



Fabio Tosques (Hrsg.)

20 Jahre digitale Sprachgeographie

Berlin 2014

Fabio Tosques (Hrsg.)

**20 Jahre digitale Sprachgeographie
Tagungsband**

Berlin 02. bis 03. November 2012

Humboldt-Universität zu Berlin

Institut für Romanistik

2014

ISBN: 978-3-00-046278-8

Inhaltsverzeichnis

<i>Vorwort</i>	I
Di Nunzio, Giorgio/Rabanus, Stefan: <i>Research on Geolinguistic Linked Data: The Test Case of Cimbrian Varieties</i>	1
Dicklberger, Alois: <i>Eine WebGIS-Anwendung für die Ortsamenforschung: Das Projekt „Die ältesten Ortsnamen im bayerisch baye- risch -tschechischen Grenzraum tschechischen Grenzraum tsche- chischen Grenzraum tschechischen Grenzraum tschechischen Grenzraum“</i>	9
Romano, Antonio/Contini, Michel: <i>L'Atlas Multimédia Prosodique de l'Espace Roman: uno strumento per lo studio della variazione geoprosodica</i>	27
Aurrekoetxea, Gotzon/Videgan, Charles: <i>Outils pour la géolinguis- tique automatisée</i>	53
Goebel, Hans/Schauer, Bernhard/Beer, Heidemarie/Staudinger, Ag- nes: <i>Reflexionen und Streiflichter zum Einsatz der EDV beim ALD-II (Ladinienatlas, 2. Teil)</i>	67
Bauer, Roland: <i>Zur Dialektometrisierung des ALD (I und II): Ein Arbeits- und Erfahrungsbericht 2000–2012</i>	95
Möller, Robert/Elspaß, Stephan: <i>Zur Erhebung und kartographi- schen Darstellung von Daten zur deutschen Alltagssprache onli- ne: Möglichkeiten und Grenzen</i>	121
Krefeld, Thomas/Lücke, Stephan: <i>Geoling 2.0 – Ein aktueller Bericht aus der Werkstatt der webbasierten Sprachgeographie</i>	133
Bibiri, Anca-Diana/Panaite, Oana/Turculeț, Adrian: <i>The Romanian Multimedia Prosodic Atlas (AMPROM)</i>	155
Diémoz, Federica/Kristol, Andres: <i>L'atlas linguistique audiovisuel du franco-provençal valaisan et les défis du polymorphisme</i>	163
Tosques, Fabio/Castellarin, Michele: <i>20 Jahre digitaler Sprachatlas VIVAio Acustico delle Lingue e dei Dialetti d'Italia (VIVALDI)</i>	185

Reflexionen und Streiflichter zum Einsatz der EDV beim ALD-II (Ladinienatlas, 2. Teil)

Hans Goebel & Bernhard Schauer & Heidemarie Beer & Agnes Staudinger, Universität Salzburg

1 Bericht des Projektleiters (*Hans Goebel*)

Wiewohl im Folgenden explizit zum 2. Teil des Ladinienatlases berichtet werden soll, muss dennoch immer auf frühere Phasen des Projektes ALD zurückgegriffen werden, da die hier besprochenen Probleme ohne genauere Kenntnisse des Gesamtprojektes ALD bzw. des Ineinandergreifens der beiden Projektteile nur teilweise verstanden werden können.

1.1 Entstehung und Realisierung des ersten Projektteiles (ALD-I)

Die Idee zu einem mit dem Kürzel ALD (für „Atlante linguistico ladino“) zu bezeichnenden Sprachatlas Nordostoberitaliens entstand im Juli 1972¹ an einem Wirtshaustisch in St. Martin / S. Martin de Tor im Gadertal. Damals traf ich erstmals – und zwar auf explizite Empfehlung durch verschiedene von mir zu sprachlichen Dingen befragte Einheimische – mit Lois Craffonara, dem späteren Gründer und langjährigen Direktor des Istitut ladin „Micurà de Rù“ (ILMdR), zusammen und konnte mit diesem nach einer sehr kurzen Zeit des ersten Beschnuppens eine weitestgehende Übereinstimmung zu diversen Streitpunkten der sattsam bekannten „Questione ladina“ herstellen. Die damals schon sehr detailliert ventilierte Idee war, diesen Sprachatlas nicht nur auf die eigentliche Ladinia zu beschränken, sondern das Untersuchungsgebiet so weit als möglich auf den ganzen Zentralbereich des *Rätoromanischen* (nach Th. Gartner) bzw. der *zona ladina* (nach G. I. Ascoli) auszudehnen und dabei generös das südliche Vorfeld einzubeziehen.

In den anschließenden Jahren (1973-1982) konnte ich im Rahmen einer Assistentenstelle an der Universität Regensburg durch zahlreiche Sondierungs-enquêtes vor Ort das Thema vertiefen und auch Ideen entwickeln, wie die Daten eines solchen Atlases zum einen im Feld gesammelt und zum anderen mit den damals immer operationeller werdenden Mitteln der „Computerei“ weiterverarbeitet und publiziert werden könnten. Angedacht war damals bereits aus arbeitspraktischen und ökonomischen Gründen eine Zweiteilung des Projekts (ALD-I: Phonetik und elementare Morphologie; ALD-II: „der Rest“); zudem stand auch schon ab 1978 das später explorierte Gebiet fest, das mit seinen 25.000 km² der Fläche Siziliens entspricht.²

Dieser sondierenden Phase folgte nach meiner Berufung an die Universität Salzburg (1982) jene der praktischen Umsetzung. In Zusammenarbeit mit Dieter Kattenbusch, meinem Nachfolger auf der Regensburger Assistentenstelle, und dessen Freund Thomas Stehl³ konnte das für den ALD-I vorzusehende Fragebuch erarbeitet werden. Lois Craff-

¹ In der „Introductio“ des ALD-I wird auf Seite VII dieses erste Zusammentreffen mit Lois Craffonara auf den Monat *September* des Jahres 1972 gelegt. Das ist nicht richtig. Ich konnte jüngst anhand der damals gemachten Photographien korrigierend feststellen, dass es sich um den Monat *Juli* gehandelt hat.

² Über alle Arbeitsfortschritte bei ALD-I und ALD-II wurde regelmäßig und sehr genau berichtet; dies geschah meistens in der Zeitschrift „Ladinia“. Eine vollständige Auflistung der betreffenden Publikationen findet man am Ende der „Introductiones“ von ALD-I und ALD-II und auf den Homepages zu ALD-I und ALD-II.

³ Thomas Stehl ist seit geraumer Zeit Professor für Romanische Philologie an der Universität Potsdam, Die-

fonara wiederum sorgte als seit 1977 amtierender Leiter des IIMdR für die finanzielle Bedeckung der im Herbst 1985 von Dieter Kattenbusch im Fassatal begonnenen regulären Enquêtes. Diese erstreckten sich dann – unter Beteiligung weiterer vier Exploratoren (darunter zwei Herren und zwei Damen) – bis zum Jahr 1992.

Der schon bei der Konzeption des Fragebuchs vorgesehene Einbezug der EDV konnte an der Universität Salzburg dank des Einsatzes zunächst von Roland Bauer und dann von Edgar Haimerl in sehr effizienter Weise realisiert werden. So wurde für die EDV-gestützte Eingabe der Transkripte der 217 Fragebücher bereits eine richtiggehende Erfassungs- und Korrekturschiene in Betrieb genommen, innerhalb derer alle zu verrichtenden Arbeitsschritte Gegenstand genauer Planung und Evaluierung waren. Edgar Haimerl hat dann seit den frühen 1990er Jahren ein exakt dazupassendes DOS-Programm namens *CARD* („*Cartography and Retrieval of Dialect Data*“) konzipiert und perfektioniert, womit nicht nur die Ersteingabe und alle nachfolgenden Korrekturen der Transkripte, sondern auch deren Einspeisung in eine relationale Datenbank und die Herstellung der eigentlichen Sprach-Karten sowohl für die Phase der Karten-Ausarbeitung und -Redaktion als auch für jene des eigentlichen Drucks bewerkstelligt werden konnten.

Dazu musste an der Salzburger ALD-Forschungsstelle namens „ALD-Archiv“ eine komplexe Kooperationskette organisiert werden, deren Glieder – bestehend aus philologisch und informatisch arbeitenden Mitarbeitern (*semper utriusque sexus*) – möglichst reibungslos und effizient zusammenarbeiten sollten, ja mussten. Die damit verbundenen fachlich-romanistischen, logistischen, menschlichen und – last but really not least – finanziellen Probleme konnten – *Deo maximas gratias!* – alle bemeistert werden, so dass der ALD-I im Jahr 1998 unter dem Schirm des Dr. L. Reichert-Verlags bei einer Stuttgarter Druckerei unter Verwendung der von uns gelieferten PDF-Dateien produziert und noch im selben Jahr in den Handel gebracht werden konnte.

Ich habe ganz bewusst auf eine Optimierung der drucktechnischen und buchbinderischen Qualität des Papier-Produkts ALD-I hingearbeitet, da schon damals klar war, dass ein derartiges *opus* die Eigenschaft *aere perennius* haben sollte, um deutlich länger (über)leben zu können als die parallel dazu in Umlauf gesetzten drei CD-ROM, die man leider – wiewohl (ebenso: *leider*) vorhersehbar – *tels quels* heute in keinem der üblichen PC mehr zum Laufen bringen kann.

Diese drei CD-ROM enthielten zunächst alle Daten des ALD-I in drei graphischen Versionen: in der vollen Lautschrift („*ALD-standard*“), in einer simplifizierten Version dazu („*ALD-light*“) und in normalen Buchstaben. Zusätzlich befand sich darauf ein von Edgar Haimerl geschaffenes Suchprogramm namens *IRS* („*Index Retrieval System*“), das es gestattete, die in *ALD light* gefassten Daten nach vom User einzugebenden Kriterien alternativ vom Anfang oder vom Ende eines Worts her zu klassifizieren. Es war das ein Tribut an die in Salzburg intensiv praktizierte Dialektometrie⁴ und die in ihrem Rahmen notwendige Zusammenfassung („Taxierung“) etymologisch identischer, lautlich aber variierender Formen in als linguistisch konvergent gedachten Gruppen („Taxatatareale“).

Auf in den frühen 1990er-Jahren in Salzburg vom Stapel gelassene EDV-technische Basteleien, die in Kooperation mit einer in Essen (Deutschland) angesiedelten Soft-

ter Kattenbusch dasselbe an der Humboldt-Universität Berlin.

⁴ Die Daten des ALD-I wurden am Beginn des laufenden Jahrtausends von Roland Bauer dialektometrisiert: cf. dazu Bauer 2009. Allerdings ist dabei m. W. diese Möglichkeit nicht oder kaum zum Einsatz gekommen. Überdies befanden sich auf den fraglichen CD-ROM auch animierte Demo-Sequenzen, die die Handhabung von IRS sehr anschaulich vorstellen.

warefirma⁵ begonnen und etappenweise perfektioniert wurden, geht eine weitere informatische Komponente zurück, die sich anfangs ebenso auf einer der drei zitierten CD-ROM befand und in späterer Folge auf eine eigene DVD überwechselte: der „Sprechende Sprachatlas“, kurz „der Sprechende“ genannt.

Die hinter dem Sprechenden stehende Idee war eine reine Linguisten-Geburt: es sollte dem *Ohr* des Fachmanns die Möglichkeit geboten werden, die lautlichen Äquivalente der auf den Karten festgehaltenen Transkriptionen in akustisch optimaler Form zu perzipieren. Und zwar in genau jener isoliert-diskreten Weise, in der sich die auf den Atlas-Karten befindlichen Transkriptionen dem *Auge* des Linguisten darbieten. Dahinter steckte also nicht nur ein *heuristisches*, sondern auch ein *sprachtheoretisches* Programm, das im Übrigen auch der geolinguistischen Gesamtkonzeption des ALD zu Grunde liegt. Dieser diente ab ovo keineswegs zur Erhebung völlig unbekannter oder besonders interessanter Materialien oder gar der Erkundung diastratischer Probleme (wie des Nebeneinanders von Basi-, Meso- und Akrolekten in den Köpfen der Einheimischen), sondern – unter Fortsetzung und Schärfung des von Jules Gilliéron bei der Schaffung des ALF (Datensammlung: 1897-1901, Publikation: 1902-1910) angewandten Prinzips – einer mit sprachlichen Mitteln (hier: exklusiv mittels der Elizitation von basilektalen Responses) vorzunehmenden Vermessung („Geodäsie“) eines bestimmten Raumes (hier: der 25 000 km² der ALD-Zone). Man kann daher ruhig zur Charakterisierung dessen, was hier geschehen ist und noch betrieben wird, das Etikett einer „Glotto-Geodäsie“ verwenden.⁶

Die angepeilte hohe akustische Qualität konnte nur über Nacherhebungen erreicht werden, die mit einem digitalen (und eben nicht mehr analogen!) Magnetophon⁷, einem exzellenten Mikrophon und im Rahmen einer maximal „entspannten“ Aufnahmesituation⁸ vorgenommen wurden. Angesichts des damit verbundenen hohen Aufwands wurde die dafür herangezogene Zone auf die 21 Messpunkte der brixnerisch-tirolischen Ladinia (= ALD-PP. 81-101) beschränkt.

In formaler Hinsicht entsprach die solcherart erhobene Datenmenge einer zweidimensionalen Matrix aus 21 Messpunkten und rund 900 Antwortreflexen, d. h. also aus rund 19 000 in Salzburg elektronisch aus den Aufnahmegesprächen herauszuschneidenden Segmenten, die nach erfolgtem Schnitt ihrerseits in eine eigene Datenbank eingespist wurden.

Die von wirklich aufopferungsbereiten Mitarbeitern (Brigitte Rührlinger und Slawomir Sobota) herauspräparierten 19 000 Tonsegmente sind noch immer hochaktuell: im Jahr 2005 wurden sie in eine dem Unternehmen VIVALDI entlehnte Netz-Technologie übernommen, wodurch der heute noch klaglos funktionierende „Netz-Sprechende“ entstand. In der zweiten Hälfte des Jahres 2012 hat sich Bernhard Schauer,

⁵ Diese stand und steht unter der Leitung des Trierer Computerlinguisten Reinhard Köhler.

⁶ Zur sprachlichen Aufdröselung der drei altgriechischen Bestandteile dieses Wortes: *glóttā* (etc.) „Sprache“, *geo-* „zur Erde gehörig“, *dáís* „Teil“, *dáiomai* „teilen“ (entspricht dem lateinischen Wort *pars*).

⁷ Damals handelte es sich um einen DAT-Rekorder. *DAT: Digital Audio Tape*.

⁸ Diese Aufnahmesituation war damit nicht das, was die traditionelle Dialektologie unter „natürlich“ versteht. Für mich als einen in den Dimensionen von Quantität und Messung denkenden Sprachgeographen hat das Konzept der Natürlichkeit keine wie immer geartete Relevanz. Keine Messung ist „natürlich“, wo immer auch gemessen wird. Überall wird anhand theoretisch vorfixierter Konzepte und auch unter „Molestierung“ der zu vermessenden Objekte beurteilt. Im vorliegenden Fall ist aus methodischer Hinsicht noch das Prinzip der „Interkomparabilität“ sehr wichtig. Die an den 217 Messpunkten gesammelten Daten müssen untereinander voll vergleichbar sein und damit über denselben messtheoretischen Status verfügen. Nur dadurch ist es möglich, bei der Auswertung der Sprachatlas-Daten (geolinguistische) Aussagen zu tätigen, die für alle Teile des Gesamtgebiets gleichermaßen relevant sind.

der Chef-Informatiker des Projektes ALD-II, mit einem eigens dazu angeheuerten Ferrialpraktikanten dieser Daten mit dem Ziel angenommen, einen neuen netz- und DVD-basierten „Sprechenden“ zu kreieren: der heute dazu schon vorliegende Prototyp ist sehr vielversprechend und wird bald in seine definitive Form übergehen.

Der „Sprechende“ ist seit 1998/1999 mit seinem gerade für Ladiner sehr attraktiven Inhalt und seinen ausgeklügelten Funktionen immer wieder in Unterricht, im Museumsbetrieb und in der Forschung zur Anwendung gekommen.

In toto dauerten die Arbeiten zum ALD-I also rund 13 Jahre (1985 bis 1998), innerhalb derer ein organismusähnliches System bis zur Erreichung des gesteckten Zieles unter Dauerspannung gehalten werden musste.

1.2 Entstehung und Realisierung des zweiten Projektteiles (ALD-II)

Im Jahr 1998 standen mir als Projektleiter angesichts der herrschenden Gesetzeslage, die meine Emeritierung ab dem 1.10.2012 vorsah, noch rund 14 Jahre an verplanbarer Arbeitszeit zur Verfügung. Daher musste hinsichtlich der ein zweites Mal zu durchlaufenden Arbeitsetappen (Erstellung des Fragebuchs, Probe-Enquêtes, reguläre Feldarbeit, Aufbau einer neuen EDV-Schiene, EDV-gestützte Erfassung der gesammelten Daten etc.) ohne Zeitverlust gehandelt werden.

Dabei kam es, was auch angesichts der finanziellen Rahmenbedingungen nicht verwunderlich ist, zur Bildung eines personell weitestgehend neu zusammengesetzten Teams, in dem nur wenige Köpfe schon beim ALD-I mitgearbeitet hatten: zu diesen zählten neben Edgar Haimerl (der durch seinen erneuten Einsatz einen im Jahre 2001 mit einer externen Programmierfirma erlittenen EDV-technischen Schiffbruch reparierte) nur noch die beiden (steirischen) Exploratoren Helga Böhmer und Tino Szekely. Ähnlich wie beim ALD-I kamen die anderen Mitarbeiter aus verschiedenen Ecken Österreichs (vornehmlich aus Salzburg), Deutschlands und Italiens.

Bei der Erstellung des Fragebuchs und den damit in Ladinien durchgeführten Explorationen tat sich ganz besonders der aus Enneberg (= ALD-P. 81) stammende und später in Innsbruck und Eichstätt zum Romanisten ausgebildete Linguist Paul Videsott⁹ hervor. Unter den Exploratoren nahm ganz unzweifelhaft die in weiterer Folge mehrfach zitierte Brigitte Rührlinger auch deshalb eine Sonderstellung ein, weil sie nach der Exploration von 32 Ortschaften noch zahlreiche Sonderaufträge für den ALD-II übernommen und bravourös erledigt hat, die sich von Nachforschungen auf Friedhöfen, über Nachenquêtes bis hin zu höchste Präzision erfordernden Tagging-Arbeiten an den Sound-Daten des ALD-I erstreckten: auch dazu wird man in der Folge (v. a. im Abschnitt 3.) Näheres lesen können. Br. Rührlinger wurde für die Belange des ALD im Rahmen des ersten von mir zu diesem Thema an der Universität Salzburg gehaltenen Seminars „entflammt“ und hat dem Gesamtprojekt bis heute unverbrüchlich die Treue gehalten.

Während beim ALD-I die insgesamt 217 Enquêtes in den Händen von fünf Exploratoren lagen, wurde die analoge Arbeit mit dem nunmehr 1 063 z. T. recht komplexe Fragen enthaltenden Questionnaire von insgesamt zehn Exploratoren (davon sechs Männer und vier Frauen) erledigt. Die Regel-Enquêtes fanden zwischen 2001 und 2007 statt und umfassten somit sieben Jahre.

Insgesamt wurden dabei 833 Gewährspersonen befragt (und deren Äußerungen akustisch dokumentiert), denen beim ALD-I nur 488 Informanten gegenüberstanden.

⁹ Paul Videsott ist derzeit Professor für Romanische Philologie an der Universität Bozen (mit Sitz in Brixen).

Der Inhalt des neuen Fragebuchs betrifft die linguistischen Kategorien (*elaborierte Morphologie, Syntax* und *Lexikon*) und versteht sich als Supplement zu jenem des ALD-I. Klarerweise sind in dessen Konzeption und auch in die layout-technische Umsetzung die beim ALD-I gemachten Erfahrungen eingeflossen, wobei diese erneut auf den optimierten Einsatz der EDV bei der Erfassung und Weiterarbeitung der Daten abzielten.

Aus einsichtigen Gründen mussten beim ALD-II dasselbe Netz und dasselbe Lautschriftsystem zum Einsatz kommen. Weitere *Ähnlichkeiten* zum ALD-I: ethnophotografische Dokumentation jeder Ortschaft; in den ersten Jahren der Feldarbeit: gemeinsame Schulungsseminare („*Seminario di trascrizione*“-*SETRA*) für die Exploratoren.

Projekt- und zeitspezifische *Unterschiede* zum ALD-I waren: nur einmalige Abfrage des Questionnaires pro Ortschaft; durchgehender Einsatz von digitalen (und damit eben nicht mehr von analogen) Aufnahmegeäten.

Angesichts der um die Jahrtausendwende vollzogenen Ablösung des Betriebssystems *MS-DOS* durch *Microsoft Windows* war klar, dass für alle einschlägigen EDV-Belange des Projektes ALD-II in *Ersetzung* (und nicht bloß: *Reparatur*) des veralteten ALD-I-Programms *CARD* eine neue Generalsoftware zu erstellen war. Dies geschah ab etwa 2002 durch das von Edgar Haimerl konzipierte und lauffähig gemachte Programm *DMG* („*Dialect Map Generator*“).

Die Funktionalitäten dieses Programms übertrafen natürlich jene von *CARD*: dies betraf ganz besonders den Bereich der Produktion der Karten, die angesichts der großen optischen Bedeutung, inhaltlichen Komplexität und vor allem Anzahl derselben in größtmöglichem Umfang automatisiert werden sollte. Dies hatte zur Voraussetzung, dass Programm-Module geschaffen werden mussten, die imstande sein sollten, die überschneidungsfreie Verteilung von Transkriptionen (samt allen Sub- und Superskripten) auf den beiden Kartenhälften des ALD-Netzes fehlerfrei vorzunehmen. Dieses Problem wurde von insgesamt vier Informatikern (Edgar Haimerl, Fabio Tosques, Andreas Wagner und Bernhard Schauer) bearbeitet und schlussendlich in wirklich überzeugender Weise gelöst.

Nur nebenbei: von den vier genannten Personen waren bzw. sind zwei (Andreas Wagner, Bernhard Schauer) ausschließlich informatisch tätig, während Edgar Haimerl und Fabio Tosques von ihrer Ausbildung her interdisziplinäre Doppelnaturen darstellen. Abgesehen von der Informatik hat sich Edgar Haimerl mit Skandinavistik, Mathematik und Philosophie befasst, während Fabio Tosques die Informatik mit der Romanistik kombiniert hat.

Wie schon erwähnt, sind beim ALD-II die Enquêtes im Jahr 2007 zu Ende gegangen. Schon zwei Jahre darnach existierte eine als „fehlerfrei“ zu betrachtende Datenbank mit dem gesamten transkribierten Feldertrag, so dass die Produktion der Probe-Karten und damit die erste Stufe von deren philologisch-linguistischer Bearbeitung anlaufen konnte.

B. Schauer, H. Beer und A. Staudinger beschreiben die Details dieser Arbeit in den Kapiteln 2.-4. In toto waren an dieser zeitlich und technisch ungemein anspruchsvollen Arbeit im Wesentlichen nur vier Personen beteiligt: die drei zuvor Genannten und meine Wenigkeit. Insgesamt gab es drei volle und eine halbe Korrekturphase, wozu jeweils der gesamte Besatz an Karten und den dazugehörenden (sehr nützlichen) Liste ausgedruckt werden musste. Agnes Staudinger gibt dazu in Kapitel 4. numerische Hinweise.

Von unschätzbarem Wert war die ab 2009 mögliche Einbindung einer Sound-Datenbank in die Redaktionsarbeit. So war es möglich, jeden irgendwo an der Korrektheit einer Transkription aufgetretenen Zweifel durch sofortiges (und beliebig ausdehnbares) Hineinhören in die fragliche Stelle auszuräumen.

Da die Konzeption des ALD-II nicht nur die Produktion von *Karten*-Bänden, sondern auch jene von zwei kleinformatigen *Zusatz*-Bänden (Supplement-Band [„Volumen supplementarium“] und Index-Band [„Index generalis“]) vorsah und auch dafür möglichst elegant und zugleich arbeitsökonomisch zu produzierende Druckvorlagen erstellt werden sollten, hat Bernhard Schauer dazu ein weiteres Spezial-Programm („SuBIREs“) kreiert.

Von der für alle diese Zwecke und Programme geschaffenen EDV-Architektur bzw. Server-Landschaft berichtet B. Schauer im nachfolgenden Kapitel.

Für den ALD-II wurde angesichts der seit der Zeit des ALD-I erfolgten EDV-technischen Fortschritte auf die Produktion eines auf Papier zu publizierenden vor- und rückwärts alphabetischen Index verzichtet und stattdessen einem netzbasierten „*Index Retrieval System*“ (*IRS*) neuer Art der Vorzug gegeben. Dieses kann nicht nur – erneut auf der Grundlage von *ALD-light* und den *Normalbuchstaben* – die erwähnten alphabetischen Indizes erstellen, sondern auch nach bestimmten positionellen Vorgaben (betreffend Anfang, Mitte und Ende eines Wortes) feststellen, ob eine bestimmte Zeichensequenz im Gesamtdatenbestand vorkommt. Derzeit gibt es diese sehr schnelle Suchmaschine für beide Teile des ALD. Über das *IRS* des ALD-II kann man auch an die PDFs ganzer Karten und aller dazugehörenden Listen herankommen.

Unter Ausnützung von für unser Projekt günstigen Neuerungen (Stichwort: *open access*) bei der Politik der Druckförderung vonseiten der österreichischen Forschungsförderungsorganisation FWF war es möglich, alle Etappen des eigentlichen Druckvorgangs in der Hand zu behalten und damit die anfallenden Kosten zu minimieren. Dabei kam es zu einer technisch recht komplexen, prozedural und menschlich gleichwohl aber stets sehr harmonischen Kooperation mit der Trentiner Druckerei Alcione (gelegen in Lavis, auf halbem Weg zwischen den ALD-Messpunkten 66 [S. Michele all’Adige] und 121 [Trient]).

Drucker und Herausgeber (also meine Wenigkeit) konnten dabei – vor allem bei der Reparatur von gar nicht wenigen, während der Druckvorbereitung entdeckten Fehlern – von der ungemeinen Effizienz der unter der Obhut von B. Schauer stehenden (und entstandenen) EDV-Maschinerie profitieren.

Dieser Effizienz, die auch auf die nicht-informatischen Mitarbeiter(innen) übersprungen ist, war es schlussendlich zu verdanken, dass die zwei für den ALD-II geschaffenen zentralen Tools – *SDB* und *IRS* – auch für die Daten des ALD-I geöffnet worden sind, so dass diese – abgesehen von ihrer ungeschmälerten Fortexistenz im (und als) Druckwerk – qua EDV einen neuen Impuls erhalten haben.¹⁰

1.3 Zukunftsperspektiven

Vom Druckwerk ALD-II wurden 330 Exemplare produziert und davon 90% für den Verkauf freigegeben. Dieser erfolgt im Rahmen des Non-Profit-Verlages „Editions de Linguistique et de Philologie“ (mit Sitz in Straßburg und Paris) und hat nach zwei von Salzburg aus gestarteten Werbe-Campagnen (via Post und via Mail) im Dezember 2012 begonnen.

¹⁰ Die Gesamtmenge der von ALD-I und ALD-II der Fachwelt zur Verfügung gestellten Sprachkarten (SK) beträgt 1950 Einheiten. Innerhalb jener Kartenmengen, die man in den anderen romanischen Regionalatlanten findet, ist das ein absoluter Spitzenwert. Nur der korsische Atlas ALEIC (publiziert: 1933-1942, 9 Karten-Bände) und der gaskognische Atlas ALG (publiziert: 1954-1974, 4 Karten-Bände) haben mit 2001 bzw. 2531 SK einen größeren Kartenbestand. Wenn man aber das Produkt aus der Anzahl der Messpunkte (MP) und jener der Karten (SK) bildet, dann liegt der ALD als Ganzes mit dem Wert 423 150 an der absoluten Spitze (ALEIC: 49 MP mal 2001 SK = 98 049; ALG: (174 MP + 155 MP)/2 mal 2531 SK = 416 349,5). Ist es angesichts dieser Fakten bloß ein historischer Zufall, dass der Autor des datenmäßig sehr umfangreichen ALG – Jean Séguéy (1914-1973) – zum Begründer der *dialectométrie* geworden ist?

Dabei wird die eigentliche Verkaufsarbeit durch einen in Baden-Württemberg angesiedelten Distributor geleistet, der nicht nur die Lagerhaltung und den Versand, sondern auch die Entgegennahme der Bestellungen und das Inkasso der ausgestellten Rechnungen besorgt.

Durch die schon erwähnte Akkumulierung günstiger Umstände – wozu auch die hohe Performance der ALD-Informatik zählt – war es möglich, einen konkurrenzlos niedrigen Verkaufspreis (von nur 200 Euro!!!) für die sieben in exzellenter Qualität produzierten Bände des ALD-II festzusetzen. Die Zukunft des papiernen ALD-II scheint damit für lange Zeit gesichert zu sein.

Doch wie schaut es mit den netzbasierten Tools¹¹ aus, über die man über unsere zwei neuen Homepages¹² herankommt? Da mit 31.12.2012 die ALD-Mannschaft in alle Winde zerstoßen ist und bedauerlicherweise für das Jahr 2013 bereits zugesagte Fördermittel für die Aufarbeitung noch unerledigter Projekt-Arbeiten und die fortdauernde Pflege der EDV-Struktur wieder zurückgezogen worden sind, ist auf dem EDV-Sektor die Zukunft des ALD-II deutlicher weniger rosig als auf jenem des Papiers.

Ich werde versuchen, trotz weitestgehender Austrocknung der (zwischen 1999 und 2012 in sehr großzügiger Weise geflossenen) finanziellen Mittel von Fall zu Fall die sich als nötig erweisenden Sanierungen zu organisieren. Mit sehr großer Dankbarkeit habe ich die diesbezüglich geäußerte Hilfsbereitschaft der Mitautoren dieses Berichts zur Kenntnis genommen.

Immerhin geht es hier um so etwas wie die „Pflege des eigenen Kindes“. Und zwar eines rundum sehr gut geratenen Kindes...

Die über das Ende von 2012 noch offen verbliebenen Agenden sind:

- Inhaltliche und sprachliche Komplettierung der beiden neuen Homepages;
- Einrichtung der Bilddatenbank und Tagging der 17 000 Dias;
- Fertigstellung der neuen Version des „Sprechenden“ zum ALD-I;
- Fortsetzung und Vollendung des Fein-Taggings der Daten der Sound-Datenbank des ALD-II.

Alle diese Agenden könnten bei Vorhandensein entsprechender Fördermittel von den ehemaligen Mitarbeiterinnen des ALD-II „nebenbei“ erledigt werden.

Noch ein Wort zu den derzeit der geolinguistischen Fachwelt zur Verfügung gestellten Funktionalitäten von ALD-I und ALD-II:

Papierversionen:

Traditionelle Auswertung von 1950 Volltext-Karten (mit feiner Transkription) durch direkte Konsultation mit (oder ohne) parallele(r) Anwendung von „stummen Karten“ (die über die beiden neuen Homepages verfügbar sind)

¹¹ Überdies findet man die PDFs des Inhalts aller sieben Bände des ALD-I in der E-Book-Library „Phaidra“ des FWF unter dem folgenden Link: https://e-book.fwf.ac.at/search_object. Man kann sich dort in der Tat – das Vorhandensein eines guten Druckers (für Farbe und das Format A2) und von sehr viel Geduld vorausgesetzt – die sieben Bände des ALD-II völlig „frei“ herunterladen und ausdrucken. Auch hier bleibt offen, wie lange dies möglich sein wird. Doch ist davon auszugehen, dass der FWF als öffentliche Institution alles daransetzen wird, um die EDV-technische Lesbarkeit des Inhalts seiner E-Book-Library auf lange Sicht sicherzustellen.

¹² ALD-I (neu): <http://ald1.sbg.ac.at/>; ALD-I (alt): <http://ald.sbg.ac.at/ald/>; ALD-II: <http://ald2.sbg.ac.at/>.

Elektronische Tools:

SDB: akustische Nachkontrolle aller Transkriptionen, sowohl punktgenau (via Eingabe von Messpunkt x [1-217] und Frage y [1-1063]) als auch durch Anhörung stundenlanger Aufnahmegespräche

IRS:

Zugriff auf die transkriptorisch auf *ALD-light* reduzierten Daten nach den folgenden Kriterien:

- vorwärts alphabetisch;
- rückwärts alphabetisch;
- freie Suche (via Definition eines frei [in *ALD-light* oder in Normalbuchstaben] definierbaren Such-Nexus) nach der Existenz eines Such-Nexus in den Gesamtdaten von ALD-I oder ALD-II in den drei folgenden Positionen eines „Worts“: Anfang, Mitte, Ende.

Beim ALD-II kann die zuletzt genannte Prozedur bis zur Auffindung (und zum nachfolgenden Ausdruck) der betreffenden Sprach-Karten und der dazugehörigen Listen verlängert werden. Damit können genuin philologisch-linguistisch orientierte Interessen von Sprachgeographen optimal zufriedengestellt werden.

Anders als beim ALD-I¹³ wird aber beim ALD-II ein Bereich nicht bedient: jener von datenanalytisch orientierten Computerlinguisten, die daran Interesse hätten, direkt auf die Transkriptionen (ob nun in *ALD-standard*, *ALD-light* oder in Normalbuchstaben) zuzugreifen. Diese Interessen können derzeit nicht durch die Konsultation der beiden Homepages bedient werden. Wer mit den Transkriptionen in diesem Sinn „rechnen“ möchte, muss sich an mich oder Bernhard Schauer wenden, damit er solcherart Zugriff auf unsere einschlägige Datenbanken erhält.

2 Zum EDV-Einsatz beim ALD (*Bernhard Schauer*)

2.1 Zum EDV-Einsatz beim ALD-I

Zur Eingabe und Korrektur von Transkriptionen sowie zum Druck der Sprachkarten wurde in der Mitte der 1990er Jahre von Edgar Haimerl eine Software namens CARD entwickelt und erfolgreich eingesetzt. In weiterer Folge wurde dieses DOS-Programm durch das Programm *IRS* („Index Retrieval System“) ergänzt, das das Durchsuchen der Transkripte in einer vereinfachten Lautschrift („ALD light“) erlaubte, die entsprechenden Vorkommen auf den Karten anzeigte und auch zu Zwecken der Klassifikation (samt nachfolgender Visualisierung der Resultate) der Daten einzelner Karten des ALD-I verwendet werden konnte.

Um dem ALD-I auch eine hörbare Komponente beizugeben, wurde gegen Ende des letzten Jahrtausends der „Sprechende Sprachatlas“ (kurz: der „Sprechende“) zuerst als CD-, dann als DVD- und schließlich – mit einer dem Projekt VIVALDI entlehnten Technologie – als Internet-Version entwickelt. Zu diesem Zweck wurden in Ladinien nach den Regel-Explorationen zusätzlich Schälle mit besonders hoher Qualität („Edelsounds“) – also unter idealen Bedingungen aufgenommene Ton-Daten – erhoben und der Allge-

¹³ Beim ALD-I war der skizzierte Zugriff auf die Gesamtdaten über die erwähnten drei CD-ROM möglich. Diese Möglichkeit wurde von mir im Jahr 2008 in Kooperation mit Thomas Zastrow (Universität Tübingen) benützt, um eine dialektometrische Verrechnung der Daten des ersten Bandes des ALD-I mittels der Levenshtein-Distanz vorzunehmen.

meinheit über den „Sprechenden“ zur Verfügung gestellt. Datenseitig handelte es sich dabei nicht um die 217 Messpunkte des Gesamtnetzes, sondern nur um die 21 Messpunkte Ladiniens (ALD-PP. 81-101).

Beim ALD-II wurde im Jahr 2010 mit zwei Digitalisierungs-Campagnen begonnen: diese betrafen zuerst die während Regel-Enquêtes bereits mit digitalen Aufnahmegegeräten erstellten Ton-Aufnahmen und ab 2012 die bei der Feldarbeiten aufgenommenen Dias („Ethnophotographie“). Weiters wurde im Sommer 2012 die Software für den DVD-Sprechenden des ALD-I neu entwickelt, da ab Windows 7 64bit das auf die Zeit vor der Jahrtausendwende zurückgehende, alte Programm nicht mehr funktionstüchtig war.

2.2 Zum EDV-Einsatz beim ALD-II

Die Erfassung und Korrektur der Transkriptionen sowie der Druck der Sprachkarten beim ALD-II erfolgte über das Softwarepaket *DMG* („Dialect Map Generator“). Auch beim ALD-II wurde dieses Programm durch ein *IRS* („Index Retrieval System“) flankiert, das aber in diesem Falle über das Internet funktioniert und das gezielte Durchsuchen des Datenbestandes anhand der schon erwähnten reduzierten Transkription („ALD light“) erlaubt. In diesem Fall geht es also nicht mehr um das Klassifizieren der Daten des ALD-II.

Zusätzlich entstand beim ALD-II eine eigene Software („SoundDatenBank“ oder „Sound-Datenbank“, *SDB*) zur Verwaltung der im Feld während der Exploration erhobenen Ton-Daten. Die *SDB* fand unter anderem schon bei der Korrektur und Redaktion der zu veröffentlichenden Daten Anwendung; dabei war es wesentlich, dass Ton-Daten auf ganz kurzem Weg aufgefunden und abgehört werden konnten.

Eine weitere Zielsetzung beim ALD-II war, die rund 17 000 Dias von ALD-I & II nach deren Digitalisierung über das Internet allgemein verfügbar zu machen. Die dazu entstehende *BDB* („BildDatenBank“) konnte allerdings bis Projektende (31.12.2012) leider nicht zur Gänze fertig gestellt werden.

3 Details zu den Strukturen der beim ALD-II verwendeten EDV

In diesem Kapitel soll ein vertiefter Einblick in die technische Struktur des ALD-II und der einzelnen Programme vermittelt werden. In einem weiteren Punkt soll auf die jeweiligen Erkenntnisse aus der Programmerstellung eingegangen werden. Um hier detailgenau informieren zu können, ist es erforderlich, das einschlägige Fachvokabular zu verwenden. Es wird aber versucht, alle verwendeten Abkürzungen und Begriffe in Fußnoten zu erklären und auch auf weiterführende Literatur hinzuweisen.

3.1 Infrastruktur

Um die technische Infrastruktur funktional zu halten und dabei möglichst hardwareunabhängig zu bleiben, wurde gemeinsam mit dem Fachbereich Romanistik der Universität Salzburg ein physikalischer Server angeschafft, auf dem die Virtualisierungslösung *VMWare ESXi*¹⁴ (Version 4) eingesetzt wird: siehe dazu die Visualisierung in Abbildung 1. Die in der Wolke zusammengefassten Server laufen dabei innerhalb der Virtualisierungslösung parallel.

¹⁴ Aktueller Produktname, ab Version 5: *VMWare VSphere Hypervisor*.

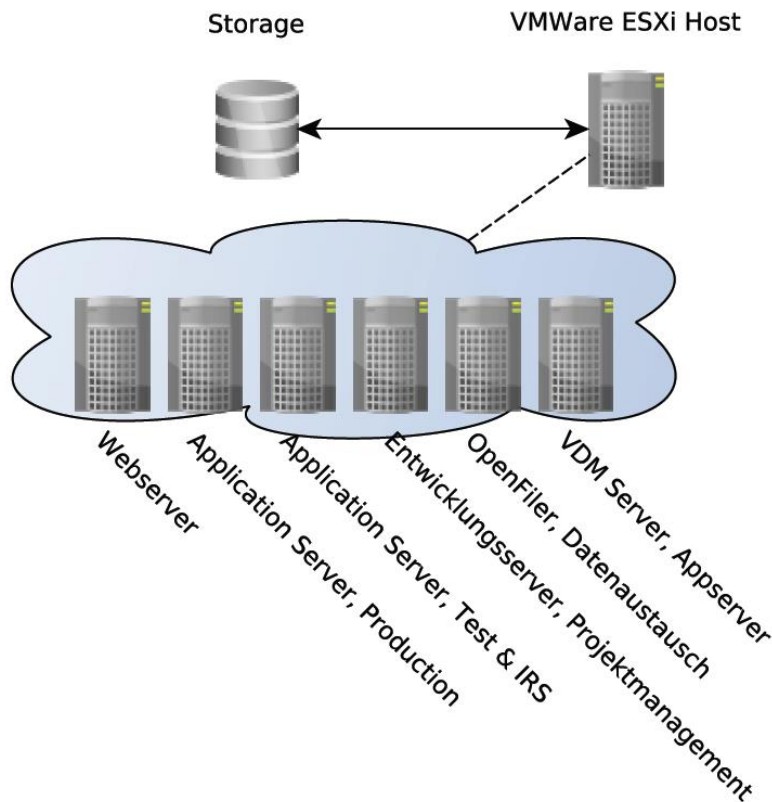


Abbildung 1: Überblick über die EDV-Infrastruktur des ALD-II sowie über die virtuellen Server

Ein großer Vorteil der Virtualisierung der Server besteht darin, dass diese einfach auf andere *VMWare ESXi* Server umgesiedelt werden können, und auch, dass die Universität Salzburg eine derartige Infrastruktur aufgebaut hat. Alternativ dazu ist es möglich, die virtuellen Server in die „Cloud“ zu übersiedeln, beispielsweise zu *Amazon EC2* oder ähnlichen Angeboten. In unserem Fall kommt dabei eine sogenannte „*private Cloud*“ zum Einsatz: siehe Mell et al. 2011.

Die digitalen Daten des ALD-I und II werden von den ITServices der Universität Salzburg gehostet und intern über *Windows Shares*¹⁵ bzw. *iSCSI*¹⁶ zur Verfügung gestellt. Im technischen Sinn handelt es sich dabei um ein *Network Storage*, das Daten sowohl als *SAN*¹⁷ bzw. *NAS*¹⁸ zur Verfügung stellen kann.

Die Aufteilung auf die einzelnen virtuellen Server¹⁹ basiert auf speziellen Funktionalitäten, wobei versucht wurde, deren Dienste anhand der gestellten Anforderungen zu separieren.

¹⁵ Darunter wird hier die Anbindung über *CIFS* verstanden. Die Daten sind dabei über die „Netzwerkumgebung“ unter *Windows* bzw. vergleichbare Dienste unter *MacOS* bzw. *Linux* zu erreichen.

¹⁶ Es ist das der Name des Protokolls, um Festplatten „blockweise“ über ein Netzwerk ansprechen zu können. D.h. es handelt sich um keinen dateibasierten Zugriff.

¹⁷ „*Storage Attached Network*“, ist der Oberbegriff für Systeme, die Daten *block*-basiert zur Verfügung stellen (beispielsweise über *iSCSI*).

¹⁸ „*Network Attached Storage*“, ist der Oberbegriff für Systeme, die Daten *datei*-basiert zur Verfügung stellen (beispielsweise über *CIFS* bzw. *NFS*).

¹⁹ Ab hier wird in diesem Text mit dem Begriff *Server* immer ein „virtueller Server“ bezeichnet.

Der Webserver ist der einzige Server der IT Infrastruktur der Universität Salzburg, der von außen erreichbar ist; er stellt unter anderem die Projekt-Homepages (zu: ALD-I, ALD-II, ALD-I Web-Sprechender, Dialektometrie) zur Verfügung. Darüber hinaus wird er auch als „*Reverse Proxy*“²⁰ für die dahinter liegenden Server verwendet.

Weiters stehen dem Projekt zwei Anwendungsserver zur Verfügung, die einerseits die Produktivumgebung für den laufenden Betrieb und andererseits eine umfassende Testumgebung bereitstellen.

Ein weiterer Server dient dem Projektmanagement, der Quellcodeverwaltung sowie als *Build/CI-Umgebung*: er wird als *Entwicklungsserver* bezeichnet.

Um die Tondateien und die Daten der digitalisierten Dias gleichzeitig mehreren Servern zur Verfügung stellen zu können, wurde eine *OpenFiler*²¹-Instanz installiert, die den Zugriff darauf über *NFS* gewährleistet.

Da der Zugriff auf die Daten nur über einen einzelnen Punkt erfolgt, lässt sich dieser besser absichern. Zudem ist es auch einfacher, im Hintergrund Server auszutauschen, da diese nur *intern* und nicht *weltweit* bekannt sein müssen. Zusätzlich nimmt diese Aufteilung auch Rechenlast vom Webserver, da diese Zugriffe nur weitergeleitet werden. Aufgrund der Virtualisierung wäre es hier möglich, den/die *Anwendungsserver* auf physikalisch andere Server umzusiedeln, um mehr Rechenleistung zur Verfügung zu haben.

3.2 Softwarearchitektur

Die für den ALD-II entwickelte Software basiert grundsätzlich auf einer *Client-Server*-Architektur mit *Datenbank-Backend*, es handelt sich also um eine *3-Tier*-Architektur (*Client, Anwendungsserver, Datenbank*): siehe dazu die Abbildung 2. Hier erkennt man auch den Unterschied zwischen internen und externen Zugriffen auf die Struktur. Serverseitig wird beim Forschungsprojekt ALD-II ein *JBoss AS*²² bzw. für die Webservices ein *Apache Tomcat Servlet Container*²³ verwendet.

²⁰ Dabei wird der interne Server nur über den Webserver indirekt angesprochen. Abfragen könnten beispielsweise damit auch am Webserver zwischengespeichert werden.

²¹ Das ist eine Softwarelösung, die als lokales *NAS* für die Server operiert. Die Datenspeicherung erfolgt in unserem Fall über das *SAN*.

²² *JBoss AS: Java EE Application Server*.

²³ Es handelt sich dabei – vereinfacht ausgedrückt – um ein Produkt, um Java Web-Anwendungen in einer „leichtgewichtigeren Umgebung“ betreiben zu können, als dies ein *Java EE Server* wäre.

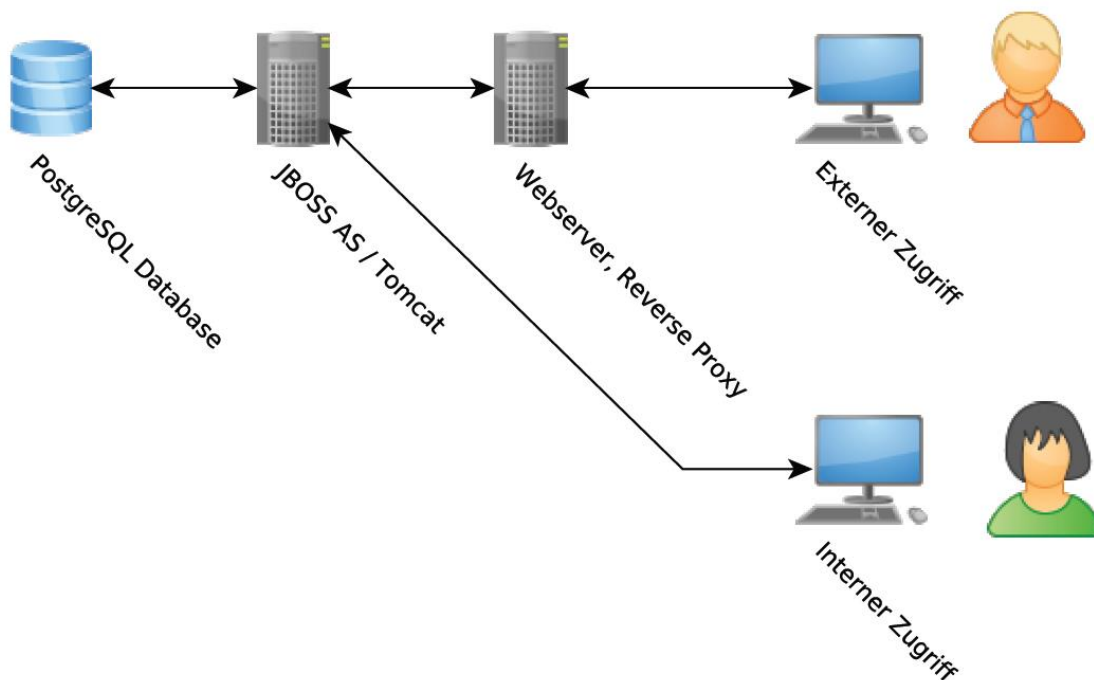


Abbildung 2: Softwarearchitektur mit Zugriffsvarianten

Die eigentlichen *Daten* werden in der dahinter liegenden Datenbank abgelegt, die unter anderem auch deren referentielle Integrität sicherstellt.

3.3 Verwendung von Software, Tools & Bibliotheken

Bei der Auswahl der zu verwendenden Software wurde auf mehrere Gesichtspunkte Rücksicht genommen: primär wurde, soweit dies möglich war, auf *Open-Source*-Lösungen abgezielt (*Apache Tomcat*, *JBoss AS*, *PostgreSQL*, *Apache Webserver*, usw.). Wenn dies nicht möglich oder auch nicht genug effizient war, wurde die Palette um kostenfrei verwendbare Produkte erweitert, was aber nur für die Virtualisierungslösung notwendig war. *VMWare ESXi* lizenziert die Serversoftware alleine kostenfrei; das Monitoring bzw. die Verwaltung mehrerer Server usw. müssten bei Bedarf kostenpflichtig nachgerüstet werden.

Ein weiterer Punkt, der bei der Softwareauswahl zu berücksichtigen war, sah vor, dass nach Möglichkeit versucht wurde, immer weit verbreitete Produkte zu verwenden, da davon eine größere Langlebigkeit erwartet werden konnte. Dazu kommt, dass weit verbreitete (Open Source) Produkte in aller Regel besser gepflegt werden, da quantitativ ein größeres Interesse am Funktionieren dieser Software besteht.

3.4 Kurzvorstellung der einzelnen Programme

3.4.1 DMG - Dialect Map Generator

Die „Hauptsoftware“ des Projektes erlaubt die Erfassung und Korrektur von Transkriptionen, sowie den Druck der Sprachkarten. Weiters wurden damit die Daten zu den Informanten und den explorierten Orten erfasst.

Die Entwicklung dieser Software sollte ab 2001 von einer externen Firma durchgeführt werden, was zu keinem Resultat führte. Daher erfolgte umgehend und unter direkter Federführung von Edgar Haimerl der „Neustart“ der Entwicklung dieser Software, die dabei durch den als *Map Generator* bezeichneten Teil, der die Kartenerzeugung unterstützt, komplettiert wurde.

Technisch gesehen erfolgte die Entwicklung auf der Grundlage von *Java EE*, im speziellen von *EJB 2.1*. Ein großes Manko dieser Software ist, dass sie nicht von Anfang an auf Nutzungseffizienz getrimmt war. Nähere Details findet man in den nachfolgenden Abschnitten 3. und 4.

3.4.2 SDB - SoundDatenBank

Um auf die während der Exploration aufgenommenen Tondaten schnell und punktgenau zugreifen zu können, entstand ab Frühjahr 2012 die zweite Version der Sounddatenbank (*SDB2*). Die erste Version (*SDB1*) war eine Behelfslösung, die vor allem mit dem Ziel der raschen und einfachen Umsetzung entstanden ist und keine vollständige Abbildung der vorhandenen Daten auf die Datenstruktur erlaubte.

Diese neue Software entstand als monolithische Anwendung und enthielt von Anfang an die folgenden Funktionen: die Neuerfassung von Tondateien, die Segmentierung längerer Ton-Strecken durch Taggen zwecks Erzeugung eigener Responses, die damit verbundenen Korrekturen sowie das Anhören der getaggteten Responses über das Netzwerk. Die Ton-Daten werden dabei nach Bedarf ganz oder ausschnittsweise übertragen. Zur besseren Orientierung wird dem Benutzer eine Schalldruckkurve präsentiert, mit deren Hilfe er innerhalb ganzer Ton-Dateien bestimmte Bereiche markieren und anhören kann, ohne dass deswegen die vollständigen Daten auf den PC des Benutzers übertragen werden müssen.

Serverseitig wurde hier auf *Webservices* gesetzt, die mit *Apache Axis 2* umgesetzt wurden und in Richtung Datenbank durch *Hibernate* ergänzt werden. Dabei handelt es sich um ein Framework, das den Zugriff über *SOAP*²⁴ übernimmt.

Aufgrund der Erfahrungen mit dem weiter beschriebenen Programm *SuBIReS* und mit der dort verwendeten *Netbeans RCP*²⁵ wurde die *SDB* mit der Version 2.1 auf diese Plattform portiert, in Module zerlegt und an die sich in der Zwischenzeit geänderten Anforderungen angepasst.

Aufgrund des großen Erfolges der *SDB2* beim ALD-II wurde beschlossen, diese auch auf die Daten des ALD-I auszuweiten. Dieser Vorgang konnte im November 2012 abgeschlossen werden. Seitdem ist es möglich, in der *Version 2.1* der *SDB* frei zwischen ALD-I und II hin- und herzuschalten.

Die aktuelle Version ist über das Internet verfügbar. Der dazu erforderliche Client kann von der Projekthomepage heruntergeladen und unter *Windows*, *MacOS* als auch *Linux* verwendet werden.

²⁴ *SOAP: Simple Object Access Protocol*: vereinfacht gesagt handelt es sich um ein Netzwerkprotokoll, um Daten über eine allgemein gültige Beschreibung unabhängig von der Programmiersprache zu übertragen. Abfragen und Daten werden dabei in *XML* gepackt und übermittelt.

²⁵ *RCP: Rich Client Platform*: in grober Näherung handelt es sich dabei um eine Programmbasis, die dem Entwickler Arbeit abnimmt, da beispielsweise Abhängigkeiten zwischen Modulen, deren Update, Inter-Modul-Kommunikation usw. von der Plattform bereits fertig implementiert angeboten werden.

3.4.3 SuBIReS - Supplementary Book and Index Retrieval System

Zur Erstellung der Druckvorlagen – vor allem für die geplanten Papier-Indices zum ALD-II – wurde eine weitere Softwarelösung benötigt, die es gestattet, bei der Vorbereitung der Druckvorlagen die einzelnen Dokumente aus der Datenbank möglichst automatisiert erzeugen zu können.

In layout-technischer Hinsicht wurde auf das automatisch gut erzeugbare *(La)TeX* gesetzt. *SuBIReS* erzeugt damit *LaTeX*-Dokumente und erlaubt es auch, diese kleinweise (d. h. Verzeichnis um Verzeichnis) nach *PDF* umzuwandeln. Damit konnten zum einen der „Supplementband“, der jene Daten enthält, die auf den bestimmten Karten des ALD-II keinen Platz mehr hatten, und zum anderen der Indexband (mit vorwärts- und rückwärts alphabetischen Indices zu den italienischen Kartentiteln) erzeugt werden.

Zusätzlich zu den Karten, die weiterhin mittels *DMG* entstanden, wurden zu jeder Atlas-Karte mehrere Listen gedruckt, deren Erstellung im Zuge der Entwicklung von *SuBIReS* ebenfalls automatisiert werden konnte. Diese Listen enthalten zu jeder Karte die vollständigen Transkriptionen inklusive aller Versionen in verschiedener Abfolge: sortiert nach Orten, tokenisiert sowie in vorwärts- und rückwärtsalphabetischer Anordnung.

SuBIReS basiert wiederum auf *Netbeans RCP* für die Anwendung, auf *Apache Axis 2* als Serveranbindung und auf *Hibernate* für den Datenbankzugriff.

3.4.4 IRS2 - Index Retrieval System des ALD-II

Um die Transkripte online durchsuchen zu können, wurde ab Frühjahr 2011 eine auf *Adobe Flex*²⁶ basierende Web-Anwendung erstellt. Aufgrund der verwendeten Technologie muss auf dem PC des Betrachters die jeweils aktuellste Version von *Flash Player* installiert sein.

Das System erlaubt es, den Datenbestand anhand der reduzierten Transkription („*ALD light*“) zu durchsuchen. Dabei gibt es zwei Abstufungen der Reduktion: *ALD light* (als simplifizierte Lautschrift) und *A-Z* (als Normalbuchstaben). Im zweiten Fall werden alle Zeichen des Transkriptionssystems auf die entsprechenden Grund-Buchstaben des Alphabets reduziert. Seit dem Frühjahr 2012 ist das System produktiv über die Projekt-homepage erreichbar und auch mit den Daten des ALD-II direkt verlinkt, wobei zusätzlich Grafiken, Listen und sogar ganze Sprach-Karten gezeigt werden können. Seit dem Sommer 2012 ist *IRS2* auch für die Daten des ALD-I verfügbar, jedoch ohne die Option direkter Links zu den eben erwähnten Grafiken, Listen und Sprach-Karten.

Serverseitig wurde hier auf *Blaze DS* anstatt auf *SOAP* gesetzt. Die Anbindung an die Datenbank erfolgte wie bei allen Webservice-Anwendungen des ALD-II über *Hibernate*.

3.4.5 BDB - Bilddatenbank

Im Sommer 2012 begannen die Arbeiten an der Bilddatenbank (*BDB*), die die Digitalisate aller bei ALD-I und ALD-II gemachten Dias verwalten und über das Web verfügbar halten sollte.

Die Bilddatenbank sollte serverseitig sowohl über *SOAP* als auch über *PHP* erreichbar sein, um die Integration der Daten in die Projekthomepage zu vereinfachen. Da die Arbeiten zur Nachbearbeitung und Kategorisierung der digitalisierten Dias nicht – wie ursprünglich geplant – schon im Herbst des Jahres 2012 begonnen werden konnten, war es

²⁶ Jetzt: *Apache Flex*.

nicht mehr möglich, die *BDB* bis zum Projektende fertig zu stellen. Eine Vollendung zu einem derzeit noch nicht feststehenden Zeitpunkt ist aber angedacht.

3.4.6 Homepages

Im Jahr 2011 begannen die Arbeiten an der Neugestaltung der Homepages der beiden Teile des Gesamtprojektes ALD. Sowohl der ALD-II als auch der ALD-I haben damit ein neues Gesicht bekommen. Die neuen Homepages bestehen nicht mehr aus statischen *HTML*-Seiten, sondern werden über ein *Content-Management-System (CMS)* namens *Concrete 5* verwaltet. Damit ist es auch für „Nicht-Techniker“ möglich, einzelne Elemente der Homepage zu ändern bzw. in verschiedene Sprachen zu übersetzen, was sich sehr positiv auf die Wartbarkeit auswirkt.

Der letzte Schritt, auch die neue Version des Web-Sprechenden des ALD-I in die neue Homepage des ALD-I zu integrieren, konnte bis Projektende leider nicht mehr realisiert werden.

3.5 Erkenntnisse

Es hat sich gezeigt, dass es sehr vorteilhaft ist, dass der EDV-Techniker seine Arbeit inmitten des Projektteams verrichtet. Auf der einen Seite erhält er dadurch ein direktes Feedback zu den eigenen Lösungen; andererseits lassen sich dadurch neue Ideen und die damit verbundenen Programme im wechselseitigen Austausch rasch entwickeln bzw. erstellen.

Die Verwendung einer Projektmanagement-Lösung wie beispielsweise *Redmine*, das bei uns zum Einsatz kam, war unverzichtbar, um rechtzeitig steuernd eingreifen zu können und die entwickelten bzw. zu entwickelnden Softwarelösungen und deren Änderungen verwalten zu können. Dass zur Quellcode-Verwaltung ein Versionsverwaltungssystem (*VCS*) wie beispielsweise *Subversion* eingesetzt wurde, war dabei eine Notwendigkeit.

Ein weiterer wichtiger Punkt war auch, die Software so *modular* wie möglich zu entwickeln, um die Wiederverwertung von Quellcodes zu vereinfachen bzw. überhaupt erst zu ermöglichen. Klare und einfache Konzepte auch im Sinne des *KISS*-Prinzips („*Keep it small and simple*“) haben sich hier immer bezahlt gemacht. Dazu gehörte aber auch, dass Ziele möglichst früh definiert werden mussten, wenngleich dies bei einem so komplex strukturierten Forschungsprojekt natürlich schwierig war, da sich laufend neue Notwendigkeiten ergeben haben bzw. bestehende Lösungsansätze abgeändert werden mussten.

4 Zukunftsperspektiven

Mit dem Ende des Jahres 2012 endete das Forschungsprojekt ALD-II. Zu diesem Zeitpunkt wurden alle Softwarekomponenten auf dem erreichten Stand eingefroren und werden hinfert – zumindest von Seiten der Universität Salzburg – nicht weiter betreut. Es steht zu erwarten, dass, solange die Hardware des *VMWare ESXi Servers* funktionsfähig bleibt, die Server-Dienste auf dem aktuellen Stand erhalten bleiben. Danach sollten diese nahtlos in die IT-Struktur der Universität Salzburg eingegliedert werden.

Dass die Software lauffähig bleibt, scheint zumindest für die Dauer der nächsten Jahre wahrscheinlich, da *Java* als etablierte Umgebung auf fast jedem PC anzutreffen ist. Dagegen ist – zumindest für die nächsten Versionen – *Flashplayer* als integraler Bestand-

teil des Browsers *Google Chrome* in *plattform-unabhängiger* Form nur eingeschränkt verfügbar.

Die Webseiten und auch die damit verbundenen *HTML5*-Inhalte dürften, wenn man ihnen dieselbe Lebensdauer wie den heute immerhin schon 15 Jahre alten *HTML*-Dokumenten zubilligt, am längsten verfügbar bleiben.

5 Zur praktischen Anwendung der für den ALD-I und ALD-II entwickelten EDV-Werkzeuge: die Sound-Datenbank (SDB) (Heidemarie Beer)

5.1 Die für den ALD-II entwickelte Sound-Datenbank

Neben den Transkripten und den bildlichen Daten (Diapositiven) gibt es beim ALD-II auch recht umfangreiches Tonmaterial, das während der Enquêtes erfasst wurde.

Es sind dies die zu den 217 Messpunkten des ALD-II und dessen 1 063 Fragen erhobenen akustischen Daten. Sie wurden von 10 Exploratoren bei 833 Informanten erhoben. Diese durchwegs in digitaler Form vorliegenden Aufnahmen umfassen insgesamt mehr als 20 000 Stunden bzw. ein Datenvolumen von rund 145 Gigabyte.

Von Anfang an war geplant, alle Tonaufnahmen in einer Sound-Datenbank (*SDB*) zur Verfügung zu stellen, die es erlaubt, bestimmte Schallereignisse in ihrer originalen Form möglichst exakt und rasch aufzufinden und abzuhören.

Mit der Einführung einer neuen Version der *SDB* im Frühjahr 2012 (*SDB2*) wurde es viel leichter, sich im Programm zu orientieren, die gewünschte Fragennummer zu finden und abzuhören.

Die Abbildung 3 zeigt die verschiedenen Felder des Programmfensters der *SDB*.

Zunächst gibt man die gewünschte Ort- bzw. Fragennummer in die entsprechenden Felder ein und klickt auf „Query“. Nun werden im rechten oberen Feld alle Sounddateien aufgelistet, die es zu diesem Ort gibt.

Wählt man (rechts oben) eine Sounddatei aus, so erscheinen im Feld darunter alle darin vorhandenen Fragennummern mit den Positionsangaben. Außerdem erscheint in der Mitte unten das Schalldruckdiagramm. Damit kann man Segmente, die man abhören möchte, markieren und sich nach Bedarf vor- oder zurückbewegen. Das Schalldruckdiagramm bildet die jeweilige Tonstelle visuell ab und bietet damit die Möglichkeit, sich noch genauer zu orientieren und das Gehörte zu vertiefen

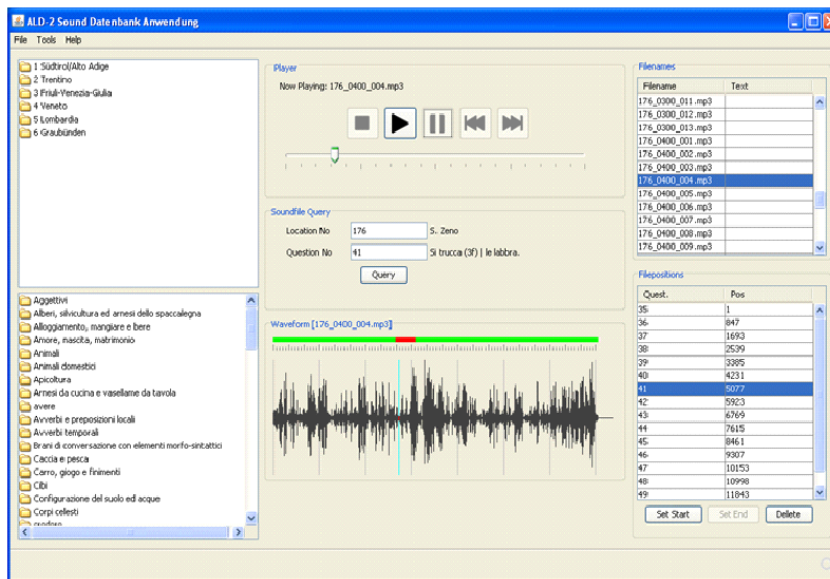


Abbildung 3: Die SDB2 des ALD-II in Betrieb (Abhörbeispiel: ALD-Ortschaft 176, Frage 41)

In den beiden linken Feldern sehen wir zusätzliche Einträge, nach denen man die Suche ebenso ausrichten kann. Im unteren Bereich sind dies die Regionen der ALD-Zone, darüber die nach Sachthemen gruppierten Fragen des ALD-II-Questionnaires. Wenn man nun beispielsweise den Ordner *Trentino* auswählt, so klappen alle dem Trentino zugehörigen Orte des ALD auf. Ebenso verhält es sich bei den anderen Regionen.

Letztlich entscheidet aber der interessierte Benutzer am besten für sich selber, welche Suchmethode (1. Suche im regionalen Ordner, 2. Eingabe der Nummern der Messpunkte und der Fragen) er zur Ingangsetzung der SDB2 bevorzugt.

Die SDB des ALD-II war für uns auch während der Redaktion der Atlas-Karten von großer Bedeutung, so vor allem für die Rekonstruktion von Teilen komplexer Fragen des Fragebuchs. Zum einen handelte es sich dabei um die Rekonstruktion der Antworten auf Stimuli, die nach den Vorgaben des Fragebuchs bei der Enquête nicht transkribiert werden sollten. Zum anderen mussten während der letzten Korrekturdurchgänge redaktionelle Unsicherheiten abgeklärt werden. So saßen der Projektleiter und ich nicht nur einmal in unserem ALD-Büro im alten Gebäude (Akademiestraße 24) und hörten uns gemeinsam die eine oder andere Frage an, um nochmals zu überprüfen, welche Transkription im speziellen Fall nun die richtige wäre.

5.2 Die beim ALD-II getätigten Nachaufnahmen

An einigen Messpunkten des Befragungsgebietes verblieben nach dem Ende der Regel-Enquêtes (im Jahr 2007) Datenlücken. Es handelte sich dabei um insgesamt 25 Ortschaften, in denen die ALD-II-Exploratorin Brigitte Rührlinger im Herbst 2010 sowie im Frühjahr und Sommer 2011 in bewährter Weise Nachenquêtes durchführte.

Bei diesen Enquêtes fielen erneut akustische Daten an, die anschließend weiterbearbeitet werden mussten. Es war nun meine Aufgabe, diese Daten abzuhören und zu schneiden. Dies bedeutete, dass die auf Speicherkarten vorliegenden und oft recht langen Dateien in etwa 5 Minuten lange Segmente zerlegt werden mussten. Nebenbei waren

Listen anzufertigen, in die einzutragen war, welche Fragennummern sich in der jeweils neu geschnittenen mp3-Datei befanden.

Teilweise wurden diese von Brigitte Rührlinger neu erfassten Daten auch in das Programm *DMG* eingegeben und – soweit möglich bzw. notwendig – sogar noch in die Korrektur und Redaktion der zum Druck bestimmten Karten einbezogen.

Nach dem Taggen galt es, die neuen mp3-Dateien in die schon bestehende Sound-Datenbank einzupflegen. Dazu waren vor allem die vorhin erwähnten systematisch angelegten Listen von Bedeutung.

Für das Taggen selbst verwendete ich das Audio- und Schneideprogramm *Audacity*, mit dem man im Allgemeinen sehr gut arbeiten und sogar Mängel in der akustischen Qualität der Daten sehr effektiv ausgleichen oder beseitigen kann (s. dazu Abbildung 4).

5.3 Die für den ALD-I nachträglich erstellte Sound-Datenbank

Nach den äußerst positiven Erfahrungen mit der Sound-Datenbank des ALD-II kam bald der Wunsch nach einer analogen Möglichkeit für die Tondaten des ALD-I auf.

Vom ALD-I gab es systematisch aufbereitetes akustisches Material zunächst nur in der Form des „Sprechenden“, der jedoch nur die 21 Messpunkte der brixnerisch-tirolischen Ladinia (ALD-PP. 81-101) abdeckt. Bei der Erstellung der Sound-Datenbank des ALD-I und dem damit verbundenen Taggen wurden wegen ihrer wesentlich besseren Qualität die dem „Sprechenden“ zu Grunde liegenden DAT-Aufnahmen und nicht die bei den Regel-Enquêtes gemachten (und mit einem deutlich hörbaren Grundrauschen behafteten) C90-Kassetten verwendet.

Sämtliche akustischen Daten des ALD-I lagen ursprünglich in analoger Form auf mehr als 1 200 C90-Kassetten vor. Diese nicht unbeträchtliche Daten-Menge wurde im Sommer 2010 von Brigitte Rührlinger zunächst in ziemlich mühseliger und zeitfressender Kleinarbeit digitalisiert und entauscht. Bei insgesamt 217 Orts-Aufnahmen kam schließlich ein Datenvolumen von insgesamt etwa 95 Gigabyte zusammen, das nach dem Schneiden immerhin noch 88 Gigabyte umfasste.

Im Herbst 2010 begann Brigitte Rührlinger – zunächst im Alleingang – die zuvor digitalisierten Daten zu taggen, wobei die jeweils 45 Minuten umfassenden Vorder- und Rückseiten der C90-Kassetten in etwa 5 Minuten lange Segmente zerlegt werden sollten. Das Abhören und Schneiden einer Kassetten-Seite dauerte durchschnittlich 40 Minuten und führte zu durchschnittlich neun Schnitten.

Allerdings waren die bei dieser Arbeit erzielbaren Fortschritte sehr unterschiedlich, da manche Aufnahmen ein rascheres, andere wiederum ein weniger zügiges Vorankommen erlaubten. Hilfreich und arbeitsökonomisch war in einer solchen Situation die Technik des „Springens“ von einer Frage zur nächsten. Ein weniger zügiges Vorankommen ergab sich dann, wenn die Gewährspersonen langatmige Digressionen zum Besten gaben, von denen die interessierenden Frage-Antwort-Komplexe sozusagen „zugeschüttet“ waren. In solchen Fällen war ein genaueres Abhören nötig: man konnte nur selten „springen“.

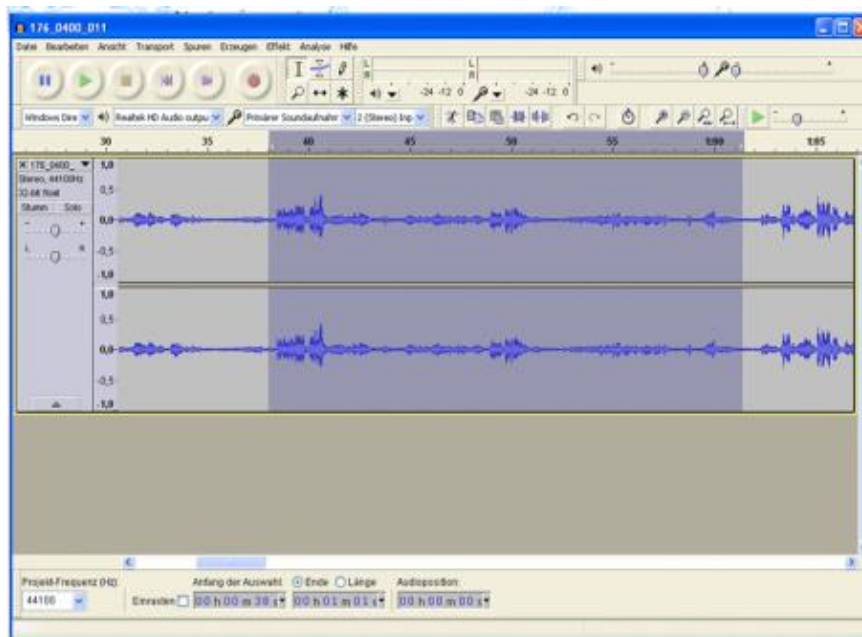


Abbildung 4: Taggen der Tondaten mit *Audacity*

Angesichts der großen Datenmenge und der Kürze des zur Verfügung stehenden Zeitrahmens – die Arbeiten sollten auf jeden Fall vor dem Ende des Jahres 2012 beendet sein – wurden zusätzliche Mitarbeiter in diese Tätigkeit einbezogen. Neben Brigitte Rührlinger waren dies in weiterer Folge Stefanie Holzner, die beiden Italienerinnen Ilaria Adami und Tiziana Gatti sowie meine Wenigkeit. Ilaria Adami war – wie Brigitte Rührlinger – ebenfalls Exploratorin beim ALD-II. Bei den ladinischen Orten unterstützte uns eine Zeitlang auch der Salzburger Studienassistent Christoph Hülsmann. Die Koordination der in und um Salzburg ablaufenden Arbeiten war meine Aufgabe, während die in Italien zu erledigenden Arbeiten von Brigitte Rührlinger (von ihrem dortigen Wohnort Sabbio Chiese aus) überwacht wurden. Allerdings wurden gar manche Details im direkten Gespräch zwischen Ilaria Adami sowie Tiziana Gatti und meiner Wenigkeit besprochen bzw. geklärt.

Insgesamt dauerten die Tagging-Arbeiten 1 ¼ Jahre, was bei einer 30-Stunden-Woche einem Personenarbeitsjahr entspricht. Ab dem Frühjahr 2012 waren wir über mehrere Monate sogar zu sechst mit der „Belauschung“ der Tondaten des ALD-I beschäftigt, wobei wiederum das Schneide- und Akustik-Programm *Audacity* zum Einsatz kam. Die Bearbeitung der letzten 20 Orte lag schließlich in den Händen von Brigitte Rührlinger, Stefanie Holzner und mir selber.

Neben dem Programm *Audacity* schien auch die Verwendung eines weiteren Programms sinnvoll zu sein, das den Arbeitsfortschritt dokumentiert und jedem Mitarbeiter einen Überblick über das Geschehen – zu seiner eigenen Tätigkeit und jener der anderen – gibt. Bernhard Schauer adaptierte dazu das Verwaltungsprogramm *Redmine*.

Die Abbildung 5 zeigt ein leeres Ticket, worin wir Mitarbeiter verschiedene Daten einzutragen hatten, so etwa:

- Zuweisen des Messpunktes an den entsprechenden Mitarbeiter;
- Anfang und Ende der Arbeit;

- Arbeitszeit (Stunden, die man für das Taggen aufgewendet hat);
- Prozentsatz der erledigten Arbeit eines Messpunktes;
- eventuelle Kommentare, wenn Auffälliges zu vermerken war.

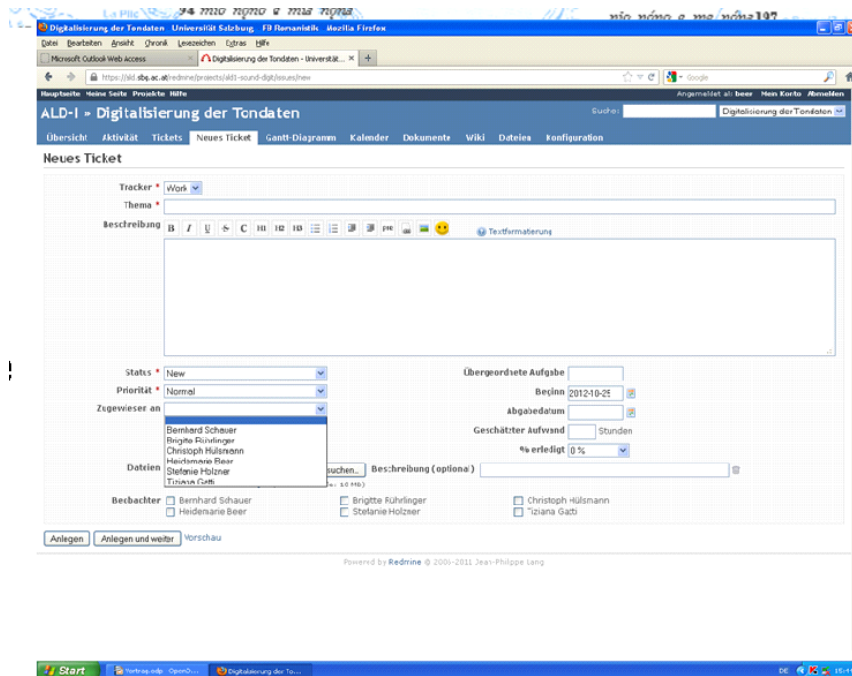


Abbildung 5: Beispielsseite aus Redmine

Für jeden der 217 Messpunkte wurde ein eigenes Ticket angelegt. Diese Tickets wurden – entsprechend dem Arbeitsfortschritt – laufend von den jeweiligen Mitarbeitern aktualisiert.

Ende September 2012 konnten die Schneidearbeiten abgeschlossen werden.

Für das nachfolgende Einpflegen aller neu geschnittenen mp3-Dateien in die Sounddatenbank waren wiederum sehr systematisch aufgebaute Listen unentbehrlich, die über den Messpunkt, die jeweilige Datei und die darin vorhandenen Fragenummern genaue Auskunft gaben.

Die Abbildung 6 zeigt ein Detail aus einer solchen Begleitliste.

Zuletzt wurde bei der Berliner Demonstration ein Hörbeispiel aus der SDB des ALD-II zum Besten gegeben: und zwar zum ALD-Messpunkt 176, San Zeno, der von Brigitte Rühlinger exploriert worden war. An diesem Messpunkt – wo auch eine Nachenquète stattfand – stand Brigitte Rühlinger ein besonders guter Informant zur Verfügung. Zudem ist die Tonqualität der Aufnahme sehr gut.

	A	B	C	
	Ort-Nr	Dateiname/mp3	Fragenummer	Kommentar
1				
2	200	200_0101_001	1-17	
3	200	200_0101_002	18-32	
4	200	200_0101_003	33-53	
5	200	200_0101_004	54-57, 59-67	58 fehlt
6	200	200_0101_005	68-78	
7	200	200_0101_006	79-94	93: il cantone s
8	200	200_0101_007	95-108	
9	200	200_0101_008	109-119	
10	200	200_0101_009	120-132	123: + la chiocc
11	200	200_0101_010	133-155	
12	200	200_0101_011	156-169	
13	200	200_0101_012	170-203	204 (detto) gefr
14				
15	200	200_0102_001	205-226	
16	200	200_0102_002	227-253	
17	200	200_0102_003	254-278	260: ev. auch la
18	200	200_0102_004	279-304	

Abbildung 6: Beispiel einer Begleitliste zur Kontrolle der einzupflegenden Sound-Dateien

Es wurde eine Passage vorgeführt, die auch Amüsantes beinhaltet. Man erkennt daraus, wie unterhaltsam sich manche Enquêtes gestalten können. Das vorgeführte Gespräch bezog sich auf einen Teilbereich der Frage 44 des ALD-II-Questionnaires – *il moccio* („der Nasenschleim“) –, den man nach Eingabe der Nummern 176 (für die Ortschaft) und 44 (für die interessierende Frage) sowie unter Konsultation der Schalldruckkurve leicht auffinden kann. Mit der Maus können nach Belieben kleinere oder größere Segmente zur Abhörung ausgewählt werden. Man kann aber auch – wenn gewünscht – rasch zu einer anderen Frage übergehen.

Jeder, der die Sounddaten des ALD-II – seit kurzem gesellen sich ja auch jene des ALD-I dazu – anhören möchte, kann jederzeit nach eigenem Gutdünken selbst in die Originalaufnahmen hineinhören. Dafür stehen fortan sämtliche Daten im Netz (mit Installations- und Bedienungsanleitung) zur allgemeinen Verfügung:

ALD-I: <http://ald1.sbg.ac.at/a/index.php/de/daten/sound-datenbank/>.

ALD-II: <http://ald2.sbg.ac.at/a/index.php/de/daten/sound-datenbank/>.

6 Zur praktischen Applikation der für ALD-II und ALD-I entwickelten EDV-Werkzeuge: das Programm DMG und die Homepages (Agnes Staudinger)

6.1 DMG (Dialect Map Generator)

Nach Abschluss der Aufnahmen vor Ort wurden die in den Fragebüchern deponierten Transkriptionen in Salzburg händisch in *DMG* übertragen. Die Eingabe der betreffenden Lautzeichen erfolgte alternativ durch Anklicken in einem auf dem Bildschirm sichtbaren Alphabetikum oder durch die Applikation von Tastenkombinationen auf der Tastatur. Die Verwendung der Tastenkombinationen erwies sich dabei als weniger zeitintensiv, allerdings unter der Voraussetzung, dass man zuvor das System gut durchschaut hatte bzw. sogar auswendig kannte: siehe dazu die Abbildung 7.

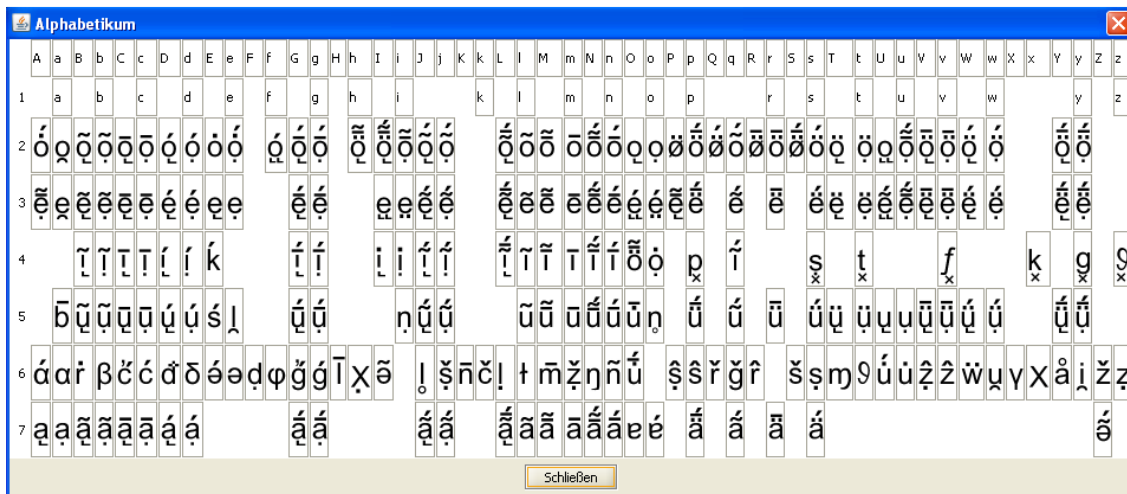


Abbildung 7: Alphabetikum der beim ALD (Teil I und Teil II) verwendeten Lautschrift

So entspricht eine *Ziffer* (1-7: als *Zeilen* organisiert) hauptsächlich den verschiedenen Varianten eines Vokals: *Zeile 2*: o-Laute, *Zeile 3*: e-Laute, *Zeile 4*: i-Laute, usw. Die *Buchstaben* (Aa-Zz; als *Spalten* organisiert) definieren dagegen den Typ: *Spalte D*: offen und betont, *Spalte d*: geschlossen und betont, usw.

Drei konkrete Anwendungsbeispiele: 2d: geschlossenes betontes o; 3d: geschlossenes betontes e; 4d: geschlossenes betontes i, etc.

Auf Abbildung 8, links oben, sieht man die Frage 2, die aus vier Teilfragen (*Dove...saranno...mio nonno e...mia nonna?*) besteht, als Baum. Es wird gerade die Transkription der dialektalen Antwort auf die Frage „Dove“ eingegeben. Man hat nun (z. B. bei zwischen Frage und Antwort divergierender Syntax) die Möglichkeit, bei der Eingabe die Reihenfolge der zu erfassenden Teil-Antworten/Fragen zu vertauschen, und kann außerdem (bei Doppel- oder Mehrfachantworten) eine weitere Version hinzuzufügen.

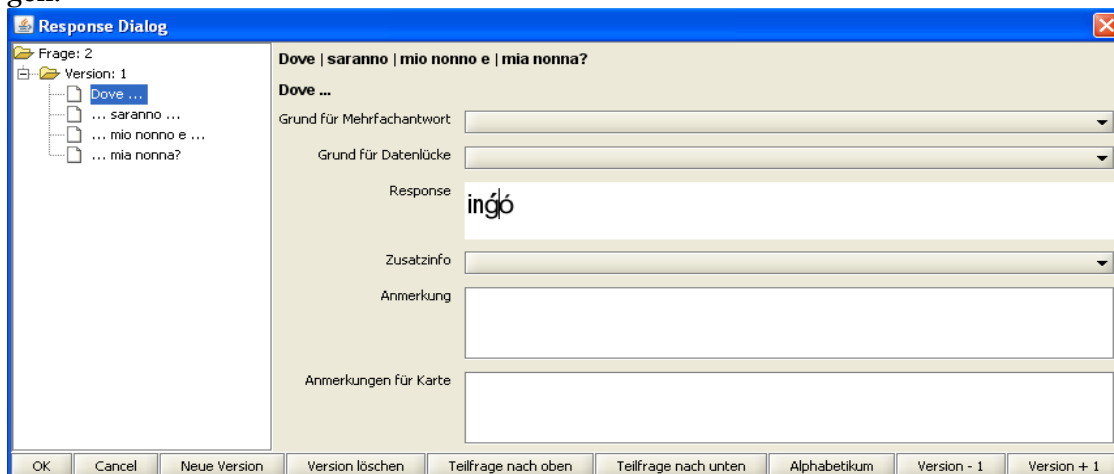


Abbildung 8: Eingabefeld von DMG zur Frage 2 des ALD-II-Fragebuchs

Nach der Eingabe und Korrektur aller Transkriptionen konnte der erste Ausdruck von Probe-Karten realisiert werden. Diese wurden daran anschließend aus philologisch-linguistischer Perspektive korrigiert. Die Erzeugung solcher Karten mittels *DMG* war

relativ komplex und damit sehr zeitaufwändig. So waren für die Produktion einer Karte bis zu 11 Mausklicks nötig, um ein ausdrucksfähiges PDF zu erzeugen.

Bei den beiden ersten der insgesamt drei Kampagnen zur Produktion von Probe-Karten wurden bei komplexeren Fragen fast alle Möglichkeiten der Kombination von Einzel- und Teilfragen (als „Kombi-“ [KK] und „Doppel“-Karten [DK]) ausprobiert. Dadurch kamen pro Produktions-Kampagne für die 1 063 Fragen des ALD-II-Questionnaires insgesamt 2 026 Karten zusammen.

In den Jahren 2009-2012 wurden die PDFs einerseits für drei intermediäre Produktions-Kampagnen (zu je 2 026 Karten) und andererseits für den finalen Durchgang (mit 1 066 zur Publikation ausgewählten Atlas-Karten) erstellt: das sind in toto (= 3 mal 2 026 + 1 066) 7 144 Karten-PDFs, für deren Generierung 22 286 Mausklicks nötig waren. Rechnet man dies auf die dafür benötigte Zeit um (das Erstellen und Abspeichern eines Karten-PDFs dauert ca. 1,5 Minuten), dann kommt man auf 178,6 Stunden reiner Arbeitszeit bzw. auf 7,14 Personenarbeitswochen (bei 30 Stunden/Woche).

Der Ausdruck einer Karte auf DIN A2 großem Papier nahm mit unserem Vierfarbendrucker *HP designjet* durchschnittlich vier Minuten in Anspruch. Alle vier Durchgänge wurden sowohl auf Prüfpfad-Vorlagen, deren Verwendung die Korrektur erleichtert, als auch auf Blaudruck-Vorlagen, die dem definitiven Erscheinungsbild mit blauem Kartenhintergrund entsprechen, realisiert. Somit kommt man auf ca. 14 300 Karten, deren Ausdruck 953 Stunden benötigte: dem entsprechen bei 30 Stunden/Woche umgerechnet 38,1 Personenarbeitswochen.

Parallel zu den Kartenausdrucken wurden zu Korrekturzwecken verschiedene Begleitlisten angefertigt. Auf ihnen wurden die Karteninhalte z.B. in vorwärts alphabetischer oder in rückwärts alphabetischer Form bzw. auch nach Orten sortiert (1 bis 217) präsentiert. Auch gab es Listen, die jene Informationen beinhalteten, die später im Supplementband („Volumen supplementarium“) veröffentlicht wurden, und solche, die nur die Anmerkungen und die Zusatzinformationen einer Karte enthielten.

Die PDFs dieser Listen wurden wie jene der Karten ebenso mit *DMG* erzeugt, was pro Liste ca. 30 Sekunden oder 13 Mausklicks in Anspruch nahm. Bei den im Verlauf von zwei Kampagnen manuell erzeugten 9 556 Listen kamen wir so auf 124 228 Klicks bzw. eine Arbeitszeit von 80 Stunden.

Den Ausdruck der Listen führten wir bei den ersten beiden Durchgängen noch am ALD-Archiv durch: dies betraf insgesamt 74 496 Seiten. Da unser *HP LaserJet* ca. 15 Seiten/Minute schaffte, ergab sich daraus eine reine Druckzeit von 82,8 Stunden.

Da der projekt-eigene Drucker damit seine Leistungsgrenze erreicht hatte und zu „schwächeln“ begann, beschlossen wir, die maschinell um vieles besser ausgerüstete Kopierstelle der Universität Salzburg mit dem Ausdruck der definitiven Listen zu betrauen, die nach dem Ausdruck archiviert wurden. Diese letzte Erzeugung von listenartig organisierten PDFs wurde außerdem nicht mehr mittels, sondern mit dem neuen, von Bernhard Schauer erstellten Programm *SuBIReS* („*Supplementary Book and Index Retrieval System*“) durchgeführt. Für die Generierung aller 4 778 Listen eines kompletten Durchgangs benötigte *SuBIReS* ca. 2,5 Stunden, wobei es im Hintergrund lief und man sich daneben anderen Dingen widmen konnte.

Der Start von *SuBIReS* dauerte pro Listenart ca. 10 Sekunden. Diese Aktion und die nachfolgende PDF-Erzeugung mussten insgesamt vier Mal repetiert werden. Daraus ergab sich eine Gesamtzeit von etwa einer Minute für den Start. Für die Erzeugung und den Druck der Listen gelangten wir somit auf 162,8 Stunden reiner Arbeitszeit, also auf 6,5 Personenarbeitswochen (bei 30 Stunden/Woche).

Insgesamt ergab sich damit ein ganzes Personenarbeitsjahr (bei 30 Stunden/Woche) für die Erzeugung und den Druck aller Karten und Listen.

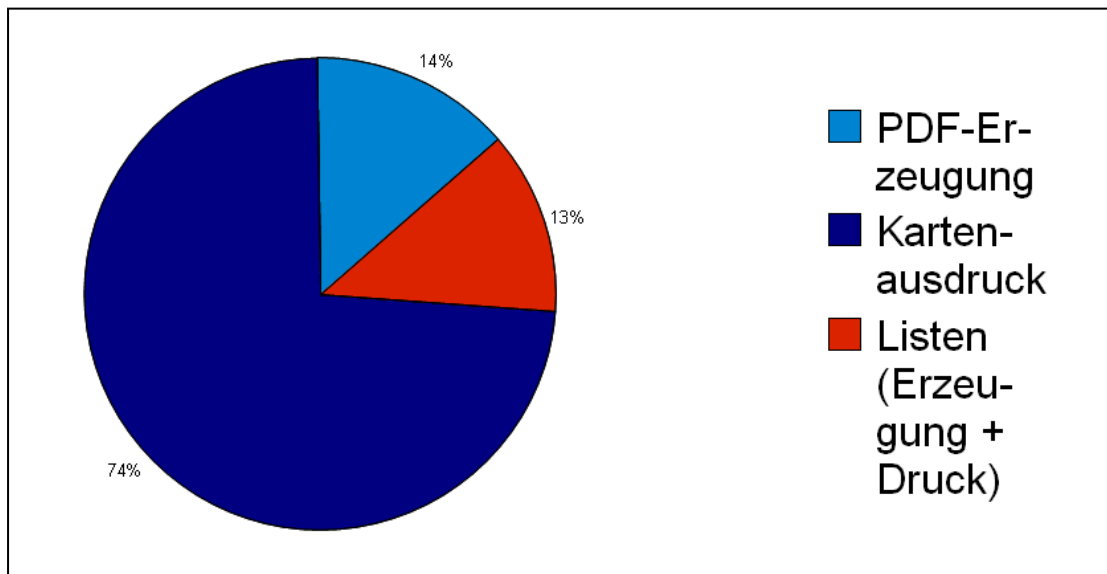


Abbildung 9: Aufteilung eines Arbeitsjahrs nach Tätigkeiten

In dieser Arbeitszeit sind allerdings das Falten der A2 großen Kartenausdrucke, das Zusammenheften der (A4 großen) Listen, das Abstempeln beider sowie deren Einordnung in die A3 großen Hängeordner nicht enthalten.

6.2 Ethnophotographische Daten

Die Exploratoren fertigten im Zuge der Enquêtes auch Dias von den besuchten Orten an. Darauf wurden unter anderem die Gewährspersonen selber, die jeweilige Dorfkirche samt Turm und Innenraum, der Friedhof, Denkmäler und das Straßenbild festgehalten. Beim Friedhof war es für uns von besonderem Interesse, ob es dort nur Erdgräber (ital. *tombe*) oder auch Nischengräber (ital. *loculi*) gab. Die Bestattung in *loculi* hat in den südlichen Teilen der ALD-Zone Tradition, breitet sich aber immer mehr nach Norden hin aus und bringt somit eine Änderung der Begräbnis- und Grabkultur mit sich.

An den 217 Messpunkten wurden sowohl beim ALD-I als auch beim ALD-II jeweils 30-40 Aufnahmen gemacht, wodurch es möglich wurde, einerseits zwei synchrone Bild-Dokumentationen aller Aufnahmeorte zu erhalten und andererseits die diachrone Entwicklung jedes Ortes über jene 15 Jahre hinweg, die zwischen ALD-I und ALD-II liegen, zu beobachten.

Insgesamt gibt es von beiden Atlasteilen ca. 17 000 Aufnahmen. Davon bestand der Großteil aus analogen Diapositiven, die von einer Mitarbeiterin (Stefanie Holzner) mittels Scan digitalisiert wurden. Unser Plan war, bis Ende des des Jahres 2012 eine (grob getaggte) Bilddatenbank aufzubauen und mit Daten zu befüllen, die es Außenstehenden erlauben sollte, darin unter Eingabe der Nummern der 217 Ortschaften und verschiedener Oberbegriffe (wie „Kirche“, „Friedhof“, „Informant“ etc.) zielorientiert oder enzyklopädisch zu suchen. Leider konnte dieses Ziel nicht erreicht werden.

6.3 Homepages

Die Homepage des ALD-II wurde mit einem CMS erstellt, dessen Vorteil darin besteht, dass Personen ohne spezielle Programmierkenntnisse Inhalte hinzufügen und bearbeiten können. Es erschien uns dabei wichtig, der Fachwelt neben grundsätzlichen Informationen zum ALD-II auch konkretes Datenmaterial zugänglich zu machen. So kann sich jeder das komplette Fragebuch, stumme Karten zum Untersuchungsnetz und sogar alle 1 066 Karten des ALD-II als PDFs herunterladen und bei Bedarf auch ausdrucken. Außerdem gibt es auf der Homepage Statistiken zu den Informanten, den Exploratoren und zu den enquêtierten Ortschaften.

Man kann nach einzelnen Karten und auch nach Wörtern, die in den (italienischen) Kartentiteln vorkommen, suchen. Über die *SDB2* des ALD-II lassen sich die 217 im Feld gemachten Aufnahmen punktgenau anhören und die 1 066 Sprachkarten mittels des digitalen Index (IRS2) in sehr gezielter Weise absuchen. Auch findet man auf unserer Homepage ein Flußdiagramm zu den über die Jahre gemachten Arbeitsfortschritten sowie eine ständig à jour gehaltene Übersicht über den jeweiligen Stand einzelner Arbeitsschritte.

The image shows the German version of the ALD-II homepage. At the top, there is a banner with a mountain landscape. The title "ALD-II" is prominently displayed in a large, stylized red font. Above and around it are several variations of the word "bondi" in different dialectal scripts: "bunđi", "bundé", "bondí", "bondi", "bunđórno", and "bonjórno". Below the banner, there are small flags representing the languages available: Italian, German, English, and French. The main content area is divided into a left sidebar and a main text area. The sidebar lists navigation options: "Das Projekt", "Fragebuch", "Untersuchungsnetz", "Mitarbeiter", "ALD-Archiv", "Aktueller Status", "Zeitplan", "Statistiken", "Daten", "Technik", "Publikationen", "Links", "Förderer", and "Impressum". The main text area has the title "Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 2. Teil" and a paragraph of text describing the project as a continuation of the 1998 project. Below this is a section titled "Theoretische Grundkonzeption" which provides more details about the project's methodology and goals.

Abbildung 10: Homepage des ALD-II, deutsche Version

Um der Homepage des ALD-II eine größtmögliche internationale Sicht- und Lesbarkeit zu verschaffen, gibt es davon fünf sprachliche identische Versionen: und zwar neben Deutsch auch in Ladinisch (Ladin dolomitan), Italienisch, Englisch und Französisch.

bunđí *bundé* *bonđí* *bondí* *bonjórno*

ALD-II

L projet
 Chestionar
 Rei de enrescida
 Colaboradours
 Archif dl ALD
 Situazion atuala
 Program
 Statistiches

Dac
 Tecnica
 Publicazions
 Links
 Promotours
 Impressum

Atlant linguistich dl ladin dolomitich y di dialec vejins, 2ª pert

L ALD-II é la continuazion dl [ALD-I](#) publiché dl 1998 (te 7 volums y 3 CD-ROM). Depierpul che per l ALD-I fôvel vegnù enresci la fonetica y la morfologia nominala y verbala elementara, se concentreia l ALD-II sun l lessich, la morfologia plu menuda y la sintassa. La conzezion de basa teorica ("filosofia dl projet"), la metoda de enrescida y la rei de analisa resta unfat sciche pro l ALD-I.

Conzezion de basa teorica

L projet ALD ti vâ do a la tradizion dla geografia linguistica romanza, pîeda via con l ALF de Jules Gilliéron y con l AIS de Karl Jæberg y Jakob Jud. L fin dl ALD é chel de cueie adum y documenté les competenzes dialetoles (vuel di: basiletoles) de persones che reiona en linia de prinzip deplù lingac y che vif (o se sent da cîasa) te un di 217 ponc de mesurazion dl raion de analisa. Chestes persones „aministreia“ si raion de vita tres en juech compliché de convergenzes y divergenzes dialetoles, che i dac coius adum per l ALD dess daidé conesce

Abbildung 11: Homepage des ALD-II, ladinische Version

Analog zur Homepage des ALD-II wurde auch die Homepage des ALD-I in einem neuen Design erstellt, da die auf die 1990er-Jahre zurückgehende alte Homepage weder optisch noch inhaltlich à jour war und eigentlich nur mehr historischen Wert besaß. Allerdings stellte sie bis Mitte 2012 die einzige netzbasierte Informationsquelle zum ALD-I dar. Sie wurde nunmehr komplett überarbeitet, sodass die beiden neuen Homepages einander nicht nur im Design, sondern auch in der Menüführung gleichen.

Am Ende des Jahres 2012 war die neue Homepage des ALD-I vorerst in den Sprachen Deutsch, Italienisch und Französisch verfügbar. Die restlichen zwei Sprachen (Ladin dolomitan und Englisch) sollen später eingearbeitet werden.



bundí *bundé* *bondí* *bondí* *bondí* *bunǰǰorno* *bondi* *bunǰǰorno*

ALD-I

Das Projekt

- Fragebuch
- Untersuchungsnetz
- Mitarbeiter
- ALD-Archiv
- Zeitplan
- Statistiken

Das Teilprojekt ALD-I wurde im Jahr 1985 begonnen und im Jahr 1998 mit der Publikation von 7 Bänden und 3 CD-ROM abgeschlossen. Auf den 884 publizierten Sprachkarten werden vornehmlich Themen der Lautlehre sowie der elementaren nominalen und verbalen Morphologie behandelt. Über die 806 Fragen des Questionnaires des ALD-I sollten nicht nur erbwörtlich entwickelte Materialien, sondern auch jüngere und deutlich „gelehrt“ (d.h. lateinisch und/oder italienisch) beeinflusste Lexeme elizitiert werden, um solcherart auch die sekundäre Wirksamkeit der örtlichen Lautgesetze bzw. -trends studieren zu können (siehe dazu die ALD-I-Karten 180 *il comune / i comuni*, 360 *il giudice*, 382 *l'inferno*, 401 *la legge* etc.). Die extensive Behandlung der elaborierten Morphologie, der Syntax und des Lexikons erfolgte im Rahmen des zwischen 1999 und 2012 realisierten zweiten Projektabschnitts (ALD-II).

Theoretische Grundkonzeption

Hinsichtlich des Forschungsziels und angewandten Methode stehen sowohl der ALD-I als auch der ALD-II zur Gänze in der Tradition der von Jules Gilliéron (1854-1926) mit dem ALF

Abbildung 12: Homepage des ALD-I, deutsche Version

6.4 Texmaker

Das Vorwort und die Deckblätter der Kartenbände sowie die Indices und der Supplementband wurden mit dem *LaTeX*-Programm *Texmaker* erstellt. Der Vorteil dieses Programms besteht in einer relativ leichten Handhabung bei gleichzeitigem professionellem Design, wobei es zu keinerlei Verrutschen oder anderen Formatierungsproblemen kommen kann, wie sie beispielsweise bei Word nicht selten sind.

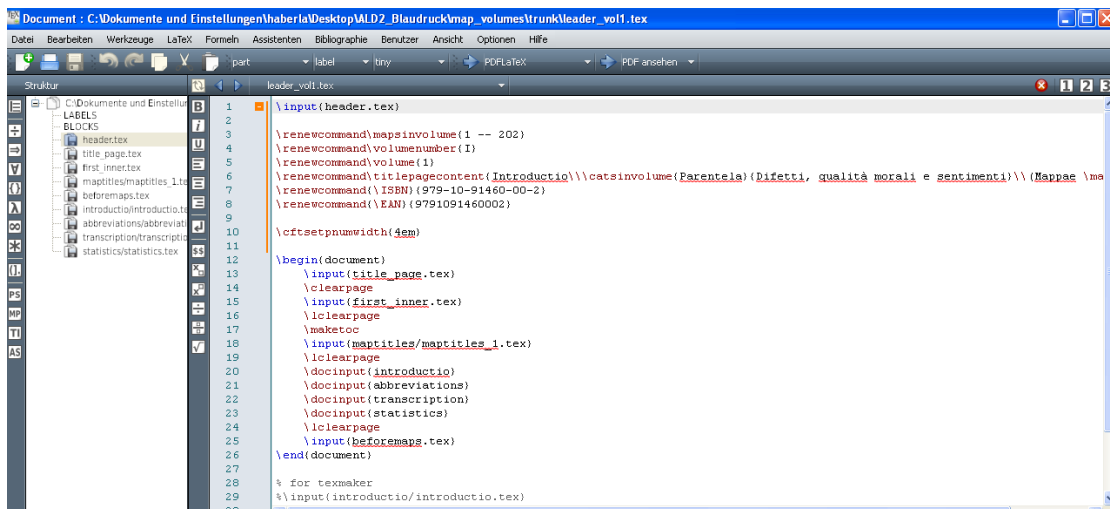


Abbildung 13: ein mit LaTeX erstelltes Dokument

Links sieht man den Aufbau des Dokuments, in der Mitte das Dokument selbst, oben lässt sich mit einem Mausklick ein PDF erzeugen („*PDFLaTeX*“) und ansehen („*PDF ansehen*“). Außerdem gibt es, ähnlich wie bei Word, im Menü viele nützliche Hilfen, so dass

man nicht alle Befehle selbst auswendig kennen muss. So ist es beispielsweise möglich, den Text mit einem Mausklick kursiv zu setzen, ohne dazu eigens den Befehl „\textit{}“ einfügen zu müssen.

Bibliographie

- ALF: Gilliéron, Jules/Edmont, Edmond (Hg.): Atlas linguistique de la France, Paris: Champion, 1902-1910, 10. vol. (Neudruck: Bologna: Forni, 1968).
- ALD-I: Goebel, Hans (Hg.), unter Mitarbeit von: Bauer, Roland/Haimerl, Edgar/Böhmer, Helga/Gislimberti, Silvio/Kattenbusch, Dieter/Perini, Elisabetta/Szekely, Tino/Dautermann, Irmgard/Heißmann, Susanne/Hofmann, Ulrike/Kozak, Anna/Pamminger, Heide Marie/Rössler, Judith: *Atlant linguistisch dl ladin dolomitich y di dialec vejins, 1^a pert. Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 1^a parte. Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 1. Teil*, Wiesbaden: Dr. L. Reichert, 1998, 4 vol. Karten (vol. I: 1-216; vol. II: 217-438; vol. III: 439-660; vol. IV: 661-884), 3 vol. Indizes (Index alphabeticus, 823 S.; Index alphabeticus inversus; 833 S.; Index etymologicus, 177 S.) mit 3 CD-ROM (Salzburg 1999-2000), 1 DVD (Salzburg 2002f.).
- ALD-II: Goebel, Hans (Hg.), unter Mitarbeit von: Haimerl, Edgar/Schauer, Bernhard/Tosques, Fabio/Wagner, Andreas/Adami, Ilaria/Böhmer, Helga/Heinemann, Axel/Jodl, Frank/Klinger, Liza/Rando, Daniele/Rührlinger, Brigitte/Strauß, Walter/Szekely, Tino/Videsott, Paul/Beer, Heidemarie/Staudinger, Agnes: *Atlant linguistisch dl ladin dolomitich y di dialec vejins, 2a pert. Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 2a parte. Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 2. Teil*, Strasbourg: Éditions de Linguistique et de Philologie, 2012, 5 vol. Karten (vol. I: 1-202; vol. II: 203-420; vol. III: 421-635; vol. IV: 636-850; V: 851-1066), 2 vol. Indizes (Volumen supplementarium 174 S.; Index generalis, 213 S.).
- Bauer, Roland (2009): *Dialektometrische Einsichten. Sprachklassifikatorische Oberflächenmuster und Tiefenstrukturen im lombardisch-venedischen Dialektraum und in der Rätorumnia*, San Martin de Tor.
- Beer, Heidemarie/Staudinger, Agnes (2012): *Die praktische Applikation der für den ALD-II entwickelten EDV-Werkzeuge*. [Online] 2012. [Zitat vom 12. 18. 2012.] http://www2.hu-berlin.de/vivaldi/tagung/fohlen/Staudinger_Beer.pdf.
- Mell, Peter M./Grance, Timothy (2011): *SP 800-145. The NIST Definition of Cloud Computing*. Gaithersburg, MD: National Institute of Standards & Technology.