

Walter Breu, Evangelia Adamou

SLAVISCHE VARIETÄTEN IN NICHTSLAVOPHONEN LÄNDERN EUROPAS. DAS DEUTSCH-FRANZÖSISCHE GEMEINSCHAFTSPROJEKT EuroSlav 2010

1. Einleitung

Der vorliegende Beitrag befaßt sich mit einer Darstellung des aktuell durchgeführten Gemeinschaftsprojekts *Elektronische Datenbank bedrohter slavischer Varietäten in nichtslavophonen Ländern Europas / Base de données électronique de variétés slaves menacées dans des pays européens non slavophones* (EuroSlav 2010). Es befindet sich beim jetzigen Stand (April 2011) etwa in der Hälfte der veranschlagten Bearbeitungszeit von 24 Monaten und soll Anfang 2012 abgeschlossen werden. Das Projekt wird in enger Kooperation zwischen dem Lehrstuhl Slavistik/Sprachwissenschaft der Universität Konstanz und dem von der französischen Wissenschaftsorganisation CNRS (Centre national de la recherche scientifique) getragenen *Laboratoire Lacito* (Langues et civilisations à tradition orale) in Paris/Villejuif durchgeführt.¹ Die Ergebnisse des Projekts sollen als Sprachkorpus in das Programm *Archivage* des Lacito integriert werden, das bereits 200 Dokumente in 44 Sprachen umfaßt, sowie in die Plattform des CRDO (Centre de ressources pour la description de l'oral) mit 90 Sprachen und 1640 öffentlich zugänglichen Dokumenten. Auf diese Weise wird das Projekt von den Erfahrungen und Möglichkeiten dieser beiden Plattformen profitieren und der dauerhafte Fortbestand der Datenbank sichergestellt, zumal Datenbanken in Frankreich als vorrangige Forschungsgebiete (*axes de recherche prioritaires*) gesehen werden. Das Lacito-Archiv zur mündlichen Rede (*Archives orales*) ist im Internet frei zugänglich.²

¹ Das Projekt war im Rahmen der „Ausschreibung eines deutsch-französischen Programms in den Geistes- und Sozialwissenschaften 2009“ beantragt worden. Es wird von der DFG und der französischen Partnerorganisation ANR (Agence nationale de la recherche) gemeinsam gefördert. Den Pariser Teil des Projekts leitet Evangelia Adamou, die Leitung des Konstanzer Teils hat Walter Breu. Weitere verantwortliche Mitarbeiter sind in Konstanz Lenka Scholze und Mía Barbara Mader Skender, in Paris Georges Drettas.

² Vgl. die Startseite des Archivs unter <http://lacito.vjf.cnrs.fr/archivage/presentation.htm>. Eine kurze Zusammenfassung des Projekts *EuroSlav 2010* ist augenblicklich über den Link <http://lacito.vjf.cnrs.fr/parteneriat/euroslav/index.htm> abrufbar. Nach der Fertigstellung der

Sowohl hinsichtlich des sprachlichen Materials wie auch hinsichtlich der Veröffentlichungsform werden in dem Projekt neue Wege eingeschlagen, um den an slavischen Minderheitensprachen interessierten Sprachwissenschaftlern mit und ohne slavistischen Hintergrund einen direkten Zugriff auf mündliche Texte zu liefern.

Die Texte werden dabei so detailliert aufbereitet, daß auch nicht mit diesen Mikrosprachen vertraute Sprachwissenschaftler einen umfassenden Einblick in das Funktionieren der in absolutem Sprachkontakt³ stehenden Varietäten gewinnen können, um verlässliche Grundlagen für eine Berücksichtigung der betreffenden Sprachsysteme in sprachübergreifenden Analysen zur Verfügung zu haben. Die notwendige Software für die Text- und Tondokumentation wurde vom Lacito zur Verfügung gestellt.⁴ Für die technische Aufbereitung steht am Lacito eine Informatikerin zur Verfügung (Séverine Guillaume), die sich auch um spezielle, bisher nicht implementierte Problemlösungen unseres Projekts kümmert.

In der vorliegenden Beschreibung soll ein geraffter Überblick über die Darstellungsverfahren und die dabei zu lösenden Probleme vermittelt werden. Es handelt sich entsprechend dem gegenwärtigen Stand des Projekts (*work in progress*) um eine Momentaufnahme, die das prinzipielle Vorgehen erläutern soll, in der Hoffnung, ähnlich strukturierte Projekte für andere Sprachgebiete anzuregen und hierfür eine Diskussionsgrundlage zu bieten, wobei sich die vorliegende Skizze schwerpunktmäßig auf den „deutschen“ Teil des Projekts beschränkt und hier wiederum auf das Moliseslavische, dessen Dokumentation schon am weitesten fortgeschritten ist. Aus dem „französischen“ Projektteil wird gegen Ende des Beitrags ein Beispieltext aus der slavischen Varietät von Liti (Griechenland) vorgestellt.

Das Projekt *EuroSlav 2010* ist ein Teil der Bemühungen in Europa und der Welt, etwa im Rahmen der Unesco, um die Bewahrung des immateriellen Erbes der Menschheit. Die Europäische Union ermuntert ebenfalls seit mehreren Jahren ihre Mitgliedsstaaten, sich ihres mehrsprachlichen Erbes bewußt zu werden. Während Deutschland, Österreich und Italien auf ihrem Territorium gesproche-

Dokumentation werden die Texte entsprechend den Prinzipien des Lacito eingruppiert, und zwar mit den projektypischen erweiterten Auswahlmöglichkeiten.

³ „Absoluter“ oder „totaler“ Sprachkontakt liegt dann vor, wenn alle Sprecher einer Sprachgemeinschaft neben ihrer Erstsprache in einem hohen Maße auch die dominante Zweitsprache beherrschen, in diesem Sinne also vollständig bilingual sind, wobei die Zweitsprache auch Dachsprachenfunktion innehat (fremdes Dach). In einer solchen Situation kann jeder Sprecher jederzeit aus der Zweitsprache in die Erstsprache entlehnen, und er kann bei Bedarf stets zwischen beiden Sprachen wechseln. Jede der beiden Kontaktsprachen kann in mehreren Varietäten auftreten.

⁴ Die betreffenden Werkzeuge sind im Prinzip unter folgendem Link öffentlich zugänglich: <http://lacito.vjf.cnrs.fr/archivage/outils.htm>.

ne traditionelle Minderheitensprachen heute in mehr oder minder expliziter Weise rechtlich anerkennen, besteht beispielsweise in Griechenland auf diesem Gebiet noch erheblicher Nachholbedarf. Die bereitzustellenden Korpora können zu den Bemühungen um eine größere Wertschätzung der in Frage stehenden Varietäten beitragen und in der Zukunft ihre Akzeptanz durch die Politik erleichtern. Das betrifft insbesondere Fragen des Schulunterrichts. Soweit überhaupt Sprachunterricht in den im Projekt untersuchten Minderheitensprachen erteilt wird, werden nämlich bisher nur literarische, in der Regel sehr stark von der Alltagssprache der Sprecher vor Ort abweichende Sprachformen gelehrt, die von Letzteren oft als fremd empfunden werden. Mit unserer öffentlich zugänglichen Präsentation soll deshalb auch eine Hebung des Prestiges der tatsächlich gebrauchten kontaktbestimmten Varietäten bei den Sprechern selbst einher gehen, was deren Selbstverständnis in sprachlicher Hinsicht fördert und damit zur Akzeptanz der gegenüber den Standardsprachen oft als minderwertig hingestellten Varietäten beiträgt. Damit könnte größere Sprachloyalität und letztendlich die Bewahrung der Minderheitensprachen überhaupt gefördert werden.

Im Folgenden seien die hauptsächlichen Zielsetzungen des Projekts *EuroSlav2010* zusammengefaßt:

- Schaffung eines annotierten und der internationalen wissenschaftlichen Öffentlichkeit (*scientific community*) online zugänglichen Korpus nicht dokumentierter und vom Aussterben bedrohter slavischer Sprachvarietäten in nichtslavischen Ländern;
- Bereitstellung eines an Sprachkontaktphänomenen reichen Datenkorpus, das zu kontaktlinguistischen Forschungen herangezogen werden kann;
- ein allgemein-sprachwissenschaftlicher Ansatz mit internationaler Sichtbarkeit der Ergebnisse;
- neben den Slavisten sollen so ganz allgemein sprachkontakt- und typologieinteressierte Linguisten zu natürlichen, analysierten und annotierten mündlichen Daten dieser nichtdokumentierten und vom Verschwinden bedrohten slavischen Sprachen Zugang erhalten;
- angesichts der Tatsache, daß die hier präsentierten slavischen Varietäten in mehr oder minder hohem Ausmaß zum Aussterben tendieren, ist schließlich die Bewahrung eines vom Verschwinden bedrohten sprachlichen und kulturellen Erbes für die Zukunft ein wichtiges, über die Sprachwissenschaft hinausgehendes Anliegen von *EuroSlav 2010*.

2. Dokumentierte Sprachen und Datengewinnung

EuroSlav 2010 ist ein Pilotprojekt und erfasst vom Aussterben bedrohte slawische Mikrosprachen im absoluten Sprachkontakt in Italien (Moliseslawisch), Österreich (Burgenlandkroatisch), Deutschland (Obersorbische Umgangssprache) und in Griechenland (Liti und Chrisa).⁵ Die an dem Projekt beteiligten Sprachwissenschaftler beschäftigen sich zum Teil schon seit langer Zeit mit den betreffenden Minderheitensprachen und haben ausführliche Sprachbeschreibungen geliefert, deren bibliographische Daten im Rahmen des Internetauftritts von *EuroSlav 2010* in einem Anhang zur Verfügung gestellt werden. Bei den bisherigen Arbeiten handelte es sich um Beschreibungen einzelner grammatischer Kategorien und um Fragen des kontaktsprachlichen Einflusses, die an Hand ausgewählter, oftmals adaptierter oder gekürzter Beispiele diskutiert wurden. Was bisher aber vollständig fehlte, sind allgemein zugängliche Texte als Basis für eine Einbeziehung dieser Mikrosprachen in einen erweiterten wissenschaftlichen (und kulturellen) Diskurs.

In diesem Sinne ist das Ziel des Projekts *EuroSlav 2010* die Schaffung einer Datenbank aus in Feldforschung gesammelten mündlichen Originaldaten, die der Forschergemeinde im Netz über die üblichen Browser wie *Firefox* oder *Internet Explorer* zugänglich sein wird. Das Datenkorpus wird weitgehend vereinheitlichte, aber die konkrete Struktur der Sprachsysteme berücksichtigende morphosyntaktische Glossierungen für alle erfaßten Mikrosprachen enthalten, und zwar in einer Kombination der Verfahren der linguistischen Typologie und der slavistischen Tradition.

Die Datenbank wird einen Verbund von Laut und Text darstellen, der fortlaufend oder Satz für Satz abgehört werden kann. Der geplante Umfang der Internetpublikation von etwa sechzig Minuten Aufnahme je Varietät mag auf den ersten Blick wenig erscheinen, tatsächlich ergibt sich hieraus aber angesichts der Vielzahl von Analyseebenen ein ziemlich umfangreiches Korpus, das aufgrund der Textauswahl auch einen einigermaßen repräsentativen Überblick über die grammatische und lexikalische Struktur der Mikrosprachen unter besonderer Berücksichtigung der Lehneinflüsse liefert. Es wird sich um die bisher ausführlichste Dokumentation im Archiv des Lacito überhaupt handeln.

Die Sammlung der Texte unterliegt in den einzelnen Kontaktgebieten spezifischen Bedingungen, die einen unterschiedlichen Umfang bzw. Unterschiede in

⁵ Zum Moliseslawischen s. die Literaturverweise weiter unten, zum System der obersorbischen Umgangssprache und seiner Rolle innerhalb des sorbischen Sprachgebiets vgl. Scholze (2008). Einen Überblick über das Burgenlandkroatische bietet etwa Neweklowsky (1978), zur heutigen soziolinguistischen Situation vgl. Szeuich (2000). Für eine strukturelle Beschreibung der slavischen Varietät von Liti (Nashta) vgl. Adamou (2006), zu derjenigen von Chrisa vgl. Drettas (1990).

der Textauswahl zur Folge haben. Das ergibt sich beispielsweise schon aus dem Grad der Bedrohtheit, bzw. der Altersstruktur der noch verbliebenen Sprecher. Eine Rolle spielen aber auch soziopolitische oder im engeren Sinne politische Vorgaben, die die Textsammlung im Einzelfall erschweren. So ist es beispielsweise in Griechenland nicht opportun, die Sprecher slavischer Varietäten als „Minderheiten“ zu erfassen. Im Bereich des Obersorbischen tritt neben das Deutsche als Kontaktsprache die obersorbische Schriftsprache, die den Zugang zur „eigentlichen“ Umgangssprache im katholischen „Kerngebiet“, die sich strukturell in charakteristischer Weise von der Schriftsprache unterscheidet, erschweren kann.⁶ Im Burgenlandkroatischen bestehen neben der eigentlich zu untersuchenden Umgangssprache in der Alltagskommunikation sogar zwei standardisierte slavische Varietäten, die burgenlandkroatische Schriftsprache und die kroatische Standardsprache, ganz abgesehen davon, daß hier aufgrund der großen geographischen und dialektalen Streuung gesprochener Varietäten eine enge lokale Auswahl getroffen werden muß. Vor allem in Süditalien und ganz besonders in Griechenland spielt es eine Rolle, daß zum Teil nur ältere Sprecher noch wirkliche Träger der Minderheitensprache sind.

Ungeachtet bereits in der Vergangenheit gesammelter Texte sind für *EuroSlav 2010* neue Datenerhebungen im Rahmen von Feldforschungsunternehmungen notwendig. Die Gründe hierfür sind vielfältig. Im Fall der slavischen Varietät von Chrysa in Griechenland muß das von Georges Drettas in den 1970er Jahren gesammelte Korpus in einem gewissen Maße aktualisiert und mit dem aktuellen Sprachstand abgeglichen werden. Bei den anderen Varietäten sind die älteren Aufnahmen zum Teil von schlechter Tonqualität, so daß die bestehenden Korpora unter optimalen Aufnahmebedingungen vervollständigt werden müssen. Außerdem ist es teilweise notwendig, die vorhandenen Themenbereiche anlässlich der Erstellung der Datenbank auszuweiten. Eine Homogenisierung der Datenbank ist dabei angestrebt, wenn auch aus Gründen, die mit der Natur der Daten sowie den Schwierigkeiten ihrer Sammlung in den einzelnen Kontaktgebieten im Zusammenhang stehen, einige individuelle Abweichungen unvermeidbar sind: es ist eben gerade für vom Aussterben bedrohte Sprachen oft schwierig, andere mündliche Äußerungen zu sammeln als Märchen oder Erzählungen. Immerhin versuchen wir durch gezielte Aufnahmen in der Linguistik verbreiteter „Standardtexte“, auch einen hiervon abweichenden homogenen Themenbereich zu erschließen, namentlich der *pear story* (mit Videovorlage) und der *frog story* (mit Bildvorlage).

Die Aufnahmen werden mit modernen digitalen Aufnahmegeräten im Taschenformat durchgeführt, und zwar mit den Digitalrekordern Roland/Edirol R09-HR bzw. Marantz PMD-660, jeweils mit auswechselbaren Speicherkarten,

⁶ Zur äußeren Situation der obersorbischen Umgangssprache vgl. Scholze (2008: 31-42).

deren Inhalt zur Sicherung kontinuierlich auch auf das mitgeführte Notebook übertragen wird.⁷ In der Regel werden gerichtete Mikrophone eingesetzt, wie das AKG C535eb, das Sony ECM-MS957 und das an R09 optimal angepaßte Stereomikrophon Roland CS-15. Ältere Aufnahmen wurden z.T. mit Tonband (Uher), z.T. mit Kassettenrekorder (Aiwa) und seit den späten 90er Jahren mit einem DAT-Rekorder realisiert und werden direkt über die Soundkarte am PC digitalisiert.

Zumeist wird für die digitalen Aufnahmen das unkompromierte *wav*-Format (44,1 kHz, 24 bit) gewählt, für die Internetpräsentation kann es in das teilweise bereits ursprünglich vorliegende *mp3*-Format überführt werden, da dieses bei einer ausreichend hohen Bitrate (44,1 kHz, 192 Kbit/sec) ebenfalls sehr gute Resultate liefert und den Vorteil eines geringeren Speicherbedarfs hat. In Einzelfällen werden Nebengeräusche oder nicht zum Haupttext gehörige Passagen herausgeschnitten, im allgemeinen erscheinen die Texte aber in der Form, wie sie aufgenommen wurden, was gesprächstypische Wiederholungen und Verbesserungen einschließt, in unvermeidbaren Fällen auch die Intervention anderer Sprecher.

Transkription und Übersetzung werden durch die beteiligten Linguisten selbst gewährleistet, unter Mithilfe von Informanten, die über Werkverträge an das Projekt gebunden werden.

Da die aus älterer Zeit vorliegenden Daten meist in Hinblick auf eine spezielle Arbeit des einzelnen Forschers oder eine nicht-audiogestützte Textpublikation gesammelt worden waren, etwa zur Anlage einer Märchensammlung, wurden die Sprecher seinerzeit in der Regel nicht um schriftliche Autorisierung zur Nutzung der Daten gebeten. Nun sind für die Online-Nutzung solche Autorisierungen zwar notwendig geworden, infolge politischer und kultureller Eigenheiten der betroffenen Gebiete sind schriftliche Verträge aber meist ein ungeeignetes Mittel, und bei alten Aufnahmen sind die Gewährleute oft schon gestorben. Man muß also eine Lösung entsprechend dem *Guide des Bonnes Pratiques pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux*⁸ anwenden und Archivierungsprogramme, die die betreffenden Überlegungen bereits durchgeführt haben, zu Rate ziehen. Beispielsweise nimmt DoBeS (Dokumentation bedrohter Sprachen) in Abhängigkeit von der Sensibilität der Daten

⁷ Diese Vorsichtsmaßnahme hat sich als dringend notwendig herausgestellt, da insbesondere SD-Karten nicht sehr robust sind und wir schon mehrere Ausfälle hatten. In einem Fall wurde die Karte sogar während der Übertragung auf den PC selbst zerstört, was entsprechenden Datenverlust zur Folge hatte. Die Wiederherstellung durch Spezialfirmen geht bei defektem Controller der Karte durchaus in den 1000€-Bereich.

⁸ Downloadbar unter: http://www.culture.gouv.fr/culture/dglf/Guide_Corpus_Oraux_2005.pdf (Zugriff April 2011).

vier Zugangsniveaus an.⁹ Bei den aktuell durchgeführten Aufnahmen werden die Sprecher direkt um ihre Zustimmung zu einer Internetveröffentlichung gebeten. Die Schriftform bleibt allerdings auch hier aus kulturbedingten Gründen ausgeschlossen.

3. Zur Situierung von *EuroSlav 2010* in der Forschungslandschaft

Seit etwa zwanzig Jahren erfahren Programme zur Sprachdokumentation einen lebhaften Aufschwung. Neben den Korpora zu weitverbreiteten Sprachen stehen Projekte zur Dokumentation bedrohter Sprachen – vor allem aus entlegenen Erdstrichen – im Mittelpunkt des Interesses. Reguläre Finanzierungen wurden von mehreren Stiftungen und Organisationen lanciert, mit dem Ziel der Veröffentlichung einer Grammatik der betreffenden Sprache sowie der Archivierung von Audio- und Videodaten: Foundation for Endangered Languages (Bath, U.K.), Endangered Languages Foundation (Yale, USA), UNESCO, Endangered Language Documentation Programme (School of Oriental and African Languages, London U.K.), Volkswagenstiftung: DoBeS (Dokumentation Bedrohter Sprachen) etc. In Projekten dieser Art wurden jedoch die vom Aussterben bedrohten Varietäten großer Sprachfamilien, wie die slavischen Minderheitensprachen im vorliegenden Projekt, nicht erfaßt.

Zwar sind slavische oder gesamteuropäische dialektologische Projekte prinzipiell imstande, solche Daten zu erarbeiten, sie haben aber unterschiedliche, zum beantragten Projekt komplementäre Zielsetzungen. Zunächst ist festzustellen, daß die große Tradition dialektologischer Studien vor allem auf phonetischen und lexikalischen Untersuchungen basiert, während Morphologie und Syntax weniger entwickelt sind. Außerdem richten sich die unter dialektologischer Perspektive bearbeiteten Texte in erster Linie an Slavisten und sind nur selten glossiert oder in sprachtypologisch erforderlicher Weise analysiert. Vom methodologischen Standpunkt her ist die Datenbeschaffung hier oft in idealisierender Weise auf eine nicht mehr zutreffende Authentizität ausgerichtet. Beispielsweise wurde diese hinsichtlich der in Griechenland gesprochenen Varietäten häufig bei Immigranten und politischen Flüchtlingen durchgeführt, unter Nichtbeachtung der von den Sprechern zwischenzeitlich aufgenommenen Einflüsse slavischer Standardsprachen. Was die anderen untersuchten Mikrosprachen angeht, so sind oft störende puristische „Verbesserungen“ zu verzeichnen oder aber es handelt sich um Dialektaufnahmen von besonders konservativen Sprechern, die der heutigen Realität nicht mehr entsprechen. Grundsätzlich sind die Varietäten des Projekts *EuroSlav 2010* bisher nur sehr spärlich dokumentiert

⁹ Vgl. http://www.mpi.nl/DOBES/archive_access/ (Zugriff April 2011).

und in aller Regel in reiner Schriftform, ohne Zugriff auf die tatsächliche Lautung.

Im Projekt *EuroSlav 2010* werden die Aufnahmen vor Ort erstellt, und die Daten werden unter Berücksichtigung der realen Äußerungssituation analysiert. Auf der Analyseebene sind der aktuelle Gebrauch und die bei einem Sprecher oder in der Gemeinschaft der Sprecher beobachtbare Variation ein Gegenstand der Untersuchung. Tatsächlich sind die genannten Varietäten nicht einfach Dialekte der ihnen am nächsten stehenden Sprachen (Kroatisch, Bulgarisch, Makedonisch, Obersorbisch), für die es genügen würde, die Abweichungen einfach als archaische Züge zu bestimmen. Sie zeigen vielmehr zahlreiche Neuerungen, mit für slavische Sprachen unerwarteten Phänomenen, bis hin zur Durchbrechung diachroner grammatischer Konstanten des Slavischen und ihrer Ersetzung durch diachrone Konstanten der jeweiligen Kontaktsprache; vgl. zum Beispiel Breu (2008a) zur Entstehung von Artikelsystemen im Moliseslavischen und in der obersorbischen Umgangssprache oder Breu (2005) zu den sprachkontaktinduzierten Änderungen im Aspektsystem dieser Sprachen. Scholze (2008) ermöglicht einen Überblick über die kontaktinduzierten „Singularitäten“ der obersorbischen Umgangssprache, Breu (2011a) faßt wichtige kontaktbedingte Veränderungen im Verbsystem mehrerer slavischer Mikrosprachen zusammen.¹⁰ Man vergleiche daneben auch Bayer (2006) zu einer kontrastiven Studie der sprachlichen Interferenzen bei diversen slavischen Minderheiten im Kontakt mit dem Deutschen, sowie die in die typologische Literatur aufgenommenen moliseslavischen Beispiele bei Heine & Kuteva (2006).

Die in dem Projekt erfaßten süd- und westslavischen Varietäten befinden sich in europäischen Staaten mit einer nicht zur slavischen Familie gehörenden offiziellen Sprache (Griechisch, Italienisch, Deutsch). In allen Fällen geben die Sprecher ihre Varietät zunehmend zugunsten der Staatssprachen auf, die für ihren sozialen Aufstieg nützlicher sind und einen Teil ihrer nationalen Identität darstellen. Die Situation variiert jedoch je nach Land. Wie bereits festgestellt, ist in Griechenland die Minderheitenfrage ein Tabu, während in Österreich, Deutschland und Italien sogar Schulunterricht vorgesehen ist. Allerdings ist dieser Unterricht außer beim Obersorbischen nur fakultativ und wird in der Realität (vor allem in Italien) nur wenig in Anspruch genommen. Außerdem wird – wie oben bereits festgestellt – in der Regel (etwa im Fall des Obersorbischen) in standardisierten und puristischen Sprachformen unterrichtet. Die Projektdaten sind in diesem Zusammenhang in zweierlei Hinsicht wertvoll, einerseits weil diese Varietäten wohl in nicht allzu ferner Zukunft verschwinden werden, ande-

¹⁰ Als Sonderentwicklung in einer durch Sprachkontakt gekennzeichneten Varietät sei auch die für eine slavische Sprache ungewöhnliche Entwicklung eines temporal-modalen Referenzsystems im Pomakischen in Nordgriechenland auf der Basis der traditionellen raumbezogenen deiktischen Dreifachdifferenzierung genannt (Adamou 2011).

erseits weil in gewissen Fällen (insbesondere in Griechenland) der Zugang zu den Gebieten aus politischen und ideologischen Gründen nicht einfach ist.

4. Aufbau und Präsentation der Datenbank von *EuroSlav 2010*

4.1. Streuung der Texte nach Dialekt, Sprecher, Genre

Jeder Aufnahme werden Metadaten vorangestellt mit Informationen über Ort, Datum, Art der Aufnahme usw. Außerdem wird für jede der Mikrosprachen ein Porträt der extralinguistischen Situation sowie der grammatischen und lexikalischen Besonderheiten erstellt.

Es wird auf eine möglichst breite Streuung von Sprechern nach Alter, Geschlecht und beruflicher Tätigkeit geachtet, wobei die tatsächlichen Möglichkeiten in den einzelnen Gebieten sehr unterschiedlich sind. Soweit nicht von Haus aus die Varietät einer einzigen Ortschaft analysiert wird, soll durch die Wahl der Herkunft der Sprecher ein Überblick über die Binnendifferenzierung der betreffenden Varietät erreicht werden. Beispielsweise werden im Fall des Moliseslavischen Sprecher aus allen drei slavischsprachigen Dörfern mit ihrer vergleichsweise starken dialektalen Differenzierung herangezogen.

Die Aufnahmen verteilen sich in den einzelnen Varietäten auf die im folgenden angegebenen Themenbereiche, die allerdings nicht immer klar voneinander getrennt werden können. Die Zeitangaben sind als maximaler Umfang gedacht, wobei sich lokaltypische Schwerpunktsetzungen ergeben; der tatsächliche Gesamtumfang sollte, wie gesagt, jeweils bei 60 Minuten liegen:

- 15 Minuten traditionelle Erzählungen, insbesondere auch solche aus dem Bereich des Aberglaubens (Vampire, Hexen)
- 10 Minuten persönliche Erlebnisse, insbesondere auch im Krieg, der bei allen Minderheiten eine wichtige Rolle spielt
- 10 Minuten Anekdoten, insbesondere solche mit einem fröhlichen Hintergrund
- 20 Minuten Traditionen und Alltag: Lokalgeschichte, Sitten und Gebräuche, Hochzeiten, Landwirtschaft, Küche, Rezepte, Seidenproduktion, Spinnen und Weben, Prozessionen, Trachten
- 10 Minuten Dialoge und Diskussionen (Teilnahme mehrerer Sprecher)
- 10 Minuten Interpretation von Bildergeschichten/Videos wie *pear story* und *frog story*

4.2. Die technische Realisierung

Die Erstellung der Transkriptionen geht von drei unterschiedlichen Voraussetzungen aus. Bei Daten, die anlässlich früherer Erhebungen gesammelt wurden, liegt der Text entweder als Manuskript vor, oder es wurde bereits ursprünglich eine Annotierung in computerisierter Form vorgenommen. Bei den aktuellen Aufnahmen bestehen natürlich überhaupt noch keine Vorlagen. In allen drei Fällen wird einheitlich auf die von M. Jacobson am Lacito erstellte Software *Interlinear Text Editor* (ITE.jnlp) zurückgegriffen, die die Präsentation durch eine erleichterte Datenerfassung unterstützt.¹¹ Genauer ist ITE ein Textverarbeitungsprogramm zur Darstellung komplexer linguistischer Informationen mit Hilfe der Auszeichnungssprache *xml*, das unter anderem die bündige Glossierung eines Dokuments erlaubt mit der Repräsentation von Sätzen, Glossierungen und Übersetzungen in interlinearer Form und auch die Vorhaltung einer Liste der verwendeten Glossen ermöglicht. Die Sprache der glossierenden Interlinearversion ist durchgehend das Englische zur Gewährleistung einer gemeinsamen, homogenen Zwischenstufe für alle Projektteile. Das Programm erlaubt auch den Einsatz von freien Fonts auf *Unicode*-Basis, die für die phonetische Darstellung der Texte in IPA unerlässlich sind, wenn ein offener Zugang über den Browser möglich sein soll. Aktuell verwenden wir die Zeichensätze (ttf-Forts) *Charis SIL*, *Doulos SIL* und *Gentium Plus* des *Summer Institute of Linguistics*, die die notwendigen Symbole enthalten.¹² Der windows-originaire Font *Lucida Sans Unicode* wäre prinzipiell ebenfalls einsetzbar, hat aber einige darstellungstechnische Nachteile (gewöhnungsbedürftiger Schriftschnitt, fehlende Zeichen).

ITE bietet mehrere Analyseebenen an, beginnend mit einer Gesanterfassung des Textes und seiner Übersetzung. Der Originaltext sollte dabei auf der Eingangsseite so eingegeben werden, daß die Einzelsätze mit einem Absatzende (Enter-Taste) abgetrennt werden. Dann kann er sehr leicht mit der Funktion „text > sentences“ (rechte Maustaste) in die Einzelsätze für die Satzanalyseebene geteilt werden. Dort (Reiter *sentence*) werden dann die einzelnen „Sätze“ (eigentlich phonologische Phrasen s.u.) übersetzt. Sobald dieser Vorgang abgeschlossen ist, kann auf der Eingangsseite unterhalb des Textes die Gesamtübersetzung mit Hilfe der Funktion „text < sentence“ aus den Satzübersetzungen generiert werden. Eine umgekehrte Funktion zur Erstellung der Einzelüberset-

¹¹ Download des (frei zugänglichen) Programms und Hinweise zu seiner Benutzung unter dem Link <http://michel.jacobson.free.fr/ITE/> (April 2011). Das Programm benötigt für seine Ausführung Java; empfohlen wird JRE 1.4.2 (*Java Runtime Environment version 1.4.2*), das in dieser Version aber nicht mehr aktuell ist. Für weitergehende Informationen vgl. <http://www.oracle.com/technetwork/java/javase/download-138376.html> (April 2011).

¹² Für Downloads: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=IP/home (Zugriff April 2011).

zungen aus der Gesamtübersetzung besteht hier nicht, d.h. die vorherige Eingabe der Gesamtübersetzung ist unnütz und birgt sogar die Gefahr von Übersetzungsdubletten in der Präsentation. Es sei auch darauf hingewiesen, daß der Vorgang der Erstellung einer Gesamtübersetzung aus den Teilübersetzungen nur funktioniert, wenn das Übersetzungsfeld nicht vollständig leer ist (eine Leertaste oder ein beliebiger Text genügt). Hier wie auch auf allen anderen Analyseebenen ist stets zu beachten, daß durch unbedacht angewählte Ersetzungsfunktionen bereits geleistete Arbeit wieder überschrieben wird!

Hinsichtlich der Aufgliederung des Textes auf „Satzebene“ ist zu beachten, daß bei der im Zentrum des Projekts stehenden mündlichen Rede nicht syntaktische Strukturen für die Segmentierung ausschlaggebend sind, sondern die phonologische Rhythmisierung, in der Hauptsache gekennzeichnet durch die Position der Pausen zwischen den einzelnen Textabschnitten. Sie fallen oft nicht mit den logischen Satzteilen oder Satzgrenzen zusammen, beispielsweise werden Relativpronomina oder Konjunktionen zuweilen direkt an das vorangehende Syntagma angeschlossen, gefolgt von einer Pause, oder aber es finden sich innerhalb eines längeren Satzes rhythmisch abgetrennte Sinnkomplexe, was auf der „Satzebene“ zur Aufteilung von Sätzen in kleinere Einheiten führen kann. Die Orientierung der als „Satz“ erfaßten Textabschnitte an der nicht notwendigerweise mit der syntaktischen Struktur zusammenfallenden lautlichen Realisierung ist eine unbedingt notwendige Voraussetzung für die Verknüpfung der segmentierten Transkription mit der Lautausgabe, weil sonst unschöne Schnitte mitten in der fortlaufenden Rede vorgenommen werden müßten.

Ausgehend von der Satzebene (Reiter *sentence*) kann nun über die Funktion „sentence > words“ wieder eine automatische Aufgliederung in nicht weiter segmentierte Wörter vorgenommen werden (Kriterium Wortabstand/Leertaste), die auf der Wortebene (Reiter *word*), falls man das will, einzeln übersetzt werden können. Das ist v.a. für die automatische Anlage eines Wörterbuchs sinnvoll. Im Projekt *EuroSlav 2010* ist ein solches nicht vorgesehen, so daß wir die Wortebene ersatzweise für andere Informationen verwenden können, s.u.

Ebenfalls von der Satzebene aus kann eine automatische Aufgliederung in zu segmentierende Wörter (Reiter *morpheme*) vorgenommen werden. Die so erstellten Wörter werden auf der Morphemebene manuell in ihre Morphe zerlegt (Funktion „split the morpheme“); vgl. hierzu den folgenden ITE-Screenshot:

File Edit Tools			
text	sentence	word	morpheme
word[1]		word[2]	
n ART.INDF	u= ACC.SG.F	vo:t time	u ACC.SG.F
word[3]		word[4]	
'bi be	x IPRF	u 1SG	di where
word[5]	word[6]	word[7]	
moj my.NOM.SG.M	tsi:l uncle	a NOM.SG.M	'pop priest.NOM.SG.M

Jedem einzelnen Morph wird hier in der 2. Zeile seine Morphembedeutung zugeordnet, in der Regel beginnend mit einer englischen Übersetzung der Lexembedeutung, gefolgt von der grammatischen Glossierung. Wir verwenden die *Leipzig Glossing Rules*,¹³ erweitert um für eine adäquate Erfassung der einzelnen Mikrosprachen notwendige zusätzliche Siglen; nähere Erläuterungen der Glossierung s.u.

Während die Glossierung im Gesamtprojekt weitestgehend homogen durchgeführt wird, einschließlich der Verwendung der IPA-Umschrift für die Wiedergabe der Morphemformen, wird auf der Text- und Satzebene aus pragmatischen Gründen unterschiedlich verfahren. Für die Mikrosprachen in Italien, Österreich und Deutschland bietet sich eine orthographische Wiedergabe an, die sich an die dort üblichen Schreibweisen der slavischen Mikro- bzw. Standardsprachen anlehnt. Die im französischen Projektteil jeweils verwendete Umschrift der Texte und Sätze ist hingegen identisch mit der IPA-Transkription der Morpheme, da für die hier untersuchten Texte slavischer Varietäten in Griechenland keine übliche Graphie besteht und man aus kulturspezifischen Gründen weder eine Anlehnung an die makedonische oder bulgarische Kyrilliza für geeignet hält noch eine an den slavistischen Transliterationszeichen orientierte Umschrift anstrebt.

Im Gegensatz zur herkömmlichen Beschränkung auf eine einzige grammatische Glossierungszeile, die in der Regel eine mangelhafte Kennzeichnung analytischer Formen mit sich bringt, haben wir uns bei *EuroSlav 2010* entschlossen, zwischen der Annotierung synthetischer und analytischer Formen explizit zu

¹³ Download: <http://www.eva.mpg.de/lingua/resources/glossing-rules.php> (April 2011).

unterscheiden, indem wir eine zusätzliche Glossierungszeile einfügen. Obwohl im Editor ITE hierfür eigentlich keine Möglichkeit vorgesehen ist, gelingt die Angabe einer solchen zusätzlichen (morphosyntaktischen) Zeile für analytische Formen durch einen Kunstgriff dennoch, und zwar auf dem Wege einer Uminterpretation der vorgesehenen Analyseebenen. Da die Morphem- und die von uns sonst nicht genutzte Wortebene (s.o.) per Numerierung direkt aufeinander bezogen sind, ergibt sich nämlich die Möglichkeit, die einer analytischen Form zugrundeliegenden Einzelformen auf der Wortebene (Reiter *word*) erneut zu glossieren, und zwar als „Übersetzung“ der „Wörter“.

Das Programm ITE erlaubt noch eine Reihe weiterer Eingaben, insbesondere über den direkten Zugriff auf die *xml*-Auszeichnung, die es unter anderem ermöglicht, Sprach- (v.a. Sprache der Übersetzung) und Schriftattribute (z.B. kursiv) explizit einzutragen. Hierzu ist zu sagen, daß das Programm bei der Öffnung oder Neuanlage einer Datei nur die Sprachauswahl „français“ und „English“ zuläßt, abgesehen von einem „default“, der aber tunlichst zu vermeiden ist, will man bei Verwendung mehrerer Übersetzungssprachen nicht ein Durcheinander verursachen. Es ist aber möglich, eine solche Sprachzuordnung mit einem *xml*-Editor im Nachhinein zu ändern, beispielsweise für das Moliseslavische die Markierung *xml:lang="fr"* (französisch) durch *xml:lang="it"* (italienisch) zu ersetzen. Die Übersetzung in einer weiteren Sprache, etwa Deutsch, kann unter „English“ eingetragen und dann ebenfalls umkategorisiert werden.

Eine nützliche Funktion innerhalb von ITE ist die über den Menüpunkt *Tools* zugängliche versuchsweise Darstellung des Ergebnisses der Textanalyse im *html*-Format, also im Prinzip so, wie es nachher im Internetbrowser erscheint. Hierzu muß die *xml*-Datei mit einem *xsl*-Schema (stylesheet) verbunden werden. Das dem ITE beigegebene Schema *it.xsl* erlaubt allerdings nur die Wiedergabe der Übersetzung in einer einzigen Sprache und die Markierung analytischer Formen erscheint vor der phonetischen Transkription. Dem ist aber durch Modifikation des *xsl*-Schemas mit Hilfe eines *xml*-Editors¹⁴ leicht abzuhelfen, was allerdings zumindest Grundkenntnisse in der Auszeichnungssprache *xml* erfordert. Die *html*-Präsentation für den Internetbrowser kann unter *Tools* mit *File/Save as* (unter Hinzufügung der Erweiterung *.html* hinter dem Dateinamen) auch direkt abgespeichert und später im Browser aufgerufen werden.

Die Synchronisierung (*time alignment*) der Transkription und der Lautung auf Satzebene wird mit *SoundIndex* (SI.tcl) durchgeführt, einer mit ITE verbundenen Software,¹⁵ die es über Markierungen mit der Maus erlaubt, Ausschnitte

¹⁴ Ein kostenloser Editor ist beispielsweise XPontus: <http://xpontus.sourceforge.net/>

¹⁵ Vgl. <http://michel.jacobson.free.fr/soundIndex/> (April 2011) für die zur Nutzung des Programms notwendigen technischen Voraussetzungen, die allerdings keineswegs trivial sind. Leider hat die aktuelle Version von *Soundindex* zudem die unangenehme Eigenschaft, beim Abspeichern die *xml*-typische Sprachauszeichnung *xml:lang* durch einfaches *lang* zu erset-

aus der zugehörigen Tondatei (möglich sind die Formate *wav* und *mp3*) fortlaufend den Sätzen oder Textabschnitten der *xml*-Datei zuzuordnen. *Soundindex* ist nicht für eine textgebundene Tonanalyse geeignet. Jedoch kann die durch die Synchronisierung um die Toninformationen erweiterte *xml*-Datei erneut im *Interlinear Text Editor* geöffnet werden. Nach zusätzlicher Öffnung der *wav*- oder *mp3*-Mediendatei ist dort nun Satz für Satz abhörbar (Lautsprechersymbol) und eine Erstellung oder Korrektur der phonetischen Transkription auf Satz-, Wort- oder Morphemebene möglich. Als technische Voraussetzung ist hierfür zusätzlich die Installation des *Java Media Framework* notwendig.¹⁶

Die endgültige Überführung der analysierten und annotierten Texte ins orale Archiv des Lacito wird von der dortigen Informatikerin geleistet, die sich auch um spezifische Informationen der Internetpublikation kümmert, wie die Angabe der vom Browser zu verwendenden *Unicode*-Schriftart oder die Interpretation auf der Morphemebene zusätzlich eingefügter Markierungen, vor allem diejenige zur Kursivkennzeichnung der Lehnwörter (s.u.). In der finalen Internetpräsentation wird der Verweis auf das verwendete *xsl*-Schema direkt in die *xml*-Datei eingegeben, so daß diese ohne den Umweg über ein *html*-Format direkt im Browser geöffnet werden kann.

4.3. Analyseebenen

Die im Abschnitt „Technische Realisierung“ angesprochenen Möglichkeiten sollen nun in ihrer konkreten Anwendung im Projekt beschrieben werden. Einzelne Überschneidungen mit den oben gegebenen Erläuterungen dienen dem besseren Verständnis des konkreten Verfahrens im Gesamtzusammenhang.

Die Texte werden in möglichst kurze Sinnabschnitte, in der Regel unterhalb der Satzebene eingeteilt, damit sie gut am Bildschirm parallelisierbar sind. Das ist vor allem wegen der Glossierungsebene notwendig, die ziemlich viel Platz in Anspruch nimmt und sonst eine Vielzahl unübersichtlicher Zeilenumbrüche hervorrufen würde. Neben der Kürze spielt allerdings die durch die Aufnahmen vorgegebene phonologische Rhythmisierung die Hauptrolle für die Segmentierung (s.o.). Wir sprechen im Folgenden analog zu dem Verfahren des Präsentationsprogramms ITE vereinfachend dennoch von „Sätzen“. Es werden insgesamt sechs Darstellungsebenen angesetzt: orthographisch, lautlich (phonetisch-phonologisch), grammatisch-synthetisch, grammatisch-analytisch, kontaktsprachliche Glossierung, Glossierung in einer weiteren Sprache mit überregionalem

zen, was dann bei der späteren Aufbereitung für den Browser zu großen Problemen führen kann. Deshalb muß nach der Verwendung von *Soundindex* die ursprüngliche Auszeichnung mit Hilfe eines *xml*-Editors wiederhergestellt werden.

¹⁶ Download unter <http://www.oracle.com/technetwork/java/javase/tech/index-jsp-140239.html> (April 2011).

Gebrauch in der Sprachwissenschaft, jedoch vorerst nicht in Englisch, da die Glossierung bereits in dieser Sprache erfolgt. Es soll später entschieden werden, ob eine zusätzliche Übersetzung ins Englische eventuell doch sinnvoll sein könnte.

1. Auf der orthographischen Ebene wird der zu analysierende „Satz“ mit der in der betreffenden Varietät üblichen Schreibung dargestellt. Da es sich im Prinzip um nicht verschriftete Sprachvarietäten handelt, wird eine der slavistischen Transliteration nahestehende Form gewählt, die auch mehr oder minder große Ähnlichkeit mit den Verschriftungen der genetisch nächstverwandten Standardsprachen aufweist, soweit diese in Lateinschrift geschrieben werden. Hierin weichen die aus Griechenland stammenden Texte ab, da die den betreffenden Varietäten am nächsten stehenden Standardsprachen Makedonisch und Bulgarisch kyrillisch geschrieben werden. Aufgrund lautlicher Besonderheiten ergeben sich aber auch sonst grundsätzlich Abweichungen von den in Frage stehenden Standardsprachen. Auf der orthographischen Ebene werden auch Verspacher (in eckigen Klammern) oder Auslassungen (in runden Klammern) angegeben, die in der phonetisch notierten Morphemform dann unterbleiben.

Konkret erfolgt die orthographische Wiedergabe der nicht aus Griechenland stammenden Texte unter Verwendung der von anderen Slavinen her bekannten Diakritika, vor allem des *Háček*, und diverser Ligaturen, etwa *lj*, *nj* für die palatalen Sonoranten [ʎ] und [ɲ]. Zum Teil treten aber auch sonst nur außerhalb des Slavischen verwendete Buchstaben(kombinationen) auf, wie *kj* [c], *hj* [ç], *gh* [ɣ] oder *ě* [ə], *ě* [ɛ], *ò* [ɔ] im Moliseslavischen sowie *ü* [y] oder *ö* [ø] in der obersorbischen Umgangssprache. Soweit sie in dem betreffenden Sprachgebiet üblich sind, kommen auf der graphischen Ebene auch Differenzierungen vor, die auf der Lautebene keine Entsprechung haben, etwa *w* = *l* [w] und *č* = *ć* [tʃ] oder *dž* [dʒ] vs. *ž* [ʒ] in der obersorbischen Umgangssprache, andererseits fehlen auf der graphischen Ebene prosodische Charakteristika wie Vokallänge, Akzentstelle, Wortintonation oder auch in dieser Varietät üblicherweise nicht geschriebene Differenzierungen wie bei den *i*-Lauten, vgl. hier etwa *pica* mit der Opposition [pitsa] ‘Futter’ vs. [pitsa] ‘Pizza’. Während die orthographische Ebene die Texterkennung erleichtern soll, spiegelt die Wiedergabe auf der Lautebene (Morphemform) die tatsächliche Aussprache wieder.

2. Auf dieser lautlichen Ebene wird als Mittelweg eine weite phonetische Transkription mit IPA-Symbolen gewählt, die weitgehend mit einer phonologischen Wiedergabe zusammenfällt, jedoch die Möglichkeit zur Unterscheidung grundlegender Allophone offenläßt. Eine enge phonetische Transkription hätte die Lesbarkeit zu stark beeinträchtigt, eine rein phonologische Transkription hätte die Beigabe von Ausspracheregeln erfordert. In Abhängigkeit von der einzelnen Varietät sind hier auch prosodische Charakterisierungen notwendig, und zwar bei allen Varietäten, einschließlich derjenigen in Griechenland, hinsicht-

lich der Akzentstelle (im Sorbischen nur bei Lehnwörtern), im Burgenlandkroatischen und Moliseslavischen auch hinsichtlich der Wortintonation (musikalischer Akzent), in den beiden letzteren Varietäten und in der obersorbischen Umgangssprache (anders als im obersorbischen Standard) auch hinsichtlich der Vokalquantität. Allegrosprechformen werden durch Hinzusetzungen weggelassener oder verschliffener Laute in Klammern vervollständigt. Nichtintegrierte Spontanentlehnungen (nicht transkribiert) wie auch die eigentlichen Lehnwörter werden durch Kursivschrift markiert. Hierfür muß angesichts der Beschränkungen von ITE wieder ein Umweg gewählt werden, hier die Einführung einer besonderen Markierung in den *xml*-Quellcode des Morphems: `<M class="i">`.

3. Auf der grammatisch-synthetischen Ebene wird zunächst eine wörtliche englische Übersetzung der Wortform gegeben und daran eine Glossierung der Flexionsformen entsprechend den *Leipzig Glossing Rules* angefügt. Diese Glossierung soll maximal explizit sein. Es sind eine Reihe einzelsprachlicher Besonderheiten zu berücksichtigen, beispielsweise, daß im Moliseslavischen die slavische derivativ-grammatische Aspektkategorie (Perfektivitätsopposition) bei bestimmten Verbgruppen, die den Aspekt nur über die auch im Romanischen vorhandene „flexivische“ Opposition von Imperfekt : Perfekt ausdrücken, nicht glossiert wird (Neutralisierung); in der obersorbischen Umgangssprache wird die derivativ-grammatische Aspektkategorie aus Vergleichsgründen über IPFV : PFV wiedergeben, obwohl es sich inhaltlich-funktional hier um eine grammatisierte Terminativitätskategorie handelt.

Der definite Artikel, der in den in Griechenland gesprochenen Varietäten – wie in den verwandten Standardsprachen Bulgarisch und Makedonisch – als postponiertes Morphem erscheint, ist in der obersorbischen Umgangssprache ein selbständiges dem Nomen vorangestelltes Wort, im Moliseslavischen wird er über das zum indefiniten Artikel in Opposition stehende Nullmorphem ausgedrückt, das aber nicht als solches notiert wird, da es kein phonologisches Korrelat hat. Das gilt auch in der obersorbischen Umgangssprache außerhalb des pragmatisch-anaphorischen Bereichs. Im Burgenlandkroatischen ist erst noch zu entscheiden, ob es sich bei den betreffenden Formen wirklich um einen Artikel handelt oder eventuell doch „nur“ um ein Demonstrativum. Die Abtrennung des (lexikalischen) Wurzelmorphems sowie die Aufgliederung und Zuordnung der grammatischen Morpheme zu den grammatischen Kategorien bilden ein eigenes Problemfeld, bei dem innerhalb des Projektes erhebliche Analyse- und Homogenisierungsarbeit zu leisten ist.

Die in allen Varietäten relevanten Klitika werden per Konvention durch ein Gleichheitszeichen in Richtung auf den phonologischen Anker hin markiert, wobei im Obersorbischen und Burgenlandkroatischen die Beibehaltung der enklitischen Positionierung an der „zweiten Stelle“ im Satz auffällig ist (Wackernagels Gesetz), im Moliseslavischen (im Gegensatz zu den Standardvarian-

ten des Bosnisch-Kroatisch-Serbischen) und in den slavischen Varietäten in Griechenland aber die am Verb ausgerichtete Pro- und Enklise.

Spontan eingefügte kontaktsprachliche Wörter oder Phrasen (*code switching*) werden nicht glossiert, sondern durch eine Doppellinie in der Glossierungszeile gekennzeichnet. Dagegen werden ins grammatische System integrierte Lehnwörter, auch solche jüngeren Datums, normal glossiert.

4. Eine eigene 4. (morphosyntaktische) Analyseebene wird für die analytischen Verbformen eingerichtet, da sie auf der Wortebene nicht als solche zu erkennen sind. Sie bestehen in der Regel aus einem Auxiliar oder einer Partikel an einer Stelle und einer finiten oder nichtfiniten Verbform an einer anderen Stelle im Satz, deren Zusammengehörigkeit sich zumindest dem Nichtslavisten nicht unmittelbar erschließt. Wenn beispielsweise ein Perfekt aus den Formen des Auxiliars SEIN und dem *l*-Partizip zusammengesetzt ist, dann werden diese etwa im Fall von moliseslavisch *je pola* 'ist gegangen' auf der (synthetischen) Morphemebene (3. Ebene) mit der Segmentierung [je 'po-l-a] als „be.PRS.3SG“ und „go-PTCP-SG.F“ glossiert. Auf der morphosyntaktischen Analyseebene (4. Ebene) wird unter der 1. synthetischen Glosse „AUX.PRF(1)“ notiert, unter der 2. synthetischen Glosse „PRF(1)“, wobei die Indizierung, hier (1), angibt, welche Formen gemeinsam das Perfekt bilden; konkrete Beispiele s.u. Wir halten eine solche Doppelmarkierung von synthetischer und analytischer Ebene für einen deutlichen Fortschritt gegenüber den bisher üblichen Verfahren. Selbstverständlich können auf diese Weise problemlos auch analytische Formen annotiert werden, die aus mehr als zwei Einzelformen bestehen, und zwar auch im Fall diskontinuierlicher Elemente wie bei dem im Sorbischen üblichen Satzrahmen.

Außer dem Perfekt bestehen unter anderem folgende analytische Formen: Plusquamperfekt, Konditional, Nezesitivfutur, Probabilitätsfutur. Das moliseslavische Plusquamperfekt ist aus drei Formen zusammengesetzt, dem Perfektauxiliar, dem *l*-Partizip und der – teilweise infigierten – Plusquamperfektpartikel, die jeweils alle 3 denselben Index erhalten, um sie von weiteren Perfekten und Plusquamperfekten im Satz (hier als syntaktische Einheit, nicht als segmentierter Textabschnitt zu verstehen) unterscheiden zu können.

5. Auf der kontaktsprachlichen Glossierungsebene wird eine möglichst wörtliche Übersetzung des analysierten Textteils in der jeweiligen Kontaktsprache (Deutsch, Italienisch, Griechisch) geliefert. Die kontaktsprachlichen Lehnwörter und Spontanübernahmen werden auch hier – durch besondere Symbolisierung – gekennzeichnet; beim augenblicklichen Stand der Diskussion durch ein Sternchen zu Beginn und am Ende der Entlehnung. Da die kontaktsprachlichen Elemente in der Regel nicht aus dem Standard, sondern aus regionalen Varianten stammen oder aber sekundär an die mikrosprachliche Lautung angepaßt wurden, ist eine orthographische und lautliche Übereinstimmung mit der kontaktsprachlichen Übersetzung eher die Ausnahme, die Übernahme als solche ist aufgrund

der besonderen Markierung aber in jedem Fall klar zu erkennen. In besonders abweichenden Einzelfällen ist eine Anmerkung zu der betreffenden regional-sprachlichen Quellform vorgesehen.

6. Mit dem Ziel einer partiellen Homogenisierung und der Erleichterung der internationalen Benutzbarkeit wird bei den Texten aller Varietäten eine Übersetzung in mindestens einer weiteren Sprache hinzugefügt, die angesichts der (englischen) Glossierung relativ frei sein kann, auch um die bei der eher wörtlichen Übersetzung in der Kontaktsprache oft verborgene Pragmatik der konkret vorliegenden Lesart klarer zu machen. Im Moment ist vorgesehen, bei den slavisches Texten aus Griechenland neben der griechischen eine französische Übersetzung beizugeben, bei den moliseslavischen Texten neben der italienischen eine deutsche und bei den burgenlandkroatischen sowie den sorbischen Texten neben der deutschen eine französische.

Eine 7. Analyseebene besteht in Form von Annotierungen sonst kaum verständlicher kultureller Besonderheiten in Form von Fuß- oder Endnoten. Bei den ersten Texten im Korpus sollen verstärkt auch sprachliche Erklärungen hinzugefügt werden, die sowohl etymologischer, struktureller als auch kontaktlinguistischer Natur sein können.

Wie schon oben beschrieben wird jedem analysierten „Satz“ die direkt per Mausclick aktivierbare Tondatei der Originalaufnahme beigegeben, die einen Klangeindruck sowie einen Einblick in die Artikulationsweise der einzelnen Varietäten (bzw. des konkreten Sprechers) vermittelt, dem Nutzer aber insbesondere einen direkten phonetischen Zugang im segmentalen wie suprasegmentalen Bereich, unter Einschluß der Satzprosodie, ermöglicht.

5. Ausschnitt aus dem moliseslavischen Teilprojekt

5.1 Die Situation des Moliseslavischen

Das Moliseslavische steht seit der vor etwa 500 Jahren von der Hercegovina her erfolgten Einwanderung in Kontakt mit italienischen Varietäten, insbesondere dem lokalen molisanischen Dialekt und seit etwa 150 Jahren auch mit Regionalformen der italienischen Standardsprache. In „genetischer“ Hinsicht gehört es der mittelsüdslavischen (serbokroatischen) Dialektgruppe an und ist dem ikavischen Štokavisch zuzuordnen. Die nächstverwandte Standardsprache ist das im Alltag keinerlei Rolle spielende Standardkroatische, von dem es sich – abgesehen von Eigenentwicklungen des Standards – gerade durch die Ergebnisse des Jahrhunderts währenden romanischen Einflusses unterscheidet.

Moliseslavisch wird heute noch in drei Dörfern gesprochen, die durch unterschiedliche Dialekte gekennzeichnet sind: *Acquaviva Collecroce* (der traditionelle Hauptort), *Montemitro* (kleiner und konservativer) und *San Felice del Mo-*

lise (nur noch wenige ältere Sprecher).¹⁷ Die Unterschiede liegen vor allem auf der lautlichen Ebene, doch bestehen durchaus auch auffällige Differenzierungen im lexikalischen und grammatischen Bereich. Zu nennen ist hier etwa die Aussprache von unbetontem *e* und *o*, insbesondere im Wortauslaut, als *a* in Acquaviva und San Felice (moliseslavisches Akanje), im Gegensatz zu ihrer Erhaltung in Montemitro. In grammatischer Hinsicht ist etwa der konservative Dativ der Feminina auf *-i* in San Felice zu nennen, verglichen mit – lautlich zu erklärender – Nullendung in Montemitro und analogisch zum Akkusativ (oder zum Dativ der Maskulina) eingeführter Endung *-u* in Acquaviva. Beim Plusquamperfekt verwendet Acquaviva die Partikel *bi*, Montemitro *ba*, San Felice hat beides.

5.2. Anwendung des Analyseverfahrens

Bevor in der Folge eine moliseslavische Anekdote als konkretes Beispiel der Präsentation gegeben wird, soll zunächst das beschriebene Präsentationsverfahren, darunter insbesondere auch die gewählte morphologische Segmentierung, anhand kurzer Ausschnitte aus diesem Text erläutert werden. Die Numerierung der Ausschnitte entspricht der Position im nachfolgenden Gesamttext. Wie oben beschrieben, gilt folgende Abfolge der Analyseebenen (dabei stehen die aufeinander bezogenen Elemente in 2, 3 und 4 linksbündig untereinander):

1. Graphische Ebene (übliche Orthographie der Mikrosprache)
2. Lautliche Ebene mit phonetischer Notation der Morphemformen
3. Morphologisch-synthetische Glossierung (Lexembedeutung + Analyse)
4. Analytisch-morphosyntaktische Glossierung; sie entfällt, wenn im Satz keine analytischen Verbformen vorhanden sind
5. Kontaktsprachliche Übersetzung (hier italienisch)
6. Freie Übersetzung (hier deutsch)

„Satz“ 3 der Anekdote beginnt mit einem indefiniten Artikel, der in den Stamm *n-* und die Endung *-a* segmentiert wird. Hierauf beziehen sich die Glossierungen links bzw. rechts des Segmentierungsstrichs. Da es sich um eine grammatische Form handelt, wird statt einer lexikalischen Bedeutung die Kennzeichnung „ART.INDF“ angegeben:¹⁸

¹⁷ Vgl. hierzu die Kurzporträts des Moliseslavischen in Breu (2008b) und in den Internetpräsentationen <http://www.uni-konstanz.de/FuF/Philo/Sprachwiss/slavistik/acqua/index.htm> sowie http://www.uni-koeln.de/gbs/Berichte/Breu_Moliseslawisch.pdf.

¹⁸ Das Moliseslavische hat ebenso wie die obersorbische Umgangssprache einen indefiniten Artikel entwickelt, jedoch im Gegensatz zu letzterer keinen definiten Artikel. Aufgrund der Opposition zum morphologisch expliziten indefiniten Artikel wird Artikellosigkeit von den durchgehend zweisprachigen Moliseslaven grundsätzlich mit dem italienischen definiten Artikel gleichgesetzt. In Breu (2008a) wurde hierfür der Terminus „definiten Nullartikel“

3. Na dan sa vidija utra na štip

n-a = 'da:n s-a = 'vidi-ja 'utra n-a = 'štip (nol)
 ART.INDF-ACC.SG.M day.ACC.SG.M be.PRS-1SG see-PTCP.SG.M in ART.INDF-ACC.SG.M sideboard.ACC.SG.M
 AUX.PRF(1) PRF(1)

I: Un giorno ho visto in una *credenza*

D: Eines Tages sah ich auf einer Anrichte

Der proklitische Charakter des indefiniten Artikels im Moliseslavischen wird durch das rechts von der Wortform stehende Gleichheitszeichen symbolisiert. Die für flektierende Sprachen typischen kumulierten Morphembedeutungen ohne entsprechende formale Segmentierung werden in der heute üblichen Weise durch einen Punkt zwischen den grammatischen Angaben symbolisiert. Im Fall des endungslosen Substantivs *dan* fehlt natürlich eine Segmentierung zwischen Stamm und Endung, sowie entsprechend auch zwischen der lexikalischen und der grammatischen Glossierung. Die proklitische Auxiliarform¹⁹ *sa* wird in einen lexikalischen Stamm *s-* mit der Bedeutung 'sein' und die Endung *-a* segmentiert, die Träger der grammatischen Bedeutung PRS.1SG ist.²⁰ Es folgt das maskuline *l*-Partizip *vidija* zu *vit* 'sehen'. Eine Segmentierung ist hier – anders als etwa bei der femininen Form *vidi-l-a* 'see-PTCP-SG.F' – nur zwischen lexikalischem Stamm (genauer Infinitivstamm) und partizipialer Endung insgesamt

(нулевой определенный артикль) verwendet. Hier wird auch eine kontaktlinguistische Erklärung für die Durchbrechung des linguistischen Universales, nach dem sich vor einem indefiniten Artikel immer erst ein definitiver Artikel entwickelt, gegeben. In der Internetpräsentation wird auf die explizite Angabe eines definiten Nullartikels \emptyset einschließlich seiner Glossierung verzichtet.

¹⁹ Im Gegensatz zu anderen Varietäten auf štokavischer Basis, insbesondere auch zu den bosnisch-kroatisch-serbischen Standardvarietäten, hat das Moliseslavische, wie gesagt, die phonologische Regelung einer „2. Position im Satz“ (Wackernagels Gesetz) durch eine Proklise (im Imperativ Enklise) mit dem Verb als Ankerform ersetzt. Es handelt sich dabei grundsätzlich um eine kontaktsprachliche Anpassung ans Italienische. Im Fall des Zusammentreffens mehrerer Klitika weicht ihre relative Anordnung (Auxiliarformen, Pronomina, Reflexivum, Negation) aber vom italienischen Modell ab. Eine genauere Untersuchung dieser partiellen Resistenz gegen das kontaktsprachliche Modell ist für die nächste Zeit geplant.

²⁰ Die Grundprinzipien für die Segmentierung sind die Distribution und die Opposition der mutmaßlichen morphologischen Einheiten. Sie führen zu der hier angeführten Segmentierung *s-a*, da etwa in der entsprechenden Pluralform *s-ma* 'wir sind' die Endung *-ma* für alle Verben im Präsens, Imperfekt, Imperativ und Konditional gilt, so daß *s-* sich als alleiniger Träger der lexikalischen Bedeutung 'sein' in der 1. Person Plural ergibt und dann in analoger Weise auch in der 1. Person Singular angesetzt werden muß, trotz der hieraus folgenden singulären Endung *-a*, die aber mit der Endung *-i* in der 2. Person Singular *si* (*s-i*) eine Parallele besitzt. Im ganzen Plural gelten dagegen auch sonst übliche Endungen, neben *s-ma* auch in *s-ta* (2PL) UND *s-u* (3PL). Nicht segmentierbar ist andererseits die Form der 3. Person Singular Präsens *je* mit der ungegliederten Glossierung *be.PRS.3SG*.

möglich.²¹ Das Lehnwort *štip* 'Anrichte, Büffet' ist in der phonetischen Analysezeile kursiv gesetzt, das in der italienischen Übersetzungszeile entsprechende *credenza* zwischen Sternchen.²² In der 4. (morphosyntaktischen) Analysezeile werden die beiden Bestandteile des Perfekts *sa vidija* in der oben beschriebenen Weise per Indizierung zusammengruppiert.

Der nachfolgende Teilsatz 4 exemplifiziert an Hand der Position des Adjektivs *zelenu* die kontaktbedingte Nachstellung von objektiv-qualitativen Attributen im Moliseslavischen, im Gegensatz zu subjektiven (ausschmückenden) wie hier *lipu*. Die lautliche Repräsentation des letztgenannten Adjektivs zeigt ebenso wie [ʃti:p] in 3. die Existenz von (phonematischen) Langvokalen. Aus der Gegenüberstellung der lautlichen Repräsentation von *lipu* mit fallendem Akzent zu dem steigenden Akzent von *zelenu* ist das gewählte Verfahren für die Markierung des Tonalakzents zu erkennen, und zwar senkrechter Strich vor der zu betonenden Silbe mit (fallendem) Normalakzent vs. Akut auf dem Vokal mit Steigton. Die Notation mit Akut symbolisiert in gewisser Weise auch die tatsächliche Realisierung des Steigtons im Moliseslavischen, nämlich in der Regel in Form einer Doppelbetonung des gekennzeichneten und des nachfolgenden Vokals (Zweisilbenakzent).²³ Wie in diesem Beispiel zu sehen, wird das Genus neben Kasus und Numerus auch beim Substantiv, also bei der Kongruenzquelle angegeben, da die moliseslavische Substantivdeklinations direkt vom Genus abhängig ist:²⁴

²¹ Das *-a* entspricht hier der generellen Entwicklung eines auslautenden historischen **-l* im Moliseslavischen, die, im Gegensatz etwa zur standardkroatischen Entwicklung von **-l > -o*, auf die Herkunft der Moliseslaven aus dem herzegovinischen Neretvagebiet verweist. Es liegt also kein Akanje *-o > -a* vor, wie das Auftreten derselben Form *nosija* auch im akanjlosen Dialekt von Montemitro beweist. Das *-j-* ist ein Hiatusilger und wird bei der morphologischen Analyse zur Endung geschlagen, weil sonst ein eigener Alternationsstamm angenommen werden müßte, statt eines einheitlichen Stammes *nosi-* für alle Formen.

²² Die tatsächliche Quelle, die anders als bei Entlehnungen aus dem Standarditalienischen nicht direkt aus der Präsentation ablesbar ist, ist hier das molisanisch-dialektale Substantiv *štipa*, was durch den Zusatz „(mol)“ gekennzeichnet ist. Diese Kennzeichnung wird nur bei substantiellen lexikalischen Abweichungen der Entlehnung von der italienischen Übersetzung vorgenommen, da etwa eine abweichende Lautung nicht nur durch dialektale Entlehnung bedingt sein kann, sondern auch durch sekundäre Adaptation während des Entlehnungsprozesses.

²³ Anders als in den bosnisch-kroatisch-serbischen Standardvarietäten ist im Moliseslavischen auch Steigton auf der letzten oder einzigen Silbe des Wortes möglich, sofern der betreffende Vokal lang ist, z.B. [kú:p] 'kauf!' : [ʃku:p] 'sammle!', begründet durch den Abfall von Vokalen, insbesondere von kurzem auslautendem *-i*. In diesem Fall wird der Steigton auf einem einzigen Vokal realisiert. Zu diesen und anderen phonologischen Regeln des Moliseslavischen vgl. Breu (1999).

²⁴ Die „Genusflexion“ des Substantivs wurde durch den Verlust des substantivischen Neutrums und den Deklinationswechsel der alten femininen *i*-Deklination erreicht, wobei im letzteren

4. nu lipu butilju zelenu

n-u = 'li:p-u bútik -u zélen-u

ART.INDF-ACC.SG.F beautiful-ACC.SG.F bottle-ACC.SG.F green-ACC.SG.F

I: una bella *bottiglia* verde

D: eine schöne grüne Flasche

Teilsatz 5 enthält ein (bezogen auf den Gesamtsatz) zweites Perfekt mit entsprechender Indizierung und zeigt außerdem die Segmentierung des Imperfekts bei *biša* in Stamm, Imperfektsuffix und die Endung für Person und Numerus. Dieses Verb stellt den einzigen Fall eines Imperfekts mit dem vokallosen Suffix *-š-* in der 2. und 3. Person Singular dar, in Alternation mit *-h-* in allen anderen Personen, z.B. *bi-h-u* : be-IPRF-1SG. Alle anderen Verben zeigen ein mit *a* (je nach den Akzentverhältnissen kurz oder lang) erweitertes Suffix, z.B. *nos-ah-u* : 'carry-IPRF-1SG', *nos-aš-a* 'carry-IPRF-3SG'. Das Verbum *bit* 'sein' bildet auch im Präsens eine Ausnahme. Wie vorher beschrieben, kann etwa in der 1SG des Präsens der Stamm *s-* nicht weiter segmentiert werden, während es sonst meist möglich ist, dem Präsens ein Suffix zuzuordnen, z.B. *nos-i-m* : 'carry-PRS-1SG', *jim-a-m* : 'have-PRS-1SG'. Eine Ausnahme bildet sonst lediglich die 3PL, soweit statt eines eindeutigen Präsenssuffixes Kumulation mit der Personen-Numerus-Endung vorliegt, z.B. in der *i*-Konjugation bei *nos-u* : 'carry-PRS.3PL, aber nicht in der *a*-Konjugation, etwa bei *jim-a-ju* : 'have-PRS-3PL.²⁵ Dieses Beispiel zeigt im übrigen, daß die Konjunktion *ka* 'daß' als *complementizer* glossiert wird:²⁶

5. a sa mislija ka biša likvor.

a s-a = 'misli-ja ka 'bi-š-a líkvor

and be.PRS-1SG think-PTCP.SG.M COMP be-IPRF-3SG liqueur.NOM.SG.M

AUX.PRF(2) PRF(2)

I: e ho pensato che fosse *liquore*.

D: und ich dachte, daß es Likör sei.

„Satz“ 9 zeigt, daß die Angabe der Perfektivitätsopposition (derivativen Aspektopposition), soweit nicht aspektuelle Präfixe oder Suffixe vorliegen, als Charakteristikum des Stammes notiert wird, hier durch die Angabe 'PFV' (perfektiv) bei dem Partizip *velena* 'vergiftet', imperfektiv wäre *velen-iva* 'poison-

Fall die betreffenden Substantive nach dem Genus ihrer italienischen Entsprechung auf die beiden übrigen Genera/Deklinationen verteilt wurden; vgl. Breu (2004).

²⁵ Die Abweichung der 3. Person Plural in der *i*- und der *e*-Konjugation von der besonders produktiven *a*-Konjugation wird zunehmend durch die analogische Übernahme der Endung der *a*-Konjugation beseitigt, vor allem in der *e*-Konjugation: *gred-u* ~ *gred-a-ju* 'go-PRS-3PL'. Die ursprüngliche Endung *-e* der *i*-Konjugation ist im Moliseslawischen sowieso schon durchgängig durch das aus den anderen Konjugationen stammende *-u* ersetzt.

²⁶ In anderen Beispielen fungiert *ka* als Relativpronomen, glossiert REL, sowie als kausale oder temporale Konjunktion, glossiert mit 'because' bzw. 'when'.

IPFV.PTCP.SG.M' mit dem Suffix *-iva* zur Bildung des Aspektstamms. Die Aspektzugehörigkeit wird, wie bereits festgestellt, nur bei paarigen Verben angegeben, da bei den unpaarigen Verben aspektuelle Informationen allein über die „flexivische“ Opposition von Imperfekt vs. Perfekt ausgedrückt werden, z.B. oben in Satz 3 bei *vidija*. In Fall von *velena* sind auch die morphologischen Informationen für das Partizip direkt dem Stamm zuzuordnen, da die ursprüngliche Endung mit diesem verschmolzen ist, im Gegensatz etwa zur femininen Form *velena-l-a* 'poison.PFV-PTCP-SG.F' oder der Pluralform *velena-l* 'poison.PFV-PTCP.PL', bei der aber andererseits infolge der Tilgung des auslautenden **-i* die Numerusangabe mit der Partizipialinformation verschmolzen ist. Auf der phonetischen Ebene ist in diesem Beispiel die Notation der auslautenden (stimmlosen) Flüstervokale mit untergesetztem Kringel zu erkennen, nämlich bei [*'misli-ja*]. Das Plusquamperfekt, das sich aus drei Elementen der synthetischen Analyseebene zusammensetzt, dem Auxiliar, der Partikel *bi* sowie dem Partizip, erscheint auf der analytischen Ebene mit der dreifachen Angabe '=PQP (1)':

9. a sa mislija ka sa bi sa velena.

a	s-a=	'misli-ja	ka	s-a=	bi=	sa=	veléna
and	be.PRS-1SG	think-PTCP.SG.M	COMP	be.PRS-1SG	PTL	REFL	poison.PFV.PTCP.NOM.SG.M
	AUX.PRF(3)	PRF(3)		AUX.PRF=PQP(1)	PST=PQP(1)		PRF=PQP(1)

I: e ho pensato che mi fossi *avvelenato*.

D: und ich dachte, daß ich mich vergiftet hätte.

Dem folgenden Ausschnitt ist die Behandlung des Konditionals zu entnehmen. Im Gegensatz zur teilweise homonymen Form des Plusquamperfekts ist *bi* im Konditional keine Partikel, sondern eine – im gegebenen Beispiel zufällig endungslose – Form in einem Flexionsparadigma – vgl. etwa *bi-ma* 'SBJV-1PL' – die wir als Konjunktivauxiliar auffassen. Der Konditional insgesamt ist eine analytische Form aus dem Auxiliar und dem Partizip, was in der morphosyntaktischen 4. Analysezeile in der bekannten Weise notiert wird:

11. ka bi bila dobra jizist na mala cukra.

ka	bi=	bi-l-a	dó:br-ǵ	jíz-is-t	n-a=	'mal-a	'tsukr -ǵ
COMP	SBJV.3SG	be-PTCP-SG.N	good-NOM.SG.N	PFV-eat-INF	ART.INDF-ACC.SG.N	little-ACC.SG.N	sugar-GEN.SG.M
		AUX.COND(1)	COND(1)				

I: che sarebbe stato bene mangiare un po' di *zucchero*.

D: daß es gut wäre, ein bißchen Zucker zu essen.

Es ist hier natürlich nicht möglich, auf alle Einzelfälle der keineswegs trivialen Segmentierungen und Glossierungen für die Gesamtdarstellung der moliseslavischen Texte einzugehen. Die beispielhaft gegebenen Interpretationen sollten

aber für das Verständnis der Analyse der im folgenden einkopierten Anekdote ausreichen.²⁷

Da diese Anekdote keine Fälle von *Code switching* enthält, soll das hierfür angewendete Verfahren an einem Beispiel aus einem anderen Text exemplifiziert werden. Er stammt aus dem Dialekt von Montemitro, was an dem fehlenden Akanje der Form *biše* (Acquaviva *biša*) gut zu erkennen ist. Die Kennzeichnung des *Code switching* besteht entsprechend dem oben beschriebenen Verfahren in einer Doppellinie statt der hier irrelevanten morphologischen Analyse, zusätzlich zu der bei allen Fremdelementen durchgeführten Kursivschreibung. Anders als bei integrierten Lehnwörtern bleibt hier außerdem statt einer phonetischen Umschrift die (italienische) orthographische Form bestehen:

Biše voda oš biše l'onore della persona.

'bi-ʃ-e 'vod-a oš 'bi-š-e l'onore della persona

be-IPRF-3SG water-NOM.SG.F and be-IPRF-3SG =====

I: C'era l'acqua e c'era *l'onore della persona*.

D: Da war das Wasser und da war die persönliche Ehre.

5.3. Eine moliseslavische Anekdote: *Der vergiftete Likör*

Im folgenden wird nun die gesamte Anekdote als Beispiel für die Internetpräsentation eines moliseslavischen Textes beim augenblicklichen Stand der Analyse wiedergegeben. Der Text stammt aus dem Dialekt von Acquaviva. Es han-

²⁷ Ergänzend sei noch darauf hingewiesen, daß beispielsweise auch die beiden Future des Moliseslavischen, das mit dem Auxiliar HABEN/MÜSSEN gebildete Neccessitativfutur und das mit WOLLEN gebildete Probabilitätsfutur zu den analytischen Verbformen zählen, z.B. *m-a-ma* 'must-PRS-1PL' + *po-kj* 'go-INF' bzw. *č-ma* 'want.PRS-1PL' + *po-kj* 'go-INF', auf der morphosyntaktischen Ebene klassifiziert als AUX.FUT.NEC(1) + FUT.NEC(1) bzw. AUX.FUT(1) + FUT(1), wobei die grundsätzlich mögliche Angabe PROB im zweiten Fall aus Redundanzgründen weggelassen wurde; zu den Modalfuturen im Moliseslavischen und zur kontaktbedingten Entwicklung dieser Differenzierung vgl. Brey (2011a, 156-158). Es ist auch noch darauf hinzuweisen, daß bei der Glossierung gewisser funktionaler Elemente, insbesondere der Personalpronomina in den Teilprojekten von *EuroSlav 2010* nicht vollkommen parallel verfahren wird. Zur besseren Lesbarkeit der Glossierungen wurde im deutschen Teilprojekt eine englische Lexemglossierung in der Nominativform mit zusätzlicher Angabe des konkreten Kasus, also etwa *hi* 'they.GEN' gewählt, während im französischen Teilprojekt einer Angabe der Art '3PL.GEN' der Vorzug gegeben wird. Die Lexemglossierung ist deshalb möglich, weil das Englische als Glossierungssprache hier mit seinen Formen 'he, she, it, they' lexikalisch zu den Mikrosprachen paßt. Der Fall liegt beispielsweise bei den Demonstrativpronomina anders, wo das Moliseslavische eine deiktische Dreifachdifferenzierung aufweist, während im Englischen nur *that* und *this* zur Verfügung stehen. Deshalb wird hier in beiden Teilprojekten eine vollständig grammatische Notation gewählt, z.B. molisesl. *ov-o* 'DEM.PROX-NOM.SG.N', *to* 'DEM.MID-NOM.SG.N' und *on-o* 'DEM.DIST-NOM.SG.N'. Dagegen ist bei *vak-o*, *tak-o*, *nak-o* wiederum eine gemischte Wiedergabe sinnvoll: 'so-PROX', 'so-MID', 'so-DIST'.

delt sich um einen männlichen Muttersprachler, geboren 1937, beide Eltern Moliseslaven, der die Sprache auch im Alltag verwendet. Die Aufnahme wurde 2002 mit einem Sony DAT Rekorder erstellt und anschließend über die Soundkarte auf den PC in eine wav-Datei übertragen.²⁸

Zusätzliche Erläuterungen werden in der Internetpräsentation auf konkret verweisende Fuß- oder Endnoten verteilt. In der Papierform nicht reproduzierbar ist die im Browser vorgesehene Möglichkeit der Reduktion durch Wegklicken einzelner Analysezeilen, um eine größere Übersichtlichkeit zu erreichen. Beispielsweise wird es möglich sein, nur die graphische Ebene und eine Übersetzungszeile auszugeben. Außerdem kann die lautliche Realisierung dazugeschaltet werden (Lautsprechersymbol).²⁹

Zur Übersicht geben wir hier zunächst den Gesamttext *en bloc* wieder, zusammen mit der freien deutschen Übersetzung:

Nu votu bihu di moj cila pop. Jimahu na desa godišť. Na dan sa vidija utra na štip nu lipu butilju zelenu a sa mislija ka biša likvor. Naka biša. Ma kada sa ga sa napija, sa čuja ka mi goraša u grla a sa mislija ka sa bi sa veleno. Za kondravelena sa mislija ka bi bila dobra jizist na mala cukra. A sa jizija kvaš nu dzukerjeru ndzan. Dan dõp moja nona na storcu je rekla: «Ja učer sa bi napunila dzukerjeru, danas je kvaš prazna!» Ja sa čija skuzu, ka nisa bi čuja aš nisa bi kapija.

Einmal war ich bei meinem Onkel dem Pfarrer. Ich war etwa zehn Jahre alt. Eines Tages sah ich auf einer Anrichte eine schöne grüne Flasche, und ich dachte, daß es Likör sei. So war es. Aber als ich ihn getrunken hatte, spürte ich, daß es mich in der Kehle brannte. Und ich dachte, daß ich mich vergiftet hätte. Als Gegenmittel, dachte ich, daß es gut wäre, ein bißchen Zucker zu essen. Und ich aß fast die ganze Zuckerdose. Am Tag danach sagte meine Oma bei Tisch: „Ich habe gestern die Zuckerdose aufgefüllt, heute ist sie fast leer!“ Ich tat so, als hätte ich nicht gehört und nicht verstanden.

1. Nu votu bihu di moj cila pop.

n-u = 'vot -u 'bi-x-u di moj tsíl -a 'pop

ART.INDF-ACC.SG.F time-ACC.SG.F be-IPRF-1SG where my.NOM.SG.M uncle-NOM.SG.M priest.NOM.SG.M

I: Una *volta* ero da mio *zio* prete.

D: Einmal war ich bei meinem Onkel dem Pfarrer.

²⁸ Dieselbe Anekdote wurde mit diesem Sprecher noch in einer etwas längeren zweiten Variante aufgenommen. Diese Version ist im 1. Teilband einer auf mehrere Bände angelegten moliseslavischen Textsammlung abgedruckt, vgl. (Breu 2011b, 298–299), allerdings ohne begleitendes Tonmaterial.

²⁹ Die vorliegende Papierversion kann ja sowieso den entscheidenden Vorteil der Internetpräsentation, die Verknüpfung von Text und Ton nicht nachvollziehen. Um dem Leser vorab eine solche Möglichkeit zur Verfügung zu stellen, wird auf der folgenden Internetseite der vertonte Text zusammen mit der Analyse – aber noch ohne weitergehende Erläuterungen – versuchsweise eingestellt. Ein Verbleib unter diesem Link ist allerdings nicht garantiert: http://www.uni-konstanz.de/FuF/Philo/Sprachwiss/slavistik/acqua/versioni_e1.htm

2. Jimahu na desa godišť.

'jim-ax-u n-a= 'desa 'godift
 have-IPRF-1SG ART.INDF-ACC.SG.N ten year.GEN.PL.M

I: Avevo una decina d'anni.

D: Ich war etwa zehn Jahre alt.

3. Na dan sa vidija utra na štip

n-a= 'da:n s-a= 'vidi-ja 'utra n-a= 'ftip (mol)
 ART.INDF-ACC.SG.M day.ACC.SG.M be.PRS-1SG see-PTCP.SG.M in ART.INDF-ACC.SG.M sideboard.ACC.SG.M
 AUX.PRF(1) PRF(1)

I: Un giorno ho visto in una *credenza*.

D: Eines Tages sah ich auf einer Anrichte

4. nu lipu butilju zelenu

n-u= 'li:p-u bútil -u zélen-u
 ART.INDF-ACC.SG.F beautiful-ACC.SG.F bottle-ACC.SG.F green-ACC.SG.F

I: una bella *bottiglia* verde

D: eine schöne grüne Flasche

5. a sa mislija ka biša likvor.

a s-a= 'misi-ja ka 'bi-ŷ-a líkwo:r
 and be.PRS-1SG think-PTCP.SG.M COMP be-IPRF-3SG liqueur.NOM.SG.M
 AUX.PRF(2) PRF(2)

I: e ho pensato che fosse *liquore*.

D: und ich dachte, daß es Likör sei.

6. Naka biša.

ná:ka 'bi-ŷ-a
 so.DIST be-IPRF-3SG

I: Così era.

D: So war es.

7. Ma kada sa ga sa napija,

ma káda s-a= ga= sa= ná-pi-ja
 but when be.PRS-1SG it.GEN.SG REFL PFV-drink-PTCP.SG.M
 AUX.PRF(1) PRF(1)

I: ma quando l'ho bevuto,

D: Aber als ich ihn getrunken hatte,

8. sa čuja ka mi goraša u grla

s-a= 'tʃu-ja ka mi= gó:r-a:ŷ-a u 'grl-a
 be.PRS-1SG feel-PTCP.SG.M COMP I.DAT burn.IPFV-IPRF-3SG in throat-ACC.SG.M
 AUX.PRF(2) PRF(2)

I: ho sentito che mi bruciava in gola

D: spürte ich, daß es mich in der Kehle brannte,

9. a sa mislija ka sa bi sa velena.

a s-a = 'misli-ǰa ka s-a = bi = sa = *veléna*
 and be.PRS-1SG think-PTCP.SG.M COMP be.PRS-1SG PTL REFL poison.PFV.PTCP.NOM.SG.M
AUX.PRF(3) PRF(3) AUX.PRF = PQP(1) PST = PQP(1) PRF = PQP(1)

I: e ho pensato che mi fossi *avvelenato*.

D: und ich dachte, daß ich mich vergiftet hätte.

10. Za kondravelena sa mislija

za *kondrave'len* -ǰa s-a = 'misli-ǰa
 for antidote-GEN.SG.M be.PRS-1SG think-PTCP.SG.M
AUX.PRF(1) PRF(1)

I: Come *contravveleno* ho pensato

D: Als Gegenmittel, dachte ich,

11. ka bi bila dobra jizist na mala cukra.

ka bi = bi-l-a dó:br-ǰa jíz-is-t n-a = 'mal-a 'tsukr -ǰa
 COMP SBJV.3SG be-PTCP-SG.N good-NOM.SG.N PFV-eat-INF ART.INDF-ACC.SG.N little-ACC.SG.N sugar-GEN.SG.M
AUX.COND(1) COND(1)

I: che sarebbe stato bene mangiare un po' di *zucchero*.

D: daß es gut wäre, ein bißchen Zucker zu essen.

12. A sa jizija kvaš nu dzukerjeru ndzan.

a s-a = jíz-i-ja 'kwaš n-u = *dzuker'jer* -u ndza:n
 and be.PRS-1SG PFV-eat-PTCP.SG.M almost ART.INDF-ACC.SG.F sugar bowl-ACC.SG.F whole
AUX.PRF(1) PRF(1)

I: E ne ho mangiato *quasi* una *zuccheriera* intera.

D: Und ich aß fast die ganze Zuckerdose.

13. Dan dòp moja nona na storcu je rekla:

dan 'dɔp moj-a 'non -ǰa na stó:rts-u je = 'rek-lǰa
 day after my-NOM.SG.F grandmother-NOM.SG.F on table-ACC.SG.F be.PRS.3SG say-PTCP.SG.F
AUX.PRF(1) PRF(1)

I: Il giorno *dopo* mia *nonna* a tavola ha detto:

D: Am Tag danach sagte meine Oma bei Tisch:

14. «Ja učer sa bi napunila dzukerjeru,

ja útjer s-a = bi = ná-puni-l-a *dzuker'jer* -u
 I.NOM yesterday be.PRS-1SG PTL PFV-fill-PTCP-SG.F sugar.bowl-ACC.SG.F
AUX.PRF = PQP(1) PST = PQP(1) PRF = PQP(1)

I: «Io ieri avevo riempito la *zuccheriera*,

D: „Ich habe gestern die Zuckerdose aufgefüllt,

15. danas je kvaš prazna!»

dánas je 'kwa:f 'pra:zn-a
 today be.PRS.3SG almost empty-NOM.SG.F

I: oggi è *quasi* vuota!»

D: heute ist sie fast leer!«

16. Ja sa čija skuzu,

ja s-a = 'tʃi-ja 'sku:z -u
 I.NOM be.PRS-1SG make-PTCP.SG.M excuse-ACC.SG.F
 AUX.PRF(1) PRF(1)

I: lo ho fatto *finta*

D: Ich tat so

17. ka nisa bi čuja aš nisa bi kapija.

ka 'ni-s-a bi = 'tʃu-ja aʃ 'ni-s-a bi = 'kapi -ja
 COMP NEG-be.PRS-1SG PTL hear-PTCP.SG.M and NEG-be.PRS-1SG PTL understand-PTCP.SG.M
 AUX.PRF = PQP(1) PST = PQP(1) PRF = PQP(1) AUX.PRF = PQP(2) PST = PQP(2) PRF = PQP(2)

I: che non avevo sentito e che non avevo *capito*.

D: als hätte ich nicht gehört und nicht verstanden.

6. Ein Text aus Liti (Griechenland): *Der Hirte und sein Schatten*

Zum Abschluß soll nun noch ein Text in der slavischen Varietät von Liti (Griechenland, etwa 10 km nördlich von Saloniki gelegen) vorgestellt werden. In diesem Fall geht es – anders als bei dem noch von Sprechern mittleren Alters im Alltag benutzten Moliseslavischen – um die Bewahrung des Erbes einer unmittelbar vor dem Sprachtod stehenden slavischen Varietät. Nur ältere Sprecher über 70, deren Hauptsprache Griechisch ist, sprechen diese Varietät noch fließend, jüngere nur noch als *Semispeaker* oder überhaupt nicht mehr. Die für *EuroSlav 2010* angefertigten Aufnahmen gehören zu den letzten überhaupt noch möglichen.

Wiederum ist die Präsentation als *work in progress* zu verstehen. Entsprechend dem oben Gesagten, bestehen verglichen mit dem moliseslavischen Beispieltext zwar einige Unterschiede in den konkreten Segmentierungen und Glossierungen,³⁰ es handelt sich aber insgesamt um dasselbe Präsentationsverfahren. Die hier abgedruckte Geschichte wurde von einer 1924 geborenen Frau erzählt

³⁰ Ein konzeptioneller Unterschied ergibt sich daraus, daß die französische Seite bei der Glossierung mit Markiertheiten arbeitet. Insbesondere werden die Glossierungen PRS beim Verb und SG beim Substantiv weggelassen, da sie aus der Abwesenheit anderer Tempus- bzw. Numerusglossierungen erschlossen werden können. Die deutsche Seite notiert die Grammembedeutungen, wie an dem moliseslavischen Beispiel leicht erkennbar, auch in diesem Fall. Beide Verfahren sind natürlich gleichermaßen aussagekräftig, es geht lediglich um eine unterschiedliche Gewichtung von Darstellungsökonomie vs. Explizitheit.

und im Jahr 2004 aufgenommen. Die Übersetzungssprachen sind entsprechend den Regeln des Projekts hier Griechisch als unmittelbare Kontaktsprache und Französisch. Zum leichteren Verständnis geben wir den Gesamttext aber im folgenden zunächst in einer deutschen Übertragung wieder, und zwar ohne den Originaltext, da die aus Griechenland stammenden Texte sowieso nur in Lautschrift präsentiert werden. Diese lautliche Repräsentation des Originaltextes kann aus der unten anschließenden Wiedergabe der Einzelsätze entnommen werden:

Also, es war einmal der (=ein) Schäfer, der seine Schafe weidete, weit weg. Dann eines Tages machte sie, machte seine Mutter Krapfen, um zu gehen und sie dem Schäfer zu bringen. Es war nachts, es gab Mond, den Mond, und er, er ging so und hinter ihm ging sein Schatten, der Schatten – hast du verstanden? – sein Schatten. „Ah“ sagt er „jemand folgt mir“. Er sagt: „Ha, ich werde dir einen Krapfen geben.“ Er wirft ihm einen hin, geht weiter: er folgt mir immer noch! Wieder wirft er welche hin, ah, bis sie ihm ausgehen. „Ha, was soll ich dir jetzt geben, ich habe sie alle verbraucht!“

Der vorliegende Text enthält zufällig keine analytischen Verbformen, weswegen die analytische (morphosyntaktische) Analysezeile hier durchgehend entfällt:

1. pæk i'dno 'vreme i'dno vəkət

pæk i'dno 'vreme i'dno vəkət

INTERJ one time.N one time.M (tur)

G: Μια φορά κι έναν *καιρό*

F: Il était une fois

2. tʃu'banat 'pasə'ʃe 'voftsete di'leko

tʃu'ban-at 'pasə-ʃe 'vofts-e-te di'leko

shepherd.M (tur) -ART.M graze-IPRF.3SG sheep-PL-ART.PL far

G: ένας *τσομπάνος* έβοσκε τα πρόβατα πέρα.

F: un berger qui faisait paître ses moutons, au loin.

3. i kini'sa i'dno den tʃi'ni la'ŋgide tʃi'ni 'maika mu

i kini'sa i'dno den tʃi'ni la'ŋgid-e tʃi'ni 'maika = mu

and get.going.AOR.3SG one day.M do.AOR.3SG fritter-PL do.AOR.3SG mother.F 3SG.M.DAT

G: και *κίνησε* μια μέρα η μητέρα του, του έφτιαξε τηγανήτες

F: Puis un jour sa mère fit des beignets

4. da poj da mu du'nese na tʃu'banat

da poj da mu = du'nese na tʃu'ban-at

COMP go.PFV.3SG COMP 3SG.M.DAT PFV-bring.3SG to shepherd.M (tur)-ART.M

G: και τις πήγε στον *τσομπάνο*.

F: et alla en apporter au berger.

5. 'beʃe 'noʃke 'jimaʃe miʃi'tʃina miʃ'tʃinta on 'xodiʃe ta'ka

'be-ʃe 'noʃke 'jima-ʃe miʃi'tʃina miʃ'tʃin-ta on 'xodi-ʃe ta'ka
 be-IPRF.3SG at.night have-IPRF.3SG moon.F moon.F-ART.F DEM.DIST.M walk-IPFV-IPRF.3SG like.this

G: 'Htane nύχta, είχε φεγγάρι, κι εκείνος πήγαινε έτσι

F: C'était la nuit, la lune brillait, et lui, il marchait ainsi

6. i uto'zat 'xodiʃe 's'aŋkata mu 's'aŋkata ra'zbra 's'aŋkata mu

i uto'zat 'xodi-ʃe 's'aŋka-ta = mu 's'aŋka-ta ra'zbra 's'aŋka-ta = mu
 and from.behind walk-IPFV-IPRF.3SG shadow.F-ART.F 3SG.M.DAT shadow.F-ART.F understand.PFV.AOR.2SG shadow.F-ART.F 3SG.M.DAT

G: και πίσω του πήγαινε η σκιά του, κατάλαβες; Η σκιά του.

F: et derrière lui marchait son ombre, tu as compris ? Son ombre.

7. a 'veʎe to mi 'sʎade 'veʎe xa da ti daj i'dna la'ŋgida

a 'veʎe to mi = 'sʎade 'veʎe xa da ti = daj i'dna la'ŋgida
 INTERJ say.3SG DEM.MID.N 1SG.DAT follow.3SG say.3SG INTERJ COMP 2SG.DAT give.PFV.1SG one.F fritter.F

G: Α! Λέει, κάποιος με ακολουθεί. Λέει: "Άντε, ας σου δώσω μια τηγανήτα!"

F: Ah, dit-il : "Quelqu'un me suit". Il dit : "Tiens, je vais lui donner un beignet."

8. mu 'fərle i'dna 'xode po 'naŋka pak mi 'sʎade

mu = 'fərle i'dna 'xode po 'naŋka pak mi = 'sʎade
 3SG.M.DAT throw.3SG one.F walk-IPFV.3SG more far again 1SG.DAT follow.3SG

G: Του ρίχνει μια και προχωράει παραπέρα. Ακόμα με ακολουθεί!

F: Il lui en jette un et il va plus loin : on me suit toujours !

9. pak mu 'fərle e dur gi biti'sa

pak mu = 'fərle e dur gi = biti'sa
 again 3SG.M.DAT throw.3SG INTERJ until 3PL.ACC finish.AOR.3SG (tur)

G: Πάλι, του ρίχνει μέχρι που του *τελειώνουν* οι τηγανήτες.

F: De nouveau, il lui en jette jusqu'à ce qu'il n'ait plus de beignet.

10. e 'ʃo da ti daj tsa'ɣa i biti'sax 's'ite

e 'ʃo da ti = daj tsa'ɣa i = biti'sax 's'ite
 INTERJ what COMP 2SG.DAT give.PFV.1SG now 3PL.ACC finish.AOR.1SG (tur) all

G: "Τι να σου δώσω τώρα; Τις *τελείωσα* όλες!"

F: "Mais alors, que vais-je te donner maintenant, je les ai tous jetés !"

7. Zusammenfassung und Ausblick

Das deutsch-französische Gemeinschaftsprojekt *EuroSlav 2010* hat zum Ziel, ein Korpus mündlicher Texte in slavischen Mikrosprachen, die sich in nichtslavischsprachigen Ländern in einer totalen Sprachkontaktsituation befinden, im Internet zur Verfügung zu stellen: Moliseslavisch (Italien), obersorbische Umgangssprache (Deutschland), Burgenlandkroatisch (Österreich) sowie die Varietäten von Liti und Chrisa (Griechenland). Die Texte werden zusammen mit den

originalen Tonaufnahmen zur Verfügung gestellt, die im Internetbrowser direkt abspielbar sein sollen. Die Art der Aufbereitung wird der Forschung eine Fülle von Primärdaten zur Verfügung stellen, wobei die Texte in Form einer mehrstufigen Analyse Sprachwissenschaftlern aller Disziplinen zugänglich gemacht werden, insbesondere auch denjenigen, die bisher slavische Beispiele nur aus der Sekundärliteratur rezipieren konnten.

Von den Mitarbeitern des Projekts wird eine Präsentation der Ergebnisse anlässlich einer Tagung zum grammatischen und lexikalischen Sprachkontakt angestrebt, die in eine gemeinsame Publikation zu Parallelen und Unterschieden in der kontaktbestimmten Entwicklung der einzelnen slavischen Mikrosprachen münden soll. Ein fernerliegendes Ziel könnte die Ausweitung der elektronischen Korpora auf in dem vorliegenden Projekt noch nicht erfaßte slavische und auch nichtslavische Minderheitensprachen in alloglotter Umgebung sein, unter Einbindung eines größeren Netzes von Sprachkontaktinteressierten. Zu denken ist etwa an die slovenisch-basierten Minderheitensprachen in Österreich, Italien und Ungarn, an makedonisch-bulgarische Varietäten in Albanien und die serbisch-kroatischen Varietäten in Ungarn und Rumänien, aber auch an die albanischen Varietäten in Griechenland und Italien. Die im Laufe der Arbeit an dem vorliegenden Projekt gewonnenen Erfahrungen und das erarbeitete Know-how im technischen Bereich liefern hierfür wichtige Grundlagen.

Literatur

- Adamou, E. 2006. *Le nashta. Description d'un parler slave de Grèce en voie de disparition*. München: Lincom.
- Adamou, E. 2011. Temporal uses of definite articles and demonstratives in Pomak (Slavic, Greece). *Lingua* 121, 871–889.
- Bayer, M. 2006. *Sprachkontakt deutsch-slavisch: Eine kontrastive Interferenzstudie am Beispiel des Ober- und Niedersorbischen, Kärntnerslovenischen und Burgenlandkroatischen*. Frankfurt etc.: Peter Lang.
- Breu, W. 1999. Phonologie und Verbkonjugation im Moliseslavischen. In: Rathmayr, R., Weitlaner, W. (eds.), *Slavistische Linguistik 1998*. München: Otto Sagner, 47–76.
- Breu, W. 2004. Die Genuskategorie im Moliseslavischen. In: Okuka, M., Schweier, U. (eds.), *Germano-Slavistische Beiträge: Festschrift für Peter Rehder zum 65. Geburtstag*. München: Otto Sagner, 29–43.
- Breu, W. 2005. Verbalaspekt und Sprachkontakt. Ein Vergleich der Systeme zweier slavischer Minderheitensprachen (SWR/MSL). In: Kempgen, S. (ed.), *Slavistische Linguistik 2003*. München: Otto Sagner, 37–95.
- Breu, W. 2008a. Развитие систем артиклей при полном контакте славянских меньшинств с немецким и итальянским языками. In: Kempgen, S. et al.

- (eds.) *Deutsche Beiträge zum 14. Internationalen Slavistenkongress Ohrid 2008*. München: Otto Sagner, 75–88.
- Breu, W. 2008b. Moliseslawisch (Slawisch in Süditalien). In: Ammon, U., Haarmann, H. (eds.), *Wieser Enzyklopädie: Sprachen des europäischen Westens. Zweiter Band*. Klagenfurt: Wieser Verlag, 241–249.
- Breu, W. 2011a. Il verbo slavomolisano in confronto con altre lingue minoritarie: mutamento contatto-dipendente, resistenza e sviluppo autonomo. In: Breu, W. (ed.), *L'influsso dell'italiano sul sistema del verbo delle lingue minoritarie: Resistenza e mutamento nella morfologia e nella sintassi*. (=Diversitas Linguarum 29). Bochum: Brockmeyer, 149–184.
- Breu, W. 2011b. *Südslavisch unter romanischem Dach. Die Moliseslaven in Geschichte und Gegenwart im Spiegel ihrer Sprache. Teil 1. Texte gesprochener Sprache aus Acquaviva Collecroce*. (=Sagners Slavistische Sammlung 32). München: Otto Sagner.
- Drettas, G. 1990. Le dialecte bulgare-macédonien de Xr. (Edhessa, Grèce). Questions de typologie. *Bull. de la Soc. de Ling. de Paris* 85/1, 227–265.
- Heine, B., Kuteva, T. (2006). *The Changing Languages of Europe*. Oxford: University Press.
- Neweklowsky, G. 1978. *Die kroatischen Dialekte des Burgenlandes und der angrenzenden Gebiete*. Wien: Verlag der Österr. Akad. der Wissenschaften.
- Scholze, L. 2008. *Das grammatische System der obersorbischen Umgangssprache im Sprachkontakt: Mit Grammatiktafeln im Anhang*. Bautzen: Domowina.
- Szucsich, L. 2000. Das Burgenlandkroatische: Sprachwandel, Sprachverfall, Sprachverschiebung und Sprachassimilation. In: Zybatov, L.N. (ed.), *Sprachwandel in der Slavia: Die slavischen Sprachen an der Schwelle zum 21. Jahrhundert*. Frankfurt/M. etc.: Peter Lang, 853–875.

Rezime

Zajednički njemačko-francuski projekt *EuroSlav 2010* ima kao cilj objaviti na internetu korpus usmenih tekstova slavenskih mikrojezika koji se nalaze u situaciji totalnog jezičnog kontakta u zemljama koje su izvan slavenskog govornog područja: moliškoslavenski/moliškohrvatski (Italija), gornjolужиčkosrpski (Njemačka), gradišćanskohrvatski (Austrija), kao i varijeteti sela Liti i Chrisa (Grčka). Tekstovi će biti spojeni s originalnim tonskim snimkama i direktno dostupni za preslušavanje u web pregledniku. Mnogobrojne lingvističke analitične razine trebale bi tekstove učiniti pristupačnima lingvistima svih disciplina, pogotovo onima koji su do sada slavenske primjere mogli recipirati jedino iz sekundarne literature. Ovaj način obrade će istraživanju jezičnog kontakta dati na raspolaganje mnoštvo primarnih podataka.