

BayDat – die bayerische Dialektdatenbank: Abschlussbericht

Ralf ZIMMERMANN

1. Das Projekt BayDat und dessen Ziele

Die bayerische Dialektdatenbank BayDat ist ein von der Deutschen Forschungsgemeinschaft und dem bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst finanziertes Projekt. Ziel des Projekts ist es, die elektronisch erfassten Erhebungsdaten aus den sechs Teilprojekten des Bayerischen Sprachatlas (BSA) in einer zentralen Datenbank zusammenzuführen und so einerseits die zukunftssichere Speicherung der Daten zu gewährleisten, andererseits einen unkomplizierten und schnellen Zugriff auf die Daten zu ermöglichen.

Das Ziel, die Daten zukunftssicher zu speichern, konnte durch die verwendete Datenbanksoftware Oracle 9i realisiert werden: Die Oracle-Datenbanksoftware erschien erstmals in den späten 70er Jahren und wird von zahlreichen internationalen Unternehmen und Institutionen eingesetzt. Es ist daher davon auszugehen, dass die Software auch in kommenden Jahren und Jahrzehnten weiter unterstützt wird. Außerdem gewährleistet Oracle durch integrierte Sicherheitssysteme eine sichere Speicherung der BSA-Daten.

Das zweite Ziel – der schnelle und möglichst unkomplizierte Zugriff auf die BSA-Daten – konnte einerseits durch die Realisierung einer Online-Schnittstelle, andererseits durch die weitere Aufbereitung der Daten, insbesondere der sprachwissenschaftliche Auszeichnung der Projektfragebücher, verwirklicht werden.

2. Die sprachwissenschaftliche Aufbereitung der Projektfragebücher

Die Projektfragebücher der BSA-Teilprojekte sind zwar im Umfang ähnlich, unterscheiden sich jedoch an einzelnen Stellen voneinander: Das Projektfragebuch des SPRACHATLAS VON UNTERFRANKEN enthält z.B. zusätzliche Fragen aus dem Bereich Weinbau, das Fragebuch des SPRACHATLAS VON NIEDERBAYERN ist um Fragen zur Holzwirtschaft ergänzt. Auch im formalen Aufbau, z.B. in der Nummerierung der einzelnen Fragen, können die Projektfragebücher voneinander abweichen. Um elektronisch auf sämtliche Einzelfragen zugreifen zu können, war es daher zunächst notwendig, die Projektfragebücher in einem Maximalfragebuch tabellarisch zusammenzufassen: Eine Spalte enthält die laufende Nummer der Frage im Maximalfragebuch, es folgen der Fragetext und die Fragennummern aus den einzelnen Teilprojekten.

Für den BSA wurde z.B. das Flexionsparadigma des Verbs *sein* erhoben (vgl. Abbildung 1). Wollte man nun im gesamten Maximalfragebuch nach Fragen suchen, die

Flexionsformen des Verbs *sein* erheben, müsste man diese einzeln suchen: Zunächst müsste nach *bin*, anschließend nach *bist*, *ist*, *seid* usw. gesucht werden. Sollen zusätzlich Fragen gefunden werden, die das Präteritum von *sein* erheben, müsste weiterhin nach *war*, *warst*, *wart* usw. gesucht werden. Ist man außerdem nur an Flexionsformen des Vollverbs *sein*, nicht jedoch an Flexionsformen des entsprechenden Hilfsverbs interessiert, müssten die Suchergebnisse anschließend vom Nutzer auf das gewünschte Phänomen hin durchsucht werden.

Hier setzt die sprachwissenschaftliche Auszeichnung des Maximalfragebuchs an: In separaten Datenbanktabellen werden Zusatzinformationen zu den Erhebungsfragen notiert. Wird in einer Frage z.B. ein Verb erhoben, so werden in der Tabelle *Verben* folgende Informationen festgehalten:

- die Grundform des erhobenen Verbs
- die Form des Verbs in der jeweiligen Frage
- die Person des Verbs
- der Numerus des erhobenen Verbs
- der Modus des jeweiligen Verbs
- das Tempus des Verbs
- die Art des Verbs
- die Klasse des erhobenen Verbs

Werden infinite Verbformen erhoben, so wird deren Art – Infinitiv oder Partizip II – sowie deren Position – freistehend, nach Hilfsverb usw. – notiert, für die Klasse der starken Verben wird außerdem deren Ablautreihe angegeben.

Durch die Auszeichnungen werden nun gezielte Suchanfragen möglich, so lassen sich durch eine entsprechende Anfrage z.B. sämtliche Fragen ermitteln, die das Vollverb *sein* im Konjunktiv II erheben. Das Vollverb *sein* wird innerhalb des BSA dreimal im Konjunktiv II erhoben, nämlich in den Fragen 535 - "*Wenn das Roß/der Gaul jünger wäre, könnte es/er besser ziehen*" -, 2067 - "*Wenn ich gesund wäre, würde ich tanzen gehen*" – und 2578 - "*Wäre ich froh, wenn ich das wüsste.*".

Die sprachwissenschaftlichen Auszeichnungen der Fragebücher beschränken sich natürlich nicht auf die Auszeichnung von Verben: Im Bereich der Morphologie sind relevante Informationen zu sämtlichen Wortarten ausgezeichnet, für Substantive z.B. Genus, Numerus und Kasus, für Adjektive Grad, Position usw. Außerdem wurden Auszeichnungen im Bereich Lautung und Wortbildung durchgeführt.

Ausgezeichnet wurden dabei nicht nur in den Fragetexten auftauchende linguistische Phänomene. Die Suggestierformen zu den einzelnen Fragen werden nach und nach in die Datenbank aufgenommen und ebenfalls in den Bereichen Morphologie, Lautung und Wortbildung ausgezeichnet.

Um auf die in den Datenbanktabellen enthaltenen Informationen zugreifen zu können, müssen Abfragen in der Datenbanksprache SQL erstellt werden. Da dem Laienbenutzer die Erstellung von SQL-Abfragen kaum zuzumuten ist, wurde im Rahmen des Projektes BAYDAT eine Onlineoberfläche erstellt, die nach der endgültigen Fertigstellung der Datenbank für jedermann frei zugänglich sein wird. Innerhalb der Onlineoberfläche lässt sich über die Menüpunkte *Fragen* und *Spezialsuche* auf zweierlei Weise nach Fragen suchen, die bestimmte sprachwissenschaftliche Phänomene erheben:

3. Die Suchmöglichkeiten innerhalb der Onlineoberfläche der BayDat-Datenbank

Der Menüpunkt *Fragen* erlaubt zunächst eine Volltextsuche innerhalb der Fragetexte und Suggestierformen. Dabei muss nicht der gesamte Fragetext bzw. die gesamte Suggestierformen in das Suchfeld eingegeben werden, die Eingabe von *wäre* gibt z.B. die Frage 535 – "Wenn das Roß/der Gaul jünger wäre, könnte es/er besser ziehen" als Ergebnis zurück. Die Leiste *Sachbereiche* stellt eine weitere Möglichkeit dar, innerhalb des Maximalfragebuchs zu suchen: Die Fragen sind innerhalb der Projektfragebücher nach Sachgruppen geordnet, diese Ordnung wurde für die BAYDAT-Datenbank übernommen. Durch Linksklick werden sämtliche Fragen aus dem gewählten Sachbereich als Suchergebnis angezeigt, z.B. alle Fragen aus dem Sachbereich *Das Vieh und seine Pflege*.

Möchte man über den Menüpunkt *Fragen* auf die linguistischen Auszeichnungen zugreifen, muss zunächst über das Suchfeld oder über die Leiste *Sachbereiche* nach Fragen gesucht werden. Um z.B. auf die Auszeichnungen des Verbs *sein* zuzugreifen, muss zunächst nach einer beliebigen Form des Verbs, z.B. dem Infinitiv oder der Flexionsform *wäre*, gesucht werden. Anschließend lassen sich durch einen Linksklick auf den Fragetext Zusatzinformationen zur jeweiligen Frage anzeigen.

Ein Linksklick auf den Fragetext der Frage 2067 – "Wenn ich gesund wäre, würde ich tanzen gehen." – führt z.B. zum in Abbildung 2 gezeigten Ergebnis: Unter dem Punkt *Details* finden sich die Fragennummern aus den jeweiligen Projektfragebüchern. Da zu jedem Projekt eine Fragennummer angegeben ist, kann man erkennen, dass die Frage in jedem Teilprojekt des Bayerischen Sprachatlas erhoben wurde. Diese Frage würde sich also dazu eignen, Quervergleiche zwischen den einzelnen Projekten anzustellen.

Der Unterpunkt *Verknüpfungen* erlaubt den Zugriff auf die sprachwissenschaftlichen Auszeichnungen: Zum Verb *gehen* sind 34 Auszeichnungen – hier Verknüpfungen genannt - vorhanden, d.h. das Verb wurde innerhalb des Maximalfragebuchs 34-mal erhoben und entsprechend ausgezeichnet. Man kann außerdem feststellen, dass nicht nur Verben, sondern auch andere Wortarten sowie Wortbildungen, Vokale und Konsonanten ausgezeichnet wurden, z.B. wurde das Verb *gehen* im Maximalfragebuch 14-mal im Rahmen von Wortbildungen erhoben.

Durch einen Linksklick auf *sein* lassen sich die entsprechenden Auszeichnungen anzeigen: Zunächst werden die Auszeichnungen für das Vollverb *sein* angezeigt, infinite Formen werden vor finiten Formen angezeigt. Weiter unten finden sich die Auszeichnungen für das Hilfsverb *sein*. Die Form *ist* in Frage 2058 – "Er ist nie still gewesen." – ist als 3. Person Singular Indikativ Präsens ausgezeichnet. Es werden nur die Tempora Präsens und Präteritum ausgezeichnet, d.h. die synthetischen Tempora. Um Fragen zu finden, die analytische Tempora, z.B. Perfekt oder Futur I, erheben, muss nach Fragen gesucht werden, die sowohl die finite Form des entsprechenden Hilfsverbs und eine infinite Form eines Vollverbs erheben. Diese komplexen Abfragen lassen sich über den Menüpunkt *Spezielsuche* erstellen.

Es soll nach Fragen gesucht werden, die die 3. Person Singular Perfekt mit dem Hilfsverb *sein* erheben. Hierzu muss aus der Kategorienleiste auf der linken Seite zunächst der Unterpunkt *Verben* angewählt werden. Es erscheinen nun verschiedene Auswahlmöglichkeiten, die den linguistischen Auszeichnungen der Verben entsprechen (vgl. Abbildung 3). Fragen, die die 3. Person Singular Perfekt mit dem Hilfsverb *sein* erheben, müssen das Hilfsverb *sein* in der 3. Person Singular Präsens Indikativ enthalten, die entsprechenden Werte können über die Suchmaske ausgewählt werden. Anschließend

können weitere Werte hinzugefügt werden, in diesem Fall soll zusätzlich zum Hilfsverb *sein* in der 3. Person Singular Indikativ Präsens nach Vollverben im Partizip II gesucht werden. Dazu wird erneut der Punkt *Verben* aus der Kategorienleiste ausgewählt, die Werte Vollverb und Partizip II sind auf der Folie bereits ausgewählt und können in die Abfrage übernommen werden. Die veränderte Abfrage wird angezeigt und kann nun abgeschickt werden. Die Abfrage ergibt das in Abbildung 4 gezeigte Ergebnis: Die 3. Person Perfekt mit dem Hilfsverb *sein* wird z.B. in Frage 2058 – "Er ist nie still gewesen" – oder in Frage 2069 – "Das Wetter ist schön geworden." – erhoben. Zu diesen Fragen sollen Belege angezeigt werden, die Auswahlkästchen vor den Fragen müssen hierzu aktiviert werden.

4. Anzeige von Belegen in den Formaten Html und Pdf

Um Belege aus der BAYDAT-Datenbank anzuzeigen, müssen zusätzlich zu den Fragen Erhebungsorte ausgewählt werden. Dies geschieht über den Menüpunkt *Orte*. Erhebungsorte lassen sich auf zweierlei Weise auswählen: über das Suchfeld oder über die darunter liegenden Leisten *Regierungsbezirke* und *Landkreise*.

Das Suchfeld erlaubt die Suche nach Ortsnamen und Postleitzahlen. Hier müssen, ähnlich wie bei der Suche nach Fragen, nicht komplette Ortsnamen bzw. Postleitzahlen eingegeben werden, es reicht, den bekannten Teil eines Ortsnamens oder einer Postleitzahl einzugeben. Um den unterfränkischen Erhebungsort *Schweinfurt* zu finden, reicht z.B. die Eingabe von *Schwein* oder *Furt* in das Suchfeld. Die Ergebnisse werden dabei alphabetisch nach Ortsnamen sortiert, es ist jedoch durch Linksklick auf die entsprechenden Spaltenköpfe auch eine Sortierung nach Postleitzahlen und Landkreisen sowie das Umschalten auf eine absteigende Sortierreihenfolge möglich. Die Änderung der Sortierreihenfolge kann sinnvoll sein, wenn die Erhebungsorte nicht über das Suchfeld, sondern über die Leisten *Regierungsbezirke* bzw. *Landkreise* gesucht werden, d.h. wenn Erhebungsorte aus ganzen Landkreisen oder Regierungsbezirken angezeigt werden. Für die Beispielbeleglisten sollen die Belege zu den zuvor ausgewählten Fragen 2058 – "Er ist nie still gewesen" – und 2069 – "Das Wetter ist schön geworden." – aus den Erhebungsorten Würzburg, Ochsenfurt und Bütthard angezeigt werden. Die entsprechenden Auswahlkästchen müssen aktiviert werden, durch Linksklick auf *Auswählen* werden die Erhebungsorte für die Beleganzeige übernommen.

Es gibt zwei Möglichkeiten, Belege bzw. Beleglisten auszugeben: Über den Menüpunkt *Ergebnisse* werden Belege im HTML-Format angezeigt, über den Menüpunkt *Belegliste* können Beleglisten im PDF-Format erstellt werden.

Bei der Ausgabe von Belegen im HTML-Format wird zunächst die Frage angegeben, zu der die jeweiligen Belege gehören, darunter findet man die Fragennummer aus den Einzelprojekten des BSA. Die Belege werden, nach Erhebungsorten sortiert, im Anschluss tabellarisch angezeigt. Neben der Anzeige der Belege in Teuthonista-Lautschrift werden außerdem – sofern vorhanden – die Lemmata zu den einzelnen Belegen ausgegeben. Sollte der Lautschriftbeleg zu einer gewählten Frage an einem Erhebungsort fehlen, wird stattdessen der Grund für den Nullbeleg ausgegeben. Im Erhebungsort Bütthard wurde zu Frage 2058 – "Er ist nie still gewesen." – z.B. "Wort unbekannt" als Begründung für den Nullbeleg notiert. Grüne Häkchen in der rechten Spalte der Tabelle zeigen an, dass zum jeweiligen Beleg ein Scan des Originalfragebuchs vorliegt. Sämtliche Fragebücher des SPRACHATLAS VON UNTERFRANKEN wurden eingescannt, die Scans wurden im GIF-Format in die Datenbank übernommen und

können über die Onlineoberfläche eingesehen werden. Durch die Aufnahme der Scans in die BAYDAT-Datenbank können diese nicht nur zukunftsicher gespeichert werden, sondern auch als unmittelbare Quelle zu Forschungsarbeiten herangezogen werden. Sollen neben Lautschrift und Lemma weitere Kommentare angezeigt werden, so können einzelne oder alle Belege für eine Vollanzeige markiert werden, es können so Beleglisten ganzer Landkreise oder gar Regierungsbezirke im HTML-Format erstellt werden. Für jeden Beleg werden dabei Erhebungsort, Frage und die einzelnen Kommentarteile in tabellarischer Form aufgeführt. Unter Umständen können die Beleglisten im HTML-Format recht lang werden, außerdem bietet sich das Format nicht für die Archivierung und den Ausdruck von Beleglisten an. Deshalb können über den Menüpunkt *Belegliste* Beleglisten im PDF-Format erstellt werden.

Über den Menüpunkt *Belegliste* kann zunächst ausgewählt werden, welche Kommentarteile in die Belegliste aufgenommen werden sollen, in jedem Fall werden die Belege in Teuthonista-Lautschrift, Lemmata sowie die Begründung für etwaige Nullbelege aufgenommen. Weitere Kommentarteile, wie z.B. Bedeutungsangaben, Explorátoremkommentare oder Gewährspersonenkommentare lassen sich nach Bedarf zuschalten oder ausblenden. Zusätzlich können der Seitenumbruch reguliert und ein Ortsverzeichnis sowie ein Transkriptionsschlüssel mit der Belegliste ausgegeben werden.

Jede Belegliste wird zunächst mit dem Zugriffsdatum versehen, im Anschluss findet sich auf jeder Belegliste ein Inhaltsverzeichnis, das die in der Belegliste enthaltenen Fragen sowie, falls vorhanden, Ortsverzeichnis und Transkriptionsschlüssel, auflistet. Zur schnelleren Navigation sind die Punkte des Inhaltsverzeichnisses mit den jeweiligen Seiten verknüpft. So kann bequem vom Inhaltsverzeichnis zur gewünschten Frage gewechselt werden. In der Folge finden sich für jede Frage zunächst deren Nummer aus dem Maximalfragebuch, Fragetext und die Fragennummern aus den einzelnen Projektfragebüchern. Darauf folgen, ähnlich wie in den HTML-Beleglisten, die einzelnen Belege nach Erhebungsort sortiert, jedoch erlaubt das PDF-Format eine genauere Seitenformatierung, die sich besser für die Druckausgabe eignet. Außerdem wird der zur Darstellung der Lautschriftbelege verwendete Font direkt in die PDF-Dokumente eingebunden, so dass sich die Beleglisten ohne vorherige Installation von Schriftarten auf jedem beliebigen Rechner mit vorhandenem PDF-Reader betrachten und ausdrucken lassen.

5. Schluss

Durch die Realisierung der BAYDAT-Datenbank ergeben sich mehrere Vorteile für weitere Arbeiten auf Grundlage der Erhebungsdaten des BSA:

Durch die Auszeichnung des Maximalfragebuchs wurde der Zugriff auf sprachwissenschaftliche Phänomene deutlich erleichtert. Während zur Erstellung der Atlasbände die Fragebücher der Einzelprojekte im Hinblick auf bestimmte Phänomene – Morphologie, Lautung usw. – immer wieder durchgesehen und einzelne Fragen extrahiert werden mussten, so geschieht dies nun in Sekundenschnelle über die BAYDAT-Datenbank. Sie wird damit zu einem interessanten Werkzeug für weitere Forschungen auf Basis der Erhebungen für den Bayerischen Sprachatlas.

Durch die Realisierung der Onlineschnittstelle und deren Oberfläche steht BAYDAT außerdem weltweit einem breiten Benutzerkreis zur Verfügung. Dem interessierten Laien, der wissen möchte, wie in seiner Heimatregion der zweite Heuschnitt genannt wird oder wie man den Stammvokal im Wort *Fass* ausspricht,

ebenso, wie dem Dialektologen, der die Erhebungsdaten des Bayerischen Sprachatlas für weitergehende Forschungen nutzen möchte.

Weitere Illustrationen zu diesem Aufsatz finden Sie unter www.unterfraenkisches-dialektinstitut.de.