge ergibt. Vielmehr lässt sich auch für die Konsonanten – ähnlich wie für die Vokale in

Kapitel 2.2. – feststellen, dass innerhalb der einzelnen Gruppen (wie Sonore, Verschlusslaute usw.) die einzelnen Phoneme ein zum Teil intern differenziertes Häufigkeitsverhalten haben. Dieses ist nunmehr genauer zu besprechen, da dies für die Frage der Modellierung (vgl. Kap. 2.6.) von entscheidender Bedeutung ist.

Das Verhalten von Konsonanten innerhalb von Subgruppen wird im Folgenden anhand von slowenischen Häufigkeiten illustriert. Zu beginnen ist mit den sonoren Phonemen des Slowenischen. Es ergibt sich – sortiert nach relativer Häufigkeit – folgende Reihenfolge: /n, r, v, j, m, l/, wobei das Phonem /n/ einen relativen Anteil von 0,05912, und das Phonem /l/ einen Anteil von 0,03827 hat. Betrachtet man nun die graphische Darstellung der Ranghäufigkeit von Sonanten (vgl. Abb. 3a), so fällt auf, dass diese in der Tat eine mehr oder weniger gleichmäßig abfallende Kurve bilden.

Hinsichtlich der Rangreihenfolge von Verschlusslauten ergibt sich ein etwas anderes Bild. Die abfallende Rangreihenfolge (vgl. Abb. 3b), besteht im Grunde genommen aus zwei Teilkomponenten: Einem hochfrequenten Bereich (es sind dies die Phoneme /t, k, p, d/) und einem davon abgetrennten Bereich mit Phonemen mit einer niedrigeren Frequenz (es sind dies die Phoneme /b, g/).

Ein ähnliches Bild wie soeben für die Verschlusslaute festgestellt, liefern Ranghäufigkeiten von slowenischen Frikativen: Auch innerhalb dieser Gruppe ist ein großer Unterschied in der Ausnutzung einzelner Frikative zu beobachten: Während das Phonem /s/ einen sehr hohen Anteil von 0,055707 hat, kommt das Phonem /f/ – und dies kann sprachhistorisch erklärt werden – auf einen sehr geringen Anteil von 0,0002. Betrachtet man die Ranghäufigkeit aller Frikative (vgl. Abb. 3c), so ergibt sich wiederum ein hochfrequenter und ein deutlich davon abgetrennter Bereich von niedrigfrequenten Frikativen.

Abschließend – um damit das Gesamtbild der Konsonantenhäufigkeiten zu vervollständigen – ist auf die Affrikaten einzugehen, die als Subgruppe den niedrigsten Gesamtanteil einnehmen, aber innerhalb der Gruppe ein bekanntes Bild zeigen: Während /č/ mit einem Anteil von 0,014809 relativ häufig ist, kommt das Phonem /dž/, dessen phonologischer Status bekannterweise im Slowenischen umstritten ist, auf einen sehr niedrigen Anteil von 0,000146, oder – um es plakativer auszudrücken – dieses Phonem kommt in einem Text mit insgesamt 129787 Phonemen gerade einmal 19 Mal vor.

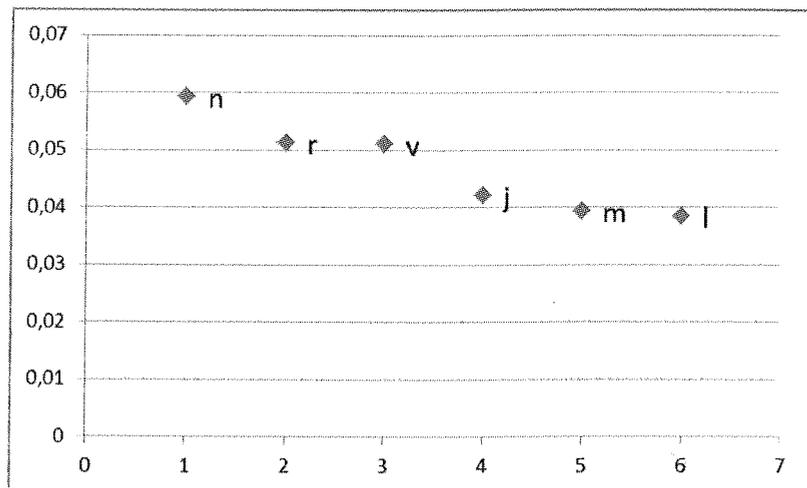


Abb. 3a: Ranghäufigkeiten: Sonanten

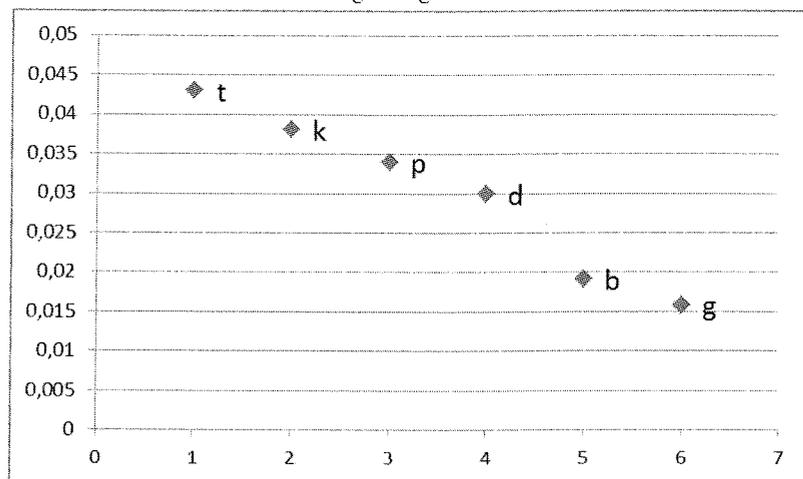


Abb. 3b: Ranghäufigkeiten: Verschlusslaute

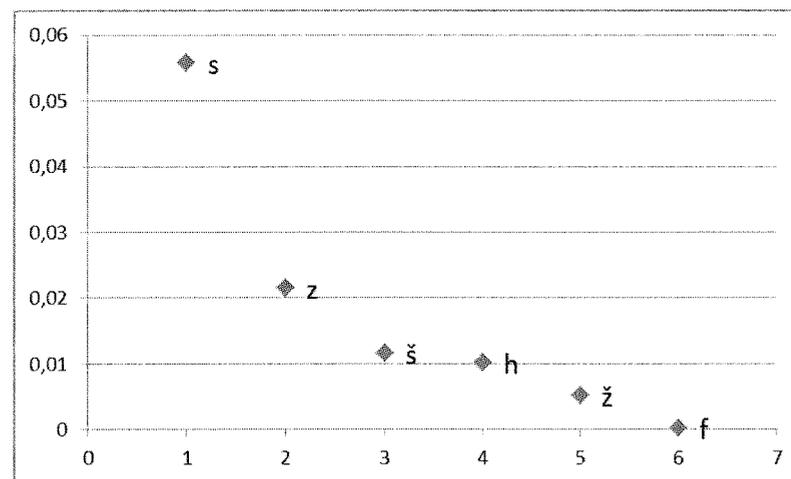


Abb. 3c: Ranghäufigkeiten: Frikative

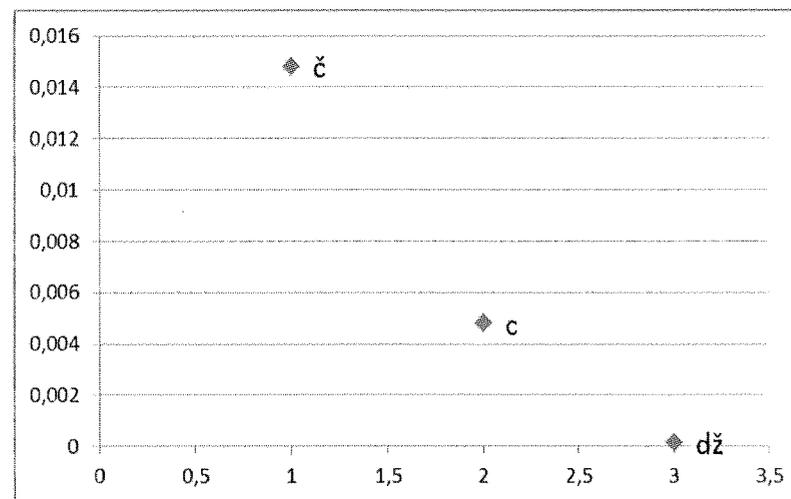
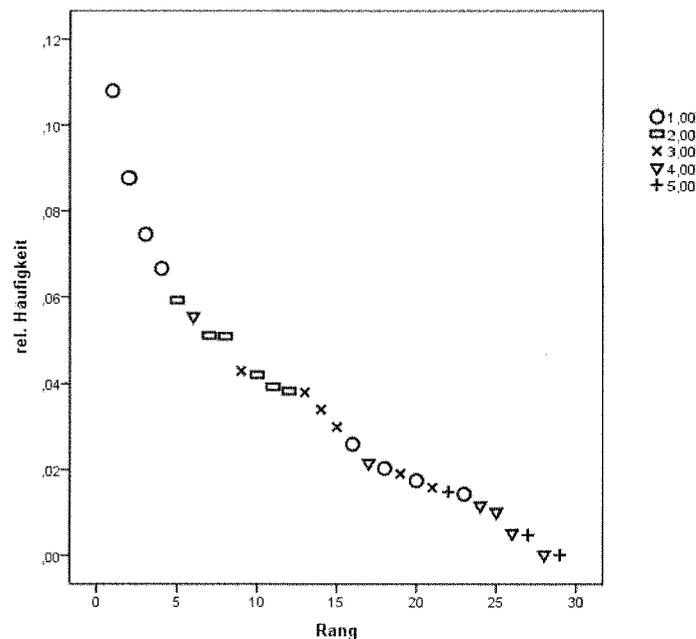


Abb. 3d: Ranghäufigkeiten: Affrikaten

Zusammenfassend lässt sich sagen, dass die Konsonanten jeder Subgruppe ein spezifisches Häufigkeitsverhalten aufweisen, welches darin besteht, dass es zu einer ungleichmäßigen Verteilung der Funktionalität im Sinne von „häufigen“ und „weniger häufigen“ Phonemen innerhalb einer Gruppe kommt. Dieses Phänomen ist aber ein gemeinsames systemlinguistisches Verhalten von sowohl Konsonanten- als auch Vokalhäufigkeiten.

2.3.2. Synthese: Slowenische Phonemhäufigkeiten als Ganzes

Nachdem nun sowohl die Vokal- als auch die Konsonantenhäufigkeit des Slowenischen im Detail besprochen worden ist, kann eine abschließende Synthese erfolgen. Betrachtet man nämlich die Rangreihenfolge aller 29 slowenischen Phoneme, so ergibt sich wiederum keine harmonische Abfolge von Vokal – Sonanten – Verschlusslaute – Frikative – Affrikaten, wie dies die Häufigkeiten der jeweiligen Subgruppen suggerieren könnten, sondern es kommt zu einer mehr oder weniger geordneten Abwechslung und Mischung der Abfolge der einzelnen Phonemgruppen. Zur Illustration dieses Sachverhalts wurde in Abb. 4 die Ranghäufigkeit des Slowenischen mit entsprechender Markierung der einzelnen Subgruppen dargestellt.



Legende: 1 = Vokale; 2 = Sonanten; 3 = Verschlusslaute; 4 = Frikative;
5 = Affrikaten

Abb. 4: Slowenische Phonemhäufigkeiten (relativer Anteil)

Die Abfolge – man ist nun mit einer mehr oder weniger harmonisch abfallenden Reihe konfrontiert – beginnt mit vier Vokalen, die sodann von einem Sonanten unterbrochen wird, der seinerseits von einem Frikativ gefolgt wird. Danach besetzen neuerlich Sonanten die entsprechenden Rangplätze, die sodann von einem Verschlusslaut usw. gefolgt werden. Diese auf den ersten Blick chaotisch erscheinende Abfolge der einzelnen Phoneme ist sicherlich nicht universell für eine Sprache gültig, sondern sie dürfte auch je nach Textsorte, Funktionalstil usw. differieren.

Darüber hinaus ist die beobachtete unterschiedliche Gewichtung der Phoneme, die zu der festgestellten Mischung in der Reihenfolge der einzelnen Phoneme führt, eine wichtige systemlinguistische Eigenschaft eines sprachlichen Systems. Offenbar wird dadurch die Redundanz und die Gleichverteilung der einzelnen Phoneme reguliert und organisiert. Im nächsten Kapitel wird die Frage einer adäquaten statistischen Modellierung der untersuchten slawischen Phonemhäufigkeiten zu diskutieren sein, um zu zeigen, dass die beobachteten Rangreihenfolgen kein chaotisches Verhalten aufweisen, sondern streng organisiert sind und mathematisch exakt beschrieben werden können.

2.4. Statistische Modellierung von Rang-Graphemhäufigkeiten

Eine Grundannahme der synergetischen Linguistik (vgl. Köhler 1986, 2005) ist, dass sprachliche Systeme auf jeder beliebigen Ebene durch Kräfte der Selbstregulation geformt werden. Eine partikuläre Ausformung dieser Selbstregulation kann anhand des Häufigkeitsverhaltens der einzelnen Subgruppen von Phonemen (Vokale, Sonanten, Affrikaten usw.) beobachtet werden. Innerhalb dieser Subgruppen bilden die einzelnen Phoneme bei Bildung einer Rangreihenfolge eine mehr oder weniger glatt abfallende Kurve, die in einigen Fällen – wie in Kapitel 2.3.1. gezeigt – in einen hoch- und niedrigfrequenten Bereich „zerfallen“ kann. Aus der Sicht der quantitativen Linguistik stellt sich an dieser Stelle die zentrale Frage, auf welche Art und Weise die Rangreihenfolge der einzelnen Phoneme mathematisch erfasst werden kann. Im Vordergrund der vorliegenden Arbeit stehen statistische Modelle (stetige Funktionen), die als adäquate Beschreibungsmöglichkeit angesehen werden können. Zuerst werden die Phoneme innerhalb der Subgruppen (Vokale, Sonanten, Verschlusslaute usw.) und im Anschluss daran die gesamte Ranghäufigkeitsverteilung modelliert.

In der quantitativen Linguistik geht man in der Regel bei der Modellierung von Ranghäufigkeiten von Phonemen von einem linguistischen Basismodell in der Form von $y = ax^b$ aus, welches besagt, dass sich die Häufigkeit von Phonemen in Abhängigkeit des Ranges proportional ändert. Der Parameter b gibt dabei Auskunft über die Schnelligkeit des Abfalls der sich ergebenden Potenz-

kurve, während der Parameter a die Verschiebung auf der y-Achse verursacht. In jedem Fall geht es dabei um einen nichtlinearen Verlauf. In sehr vielen Fällen – und Details müssen hier nicht diskutiert werden – reicht allerdings eine derart einfache Funktion nicht aus, um Phonem- bzw. Graphemhäufigkeiten zu modellieren (vgl. dazu Kelih 2009b), sondern es müssen weitere Parameter eingeführt werden.

Wie aus den bisherigen Analysen der empirischen Gegebenheiten deutlich geworden ist, ist es möglich, die einzelnen Teilgruppen (Vokale, Frikative, Verschlusslaute usw.) als gesonderte Entitäten zu verstehen, deren Häufigkeitsverhalten als „einfacher“ zu interpretieren ist als die zu einer gesamten Ranghäufigkeitsverteilung zusammengeführten Teilsysteme. In letzterem Fall kann also von einer Mischung von Teilsystemen gesprochen werden. Die Frage, die sich aus systemlinguistischer Sicht stellt, ist, welche Konsequenzen aus dieser Mischung von Teilsystemen für die theoretische Modellierung gezogen werden.

An dieser Stelle lässt sich ein zentraler Querbezug zu einem von Popescu/Altmann/Köhler (2010) entwickelten Schichtenmodell herstellen. In dieses Modell geht in erster Linie die Überlegung ein, dass linguistische Daten eine Mischung von mehreren Strata, d.h. Teilsystemen, sind. So wird bezogen auf die Häufigkeit von lexikalischen Einheiten argumentiert, dass Auto- und Synsemantika aufgrund grammatikalischer Faktoren ein unterschiedliches Häufigkeitsverhalten aufweisen und demnach Lexemranghäufigkeiten kaum durch ein einfaches Modell zu erfassen sind, sondern Mischungen auch auf Modellenebene anzunehmen sind. Mathematisch ausgedrückt wird diese Überlegung durch folgende Formel:

$$y = 1 + ae^{-bx} + ce^{-dx} + \dots$$

In diesem Fall sind a , b , c und d iterativ zu bestimmende Parameter, während e die Exponentialfunktion ist.

Es wird deutlich, dass es sich hier um eine Addition von zwei Teilmodellen (Exponentialfunktionen) handelt, gleichzeitig ist aber dieses Modell – je nach linguistischen Bedürfnissen – beliebig erweiterbar. Da dieses Modell bislang nur in wenigen Fällen auf Phonemhäufigkeiten angewandt worden ist (vgl. Rajyashree 2008), bleibt zu klären, welche Arten von Schichten, Strata bzw. Teilsystemen für Phonemhäufigkeiten angenommen werden können.

Wie aus der in den Abschnitten 2.2. und 2.3. durchgeführten Analyse der empirischen Häufigkeiten von Teilgruppen von Vokalen und Konsonanten deutlich geworden ist, besteht eine Möglichkeit darin, genau diese Teilgruppen als Teilsysteme einer vollständigen Phonemranghäufigkeit aufzufassen. Die jeweiligen Teilgruppen sind dabei jeweils durch eine Teilkomponente des

Modells zu erfassen. Im Anschluss daran kann sodann die gesamte Phonemhäufigkeit modelliert werden, wobei eine unterschiedlich hohe Anzahl von Schichten (d.h. Teilmodellen) verwendet werden soll. Dieses Vorgehen einer schrittweisen Modellierung soll zuerst anhand des slowenischen Datenmaterials und sodann im Anschluss für alle Sprachen durchgeführt werden.

Für alle Teilsysteme (d.h. Ranghäufigkeiten von Vokalen, Sonanten, Verschlusslauten usw.) wird in einem ersten Schritt das Teilmodell $y = 1 + ae^{-bx}$ angesetzt werden, um empirisch zu überprüfen, ob die entsprechenden Häufigkeiten der Teilgruppen dadurch zu erfassen sind. Die Parameter a und b werden aus den Daten geschätzt.

Ob nun diese Modelle für die einzelnen Subgruppen von Phonemen relevant sind, lässt sich statistisch gesehen dadurch absichern, dass für die Güte der Übereinstimmung zwischen empirischen Häufigkeiten und den theoretischen Werten der sogenannte Determinationskoeffizient R^2 berechnet wird. Dieser erreicht bei einer vollständigen Übereinstimmung zwischen empirischen und theoretischen Werten den Wert 1. Entsprechend einer Konvention in der quantitativen Linguistik und sonstigen empirischen Wissenschaften werden aber bereits R^2 Werte > 0.85 als hervorragende Übereinstimmung interpretiert.

Versucht man nun die einzelnen Teilgruppen mit dem soeben diskutierten Modell zu erfassen, so zeigt sich folgendes Bild: Alle Subgruppen können durch das obige Potenzmodell zufriedenstellend beschrieben werden. Es ergeben sich in allen Fällen R^2 Werte > 0.93 . Demnach erweist sich, dass in der Tat ein Teilsystem durch das vorgeschlagene Teilmodell erfasst werden kann. In Tab. 4 sind die berechneten Parameter-Werte und der R^2 angeführt.

Tab. 4 : Slowenisch: Modelle für Ranghäufigkeiten von Teilsystemen

Subsystem	a	b	R^2
Vokale	19533,76	0,27	0,93
Sonanten	8305,04	0,09	0,94
Stopp	6988,13	0,17	0,93
Frikative	15255,91	0,76	0,97
Affrikaten	6844,15	1,26	0,98

Die fünf angeführten Teilgruppen lassen sich jeweils durch ein einfaches statistisches Modell erfassen. Über die Modellierung der Ranghäufigkeiten innerhalb der einzelnen Gruppen hinaus, stellt sich abschließend die zentrale Frage, welche Auswirkungen dieser Befund für die Modellierung der Gesamtrangverteilung hat. In jedem Fall ist davon auszugehen, dass eine Gesamtrang-

verteilung von Phonemen eine Mischung von unterschiedlichen Subsystemen darstellt.

Zu diesem Zweck soll nunmehr das Schichtenmodell von Popescu/Altmann/Köhler (2010) empirisch in folgenden Schritten getestet werden:

$$(1) y = 1 + ae^{-bx}$$

$$(2) y = 1 + ae^{-bx} + ce^{-dx} + \dots$$

$$(3) y = 1 + ae^{-bx} + ce^{-dx} + fe^{-ex} \text{ .}^{11}$$

D.h. es wird mit jedem Schritt ein weiteres Teilmodell hinzugefügt, um induktiv festzustellen, wie viele Schichtungen dem untersuchten Material zugrunde liegen.

Tab. 5: Anpassungen: Slowenische Phonemhäufigkeiten

Schichten	Parameter						R ²
	a	b	c	d	f	g	
1	13710,15	0,09					0,97
2	9751,19	1,31	12458,74	0,08			0,98
3	4623,06	0,08	9751,20	1,31	7835,68	0,08	0,98

Das Ergebnis dieser schrittweisen empirischen Validierung lässt folgende Interpretation zu: Für alle drei verwendeten Modelle ergeben sich hohe R²-Werte, die zwischen 0,97 und 0,98 liegen. In dieser Hinsicht ist es überraschend, dass die gesamte Phonemhäufigkeit zufriedenstellend durch eine einzige Komponente erfasst werden kann (R² = 0,97). Eine geringfügige Verbesserung ergibt sich, wenn man zwei Komponenten ansetzt. In diesem Fall ergibt sich ein R² von 0,98. Das Interessante und Hervorhebenswerte ist aber, dass eine dritte Komponente nicht notwendig ist, da sich der R²-Wert nicht mehr weiter verbessert, sondern auf dem Niveau des Zwei-Komponentenmodells bleibt. Darüber hinaus haben im Fall des Drei-Schichtenmodells die Exponenten *b* und *g* den gleichen Wert (in beiden Fällen 0,08). Aufgrund dieses Parameterverhal-

¹¹ In dieser Formel ist die additive Konstante 1 in der Funktion als minimale Häufigkeit eingesetzt worden. In dem Fall, dass kein Phonem genau einmal vorkommt, wird stattdessen ein *k* (= minimale Häufigkeit) eingesetzt.

tens kann geschlossen werden, dass eine Kombination von zwei Teilmodellen für das Slowenische ausreicht, um ein optimales Ergebnis zu erhalten. Dieser Befund der Notwendigkeit eines „gemischten“ Modells weist somit darauf hin, dass die Gesamtverteilung keineswegs eine Summe von sechs Einzelgruppen ist, sondern die Phonemhäufigkeiten durch die Kombination von zwei Teilmodellen zufriedenstellend zu erfassen sind.

Mit anderen Worten, es bewahrheitet sich nunmehr die Annahme, dass ein System nicht als die Summe von Einzelkomponenten in Erscheinung tritt, sondern – wie im vorliegenden Fall – eine Selbstregulation hinsichtlich der Rangreihenfolge der einzelnen Elemente in Gang gesetzt wird, die durch zwei unterschiedliche Regulationskräfte gesteuert wird.

Aus empirischer Sicht zeigt sich für das Slowenische, dass mit dem Zwei-Schichtenmodell ein R² von 0,98 erreicht wird, was eine nahezu perfekte Übereinstimmung der empirischen und theoretischen Werte signalisiert. Diese hervorragende Anpassung und somit Validität des vorgeschlagenen Modells ist in Abb. 5 dargestellt.

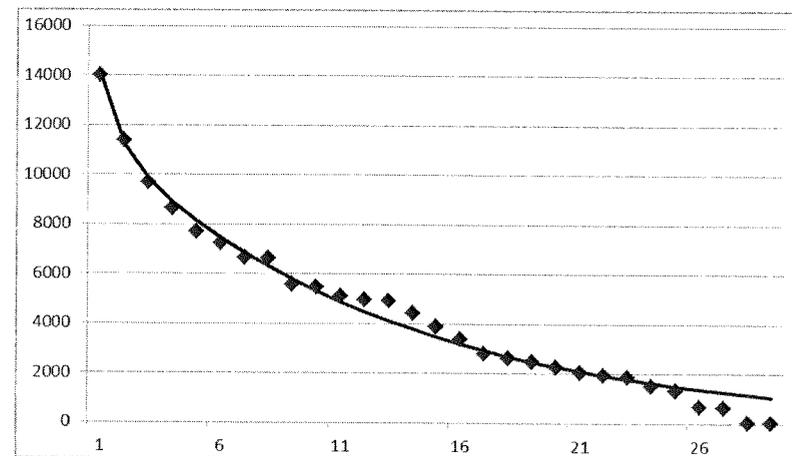


Abb. 5: Slowenische Phonemranghäufigkeiten: Empirische Werte und theoretische Kurve

Abschließend ist die offene Frage zu beantworten, ob das anhand der slowenischen Phonemhäufigkeiten illustrierte und empirisch nachgewiesene Modell auch seine Gültigkeit in den anderen slawischen Sprachen hat. Es werden daher drei unterschiedliche Modelle, die jeweils eine unterschiedliche Anzahl von Schichten beinhalten, schrittweise getestet werden.

Die Ergebnisse der entsprechenden Anpassungen sind in Tab. 5 zu finden. Die Ergebnisse lassen sich folgendermaßen interpretieren: Zu beginnen ist mit

dem Ein-Schichtenmodell, welches in Summe einen durchschnittlichen R^2 von 0,9047 erreicht. Dieser niedrige Durchschnittswert ergibt sich vor allem aufgrund der nicht überzeugenden Modellierung des Russischen ($R^2 = 0,76$) und des Weißrussischen ($R^2 = 0,64$). Ohne diese beiden Sprachen wird ein durchschnittlicher R^2 von 0,95 erreicht, was in Summe als gutes Ergebnis zu interpretieren ist. Dennoch zeigt die Modellierung durch ein Ein-Schichtenmodell, dass für die hier untersuchten slawischen Sprachen kein gemeinsames Modell zugrunde gelegt werden kann. Auffällig ist auch, dass vor allem das Russische und das Weißrussische, die eine starke Ausprägung der Palatalisierungskorrelation aufweisen, in diesem Fall als Ausreißer auftreten. Ob dies eine adäquate Begründung ist, wird in Zukunft näher zu untersuchen sein.

Bevor das Zwei-Schichtenmodell näher kommentiert wird, soll auf das Drei-Komponentenmodell eingegangen werden. Hier zeigt sich in Summe neuerlich ein Phänomen, welches sich bereits im Slowenischen gezeigt hat: Das Drei-Schichtenmodell ist im Grunde genommen kein adäquates Modell, obwohl die R^2 -Werte insgesamt sehr gut ausfallen (der R^2 -Wert liegt im Durchschnitt bei ca. 0,97) und auch keinerlei Ausreißer zu beobachten sind. Wie ein Blick in Tab. 5 zeigt, ist in allen untersuchten Sprachen – außer dem Bulgarischen – zu beobachten, dass zumindest zwei Exponenten (die Parameter b , d und g) jeweils den gleichen Wert haben. Die entsprechenden Werte sind in der Tabelle entsprechend (fett) markiert. So gilt z.B. für das Makedonische, das Polnische, das Niedersorbische usw. $b = d$ und für das Tschechische, das Slowakische usw. gilt $b = g$. Dieses Verhalten ist ein zentraler Hinweis darauf, dass im Drei-Schichtenmodell einige der Parameter z.T. schlichtweg redundant und daher eliminierbar sind. Einzig für das Bulgarische scheint ein Drei-Schichtenmodell notwendig zu sein.

Damit ist zum Zwei-Schichten-Modell überzugehen, welches für die vorliegenden Sprachen und Datensätze als durchwegs optimal anzusehen ist: Es ergibt sich ein durchschnittlicher R^2 von 0,9746, was im Vergleich zum Drei-Schichtenmodell ($R^2 = 0,9780$) eine nur äußerst geringfügige Verschlechterung darstellt und als hervorragendes Ergebnis interpretiert werden muss. Darüber hinaus zeigt sich hinsichtlich der Parameter-Werte, dass diese in keinem einzigen Fall übereinstimmen.

Tab. 5: Theoretische Parameter und Anpassungsergebnisse: 10 slawische Standardsprachen

Sprache	Schichten	Parameter						R^2
		a	b	c	d	f	g	
Slowenisch	1	13710,15	0,09					0,97
	2	9751,19	1,31	12458,74	0,08			0,98
	3	4623,06	0,08	9751,20	1,31	7835,68	0,08	0,98
Serbisch	1	138936,45	0,11					0,96
	2	119495,74	0,10	34985,98	0,45			0,96
	3	34985,98	0,45	24934,25	0,10	94561,44	0,10	0,97
Makedonisch	1	1487,65	0,14					0,98
	2	1076,10	0,11	665,45	0,44			0,98
	3	539,66	0,11	665,44	0,44	536,4318	0,11	0,98
Bulgarisch	1	2569,57	0,13					0,98
	2	257022327	13,13	2241,70	0,12			0,99
	3	445764,78	2,17	2486,60	0,12	415857,68	2,11	0,99
Polnisch	1	9907,23	0,10					0,90
	2	9461,22	0,48	5667,98	0,06			0,97
	3	6041,50	0,47	3419,71	0,47	5667,98	0,06	0,97
Tschechisch	1	723,48	0,07					0,95
	2	12312,63	0,07	153201,13	3,70			0,98
	3	3417,63	0,07	88651,26	3,14	8825,52	0,07	0,98
Slowakisch	1	2023,69	0,09					0,91
	2	1191,98	0,06	1717,33	0,43			0,96
	3	594,17	0,06	1717,33	0,42	597,80	0,06	0,96
Niedersorbisch	1	112,40	0,11					0,94
	2	89,36	0,83	87,59	0,09			0,97
	3	41,22	0,09	46,368	0,09	89,36	0,82	0,97
Russisch	1	19337,45	0,245					0,76
	2	4609,50	0,05	29698,03	0,75			0,94
	3	14822,19	0,74	4609,50	0,05	14875,83	0,74	0,94
Weißrussisch	1	1148,95	0,123					0,64
	2	593,93	0,06	93569,91	4,26			0,99
	3	44346,76	4,26	593,93	0,06	49234,91	4,26	0,99

Damit scheint ein adäquates Modell für die Modellierung aller hier untersuchten slawischen Phonemhäufigkeiten gefunden worden zu sein. In jedem Fall zeigt sich, dass die Regulation der Phonemhäufigkeiten von zumindest zwei

unterschiedlichen Selbstregulations-Mechanismen gesteuert wird. Es wird deutlich, dass Phonemhäufigkeiten als selbstorganisiertes Ganzes zu verstehen sind, in denen die Teilsysteme von Vokalen und Konsonanten ein aufeinander abgestimmtes Häufigkeitsverhalten zeigen.

3. Zusammenfassung

Die vorliegende Untersuchung versteht sich als Pilotstudie zu ausgewählten Aspekten der quantitativen Phonologie slawischer Sprachen. Ein erstes wichtiges Resultat – welches in Zukunft noch anhand anderen Materials zu zeigen sein wird – ist die fehlende Wechselbeziehung zwischen der paradigmatischen und der syntagmatischen Ebene. Bezogen auf die Phonologie der slawischen Sprachen bedeutet dies, dass die Anzahl der Phoneme, über die eine Sprache verfügt (= Phoneminventarumfang), weder in kausaler noch in statistischer Wechselwirkung zur Häufigkeit der einzelnen Phoneme steht. In diesem Sinne sind diese beiden Eigenschaften als unabhängig zu bezeichnen. Neben der Deskription und der Untersuchung von Korrelationen ist die statistische Modellierung von Phonemranghäufigkeiten ein erster Versuch in Richtung einer Erklärung von beobachtbaren Phänomenen. Es zeigt sich, dass die Ranghäufigkeit von Phonemen in allen untersuchten Sprachen durch ein bestimmtes statistisches Modell erfasst werden kann und somit von einem gemeinsamen gesetzmäßigen Verhalten auszugehen ist. Das verwendete Modell erlaubt auch die induktive Extraktion der den jeweiligen Daten zugrundeliegenden Schichten. Vorläufig gibt es Anzeichen dafür, dass als entsprechende Schicht die jeweilige Frequenz von Vokalen und Konsonanten in Frage kommt. Diese ersten Hinweise sind als Motivation für eine weiterführende linguistische Interpretation der Ähnlichkeit und gleichzeitigen Individualität der slawischen Einzelsprachen auf der Ebene von Phonemhäufigkeiten zu verstehen.

4. Literatur

4.1. Verwendete Literatur

- Altmann, G. (1972): „Status und Ziele der quantitativen Sprachwissenschaft.“ In: Jäger, S. (ed.) (1972): *Linguistik und Statistik*. Braunschweig, 1-9.
- Altmann, G. (1973): „Mathematische Linguistik.“ In: Koch (ed.) (1973), 208-232.
- Altmann, G. (1978): „Towards a theory of language.“ In: Altmann (ed.) (1978): *Glottometrika* 1. Trier, 1-25.

- Altmann, G. (1997): „The art of quantitative linguistics.“, in: *Journal of Quantitative Linguistics* 4, 1/3; 13-22.
- Altmann, G.; Zadorozhna, I., Matskulyak, J. (eds.) (2008): *Problems of General, Germanic and Slavic Linguistics. Papers for the 70th Anniversary of Professor V. Levicikij*. Chernivtsi.
- Andersen, H. (1978): „Vocalic and consonantal languages.“ In: Birnbaum et al. (eds.) (1978), 1-12.
- Birnbaum, H. et al. (eds.) (1978): *Studia linguistica Alexandro Vasilii filio Is-satschenko a collegis amicisque oblata*. Lisse
- Budilovič, A.S. (1883): *Načertanie cerkovnoslavjanskoj grammatiki, priměnitel'no k obščej teorii russkago i drugih rodstvennych jazykov*. Varšava.
- Burlakova, M.I.; Nikolaeva, T.M.; Segal, D.M.; Toporov, V.N. (1962): „Strukturnaja tipologija i slavjanskoe jazykoznanie.“ In: *Strukturno-tipologičeskie issledovanija. Sbornik statej*. Moskva, 3-18.
- Comrie, B.; Corbett, G.G. (eds.) (1993): *The Slavonic languages*. London, New York.
- Dikarev, M.A. (1891/1892): „Očerk voronežskogo meščanskogo govora sravnitel'no s ukraino-russkim narečiem (Opyt statističeskogo issledovanija zvukovyh javlenij).“ In: Dikarev, M.A. (ed.) (1891/1892): *Voronežskij etnografičeskij sbornik*. Izdanie Voronežskago Gubernskago Statističeskago Komiteta. Voronež, 1-68.
- Fenk-Oczlon, G.; Fenk, A. (2008): „Complexity trade-offs between the subsystems of language.“ In: Miestamo/Sinnemäki/Karlsson (eds.) (2008), 43-65.
- Fischer, K. B.; Krumbholz, G.; Lazar, M.; Rabięga-Wisniewska, J. (eds.) (2010): *Beiträge der europäischen slavistischen Linguistik (Polyslav)*. München. [= Die Welt der Slaven Sammelbände/Sborniki, 40]
- Grzybek, P.; Kelih, E. (2003): „Graphemhäufigkeiten (am Beispiel des Russischen). Teil I: Methodologische Vor-Bemerkungen und Anmerkungen zur Geschichte der Erforschung von Graphemhäufigkeiten im Russischen.“, in: *Anzeiger für Slavische Philologie* 31, 131-162.
- Gryzbek, P.; Kelih, E. (2004): „Towards a General Model of Grapheme Frequencies for Slavic Languages.“ In: Garabik, R. (ed.) (2004): *Computer Treatment of Slavic and East European Languages*. Bratislava, 73-87.
- Grzybek, P.; Kelih, E.; Stadlober, E. (2006): „Graphemhäufigkeiten des Slowenischen (und anderer slawischer Sprachen). Ein Beitrag zur theoretischen Begründung der sog. Schriftlinguistik.“, in: *Anzeiger für Slavische Philologie*, 34, 41-74.
- Hartenstein, K.; Jachnow, H. (eds.) (1991): *Slavistische Linguistik 1990*. München. [= Slavistische Beiträge, 274]

- Isačenko, A.V. (1939/1940): „Versuch einer Typologie der slavischen Sprachen.“, in: *Linguistica Slovaca*, 64-79.
- Kelih, E. (2009a): „Phonemverbindungen und Inventarumfang: Empirische Evidenz und Modellentwicklung.“, in: *Glottology*, Jg. 1, H. 2, 60-74.
- Kelih, E. (2009b): „Graphemhäufigkeiten in slawischen Sprachen: Stetige Modelle.“, in: *Glottometrics*, 18, 53-69.
- Kelih, E. (2010a): „Vokal- und Konsonantenanteil als sprachtypologisches Merkmal slawischer Literatursprachen.“ In: Fischer et al. (eds.) (2010), 70-77.
- Kelih, E. (2010b): „Wortlänge und Vokal- Konsonantenhäufigkeit: Evidenz aus slowenischen, makedonischen, tschechischen und russischen Paralleltexten.“, in: *Anzeiger für Slavische Philologie*, Jg. 36, 7-27.
- Kelih, E. (2015): *Wechselbeziehungen zwischen Phonologie, Morphologie und Syntax: Die Rolle des Phoneminventars*. [im Druck]
- Kempgen, S. (1991): „Isačenkos Typologie der slavischen Sprachen aus heutiger Sicht.“ In: Hartenstein/Jachnow (eds.) (1991), 146-163.
- Koch, W.A. (ed.) (1973): *Perspektiven der Linguistik*. Stuttgart.
- Köhler, R. (1986): *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum. [= Quantitative Linguistics, 31]
- Köhler, R. (2005): „Gegenstand und Arbeitsweise der Quantitativen Linguistik.“ In: Köhler/Altmann/Piotrowski (eds.) (2005), 1-16.
- Köhler, R.; Altmann, G.; Piotrowski, R.G. (eds.) (2005): *Quantitative Linguistik. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook*. Berlin u.a. [= Handbücher zur Sprach- und Kommunikationswissenschaft, 27]
- Krámský, J. (1946/1948): „Fonologické využití samohláskových foném.“, in: *Linguistica Slovaca*, 4-6, 39-43.
- Krámský, J. (1959): „A quantitative typology of languages.“, in: *Language and Speech*, 2, 72-85.
- Kubáček, L. (1994): „Confidence limits for proportions of linguistic entities.“, in: *Journal of Quantitative Linguistics* 1 (1), 56-61.
- Majewicz, A. F. (1989): *Języki świata i ich klasyfikowanie*. Warszawa.
- Miestamo, M.; Sinnemäki, K.; Karlsson F. (eds.) (2008): *Language complexity: typology, contact, change*. Amsterdam, Philadelphia
- Perebyjnis, V.S. (1970): *Kil'kisini ta jakisni charakterystyky systemy fonem sučasnoï ukraiïnskoj literaturnoï movy*. Kyïv.
- Popescu, I.-I., Altmann, G., Köhler, R. (2010): „Zipf's law – another view.“, in: *Quality and Quantity* 44 (4), 713-731.
- Rajyashree, K.S. (2008): „The Phoneme- Grapheme Correspondence in Marathi.“ In: Altmann et al. (eds.) (2008), 503-517.

- Rehder, P. (ed.) (1998): *Einführung in die slavischen Sprachen. (Mit einer Einführung in die Balkanphilologie)*. Darmstadt.
- Segal, D.M. (1972): *Osnovy fonologičeskoj statistiki (na materiale pol'skogo jazyka)*. Moskva.
- Sawicka, I. (2008): „The phonetics of Macedonian.“, in: *Language Typology and Universals*, 61, 2, 2008, 147-152.
- Sawicka, I.; Holvoet, A. (eds.) (1991): *Studies in the phonetic typology of the Slavic languages*. Warszawa.
- Trubetzkoy, N.S. (1939): *Grundzüge der Phonologie*. Prague. [= Travaux du Cercle Linguistique de Prague, 7] [zitiert nach 7. Auflage 1989, Göttingen]
- Trummer, M. (1993): „Zur Wechselwirkung von Konsonantismus und Vokalismus am Balkan: Akkomodation (Palatalisierung und Velarisierung) und Reduktion.“, in: *Balkansko ezikoznanie*, 36, 1, 73-78.

4.2. Quellenangaben (untersuchte Sprachen)

Slowenisch:

- Kolter, F. (1994): *Kvantitativna fonološka analiza slovenščine. Kosmačev roman Pomlachi dan*. Klagenfurt: Univ.-Diplomarbeit.

Serbisch:

- Ković, I. (1982): „Neki odnosi unutar relativnih frekvencija fonema i distribucija fonema prema mestu artikulacije.“ In: *Zbornik II. znanstvenega srečanja „Računalniška obdelava lingvističnih podatkov“*. Bled, 53-61.

Makedonisch:

- Džukeski, A. (1955): „Procentualnata sostojba na glasovite vo makedonskiot literaturni jazik.“, in: *Makedonski jazik*, VI, 2, 165-200.

Bulgarisch:

- Marinova, M., Marinov, A. (1964): „Statističeski izsledvanija na fonemite v bälgarskija knižoven ezik.“, in: *Bälgarski ezik* 1964, 2-3, 173-179.

Russisch:

- Kučera, H.; Monro, G.K. (1968): *A Comparative Quantitative Phonology of Russian, Czech and German*. New York.

Slowenisch	Serbisch	Makedonisch	Bulgarisch
			x 140
			x' 11
			c 106
			c' 4
			č 235
			dž 2
			š 285
			ă 425
			abs. 18213

2. Ostslawische Sprachen

Weißrussisch		Russisch	
a	1879	p	2309
i	547	p'	475
e	448	b	1096
y	445	b'	368
o	432	t	4266
j	425	t'	1885
r	412	d	1665
ŷ	177	d'	1038
c'	170	k	3175
č	160	k'	534
s'	158	g	1304
l	143	c	562
š	142	č	1633
x	129	f	948
n	386	f'	59
u	364	v	2978
t	359	v'	1016
k	302	s	3093
d	283	s'	1863
n'	259	z	1367
s	249	z'	322
m	247	š	1564
p	246	ž	951
l'	214	x	987

Weißrussisch		Russisch	
v	208	m	2317
w	195	m'	805
z	140	n	4100
v'	106	n'	2297
b	105	l	2664
c	105	l'	2081
k'	79	r	2907
š	74	r'	1381
m'	74	j	4137
dz'	59	i	14132
p'	53	a	17586
z'	43	u	3495
b'	38	e	2614
ŷ'	16	o	4026
f	13	abs.	100000
f'	8		
x'	6		
dž	3		
dz	2		
abs.	9903		

3. Westslawische Sprachen

Nieder-sorbisch		Polnisch		Slowakisch		Tschechisch	
i	50	i	3391	a	2058	e	15236
ê	15	y	3841	o	2056	o	10739
e	75	e	10612	e	1656	a	10379
a	117	a	9651	i	1196	i	9358
o	101	o	7981	t	818	t	7415
ó	19	u	2830	s	786	s	7316
u	39	ę	100	v	760	n	7092
z	12	ą	510	m	710	l	6760
ž	4	j	4371	n	680	i	6275
f	1	w	1810	r	678	k	6260
x	3	l	1920	k	654	v	6038
s	50	r	2940	l	646	m	6333
š	16	m	3501	p	616	r	5627

Das Plusquamperfekt im Ukrainischen im Vergleich zum Bulgarischen

Yana Movchan (Kiel)

1. Einleitung

Das Plusquamperfekt hat – wie auch andere Tempora – seinen Ursprung im Altkirchenslavischen bzw. im Gemeinslavischen. Somit verfügten die Vorläufer aller slavischen Sprachen am Anfang ihrer Entwicklung über vier Tempora im Präteritalbereich: die synthetischen Tempora Imperfekt und Aorist sowie die analytischen Tempora Perfekt und Plusquamperfekt. Allerdings haben nur einige der modernen Slavinen wie z.B. das Bulgarische, Makedonische, Serbische und Kroatische diese vier Tempora bis heute beibehalten. Andere slavische Sprachen wie z.B. das Tschechische, Polnische und Russische haben dagegen das komplizierte Tempussystem auf ein Präteritaltempus, das sog. Präteritum, reduziert. Damit ist heute auch das Plusquamperfekt in vielen Sprachen aus dem Gebrauch verschwunden. Ungeklärt sind in dieser Hinsicht die Verhältnisse in der ukrainischen Sprache, in der das Plusquamperfekt sich heute an der Grenze seiner Existenz zu befinden scheint. Nichtsdestoweniger lässt sich auch im Ukrainischen die Existenz von Verbformen des Typs *вчив був* „ich hatte studiert“ nicht leugnen. Im folgenden Artikel wird deshalb das Plusquamperfekt im heutigen Ukrainischen näher untersucht. Dies geschieht auf der Grundlage eines Vergleichs mit dem Bulgarischen. Diese Vorgehensweise wurde ausgewählt, weil das Plusquamperfekt im Bulgarischen recht ausführlich in den Grammatiken beschrieben ist und in der modernen Sprache reichlich verwendet wird. Ziel der Untersuchung ist es einerseits festzustellen, wie umfangreich das Plusquamperfekt in den ukrainischen Grammatiken dargestellt wird, und andererseits, ob diese Darstellung (wenn vorhanden) dem heutigen Gebrauch dieser Tempusform entspricht. Anschließend wird mit zwei

ANZEIGER FÜR
SLAVISCHE PHILOLOGIE

BAND XLI

BEGRÜNDET VON
RUDOLF AITZETMÜLLER UND
LINDA SADNIK

HERAUSGEGEBEN VON
WOLFGANG EISMANN, RENATE HANSEN-KOKORUŠ
UND NORBERT NÜBLER



Akademische
Druck- u. Verlagsanstalt
Graz/Austria
2015