

**ZUR ROLLE DES COMPUTERS BEIM  
DOLOMITENLADINISCHEN SPRACHATLAS (AD-1)\***

Sprachatlasunternehmungen sind überaus corpusintensiv und gehören somit zu jener Art von Forschungsprojekten, die riesige Datenmengen zu bewältigen haben. Auch oder gerade deshalb geht ihre Erstellung seit geraumer Zeit (etwa seit dem Ende der 80-er Jahre) mit massiver EDV-Unterstützung vor sich. Wo früher die Herstellung und Verwaltung voluminöser Zettelkästen den Arbeitsalltag dominierte, stehen heute Mikrocomputer zur Verfügung, mit deren Hilfe *a priori* viele der anfallenden Arbeiten besser, schneller und sicherer, somit insgesamt auch effizienter erledigt werden können. Die ehemaligen Handzettelkästen sind in vielen Fällen bereits ihren elektronischen Nachfolgern gewichen, doch darin liegt noch kein wesentlicher Unterschied zwischen Atlasarbeit einst und jetzt. Früher diente die händische Aufbereitung von Sprachatlasdaten v.a. einem Ziel, nämlich der Publikation des Druckwerks Atlas selbst. Die entsprechenden Archive blieben als Zeugen für die getätigte Knochenarbeit in seltenen Fällen erhalten, für Außenstehende jedoch meist schwer zugänglich. Heute freilich sieht die Situation völlig anders aus. Die bloße Produktion des Druckwerks Sprachatlas ist ein wohlgerneht wichtiges Ziel, aber nur ein Teilziel der EDV-gestützten Aufarbeitung geolinguistischer Datencorpora, was von den Betreibern der Einzelprojekte leider noch allzu oft übersehen wird. Neben einem Atlas *in folio*, dessen Wert durch computerunterstützte Erstellung noch lange nicht automatisch steigt, müssen weitere Datenkonserven übrigbleiben, die im Gegensatz zu traditionell aufbereiteten Sprachatlasdaten (Karten und Indizes auf Papier) eine neue Art des Datenkonsums von seiten der interessierten Fachwelt zulassen. Die Notwendigkeit, Sprachatlasdaten *ex toto* auch auf elektronischen Datenträgern zu publizieren, scheint *a priori* keiner besonderen Rechtfertigung zu bedürfen, liegen doch die daraus für den allgemeinen wie den besonderen Umgang mit den Daten erwachsenden Vorteile auf der Hand. Trotzdem finden sich in den Reihen der internationalen Sprachatlasszene immer wieder Leute, die den Einsatz des Computers zur (vermeintlich) schnelleren, billigeren und mitunter auch schöneren Produktion des Kartenwerks zwar bejahen, jede weitere Nutzung elektronisch gespeicherten geolinguistischer Daten aber kategorisch ablehnen, ja oft nicht einmal dazu bereit sind, die entsprechenden maschinenlesbaren Rohdaten an die interessierte Fachwelt weiterzugeben. In diesem Zusammenhang schwingt auch ein *forschungsethischer* Aspekt mit, dessen Diskussion nicht nur den Rahmen dieses Beitrags, sondern v.a. die Bezugnahme auf die Rätoromanistik bzw. auf die Sprachgeographie sprengen würde. Daher

\*) Vortrag, gehalten anlässlich des *Rätoromanischen Kolloquiums* (21.-24. März 1996) in Gießen/Rauischholzhausen.

konzentrieren wir uns hier auf die *forschungsstrategischen* Perspektiven, die sich aus der Verquickung von genuiner (und somit durchaus in einer jahrzehntelangen disziplininternen Tradition stehender) Sprachatlasarbeit mit innovativem, häufigen und raschen Paradigmatawechselln unterworfenen Computereinsatz ergeben.

Eine wesentliche Rolle spielt in diesem Kontext der Startzeitpunkt eines Sprachatlasprojekts, wobei man mit Bezug auf die letzten Dezennien durchaus vom Glück, um nicht zu sagen von einer Art "Gnade der späten Geburt" sprechen kann. Damit soll zum Ausdruck kommen, daß ein EDV-gestütztes Projekt - dank der immer breiter um sich greifenden Etablierung elementarer informatischer Strukturen in unser aller Berufs- und auch Privatwelt - einen umso leichteren Einstand findet, je jünger es ist. So war im Falle des in Salzburg entwickelten dolomitenladinischen Atlases (*AD-I*) massiver Computereinsatz zwar von Beginn an, d.h. ab etwa Mitte der 80-er Jahre, Hauptbestandteil des Projektdesigns, doch standen damals weder breit einsetzbare und zukunftsichere Strukturen (geschweige denn Standards) im Bereich der Hardware, noch ausreichend erprobte Software zur Verfügung, so daß unsere erste Marktorientierung in vielen Punkten einem "Sichtflug im Nebel" gleichkam.

Soviel stand für uns aber immer schon fest: Die EDV sollte uns helfen, den Atlas in mehrfacher Form zu veröffentlichen, und zwar erstens als gedrucktes Kartenwerk, zweitens als dazupassende Datenbank und drittens als *Sprechenden Sprachatlas*. Aus heutiger Sicht können wir rückblickend zweierlei bemerken: einerseits, daß wir offensichtlich einigermaßen Intuition betreffend Produkt- und Systemauswahl beweisen konnten, und andererseits, daß uns bei alledem wohl auch Fortuna kräftig zur Seite gestanden ist. Nur einer Kombination dieser beiden Faktoren ist es zu verdanken, daß die EDV-Strukturen des *AD-I* auch nach zehn Jahren noch immer zeitgemäß und an aktuellen Standards orientierbar geblieben sind.

Dazu ein Nachsatz: alles bislang mit Computerunterstützung Entwickelte ist mit extrem kurzen Halbwertszeiten behaftet und wird es wohl auch in Zukunft bleiben. Sowohl unser Maschinen- als auch der Programmpark haben in den letzten zehn Jahren gut und gern fünf Entwicklungsgenerationen durchlaufen, wobei den ersten Gerätekonfigurationen mittlerweile Schrottwert bzw. Museumstauglichkeit zugesprochen werden muß. Dies bedeutet, daß wir es im Schnitt *alle zwei Jahre* mit einem mehr oder weniger tiefgreifenden Systemwechsel auf Hard- und/oder Softwareebene zu tun haben! Einst mühsam und aufwendig im Pilotverfahren für eine Einzelanwendung entwickelte Module gehören heute bereits vielfach zum Lieferumfang eines Standard-PCs und sind somit jedem Anwender leicht zugänglich; denken wir beispielsweise an Multimedia-Anwendungen und Netzwerke. Andere, seinerzeit unter ebenfalls hohem Ressourcenaufwand erstellte Lösungen hingegen harren *bis dato* einer weltweiten Vereinheitlichung bzw. einer Standardisierung seitens der großen Softwarehäuser; es sei lediglich auf das für alle Linguisten bzw. Textwissenschaftler leidige Problem der Kodierung bzw. Darstellung philologischer Sonderzeichen (welcher Art auch immer) verwiesen.

Doch nun zum *AD-I*. Die Projektgeschichte unseres Atlases läßt sich grob gesprochen in vier Phasen unterteilen, und zwar erstens in jene der Vorbereitung (etwa 1972-85), zweitens in jene der Datensammlung (1985-92), drittens in jene der Datenverarbeitung (1987-1996) und schließlich viertens in die Publikationsphase (1996/98). Das Untersuchungsnetz umfaßt 217 Ortspunkte, an welchen ein rund 1.500 Einzelfragen beinhaltendes Questionnaire unter Einbeziehung von jeweils mindestens zwei Informanten mittels standardisiertem Interview vor Ort abgefragt wurde. An 214 Meßpunkten wurden die lokalen Basilekte ( $L_1$ ), in drei südbairischen Sprachinseln (Florutz / Fierozzo, Lusern / Luserna, Zahre / Sauris) hingegen die romanischen Mesolekte ( $L_2$ ) erhoben. Die gesammelten Daten wurden, ebenfalls vor Ort, mittranskribiert und auf Band aufgezeichnet. Das Fragebuch des *AD-I* konzentriert sich auf v.a. phonetische Aspekte. Lexikon, elaborierte Morphologie und Syntax sind einem zweiten, nach Abschluß von *AD-I* in Angriff zu nehmenden Projektteil (ALD II) vorbehalten. Unser Transkriptionssystem hält sich grosso modo an das von K. Jaberg und J. Jud beim AIS verwendete Schema (für weiterführende Informationen zum *AD-I* vgl. Bibliographie).

Wir haben eingangs von mehreren zu erstellenden Varianten des *AD-I* gesprochen. Sie alle greifen mittlerweile auf ein- und dieselbe Datenbasis zu, nämlich auf eine etwa 50 MB umfassende, *X-base-kompatibel* strukturierte Datenbank mit knapp 350.000 Datensätzen. In der Übersicht zu unserem Datenbankverwaltungssystem (siehe Abb.1) scheint diese Datenbank an zentraler Stelle auf und ist durch Schraffur hervorgehoben. Durch Verbindungslinien angedeutet bestehen weiters vielfache Relationen zwischen der Hauptbasis und ausgelagerten Datenbanken. So finden wir z.B. Verbindungen zu den Informanten (Abb.1, ganz links), zu den Meßpunkten (Abb.1, ganz rechts) oder zum sprechenden Sprachatlas (Abb.1, oben Mitte). Die Relationen zwischen Haupt- und Nebendatenbanken werden meist über drei numerische Felder gebildet, die es erlauben, jedes einzelne Datum eindeutig zu identifizieren. Es sind dies die Nummer des Abfrageortes, die dem Fragebuch entsprechende Nummer des Abfragestimulus und die Nummer der zugehörigen Antwort, hier als Versionsnummer bezeichnet. Die drei genannten Felder bilden also eine Art Identifikationscode, über welchen Zuordnungen zu anderen Datentypen festgelegt und alle Daten eindeutig lokalisiert werden können. Was die Genese unseres Datenbestandes betrifft, so ist ein Punkt besonders hervorzuheben. Die Struktur der Daten hat sich seit Projektbeginn nicht verändert, wiewohl sich die heutigen Rahmenbedingungen auf Hard- und Softwareseite grundlegend von jenen, die noch vor zehn Jahren üblich waren, unterscheiden. Dadurch soll auch zum Ausdruck kommen, wie wichtig eine transparente und eineindeutige Strukturierung von Daten ist. Denn nur eine solche gewährleistet Kompatibilität auch für die Zukunft, d.h. daß die in den 80-er Jahren unseres Jahrhunderts

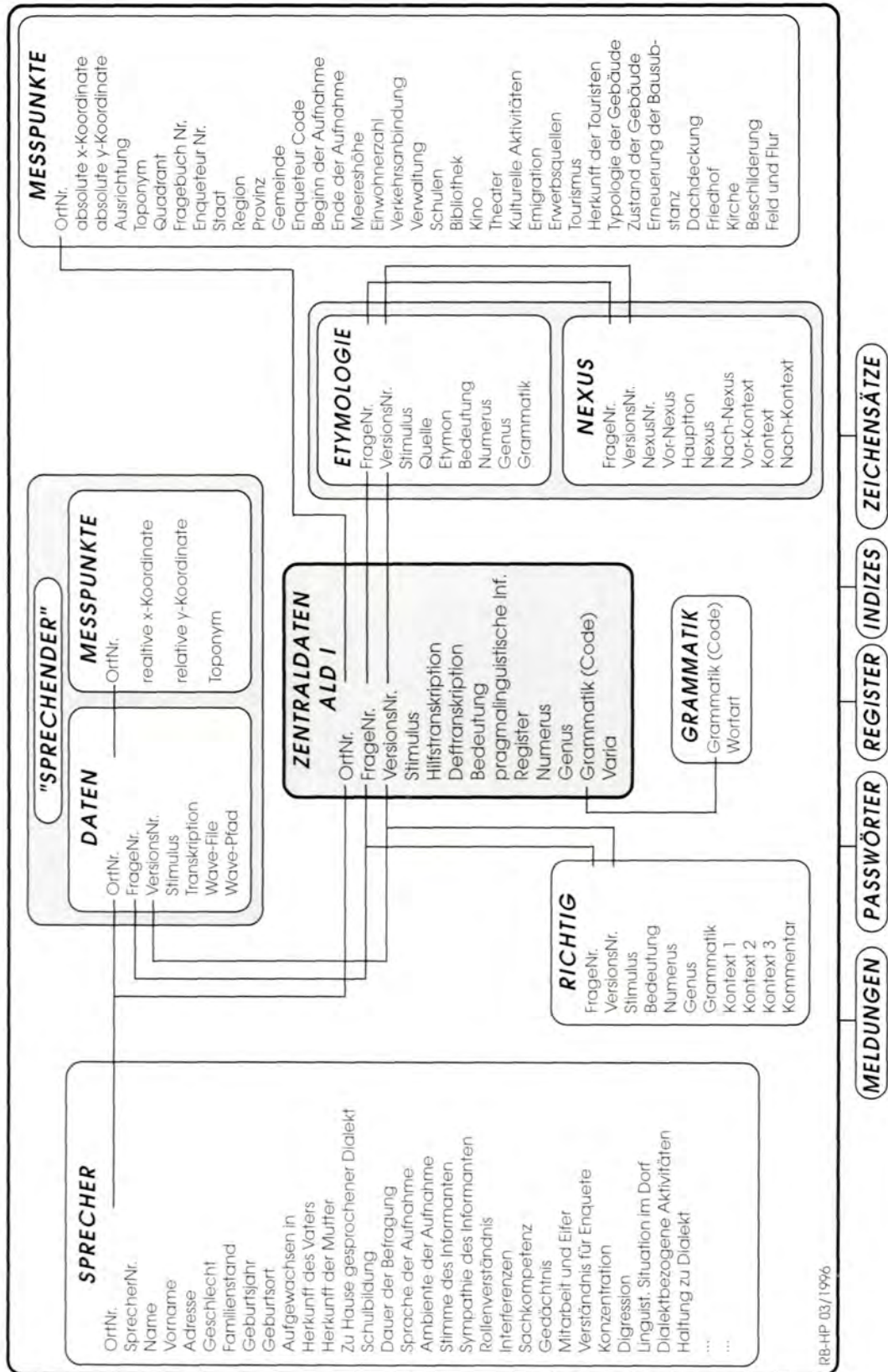


Abbildung 1: Datenbankverwaltungssystem des *Ap-1*

erfaßten Rohdaten mit neuen Methoden und Instrumenten auch im nächsten Jahrtausend weiterbehandelt werden können.

Die mit dem (aus dem englischen *Data Base Management System* = DBMS abgeleiteten) deutschen Kompositum *Datenbankverwaltungssystem* umschriebenen Strukturen sind schematisch in Abb.1 dargestellt. Die *AD-I*-Datenbanken stellen Futter für verschiedene Anwendungsprogramme dar. Das in der Informatik gutbekannte Prinzip *Garbage In - Garbage Out* stellt zunächst große Anforderungen an die Datensammlung in puncto Strukturtreue und Fehlerarmut. Daher wurden sämtliche Datenbestände vor ihrer Weiterverarbeitung einer in mehreren Schleifen ablaufenden Datenkorrektur unterzogen. Allein dieser Arbeitsgang hat ein mehrköpfiges *AD-I*-Team über knapp vier Jahre beschäftigt (vgl. dazu die graphische Übersicht in *Ladinia* XIX (1995), 314). Die im Zusammenspiel von manuellen und automatisierten Routinen realisierte Datenbereinigung stellt unseres Wissens nach, zumindest in dieser Größenordnung, ein Novum dar, weshalb der entsprechende Arbeitsablauf in unseren periodischen Arbeitsberichten (vgl. *Ladinia* XV/1991, XVI/1992, XVII/1993 und XIX/1995) auch detailliert kommentiert wurde. Wir sind uns zwar bewußt, daß auch ein fertiger *AD-I* niemals gänzlich fehlerfrei oder lückenlos sein wird. Die uns durch Computersysteme auferlegte Sorgfaltspflicht wird jedoch im Endeffekt wesentlich dazu beitragen, daß die in der Vergangenheit oft durch Vernachlässigung elementarer Datenrevisionen bekannten Nebenwirkungen im Falle unseres Atlases verträglicher ausfallen werden.

Eines jener Anwendungsprogramme, das sich auf unsere zentrale Datenbasis stützt, ist mit dem Sigel CARD überschrieben. Dieses kann einerseits in romanischsprachige Vollformen wie *Cartografia dialettale/dialettologica* oder *Cartographie dialectale/dialectologique* aufgelöst werden, kann aber, je nach Gefallen, andererseits auch im Sinne der im Angloamerikanischen hochfrequenten CA-Abkürzungen als "*Computer Aided Retrieval of Dialect Data*" interpretiert werden. CARD leistet in der Tat zwei Hauptfunktionen, nämlich zunächst die Suche nach und die Zusammenstellung von Daten nach mehr oder weniger geolinguistischen bzw. auch genuin innerlinguistischen Kriterien. So wird das Programm in seiner endgültigen Version die Rolle eines elaborierten *Data Retrieval*-Werkzeugs übernehmen, wenn es um die *On Line*-Benutzung des elektronischen *AD-I* geht. Es wird Datenauswahl, elementare Datenanalyse und multiple Datenausgabe nach individuellen Kriterien gestatten. Die zweite Hauptfunktion betrifft die Erstellung von elektronisch lesbaren Druckvorlagen für Sprachatlaskarten. Die einzelnen Menüpunkte von CARD sind in unseren Arbeitsberichten im Detail beschrieben (vgl. *Ladinia* XVI/1992, XVII/1993 und XIX/1995). Das Programm ist übrigens in *Clipper* und *C++* erstellt und somit auf jedem sogenannten IBM-kompatiblen Rechner lauffähig, vorausgesetzt die Kapazität der Massenspeicher reicht für die Beherbergung von rund 130 MB an

Daten-, Index- und Programmfiles aus. Hat CARD einmal *PostScript-Files* zu den gewünschten Sprachatlaskarten erzeugt, so steht uns, in der Salzburger *AD-I*-Werkstatt, ein leistungsfähiger Tintenstrahldrucker zur Verfügung, auf dem die Formate DIN A4, A3 und A2 ausgegeben werden können. Solcherart können wir Muster der definitiven Atlaskarten herstellen und endredigieren (sechs Beispiele von entsprechend generierten A2-Karten sind übrigens in Beilage zu unserem 8. Arbeitsbericht in *Ladinia* XVII/1993 erschienen). Für den Druck des *AD-I* werden die von CARD generierten *PostScript-Files* direkt auf Lichtsatzmaschinen geschickt und als Filme weiterverarbeitet. Entsprechende Tests mit mehreren Druckereien haben inzwischen eine 100%-ige Kompatibilität zwischen unseren Erzeugnissen und den Anforderungen professioneller Satzstudios unter Beweis gestellt. Zusammenfassend sei noch betont, daß die künftige elektronische Version des *AD-I* auf CD-ROM auch eine Vollversion des Kartierungsprogrammes enthalten wird, so daß es jedem Benutzer möglich sein wird, eigene (d.h. u.U. auch nicht im *AD-I* veröffentlichte) Sprachatlaskarten zu generieren und auszudrucken. Ausschnittkarten sind dabei ebenso vorgesehen wie die Verwendung individueller Layoutoptionen. So werden beispielsweise auch *Thiessen-Polygon-Karten* des *AD-I*-Netzes zum Einsatz kommen. Sie werden eine der kartographischen Grundlagen für weitere computergestützte Auswertungen der *AD-I*-Daten darstellen und einfache Abbildungen räumlicher Merkmalsverteilungen oder Raumstrukturen ebenso aufnehmen können wie komplexe, aus genuin dialektometrischen Analysen resultierende Kartierungen.

Eine weitere informatische Großanwendung beim *AD-I* betrifft die Erstellung eines "*Sprechenden Sprachatlases*" als akustische Ergänzung zu den beiden auf graphischer Basis realisierten Varianten des *AD-I*. Im Vergleich zu seiner "Weltpremiere" (1990) präsentiert sich der "Sprechende" heute als modernes, den weltweit etablierten Multimediastandards entsprechendes Instrument mit mehrsprachiger Menüoberfläche, hochauflösender Graphik und einer individuell abstimmbaren Konfiguration. Bislang hat auch diese Applikation mehrere Softwaregenerationen durchlaufen. Basierte der Erstling von 1990 noch auf *Microsoft-Windows 286*, so stützten sich die Nachfolgeversionen auf *Windows 3.0*, *3.1* und *Windows für Workgroups 3.11*. Derzeit (1996) steht eine Portierung auf *Windows 95* bzw. *NT* an (für weitere Informationen zum Sprechenden vgl. Bauer 1991, 1992, 1993 und 1995, Goebel 1992 sowie die in dieser Zeitschrift veröffentlichten Arbeitsberichte zum *AD-I*).

Abschließend noch ein paar Worte zum aktuellen Stand der Arbeiten bzw. zu den Zukunftsperspektiven. Derzeit (April 1996) werden im *AD-I*-Archiv der Universität Salzburg die letzten 50 von insgesamt rund 1.500 DIN-A2-großen Einzelkarten hergestellt. Parallel dazu erfolgt die Endredaktion der einzelnen Kartenblätter. In EDV-technischer Hinsicht bereiten wir die Publikation einer CD-ROM vor, die das Kartierungs- und Datensuchprogramm CARD, weitere

Module zur elementaren Datentypisierung, einen elektronisch konsultierbaren (provisorischen) Index sowie die aktuelle Version des *Sprechenden* enthalten wird. An akustischen Daten werden wir erstmals mit im Rahmen von Nachenqueten an unseren 21 dolomitenladinischen Meßpunkten aufgenommenem Qualitätsschall aufwarten können. Die zugehörigen DAT-Aufnahmen werden im Laufe der Jahre 1996/97 digitalisiert und portioniert, um sukzessive in unser Multimediasystem eingebunden werden zu können.

Neben den bereits erwähnten Publikationen zum [mittlerweile in 7 Bänden erschienenen] *AD-I* [Dr. Ludwig Reichert Verlag Wiesbaden] kann sich der interessierte Leser auch über eine *World Wide Web-Home Page* mit den jeweils aktuellsten Projektinformationen versorgen. Die entsprechende URL lautet:

[HTTP://www.sbg.ac.at/rom/people/proj/ald/ald\\_home.htm](http://www.sbg.ac.at/rom/people/proj/ald/ald_home.htm)

### Auswahlbibliographie *AD-I*

Es werden lediglich die wichtigsten Titel zu EDV-spezifischen Themen aufgeführt. Für weitere Hinweise vgl. die im 10. *Arbeitsbericht zum ALD I* in *Ladinia* XX (219-221) abgedruckte Projektbibliographie.

Bauer, Roland (1991): *ALD I - CD / CD - ALD I. 98 Tonproben zum ALD I / 98 campioni fonici per l'ALD I*. Salzburg; Compact-Disk mit zweispr. (dt.-ital.) Booklet, 44 Seiten + 3 Karten.

Bauer, Roland (1992): L'informatizzazione dell'atlante linguistico sonoro ALD I. In: *Linguistica* XXXII, (= *Paolo Tekavčić sexagenario in honorem oblata II*), 197-212.

Bauer, Roland (1993): Neue Perspektiven der LDV im Forschungsprojekt ALD I - Der sprechende Sprachatlas. In: Viereck, Wolfgang (Hg.): *Verhandlungen des Internationalen Dialektologenkongresses Bamberg 1990. Band 1: Plenarvorträge, Computative Datenverarbeitung und Dialektklassifikation*. Stuttgart, (= ZDL-Beiheft 74), 124-146.

Bauer, Roland (1995): Dialettologia computazionale ed atlanti linguistici: la dimensione sonora. In: Romanello, Maria Teresa / Tempesta, Immacolata (Hgg.): *Dialetti e lingue nazionali. Atti del XXVII Congresso della SLI* [Lecce 1993]. Roma, (= Pubblicazioni SLI 35), 155-167.

Bauer, Roland (1996): Le système de gestion de base de données de l'Atlas linguistique du ladin central (ALD I). In: Moracchini, Georges (Hg.): *Bases de données linguistiques: conceptions, réalisations, exploitations. Actes du Colloque International* [Corte/Corti 1995], Corte, 195-211.

Goebel, Hans (1992): L'atlas parlant dans le cadre de l'Atlas linguistique du ladin central et des dialectes limitrophes. In: *Nazioarteko dialektologia biltzarra. Agiriak. / Actas del Congreso internacional de dialectología* [Bilbo/Bilbao 1991]. Bilbo/Bilbao, (= iker-7), 397-412.

Haimperl, Edgar (1996): A Database Application for the Generation of Phonetic Atlas Maps. In: Nerbonne, John (Hg.): *Linguistic Databases* [Groningen 1995]. Stanford, 103-116.