

**Branko Tošović**

## **Die Distanz zwischen den slawischen Sprachen**

**0.** Den Gegenstand dieser Analyse bildet die Untersuchung der Distanz zwischen slawischen Standardsprachen (in weiterer Folge: SSS) und deren Verhältnis zur deutschen Sprache (De), wobei die Analyse auf der Basis von Korpusmaterialien, die einen Zeitraum von 40 Jahren (1970–2010) umfassen, und von daraus entstehenden Wörterbüchern, Verzeichnissen, Datenbanken u. Ä. erfolgt. Unter dem Begriff Korpus wird eine elektronische Sammlung von Texten (a) in einer oder mehreren Sprache(n) verstanden, die (b) über extralinguistische (Angaben über den Texttyp, die/den VerfasserIn, Datum und Ort der Veröffentlichung...) und/oder intralinguistische Informationen (lexikalisch-semantische, grammatikalische...) verfügen mit (c) daraus resultierenden Verzeichnissen, Wörterbüchern und anderen Datenbanken, die (d) allen oder einer begrenzten Zahl an NutzerInnen online zur Verfügung stehen und (e) zu wissenschaftlichen Forschungen und für den Unterricht herangezogen werden. Man unterscheidet zwischen einsprachigen, zweisprachigen, mehrsprachigen, parallelen, annotierten und nichtannotierten Korpora, zwischen Korpora schriftlicher Texte und gesprochener Sprache u. a. Im Rahmen dieser Analyse wird für die untersuchten Sprachen (im Folgenden: US) ein mehrsprachiges Parallelkorpus erstellt werden, das über eine extralinguistische und intralinguistische Annotierung verfügt.

**1.** Unter dem Begriff der Distanz versteht man das Verhältnis zwischen dem Objekt **A** und dem Objekt **B**, wobei auf den Grad von deren Nähe/Entfernung, wie auch auf den sich zwischen diesen Objekten befindlichen Raum hingewiesen wird. Die Sprache **A** kann demnach über Eigenschaften verfügen, die eine Distanz und einen Zwischenraum in Bezug auf **B**, **C** ... **X** schaffen. Jede Sprache nimmt eine bestimmte Position auf einer Skala zwischen „nahe“ und „fern“ ein. Als Maßeinheit für diese Nähe bzw. Entfernung dient die Distanz, die über eine gewisse Zahl an strukturellen Charakteristiken verfügt: Qualität, Quantität, Intensivität, Niveau, Grad, Ursache, Richtung ( $A \rightarrow B$ ,  $B \rightarrow A$ ,  $A \leftrightarrow B$ ) u. Ä.

Es gibt unterschiedliche Arten der Distanz: eine strukturell-disziplinäre (intellektuelle, linguistische, kulturelle, mathematische, mathematisch-linguistische, politische, psychische, soziolinguistische, ethnische...), eine quantitative (minimale, unbemerkbare, unbedeutende, winzige, große, riesige...), eine intensionelle (zielgerichtete, progressive, regressive), eine wertende (analytische, /un/günstige, relevante, /un/erwartete, /nicht/vorhergesehene, /nicht/messbare, merkliche, optimale, prognostizierte, divergente, konvergente, reale, perzeptive...) u. a. Die Distanz ist mit verschiedenen Prozessen verbunden, von denen vor allem die Bremsung, Beschleunigung, Anziehung und Abstoßung, die Konvergenz und die Divergenz zu nennen wären, die mit zwei Kräften – der Zentrifugal- und der Zentripetalkraft – einhergehen. Die Distanz nimmt Einfluss auf den Grad des Erlernens und die Schwelle der Verständlichkeit von Sprachen.

William Mackey unterscheidet einige Typen der Distanz zwischen Sprachen: 1. Distanz zwischen Sprachsystemen einerseits und Distanz im Diskurs, 2. statische vs. dynamische Distanz, wobei die statische als paradigmatische erscheint, die Unterschiede bei Elementen und Konstruktionen zwischen Sprachen aufweist, wogegen sich die dynamische syntagmatisch präsentiert, indem sie ein Ergebnis der konkreten Parole (speech) darstellt, 3. Distanz als Differenz zweier Sprachsysteme (Teilsysteme) vs. Distanz als Konversion von einer Sprache in die andere, 4. taxonomische vs. integrale Distanz, 5. inhaltliche vs. formale Distanz (gleiche Formen können unterschiedliche Bedeutungen haben), 6. Distanz in der Intensivität (Diversität, Intensivität und Produktivität der sprachlichen Differenzierung), indem in manchen Sprachen einige Kategorien

stärker differenziert sind als in anderen (Mackey 1971: 105–106). „Insgesamt erlauben die genannten Unterscheidungen und die durchgeführten Differenzmessungen ein Urteil über die kontrastive Beziehung von Sprachen im Kontakt und indirekt Prognosen über Schwierigkeiten und Distanzen in den kommunikativen Kontakten zwischen Sprechern der jeweiligen Sprachen prognostizieren“ (Mackey 1971: 106–107). Weiters werden von Mackey Beispiele zur Illustration der Möglichkeiten einer Bestimmung des Grades an Distanz mithilfe von Berechnungen einer quantitativen Indikation angeführt.

Bei der Definition von polyzentrischen Sprachen werden von Ulrich Ammon drei Typen von sprachlicher Distanz unterschieden: eine kleine – typisch für Standardvarianten einer polyzentrischen Sprache (so etwa die Distanz zwischen dem österreichischen und dem deutschen Standard der deutschen Sprache), eine mittlere – minimale sprachliche Distanz zwischen Standardvarianten unterschiedlicher Sprachen (die so genannten Ausbau-Sprachen mit z. B. der Distanz zwischen dem Letzeburgischen und dem Deutschen) und eine große – zwischen zwei beliebigen Varianten, bei denen es sich um verschiedene Sprachen handelt (die so genannten Abstand-Sprachen) – Ammon 2005: 1538. Analog dazu unterscheidet Ammon auch die Grade des Einflusses der Distanz auf das gegenseitige Verstehen, wobei ein hoher Grad kein Verstehen zulässt, ein mittlerer eine Verständigung mit erheblichen Schwierigkeiten ermöglicht, während ein kleiner Grad keinerlei Verständigungsprobleme mit sich bringt.

S. E. Jachontov vertritt die Meinung, dass die Sprachwissenschaft (in erster Linie die Soziolinguistik) eine eigene Skala zur Messung der Ähnlichkeiten der Sprachen benötigen würde (Jachontov 1980), die auf der praktischen Bedeutung für die Sprechenden und ForscherInnen beruhen müsste. Dieses Modell würde aus fünf Ebenen bestehen: a) Die TrägerInnen unterschiedlicher Idiome kommunizieren frei miteinander, b) die TrägerInnen unterschiedlicher Idiome kommunizieren ohne größere Mühe miteinander, obwohl einzelne Fälle von Nichtverstehen auftreten können, c) die TrägerInnen unterschiedlicher Idiome können sich nicht frei unterhalten, d) eine Verständigung ist nicht möglich und e) die Verwandtschaft zweier Sprachen kann nur von Fachleuten entdeckt werden. Jachontov definiert dies wie folgt:

„1. Die kleinsten sprachlichen Unterschiede, die für Sprechende bemerkbar sind: Die TrägerInnen unterschiedlicher Idiome kommunizieren frei miteinander, doch können sie anhand von Besonderheiten in der Aussprache und zum Teil in der Lexik in etwa bestimmen, woher die jeweils andere Person stammt. [...] Für eine Bestimmung solcher Unterschiede reichen 200 Jahre aus: Die englische Sprache in den USA verfügt bereits über ihre eigenen Dialekte, die in keinem Zusammenhang mit den englischen Dialekten in Großbritannien stehen. [...] 2. Die TrägerInnen unterschiedlicher Idiome kommunizieren ohne größere Mühe miteinander, obwohl einzelne Fälle von Nichtverstehen auftreten können. Einen solchen Charakter weist etwa das Verhältnis zwischen dem Russischen und Ukrainischen, dem Tatarischen und dem Baschkirischen sowie dem Usbekischen und dem Uigurischen auf. Das Alter dieser Unterschiede beträgt rund 500 oder etwas mehr Jahre. [...]. 3. Die TrägerInnen unterschiedlicher Idiome können sich nicht frei unterhalten, hören jedoch im Gespräch laufend ihnen bekannte Wörter und sogar kurze Phrasen. [...] Ein derartiges Verhältnis besteht etwa zwischen dem Russischen und dem Bulgarischen oder Polnischen, dem Türkischen und Tatarischen [...]. Das Alter dieser Unterschiede liegt bei 1000–1500 Jahren. [...] 4. Eine Verständigung ist nicht möglich, doch bei einem systematischen Studium der Sprachen entdeckt man eine Vielzahl an gemeinsamen Wörtern (die beinahe aller grundlegenden Begriffe betreffen) und grammatikalischen Regeln, einschließlich der Ähnlichkeit sehr vieler grammatikalischer Morpheme [...]. Dieser Grad an Verwandtschaft lässt sich mit geringerer Bestimmtheit festmachen als die drei zuvor genannten, weil er stärker von subjektiven Empfindungen geprägt ist. Als Beispiel dafür könnten das Englische und das Schwedische dienen, aber eventuell

auch die wesentlich weiter entfernten Sprachen russisch und litauisch. Das Alter der Unterschiede dieses Grades an Verwandtschaft beläuft sich auf maximal 2000 Jahre: Germanische Stämme mit eigenen Dialekten gab es bereits knapp nach Christi Geburt. [...] **5.** Die Verwandtschaft zweier Sprachen kann nur von Fachleuten entdeckt werden; Wörter, die gemeinsames Gut zu sein scheinen, sind eher vor kurzem übernommene Entlehnungen von einer Sprache in die andere oder aus einer gemeinsamen Quelle oder auch nur zufällige Übereinstimmungen. [...] In Sprachen, die sich auf der fünften Ebene befinden, können bei einem derartigen auszugsweisen Vergleich der Lexik jedoch ebenfalls ausreichend gemeinsame Wurzeln entdeckt werden, wie etwa einige gemeinsame typologischen Charakteristiken (z. B. in erster Linie die Tendenz zur Suffigierung oder Präfigierung, spezifische Wurzelstrukturen, das Vorhandensein oder Fehlen einer Harmonie von Vokalen, des Tons oder gewisse, selten auftretende Phonemklassen) und vielleicht auch einige (wenige) grammatikalische Morpheme. Eine Verwandtschaft auf dieser Ebene liegt etwa zwischen dem Russischen und dem Englischen vor“ (ЯХОПТОВ 1980-www).

**2.** Im Rahmen der Analyse erfolgen korpusbasierte Untersuchungen zum **G r a d d e r N ä h e b z w . E n t f e r n u n g (a)** des Bosnischen/Bosniakischen (Bs), Bulgarischen (Bg), Kroatischen (Hr), Mazedonischen (Mk), Montenegrischen (Mo), Polnischen (Pl), Russischen (Ru), Serbischen (Sr), Slowenischen (Sl), Slowakischen (Sk), Sorbischen (Ls), Tschechischen (Cs), Ukrainischen (Uk) und Weißrussischen (Be) wie auch **(b)** die Beziehung genannter Sprachen zum Deutschen (De). Daneben werden, sollte sich dies als durchführbar herausstellen, auch so genannte slawische Mikrosprachen (Burgenlandkroatisch – Hg, Kaschubisch – Ks und Russinisch – Rs) in die Untersuchungen einbezogen.

**3.** Zum Grad an Nähe bzw. Entfernung zwischen den SSS liegen unterschiedliche Positionen vor, die im Falle von sehr nahen Sprachen wie etwa des Bs, Hr, Mo und Sr oder auch des Mk und Bg von überaus starker Polarisierung geprägt sein können. Entgegengesetzte Meinungen können auch hinsichtlich der soziolinguistischen Bestimmung des Charakters von SSS zu Tage treten: Handelt es sich bei Ihnen um (a) gänzlich eigene Sprachen, (b) Varianten einer (polyzentrischen) Sprache, (c) Dialekte einer Nationalsprache, (d) eine Sprache mit (politisch bedingten) unterschiedlichen Benennungen u. Ä. Das Fehlen konkreter Untersuchungen, die das Verhältnis zwischen den SSS objektiv aufzeigen und relevantes Material zur Klärung aller grundlegenden Fragen bereitstellen würde, begünstigt subjektive und tendenziöse Interpretationen des Verhältnisses zwischen den SSS sowie eine politisch gefärbte Deutung der zwischensprachlichen Korrelationen, an denen sich nicht nur SprachwissenschaftlerInnen, sondern häufig auch Nicht-PhilologInnen (in erster Linie PolitikerInnen) beteiligen. Dieser Umstand erfuhr durch den Zerfall drei slawischer föderativer Staaten (die UdSSR, die Tschechoslowakei und Jugoslawien) in den 1990er Jahren und die militärischen Konflikte in Südosteuropa eine erhebliche Zuspitzung. Besonders erschwert wurde die Situation durch unterschiedliche Auslegungen hinsichtlich der Rechtmäßigkeit bzw. Unrechtmäßigkeit der offiziellen Ernennung und Kodifizierung einiger SSS im 20. Jahrhundert (Mitte dieses Jahrhunderts Mk und in den 90er Jahren Bs, Hr und Sr). Auf die Verkomplizierung der interslawischen Sprachbeziehungen nahm auch die Änderung des Status einiger SSS Einfluss. So etwa wurden im Falle von (a) Cs und Sk, (b) Be und Uk und (c) Sl Sprachen von Teilrepubliken zu Sprachen neu gegründeter Staaten. Andererseits verlor Ru in der Ukraine und in Weißrussland seinen Status als Sprache einer überregionalen Kommunikation, sodass sich die russischsprechende Bevölkerung in diesen Staaten in einer Konfliktsituation wiederfand, in der sich Tendenzen zu einer gänzlichen Verdrängung des Ru aus beinahe allen Kommunikationssphären verstärken. Seinen Status als Verkehrssprache innerhalb eines Staates büßte auch das Serbokroatische ein, aus dessen Erbe nach dem Zerfall Jugoslawiens drei Sprachen (Bs, Hr und Sr) kodifiziert wurden, während sich eine vierte Sprache (das im Jahre 2007 in den Verfassungsrahmen erhobene Mo) gegenwärtig im Prozess der Standardisierung befindet. Einige der so genannten Mikrosprachen fordern in den neu entstandenen Staaten hingegen eine Änderung ihres

Status ein (das Russinische in der Ukraine). Zu all dem kommt ein weiterer Aspekt hinzu, der sich dadurch manifestiert, dass mit dem Zerfall von Staaten einige SSS zu Sprachen zu Sprachen nationaler Minderheiten wurden (z. B. Hr in Serbien oder Sr in Kroatien). Ein bestimmter und in einigen Fällen erheblicher Anteil der Russischsprechenden fand sich außerhalb der Grenzen der Russischen Föderation wieder, und trat mit der Erweiterung der Europäischen Union (um die Staaten Estland, Lettland und Litauen) in die EU ein. All dies spiegelt sich auf diese oder jene Weise in den Prozessen des Zusammenwachsens Europas wider, wobei sich die Europäische Union mit den schwierigen interslawischen sprachlichen Beziehungen und den Herausforderungen zu deren Lösung konfrontiert sieht. Die vielschichtige und komplexe Sprachsituation in slawischsprachigen Staaten und Regionen macht objektive Untersuchungen der intralinguistischen, soziolinguistischen und psycholinguistischen Aspekte der wechselseitigen Beziehungen erforderlich, zumal Analysen zu sprachlichen Distanzen bis dato kaum durchgeführt wurden. Die Tatsache, dass sich De in direktem oder indirektem Kontakt mit sämtlichen SSS befindet und deutschsprachige Länder eine vielschichtige Zusammenarbeit mit der slawischen Welt unterhalten, bedingt die Notwendigkeit, die deutsch-slawischen Beziehungen in diese Untersuchungen einzubeziehen. Alle angeführten Gründe führten zu einem Reifen der Überzeugung, eine Forschung zu einer systemlinguistischen, soziolinguistischen und psycholinguistischen Analyse zumindest eines der Aspekte im Rahmen dieser komplexen Thematik aufzugreifen, das die wechselseitige Nähe bzw. Entfernung (a) der SSS und (b) der SSS und De behandelt.

№	Sprache	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Insg.
		De	Bg	Be	Bs	Cs	Hr	Ls	Mk	Mo	Pl	Ru	Sk	Sl	Sr	Uk	
1.	De	■															15
2.	Bg		■														15
3.	Be			■													15
4.	Bs				■												15
5.	Cs					■											15
6.	Hr						■										15
7.	Ls							■									15
8.	Mk								■								15
9.	Mo									■							15
10.	Pl										■						15
11.	Ru											■					15
12.	Sk												■				15
13.	Sl													■			15
14.	Sr														■		15
15.	Uk															■	15
<b>Summe</b>		<b>15</b>	<b>225</b>														

Abb. 1: Beziehungen zwischen der Sprachen

Bs – Bosnisch/Bosniakisch, Bg – Bulgarisch, Be – Weißrussisch, Cs –Tschechisch, Hr – Kroatisch, LS – Sorbisch, Mk – Mazedonisch, Mo – Montenegrinisch, Pl – Polnisch, Ru – Russisch, Sk – Slowakisch, Sl – Slowenisch, Sr – Serbisch, Uk – Ukrainisch; De – Deutsch)

4. Die Durchführung dieser Analyse und die zu erwartenden Ergebnisse besitzen elementare Bedeutung sowohl für die Disziplinen Slawistik und Germanistik als auch für slawische und deutschsprachige Staaten, die enge Kontakte zu slawischsprachigen Ländern unterhalten. Die vorliegende Untersuchung ist auch wichtig für die Slawistik in ihrer Gesamtheit, weil es das erste Mal wäre, dass auf der Basis von einheitlichem und unifiziertem Korpus texten eine systematische und komplexe Analyse der Distanz zwischen den SSS und zwischen diesen Sprachen und De durchgeführt würde. Die Ergebnisse der Analyse werden von für jede der untersuchten

Sprachen von Nutzen sein, weil sie auf die individuelle slawische und nichtslawische (slawisch-deutsche) intersprachliche Nähe bzw. Entfernung hinweisen werden. Das interslawische parallele Online-Korpus wird eine umfassende Materialbasis für weitere Forschungen zu den SSS und zum De sowie zu deren gegenseitigen Beziehungen darstellen. Für die Germanistik wird diese Untersuchung die erste groß angelegte Forschung zur Distanz zwischen De und Sprachen sein, die in genetischer Hinsicht einer anderen Sprachfamilie angehören. In Österreich, ein Land, (a) in dem auf diese oder jene Weise beinahe alle SSS gelehrt werden, (b) in dem zwei Sprachen nationaler Minderheiten beheimatet sind (Slowenisch und Burgenlandkroatisch) und (c) dass an die Slowakei, Slowenien und Tschechien grenzt und traditionell enge Beziehungen mit Staaten wie Bosnien und Herzegowina, Kroatien, Montenegro und Serbien unterhält, werden die Ergebnisse für all jene von Interesse sein, die über Kontakte zu slawischsprachigen Staaten verfügen und auf politischer, wirtschaftlicher und kultureller Ebene mit diesen zusammenarbeiten. Im Bezug auf die Steiermark fügt sich diese Analyse zur Gänze in die Orientierung dieses Bundeslandes ein, die auf eine Vertiefung der Beziehungen zu den Ländern Südosteuropas abzielt, deren BewohnerInnen acht SSS sprechen (Bg, Bs, Cs, Hr, Mk, Mo, Sl und Sr). Für die Europäische Union könnten die Untersuchungen eine Hilfestellung zur Findung einer objektiven Lösung der Frage nach dem Status der slawischen Sprachen im vereinten Europa darstellen. Gegenwärtig befinden sich fünf SSS als Sprachen von Mitgliedsländern innerhalb der EU: Bg, Cs, Pl, Sk und Sl, wobei in baldiger Zukunft auch Hr diesen Status erreichen wird, der auch von Bs, Mk, Mo, Sr u. a. angestrebt wird, sodass auf diese Weise schon bald zehn SSS der EU angehören könnten. Im Zusammenhang damit gilt es festzuhalten, dass ein erheblicher Anteil der Sprachen jener Länder, die um einen EU-Beitritt bemüht sind, der Gruppe nahe verwandter Sprachen angehören, deren Sprechende sich frei und ohne Dolmetschung unterhalten können (ein typisches Beispiel dafür sind BosniakInnen, KroatInnen, MontenegrinerInnen und SerbInnen). Aus diesem Grund werden in der EU bereits zum gegenwärtigen Zeitpunkt Diskussionen darüber geführt, wie man diese Sprachen behandeln könne und ob die Einrichtung eines eigenen Übersetzungsdienstes sinnvoll sei. Das vorliegende Analyse wäre in der Lage, ein objektives Bild der Distanz zwischen diesen Sprachen zu erstellen und aufzuzeigen, in welchem Maße Übersetzungen sinnvoll und zielführend sind. Einer Analyse der Distanz zwischen den genannten Sprachen kommt auch in Bezug auf den Unterricht, vor allem an höheren Bildungsanstalten, wesentliche Bedeutung zu. Die Analyse wird all jenen wertvolles Material liefern, die sich mit Fragen der Übersetzung, komparativen linguistischen Untersuchungen, Spracherkennung und automatischer Sprachbestimmung befassen. Für Studierende als besonders wertvoll würden sich dabei das multilinguale Korpus, die daraus resultierenden Datenbanken und bestimmte Programme erweisen, weil sie Texte zu allen oder zumindest einigen SSS beinhalten, mit denen eine Untersuchung der Ähnlichkeiten und Unterschiede zu De ermöglicht würde. Daneben wird Studierenden die Möglichkeit eingeräumt, diesbezügliche Fragestellungen im Rahmen von Seminar- und Diplomarbeiten sowie von Dissertationen zu erörtern.

**5.** Das **Hauptziel** dieser Analyse liegt darin, anhand umfangreichen Materials den Grad der Distanz zwischen den angeführten Sprachen zu untersuchen und zu bestimmen. Die Distanzen zwischen den Sprachen werden dabei auf phonetisch-phonologischer, lexikalisch-semanticischer und grammatikalischer Ebene erhoben. Die Untersuchung soll Antwort auf die Frage geben, wie nahe bzw. ferne sich Sprachen stehen, die einerseits einer genetischen Sprachfamilie angehören (der slawischen) und andererseits verschiedenen Sprachfamilien (der germanischen und der slawischen) zuzurechnen sind. Die Analyse erfolgt im Rahmen des Modells von sprachlichen Korrelationen, die im Buch „Korrelationale Syntax“ (Tošović 2001), in einer Reihe von Aufsätzen und auch auf Grundlage theoretischer Abhandlungen über den Unterschied in der Sprache und den Unterschied als allgemeinen Begriff dargelegt werden (Tošović 2008a). Weiters soll die Analyse Auskunft darüber erteilen, (1) welche linguistischen, soziolinguistischen und psycholinguistischen Prozesse die Nähe bzw. Entfernung zwischen den SSS beeinflussen, (2) in

welchem Maße die linguistische Distanz zwischen den US auf sozio- und psycholinguistische Prozesse (in erster Linie auf die Schwelle des Erkennens und Verstehens) wirkt und (3) wie sehr sich die SSS einander annähern oder sich voneinander entfernen, wobei dies im Rahmen von vier Gruppen erhoben werden soll: innerhalb einer Gruppe sehr nahe verwandter Sprachen (z. B. Bg und Mk), in einer Gruppe von territorial nahen Sprachen (etwa Be, Ru, Uk) und innerhalb einer Gruppe territorial weiter entfernter Sprachen (z. B. Hr und Cs). Besonderes Augenmerk wird den Beziehungen zwischen den SSS und De zukommen.

**6.** Sofern bekannt liegen in den SSS bis dato sowohl hinsichtlich dieser Analyse als auch in Anbetracht der konzeptuell vereinheitlichten, zur Verfügung stehenden Korporaquellen keine vergleichbaren Untersuchungen zu diesem Thema vor. Von der Analyse, deren Ergebnisse teilweise für die durchzuführenden Analysen herangezogen werden können, gilt es vor allem das in Opole (Polen) initiierte, internationale Programm mit dem Titel „Współczesne przemiany języków słowiańskich (1945–1997)“ [Aktuelle Änderungen in den slawischen Sprachen] unter der Leitung von Stanisław Gajda zu erwähnen. In den veröffentlichten 13 Bänden, die jeweils eine SSS zum Thema haben, werden sprachliche Veränderungen in der zweiten Hälfte des 20. Jahrhunderts dargestellt, wobei eine komparative Analyse zwischen den SSS jedoch ausbleibt, zumal deren gegenseitige Distanz auch nicht Thema der Untersuchungen war. In einigen, im Rahmen dieses Programmes veröffentlichten Arbeiten wurde der Versuch einer synthetischen und vergleichenden Analyse von Prozessen unternommen, deren Ergebnisse auch für diese Analyse von Interesse sind. In den bislang veröffentlichten Publikationen (zu allgemeinen und auch speziellen Fragen – wie etwa in (Gajda 2000, Ohnheiser 2003, Sawicka 2007, Mokienko/Walter 2008) – lassen sich keine exakten Antworten auf den Grad an Distanz zwischen den behandelten Sprachen finden. Prozesse aus jüngster Zeit sollen nur dann untersucht werden, wenn sie auf eine Annäherung oder ein Entfernen der SSS Einfluss nehmen. Als ein anderes, für die vorliegende Untersuchung kann ein gegenwärtig in der Slowakei durchgeführtes mit dem Titel „Konvergente und divergente Prozess in der slawischen Staaten“ genannt werden.

Als Grundlage der Analyse wird allgemeine und grundlegende Literatur über die SSS und deren wechselseitige Beziehungen (hinsichtlich Nähe bzw. Entfernung) herangezogen, wie etwa aus Werken zur komparativen Grammatik der SSS, Lehrbüchern zur komparativen Grammatik einzelner Gruppen von Sprachen (wie etwa der ostslawischen), Publikationen mit kontrastiven Analysen zweier oder mehrerer slawischer Sprachen sowie Lehrbücher und Sammelbände, in denen ein Vergleich zwischen SSS und De hergestellt wird. (In diesen Untersuchungen wird diese Distanzanalyse in der Regel nicht explizit ausgeführt, sodass auch keine direkten Antworten auf Grad und Charakter der Distanz zwischen den SSS erteilt werden.)

Ein für vorliegende Forschungen themenrelevantes Projekt trägt den Titel „Measuring linguistic unity and diversity in Europe“ und wurde von Erhard Hinrichs und Dale Gerdemann aus Deutschland (Abteilung für Computerlinguistik des Seminars für Sprachwissenschaft an der Eberhard-Karls-Universität Tübingen) sowie von John Nerbonne (Alfa-informatica der Faculté der Letteren an der Rijksuniversiteit Groningen – Hinrichs/Gerdemann/Nerbonne-www) in Zusammenarbeit mit der Kliment-Ohridski-Universität Sofia (Bulgarien) und der Bulgarischen Akademie der Wissenschaften durchgeführt. Das Ziel lag dabei in der Schaffung quantitativer Methoden zur Messung der linguistischen Vielfalt und in deren Anwendung auf sprachliche Varianten in Bulgarien und angrenzenden Regionen, wobei der Gebrauch quantitativer Methoden zur Beschreibung sprachlicher Zustände einer kritischen Überprüfung unterzogen wurde. Diese Methode wurde in Westeuropa bereits erfolgreich zur Untersuchung der sprachlichen Vielfalt und Einheit herangezogen und führte in diesem Projekt zu völlig anderen Ergebnissen. Diese Methode wurde in Westeuropa bereits erfolgreich zur Untersuchung der sprachlichen Vielfalt und Einheit herangezogen. Im genannten deutsch-niederländisch-bulgarischen Projekt wurde die sprachliche Entfernung zwischen nicht dem Standard angehörenden Varianten des Bulgarischen, zwischen

diesen Varianten und der bulgarischen Standardsprache und auch zwischen diesen Varianten und der mazedonischen Standardsprache gemessen, sodass eine Messung einer Distanz zwischen Paaren vorliegt. Bei den Untersuchungen zur sprachlichen Einheit und Vielfalt auf der Balkanhalbinsel wurde von den Wissenschaftlern festgehalten, dass die Erforschungen der Parallelen und Divergenzen zwischen denen der zuvor untersuchten westeuropäischen und denen der südosteuropäischen Sprachen die Ansichten über die sprachliche Vielfalt ausweiten würden. Zur Durchführung der Messung bediente man sich der Distanz-Methode von Levenshtein (Levenshtein distance – Levenshtein 1965). Im Zuge der Analyse wurde ein Index der sprachlichen Unabhängigkeit (index of linguistic independence), der die mittlere Distanz zwischen einer Gruppe von Varianten und einer nahen Standardsprache bezeichnet. Die Autoren unterschieden dabei die äußere Vielfalt und die innere Einheit. Unter äußerer Vielfalt (external diversity) versteht man dabei die Unterschiede zwischen regionalen Sprachen, Sprachen von Minderheiten und Standardsprachen (die Distanz zwischen offiziellen Sprachen). Die Analyse sollte zeigen, dass etwa das Friesische eine Distanz von 43,5% zum Standardniederländischen (Dutch) aufweist, während die Distanz zwischen dem Standardniederländischen und Deutschen 43,4% beträgt. Als innere Einheit (internal unity) wurde die Kohärenz von Gruppen definiert.

Als für die Analyse relevant erweisen sich auch die Ergebnisse einer in den Proceedings of the Workshop – Coling ACL (Nerbonne/Hinrichs 2006) abgedruckten Untersuchung zur sprachlichen Distanz. Im Artikel „Linguistic Distances“ weisen die Herausgeber dieser Publikation darauf hin, dass das Hauptaugenmerk auf der sprachlichen Distanz oder Ähnlichkeit liegt.

Daneben trifft man auch auf eine Reihe linguistischer Versuche, die Distanz zwischen bestimmten Sprachen zu erheben. Einer dieser Versuche geht auf O. Revzina zurück, die eine Messung der Distanz zwischen verwandten Systemen slawischer Sprachen vornahm (Revzina 1970). In Anbetracht des Grades dieser Distanzen nahm sie eine Einteilung in fünf Hauptgruppen verwandter Systeme vor: eine polnische (unter Einbeziehung des Obersorbischen und Slowakischen), eine russische (inklusive dem Ukrainischen), eine serbische, slowenische und tschechische. Der größte Gegensatz ließe sich laut Revzina zwischen dem russischen und polnischen Typus ausmachen: „Es ist interessant, dass der tschechische Typus nur leicht in Richtung des Polnischen geht und die zwei südslawischen Typen – der serbische und der slowenische – eine verhältnismäßige symmetrische Position zwischen dem russischen und dem polnischen einnehmen, wobei der slowenische Typ weiter von den beiden anderen Typen abweicht. [...] Eine verhältnismäßig symmetrische Position hinsichtlich der Genera Maskulinum, Femininum und Neutrum mit erheblichen gegenseitigen Abweichungen sind für die serbische und tschechische Sprache charakteristisch, die keine neuen Genera entwickelten. Für den polnischen und russischen Typus, in denen die Auflösung der alten Genera am und dementsprechend auch die Ausweitung ihres Inventars am weitesten fortschritt, sind einerseits die asymmetrische Positionierung der Genera im System und andererseits eine Verkleinerung der Distanz zwischen den Genera typisch“ (Revzina 1970: 30).

Pavle Ivić führte eine Analyse über die genetische Nähe der slawischen Sprachen auf phonologischem Niveau durch (Ivić 1998: 66–67), im Rahmen derer er das Serbokroatische und das Russische verglich. Zu diesem Zweck nahm er eine Analyse von zehn einsilbigen Wörtern mit urslawischer Etymologie vor, die zum allgemeinen lexikalischen Erbe zählen. Es waren dies die Wörter *sin* – сын, *list* – лист, *lek* – лек, *red* – ряд, *led* – лед, *naš* – наш, *luk* – лук, *bok* – бок, *san* – сон und *lan* – лен. Ivić gelangte dabei zur Schlussfolgerung, dass ein System aus zehn Vokalen beinahe den gesamten Phonembestand des Späturslawischen abzudecken vermag. Der Index der Änderungen, der als Maß für die genetische Entfernung herangezogen werden kann, ergibt sich aus folgender Formel

$$\frac{i(*N)}{\Sigma Ni},$$

in der  $i$  die Anzahl der Sprachen und  $N$  die Anzahl der Phoneme bezeichnet. Das Ergebnis für die serbokroatische Sprache stellt sich wie folgt dar:

$$\frac{2 \times 10}{5+5} = 2$$

Auf V. Ginsburgh, I. Ortuño-Ortín und S. Weber (Ginsburgh et al. 2007) geht eine Untersuchung über die Distanz zwischen Sprachen hinsichtlich der Nützlichkeit von deren Erlernen zurück, wobei eine größere Distanz den Nutzen eines Lernens erhöhe. Daneben würden die Kosten des Erlernens einzig von der Distanz zwischen Sprachen abhängen: Je näher die Sprachen, desto kleiner die Kosten, wobei die Autoren feststellen, dass die Distanz zwischen Sprachen auch den Lernerfolg beeinflusst.

Morris Swadesh erstellte Mitte des 20. Jahrhunderts eine 100 Wörter umfassende Liste der lexikalischen Kernelemente (die anfangs 215 Wörter zählte), und vertrat die Meinung, dass bei 300 Elementen die Obergrenze liegen würde (Swadesh 1999, Swadesh\_lists1-www). Zum elementaren lexikalischen Kern zählte er die Pronomina, die Zahlwörter und auch die Benennungen von Körperteilen, geographischen Gegebenheiten, gewissen elementaren Naturerscheinungen sowie von allen Menschen eigenen Tätigkeiten, denen in jeder Gesellschaft universale Bedeutung zukommt und in jeder Sprache Ausdruck verliehen wird. Dazu würden im Besonderen die Wörter *ich, ihr, wir, dieser, jener, wer?, was?, nein/nicht, alles* u. a. zählen. Swadesh wollte ein für alle Sprachen zutreffendes Verzeichnis erstellen, was sich jedoch als nicht durchführbar herausstellen sollte. Die grundlegenden Ansätze von Swadesh' Theorie können wie folgt zusammengefasst werden: 1. Im Wörterbuch jeder beliebigen Sprache findet sich ein bestimmter Teil, der als grundlegend und stabil angesehen werden kann und die elementaren, grundlegenden und gleichzeitig auch alltäglichen Begriffe umfasst. 2. Es gibt Bedeutungen, die in jeder beliebigen Sprache unbedingt mit Wörtern aus diesem elementaren Teil zum Ausdruck gebracht werden. Beim Vergleich des prozentuellen Anteils der Wörter aus diesem Kern, die in Paaren nahe verwandter Sprachen erhalten blieben, unternahm Swadesh den Versuch, jene Zeitspanne zu messen, die seit dem Moment der Trennung dieser beiden Sprachen vergangen war. Dabei stellte er im Besonderen fest, dass sich der grundlegende Teil des Wörterbuches mit beständiger Geschwindigkeit laufend verändert. Diese widersprüchliche Methode war vor allem in den 1960er und 1970er Jahren populär. Doch gerade in letzter Zeit wird Swadeshs Methode erneut aufgegriffen, wobei auf ihr basierend versucht wird, neue Konzeptionen auszuarbeiten. Dieser Weg wurde etwa von Viktor Kromer besprochen (Kromer 2004, 2005), der im Rahmen der Analyse die von Swadesh genannten Gesetzmäßigkeiten unter Hinzuziehung einiger präzisierender Methoden untersucht. Die Konzeption Kromers setzt sich aus vier Punkten zusammen. **(1)** Neben dem Faktor der Divergenz von Wörterbüchern und gemäß den Postulaten von Swadesh wird vom Vorhandensein des Faktors einer gleichzeitigen sprunghaften Ersetzung eines Teils der Wörter der grundlegenden lexikalischen Liste ausgegangen. Die modifizierte Methode erlaubt somit auch eine Untersuchung von Pidgin und Kreolsprachen. Ebenfalls ermöglicht sie für jede Sprachgruppe (die ebenfalls nach dieser neuen Methode definiert werden) die Rekonstruktion einer Sprache (oder von mehreren Sprachen), die direkt zur Ursprache dieser Gruppe führt (bzw. führen). Die bis dato erhaltenen Ergebnisse zeigen, dass im Hinblick auf die keltischen Sprachen das Bretonische der Ursprache am nächsten steht, unter den germanischen Sprachen das Dänische und Deutsche und hinsichtlich der slawischen Sprachen das Slowenische. In Bezug auf die Dendrogramme kann festgehalten werden, dass neben der Divergenz auch die Konvergenz die Notwendigkeit einer zweiten Messung erforderlich macht. **(2)** Es ist notwendig, die in der Praxis auftretende Verzerrung von Den-

drogrammen abzuklären, weil in der Praxis zur Analyse in höherem oder geringerem Ausmaß auch Mischsprachen herangezogen werden. (3) Das Ziel dieser Untersuchung liegt schließlich in der Entwicklung eines von Viktor Kromer bereits früher vorgeschlagenen Drei-Parameter-Modells zur Wortlänge, um dadurch den Grad der Nähe von Sprachen auf Basis eines Parameter-Vergleichs zu bestimmen.

Die Distanz zwischen Sprachen wurde auch ausgehend von der Interferenz untersucht. Betreffend diesen Zugang wird seitens einiger WissenschaftlerInnen die Meinung vertreten, dass eine kleine typologische Distanz zwischen verwandten Sprachen, d. h. ein hoher Grad an Ähnlichkeiten und geringe Unterschiede, die Wahrscheinlichkeit des Auftretens von Interferenzen erhöhe. Auf diese Weise würde eine große Distanz, wie etwa zwischen genetisch nicht verwandten Sprachen, die Fälle von automatischer Übernahme und somit auch von Fehlern verringern (Dmitrijeva-www).

Zur Erhebung der Distanz zwischen Sprachen auf semantischer Ebene (bzw. zur Messung der Beurteilung der Botschaft lexikalischer Einheiten) kann das semantische Differential von Osgood (Osgood et al. 1957) herangezogen werden, wie auch eine von Danko Šipka durchgeführte Analyse belegt, in der er diese Methode zur Bestimmung der Unterschiede zwischen Hr und Sr hinsichtlich der Auslegung von Wörtern als „eigene“ und „fremde“ zur Anwendung brachte (Šipka 2008).

Mit der Frage der Distanz beschäftigte sich auch die *S o z i o l i n g u i s t i k*, indem sie sich in erster Linie der Erklärung der Unterschiede zwischen nahe verwandten Sprachen und auch zwischen Varietäten einer Standardsprache widmet. Zu nennen wären dabei vor allem die Arbeiten Ulrich Ammons, die im Besonderen die genetische, typologische und linguistische Distanz zum Inhalt haben (Ammon 1987b). Ammon zufolge sei für eine genetische Klassifizierung die genetische Distanz von Bedeutung, während für eine typologische die linguistische (d. h. eine auf der Grammatik basierende) ausschlaggebend wäre. Er vertritt die Meinung, dass in Fällen, in denen sich LinguistInnen hinsichtlich der Zahl an Varianten einer Sprache uneinig seien, das Kriterium der linguistischen Distanz wesentlich sei und man sich in strittigen Fällen auf dieses stützen müsse. Dabei gelte es die allgemeine Regel zu berücksichtigen, wonach sich der Status einer Variante auf eine Sprache bezieht, zu der auch eine beliebige andere Variante gehört, wenn zwischen diesen beiden Varianten eine relativ geringe linguistische Distanz besteht. In einem solchen Falle handle es sich nicht um unterschiedliche Sprachen, sondern um zwei Varianten einer Sprache. Doch auch diese Regel weise erhebliche Unzulänglichkeiten auf, weil etwa die Methoden zur Messung der Distanz äußerst unvollständig und zum Teil nur mit Mühe anwendbar seien (Ammon 1987a: 247). Ammons Distanzbegriff korreliert mit Heteronomie (mit der Abhängigkeit einiger Varianten in Bezug auf andere), wobei einer standardsprachlichen Variante in dieser Hinsicht eine privilegierte Stellung zukomme. Sämtliche Varianten, die in Bezug auf eine bestimmte Standardvariante als heteronom erscheinen, gehören zur gleichen Sprache, der auch die Standardvariante zuzurechnen ist. Im Zusammenhang mit dem Konkurrenzverhältnis zwischen Autonomie und Heteronomie und dem Kriterium der linguistischen Distanz wird bei der Ammon-Distanz betont, dass Heteronomie die Kriterien einer linguistischen Distanz nicht disqualifiziere, sondern zu einer Modifizierung beitragen könne (Ammon 1987b: 326). Ammon hebt weiters hervor, dass unter den SprachwissenschaftlerInnen Einigkeit darüber bestehe, dass all jene Varianten, zwischen denen eine große linguistische Distanz besteht, trotz der Heteronomie verschiedenen Sprachen angehören, wogegen es im Falle einer kleinen Distanz unterschiedliche Auffassungen gebe. Ammon führt das Beispiel des niederdeutschen Dialektes an, der gewöhnlich dem Deutschen zugerechnet wird, obwohl seine Distanz zur deutschen Standardsprache größer sei als die zum niederländischen Standard. Der Konflikt zwischen dem Kriterium der linguistischen Distanz und der Heteronomie werde oftmals zu Gunsten der Heteronomie entschieden. Auf diese Weise könnten Fälle beobachtet werden, in denen eine kleine linguistische Distanz zu vergrößern versucht wer-

de, um einen besonderen Status einer Sprache zu erlangen, der als Symbol einer nationalen Unabhängigkeit dienen kann. Andererseits gebe es Bemühungen, aus Gründen einer Assimilierung große linguistische Distanzen zu verringern. Außerdem treffe man auch auf Fälle, in denen Varianten mit einer kleinen Zahl an Sprechenden aus politischen und wirtschaftlichen Interessen einer einzigen Sprache zugerechnet werden, ungeachtet dessen, dass zwischen diesen Varianten eine erhebliche linguistische Distanz vorliege. Ammon zweifelt daran, dass in naher Zukunft ein universal einsetzbares Maß zur Messung der Distanz geschaffen werden könne (Ammon 1986).

Einen wichtigen soziolinguistischen Aspekt stellt der Einfluss der Sprachpolitik auf eine Vergrößerung oder Verkleinerung der Distanz zwischen den SSS dar. Es sind uns keine Arbeiten bekannt, in denen der Aspekt, inwieweit Sprachpolitik Einfluss auf eine Prozesse einer Annäherung bzw. Entfernung von SSS, einer komplexen Behandlung unterzogen worden wäre. So liegen bloß Arbeiten vor, in denen die Beziehungen zwischen zwei oder drei Sprachen analysiert werden. In diesem Zusammenhang wären einige Artikel von Snježana Kordić zu nennen, in denen sie die Frage der Distanz zwischen Bs, Hr und Sr im Rahmen von deren Definition als eigene Sprachen als Standardsprachen oder Varianten einer polyzentrischen Sprache behandelte (wobei sie sich in erster Linie auf Ammon berief).

Auf den Grad an Distanz zwischen den SSS wirkt in hohem Maße auch der Purismus ein, wobei dieser in den einzelnen slawischen Sprachen in heterogener Weise zum Ausdruck kommt. So etwa weisen einige SSS (z. B. Bg, Ru oder Sr) keine stark ausgeprägten puristischen Tendenzen auf, während anderen traditionellerweise ein starker Purismus innewohnt (z. B. Hr und Sl), der eine Vergrößerung der Distanz zwischen den Sprachen begünstigt. Leider wird der Purismus in der Literatur in der Regel intralinguistisch betrachtet, d. h. einzig im Rahmen einer einzigen Sprache und ohne Analyse seiner Wirkung auf andere Sprachen. Eine nicht unwesentliche Rolle kommt dabei auch der traditionellen Norm und den Regeln der Sprachetikette zu.

Auch im Rahmen der *e x a k t e n W i s s e n s c h a f t e n* trifft man eine Reihe von Methoden zur Messung der Distanz zwischen Sprachen. Einer der ersten diesbezüglichen Versuche geht auf den polnischen Universalforscher Jan Czekanowski zurück, der eine quantitative Bestimmung des Verwandtschaftsgrades von Sprachen vornahm (Czekanowski 1957, erstmals veröffentlicht 1927). Unter Heranziehung der anthropologischen Forschungen von Edward Burnett Tylor (1888) unternahm er einen Versuch zur Darstellung des Verwandtschaftsgrades der slawischen und später auch einiger anderer indoeuropäischen Sprachen. Czekanowskis Methode wurde von der Mehrzahl der LinguistInnen angesichts der willkürlichen Wahl von Parametern verworfen, doch wurden seine Ideen von einem anderen polnischen Wissenschaftler, Witold Mańczak, fortgeführt, der zur Erkenntnis gelangte, dass zwischen dem Verwandtschaftsgrad von Sprachen und deren geographischer Lage eine bestimmte Verbindung besteht (Mańczak 1981).

Eine Methode gründet sich auf den so genannten „Distanzen der Editierung“, d. h. auf dem Aufwand bezüglich Übersetzungen einer Zeichenkette in eine andere. Als einfachste und am weitesten verbreitete Editierdistanz gilt die Levenshtein-Distanz (Levenshtein 1965), deren Wesen auf der Umwandlung einer Zeichenkette in eine andere beruht.

Bei dieser Methode kommen drei Operationen – Ersetzen, Löschen und Einfügen – zur Anwendung, wobei sich die Distanz durch die Zahl der durchzuführenden Operationen definiert. So etwa sind bei einer Umwandlung des russischen Substantivs *диалог* (Dialog) hin zu *одеяло* (Decke) folgende Operationen erforderlich: Hinzufügung von **о**, Ersetzen von **и** durch **е**, Ersetzen von **а** durch **я** und Löschen von **г**, sodass das Ergebnis gemäß der Levenshtein-Distanz in diesem Falle bei 4 liegt. Die Distanz zwischen Sprachen kann in vorliegendem Falle als mittlere Distanz nach Levenshtein bezeichnet werden, als Distanz zwischen Wörtern paralleler Listen der vorliegenden Sprachen. Anstatt in orthographischer empfiehlt es sich die Wörter in phonetischer Niederschrift darzustellen, um dadurch eine Unabhängigkeit von den orthographischen Traditionen

der einzelnen Sprachen zu erzielen. Für eine solche Berechnung wurde ein Programm in der Programmiersprache PERL entwickelt, das die phono-morpho-lexikalische Distanz zwischen Dialekten des Lettischen misst (Bērziņš 2006). Levenshteins Methode findet vor allem bei der Messung der Distanz zwischen Dialekten (im Rahmen der Dialektometrie) Anwendung, wobei in diesem Zusammenhang in erster Linie die Dissertation von Wilbert Heeringa (Heeringa 2004) und ein gemeinsam mit Charlotte Gooskens verfasster Artikel (Gooskens/Heeringa 2004) zu erwähnen sind.

Eine tiefer greifende Bewertung der Distanz zwischen Sprachen lässt sich erzielen, wenn statt der Levenshtein-Distanz das Wagner-Fischer-Distanzmodell zur Anwendung gebracht wird. Bei diesem Verfahren definiert sich der Ersetzungsprozess zusätzlich auch durch die Nähe der zu ersetzenden Buchstaben (Phoneme). So etwa wiegt ein Ersatz eines Vokals durch einen anderen Vokal geringer als das Ersetzen eines Vokals durch einen Konsonanten (Bērziņš 2006). Unter konkreter Anwendung der Wagner-Fischer-Distanz ein achtgliedriges phonetisches Modell für alle Phoneme des Lettischen und Lettgallischen erarbeitet (Bērziņš/Grigorjevs 2007), dem auch alle Phoneme der slawischen Sprachen zugeordnet werden können.

William Cavnar und John Trenkle schlugen zur Kategorisierung von Texten eine Heranziehung von Zeichenfolgen im Rahmen eines N-Gramms vor (Cavnar/Trenkle 1994). Gemäß dem Zipfschen Gesetz können Wörter, Zeichen- und Lautverbindungen u. a. entsprechend deren Vorkommensfrequenz in eine Reihe gebracht werden, wobei die genannten Autoren die Anlegung von Frequenzlisten und N-Grammen für unterschiedliche Texte vorschlagen, um durch einen Vergleich der Listen und des N-Gramms des Ausgangstextes die Kategorie festzulegen, der der Text zuzurechnen ist. Bei Anwendung einer solchen Methode kann automatisch und mit hoher Trefferquote die Sprache, Kodierung und das Thema eines Textes bestimmt werden. A. A. Bērziņš schlägt die Verwendung von Frequenzlisten und N-Grammen auch für eine Messung der Distanz zwischen Sprachen vor (Bērziņš 2004a, 2004b). In diesem Falle dienen als Ausgangsdaten willkürliche, unmarkierte Textkorpora der zu untersuchenden Sprachen, wobei diesbezüglich (bislang nicht publizierte) positive Ergebnisse von A. A. Bērziņš zur Messung der sprachlichen Abstände anhand von Phonogrammen vorliegen (als Ausgangsdaten werden dabei untranskribierte, mehrsprachige Audioaufnahmen zahlreicher Sprechender herangezogen und mit dem Hidden-Markov-Modell analysiert). Der N-Gramm-Methode bedienen sich eine Reihe von WissenschaftlerInnen (z. B. Cavnar/Trenkle 1994, Cavnar/Vayda 1992, Cavnar/Vayda 1993, Kondrak 2005).

A. A. Bērziņš beschäftigt sich mit einer Methode zur Bestimmung der Distanz zwischen Sprachen, bei der jeder Laut mit einer gleichen Zahl an Informationseinheiten beschrieben wird (Bērziņš-www). Da ein Byte (256 Permutationen, d. h. Vertauschung von Elementen) für die Beschreibungen aller Laute/Phoneme der menschlichen Sprache nicht ausreicht, wäre es Bērziņš zufolge vorteilhaft, diese mit zwei Bytes zu beschreiben (65536 Permutationen). Er meint dazu: „Natürlich wäre theoretisch auch eine feinere Lautunterteilung möglich, doch erscheint uns diese als nicht zielführend, weil bei einer zu feinen Teilung im Zuge der Transkription die Fehleranfälligkeit und die Abhängigkeit vom individuellen Verständnis der Fachleute, die die Transkription vornehmen, ansteigt. Außerdem wird, soweit und bekannt ist, eine feinere Teilung auch nicht beim Verzeichnen bereits bestehender Standards phonetischer Transkriptionen durchgeführt [...] Gegenwärtig haben wir und in der Arbeit mit Phontexten baltischer Sprachen auf die Verwendung eines Pseudocodes von zwei Bytes beschränkt: Jeder vorhandene Laut wird mit zwei ASCII-Symbolen entsprechend einem bestimmten Schema erfasst“ (Bērziņš-www). Bei einer Ausweitung auf alle Sprache schlägt er die Einführung eines gewissen Maßes vor, das die phonetische Distanz zwischen zwei beliebigen Lauten, die auf physikalischen Eigenschaften beruht, durch physiologische Charakteristiken ihrer Aussprache beschreibt.

Einen besonderen Platz nehmen die bis dato noch nicht veröffentlichten, positiven Ergebnisse von Bērziņš im Hinblick auf die Messung der Distanz zwischen Sprachen anhand von

Phonogrammen ein. In diesem Fall werden Inputfile nichttranskribierte Lautschriften einer möglichst großen Zahl an Sprechenden der verglichenen Sprachen verwendet, und in methodischer Hinsicht werden die zur Spracherkennung häufig eingesetzten Hidden-Markov-Modelle eingesetzt.

Zum Analyse der Nähe von 95 indoeuropäischen Sprachen wurde eine auf 200 grundlegenden Begriffen, d. h. gemeinsamen Wurzeln fußende Distanzmatrix angelegt (Dy-en/Kruskal/Black 1992), in der die Distanz zwischen De und Ru bei 0,76 lag. Die Distanz zwischen dem Deutschen und anderen SSS stellte sich wie folgt dar: Sl 0,73, Cs und Sk 0,74, Pl 0,75, Uk 0,76. Die Distanz zwischen Ru und anderen SSS beträgt im Vergleich dazu bei Sl 0,39, Cs und Sk 0,26, Pl 0,27 und bei Uk 0,22. Hinsichtlich der europäischen Sprachen kann eine minimale Distanz zwischen De und dem Niederländischen (162), dem Dänischen (293), Schwedischen (305) und dem Englischen (422) beobachtet werden. während eine maximale Distanz zum Finnischen (1000) und Griechischen (812) vorliegt. Eine mittlere Position nehmen das Französische (756), Spanische (747), Portugiesische (753) und das Italienische (735) ein.

Eine besondere Methode schlägt Valentin Stecjuk vor, der innerhalb eines Sprachenpaares eine Erhebung aller gemeinsamen Merkmale vornimmt (Stecjuk-www). Nach einer Formel aus der Logik wird sodann die linguistische Distanz zwischen den beiden Sprachen verglichen. Diese Methode erweist sich im Besonderen zur Analyse der slawisch-nichtslawischen Verhältnisse von Interesse, die anhand von Sprachpaaren mit großer linguistischer Distanz (z. B. deutsch-russisch) hin zu einer kleineren (mit kleinen linguistischen Distanzen, etwa russisch-ukrainisch) durchgeführt werden kann. Als Summe erhält man eine gewisse Zahl an gemeinsamen Merkmalen innerhalb des deutsch-slawischen und des slawisch-slawischen Paares.

Ähnlich den Methoden zur Messung der Distanz zwischen Sprachen stellt sich der Versuch zur Bestimmung des Alters einzelner Sprachen dar, der in den 1970er Jahren von M. V. Arapov und M. M. Cherc durchgeführt wurde (Arapov/Cherc1974). Die beiden Wissenschaftler stellten sich die Aufgabe, ein mathematisches Modell der Änderungen im Wörterbuchbestand zu entwickeln und auf dessen Basis theoretisch die Abhängigkeit zwischen der Zeit des Entstehens des Wortes und seiner Position im Frequenzwörterbuch erhalten (Arapov/Cherc1974: 3).

Es gibt eine weitere, auf einer lexikalischen Datenbank fußende Methode mit semantischer Matrix, die das Ziel der Berechnung eines Algorithmus zur Bestimmung der Distanz zwischen zwei russischen Wörtern zum Ziel hat, die mit englischen Äquivalenten verbunden sind, wobei die Zahl der englischen Äquivalente, die gleichzeitig beiden Wörtern zugerechnet werden können, bestimmt wird (Potemkin-www).

An dieser Stelle kann man auch eine Methode zur Erhebung des Klangs von Sprachen erwähnen, die vom Klangbild der Sprachen der Welt ausgeht und auf der von Karl Pearson begründeten Chi-Quadrat-Verteilung basiert. (Bei dieser Verteilung handelt es sich um eine stetige Wahrscheinlichkeitsverteilung über der Menge der positiven reellen Zahlen, deren einziger Parameter  $n$  eine natürliche Zahl sein muss. Die Chi-Quadrat-Verteilung berechnet die Wahrscheinlichkeit des Auftretens bestimmter Werte der Summe unabhängiger standardnormalverteilter Zufallsvariablen, wie sie etwa bei der Ermittlung der Varianz einer Stichprobe auftreten können.) Einen diesbezüglichen Versuch unternahm eine Gruppe von Wissenschaftlern des Sibirischen Instituts für internationale Beziehungen und regionale Angelegenheiten mit Jurij A. Tambovcev an der Spitze, die die Distanz zwischen 106 Sprachen und dem Japanischen berechnete, um damit auf phonetischem Wege die Theorie der altaischen Herkunft des Japanischen zu untermauern (Zvuk-www). Die Messung der Ähnlichkeiten zwischen den Sprachen auf der Basis des Klangbildes untermauerte die Theorie. Tambovcev untersuchte im Besonderen die Kompaktheit von Gruppen und Untergruppen, d. h. in welchem Maße mehr oder weniger kompakt die Sprachen in diesen eingebettet sind (Tambovcev 2002). Der Wissenschaftler verfolgt das Ziel, ein Maß der

Verteilung bzw. ein Maß der Diffusion zu erhalten (Unter Diffusion versteht man den Prozess eines Übertragens von Materie oder Energie aus einer Masse hoher Konzentration in eine Masse niedriger Konzentration.) Mithilfe der Größe der Kompaktheit wollte Tambovcev erfahren, ob eine bestimmte Sprache einer bestimmten Gruppe oder Sprachfamilie angehört. Der Forscher hob dabei hervor, dass diese Methode einen Vergleich einer beliebigen Sprache mit einer beliebigen Gruppe erlaubt. Jurij Tambovcev erhob so z. B. die phonotypologische Distanz zwischen dem Baskischen und einer Reihe von anderen Sprachen unterschiedlicher Familien gemäß der Euklidischen Formel der Distanz in achtfacher Dimension, wobei die Frequenz des Vorkommens einzelner Vokale als maßgeblich angesehen wurde (Tambovcev 2003). Dabei ging er von der Annahme aus, dass wenn eine beliebige Sprache einer gewissen anderen Sprachfamilie zugeordnet wird, (in unserem Fall das Deutsche zur slawischen Familie), dies entweder zu einer Erhöhung oder zu einer Verringerung der Kompaktheit führt. Ist eine Sprache ähnlich, dann führt sie kein Sinken der Kompaktheit herbei, d. h. es tritt kein Ansteigen der Diffusität ein. Tambovcev weist darauf hin, dass seine Methode der Bestimmung der Ähnlichkeiten zwischen Sprachen in phonetischer Hinsicht auch auf Gruppen jene Sprachen angewendet werden kann, die dem Balkan-Sprachbund angehören. Er errechnete, dass eine Hinzuziehung des Armenischen zu einer gewissen Senkung von deren Diffusität führt – ein Prozess, der durch eine Hinzuziehung des Türkischen noch weiter vorangetrieben wird, was von einer näheren Verwandtschaft des Türkischen zu den Sprachen des Balkan-Sprachbundes zeugt.

Одним из методов является измерение общей части словаря (Read/Spicer 1952).

Für die vorliegende Analyse können auch Ergebnisse herangezogen werden, die bei Untersuchungen zur automatischen Identifizierung von Sprachen gewonnen wurden (etwa Arjen Poutsma 2001, Souter et al 1994, Combrinck/Botha 1994, Kranig 2005, Muthusamy/Lawrence 1997, Norman 1976, Newman 1987, Dunning 1994, Muthusamy et al 1994...). Damir Boras und seine Mitarbeiter entwickelten ein kleines Programm, mithilfe dessen ein bestimmter Text in den Sprachen Bosnisch/Bosniakisch, Kroatisch oder Serbisch einer der drei Sprachen zugeschrieben werden kann (Ljubešić/Nives/Boras-www).

In Arbeiten zur *Kontaktlinguistik* erweisen sich Auslegungen betreffend Lehnwörter und die sprachliche Interferenz als nutzbringend. In diesen Untersuchungen wird im Besonderen darauf hingewiesen, dass in Bezug auf die Distanz dem Vorhandensein und dem Fehlen eines direkten oder indirekten Kontaktes eine wesentliche Bedeutung zukommt (ein engerer und intensiverer Kontakt schafft die Voraussetzungen für ein Sinken der Schwelle der Verständlichkeit).

Für das vorliegende Thema als wichtig erweisen sich auch Publikationen, die psycholinguistischen Aspekten der Distanz zwischen Sprachen gewidmet sind, wobei dies vor allem auf jene Arbeiten zutrifft, in denen die Wahrnehmung sowie das Verstehen von – in erster Linie nahe verwandten – Sprachen untersucht wird. So etwa stellte Uriel Weinreich fest, dass ein Sprachkontakt nur in einem weiteren psychologischen und kulturologischen Kontext verstanden werden könne (Weinreich 1953). Er ist der Meinung, dass als Ergebnis akkordierter Bemühungen von VertreterInnen unterschiedlicher Disziplinen im Rahmen einer interdisziplinären Herangehensweise (unter Einbeziehung der Linguistik, Psycholinguistik, Soziolinguistik u. a.) ein Erzielen aussagekräftigerer Resultate zu erwarten sei.

In unserer Analyse käme es zu einer Verifizierung des Zutreffens sämtlicher beschriebener Methoden, wobei sodann all jene zur eingehenderen Anwendung gebracht würden, die in bestmöglichem Grade der Konzeption entsprechen und sich hinsichtlich zeitlichem Aufwand, Zahl der Mitarbeitenden und der zur Verfügung gestellten finanziellen Mittel als durchführbar und analysetauglich erweisen. Eine Besonderheit in Bezug auf die genannten Methoden liegt darin, dass die Distanz anhand eines unifizierten und mehrsprachigen Korpus eruiert wird – eine

Vorgehensweise, die den vorliegenden Informationen zufolge (zumindest in der Slawistik) bis dato nicht beschränkt wurde.

7. In der Analyse soll die Distanz zwischen Sprachen anhand von fünf Korrelationssystemen untersucht werden, die als das Intrakorrelational, Interkorrelational, Suprakorrelational, Superkorrelational und das Extrakorrelational bezeichnet werden. Unter dem Intrakorrelational wird ein Beziehungsnetz innerhalb einer Sprache verstanden (z. B. die Veränderungen in Ru und wie diese auf die Distanz etwa zwischen Ru und Pl einwirken). Eine interkorrelationale Korrelation entsteht durch die wechselseitigen Beziehungen zweier sehr naher Sprachen, die einen überaus hohen Grad an Verständlichkeit aufweisen (etwa Bs, Hr, Mo und Sr). Im Rahmen eines suprakorrelationalen Systems stellt den Forschungsgegenstand das Verhältnis von zwei territorial nahe stehenden SSS dar (ostslawisch, westslawisch, südslawisch) dar, während mit dem Superkorrelational eine Beziehung zwischen Sprachen beschrieben wird, die unterschiedlichen slawischen Arealen angehören (z. B. dem ostslawischen und dem südslawischen, wie etwa im Falle von Ru ↔ Bg). Als Extrakorrelational wird das Beziehungssystem zwischen genetisch unterschiedlichen bzw. nicht unmittelbar nahe verwandten Sprachen bezeichnet, was in vorliegendem Falle auf das Verhältnis zwischen SSS und De zutrifft.

№	Sprache <sup>1</sup>	Periode <sup>1</sup>	Sprache <sup>2</sup>	Periode <sup>2</sup>
1.	De <sup>1</sup>	1969-1989	De <sup>2</sup>	1990-2010
2.	Bg <sup>1</sup>	1971-1991	Bg <sup>2</sup>	1992-2012
3.	Bs <sup>1</sup>		Bs <sup>2</sup>	
4.	Hr <sup>1</sup>		Hr <sup>2</sup>	
5.	Mk <sup>1</sup>		Mk <sup>2</sup>	
6.	Mo <sup>1</sup>		Mo <sup>2</sup>	
7.	Sl <sup>1</sup>		Sl <sup>2</sup>	
8.	Sr <sup>1</sup>		Sr <sup>2</sup>	
9.	Cs <sup>1</sup>	1972-1992	Cs <sup>2</sup>	1993-2013
10.	Ls <sup>1</sup>		Ls <sup>2</sup>	
11.	Pl <sup>1</sup>		Pl <sup>2</sup>	
12.	Sk <sup>1</sup>		Sk <sup>2</sup>	
13.	Be <sup>1</sup>	1968-1988	Be <sup>2</sup>	1989-2009
14.	Ru <sup>1</sup>		Ru <sup>2</sup>	
15.	Uk <sup>1</sup>		Uk <sup>2</sup>	

Tab. 1: Korrelation zwischen den Entwicklungsperioden einer Sprache

Das erste Korrelationssystem, das **Intrakorrelational**, umfasst einzig Beziehungen innerhalb einer der US. Als Gegenstand der Analyse dienen dabei **(a)** dynamische Prozesse, indem Untersuchungen darüber angestellt werden, in welchem Maße in den letzten 40 Jahren (1970–2010; unterteilt in zwei Perioden: 1970–1990 ↔ 1991–2010) eingetretene Veränderungen in den US Einfluss auf die zwischensprachliche Distanz nehmen und zu welchen konkreten Ergebnissen diese führen. Diese können sich etwa in einer Vergrößerung oder Verkleinerung der Distanz äußern, zu einem Nichtverstehen von Neologismen, einer positiven oder negativen Wahrnehmung führen, das Schwelle der Verständlichkeit zwischen Sprechenden unterschiedlicher slawischer Sprachen erhöhen oder herabsetzen und **(b)** statistische Prozesse, von denen vor allem der Frage, wie sehr bereits vor der untersuchten Periode aufgetretene strukturell-typologische Eigenschaften auf die zwischensprachlichen Distanzen wirken, besondere Bedeutung beigemessen wird.

Nr.	Sprache	
	De <sup>1</sup>	De <sup>2</sup>
1.	De <sup>1</sup>	De <sup>2</sup>
2.	Bg <sup>1</sup>	Bg <sup>2</sup>
3.	Be <sup>1</sup>	Be <sup>2</sup>
4.	Bs <sup>1</sup>	Bs <sup>2</sup>
5.	Cs <sup>1</sup>	Cs <sup>2</sup>
6.	Mo <sup>1</sup>	Mo <sup>2</sup>
7.	Hr <sup>1</sup>	Hr <sup>2</sup>
8.	Ls <sup>1</sup>	Ls <sup>2</sup>
9.	Mk <sup>1</sup>	Mk <sup>2</sup>
10.	Pl <sup>1</sup>	Pl <sup>2</sup>
11.	Ru <sup>1</sup>	Ru <sup>2</sup>
12.	Sk <sup>1</sup>	Sk <sup>2</sup>
13.	Sl <sup>1</sup>	Sl <sup>2</sup>
14.	Sr <sup>1</sup>	Sr <sup>2</sup>
15.	Uk <sup>1</sup>	Uk <sup>2</sup>

Tab. 2: Intrakorrelative Beziehungen

Im Zuge der Bearbeitung des Korpusmaterials werden folgende Arbeitsschritte durchgeführt: **(a)** Auswahl von einsprachigen Texten aus den beiden untersuchten Zeitabschnitten (1970–1990, 1991–2010) und Auffindung von Veränderungen, die Einfluss (Vergrößerung oder Verkleinerung) auf den Grad der Distanz zwischen den US nehmen bzw. nehmen könnten, **(b)** Bestimmung des Charakters dieser Veränderungen, des Grundes und der Art ihres Auftretens (im Sinne eines ökonomischeren und effizienteren sprachlichen Ausdrucks, spontan oder zielgerichtet, mit politischen Zielen, zur Vergrößerung der zwischensprachlichen Konvergenz u. Ä.) sowie Klärung der Frage, ob es sich bei diesen Veränderungen um relevante, zufällige, spontane, geplante, zielgerichtete oder andere Prozesse handelt.

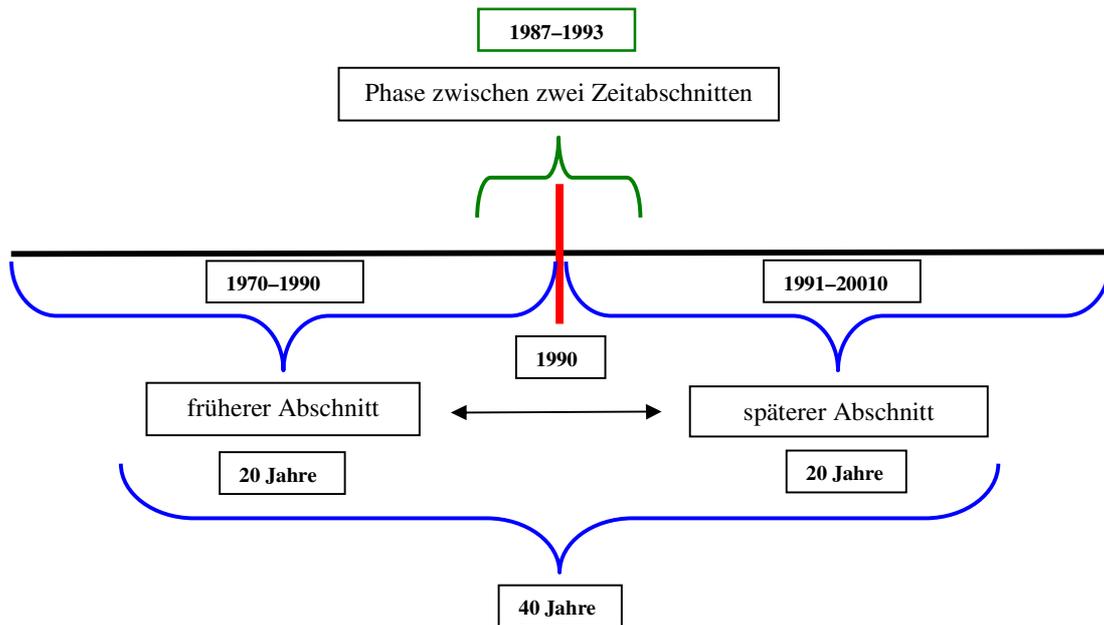


Abb. 2: Intra-korrelational

Im Rahmen des Intra-korrelational gibt es für jede Sprache eine zeitliche Zäsurlinie in Form eines Jahres, in dem ein für die Sprachfrage wesentliches Ereignis eintrat. Es ist die für Be, Ru und Uk der Zerfall der UdSSR (1989), für Cs und Sk das Auseinanderbrechen der ČSFR (1993), für Bs, Hr, Mo und Sr der Zerfall Jugoslawiens (1991) sowie für De die Wiedervereinigung Deutschlands (1990). Bei Nichtvorliegen derart markanter Einschnitte, werden andere Anhaltspunkte zur zeitlichen Trennung der analysierten Periode herangezogen (für Pl und Ls innerhalb der Gruppe der westslawischen Sprachen das Jahr 1993, für Bg im Rahmen der südslawischen Gruppe das Jahr 1992). Auf diese Weise werden die Jahre von 1989 bis 1993 zu einer zeitlichen Trennlinie für die Messung des Einflusses von Veränderungen in einer Sprache in Bezug auf die Distanz zu einer anderen. Zur Bestimmung der **intra-korrelationalen** Distanz (eingetreten durch Veränderungen in den slawischen Sprachen von 1970 bis 2010) werden Übersetzungen aus einer untersuchten Sprache angefertigt, die als Minimum über zwei Versionen verfügen – eine aus dem Zeitraum zwischen 1970 und 1990, die andere aus der Periode von 1991 bis 2010.

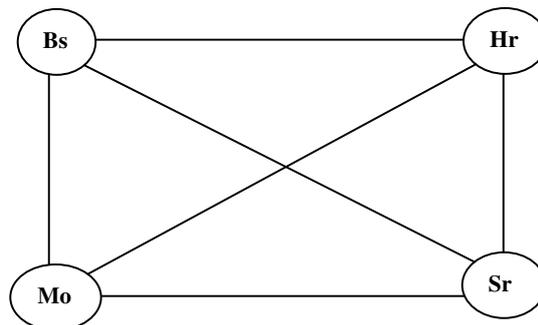
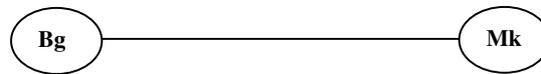
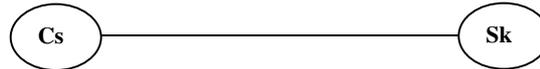


Abb. 3: Inter-korrelational A

Abb. 4: Interkorrelational **B**Abb. 5: Interkorrelational **C**

Das zweite Korrelationssystem, das Interkorrelational beschreibt das gegenseitige Verhältnis zweier sehr naher SSS, wie etwa **(a)** Bs, Hr, Mo, Sr, **(b)** Bg und Mk sowie **(c)** Cs und Sk. Die im Rahmen der ersten Gruppe bereits durchgeführten Forschungen (Tošović 2008a) zeugen davon, dass sich der Prozess der Divergenz nach dem Zerfall des ehemaligen Jugoslawien verstärkte, wodurch es zu einer Vergrößerung der Distanz zwischen Bs, Hr, Mo und Sr kam. Dies wurde und wird im Besonderen durch die radikalen puristischen Tendenzen in einigen Staaten, die Zuspitzung gewisser gesellschaftlicher Prozesse (Anwachsen des Nationalismus und Chauvinismus, Ausbruch bewaffneter Konflikte u. Ä.) und durch psychische Faktoren (z. B. Hass gegenüber anderen Völkern und Sprachen) verstärkt. Das Ziel dieser Untersuchungsebene liegt nicht nur in einer Erörterung der Frage des Abstandes zwischen den SSS innerhalb jedes einzelnen Interkorrelationals (a, b und c), sondern auch in der Erhebung der Distanz zwischen den drei genannten Gruppen (ob etwa die Distanz zwischen Hr und Sr größer ist als zwischen Bg und Mk oder zwischen Cs und Sk).

Das **Suprakorrelational** definiert sich als Verhältnis zwischen territorial nahen SSS. Es besteht aus dem östlichen – Be, Ru, Uk (**A**) –, dem westlichen – Cs, Ls, Pl, Sk (**B**) – und den südlichen Suprakorrelational – Bg, Bs, Hr, Mk, Mo, Sr (**C**). Das Ziel der Analysen im Rahmen des Suprakorrelationals liegt in der Bestimmung der Distanz zwischen denjenigen Sprachen, die sich innerhalb dieses Beziehungssystems befinden (A, B, C, so z. B. die Nähe bzw. Entfernung zwischen Ls, Pl, Sk und Cs). An dieser Stelle erhebt sich die Frage, ob sich Sprachen innerhalb eines bestimmten Suprakorrelationals näher stehen als Sprachen eines anderen Suprakorrelationals (z. B. Bg, Mk ↔ Cs, Sk; Cs, Pl ↔ Ru, Uk; Be, Ru ↔ Mk, Sl).

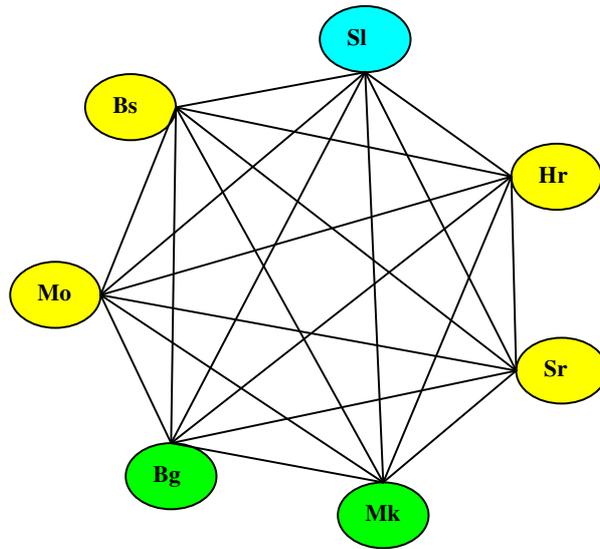


Abb. 6: Suprakorrelational A

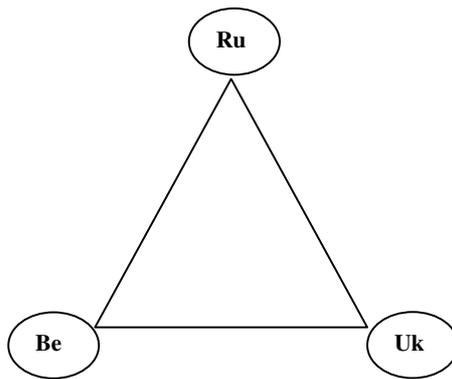


Abb. 7: Suprakorrelational B

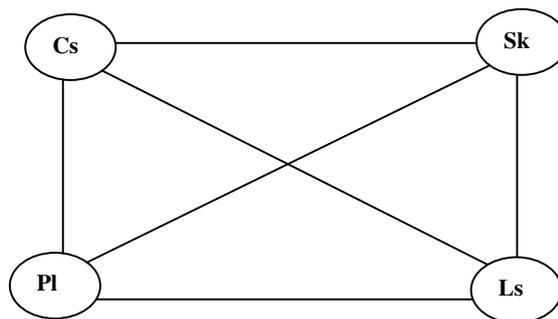


Abb. 8: Suprakorrelational C

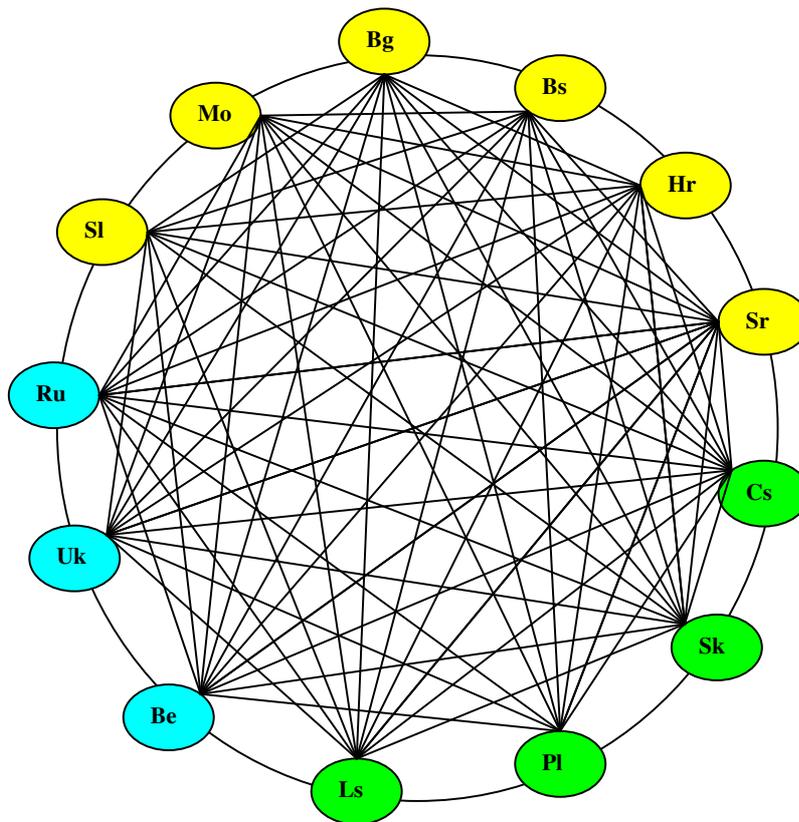


Abb. 9: Suprakorrelational

Innerhalb des **Superkorrelationals** befinden sich SSS, die unterschiedlichen Sprachgruppen (bzw. Arealen) angehören, nämlich 1. der ost-westslawischen: a) Be, Ru, Uk und b) Ls, Pl, Sk, Cs; 2. der ost-südslawischen: a) Be, Ru, Uk und b) Bg, Bs, Hr, Mk, Mo, Sl, Sr sowie 3. der west-südslawischen: a) Cs, Ls, Pl, Sk und b) Bg, Bs, Hr, Mk, Mo, Sl, Sr. Eine Besonderheit bei den Untersuchungen im Rahmen dieses Systems liegt dabei in der Frage, in welchem Maße interkorrelationale Veränderungen Einfluss auf die superkorrelationale Distanz nehmen. In diesem Zusammenhang erscheint es sinnvoll, die Hypothese, wonach eine Vergrößerung der interkorrelationalen Distanz den Charakter der superkorrelationalen Distanz beeinflusst, auf ihre Richtigkeit hin zu überprüfen. Die auf diese Weise vorgenommene Untersuchung kann Auskunft darüber erteilen, ob einige Prozesse im Falle von Hr, die auf eine bewusste Distanzierung von Sr abzielen (d. h. Vergrößerung der interkorrelationalen Distanz) zu einer Annäherung von Hr an Ru führen (d. h. Verringerung der superkorrelationalen Distanz).

Zur Bestimmung der **inter-**, **supra-** und **superkorrelationalen** Distanz werden Texte erarbeitet, die nach Möglichkeit in alle bzw. in eine größtmögliche Zahl der US übersetzt werden, wobei jede slawische Sprache nach so gut als möglich vertreten sein soll.

Ein besonderes Beziehungssystem gehen im Rahmen des **Extrakorrelationals** die SSS und die deutsche Sprache ein. Besonderes Augenmerk liegt hierbei auf der Distanz zwischen dem Deutschen und den slawischen Sprachen im Rahmen des Interkorrelationals (z. B. Bs, Hr, Mo, Sr), Suprakorrelationals (z. B. Cs, Ls, Pl, Sk) und des Superkorrelationals (z. B. Sl und Uk, Pl und Be, Mk und Cs). Dieser Teil der Untersuchungen verfolgt das Ziel einer Beantwortung der Frage, welche SSS die kleinste und welche die größte Distanz zum Deutschen aufweist. In diesem Zusammenhang gilt es die Annahme zu verifizieren, ob ein direkter territorialer Kontakt (d. h. eine

gemeinsame Grenze) auf eine Verringerung der Distanz Einfluss nimmt. Zur Bestimmung der extrakorrelationalen Distanz zwischen den US werden (a) slawische Texte mit Übersetzungen ins Deutsche (die zur Erhebung der inter-, supra- und superkorrelationalen Distanz nach Möglichkeit bereits in das Korpus aufgenommen wurden) und (b) Texte in deutscher Sprache herangezogen, die in die SSS übersetzt wurden.

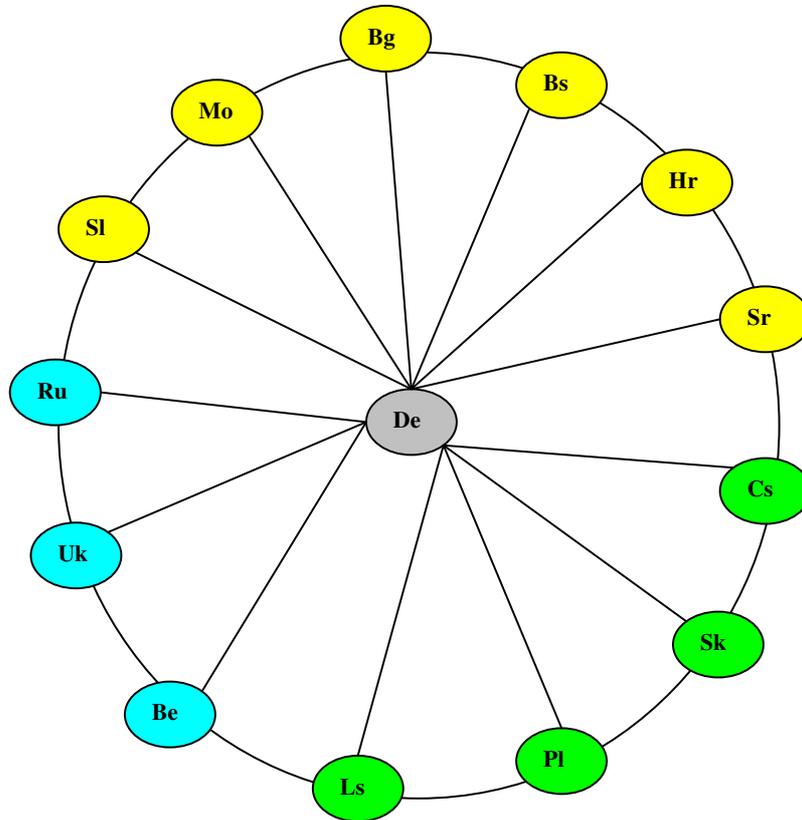


Abb. 10: Extrakorrelational

Zur Eruiierung der Distanz im Rahmen sämtlicher Korrelationssysteme werden Übersetzungen herangezogen, die in den Jahren zwischen 1970 und 2010 angefertigt wurden, wobei bei Vorliegen von zwei oder mehreren Übersetzungen zumindest eine aus dem untersuchten Zeitraum stammen muss. Im Zuge der Auswahl wird weiters berücksichtigt, welche Texte am häufigsten in die US übersetzt wurden. Für die Bestimmung der Distanz im literarisch-künstlerischen Stil werden Übersetzungen von Prosawerken herangezogen. Die Distanzanalyse im publizistischen Stil erfolgt mithilfe von Texten gleichen Inhalts aus Online-Portalen wie „Southeast European Times“, „Deutsche Welle“, „Voice of America“ u. a. Angesichts dessen, dass diese Texte nur für manche US vorliegen, werden für die fehlenden entsprechende Übersetzungen angefertigt. Als Material für die Erhebung der Distanz im Rahmen des wissenschaftlichen Stils dienen Übersetzungen streng wissenschaftlicher Publikationen, wobei als Kriterium hier auch die Zeit (die Übersetzungen müssen dem Zeitraum zwischen 1970 und 2010 entspringen) und die Anzahl der sprachlichen Versionen (je mehr, desto besser) zum Tragen kommt. Für die Distanzanalyse im administrativen Stil wird auf unterschiedliche Versionen grundlegender Dokumente internationaler Organisationen (UNO, EU, UNESCO u. a.) zurückgegriffen werden. Zur Untersuchung des umgangssprachlichen Stils ist die Entwicklung eines „SlawSpeech-Korpus“ mit Aufnahmen gesprochener Sprache geplant, das aus drei Teilen bestehen soll: einem Wort-, Fix- und einem Freikorpus, wobei die beiden erstgenannten Subkorpora für phonetische und prosodische Analysen

dienen. Das Wort-Korpus setzt sich aus Aufnahmen einzelner Wörter zusammen, die allen SSS gemeinsam sind. In das Fix-Korpus werden Audiofiles zusammenhängender Wörter und Sätze aufgenommen, die in allen SSS über die gleiche oder eine ähnliche Struktur verfügen. Diese Texte werden in der Regel kleineren Umfanges sein (nicht mehr als 20 Sätze). Das dritte Subkorpus, das Frei-Korpus, ist für eine Analyse der Distanz auf textueller und stilistischer Ebene vorgesehen und wird spontane Aussagen zu bestimmten Themen umfassen, wodurch eine von äußeren Faktoren (Zensur) unabhängige Messung der Distanz ermöglicht wird. So etwa werden ProbandInnen Zeichnungen oder Abfolgen von Bildern vorgelegt, gefolgt von der Bitte, das Betrachtete mit eigenen Worten zu beschreiben. Als elementarer Bestandteil des Speech-Korpus und zur Analyse der Distanz in prosodischer Hinsicht wird ein auf einer relationalen Datenbank fußendes Programm mit der Bezeichnung „Akzentarium“ dienen (zu dessen Struktur für Bs, Hr und Sr siehe Tošović 2008a: 770–776). Im Zuge der Analyse des mündlichen Sprachmaterials werden zwei Methoden zur Anwendung kommen, nämlich einerseits eine akustische (unter Zuhilfenahme von Programmen wie „Praat“ wird dabei nach grundlegenden Parametern wie etwa Länge der ausgesprochenen Laute, Änderungen in der Intonation u. a. die Distanz zwischen den US gemessen) und andererseits eine auditive.

**8.** Das Netz der slawischen Sprachen umfasst ein linguales (Gesamtkomplex der Sprachen) und ein strukturelles Megasystem (Komplex der sprachlichen Einheiten). Im ersten Fall handelt es sich um die Beziehungen zwischen den Sprachen (linguales Korrelational), im zweiten um die von deren Einheiten (strukturelles Korrelational).

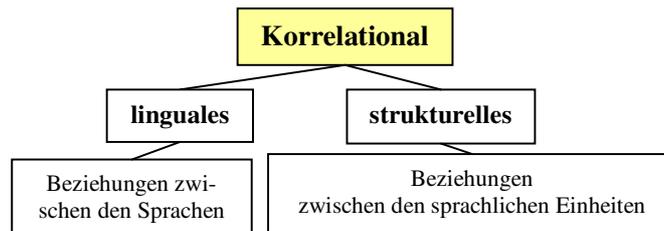


Abb. 11: Korrelational

**9.** Es gibt verschiedene Gliederungen der slawischen Sprachen, die vom Areal (ost-, süd- und westslawische Sprachen), von Besonderheiten im phonetisch-phonologischen System (vokalische und konsonantische Sprachen), von morphologischen Charakteristiken (analytische und synthetische Sprachen, Sprachen mit und ohne Artikel, Sprachen mit umfassender, teilweiser oder geschwundener Deklination u. a.) ausgehen. Für die vorliegende Analyse erweisen sich diejenigen Veränderungen in der einen oder anderen Sprache als relevant, die auf den Grad der wechselseitigen Nähe bzw. Entfernung Einfluss nehmen. Ausgehend davon werden im Rahmen des lingualen Korrelationalen fünf zwischensprachliche Systeme unterschieden: Intra-, Inter-, Supra-, Super- und Extrakorrelational.

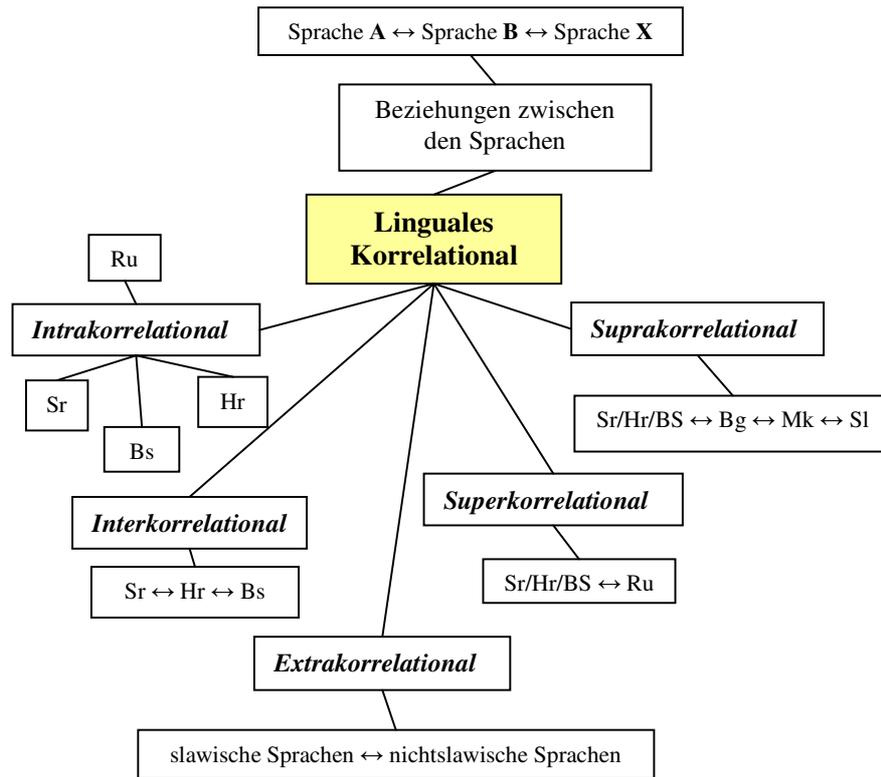


Abb. 12: Das linguale Korrelational

**10.** Zum Verbund des linguales Intrakorrelational gehören Beziehungen innerhalb einer isolierten slawischen Sprache, wobei im Besonderen die Beziehungen zwischen deren regionalen Varietäten zum Tragen kommen. An dieser Stelle stellt sich die wesentlicher Frage, wie sehr innere Veränderungen in einer Sprache (z. B. in Ru) auf die äußeren Beziehungen zwischen zwei oder mehreren Sprachen (etwa Ru und Sr) Einfluß ausüben. Das linguale Interkorrelational umfasst Verbindungen zwischen genetisch sehr nahe verwandten slawischen Sprachen, wie beispielsweise Sr, Hr und Bs. Zum linguales Suprakorrelational zählen Sprachen ein und desselben Areals (z. B. die süd- oder ostslawischen Sprachen). Suprakorrelative Sprachen wären etwa Bs, Hr und Sr einer- und Bg, Mk und Sl andererseits. Slawische Sprachen, die zu verschiedenen Arealen angehören, bilden das linguale Suprakorrelational. Hierzu zählen die Beziehungen zwischen Bs, Hr und Sr mit den ostslawischen Sprachen Be, Ru und Uk. Die Korrelationen zwischen slawischen Sprachen und nichtslawischen Sprachen sind Teil des linguales Extrakorrelational. All das, was zwischen den Sprachen gemeinsam ist, wird zum linguales Kategorial gezählt. Neben dem synchronen Aspekt korrelieren die slawischen Sprachen auch auf der diachronen Ebene, wobei derartige Beziehungen zum linguales Retrokorrelational zählen.

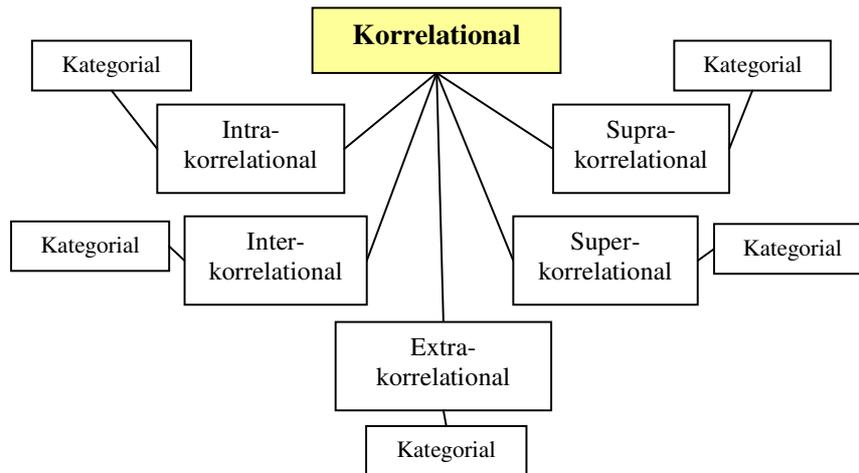


Abb. 13: Korrelational

**11.** Geht man davon aus, dass die Distanz zwischen interkorrelationalen Sprachen (wie etwa im Falle von Bs, Hr und Sr) minimal und zwischen superkorrelationalen (wie Bs, Hr und Sr auf der einen und Ru auf der anderen Seite) maximal ist, kann die Schlussfolgerung getroffen werden, dass sich interkorrelationale Sprachen wesentlich weniger voneinander unterscheiden als superkorrelationale.

**12.** Das strukturelle Korrelational umfasst Beziehungen formaler, semantischer, kategorialer und funktionaler Einheiten einer Sprache oder von korrelierenden Sprachen. So zum Beispiel werden auf der Ebene der Phonetik und Phonologie zwei grundlegende Systeme unterschieden – das System der Vokale und das System der Konsonanten. Diese bilden jeweilige intra-, inter-, supra-, super- und extrakorrelationale Komplexe.

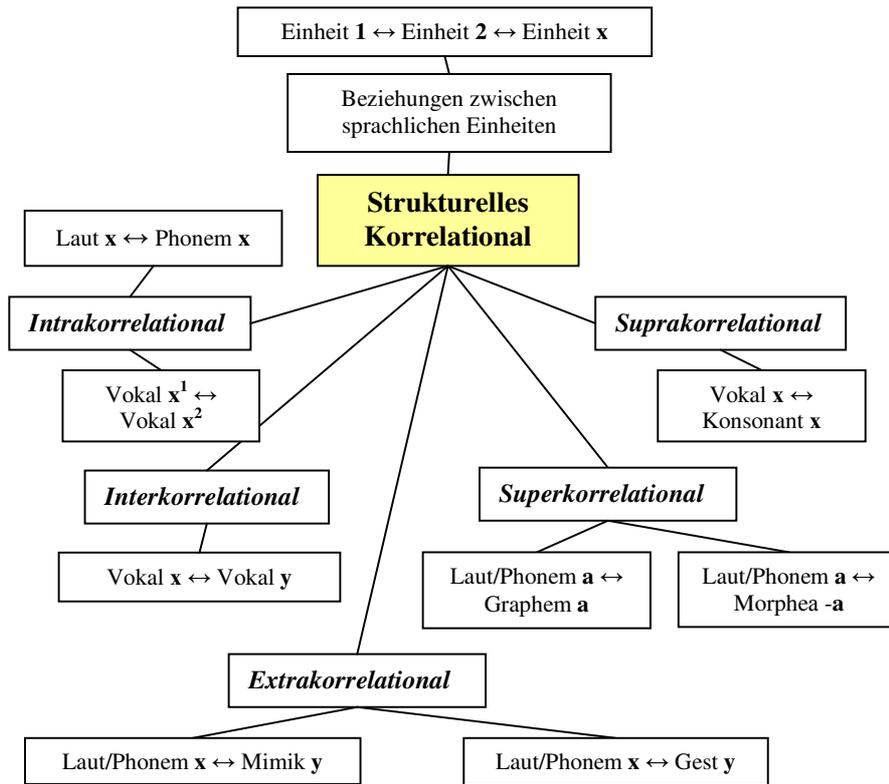


Abb. 14: Das strukturelle Korrelational

Das phonetisch-phonologische Intrakorrelational beinhaltet Beziehungen zwischen a) Lauten und Phonemen sowie b) Varianten (Allophonen) ein und desselben Lautes/Phonems. Weil es sich dabei um artikulatorische und funktionale Varianten desselben Lautes/Phonems handelt, umfasst ein solches System sämtliche variierenden Typen eines Lautes/Phonems (z. B.: offenes und geschlossenes e, betontes und unbetontes e, reduziertes und nichtreduziertes e, harter und weicher Zischlaut u. a.).

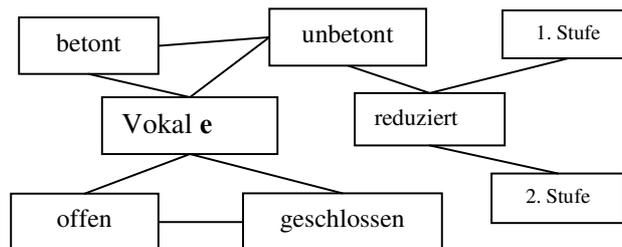


Abb. 15: Das strukturelle Intrakorrelational des Vokales e

Das phonetisch-phonologische Interkorrelational besteht aus Verbindungen zwischen Einheiten einer bestimmten Sprache, die sich auf ein und dieselbe phonetisch-phonologische Unterkategorie beziehen, so etwa im Vokal- ( $a \leftrightarrow e$ ,  $o \leftrightarrow u$ ...) oder im Konsonantensystem ( $s \leftrightarrow \check{s}$ ,  $t \leftrightarrow d$ ...).

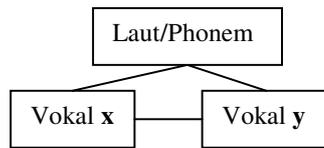


Abb. 16: Das strukturelle Interkorrelational

Zum phonetisch-phonologischen Suprakorrelational gehören Beziehungen zwischen Phonemen/Lauten und prosodischen Einheiten wie etwa Betonung, Intonation, Pause, Rhythmus u. a.

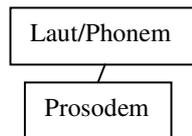


Abb. 17: Das strukturelle Suprakorrelational

Das phonetisch-phonologische Superkorrelational besteht aus Verhältnissen zwischen Einheiten auf unterschiedlichen Ebenen, wie z. B. auf der phonetischen, phonologischen, morphologischen u. a., so etwa 1. Phonem und Morphem (im Besonderen morphologische Alternationen), die einen Gegenstand der Morphologie darstellen, und 2. Laut/Phonem und Graphem.

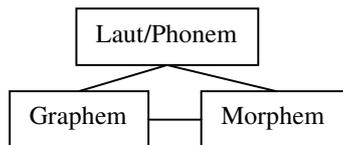


Abb. 18: Das strukturelle Superkorrelational

Das phonetisch-phonologische Extrakorrelational wird von sprachlichen Einheiten gebildet, die образуют unterschiedlichen Hyperkategorien angehören. Dabei geht es um Beziehungen folgender Natur: phonetisch-phonologische Einheit  $\leftrightarrow$  nicht phonetisch-phonologische Einheiten (morphologische, lexikalische, derivative, syntaktische u. a.).

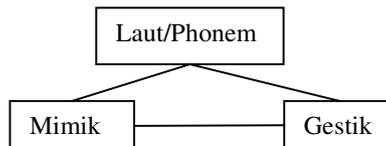


Abb. 19: Das strukturelle Extrakorrelational

**13.** Innerhalb des phonologischen Systems der slawischen Sprachen trifft man auf universale, spezifische und exklusive Phoneme. Als universale werden dabei diejenigen Phoneme bezeichnet, die in allen slawischen Sprachen angetroffen werden können. Spezifische Phoneme sind nur in manchen Sprachen vorhanden, während exklusive nur in einer Sprache vorkommen.

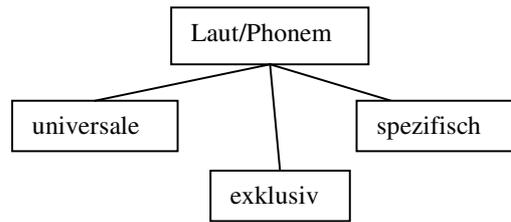


Abb. 20: Slawische Laute/Phoneme

**14. Forschungslinien.** Die Analyse der Distanz zwischen den SSS kann im Rahmen dreier Forschungslinien erfolgt: einer korpuslinguistischen, einer soziolinguistischen und einer psycholinguistischen.

Die korpuslinguistische Linie verfolgt das Ziel, (a) für die Distanzanalyse erforderliches Material zu sichten und zu bearbeiten, (b) die aufbereiteten Texte in das online verfügbare und zugängliche „Slawische Korpus“ (kurz: Slaw-Kor) zu integrieren und (c) mithilfe des Korpus Untersuchungen zur zwischensprachlichen Distanz durchzuführen. Jedes der oben genannten Korrelationssysteme wird über ein eigenes Subkorpus verfügen: das Intra-Kor (für nur eine Sprache), Inter-Kor (für sehr nahe verwandte SSS), Supra-Kor (für Sprachen, die einem Areal der Slawia angehören), Super-Kor (für Sprachen unterschiedlicher Areale) und das Extra-Kor (für die SSS und das Deutsche). Einsprachig wird einzig das erste Subkorpus (Intra-Kor) sein, wobei jedoch auch dieses Korpus aus Paralleltexten besteht, die die Entwicklung der Sprache innerhalb des untersuchten Zeitraumes zeigen. Sämtliche übrigen Subkorpora werden im Hinblick auf ihren Inhalt vereinheitlicht und funktionell angeglichen, um dadurch anhand des gleichen Materials die Distanz in allen oben angeführten Beziehungssystemen (inter-, supra-, super- und extra-) und im Rahmen aller elementarer Varianten analysieren zu können. Als Online-Plattform wird die Website Gralis – Das linguistische Slawistik-Portal der Karl-Franzens-Universität Graz (Gralis-www) – dienen, das bereits über einen Parallelkorpus mit der Bezeichnung „Gralis-Korpus“ für Bs, Hr und Sr verfügt, das im Rahmen des FWF-Projektes P19158-G03 erstellt wurde, wobei die geschaffene Infrastruktur auch für die Entwicklung des multilingualen Korpus für alle SSS herangezogen wird. In dieser Forschungslinie werden zwei Teilbereiche unterschieden: a) die Erstellung, Befüllung und Weiterentwicklung des Korpus und b) die linguistische, soziolinguistische und psycholinguistische Korpusanalyse.

Die Linie zur Entwicklung des Korpus umfasst die Sammlung von Material, dessen Bearbeitung und Annotierung, wobei die Hauptaufgabe in der Schaffung eines Parallelkorpus für alle US liegt.

Gegenwärtig gibt es eine nur kleine Zahl an Parallelkorpora für die SSS. Eines dieser Korpora stellt in Anbetracht seiner allgemeinen Struktur und seines Umfangs das Gralis-Korpus dar (Gralis-Korpus-www), das für eine Analyse der Sprachen Bs, Hr und Sr dient und im Rahmen des FWF-Projektes P19158-G03 (2006–2009) entwickelt wurde (Tošović 2008a, 2008b). Dieses Korpus besteht einerseits aus Texten (Text-Korpus) und andererseits aus einem Korpus der gesprochenen Sprache (Speech-Korpus). In das Text-Korpus (Wonisch 2008b) wurden Texte aller funktionalen Stile (literarisch-künstlerisch, publizistisch, wissenschaftlich und administrativ – siehe dazu Tošović 2002) aufgenommen. Mit dem Stand von Februar 2009 umfasst dieses Korpus rund drei Millionen Tokens, wobei die Texte mit einer grundlegenden metasprachlichen und grammatikalischen Annotierung versehen wurden. Das Speech-Korpus gliedert sich in drei Subkorpora, nämlich das Wort-, Fix- und Frei-Korpus (Forić 2008, Wonisch/Just 2008). Das Wort-Korpus stellt eine Sammlung isoliert ausgesprochener Wörter dar; das Fix-Korpus setzt sich aus Audioaufnahmen kürzerer Texte zusammen (z. B. der Text „Jutro“ mit 18 Sätzen), die keinerlei lexikalischen und grammatikalischen Unterschiede zwischen Bs, Hr und Sr aufweisen, während das Frei-Korpus etwa 300 Aufnahmen frei und spontan gesprochener Sprache beinhaltet. Damit

im Zusammenhang verdient das vergleichbare „Russisch-slowakische Parallelkorpus“ Erwähnung.

Im Rahmen der russisch-nichtslawischen Parallelkorpora gilt es das „Englisch-russische Parallelkorpus“ im Verband des „Nationalkorpus der russischen Sprache“ (Ruscorpora) zu erwähnen, das sich gegenwärtig in der Frühphase seiner Entwicklung befindet. In dieses Korpus wurden in kleinerem Umfang literarische Werke im Original und in ihrer Übersetzung aufgenommen. Im Rahmen eines Projektes mit dem Titel „Opus“ wurde eine auf der IMS Corpus Workbench basierende Sammlung frei verfügbarer Paralleltex-te angelegt (technische Dokumentationen, Korpus an Subtiteln). An dieser Stelle zu nennen wären auch das mehrsprachige Korpus mit Übersetzungen des altrussischen Textes „Das Lied vom Heerzug Igors („Slova o polku Igoreve“ – SPI-www) und „Lilabar“, ein englisch-russisches Korpus paralleler Sätze mit 8500 Sprichwörtern und 130.000 Phrasen (Lilabar-www). An der Abteilung für Translationsstudien der Universität Tampere (Finnland) wurde ein russisch-finnisches Parallelkorpus künstlerischer Texte mit dem Namen „ParRus“ entwickelt. Einen größeren Umfang (für jede Sprache 1,5 Millionen Tokens) als dieses Korpus weist das russisch-englische Korpus auf, in das Werke der russischen Literatur aus dem 19. und 20. Jahrhundert im Original und ihrer englischen Übersetzung aufgenommen wurden.

Ein weiteres Parallelkorpus, das einerseits aus Texten in deutscher und englischer Sprache und andererseits aus Texten in einigen slawischen Sprachen (Bg, Be, Bs, Cs, Hr, Pl, Ru, Sk, Sr und Uk) besteht, ist am Institut für Slawistik der Universität Regensburg beheimatet und wurde mit der Bezeichnung „The Regensburg Parallel Corpus (RPC)“ versehen (RPC-www).

Im der Phase seiner Entwicklung befindet sich ein russisch-deutsches und deutsch-russisches Korpus im Rahmen des „Nationalkorpus der russischen Sprache“, wie auch ein russisch-deutsches Korpus paralleler Texte innerhalb des „Austrian Academy Corpus der Österreichischen Akademie der Wissenschaften“. Letztgenanntes enthält bloß einen einzigen Roman – „Der Idiot“ von F. M. Dostoevskij (entstanden in den Jahren 1868–1869) – zusammen und eine von dessen Übersetzungen ins Deutsche. In Vorbereitung befindet sich weiters ein deutsch-russisches Korpus mit der Bezeichnung „Traumdeutung“, bestehend aus dem gleichnamigen Text Sigmund Freuds (Traumdeutung-www).

Von den Parallelkorpora ohne slawische Sprachen gilt es in erster Linie das „Europarl Parallel Corpus“ (EPC-www) zu erwähnen, in dessen Rahmen Subkorpora für folgende Sprachpaare angelegt wurden: Danish-English, German-English, Greek-English, Spanish-English, Finnish-English, French-English, Italian-English, Dutch-English, Portuguese-English, Swedish-English. Am Zentrum für Translationsstudien der Universität Leeds (Großbritannien) wurde das „Leeds Corpus“ (Leeds-www) Corpus entwickelt, das sich aus mehreren Sprachen zusammensetzt (chinesisch, deutsch, englisch, französisch, italienisch, japanisch, russisch und spanisch). Ein an der Universität Augsburg erarbeitetes und mit dem Namen „MAASTR“ versehenes Parallelkorpus (Maastr-www) enthält den Maastrichter Vertrag in deutscher und englischer Version. Im Zuge des deutsch-französischen Projektes „Kollokationen im Kontext“ kam es zur Entwicklung eines Korpus, das Texte in deutscher Sprache und deren französische Übersetzung sowie französische Texte mit deutscher Übersetzung umfasst. Der Aufzählung hinzugefügt werden kann abschließend das offen zugängliche Online-Korpus mit dem Titel „Parallel Corpus of Portuguese and English“ (COMPARA – Compara-www).

Die Bearbeitung des Materials erfolgt in drei Etappen. Im Rahmen der ersten werden Texte für jede SSS vorbereitet, indem als wesentliches Element zusätzlich eine metasprachliche, lexikalisch-semantiche und grammatikalische Annotierung vorgenommen wird. Die metasprachliche Annotierung setzt sich aus der Ausweisung der Quelle (Autorin/Autor, Textkapitel, Ort und Jahr der Entstehung, Seitenzahl, Angaben zum Verlag, zur Übersetzung u. a.). Die lexikalisch-

semantische Annotierung stellt die wesentlichen lexikalischen und semantischen Charakteristiken von Wörtern dar. Aus der grammatikalischen Annotierung geht schließlich hervor, welche morphologische Struktur und welcher Typ von Valenzen zu anderen Wörtern auf der Ebene von Syntagmen und Sätzen bestehen. Angesichts dessen, dass die Untersuchungen (zur Auffindung von Informationen in allen sprachlichen Versionen der Texte) einer gänzlich vereinheitlichten Annotierung bedürfen, muss diese in höchstmöglichem Maße an die Besonderheiten der US angepasst werden. Die Erfahrung und die bis dato durchgeführten Arbeitsschritte bei der Vorbereitung des Parallelkorpus für Bs, Hr und Sr (Gralis-Korpus) zeugen davon, dass sich für diese Zwecke der von einer Gruppe von WissenschaftlerInnen mit Tomaž Erjavec an der Spitze im Jahre 2004 ausgearbeitete Standard „Multext-East“ (Multilingual Texts and Corpora for Eastern and Central European Languages – multilingual dataset for language engineering research and development: MultiText East-www, Erjavec 2004, 2006) am besten eignet. Im Rahmen von Multext-East wurde ein System einer morpho-syntaktischen Annotierung für (nur) einige slawische Sprachen geschaffen (Bg, Cs, Hr, Ru, Sl und Sr), sodass dieses auf alle SSS ausgeweitet wird (dies wird Aufgabe von Tomaž Erjavec sein, der als internationaler Kooperationspartner fungieren wird).

Nach Vornahme der Annotierung beginnt die zweite Etappe, in der die Texte in Sätze untergliedert werden, indem als Ergebnis ein System geschaffen wird, in dem jedem Satz der Sprache **A** ein Satz in den Sprachen **B, C ...** entspricht. Sollte dabei ein Absatz in einer Sprache etwa aus drei Sätzen bestehen und in einer anderen aus fünf, wird es erforderlich sein, dieses Missverhältnis auszugleichen. Im Sinne einer Automatisierung dieses Prozesses werden (a) bereits bestehende oder aus anderen Quellen bekannte Vorgehensweisen übernommen und adaptiert und/oder (b) neue Programme entwickelt.

Die dritte Etappe besteht darin, aus den Korpus-Texten Listen der sprachlichen Einheiten zu extrahieren, die sodann unter Nutzung des Programms „MySQL“ in relationale Datenbanken umgewandelt werden, sodass als Ergebnis Wörterbücher unterschiedlichen Typs angefertigt werden können.

Für die Generierung des Korpus kommen als Software die „IMS Corpus Workbench (CWB)“ und als webbasierter Workflow zur Informationsmodellierung „Asset Management“-Systeme (AMS) zum Einsatz. Bei der IMS Corpus Workbench handelt es sich um ein multifunktionales Tool für die Verwaltung, Aufbereitung und Durchführung der Suche in großen Textkorpora mit linguistischer Annotierung. Die Hauptkomponente dieser Workbench bildet das userbasierte und umfassende Suchprogramm CWB (Corpus Query Processor), das folgende Elemente beinhaltet: Instrumente zur Kodierung, Indexierung, Kontraktion, Dekodierung und Darstellung von Frequenzen, ein Gesamtregister, in dem die Informationen über das Korpus gespeichert werden (Name, Attribute, Standort), sowie das Suchprogramm CQP mit seiner Syntax von regular expressions.

Im Zuge der Erstellung des Korpus müssen neben der Vorbereitung der Texte und der Annotierung der Metadaten im XML-Format auch Datenmaterialien in Audiodateiformaten (mp3, wav u. a.) gesammelt werden, die zur Messung der Distanz auf phonetischer und prosodischer Ebene herangezogen werden. In Bezug auf die konkrete Anwendung der Asset Management-Systeme kann festgehalten werden, dass diese den aufbereiteten Datenmaterialien eine zentrale IT-Struktur verleihen und damit deren nachhaltige und umfangreich modulierbare Archivierung ermöglichen. Als zentrales Tool für den Materialtransfer, die Verwaltung und Sicherung der Webressourcen dient das Open Source-Projekt „Fedora“ (Flexible Extensible Digital Object Repository Architecture).

Die Forschungslinie zur *A n a l y s e* *d e s* *K o r p u s* verfolgt das Ziel einer umfassenden Auswertung des gesammelten Korpusmaterials, bei der die elementaren Charakteristiken

der Distanz (phonetisch-phonologische, orthographische, lexikalisch-semantische und grammatikalische) in allen funktionalen Stilen analysiert werden.

Bei der Auswertung des Korpus werden nicht nur rein linguistische Fragen und Aspekte der Distanz erörtert, sondern darüber hinaus auch etwa der Einfluss von Sprachpolitik, Standardisierung und Kodifizierung und deren Einwirkung auf eine Vergrößerung oder Verringerung der Distanz zwischen den US. Besonderes Augenmerk wird dabei dem Purismus und Lehnwörtern beigemessen werden. Als zusätzliches Hilfsmittel zur Sammlung und zum Studium von Sprachdaten wird ein online verfügbares Programm zur Erstellung, Durchführung und Bearbeitung von Umfragen mit der Bezeichnung „Gralis-Anketarium“ (entwickelt im Rahmen des FWF-Projektes P19158-G03) – Thomann 2008 und das Programm „Gralis-Präskriptarium“ zum Vergleich der orthographischen Normen in den US dienen, dessen Grundkonstruktion bereits ausgearbeitet wurde (Tošović 2008: 822–825).

Auf Basis des schriftlichen wie mündlichen und im Korpus versammelten Sprachmaterials erfolgt eine Erörterung der Frage, wie hoch der Grad der Verständigungsmöglichkeit zwischen TrägerInnen unterschiedlicher SSS und die Distanz zwischen den US beziffert werden können. Im Zentrum dieser Analyse steht das Kriterium des gegenseitigen Verstehens, wie auch der Einfluss der Distanz auf Sprachwechsel im Sinne eines Code-Switching. In diesem Zusammenhang kommt nicht nur dem reinen Verstehen, sondern auch der Ebene (dem Grad) des Verstehens auf einer Skala eine wesentliche Rolle zu.

**15.** Die Analyse der Distanz zwischen den SSS auf unterschiedlichen Ebenen erfordert eine Anwendung verschiedener Untersuchungsmethoden. Die Messung der phonetisch-phonologischen Distanz erfolgt auf der Grundlage von Phonemlisten und -datenbanken, die aus dem Korpus der gesprochenen Sprache extrahiert werden. Im Zuge der Untersuchung werden dabei auch die bei einer Analyse der phonetischen Distanz zwischen verschiedenen Sprachen und Dialekten gewonnenen Ergebnisse berücksichtigt (z. B. Wildgen 1977, Nerbonne/Hinrichs 2006, Nerbonne/Heeringa 1997, Tambovtsev 2002). Im Hinblick auf die Ebene der Lexik wird vor allem die semantische Distanz (Messung von Bedeutungen, der semantischen Ähnlichkeit) im Mittelpunkt stehen, die gewöhnlich einzig im Rahmen einer Sprache untersucht wird (siehe dazu: Osgood et al. 1957, Wildgen 1977). Zur Bestimmung der semantischen Distanz zwischen zwei oder mehreren Sprachen (Messung der gefühlten Beurteilung des Inhalts einer lexikalischen Einheit) kann das semantische Differential von Osgood herangezogen werden (Osgood et al. 1957), was auch in einer Analyse von Danko Šipka bestätigt wird, der diese Methode zur Bestimmung der Unterschiede zwischen Hr und Sr bezüglich der Klassifizierung von Wörtern als „eigene“ und „fremde“ anwandte (Šipka 2008). Für die Durchführung der grammatikalischen Analyse wird man sich der grammatikalischen Annotierung von Texten bedienen, auf Grundlage derer man automatisch Listen aller Tokens erhält, durch die eine Bestimmung der zwischensprachlichen Distanz ermöglicht wird. Dazu wird aus dem Korpusmaterial anfangs eine Datenbank angefertigt, die aus den Tokens jeder einzelnen Sprache besteht und als Basis für eine paradigmatische und syntagmatische Analyse dient. Während im Rahmen der paradigmatischen Analyse die Frequenz einzelner Tokens eruiert wird, verfolgt die kontextuale Analyse das Ziel einer Bestimmung der Distanz auf der Ebene von Sätzen zweier oder mehrerer Sprachen, wobei die Erfahrungen vorangegangener Untersuchungen berücksichtigt wird (im Besonderen Nerbonne/Hinrichs 2006, Nerbonne/Wiersma 2006).

Im Zuge der orthographischen Analyse werden die aktuellen Rechtschreibregeln der US in einer Datenbank (Präskriptarium) erfasst. Über ein nach Sprachen, Regeln und Keywords gegliedertes Interface können in diesem Programm Suchabfragen nach bestimmten Regeln (z. B. Getrennt- und Zusammenschreibung) oder Keywords (z. B. Beistrich) in allen Rechtschreibungen durchgeführt werden. Wenn es in einer Sprache zwei oder mehrere Rechtschreibwerke gibt, wird dasjenige ausgewählt, das für die durchzuführende Untersuchung geeigneter erscheint. Für die

sozio- und psycholinguistische Analyse werden die metasprachliche Annotierung des Korpus und Online-Umfragen herangezogen.

### L i t e r a t u r

- Ammon 1987: Ammon, Ulrich. Language – Variety/Standard Variety – Dialect. In: Ammon, Ulrich et al. (ed.). *Sociolinguistics. An International Handbook of the Science of Language and Society*. Bd. 1. Berlin – New York. S. 316–335.
- Ammon 2005: Ammon, Ulrich. Pluricentric and Divided Languages. In: Ammon, Ulrich et al. (ed.). *Sociolinguistics*, Vol. 2, Berlin/New York. S. 1536–1543.
- Арапов/Cherc1974: Арапов, М.В.; Херц М.М. Математические методы в исторической лингвистике. Moskau.
- Bērziņš 2004a: Berzinch, A.A. La comparaison de typologie traditionnelle et de typologie phonolexique, basée sur la méthode des n-grammes, dans les dialectes baltes. In: *Identification des langues et des variétés dialectales par les humains et par les machines*. Paris. S. 103–104.
- Bērziņš 2004b: Берзиньш, А.А. Сравнение балтийских языков методом n-грамм. In: *Труды международной конференции „Корпусная лингвистика“ – 2004*. Sankt Peterburg. S. 65–71.
- Bērziņš/Grigorjevs 2007: Берзиньш, А.А.; Grigorjevs J. Latviešu izloksnēs sastopamo fonēmu telpa. In: *Iesniegts publicēšanai Linguistica Lettica 2007. gadā*. Rīga – <http://ansis.lv/raksti/endz2007.pdf>
- Bosák 1998: Bosák, Jan. *Slovenský jazyk*. Opole.
- Cavnar/Trenkle 1994: Cavnar, William B.; Trenkle, John M. Ngram-based text categorization. In: *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval, Las Vegas*. S. 161–175.
- Cavnar/Vayda 1992: Cavnar, William B.; Vayda, Alan J. Using superimposed coding of N-gram lists for Efficient Inexact Matching. In: *Proceedings of the Fifth USPS Advanced Technology Conference, Washington D.C.*
- Cavnar/Vayda 1993: Cavnar, William B.; Vayda, Alan J. Ngram-based matching for multi-field database access in postal applications. In: *Proceedings of the 1993 Symposium on Document Analysis and Information Retrieval, University of Nevada, Las Vegas*.
- Dmitrijeva-www: Дмитриева, Ю. В. Проблемы двуязычия и интерференции. In: <http://nakhodka.wl.dvgu.ru/forum/section7/7-08.htm>
- Dimitrova 1997: Dimitrova, Stefana (ed.). *Български език*. Opole.
- Duličenko 1981: Дуличенко, А.Д. Славянские литературные микроязыки: Вопросы формирования и развития. Tallinn.
- Dyen/Kruskal/Black 1992: Dyen, Isidore; Kruskal, Joseph B.; Black, Paul. An Indo-European Classification: a Lexicostatistical Experiment. In: *Transactions of the American Philosophical Society*. Nr. 82 (5).
- Erjavec 2004: Erjavec, Tomaž. Multext-East Version 3: multilingual morphosyntactic specification, lexicons and corpora. In: Lino, Maria Teresa; Xavier, Maria Francisca (ed.). *Fourth international conference on language resources and evaluation, Lisbon, 26th*,

- 27th & 28th May 2004. Proceedings: held in memory of Antonio Zampolli. Paris. S. 1535–1538.
- Erjavec 2006: Erjavec, Tomaž. Multitext-East Morphosyntactic specifications and XML. In: Slavcheva, Milena; Simov, Kiril, Angelova, Galia. Readings in multilinguality: selected papers for young researchers. Sofia. S. 41-48.
- Faska 1998: Faska, Helmut. Serbščina. Opole.
- Forić 2008: Forić, Sandra. Das Gralis Speech-Korpus. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen – Münster et al. S. 755–764.
- Gajda 1998: Gajda, Stanisław. Język polski. Opole.
- Gajda 2000: Gajda, Stanisław (ed.). Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Opole.
- Gladrow 1989: Gladrow, Wolfgang. Russisch im Spiegel des Deutschen. Leipzig.
- Ginsburgh/Ortuño-Ortin/Weber 2007: Ginsburgh, Victor; Ortuño-Ortin, Ignacio; Weber, Shlomo. Why do People Learn Foreign Languages? In: Journal of Economic Behavior and Organizations. Nr. 64 (3). S. 337–347.
- Gladková/Likomanova 2002: Гладкова, Хана; Ликоманова, Искра. Языковая ситуация: истоки и перспективы (болгарско-чешские параллели). Prag.
- Gooskens/Heeringa 2004: Gooskens, Charlotte; Heeringa, Wilbert. Perceptual evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. In: Language Variation and Change, 16(3). S. 189–207.
- Gutschmidt 2002: Gutschmidt, Karl. Möglichkeiten und Grenzen der Standardisierung slavischer Schriftsprachen der Gegenwart. Dresden.
- Heeringa 2004: Heeringa, Wilbert J. Measuring Dialect Pronunciation Differences Using Levenshtein Distance. Rijksuniversiteit Groningen. Univ.-Dissertation.
- Hinrichs 1999: Hinrichs, Uwe (ed.). Handbuch der Südosteuropa-Linguistik: Handbuch der Südosteuropa-Linguistik. Harrassowitz Verlag, Wiesbaden.
- Hinrichs/Gerdemann/Nerbonne-www: Hinrichs, Erhard; Gerdemann, Dale; Nerbonne, John. Measuring linguistic unity and diversity in Europe. In: <http://www.sfs.uni-tuebingen.de/dialectometry/docs/VW-dialect-proposal.pdf>. 31. 1. 2008
- Ivić 1998: Ivić, Pavle. Rasprave, studije članci: 1. O fonologiji. Sremski Karlovci – Novi Sad.
- Jachontov-www: Яхонтов, С.Е. Оценка степени близости родственных языков. In: Теоретические основы классификации языков мира. Москва 1980. S. 148–157. Zit. n: [www.philology.ru/linguistics1/yakhontov-80.htm](http://www.philology.ru/linguistics1/yakhontov-80.htm)
- Jermolenko 1999: Jermolenko, Svitlana. Українська мова. Opole.
- Koch/Oesterreicher 1985: Koch, Peter; Oesterreicher, Wulf. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Romanistisches Jahrbuch 36. S. 15–43.
- Kofler/Wonisch 2008: Kofler, Stefan; Wonisch, Arno. Das Gralis-Rezensarium. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 803–806.

- Kondrak 2005: Kondrak, Grzegorz. N-gram similarity and distance. In: Proceedings of the Twelfth International Conference on String Processing and Information Retrieval (SPIRE 2005), Buenos Aires. S. 115–126.
- Kořenský 1998: Kořenský, Jan (ed.). Český jazyk. Opole.
- Kromer 2004: Кромер, В.В. Глоттохронологическая ретрогностика языковой системы. In: Проблемы лингвистической прогностики. Воронеж. S. 136–144.
- Kromer 2005: Кромер, В.В. Об одном методе оценивания степени смешанности языков. In: Актуальные проблемы компьютерной лингвистики. Минск. S. 104–110.
- Kunzmann-Müller 2000: Kunzmann-Müller, Barbara et al. (ed.). Die Sprachen Südosteuropas heute. Umbrüche und Aufbruch. Frankfurt am Main.
- Laškova 1996: Laškova, Lili. On the Phenomenon of Slavic Languages in the Balkans. – In: Linguistique Balkanique. Sofija. 38/3. Pp. 231–237.
- Lehner 2008: Lehner, Olga. Die technische Entwicklung des Gralis Speech-Korpus. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 777–779.
- Levenshtein 1965: Levenshtein, Vladimir I. Binary codes capable of correcting spurious insertions and deletions of ones, Problems of Information Transmission 1(1). S. 8–17.
- Ljubešić/Nives/Boras-www: Ljubešić, Nikola; Mikelić, Nives; Boras, Damir. Language identification: how to distinguish similar languages? In: [http://infoz.ffzg.hr/ljubestic/nlnmdb\\_iti07.pdf](http://infoz.ffzg.hr/ljubestic/nlnmdb_iti07.pdf)
- Lončarić 1998: Lončarić, Mijo (ed.). Hrvatski jezik. Opole.
- Lukašanec et al. 1998: Lukašanec, Aljaksandr. Беларуская мова. Opole.
- Mackey 1971: Mackey, William F. La distance interlinguistique. Quebec: Les Presses de l' Université Laval.
- Magocsi 2004: Magocsi, Paul Robert. Русинський язык. Opole.
- Marti 2000: Marti, Roland. Slavische Standardsprachen im Kontakt. Das Neben-, Mit- und Gegeneinander slavischer Standardsprachen. In: Zybatow, L.N. (ed.). Sprachwandel in der Slavia. Teil 2. S. 527–541.
- Minova-Đurkova 1998: Minova-Đurkova, Liljana (ed.). Македонски јазик. Opole.
- Mokienko/Walter 2008: Mokienko, Valerij; Walter, Harry (ed.). Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Bd. 3: Frazeologia. Opole.
- Multext-East-www: Multext East. In: <http://nl.ijs.si/ME/V3/>
- Nerbonne/Heeringa 1997: Nerbonne, John; Heeringa, Wilbert. Measuring dialect distance phonetically. In Proceedings of SIGPHON-97: 3rd Meeting of the ACL Special Interest Group in Computational Phonology, Madrid. S. 11–18.
- Nerbonne/Hinrichs 2006: Nerbonne, John; Hinrichs, Erhard. Linguistic Distances. In: Nerbonne, John; Hinrichs, Erhard (ed.). Linguistic Distances Workshop at the joint conference of International Committee on Computational Linguistics and the Association for Computational Linguistics, Sydney, July, 2006. 1–6.
- Nerbonne/Wiersma 2006: Nerbonne, John; Wiersma, Wybo. Measure of Aggregate Syntactic Distance. In: Nerbonne, John; Wiersma, Wybo (ed.). Linguistic Distances Workshop at

the joint conference of International Committee on Computational Linguistics and the Association for Computational Linguistics, Sydney. S. 82–90.

- Neščimenko 2003: Нещименко, Г.П. Языковая ситуация в славянских странах. Опыт описания. Анализ концепций. Moskau.
- Nikitevič 2003: Никитевич, А.В. К сопоставлению деривационных подсистем глагола в славянских языках. In: Мовознаўства. Літаратура. Культуралогія. Фалькларыстыка. XIII Міжнар. з'езд славістаў (Любляна, 2003). Minsk. S. 144–158.
- Ohnheiser 2003: Ohnheiser, Ingeborg (ed.). Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Bd. 1: Słowotwórstwo/Nominacja. Opole.
- Osgood et al 1957: Osgood, Charles E. et al. The measurement of meaning. Urbana: University of Illinois Press.
- Potemkin-www: Потемкин, С.Б. Лексическая база данных с наложенной семантической метрикой. In: <http://www.philol.msu.ru/~rlc2004/files/sec/19.doc>
- Radovanović 1996: Radovanović, Milorad (ed.). Српски језик. Opole.
- Revzina 1970: Ревзина, О.Г. Типологический анализ грамматической категории рода. (На материале славянских языков). Moskau, 1970: Univ.-Dissertation (Zusammenfassung, AKD).
- Sawicka 2007: Sawicka, Irena (ed.). Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Bd. 2: Fonetyka/Fonologia. Opole.
- Šipka 2008: Šipka, Danko. Varijantske razlike u semantičkom diferencijalu. In: Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 130–142.
- Širjaev 1997: Širjaev, Evgenij. Русский язык. Opole.
- Stecjuk-www: Стецюк, Валентин. Лексика как материал для реконструкции истории языка. In: <http://www.inauka.ru/blogs/article71407.html>
- Stigler 2008: Stigler, Hubert. XML-Frameworks im Korpusmanagement. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 617–629.
- Swadesh 1952: Swadesh, Morris. Lexico-statistic dating of prehistoric ethnic contacts. In: Proceedings of the American philosophical society. Nr. 36. S. 452–463.
- Tambovtsev 2002: Tambovtsev, Yuri. Comparative typological study of language distances based on the consonants in sound chains of various languages. In: Elliot, John (ed.). The 5th National Colloquium for Computational Linguistics in the UK. Proceedings of the Conference. 8-9 January. University of Leeds, UK. Leeds. S.77–80.
- Tambovcev 2002: Тамбовцев, Ю.А. Фонологическая схожесть и фонологические расстояния. In: сб. Гуманитарные проблемы миграции: социально-правовые аспекты адаптации соотечественников в Тюменской области. Тюмень. S. 274–277.
- Thomann 2008: Tomman, Robert. Das Gralis Anketarium. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 796–802.
- Tošović 2001: Tošović, Branko. Korelaciona sintaksa. Projektional. Graz.
- Tošović 2002: Tošović, Branko. Funkcionalni stilovi. Funktionale Stile. Graz.

- Tošović 2008a: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. 1/3. Münster et al. Reihe Slawische Sprachkorrelationen. Bd. 1.
- Tošović 2008b: Тошович, Бранко. Сопоставительное изучение славянских языков при помощи многоязычного „Гралис-Корпуса“. In: Stanković, Bogoljub (ed.). Izučavanje slovenskih jezika, književnosti i kultura kao inoslovenskih i stranih. Beograd. S. 336–340.
- Vidovič-Muha 1998: Vidovič-Muha, Ada (ed.). Slovenski jezik. Opole.
- Vojvodić 1997: Vojvodić, Dojčil. О еллиптичним конструкцијама у словенским језицима. In: Славистика, књ. I (1997). S. 7–14.
- Wagner/Fischer 1974: Wagner, Robert A.; Fischer, Michael J. The string-to-string correction problem. In: Journal of the Association for Computing Machinery. Nr. 21(1). S. 168–173.
- Weinreich 1953: Weinreich, Uriel. Languages in Contact. The Hague.
- Wildgen 1977: Wildgen, Wolfgang. Differentielle Linguistik. Entwurf eines Modells zur Beschreibung und Messung semantischer und pragmatischer Variation. Tübingen.
- Wingender 1998: Wingender, Monika. Standardsprachlichkeit in der Slavia. In: Zeitschrift für Slawistik. Bd. 43. S. 127–139.
- Wonisch 2008a: Wonisch, Arno. Das Gralis Personalium. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 813–821.
- Wonisch 2008b: Wonisch, Arno. Das Gralis Text-Korpus. In: Tošović, Branko (ed.). Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen. Münster et al. S. 724–749.
- Zieniukowa 1992: Zieniukowa, J. (ed.) Procesy rozwojowe w językach sowiaskich. Warschau.
- Zybatow 1998: Zybatow, Lew N. Zu neuen Horizonten der slavistischen Sprachkontakt und Sprachinselforschung. In: Die Welt der Slawen: München. Jg. XLIII/2. S 323–338.

Parallel (multilingual) corpora

- Compara-www: <http://www.linguateca.pt/COMPARA/index.php>
- EPC-www: <http://www.statmt.org/europarl>
- Gralis-Korpus-www: <http://www-gewi.kfunigraz.ac.at/gralis/0.Projektarium/Gralis-Korpus/korpus.html>
- Kollokation-www: <http://www.kokken.go.jp/public/world/mirror/www.ids-mannheim.de/gra/kollokation.html>
- Leeds-www: <http://corpus.leeds.ac.uk>
- Lilabar-www: <http://lilabar.com/index.php>
- Maastr-www: <http://www.philhist.uni-augsburg.de/lehrstuehle/anglistik/sprachwissenschaft/mitarbeiter/stoll/elekhilf>
- RPC-www: [http://www.uni-regensburg.de/Fakultaeten/phil\\_Fak\\_IV/Slavistik/RPC](http://www.uni-regensburg.de/Fakultaeten/phil_Fak_IV/Slavistik/RPC)
- Ruscorpora-www: <http://ruscorpora.ru>; <http://ruscorpora.ru/search-para.html>

SPI-www: <http://nevmenandr.net/slovo>

Traumdeutung-www: [http://www.aac.ac.at/lab\\_parallel\\_freud.html](http://www.aac.ac.at/lab_parallel_freud.html)

#### Sources

Deutsche Welle: <http://www2.dw-world.de>

Glas Amerike: <http://www.voanews.com/serbian/>

Gralis-www: <http://www-gewi.uni-graz.at/gralis/>

Southeast European Times <http://www.setimes.com>

#### Abkürzungen

Be – Weißrussisch

Bg – Bulgarisch

Bs – Bosnisch

Cs – Tschechisch

De – Deutsch

Hg – Burgenlandkroatisch

Hr – Kroatisch

KS – Kaschubisch

Mk – Mazedonisch

Mo – Montenegrinisch

Pl – Polnisch

RS – Russinisch

Ru – Russisch

Sk – Slowakisch

Sl – Slowenisch

So – Sorbisch

Sr – Serbisch

Uk – Ukrainisch

SSS –slawische Standardsprachen

