



WespA

Würzburger elektronische sprachwissenschaftliche Arbeiten

Band 1

Ralf Zimmermann

BAYDAT – Die bayerische Dialektdatenbank

WespA

Würzburger elektronische sprachwissenschaftliche Arbeiten

Die „Würzburger elektronischen sprachwissenschaftlichen Arbeiten“ sind ein Publikationsforum für Arbeiten, die am oder in Zusammenarbeit mit dem Lehrstuhl für deutsche Sprachwissenschaft der Universität Würzburg entstanden sind. Auf diese Weise werden Forschungsergebnisse schnell veröffentlicht, um die sprachwissenschaftliche Diskussion zu intensivieren. Die Herausgeber sind für jede Reaktion dankbar.

Herausgeber:

Johannes Schwitalla, Peter Stahl, Werner Wegstein, Norbert Richard Wolf

<http://www.spr.germanistik.uni-wuerzburg.de/wespa>

WespA. Würzburger elektronische sprachwissenschaftliche Arbeiten

Nr. 1 (Mai 2007)

Ralf Zimmermann:

BAYDAT – Die bayerische Dialektdatenbank

ISSN:

ISBN: 978-3-923959-34-1

© Lehrstuhl für deutsche Sprachwissenschaft

Julius-Maximilians-Universität Würzburg

Institut für deutsche Philologie

Am Hubland

97074 Würzburg

Tel.: +49 (0) 931 - 888 - 56 30

Fax: +49 (0) 931 - 888 - 46 16

<http://www.spr.germanistik.uni-wuerzburg.de>

Alle Rechte vorbehalten.

Würzburg 2007.

Dieses Dokument wird bereitgestellt durch den Online-Publikationsserver der Universität Würzburg.

Universitätsbibliothek Würzburg

Am Hubland

97074 Würzburg

Tel.: +49 (0) 931 - 888 - 59 17

Fax: +49 (0) 931 - 888 - 59 70

opus@bibliothek.uni-wuerzburg.de

<http://opus.bibliothek.uni-wuerzburg.de/>

Deckblattgestaltung: Dagmar Rußner-Blank

BAYDAT – Die bayerische Dialektdatenbank

Inaugural-Dissertation
zur Erlangung der Doktorwürde der
Philosophischen Fakultät II
der
Julius-Maximilians-Universität Würzburg

vorgelegt von

Ralf Zimmermann
aus Bad Mergentheim

Würzburg 2006

Inhaltsverzeichnis

1. Einleitung	5
2. Datenerhebung und -kodierung im Projekt BAYERISCHER SPRACHATLAS	7
2.1. Die BSA-Fragekonkordanz	8
2.2. Die BSA-Ortsliste	10
2.3. Die ASCII-Ortsdateien	11
3. ORACLE – Ein relationales Datenbankmanagementsystem	17
4. Aufbereitung der Quelldateien für die bayerische Dialektdaten- bank BAYDAT	22
4.1. Umsetzung der BSA-Fragekonkordanz innerhalb der BAYDAT-Datenbank	22
4.1.1 Beschreibung der Datenbanktabellen	22
4.1.2 Beziehungen zwischen den Tabellen zur BSA-Fragekonkordanz	26
4.1.3 Beispielabfragen zur BSA-Fragekonkordanz	26
4.2. Umsetzung der BSA-Ortsliste innerhalb der BAYDAT-Datenbank	29
4.2.1. Beschreibung der Datenbanktabellen	29
4.2.2. Beziehungen zwischen den Tabellen zur BSA-Ortsliste	30
4.2.3. Beispielabfragen zur BSA-Ortsliste	31
4.3. Umsetzung der ASCII-Ortsdateien innerhalb der BAYDAT-Datenbank	33
4.3.1. Beschreibung der Datenbanktabellen	33
4.3.2. Behandlung von Unterbelegen und Spontanbelegen in der BAYDAT- Datenbank	39
4.3.3. Beziehungen zwischen den Tabellen zu den ASCII-Ortsdateien	40
4.3.4. Beziehungen der Tabelle <i>Belege</i> zu anderen Tabellen der BAYDAT-Datenbank	41
4.3.5. Beispielabfragen zu den ASCII-Ortsdateien	41
5. Auszeichnung der Fragen unter sprachwissenschaftlichen Gesichtspunkten	46
5.1. Auszeichnung lautlicher Phänomene	47
5.1.1. Vokale	47
5.1.1.1. Generelles zur Auszeichnung der Vokale	47
5.1.1.2. Datenbanktabellen zur Auszeichnung der Vokale	49
5.1.1.3. Beziehungen der Tabelle <i>Link_Vokale</i> zu anderen Tabellen der BAYDAT- Datenbank	50
5.1.1.4. Beispielabfragen	51

5.1.2. Konsonanten	53
5.1.2.1. Generelles zur Auszeichnung der Konsonanten	53
5.1.2.2. Datenbanktabellen zur Auszeichnung der Konsonanten	54
5.1.2.3. Beziehungen der Tabelle <i>Link_Konsonanten</i> zu anderen Tabellen der BAYDAT-Datenbank	55
5.1.2.4. Beispielabfragen	56
5.2. Auszeichnung morphologischer Phänomene	58
5.2.1. Verben	59
5.2.1.1. Generelles zur Auszeichnung der Verben	59
5.2.1.2. Datenbanktabellen zur Auszeichnung der Verben	61
5.2.1.3. Beziehungen der Tabelle <i>Link_Verben</i> zu anderen Tabellen der BAYDAT- Datenbank	62
5.2.1.4. Beispielabfragen	63
5.2.2. Substantive	66
5.2.2.1. Generelles zur Auszeichnung der Substantive	66
5.2.2.2. Datenbanktabellen zur Auszeichnung der Substantive	67
5.2.2.3. Beispielabfragen	68
5.2.3. Adjektive	72
5.2.3.1. Generelles zur Auszeichnung der Adjektive	72
5.2.3.2. Datenbanktabellen zur Auszeichnung der Adjektive	73
5.2.3.3. Beispielabfragen	74
5.2.4. Pronomen	78
5.2.4.1. Generelles zur Auszeichnung der Pronomen	78
5.2.4.2. Datenbanktabellen zur Auszeichnung der Pronomen	79
5.2.4.3. Beispielabfragen	80
5.2.5. Artikelwörter	83
5.2.5.1. Generelles zur Auszeichnung der Artikelwörter	83
5.2.5.2. Datenbanktabellen zur Auszeichnung der Artikelwörter	84
5.2.5.3. Beispielabfragen	85
5.2.6. Junktionen	89
5.2.6.1. Generelles zur Auszeichnung von Junktionen	89
5.2.6.2. Datenbanktabellen zur Auszeichnung der Junktionen	89
5.2.6.3. Beispielabfragen	90
5.2.7. Präpositionen	91
5.2.7.1. Generelles zur Auszeichnung der Präpositionen	91
5.2.7.2. Datenbanktabellen zur Auszeichnung der Präpositionen	91

5.2.7.3. Beispielabfragen	92
5.2.8. Adverbien	94
5.2.8.1. Generelles zur Auszeichnung der Adverbien	94
5.2.8.2. Datenbanktabellen zur Auszeichnung der Adverbien	95
5.2.8.3. Beispielabfragen	96
5.2.9. Partikeln	97
5.2.9.1. Generelles zur Auszeichnung der Partikeln	97
5.2.9.2. Datenbanktabellen zur Auszeichnung der Partikeln	97
5.2.9.3. Beispielabfragen	98
5.3. Auszeichnung von Wortbildungen	99
5.3.1. Generelles zur Auszeichnung der Wortbildungen	99
5.3.2. Datenbanktabellen zur Auszeichnung der Wortbildungen	103
5.3.3. Beziehungen der Tabelle <i>Link_Wortbildungen</i> zu anderen Tabellen der BAYDAT-Datenbank	104
5.3.4. Beispielabfragen	105
6. Zusätzliche Elemente	109
6.1. Scans der Originalfragebücher	109
6.1.1. Beschreibung der Datenbanktabellen	109
6.1.2. Beziehungen der Tabellen zu den Scans der Originalfragebücher	110
6.1.3. Beziehungen der Tabellen zu den Scans der Originalfragebücher zu anderen Tabellen der BAYDAT-Datenbank	110
6.1.4. Beispielabfrage zu den Scans der Originalfragebücher	111
6.2. Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKT- INSTITUTS	112
6.2.1. Beschreibung der Datenbanktabellen	113
6.2.2. Beziehungen der Tabellen zu den Fotografien aus dem Bildarchiv des UNTER- FRÄNKISCHEN DIALEKTINSTITUTS	114
6.2.3. Beziehungen der Tabellen zu den Fotografien aus dem Bildarchiv des UNTER- FRÄNKISCHEN DIALEKTINSTITUTS zu anderen Tabellen der BAYDAT-Daten- bank	114
6.2.4. Beispielabfrage zu den Fotografien aus dem Bildarchiv des UNTER- FRÄNKISCHEN DIALEKTINSTITUTS	115
7. Zusammenfassendes zum Aufbau der Datenbank	116
8. Die Online-Anbindung der BAYDAT-Datenbank	118
8.1. COCOON – Eine Middleware auf XML-Basis	118
8.2. Umsetzung der kodierten Belege in TEUTHOBD/TEUTHOUC	121
8.3. Die grafische Benutzeroberfläche	129

8.3.1. Auswahl von Orten und Fragen	130
8.3.1.1. Der Menüpunkt <i>Orte</i>	130
8.3.1.2. Der Menüpunkt <i>Fragen</i>	133
8.3.1.3. Der Menüpunkt <i>Spezialsuche</i>	137
8.3.2. Die Anzeige der Suchergebnisse	142
8.3.2.1. Der Menüpunkt <i>Ergebnisse</i>	142
8.3.2.2. Der Menüpunkt <i>Belegliste</i>	145
9. Abschluss und Ausblick	147
Literaturverzeichnis	149
Tabellenverzeichnis	153
Abbildungsverzeichnis	157

Anhang A: Ausdruck einer gescannten Fragebuchseite

Anhang B: Belegliste (entsprechend Abfrage Kp. 7)

1. Einleitung

Was 1984 mit den Erhebungen für den SPRACHATLAS VON BAYERISCH-SCHWABEN an der Universität Augsburg begann, steht heute, 2006, kurz vor seiner Vollendung: Die Sprachatlasbände der sechs Teilprojekte des BAYERISCHEN SPRACHATLAS – SPRACHATLAS VON BAYERISCH-SCHWABEN, NORDOSTBAYERN, NIEDERBAYERN, OBERBAYERN, MITTELFRAKEN und UNTERFRANKEN – sind veröffentlicht oder stehen vor der Veröffentlichung. In über 1500 bayerischen Gemeinden erhoben Dialektologen fünf bayerischer Universitäten den Dialekt der älteren Generation und erstellten auf Grundlage der Erhebungen Sprachkarten zur Wort-, Formen- und Lautgeographie, die in o.g. Sprachatlasbänden publiziert wurden bzw. werden und so die Dialekte Bayerns flächendeckend dokumentieren.

Trotz der sehr umfangreichen Publikationen – das umfangreichste Teilprojekt des BAYERISCHEN SPRACHATLAS, der SPRACHATLAS VON BAYERISCH-SCHWABEN, veröffentlicht die Ergebnisse seiner Arbeit in insgesamt 13 Atlasbänden – konnte im Rahmen der Sprachatlanten nur ein Teil der Erhebungsdaten ausgewertet werden. Um die Daten für weitere Recherchen und Forschungsarbeiten verfügbar zu machen und sie zukunftssicher zu speichern, beabsichtigten die Projektleiter des BAYERISCHEN SPRACHATLAS und dessen Geldgeber – die DEUTSCHE FORSCHUNGSGEMEINSCHAFT und das BAYERISCHE STAATSMINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST – von Anfang an, diese nach Abschluss der Arbeiten an den Sprachatlanten der Allgemeinheit zugänglich zu machen.

Aus dieser Absicht entstand das von der DEUTSCHEN FORSCHUNGSGEMEINSCHAFT und dem BAYERISCHEN STAATSMINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST geförderte Projekt BAYDAT, das im Juli 2000 am Lehrstuhl für deutsche Sprachwissenschaft der BAYERISCHEN JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG die Arbeit aufnahm. Ziel von BAYDAT ist es, die Erhebungsdaten der sechs Teilprojekte des BAYERISCHEN SPRACHATLAS, die bisher in unterschiedlicher Form dezentral an den fünf beteiligten Universitäten Augsburg, Bayreuth, Passau, Erlangen und Würzburg gespeichert waren, in einer zentralen Datenbank zusammenzuführen, um so einerseits eine zukunftssichere Speicherung der Daten zu garantieren. Andererseits sollen die Daten im Rahmen des Projektes unter sprachwissenschaftlichen Gesichtspunkten weiter aufbereitet werden, um sie für zukünftige Forschungen effektiver nutzbar zu machen. Ein weiteres Ziel des Projektes BAYDAT war und ist es, die Erhebungsdaten des BAYERISCHEN SPRACHATLAS über eine Onlineanbindung einem breiten Nutzerkreis zugänglich zu machen, von Sprachwissenschaftlern über Kultur- und

Heimatpfleger, Mundartdichter und -musiker, Lehrer und Schüler, die sich im Unterricht mit dem Thema Dialekt befassen, bis hin zu dialektinteressierten Laien.

Mittlerweile sind die Arbeiten am Projekt BAYDAT abgeschlossen. Die vorliegende Arbeit dokumentiert das Projekt BAYDAT, angefangen von der Datenerhebung und -kodierung in den sechs Teilprojekten des BAYERISCHEN SPRACHATLAS über ihre Umsetzung innerhalb der BAYDAT-Datenbank, die sprachwissenschaftliche Aufbereitung der Daten sowie die Realisierung und Funktionsweise der Onlineschnittstelle der BAYDAT-Datenbank und gibt eine kurze Einführung in das verwendete Datenbankmanagementsystem ORACLE 9i. Über die Dokumentation der Arbeiten hinaus bietet die vorliegende Arbeit durch zahlreiche Beispiele einen unmittelbaren Einblick in Funktionsweise und Recherchemöglichkeiten der BAYDAT-Datenbank und liefert damit Anreiz für eine weitere Nutzung der Erhebungsdaten des BAYERISCHEN SPRACHATLAS im Rahmen zukünftiger dialektologischer Forschungen.

2. Datenerhebung und -kodierung im Projekt BAYERISCHER SPRACHATLAS

Dieses Kapitel beschreibt die elektronische Erfassung der Daten, die im Rahmen des Projektes BAYERISCHER SPRACHATLAS (=BSA) durchgeführt wurde und die der BAYDAT-Datenbank als Quelle und Datengrundlage dient. Im Einzelnen sind dies die Zusammenführung der Projektfragebücher zur BSA-Fragekonkordanz, die Zusammenführung der Informationen aller BSA-Erhebungsorte in der BSA-Ortsliste und die Erfassung der erhobenen Dialektdaten in den ASCII-Ortsdateien. Dabei soll insbesondere auf die Strukturierung der erfassten Daten eingegangen werden, da ebendiese Strukturierung den Ausgangspunkt für die automatische Weiterverarbeitung der Daten zur Erstellung der BAYDAT-Datenbank darstellt.

Im Jahr 1984 begannen Sprachwissenschaftler der Universität Augsburg, Dialektdaten für ein groß angelegtes Sprachatlasprojekt, den SPRACHATLAS VON BAYERISCH-SCHWABEN (=SBS) zu erheben. Dies geschah in einer direkten Erhebung, d.h. Exploratoren erhoben die Daten anhand eines Fragebuches bei Gewährspersonen – am Erhebungsort ansässige Dialektsprecher, die zum Zeitpunkt der Erhebung 60 Jahre oder älter waren¹ – vor Ort. Für die Auswahl der Orte wurde ein Gitternetz auf die Karte des Regierungsbezirks Bayerisch-Schwaben projiziert, aus jedem so entstandenen Planquadrat mit einer Seitenlänge von sieben Kilometern wurde ein Erhebungsort ausgewählt².

In den folgenden Jahren entstanden an vier weiteren bayerischen Universitäten Partnerprojekte des SBS, die Dialektdaten zu den zugehörigen bayerischen Regierungsbezirken nach den Methoden des SBS erhoben: der SPRACHATLAS VON NORDOST-BAYERN (=SNOB, Universität Bayreuth), der SPRACHATLAS VON MITTELFRANKEN (=SMF, Universität Erlangen), der SPRACHATLAS VON UNTERFRANKEN (=SUF, Universität Würzburg), der SPRACHATLAS VON NIEDERBAYERN und der SPRACHATLAS VON OBERBAYERN (=SNIB bzw. SOB, beide Universität Passau). Zusammengefasst wurden die Projekte im BSA.

Um die erhobenen Dialektdaten für die Erstellung der verschiedenen Sprachatlanten sinnvoll weiterverarbeiten zu können, entschied man sich früh, die Materialien – ausgehend von den Fragebüchern, über die tatsächlichen Erhebungen, bis hin zu den Ortslisten der Einzelprojekte – elektronisch zu erfassen. Ziel dieses Kapitels ist es nun, die elektronische Erfassung, wie sie im Rahmen des BSA durchgeführt wurde, näher zu beschreiben. Die elektronische Erfassung der Daten aus den Einzelprojekten des BSA bildet die Grundlage und das Quellenmaterial der bayerischen Dialektdatenbank BAYDAT.

¹ zur Auswahl der Gewährspersonen vgl. König, 1997, S.20 ff.

² zur Auswahl der Erhebungsorte vgl. König, 1997, S.17 f.

2.1. Die BSA-Fragekonkordanz

Sämtliche Fragebücher des BSA lassen sich auf das Fragebuch des SPRACHATLAS DER DEUTSCHEN SCHWEIZ (=SDS) zurückführen³. Durch die vom SDS erstmals angewandte direkte Erhebungsmethode entstand hier ein umfangreiches, sachlich gegliedertes Fragekorpus, dessen prinzipieller Aufbau aus Gründen der Vergleichbarkeit von darauf folgenden Sprachatlasprojekten, z.B. dem SÜDWESTDEUTSCHEN SPRACHATLAS, dem VORALBERGER SPRACHATLAS und eben auch dem BSA übernommen wurde. Als Grundfragebuch des BSA diente dabei das Fragebuch des SBS, dessen inhaltlicher und formaler Aufbau von den übrigen Projekten weitestgehend übernommen wurde. An einigen Stellen wurde das Fragebuch des SBS in den übrigen Teilprojekten jedoch modifiziert:

Inhaltliche Modifikationen ergaben sich z.B. aufgrund unterschiedlicher kultureller, geographischer oder historischer Umstände in den jeweiligen Regierungsbezirken der Teilprojekte. Konnte der Sachbereich *Weinbau* im Fragebuch des SBS mit lediglich sechs Fragen erhoben werden, wurde demselben Sachbereich im Fragebuch des SUF aufgrund des in Unterfranken weiter verbreiteten Weinanbaus deutlich mehr Raum gegeben: Anhand von insgesamt 88 Fragen erhoben die Exploratoren des SUF den dialektalen Fachwortschatz im Bereich *Weinbau*. Die Nummerierung der mit dem Fragebuch des SBS identischen Fragen (SBS 180.3 – SBS 180.8) wurde dabei beibehalten, die übrigen Fragen (insbesondere SUF 180.8Wa – 180.13Wu) wurden eingeschoben, die Zugehörigkeit zum SUF wird durch das Kürzel *W* angezeigt. Modifikationen ergaben sich auch durch die Tatsache, dass Fragen des SBS in anderen Projekten nicht übernommen wurden. So fehlt z.B. die Frage nach dem hornlosen Tier (SBS 40.4) in den übrigen Projekten des BSA. Teilweise sind den entsprechenden Nummern in den übrigen Fragebüchern andere Fragen zugeordnet, die Frage SNIB 40.4 erhebt z.B. das Vorkommen und die Bedeutung von *Schranne*.

Die größte formale Modifikation im Aufbau der Einzelfragebücher findet sich im Fragebuch des SMF. Das Fragebuch des SBS ist einseitig bedruckt, auf den jeweils geraden Seiten finden sich die Fragen, die ungeraden Seiten bieten Raum für Abbildungen, Kommentare, Spontanbelege u.ä. Um den Inhalt der ungeraden Seiten exakt zitieren zu können, wurden die Seiten fortlaufend durchnummeriert, die Fragen sind also zunächst durch die gerade Seitenzahl nummeriert, im Folgenden werden die Fragen jeder Seite laufend durchnummeriert, z.B. SBS 2.1 – 2.9, darauf folgen SBS 4.1 – 4.10. Das Fragebuch des SMF weicht von dieser Vorgehensweise gravierend ab. Hier bleiben die ungeraden Seiten ebenfalls frei, um Raum für Kommentare usw. zu geben, nummeriert werden jedoch nur die Seiten, die

³ vgl. König, 1997, S.18 f.

Fragen enthalten. Aus der Frage SBS 2.1. „Aufnahmeort (mdal. Form)“ wird die Frage SMF 1.1. Außerdem werden die Fragen auf den jeweiligen Seiten des SMF laufend durchnummeriert. Fragen, die von anderen Projekten eingefügt und durch Kleinbuchstaben bzw. Kürzel und Kleinbuchstaben gekennzeichnet wurden, z.B. SNOB 2.3a „Spottnamen der Ortseinwohner und Leute der Nachbarortschaften“ – das nachgestellte *a* zeigt an, dass die Frage eingefügt wurde –, verlieren ihre Kennzeichnung, aus SNOB 2.3a wird SMF 1.5, die übrigen Nummerierungen verschieben sich entsprechend⁴.

Um eine eindeutige Zuordnung eines Beleges zur entsprechenden Frage (und umgekehrt) zu ermöglichen, wurde von Mitarbeitern des SUF eine tabellarische Konkordanz sämtlicher im BSA erhobener Fragen und der entsprechenden Fragennummern in den einzelnen Projektfragebüchern erstellt. Die aus dieser Arbeit entstandene BSA-Fragekonkordanz enthält zunächst den Sachbereich und Text der jeweiligen Frage, danach folgen als Referenz die entsprechenden Fragennummern aus den Fragebüchern des SPRACHATLAS SCHWABEN-ALLGÄU (SSA), des TIROLER SPRACHATLAS (TSA) und des SDS, anschließend die Fragennummern aus den Teilprojekten des BSA in der Reihenfolge SBS, SNOB, SNIB, SOB, SMF und SUF. Zusätzlich wurden die Frageintention und Suggestierformen mit aufgenommen:

Sachbereich	Frage	SSA	TSA	SDS	SBS	SNOB	SNIB	SOB	SMF	SUF	Int.	Suggestierform
Das Vieh und seine Pflege	Melkgefäß (Name, Form, Material; mit oder ohne Schnabel?)	34.6	15.6	15.8/ 15.9/ 15.9a	28.1	28.1	28.1	28.1	15.8	28.1	Wo	Kübel/ Eimer/ Melter(n)/ Gelte
Das Vieh und seine Pflege	"Blech"	34.7	15.6		28.2	28.2	28.2	28.2	15.9	28.2	Ph	
Das Vieh und seine Pflege	"ein blecherner Eimer"			93.1		28.2a			15.10	28.2Ba	Mo	

Tabelle 2.1: Auszug aus der BSA-Fragekonkordanz

Die Frageintention zeigt an, auf welchen linguistischen Teilbereich die jeweilige Frage abzielt und diente den Exploratoren als Orientierung. Abhängig von den gegebenen Antworten konnte die eigentliche Intention bei den tatsächlichen Erhebungen vor Ort nicht immer eingehalten werden. Eine Lautfrage konnte z.B. aufgrund der gegebenen Antworten zur Lexikfrage werden und umgekehrt. Die Frageintentionen bilden für die BAYDAT-Datenbank den einzig möglichen Ausgangspunkt für die Recherche innerhalb linguistischer Teilbereiche,

⁴ vgl. Klepsch, S.36 f.

in einzelnen Fällen können die gefundenen Belege jedoch nicht dem intendierten Teilbereich entsprechen. Mögliche Werte für die Frageintention sind *Ph* (=Phonetik), *Mo* (=Morphologie), *Na* (=Eigennamen), *Soz* (=Sprachsoziologie), *Bed* (=Wortbedeutung), *Syn* (=Syntax) und *Wo* (=Lexik). Je nach Frage können mehrere Werte angegeben sein, die Frage SBS 10.5 "Kälblein" zielt z.B. sowohl auf die Bildung als auch die Lautung des Diminutivums ab, hier sind *Mo* und *Ph* als Frageintentionen angegeben.

Die Suggestierformen dienen als Hilfsmittel für die Erhebung von Lexikfragen. Gab die Gewährsperson nicht spontan Antwort, konnten die Exploratoren die Suggestierformen als Antwortmöglichkeiten vorgeben. Aufgrund des unterschiedlichen dialektalen Wortschatzes in den Dialekträumen des BSA können die Suggestierformen in den Einzelfragebüchern variieren. In die BSA-Fragekonkordanz wurden sämtliche Suggestierformen aus den Einzelfragebüchern übernommen. Sind wie in obigem Beispiel unter der Frage "Melkgefäß [...]" für eine Frage mehrere Suggestierformen angegeben, werden sie für die weitere elektronische Verarbeitung der Konkordanz durch Schrägstriche getrennt.

In den Einzelfragebüchern sind so genannte obligatorische Fragen – Fragen, deren Lexik und Flexionsform im Gegensatz zu Wort- und Bedeutungsfragen vorgegeben ist, gesondert gekennzeichnet. Hierzu dienen in der Regel – in den Fragebüchern der Einzelprojekte können andere typographische Kennzeichnungen verwendet werden – doppelte Anführungszeichen. Diese Kennzeichnung wurde in der BSA-Fragekonkordanz beibehalten (vgl. "Blech" und "ein blecherner Eimer" in obigem Beispiel). Obligatorische Fragen lassen sich auf diese Weise bei der elektronischen Weiterverarbeitung leicht von den übrigen Fragen trennen.

2.2. Die BSA-Ortsliste

1	suf1leu	97650	Leubach	NES	B12
2	suf2wei	97650	Weimarschmieden	NES	B13
3	suf3ror	97647	Roth	NES	C12
4	suf4ost	97461	Ostheim	NES	C13
5	suf5sto	97640	Stockheim	NES	C14
6	suf6mot	97786	Motten	NES	D9
7	suf7hal	97653	Haselbach	NES	D11
8	suf8obe	97656	Oberelsbach	NES	D12
9	suf9bas	97654	Bastheim	NES	D13
10	suf10mel	97638	Mellrichstadt	NES	D14

Abbildung 2.1: Auszug aus der BSA-Ortsliste

Die BSA-Ortsliste fasst sämtliche Erhebungsorte des BSA in einer Textdatei zusammen. Sie enthält – durch Tabulatoren getrennt – eine laufende Nummer, die Ortssigle, die Postleitzahl und den Namen des jeweiligen Erhebungsortes sowie das Kfz-Kennzeichen des jeweiligen Landkreises und das Planquadrat des Erhebungsortes. Die Ortssigle dient dazu, die BSA-

Erhebungsorte eindeutig kenntlich zu machen. Sie besteht aus einem Projektkürzel (SBS, SMF, SNB für den SPRACHATLAS VON NIEDERBAYERN, SNO für den SPRACHATLAS VON NORDOSTBAYERN, SOB oder SUF), einer laufenden Nummer, wobei die Erhebungsorte des jeweiligen Projektes von Nordwesten nach Südosten durchnummeriert wurden, und einem dreistelligen Ortskürzel. Durch die Ortssigle lassen sich gleichnamige Erhebungsorte in verschiedenen Projekten unterscheiden. So existieren Erhebungsorte mit dem Ortsnamen Sulzbach sowohl im SUF als auch im SNIB. Durch die Ortssiglen *suf49sul* bzw. *snb193sul* lassen sie sich auf den ersten Blick unterscheiden. Ebenso werden Zweideutigkeiten innerhalb der Einzelprojekte vermieden, z.B. gibt es im SUF einen Erhebungsort Sulzbach und einen Erhebungsort Sulzbach am Main, die durch die Ortssiglen *suf49sul* bzw. *suf109sam* eindeutig unterschieden werden können.

2.3. Die ASCII-Ortsdateien

Um die vor Ort erhobenen Dialektdaten im Rahmen der Kartenerstellung in den Teilprojekten des BSA sinnvoll und effizient weiterverarbeiten zu können, wurden sämtliche Fragebücher nach festgelegten Vorgaben elektronisch kodiert. So entstand für jeden Erhebungsort eine so genannte ASCII-Ortsdatei. Als Dateiname für die ASCII-Ortsdatei wurde jeweils der Name des Erhebungsortes gewählt, bei Ortsnamen mit mehr als acht Buchstaben Länge wurde der Dateiname systembedingt auf die ersten acht Buchstaben des Ortsnamens gekürzt, z.B. *Burghaus* für den SUF-Erhebungsort Burghausen oder *Tauberze* für den SMF-Erhebungsort Tauberzell. Umlautbuchstaben wurden durch *ae*, *oe* und *ue* ersetzt, die *sz*-Ligatur durch *ss*, z.B. *Wuerzbu* für den SUF-Erhebungsort Würzburg.

2.3.1. Grobstruktur der ASCII-Ortsdateien

Die Grobstruktur der ASCII-Ortsdateien orientiert sich am Aufbau des jeweiligen Projektfragebuchs:

```

16
1   / BRU=LD /
2   / DE5SSI5NS7O=-NE, KHU=- /
3   / DESSI5NS7O=-NE, KHE5.LBLI5 /
4   / DI-2A6-LDE, KHU- /
5   / O-2E1DE, 5R& ( #N# , SUGG ) /
6   / DSI5.DSN4 ( SUGG ) /
7   / S7WANS2 / S7WA-2E1V2 ( SUGG , "EVENTUELL NUR BEI PFERD" ) /
8   / S7WE5ND&S /
9   / S7WE5ND&SLE, /
10  / WE-DLN4 ( SUGG ) /
11  / VO-2E, &R&DE, 5V2U=-S ( <VORDER...> ) / BE2.- ( "BEI TIEREN NICHT
GEBRAEUCHLICH" ) /
12  / HINDE, 5R&V2U=-S ( <HINTER...> ) /

```

Abbildung 2.2: Auszug aus der ASCII-Ortsdatei für den SUF-Erhebungsort Würzburg, *Wuerzbu*

Zunächst findet man die jeweilige Seitenzahl, im obigen Beispiel "16". Da die Seitenzahlen isoliert in einer eigenen Zeile stehen, lassen sie sich in der elektronischen Weiterverarbeitung leicht von Fragenummern oder sonstigen im Kodat enthaltenen Zahlen unterscheiden. Darauf folgen die auf der jeweiligen Seite enthaltenen Fragen und die entsprechenden Belege und Kommentare.

Um die Fragenummern in der Folge leichter weiterverarbeiten zu können, sind sie – anders als in den meisten Projektfragebüchern – laufend durchnummeriert. Eingeschobene Fragen verlieren dadurch ihre Kennzeichnung, die Frage 16.3Ba "Das sind schöne Kälblein." aus dem Fragebuch des SUF wird in den ASCII-Ortsdateien zu Frage 16.3. Auch lässt sich anhand der ASCII-Ortsdateien nicht überprüfen, ob Fragen aus vorhergehenden Fragebüchern nicht übernommen wurden. So fehlt die Frage SBS 16.2 "Lockruf für die Kühe" im Fragebuch des SUF, in den ASCII-Ortsdateien des SUF entspricht die Frage SBS 16.3 der Frage SUF 16.2 "Das sind schöne Kühe".

Die Fragenummern sind durch vier Leerzeichen (bei einstelligen Fragenummern) bzw. drei Leerzeichen (bei zweistelligen Fragenummern) vom Beleg- und Kommentarteil getrennt. Sie lassen sich auf diese Weise in der weiteren Verarbeitung von anderen im Kodat enthaltenen Zahlen unterscheiden.

Die zu einer Frage erhobenen Belege und Kommentare stehen – jeweils durch Schrägstrich und Leerzeichen getrennt – im Anschluss an die zugehörige Fragennummer, bei längeren Belegen und Kommentaren kann ein Zeilenumbruch erfolgen. Um Beleg- und Kommentarteil unterscheidbar zu machen, steht der Kommentarteil jeweils in runden Klammern.

2.3.2. Elektronische Kodierung der Teuthonista-Lautschrift

Die Dialektdaten wurden im Rahmen der Erhebungen zum BSA in einer weiterentwickelten Form der Lautschrift TEUTHONISTA transkribiert. Die TEUTHONISTA-Lautschrift wurde 1924 erstmals in der Zeitschrift TEUTHONISTA vorgestellt. Als Grundzeichen verwendet sie bis auf wenige Ausnahmen die Buchstaben des lateinischen Alphabets, lautliche Phänomene wie Lautquantität, Öffnungsgrad usw. werden durch Diakritika unter- und oberhalb der Grundzeichen dargestellt.⁵

Da zur Kodierung der ASCII-Ortsdateien nur Zeichen aus dem ASCII-Zeichensatz (AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE), d.h. weitestgehend die Zeichen aus dem lateinischen Alphabet, Ziffern und gebräuchliche Sonderzeichen, verwendet werden konnten, mussten für die elektronische Erfassung insbesondere die Diakritika der

⁵ vgl. Klepsch, 2003, S.77

TEUTHONISTA-Lautschrift nach bestimmten Vorgaben in eine ASCII-konforme Kodierung umgesetzt werden. So wird z.B. der Öffnungsgrad eines Vokals durch dem Grundzeichen nachgestellte Ziffern angezeigt, die Ziffer 1 markiert eine sehr geschlossene, die Ziffer 6 eine sehr offene Artikulation⁶.

2.3.3. Untergliederung der Kommentare

Der Kommentarteil, der in den ASCII-Ortsdateien in runde Klammern gesetzt wurde, um ihn vom Lautschriftteil abzugrenzen, dient der Kommentierung der Lautschriftbelege. Die Kommentare lassen sich dabei grob in folgende Teile gliedern:

a) Lemmata

Ein Lemma gibt den lautschriftlichen Beleg schriftsprachlich und standardnah⁷ wieder, lemmatisiert werden vor allem Belege zu Lexikfragen. Bei obligatorischen Fragen werden Teillemmata (angezeigt durch Auslassungspunkte) angegeben, wenn die gegebene Antwort vom Fragetext abweicht. Um unsichere Lemmata anzuzeigen, kann dem Lemma ein Fragezeichen beigestellt werden.

b) Grammatische Angaben

Ist die Flexionsform eines Belegs nicht durch den Fragetext oder aufgrund des Kontextes direkt erkennbar, so muss diese im Kommentar angegeben werden.

c) Bedeutungsangaben

Bedeutungsangaben paraphrasieren die Bedeutung des Belegs, falls diese mehrdeutig oder aus sonstigen Gründen nicht ersichtlich ist.

d) Exploratorenkommentare

Exploratorenkommentare geben z.B. wieder, ob ein lautliches Phänomen sicher gehört wurde, ob spontan auf die Frage geantwortet wurde oder ob die Antwort extorquierte wurde.

e) Gewährspersonenkommentare

Gewährspersonenkommentare geben zusätzliche Angaben wieder, die die Gewährsperson zur gestellten Frage gemacht hat. Diese Angaben können teilweise recht umfangreich sein, teilweise sind sie mit weiteren Angaben (Lemmata, grammatische Angaben usw.) versehen und bilden damit Unterbelege zur jeweiligen Frage, z.B. in der ASCII-Ortsdatei für den SUF-Erhebungsort Burgwallbach, *Burgwall*, im Kommentar zu Frage SUF 18.11:

⁶ Eine vollständige Übersicht zur elektronischen Kodierung der TEUTHONISTA-Lautschrift findet sich bei König 1997, S.167 f.

⁷ zur Lemmatisierung vgl. Fritz-Scheuplein/König/Krämer-Neubert/Wolf, S.10 ff.

11 / DRE2N7GE,LE,5. (<TRAENKLEIN> , [KEIN DIMINUTIV]) /
 GLAŞ-2I;E,DRI5N7GE, ("WAR EIN S7DU=5GBRU-E,D <STUECK BROT> MIT DABEI") /

Um die einzelnen Kommentarteile unterscheidbar zu machen, wurden diese in den ASCII-Ortsdateien mit unterschiedlichen Trennzeichen versehen:

Trennzeichen	Komentarteil
<...>	Lemmata
#...#	Grammatische Angaben
...	Bedeutungsangaben
[...]	Exploratorenkommentar
"..."	Gewährspersonenkommentar

Tabelle 2.2: Kommentartrennzeichen in den ASCII-Ortsdateien

Eckige Klammern werden jedoch nicht nur für Exploratorenkommentare verwendet. Hier finden sich auch Ortsangaben, entweder in Form von Planquadraten oder unter Angabe von Ortsnamen, z.B. in der ASCII-Ortsdatei für den SUF-Erhebungsort Albersdorf, *Albersdo* in der Frage SUF 2.3:

A-2U5VBRA5ME,RI5X7 (<AUF BRAMBERG> , ORTSNAME , [I18])

Des Weiteren werden eckige Klammern für Zweckangaben benutzt, insbesondere bei Spontanbelegen (in den ASCII-Ortsdateien durch vorangestelltes SS gekennzeichnet), z.B. in o. g. ASCII-Ortsdatei auf S. 50:

SS / GRI-SE,LI5X7 (<?> , "BEVOR ES BUTTER GEWORDEN IST, WAREN ES ..." , [SACHKULTUR]) /

Außerdem stehen in eckigen Klammern Verweise auf andere Fragen, in o.g. Ortsdatei z.B. im Spontanbeleg auf S.86:

SS / U=NDE,LE5.NDEE:, (<UNTERLAENDER> , "HINTER HOFHEIM" , [478.6]) /

Daneben finden sich in eckigen Klammern Kommentarteile mit nicht eindeutigen Status. So kann der in eckige Klammern gesetzte Kommentarteil "Schibboleth-Satz für Randersacker" im Spontanbeleg auf S.50 der Ortsdatei für den SUF-Erhebungsort Würzburg, *Wuerzbur* im folgenden Beispiel eher als Bedeutungsangabe denn als Exploratoren-Kommentar gewertet werden:

SS / I-HO-A2.-A,E6.-U=- (<ICH HABE AUCH EIN EI UEBRIG> , [SCHIBBOLETH-SATZ FUER RANDERSACKER] , AO , [N11] , [LAUTUNG]) /

Ebenso ist der in eckigen Klammern gesetzte Kommentar "Ortsname" im dritten Spontanbeleg auf S. 142 der ASCII-Ortsdatei für den SUF-Erhebungsort Burgwallbach, *Burgwall* als Bedeutungsangabe zu werten:

SS / ME2DSGE,S7 {METZGERS} , [ORTSNAME]) /

Im Gegensatz zu anderen Kommentartrennern, die eine eindeutige Zuordnung zu ihrem jeweiligen Kommentarteil erlauben, stellen in eckige Klammern gesetzte Kommentaranteile für die elektronische Weiterverarbeitung ein Problem dar, eine eindeutige Zuordnung ist im Zweifelsfall nicht möglich.

Neben geklammerten Kommentaranteilen finden sich Kommentaranteile, die nicht durch Trennzeichen unterschieden sind, sondern frei zwischen runden Klammern bzw. zwischen den geklammerten Kommentaranteilen stehen, z.B. in Frage SUF 4.1 in der Ortsdatei für Erbshausen-Sulzwiesen, *Erbshaus*:

1 / A5NS7DE2.- (<ARNSTEIN> , ORTSNAME , [K11]) /

oder in Frage SUF 4.4 in o. g. Ortsdatei:

4 / 0 (WOUB) /

Bei diesen ungeklammerten Kommentaranteilen handelt es sich in der Regel um feste Abkürzungen, die sich weitestgehend den geklammerten Kommentaranteilen zuordnen lassen. So können die Abkürzungen *HD* für 'Hochdeutsch' oder *SUGG* für 'suggeriert' als Exploratorenkommentare gewertet werden, Bedeutungsangaben wie *EIGENNAME* oder *REDENSART* können dem entsprechenden geklammerten Kommentarteil für Bedeutungsangaben zugeordnet werden. Andere ungeklammerte Kommentaranteile lassen sich nur schwer unter die geklammerten Kommentaranteile einordnen. So geben die Abkürzungen *WOUB* ('Wort unbekannt'), *SAUB* ('Sache unbekannt') und *WOSAUB* ('Wort und Sache unbekannt') Begründungen für Nullbelege. Daneben finden sich im ungeklammerten Kommentar Verweise auf andere Fragen oder auf Abbildungen. Ebenfalls ungeklammert

bleiben Antworten auf Sachfragen, z.B. Antworten auf die Frage SUF 478.1, "In welchem Ort/welchen Orten spricht man genauso wie hier?". Eine klare Zuordnung ungeklammerter Kommentarteile ist im Zweifelsfall nicht möglich.

3. ORACLE – Ein relationales Datenbankmanagementsystem

Die erste Version des relationalen Datenbankmanagementsystems ORACLE kam in den späten 1970er Jahren auf den Markt. ORACLE wird seitdem von zahlreichen internationalen Firmen benutzt, in regelmäßigen Abständen wurden neue Versionen der Software veröffentlicht.⁸ Es ist daher davon auszugehen, dass die Software auch in den nächsten Jahren weiter gepflegt und unterstützt wird. Sie gewährleistet also eine langfristige, sichere Speicherung der BSA-Erhebungsdaten und ist daher die ideale Softwarelösung für die Zwecke der BAYDAT-Datenbank. Zum besseren Verständnis der folgenden Kapitel sollen hier knapp die grundlegenden Prinzipien des relationalen Datenmodells einerseits sowie das Konzept eines Datenbankmanagementsystems andererseits erläutert werden.

Das relationale Datenmodell wurde in den frühen 1970er Jahren durch den US-amerikanischen Mathematiker Edgar Frank Codd entwickelt⁹. Es geht zurück auf die Mengenlehre des russischen Mathematikers Georg Cantor¹⁰. Eine Menge ist nach Cantor eine Sammlung von einem oder mehreren Elementen des gleichen Typs, für die BAYDAT-Datenbank z.B. die Liste der Namen der BSA-Erhebungsorte. Innerhalb der Menge darf jedes Element nur einmal vorkommen, die Ordnung der Elemente innerhalb der Menge ist irrelevant. Die Elemente zweier Mengen lassen sich in einer so genannten Relation zusammenfassen, z.B. die Liste der Namen der BSA-Erhebungsorte und eine Liste der deutschen Postleitzahlen:

Postleitzahl	BSA-Erhebungsort
94542	Sachsenham
94496	Ortenburg
94496	Dorfbach

Tabelle 3.1: BSA-Erhebungsorte mit Postleitzahlen

Zu den genannten Relationen können weitere hinzutreten, z.B. die bayerischen Landkreise und Regierungsbezirke:

Postleitzahl	BSA-Erhebungsort	Landkreis	Regierungsbezirk
94542	Sachsenham	Passau	Niederbayern
94496	Ortenburg	Passau	Niederbayern
94496	Dorfbach	Passau	Niederbayern

Tabelle 3.2: BSA-Erhebungsorte mit Postleitzahl, Landkreis und Regierungsbezirk

Wie gezeigt, lassen sich Relationen physikalisch als Tabellen darstellen. Jede Zeile repräsentiert dabei einen so genannten Record (auch Tupel genannt), jede Spalte ein Feld (auch Attribut genannt)¹¹. Für die Zeilen einer Tabelle gilt dabei, wie auch schon für die

⁸ vgl. Abbey/Corey/Abramson, S.3 ff.

⁹ vgl. Codd

¹⁰ vgl. Warrender, S.13 ff.

¹¹ zur Terminologie vgl. Connolly/Begg, S.74

zugrunde liegenden Mengen, dass jede Zeile eindeutig sein muss und die Reihenfolge der Zeilen keine Rolle spielt.

Um die Eindeutigkeit jeder Zeile zu gewährleisten, verwendet man so genannte Primärschlüssel. Primärschlüssel sind einzelne Spalten oder die Kombination mehrerer Spalten, deren Werte die Zeile eindeutig definieren¹². Für o.g. Tabelle könnte die Spalte BSA-Erhebungsort den Primärschlüssel darstellen: Die einzelnen Zeilen sind durch die Ortsnamen *Sachsenham*, *Ortenburg* und *Dorfbach* eindeutig definiert, über die Postleitzahl ließen sich die Reihen nicht eindeutig festlegen, da z.B. die Postleitzahl 94496 innerhalb der Tabelle mehrmals vorkommt.

Außerdem ist, basierend auf Cantors Mengenlehre, festgelegt, dass die Werte innerhalb eines Feldes vom gleichen Typ sein müssen. Diese Regel lässt sich innerhalb eines relationalen Datenbankmanagementsystems rudimentär durch so genannte Datentypen überprüfen. Innerhalb der BAYDAT-Datenbank kommen dabei folgende Datentypen zum Einsatz:

Datentyp	Beschreibung
Char	Sequenz von Zeichen aus einem bestimmten Zeichensatz (für die BAYDAT-Datenbank ISO-8859-1, d.h. die Zeichen des lateinischen Alphabets inklusive westeuropäischer Sonderzeichen) mit einer festgelegten Länge (angegeben durch eine nachgestellte Zahl in Klammern). Z.B. char(5) für o.g. Postleitzahlen
Varchar2	Sequenz von Zeichen aus einem bestimmten Zeichensatz mit variabler Länge. Die maximale Länge kann durch eine nachgestellte Zahl in Klammern angegeben werden. Z.B. varchar2(30) für o.g. Ortsnamen, Landkreise und Regierungsbezirke
BLOB	= Binary Large Objects, d.h. Binärdateien von bis zu vier Gigabyte Größe. BLOBs werden in der BAYDAT-Datenbank u.a. zur Speicherung der originalen Fragebuchscans verwendet.

Tabelle 3.3: Datentypen innerhalb der BAYDAT-Datenbank

Indem Primärschlüssel als Spalten in anderen Tabellen verwendet werden – sie bilden dadurch so genannte Fremdschlüssel – können Tabellen miteinander in Verbindung gesetzt werden. Z.B. könnte in o.g. Beispiel der Name des Landkreises durch dessen Kfz-Kennzeichen ersetzt werden:

Postleitzahl	Ortsname	Landkreis	Regierungsbezirk
94542	Sachsenham	PA	Niederbayern
94496	Ortenburg	PA	Niederbayern
94496	Dorfbach	PA	Niederbayern

Tabelle 3.4: BSA-Erhebungsorte mit Postleitzahl, Regierungsbezirk und Kfz-Kennzeichen

¹² vgl. Warrander, S.15

Die Spalte *Landkreis* kann als Fremdschlüssel auf den Primärschlüssel einer weiteren Tabelle verweisen, die die Namen der Landkreise enthält:

Kfz-Kennzeichen	Landkreis
OAL	Ostallgäu
PA	Passau
PAF	Pfaffenhofen an der Ilm

Tabelle 3.5: Landkreise und zugehörige Kfz-Kennzeichen

Durch die Verteilung der Informationen auf zwei Tabellen kann die Konsistenz innerhalb der Datenbank gewährleistet werden: Da Fremdschlüssel auf tatsächlich existierende Primärschlüssel verweisen müssen, sind Fehler in der Zuweisung von Feldwerten nahezu ausgeschlossen.

Durch die Speicherung der Daten in einem Datenbankmanagementsystem ergibt sich ein entscheidender Vorteil für den Anwender: Sie sind softwareunabhängig¹³. Über standardisierte Schnittstellen – für die BAYDAT-Datenbank sind dies ODBC (OPEN DATABASE CONNECTIVITY STANDARD)¹⁴ und JDBC (JAVA DATABASE CONNECTIVITY)¹⁵, eine Schnittstelle, die es erlaubt, mit der Programmiersprache JAVA auf Datenbanken zuzugreifen – können beliebige Anwendungen, z.B. Textverarbeitungen, Tabellenkalkulationen oder Web-Anwendungen, auf die Datenbank zugreifen und benötigen nicht ihren jeweils eigenen Datenbestand. Redundanzen und Inkonsistenzen des Datenbestandes sind dadurch ausgeschlossen. Für die BAYDAT-Datenbank werden z.B. die Skriptsprache PERL zur Erstellung der Datenbank sowie die Middleware COCOON für die Online-Schnittstelle benutzt.

Für die Erstellung und Abfrage von relationalen Datenbanken hat sich die Mitte der 70er Jahre entwickelte STRUCTURED QUERY LANGUAGE (SQL) als De-facto-Standard etabliert, 1987 wurde SQL von der ISO (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION) als Standard festgelegt (ISO-9075)¹⁶. Für die Erstellung und Nutzung der BAYDAT-Datenbank sind zwei Komponenten von SQL besonders wichtig: die DATA DEFINITION LANGUAGE (DDL) und die DATA MANIPULATION LANGUAGE (DML).

Die DATA DEFINITION LANGUAGE dient der Erstellung von Datenbankobjekten, z.B. Tabellen¹⁷. Um die Datenbanktabellen aus o.g. Beispiel zu erzeugen, verwendet man folgende SQL-Anweisung:

¹³ vgl. Warrander, S.4

¹⁴ vgl. Connolly/Begg, S.676 ff.

¹⁵ vgl. Connolly/Begg, S.978 ff.

¹⁶ vgl. Connolly/Begg, S.112 ff.

¹⁷ vgl. Connolly/Begg, S.167 ff.

```
SQL:      1      create table orte (
           2          postleitzahl char(5),
           3          ortsname varchar2(30),
           4          landkreis varchar2(3),
           5          regierungsbezirk varchar2(20),
           6          primary key(ortsname),
           7          foreign key(landkreis) references landkreise
           8      );
           9
          10      create table landkreise (
          11          kfzkennzeichen varchar2(3),
          12          landkreis varchar2(30),
          13          primary key(kfzkennzeichen)
          14      );
```

Kommentar: In beiden Anweisungen wird zunächst festgelegt, welche Art von Datenbankobjekt (*table*, Z.1/10) erstellt werden soll, anschließend wird der Name des zu erstellenden Objekts (*orte* bzw. *landkreise*¹⁸) angegeben. Darauf folgen die einzelnen Spalten mit ihren jeweiligen Datentypen (z.B. *postleitzahl char(5)*, Z.2). Durch die Anweisungen *primary key(ortsname)* in Z.6 bzw. *primary key(kfzkennzeichen)* in Z.13 werden die Primärschlüssel der jeweiligen Tabelle festgelegt. Die Anweisung *foreign key(landkreis) references landkreise* in Z.7 legt in der Tabelle *Orte* die Spalte *Landkreis* als Fremdschlüssel fest, der zur Tabelle *Landkreise* in Beziehung steht.

Die DATA MANIPULATION LANGUAGE dient dazu, Daten in Tabellen einzufügen, sie bei Bedarf zu ändern und enthält Befehle zur Abfrage der Datenbank¹⁹. Um die in den Beispielen genannten Daten in die Tabellen einzufügen, kommt folgende SQL-Anweisung zur Anwendung:

```
SQL:      1      insert into landkreise values ('OAL', 'Ostallgäu');
           2      insert into landkreise values ('PA', 'Passau');
           3      insert into landkreise values ('KT', 'Kitzingen');
           4      insert into orte values ('94542', 'Sachsenham', 'PA',
           5          'Niederbayern');
           6      insert into orte values ('94496', 'Ortenburg', 'PA',
           7          'Niederbayern');
           8      insert into orte values ('94496', 'Dorfbach', 'PA',
           9          'Niederbayern');
          10      insert into orte values ('97318', 'Kitzingen', 'KT',
          11          'Unterfranken');
```

¹⁸ Die Groß- und Kleinschreibung spielt bei der Benennung von Datenbankobjekten keine Rolle, die Tabellen *Orte* bzw. *Landkreise* können innerhalb einer SQL-Anweisung als *Orte*, *orte/Landkreise*, *landkreise* oder mit gemischter Groß-/Kleinschreibung referiert werden.

¹⁹ vgl. Connolly/Begg, S.115 ff.

Kommentar: Nach dem Befehl *insert* wird jeweils festgelegt, in welche Tabelle Daten eingefügt werden sollen. Die entsprechenden Werte stehen zwischen runden Klammern, die Werte einzelner Felder sind durch Kommata getrennt, Werte für Felder des Typs *Char* bzw. *Varchar2* werden in einfache Anführungszeichen gestellt.

Um die Daten der Tabellen abzufragen, kann z.B. folgende Anweisung verwendet werden – sie wählt alle Ortsnamen aus der Tabelle *Orte* aus:

SQL:

```
select ortsname
from orte
```

Ergebnis:

Ortsname
Sachsenham
Ortenburg
Dorfbach
Kitzingen

Tabelle 3.6: Ergebnisse der SQL-Beispielabfrage 1

Die Ergebnisse lassen sich durch eine *where*-Klausel einschränken, in folgendem Beispiel werden nur die Ortsnamen von Orten mit der Postleitzahl 94496 ausgegeben:

SQL:

```
select ortsname
from orte
where postleitzahl = '94496'
```

Ergebnis:

Ortsname
Ortenburg
Dorfbach

Tabelle 3.7: Ergebnisse der SQL-Beispielabfrage 2

Schließlich lassen sich durch Unterabfragen mehrere Tabellen gleichzeitig abfragen, durch folgende Abfrage werden Ortsnamen von Orten aus dem Landkreis Passau ausgegeben:

SQL:

```
select ortsname
from orte
where landkreis = (
    select kfzkennzeichen
    from landkreise
    where landkreis = 'Passau'
)
```

Ergebnis:

Ortsname
Sachsenham
Ortenburg
Dorfbach

Tabelle 3.8: Ergebnisse der SQL-Beispielabfrage 3

4. Aufbereitung der Quelldateien für die bayerische Dialektdatenbank BAYDAT

In Kapitel 2 wurde die Datenerfassung und -kodierung im BSA beschrieben. Folgendes Kapitel beschreibt die Anpassung und Übernahme der Daten, von der BSA-Fragekonkordanz über die BSA-Orstliste hin zu den ASCII-Orstdateien, für und in die BAYDAT-Datenbank.

4.1. Umsetzung der BSA-Fragekonkordanz innerhalb der BAYDAT-Datenbank

4.1.1. Beschreibung der Datenbanktabellen

Um die BSA-Fragekonkordanz innerhalb eines relationalen Datenbankmanagementsystems sinnvoll nutzbar zu machen, müssen die in der Konkordanz enthaltenen Informationen auf mehrere Tabellen verteilt werden:

- Die Suggestierformen stehen in der BSA-Fragekonkordanz gruppiert in einer Spalte, jeweils getrennt durch einen Schrägstrich. Da innerhalb der BAYDAT-Datenbank Abfragen der einzelnen Suggestierformen möglich sein sollen, wird eine separate Tabelle für die Suggestierformen benötigt, die Tabelle *Suggestierformen*.
- Einzelnen Fragen können, wie unter 2.1 beschrieben, mehrere Frageintentionen zugewiesen sein. Die Frageintentionen stehen in diesem Fall in der BSA-Fragekonkordanz ebenfalls in einer Spalte gruppiert und sind durch Kommata getrennt. Da Datenbankabfragen nach Frageintentionen möglich sein sollen, wird auch hier eine separate Tabelle benötigt, die Tabelle *Frageintentionen*. Die Frageintentionen sind nur in abgekürzter Form in die BSA-Fragekonkordanz aufgenommen. Daher enthält eine weitere Tabelle – *Frageintentionen_Klartext* – die Frageintentionen in ausgeschriebener Form.
- Um Redundanzen zu vermeiden, werden die 59 Sachbereiche, welchen die Fragen der BSA-Fragekonkordanz zugeordnet sind, nicht in ausformulierter Form in die Haupttabelle *Fragen* übernommen. Für die Sachbereiche wird eine separate Tabelle *Sachbereiche* erstellt.

Insgesamt verteilt sich die BSA-Fragekonkordanz innerhalb der BAYDAT-Datenbank auf fünf Tabellen: *Fragen*, *Suggestierformen*, *Frageintentionen*, *Frageintentionen_Klartext* und *Sachbereiche*.

a) Die Tabelle *Fragen*

Feld	Datentyp	Inhalt
Fragenr	Char(4)	laufende Fragennummer
Frage	Varchar2(500)	Fragetext
Sachbereich	Char(2)	Nummer des Sachbereichs
Obligatorisch	Char(1)	Frage obligatorisch oder nicht?
SBS	Char(7)	Fragennummer aus dem Fragebuch des SBS
SNO	Char(7)	Fragennummer aus dem Fragebuch des SNOB
SNB	Char(7)	Fragennummer aus dem Fragebuch des SNIB
SOB	Char(7)	Fragennummer aus dem Fragebuch des SOB
SMF	Char(7)	Fragennummer aus dem Fragebuch des SMF
SUF	Char(7)	Fragennummer aus dem Fragebuch des SUF
Link	Char(1)	Gehen von der Frage weiterführende Links aus?
Bild	Char(1)	Sind mit der Frage Bilder verknüpft?

Tabelle 4.1: Datenfelder und -typen der Tabelle *Fragen*

Die Tabelle *Fragen* stellt die wichtigste Tabelle bei der Übernahme der BSA-Fragekonkordanz in die BAYDAT-Datenbank dar. Sie enthält die Fragetexte der einzelnen Fragen sowie die Fragennummern aus den einzelnen Projektfragebüchern.

Die Fragetexte konnten ohne Veränderung aus der BSA-Fragekonkordanz in das Feld *Frage* der Tabelle übernommen werden. Die Fragennummern aus den Projektfragebüchern wurden vor der Übernahme in die einzelnen Felder (*SBS*, *SNO*, *SNB*, *SOB*, *SMF* und *SUF*) der BAYDAT-Datenbank weiter aufbereitet: Die Fragennummern in der BSA-Fragekonkordanz wurden direkt aus den Projektfragebüchern übernommen, hier finden sich also – wie unter 2.1. beschrieben – Projektkürzel und Kleinbuchstaben als Kennzeichnung für Fragen, die im Fragebuch des SBS nicht enthalten sind. Die Fragennummerierung innerhalb der ASCII-Ortsdateien weicht hiervon – wie unter 2.2. gezeigt – ab, die Fragen sind auf der jeweiligen Seite laufend durchnummeriert. Um die Nummerierung der Fragen innerhalb der BAYDAT-Datenbank zu vereinheitlichen, wurde die Nummerierung innerhalb der Tabelle *Fragen* der Nummerierung in den ASCII-Ortsdateien angepasst. Zusätzlich wurden den jeweiligen Seitenzahlen und Fragennummern bis zu zwei führende Nullen vorangestellt, Seitenzahl und Fragennummer werden durch einen Punkt getrennt, damit wird z.B. die Fragennummer *SUF* 258.2Ba im Feld *SUF* der Tabelle *Fragen* zu 258.003, die Fragennummer *SBS* 2.1 wird im Feld *SBS* zu 002.001. Durch das Voranstellen führender Nullen vor Seitenzahl und Fragennummer erhalten die Felder *SBS*, *SNO*, *SNB*, *SOB*, *SMF* und *SUF* eine feste Breite von sieben Zeichen. Dies erleichtert einerseits die Abfragen innerhalb der BAYDAT-Datenbank, andererseits

können auf diese Weise falsch nummerierte Fragen bei der Übernahme der Daten in die BAYDAT-Datenbank schnell erkannt werden.

Zusätzlich zu den Fragenummern der einzelnen Projektfragebücher enthält die Tabelle *Fragen* im Feld *Fragenr* eine laufende Fragennummer. Diese entspricht der Position der jeweiligen Frage in der BSA-Fragekonkordanz. Sie dient einerseits als Primärschlüssel innerhalb der Tabelle *Fragen*, andererseits dient sie als Grundlage der in Kapitel 5 beschriebenen Auszeichnung der Fragen unter sprachwissenschaftlichen Gesichtspunkten. Um Abfragen innerhalb der BAYDAT-Datenbank zu erleichtern, erhält das Feld *Fragenr* eine feste Breite von vier Zeichen. Fragennummern, die weniger als vier Ziffern lang sind, werden führende Nullen vorangestellt.

Informationen darüber, ob eine Frage obligatorisch ist oder nicht, sind in der BSA-Fragekonkordanz nur implizit enthalten, indem obligatorische Fragen in Anführungszeichen gesetzt werden. Das Feld *Obligatorisch* macht diese Information explizit: Durch die beiden Abkürzungen "j" und "n" wird angezeigt, ob es sich bei der entsprechenden Frage um eine obligatorische ("j") oder nicht-obligatorische Frage ("n") handelt. Entsprechend kann bei Abfragen explizit auf obligatorische Fragen zugegriffen werden.

Die Sachbereiche sind in der Tabelle *Fragen* nicht im Klartext aufgenommen. Vielmehr enthält das Feld *Sachbereich* die laufende Nummer des zur Frage gehörigen Sachbereichs. Über diese laufende Nummer wird auf die Tabelle *Sachbereiche* verwiesen, welche die Sachbereiche im Klartext enthält.

Zusätzlich zu den aus der BSA-Fragekonkordanz übernommenen Daten enthält die Tabelle *Fragen* ein Feld *Link*. Durch den Wert "j" im Feld *Link* wird angezeigt, dass die entsprechende Frage mit weiteren Fragen verknüpft ist. So ist Frage 529 "Reiten" mit Fragen verknüpft, die flektierte Formen des Verbs *reiten* beinhalten, entsprechend trägt das Feld *Link* für Frage 529 den Wert "j". Außerdem wird im Feld *Bild* durch den Wert "j" angezeigt, dass die entsprechende Frage mit einem Bild aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS verknüpft ist (vgl. 6.2.).

b) Die Tabelle *Suggestierformen*

Feld	Datentyp	Inhalt
Fragenr	Char(4)	Fragennummer der zur Suggestierform gehörenden Frage
Suggestierform	Varchar2(500)	Suggestierform

Tabelle 4.2: Datenfelder und -typen der Tabelle *Suggestierformen*

Die Tabelle *Suggestierformen* enthält die beiden Felder *Fragenr* und *Suggestierform*. Das Feld *Fragenr* verweist dabei auf das Feld *Fragenr* in der Tabelle *Fragen*. Durch das Auslagern der Suggestierformen in eine eigene Tabelle lassen sich diese einzeln abfragen. Durch den Verweis auf die Tabelle *Fragen* lässt sich die zu einer Suggestierform gehörende Frage leicht ermitteln. Umgekehrt können die zu einer Frage gehörenden Suggestierformen ermittelt werden.

c) Die Tabellen *Frageintentionen* und *Frageintentionen_Klartext*

Feld	Datentyp	Inhalt
Fragenr	Char(4)	Fragennummer der zur Frageintention gehörenden Frage
Frageintention	Varchar(3)	Kürzel der jeweiligen Frageintention

Tabelle 4.3: Datenfelder und -typen der Tabelle *Frageintentionen*

Feld	Datentyp	Inhalt
Frageintention_Abk	Varchar2(3)	Kürzel der Frageintention
Frageintention	Varchar2(500)	Frageintention im Klartext

Tabelle 4.4: Datenfelder und -typen der Tabelle *Frageintentionen_Klartext*

Die Tabelle *Frageintentionen* enthält zwei Felder, *Fragenr* und *Frageintention*. Das Feld *Fragenr* verweist auf das Feld *Fragenr* in der Tabelle *Fragen*. Durch Kreuzabfragen lassen sich so einerseits die zu einer Frage gehörenden Frageintentionen ermitteln, andererseits können sämtliche Fragen mit einer bestimmten Frageintention, z.B. sämtliche Lautfragen, ermittelt werden.

Das Feld *Frageintention* enthält lediglich ein Kürzel der jeweiligen Frageintention und verweist auf das Feld *Frageintention_Abk* in der Tabelle *Frageintentionen_Klartext*. Durch die Abfrage der Tabelle *Frageintentionen_Klartext* können die entsprechenden Frageintentionen im Klartext ermittelt werden.

d) Die Tabelle *Sachbereiche*

Feld	Datentyp	Inhalt
Sachberechnr	Char(2)	Nummer des Sachbereichs
Sachbereich	Varchar2(500)	Titel des Sachbereichs im Klartext

Tabelle 4.5: Datenfelder und -typen der Tabelle *Sachbereiche*

Um Redundanzen zu vermeiden, wurde im Feld *Sachbereich* der Tabelle *Fragen* nur die laufende Nummer des zur jeweiligen Frage gehörenden Sachbereichs aufgenommen. Die Tabelle *Sachbereiche* enthält im Feld *Sachbereich* die Titel der einzelnen Sachbereiche im Klartext. Durch die Verknüpfung des Felds *Sachberechnr* mit dem Feld *Sachbereich* in der Tabelle *Fragen* lässt sich der Sachbereich einer Frage leicht ermitteln. Umgekehrt können sämtliche Fragen, die zu einem bestimmten Sachbereich gehören, gefunden werden.

4.1.2. Beziehungen zwischen den Tabellen zur BSA-Fragekonkordanz

Die Beziehungen zwischen den Tabellen zur BSA-Fragekonkordanz lassen sich folgendermaßen schematisch darstellen:

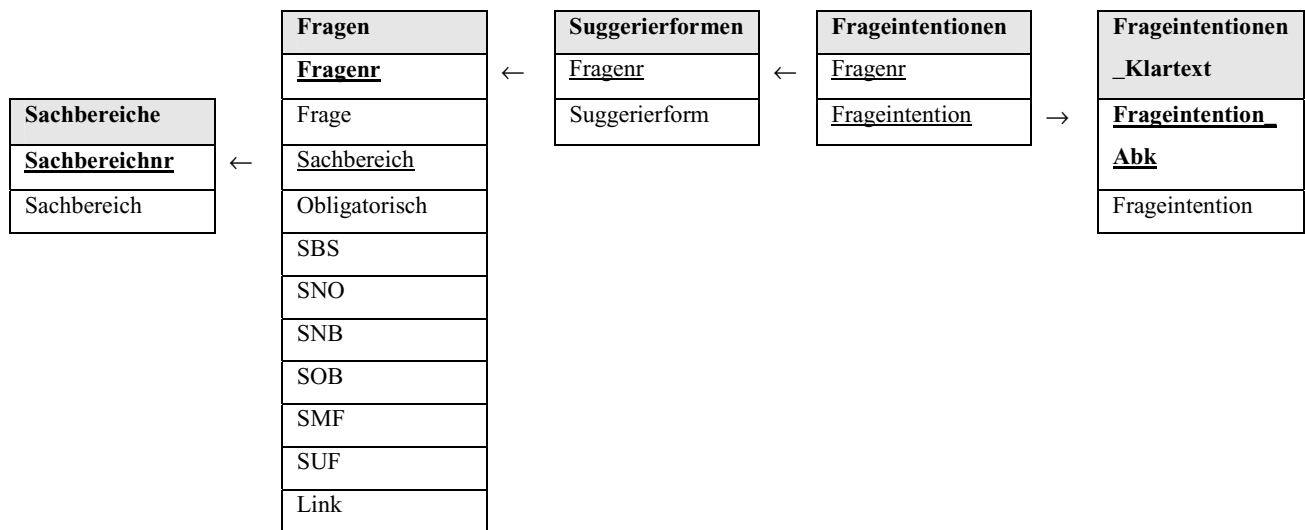


Abbildung 4.1: Beziehungen zwischen den Tabellen zur BSA-Fragekonkordanz

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

4.1.3. Beispielabfragen zur BSA-Fragekonkordanz

a) Beispielabfrage 1: Obligatorische Fragen mit der Zeichenfolge *Kalb*

Abfrage: Es sollen alle im SUF erhobenen obligatorischen Fragen (Fragennummer, SUF-Fragennummer, Fragetext) ermittelt werden, die die Zeichenfolge *Kalb* enthalten.

SQL:

```

1  select fragenr, suf, frage
2  from fragen
3  where obligatorisch = 'j'
4  and suf is not null
5  and frage like '%Kalb%'
    
```

Kommentar: Aus der Tabelle *Fragen* werden diejenigen Zeilen ausgewählt, die im Feld *Frage* die Zeichenfolge *Kalb* enthalten (*frage like '%Kalb%', Z.5*)²⁰. Wurden die entsprechenden Fragen im SUF (*suf is not null, Z.4*) obligatorisch (*obligatorisch = 'j', Z.3*) erhoben, werden deren Fragennummern, SUF-Fragennummern und Fragetexte als Ergebnis ausgegeben.

Ergebnis:

Fragenr	SUF	Frage
0062	008.002	"Kalben"
0070	010.002	"Kalb"

Tabelle 4.6: Ergebnisse der Abfrage 1 "Obligatorische Fragen mit der Zeichenfolge *Kalb*"

²⁰ '%' ist innerhalb des in ORACLE verwendeten SQL-Dialekts das Kürzel für keine oder beliebig viele beliebige Zeichen; vgl. Warrender, S.125

b) Beispielabfrage 2: Morphologiefragen aus dem Sachbereich *Gemüse*

Abfrage: Es sollen alle Morphologiefragen (Fragennummer und -text) aus dem Sachbereich *Gemüse* ausgewählt werden.

```
SQL:  1   select fragenr, frage
      2   from fragen
      3   where sachbereich = (
      4       select sachbereichnr
      5       from sachbereiche
      6       where sachbereich = 'Gemüse'
      7   ) and fragenr in (
      8       select fragenr
      9       from frageintentionen
     10       where frageintention = 'Mo'
     11   )
```

Kommentar: Aus der Tabelle *Fragen* werden diejenigen Zeilen ausgewählt, deren Werte im Feld *Sachbereich* mit der Sachbereichsnummer des Sachbereichs *Gemüse* in der Tabelle *Sachbereiche* übereinstimmen (Z.3 – 7) und deren Werte im Feld *Fragenr* mit den Werten des Felds *Fragenr* aus der Tabelle *Frageintentionen* übereinstimmt, deren Zeilen im Feld *frageintention* den Wert 'Mo' enthalten (Z.7 – 11). Für die entsprechenden Zeilen werden Fragennummer und -text als Ergebnis ausgegeben.

Ergebnis:

Fragenr	Frage
2964	"Garten" (oder Hausgarten?)
2965	"Gärten"
2966	"Gärtlein"
2977	"Säcklein"
2978	"Säcke"
2984	Sellerie (Anlaut!)
2986	"Wirsing(kohl)" (mit krausen Blättern)
2991	"Erbsen"
3002	"Gelb" (ein gelber Kürbis)
3003	Krautkopf
3004	"Ein Kopf Blaukraut"
3005	"Ein Kopf Weißkraut"
3006	"Ein Kopf Wirsing"
3013	Feuchtigkeit

Tabelle 4.7: Ergebnisse der Abfrage 2 "Morphologiefragen aus dem Sachbereich *Gemüse*"

c) Beispielabfrage 3: Suggestierformen mit der Zeichenfolge *Haus*

Abfrage: Es sollen alle Fragen (Fragennummer, -text, Suggestierform) zu Suggestierformen ermittelt werden, die die Zeichenfolge *Haus* enthalten.

```
SQL:  1   select fragen.fragenr, frage, suggestierform
      2   from fragen, suggestierformen
      3   where fragen.fragenr = suggestierformen.fragenr
      4   and suggestierform like '%Haus%'
```

Kommentar: Aus der Tabelle *suggestierformen* werden die Zeilen ausgewählt, deren Werte im Feld *suggestierform* die Zeichenfolge *Haus* enthalten (*suggestierform like*

'%Haus%', Z.4). Die Werte im Feld *Fragenr* der ausgewählten Zeilen werden mit den Werten im Feld *Fragenr* der Tabelle *Fragen* abgeglichen (*fragen.fragenr* = *suggerierformen.fragenr*, Z.3), anschließend werden die entsprechenden Fragenummern aus der Tabelle *Fragen* (*fragen.fragenr*, Z.1), Fragetexte und Suggestierformen als Ergebnis ausgegeben.

Ergebnis:

Fragenr	Frage	Suggestierform
1773	Wein aus Trester mit Zuckerwasser versetzt	Haustrunk
1885	"Das Haus des Vaters"	an Vater sei Haus
2444	Der Hase, den man im Stall hält	Haushase

Tabelle 4.8: Ergebnisse der Abfrage 3 "Suggestierformen mit der Zeichenfolge Haus"

d) Beispielabfrage 4: Obligatorische Fragen im Sachbereich *Katze*

Abfrage: Es sollen alle obligatorischen Fragen (Fragennummer, SMF-Fragennummer, Fragetext) ermittelt werden, die im SMF im Sachbereich *Katze* erhoben wurden.

SQL:

```

1   select fragenr, smf, frage
2   from fragen
3   where obligatorisch = 'j'
4   and smf is not null
5   and sachbereich = (
6       select sachbereichnr
7       from sachbereiche
8       where sachbereich = 'Katze'
9   )

```

Kommentar: Aus der Tabelle *Fragen* werden die Zeilen ausgewählt, deren Wert im Feld *Sachbereich* der Sachbereichsnummer des Sachbereichs *Katze* aus der Tabelle *Sachbereiche* entspricht (Z.5 – 9). Wurden die entsprechenden Fragen im SMF (*smf is not null*, Z.4) obligatorisch (*obligatorisch* = 'j', Z.3) erhoben, werden Fragennummer, SMF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

Fragenr	SMF	Frage
0552	034.006	"Unsere Katze"
0553	034.007	"Unsere Katzen"
0554	034.008	"Mit unseren (unsernen) Katzen sind wir zufrieden"

Tabelle 4.9: Ergebnisse der Abfrage 4 "Obligatorische Fragen im Sachbereich Katze"

4.2. Umsetzung der BSA-Ortsliste innerhalb der BAYDAT-Datenbank

4.2.1. Beschreibung der Datenbanktabellen

Die BSA-Ortsliste enthält, wie unter 2.3. gezeigt, Informationen zu den Erhebungsorten des BSA. Um diese im Rahmen der BAYDAT-Datenbank sinnvoll nutzbar zu machen bzw. Redundanzen innerhalb der Datenbank zu vermeiden, wurden diese auf mehrere Tabellen verteilt:

- In der BSA-Ortsliste zeigen Kfz-Kennzeichen den Landkreis des jeweiligen Erhebungsortes an. Zur Ermittlung der Landkreisnamen wird eine Tabelle *Lk* erstellt.
- Als zusätzliche Information wird in der BAYDAT-Datenbank der Regierungsbezirk jedes Erhebungsortes angegeben. Um Redundanzen zu vermeiden, wurden innerhalb der Haupttabelle *Orte* Kürzel für die Regierungsbezirke aufgenommen, die Tabelle *Regbez* enthält die Namen der Regierungsbezirke im Klartext.

Insgesamt verteilen sich die Informationen aus der BSA-Ortsliste in der BAYDAT-Datenbank auf drei Tabellen: *Orte*, *Lk* und *Regbez*.

a) Die Tabelle *Orte*

Feld	Datentyp	Inhalt
Signle	Char(9)	Ortssigle
Plz	Char(5)	Postleitzahl
Name	Varchar2(40)	Ortsname
Pq	Varchar2(5)	Planquadrat
Lk	Varchar2(3)	Kfz-Kennzeichen des jeweiligen Landkreises
Rb	Char(2)	Kürzel des jeweiligen Regierungsbezirks
Auflage	Char(1)	zur Erhebung verwendete Auflage des SUF-Fragebuchs

Tabelle 4.10: Datenfelder und -typen der Tabelle *Orte*

Die Tabelle *Orte* enthält die Hauptinformationen aus der BSA-Ortsliste. Die Daten wurden weitestgehend ohne Veränderung aus der BSA-Ortsliste übernommen. Der Wert im Feld *Rb* ergibt sich aus den ersten drei Zeichen der Ortssigle, zur Vermeidung von Redundanzen wurde das dreistellige Projektkürzel auf zwei Stellen gekürzt, SUF wird entsprechend zu "uf", SMF zu "mf", SOB zu "ob", SNB zu "nb", SBS zu "bs" und SNO zu "no". Als zusätzliche Information enthält die Tabelle *Orte* im Feld *Auflage* die bei der jeweiligen Erhebung verwendete Auflage des SUF-Fragebuchs. Der in diesem Feld angegebene Wert dient als Ausgangspunkt zu der unter 6.1 beschriebenen Verknüpfung mit den Scans der SUF-Originalfragebücher.

b) Die Tabelle *Lk*

Feld	Datentyp	Inhalt
Id	Varchar2(3)	Kfz-Kennzeichen des jeweiligen Landkreises
Landkreis	Varchar2(40)	Name des Landkreises

Tabelle 4.11: Datenfelder und -typen der Tabelle *Lk*

Zur Vermeidung von Redundanzen innerhalb der BAYDAT-Datenbank wurde im Feld *Landkreis* der Tabelle *Orte* nur das Kfz-Kennzeichen des jeweiligen Erhebungsortes aufgenommen. Um den Namen eines Landkreises zu ermitteln, kann auf die Tabelle *Lk* zurückgegriffen werden, die jeweils das Kfz-Kennzeichen und den dazugehörigen Namen des Landkreises enthält.

c) Die Tabelle *Regbez*

Feld	Datentyp	Inhalt
Id	Char(2)	Kürzel des jeweiligen Regierungsbezirks
Regierungsbezirk	Varchar2(40)	Name des Regierungsbezirks

Tabelle 4.12: Datenfelder und -typen der Tabelle *Regbez*

Die Tabelle *Regbez* dient ebenfalls zur Vermeidung von Redundanzen innerhalb der BAYDAT-Datenbank. Um die im Feld *Rb* der Tabelle *Orte* enthaltenen zweistelligen Kürzel der Regierungsbezirke zu entschlüsseln, wird bei Abfragen auf die Tabelle *Regbez* zugegriffen.

4.2.2. Beziehungen zwischen den Tabellen zur BSA-Ortsliste

Die Beziehungen der Tabellen zur BSA-Ortsliste lassen sich folgendermaßen schematisch darstellen:

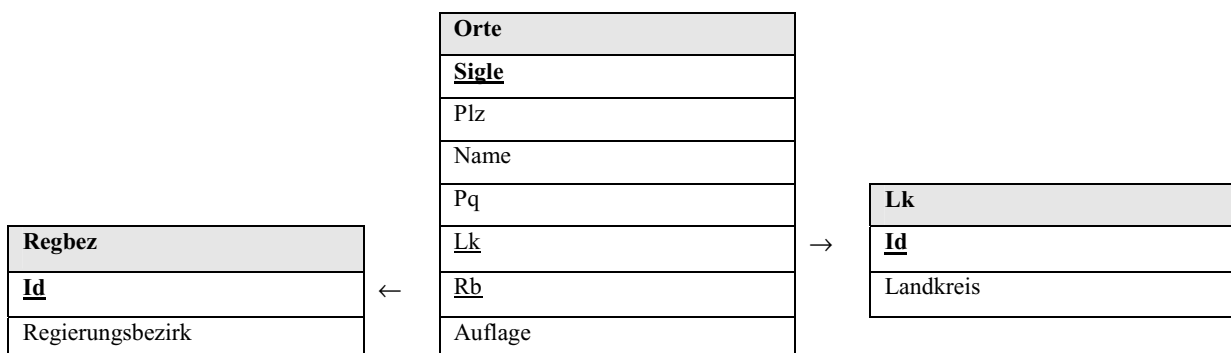


Abbildung 4.2: Beziehungen zwischen den Tabellen zur BSA-Ortsliste

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

4.2.3. Beispielabfragen zur BSA-Ortsliste

a) Beispielabfrage 1: Erhebungsorte mit Postleitzahlen zwischen 97000 und 97200

Abfrage: Es sollen alle Erhebungsorte (Postleitzahl, Ortsname, Landkreis) ermittelt werden, deren Postleitzahlen zwischen 97000 und 97200 liegen. Die Ergebnisse sollen nach Postleitzahlen sortiert ausgegeben werden.

SQL:

```

1   select plz, name, landkreis
2   from orte, lk
3   where plz between '97000' and '97200'
4   and lk = id
5   order by plz

```

Kommentar: Aus der Tabelle *Orte* werden diejenigen Zeilen ausgewählt, in welchen die Werte im Feld *Plz* zwischen 97000 und 97200 (*plz between '97000' and '97200', Z.3*) liegen. Die Anweisung in Z.4 stellt sicher, dass nur die Landkreise für die zuvor ausgewählten Zeilen als Ergebnis ausgegeben werden. Als Ergebnis werden – nach Postleitzahl sortiert (*order by plz, Z.5*) – Postleitzahl, Ortsname und der Landkreis des entsprechenden Ortes ausgegeben.

Ergebnis:

Plz	Name	Landkreis
97074	Würzburg	Würzburg
97084	Heidingsfeld	Würzburg
97199	Erlach	Würzburg
97199	Ochsenfurt	Würzburg

Tabelle 4.13: Ergebnisse der Abfrage 1 "Erhebungsorte mit Postleitzahlen zwischen 97000 und 97200"

b) Beispielabfrage 2: Erhebungsorte im Landkreis Fürth

Abfrage: Es sollen alle Erhebungsorte (Sigle, Postleitzahl, Ortsname) aus dem Landkreis Fürth ermittelt werden. Die Ergebnisse sollen nach Ortsnamen sortiert ausgegeben werden.

SQL:

```

1   select sigle, plz, name
2   from orte, lk
3   where landkreis = 'Fürth'
4   and lk = id
5   order by name

```

Kommentar: Aus der Tabelle *Lk* wird die Zeile ausgewählt, die im Feld *Landkreis* den Wert 'Fürth' enthält (*landkreis = 'Fürth', Z.3*). Aus der Tabelle *Orte* werden diejenigen Zeilen ausgewählt, deren Wert im Feld *Lk* dem Wert im Feld *Id* der oben ausgewählten Zeile entspricht (*lk = id, Z.4*). Aus den entsprechenden Zeilen werden Ortssigle, Postleitzahl und Ortsname als Ergebnis ausgegeben.

Ergebnis:	Sigle	Plz	Name
	smf056cad	90556	Cadolzburg
	smf054lan	90579	Langenzenn
	smf040omb	90587	Obermichelbach
	smf039pus	90617	Puschendorf
	smf071roß	90574	Roßtal
	smf055sek	90556	Seckendorf (=Cadolzburg)
	smf070vin	90613	Vincenzbronn
	smf053whd	91452	Wilhermsdorf

Tabelle 4.14: Ergebnisse der Abfrage 2 "Erhebungsorte im Landkreis Fürth"

c) Beispielabfrage 3: Unterfränkische Erhebungsorte mit Anfangsbuchstaben *P*

Abfrage: Es sollen alle Erhebungsorte (Postleitzahl, Ortsname, Landkreis) aus dem Regierungsbezirk Unterfranken ermittelt werden, deren Namen mit dem Buchstaben *P* beginnen. Die Ergebnisse sollen nach Ortsnamen sortiert ausgegeben werden.

SQL:

```

1  select plz, name, landkreis
2  from orte, lk
3  where name like 'P%'
4  and rb = 'uf'
5  and lk = id
6  order by name

```

Kommentar: Aus der Tabelle *Orte* werden diejenigen Zeilen ausgewählt, deren Wert im Feld *Name* mit *P* beginnt (*name like 'P%'*, Z.3) und die im Feld *Rb* den Wert *'uf'* enthalten (*rb = 'uf'*, Z.4). Die Anweisung in Z.5 stellt sicher, dass nur die Landkreise zu den vorher ausgewählten Zeilen ausgegeben werden. Als Ergebnis werden – nach Ortsnamen sortiert (*order by name*, Z.6) – Postleitzahl, Ortsname und Landkreis ausgegeben.

Ergebnis:	Plz	Name	Landkreis
	97846	Partenstein	Main-Spessart
	96176	Pfarrweisach	Haßberge
	97490	Poppenhausen	Schweinfurt
	97688	Poppenroth	Bad Kissingen
	97279	Püssensheim	Würzburg

Tabelle 4.15: Ergebnisse der Abfrage 3 "Unterfränkische Erhebungsorte mit Anfangsbuchstaben *P*"

4.3. Umsetzung der ASCII-Ortsdateien innerhalb der BAYDAT-Datenbank

4.3.1. Beschreibung der Datenbanktabellen

Die in den ASCII-Ortsdateien enthaltenen Informationen sind in der BAYDAT-Datenbank in der Tabelle *Belege* enthalten. Eine weitere Tabelle *Verweise* dient der Verknüpfung von automatisch erzeugten Unterbelegen mit den jeweiligen Ausgangsbelegen bzw. der Verknüpfung von Ausgangs- und Zielbelegen von im Kommentar angegebenen Verweisen.

a) Die Tabelle *Belege*

Feld	Datentyp	Inhalt
Ortssigle	Char(9)	Ortssigle des Belegs
Frage_Nr	Char(7)	Fragennummer des Belegs
Beleg_Nr	Char(3)	Nummer des Belegs
Lautschrift	Varchar2(500)	Lautschrift in ASCII-Kodierung
Lemma	Varchar2(500)	Lemma
Sicher	Char(1)	Lemma sicher?
Vollstaendig	Char(1)	Lemma vollständig?
Gp_Kom	Varchar2(1000)	Gewährspersonenkommentar
Bedeutung	Varchar2(500)	Bedeutungsangabe
Grammatik	Varchar2(500)	Grammatische Angabe
Expl_Kom	Varchar2(1000)	Exploratorenkommentar
Zweck	Varchar2(500)	Zweckangabe
Geo_Pq	Varchar2(500)	Geografische Angabe in Form von Planquadraten
Geo_On	Varchar2(500)	Geografische Angabe in Form von Ortsnamen
Link	Char(1)	Beleg mit weiteren Belegen verknüpft?
Gp	Varchar2(500)	Angabe zur Gewährsperson
Expl	Varchar2(500)	Angabe zur Exploration
Antwort_Kom	Varchar2(500)	Kommentar zur Geltung der Antwort
Nullbeleg	Varchar2(500)	Grund für Nullbeleg
Antwort_Sf	Varchar2(1000)	Antwort auf Sachfrage
TeuthoUC	Varchar2(500)	Lautschrift in TEUTHOUC (UNICODE-Zeichensatz)
TeuthoBD	Varchar2(500)	Lautschrift in TEUTHOBD (ASCII-Zeichensatz)

Tabelle 4.16: Datenfelder und -typen der Tabelle *Belege*

Die Werte innerhalb des Feldes *Ortssigle* entsprechen den Ortssiglen aus der BSA-Ortsliste. Sie werden durch den Abgleich der Dateinamen der ASCII-Ortsdateien mit den Ortsnamen in der BSA-Ortsliste ermittelt: Ein Abgleich des Dateinamen der ASCII-Ortsdatei für den SUF-Erhebungsort Würzburg, *Wuerzbur*, mit der BSA-Ortsliste ergibt die Ortssigle *suf149wue*. Alle Belege des SUF-Erhebungsortes Würzburg erhalten im Feld *Ortssigle* der Tabelle *Belege* den Wert *suf149wue*. Durch eine Kreuzabfrage der Tabelle *Belege* und der Tabelle *Orte* lassen sich so Belege aus ausgewählten Erhebungsorten ermitteln.

Die Werte im Feld *Frage_Nr* ergeben sich aus einer Kombination der Seiten- und Frage-Nummer des jeweiligen Belegs. Seiten- und Fragenummern mit einer Länge von weniger als drei Zeichen werden führende Nullen vorangestellt, zwischen Seiten- und Fragenummer steht ein Punkt. Damit entsprechen die Werte im Feld *Frage_Nr* den jeweiligen Fragenummern in den Feldern *SBS*, *SNO*, *SNB*, *SOB*, *SMF* oder *SUF* in der Tabelle *Fragen*. Mittels einer Kreuzabfrage der Tabellen *Fragen* und *Belege* lassen sich so gezielt Belege zu ausgewählten Fragen finden.

Wie unter 2.2.1. gezeigt, können zu einzelnen Fragen mehrere Belege erhoben worden sein. In der Tabelle *Belege* wird jedem Einzelbeleg eine eigene Zeile zugewiesen. Die Einzelbelege zu einer Frage werden im Feld *Beleg_Nr* laufend dreistellig durchnummeriert, kürzeren Belegnummern werden führende Nullen vorangestellt.

Über die Werte in den Feldern *Ortssigle*, *Frage_Nr* und *Beleg_Nr* ist jeder Beleg innerhalb der BAYDAT-Datenbank eindeutig identifizierbar, die drei Felder bilden zusammen den Primärschlüssel der Tabelle *Belege*.

Lemmata werden ohne Veränderung aus den ASCII-Ortsdateien in das Feld *Lemma* der Tabelle *Belege* übernommen. Zusätzlich enthält die Tabelle *Belege* zwei weitere Felder, die den Status des jeweiligen Lemmas anzeigen: Dem Feld *Sicher* wird der Wert "n" zugewiesen, falls das entsprechende Lemma in den ASCII-Ortsdateien mit einem Fragezeichen versehen und damit als unsicher markiert ist, andernfalls wird dem Feld *Sicher* der Wert "j" zugewiesen. Das Feld *Vollstaendig* erhält den Wert "n", wenn das entsprechende Lemma in den ASCII-Ortsdateien durch Auslassungspunkte als unvollständig markiert ist, andernfalls erhält es den Wert "j".

Gewährspersonenkommentare, Bedeutungsangaben und grammatische Angaben wurden ohne Veränderungen aus den ASCII-Ortsdateien übernommen und in der Tabelle *Belege* den entsprechenden Feldern *Gp_Kom*, *Bedeutung* und *Grammatik* zugewiesen.

Teilweise können die Gewährspersonenkommentare, wie unter 2.2.3. gezeigt, Lautschriftpassagen mit weiteren Angaben enthalten. Diese werden als Unterbelege in eine eigene Zeile überführt und damit – mit einigen Abweichungen, vgl. hierzu 3.3.2. – regulären Belegen gleichgesetzt.

Enthält ein Kommentar mehrere grammatische Angaben, so werden diese gruppiert in das Feld *Grammatik* übernommen und durch Kommata getrennt.

In der Regel enthalten eckige Klammern innerhalb der Kommentarteile der ASCII-Ortslisten Kommentare der jeweiligen Exploratoren. Wie unter 2.2.3. beschrieben, können hier jedoch auch andere Angaben stehen, im Einzelnen sind dies:

- Ortsangaben

Ortsangaben finden sich in Form von Planquadraten oder in Form von Ortsnamen. Aufgrund ihrer Struktur – ein Buchstabe, dem ein oder zwei Ziffern nachgestellt sind – lassen sich Planquadrate leicht von anderen Angaben, die in eckigen Klammern stehen können, unterscheiden. Sie werden automatisch dem Feld *Geo_Pq* zugewiesen. Um die Planquadrate der Einzelprojekte unterscheiden zu können, wird ihnen die jeweilige dreistellige Projektsigle vorangestellt.

Problematischer erweist sich die automatische Erkennung von Ortsangaben in Form von Ortsnamen. Diese sind nur durch Kompositionszweitglieder wie *-burg*, *-heim*, *-brunn*, *-dorf* usw. erkennbar, auch Regierungsbezirke und an Bayern angrenzende Bundesländer lassen sich automatisch ermitteln. Sind Ortsnamen als solche erkennbar, werden sie dem Feld *Geo_On* zugewiesen, andernfalls werden sie in das Feld *Expl_Kom* aufgenommen.

Enthält ein Kommentar mehrere Ortsangaben, so werden diese zusammen in die entsprechenden Felder *Geo_Pq* und *Geo_On* übernommen und durch Kommata getrennt.

- Verweise

Verweise lassen sich ebenso wie Planquadrate aufgrund ihrer Struktur leicht automatisch ermitteln: Auf die ein- bis dreistellige Seitennummer folgt, durch einen Punkt getrennt, die Fragenummer. Da diese Fragenummer der Fragenummer aus den Projektfragebüchern entspricht, kann sie aus Ziffern und Buchstaben bestehen. Nachdem die Verweise in ihrer Form an die Fragenummern innerhalb der BAYDAT-Datenbank (Projektsigle, dreistellige Seitennummer, Punkt, laufende dreistellige Fragenummer) angeglichen wurden, können diese in die Tabelle *Verweise* übernommen werden.

- Zweckangaben

Die Zweckangaben entsprechen weitestgehend festgelegten Kennungen (*FREMDWORT*, *IDIOMATIK*, *LAUTUNG*, *LEXIK*, *LYRIK*, *METASPRACHE*, *MORPHOLOGIE*, *SACHKULTUR*, *SEMANTIK*, *SYNTAX* und *WORTBILDUNG*). In diesen Fällen lassen sie sich leicht automatisch erkennen und in das Feld *Zweck* überführen. Sind innerhalb eines Kommentars mehrere Zweckangaben notiert, werden diese im Feld *Zweck* gruppiert aufgenommen und durch Kommata getrennt. Teilweise finden sich selbst gewählte Kennungen bzw. Abkürzungen normierter Kennungen,

diese lassen sich nur bedingt automatisch erkennen. Im Zweifelsfall werden diese in das Feld *Expl_Kom* übernommen.

- Bedeutungsangaben

Bedeutungsangaben lassen sich nur bedingt automatisch erkennen. Die Bedeutungsangaben *FLURNAME*, *ORTSNAME*, *HAUSNAME*, *EIGENNAME* und *UEBERNAME* treten gehäuft in eckigen Klammern auf und können automatisch dem Feld *Bedeutung* zugewiesen werden. Andere Bedeutungsangaben lassen sich nicht automatisch erkennen und werden damit in das Feld *Expl_Kom* übernommen.

Entspricht der Inhalt eckiger Klammern nicht den oben angegebenen Mustern, wird davon ausgegangen, dass es sich dabei um einen Exploratorenkomentar handelt, er wird entsprechend in das Feld *Expl_Kom* übernommen.

Als freistehende Kommentare finden sich zahlreiche feste Abkürzungen, die den Feldern *Gp*, *Expl*, *Antwort_Kom* und *Nullbeleg* zugeordnet werden können:

In das Feld *Gp* werden Angaben zur Gewährsperson aufgenommen. Als Abkürzungen finden sich hier:

Abkürzung	Ausformulierte Form
AEGP/AEGF/AEGM	ältere Gewährsperson/ältere Gewährsfrau/älterer Gewährsmann
AGP	andere Gewährsperson
GF/GM	Gewährsfrau/Gewährsmann
GPI/GP2 usw.	Gewährsperson 1/Gewährsperson 2 usw.
JGP/JGF/JGM	jüngere Gewährsperson/jüngere Gewährsfrau/jüngerer Gewährsmann
JJGP	viel jüngere Gewährsperson

Tabelle 4.17: Abkürzungen für Angaben zur Gewährsperson

Teilweise ist den Abkürzungen die Kennzeichnung *ID* vorangestellt, um anzuzeigen, dass eine zweite Gewährsperson eine Antwort gegeben hat, die mit derjenigen der ersten Gewährsperson identisch ist.

Dem Feld *Expl* werden Angaben zum Explorator und zur Exploration zugewiesen. Abkürzungen sind:

Abkürzung	Ausformulierte Form
AK (SUF)	Koexploratorin Almut König
AK (SMF)	Explorator Alfred Klepsch
AM	Explorator Alexander Mang
CK	Koexploration Cathrin Krüger
GD	Exploratorin Gabriele Diewald
GS	Explorator Gunther Schunk
HB	Explorator Hannes Bauer
IW (Kürzel zu Projektbeginn)	Exploratorin Isabel Knörrich
MF (SMF, Kürzel zu Projektbeginn)	Exploratorin Margit Wittowski
MF (SUF)	Koexploration Monika Fritz-Scheuplein
MH	Explorator Markus Hartfiel
MR	Koexploration Manfred Renn

RK	Explorator Rudolf Kleinöder
SF	Exploratorin Stefanie Fuchs
SK	Koexploration Sabine Krämer-Neubert
SR	Exploratorin Sibylle Reichel
TB	Tonbandaufnahme
TV	Videoaufnahme existiert
WG	Explorator Wolfgang Gräbel
WK	Koexploration Werner König

Tabelle 4.18: Abkürzungen für Angaben zum Explorator und zur Exploration

Das Feld *Antwort_Kom* enthält Angaben zur Geltung der Antwort. Hierfür finden sich folgende Abkürzungen bzw. standardisierte Kennungen:

Abkürzung	Ausformulierte Form
AFFEKT	Affekthafte Antwort
ALLEGRO	Allegro
ANL	Anlaut
AO	Anderer Ort
AOE	Am Objekt erlebt
AUSL	Auslaut
E/EE	Erinnerungsform/sehr alte Erinnerungsform
EXPL KORR	Explorator korrigiert
EXT(ORQUIERT)	Extorquiert
GP KORR	Gewährsperson korrigiert
HD/HOCHDEUTSCH	Hochdeutsch
HEITER	Heiter
HEUTE	Heute
LENTO	Lento
MDA	Mundartlich
NACHFRAGE/NF	Nachfrage
PAUSE	Pause
REP	Repetiert
SCHRIFTSPRACHE	Schriftsprache
(ANL/AUSL) (UN)SICHER	(Anlaut/Auslaut) (un)sicher
SOZ	Sprachsoziologisch
SPONTAN	Spontan
SUGG	Suggestierte Antwort
SUGG 1	Zögernde Antwort
SUGG 2	Sehr zögernde Antwort
UMGANGSSPRACHE	Umgangssprache
VERALTET	Veraltet

Tabelle 4.19: Abkürzungen für Angaben zur Geltung der Antwort

Den Angaben *(ANL/AUSL) SICHER* und *(ANL/AUSL) UNSICHER* wird häufig eine Lautschriftpassage vorangestellt, die als sicher bzw. unsicher gehört dargestellt werden soll. Zur Abgrenzung gegen den schriftsprachlichen Kommentar wird der Lautschriftteil in eckige Klammern gesetzt. Schließlich werden dem Feld *Nullbeleg* Begründungen für Nullbelege zugewiesen. Es können folgende Abkürzungen vorkommen:

Abkürzung	Ausformulierte Form
WOUB	Wort unbekannt
SAUB	Sache unbekannt
WOSAUB	Wort und Sache unbekannt
SUGG ... ABGELEHNT	Suggestierform ... abgelehnt

Tabelle 4.20: Abkürzungen für Begründungen von Nullbelegen

Vor der Übernahme in die entsprechenden Felder werden die Abkürzungen in ihre ausformulierte Form umgesetzt. Enthält ein Kommentar mehrere Angaben für ein Feld, werden diese gruppiert in das entsprechende Feld aufgenommen und durch Kommata getrennt.

Neben oben angegebenen Abkürzungen finden sich im ungeklammerten Kommentar die Kennungen *EIGENNAME*, *FLURNAME*, *HAUSNAME* und *ORTSNAME*. Diese werden in das Feld *Bedeutungsangabe* überführt.

Außerdem können im ungeklammerten Kommentar Verweise auf in den Projektfragebüchern enthaltene Abbildungen stehen. Diese lassen sich aufgrund ihrer Form – *ABB A/B/C*, *ABB 1/2/3*, *ABB UNBEKANNT*, *ABB 1 UND/ODER 3* usw. – automatisch erkennen und werden dem Feld *Expl_Kom* zugewiesen. Zur Unterscheidung von anderen Exploratorenkomentaren werden sie in eckige Klammern gesetzt.

Sonstige ungeklammerte Kommentarteile werden dem Feld *Antwort_Sf* zugewiesen, da davon auszugehen ist, dass es sich bei formal nicht standardisierten ungeklammerten Kommentarteilen um Antworten auf Sachfragen handelt bzw. eine automatische Erkennung nicht standardisierter ungeklammerter Kommentarteile und die Zuweisung in entsprechende Felder nicht möglich ist.

Die kodierten Lautschriftbelege wurden in ihrer ursprünglichen Form in das Feld *Lautschrift* übernommen. Zwei zusätzliche Felder, *TeuthoUC* und *TeuthoBD*, enthalten eine Umsetzung der Originalkodierung in die TRUETYPE-Schriftarten TEUTHOUC und TEUTHOBD. Diese Umsetzung ermöglicht eine Anzeige der Lautschriftbelege in gängigen Webbrowsern bzw. PDF-Readern²¹.

Das Feld *Link* zeigt lediglich an, ob vom entsprechenden Beleg aus auf andere Belege (reguläre Belege oder Unterbelege) verwiesen wird. In diesem Fall wird dem Feld der Wert "j" zugewiesen, andernfalls bleibt es leer. Die tatsächlichen Verweise finden sich in der Tabelle *Verweise*:

b) Die Tabelle Verweise

Feld	Datentyp	Inhalt
Sigle_Quelle	Char(9)	Ortssigle des Ausgangsbelegs
Frage_Quelle	Char(7)	Fragennummer des Ausgangsbelegs
Beleg_Quelle	Char(3)	Belegnummer des Ausgangsbelegs
Sigle_Ziel	Char(9)	Ortssigle des Zielbelegs
Frage_Ziel	Char(7)	Fragennummer des Zielbelegs

Tabelle 4.21: Datenfelder und -typen der Tabelle *Verweise*

²¹ Zu den TRUETYPE-Schriftarten TEUTHOBD und TEUTHOBC sowie ihrer Umsetzung für die BAYDAT-Datenbank vgl. 8.2.

Einzelne Belege können im Kommentarteil – in der Regel in eckigen Klammern – Verweise auf Belege zu anderen Fragen enthalten. Außerdem werden automatisch Verweise erzeugt, wenn aus Gewährspersonenkommentaren Unterbelege ausgegliedert werden. Da ein Beleg mehrere Verweise enthalten kann, werden diese in die separate Tabelle *Verweise* übernommen. Die Tabelle enthält in den Feldern *Sigle_Quelle*, *Frage_Quelle* und *Beleg_Quelle* alle Informationen, die zur eindeutigen Identifikation des Ausgangsbelegs notwendig sind. In den Spalten *Sigle_Ziel* und *Frage_Ziel* werden die Ortssigle und Fragenummer des Verweisziels aufgenommen. Eine Angabe der Belegnummer ist hier nicht möglich, da in den Kommentaren nicht auf Einzelbelege, sondern auf Fragenummern verwiesen wird.

4.3.2. Behandlung von Unterbelegen und Spontanbelegen in der BAYDAT-Datenbank

a) Behandlung von Unterbelegen

Unterbelege werden automatisch erzeugt, wenn Gewährspersonenkommentaren weitere Kommentarteile untergeordnet sind. So enthält der erste Beleg zu Frage 102.4 in der ASCII-Ortsdatei für den SUF-Erhebungsort Ostheim vor der Rhön, *OSTHEIM*, im Gewährspersonenkommentar eine Lautschriftpassage "DU=-E,LI5X7", die mit "<TUERLEIN>" lemmatisiert wird:

```
4 / A$-2UV2SEDSB4RE6.DE5E:, ( {AUFSETZBRETT} , #PL# , "MIT DU=-E,LI5X7
<TUERLEIN> VORNE UND HINTEN" ) / A$UVSEDSB4RE5DE5E:, /
```

Da der Gewährspersonenkommentar keine Lautschriftpassagen enthalten soll, wird im Ausgangsbeleg der BAYDAT-Datenbank die Lautschriftpassage durch das entsprechende Lemma ersetzt. Um durch Lemmata ersetzte Lautschriftpassagen kenntlich zu machen, wird das Lemma in eckige Klammern gesetzt, zusätzlich wird dem Lemma die Fragenummer des automatisch erzeugten Unterbelegs nachgestellt. Die Fragenummer des Unterbelegs setzt sich dabei aus der Seitennummer des Ausgangsbelegs und einer laufenden Fragenummer mit vorangestelltem "U" zusammen. Der Gewährspersonenkommentar zu oben genanntem Beleg lautet in der BAYDAT-Datenbank " MIT [TUERLEIN]102.U01 VORNE UND HINTEN."

b) Behandlung von Spontanbelegen

Spontanbelege sind in den ASCII-Ortsdateien dadurch gekennzeichnet, dass sie statt einer Fragenummer die Kennung "SS" erhalten. In der ASCII-Ortsdatei für den SUF-Erhebungsort Ostheim vor der Rhön, *OSTHEIM*, findet sich z.B. auf Seite 4 folgender Spontanbeleg:

SS / S7NE5BV (<?> , "IST EINE IN EINEM BALKEN EINGESCHNITZTE FIGUR" , [LEXIK]) /

In der BAYDAT-Datenbank werden Spontanbelege durch ihre Fragenummer gekennzeichnet: Die Seitennummer entspricht dabei der Seite, auf der sich der jeweilige Spontanbeleg befindet, die Fragenummer setzt sich aus der Kennung "S" und einer laufenden Nummer zusammen. Der Spontanbeleg aus oben genanntem Beispiel erhält in der BAYDAT-Datenbank die Fragenummer "004.S01."

4.3.3. Beziehungen zwischen den Tabellen zu den ASCII-Ortsdateien

Die Beziehungen zwischen den Tabellen zu den ASCII-Ortsdateien lassen sich schematisch wie folgt darstellen:

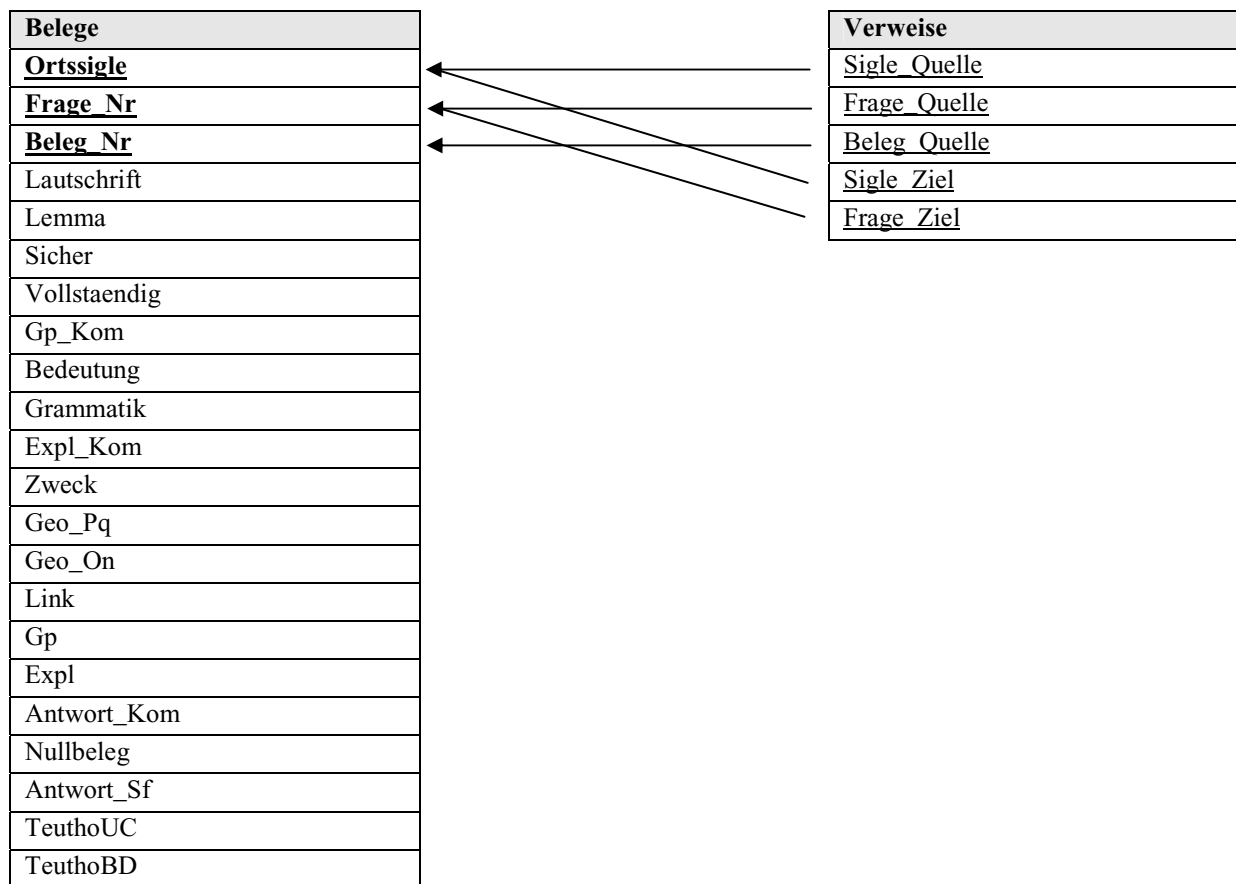


Abbildung 4.3: Beziehungen zwischen den Tabellen zu den ASCII-Ortsdateien

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

4.3.4. Beziehungen der Tabelle *Belege* zu anderen Tabellen der BAYDAT-Datenbank

Die Tabellen *Belege*, *Fragen* und *Orte* bilden den Kern der BAYDAT-Datenbank. Durch Kreuzabfragen der drei Tabellen lassen sich sämtliche Belege der BAYDAT-Datenbank abfragen. Folgende Abbildung stellt die Beziehungen der Tabellen *Belege*, *Fragen* und *Orte* dar:

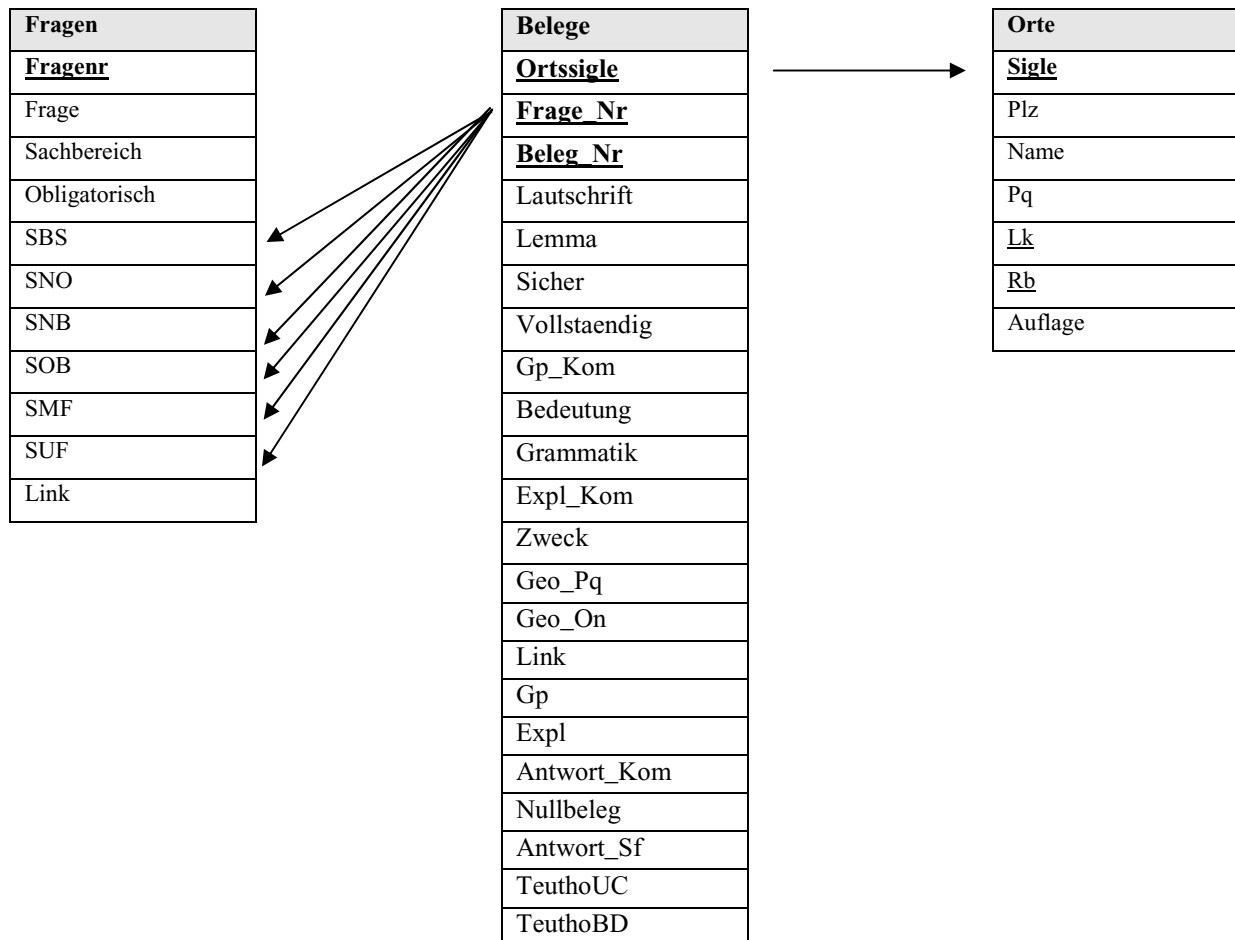


Abbildung 4.4: Beziehungen der Tabelle *Belege* zu anderen Tabellen

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

4.3.5. Beispielabfragen zu den ASCII-Ortsdateien

a) Beispielabfrage 1: Belege aus dem Sachbereich *Brot*

Abfrage: Es sollen alle Belege (Fragetext, Postleitzahl, Ort, Lemma) zu Lexikfragen aus dem Sachbereich *Brot* für Erhebungsorte im Postleitzahlenbereich zwischen 96190 und 96200 ermittelt werden. Die Belege sollen nach Fragen und Postleitzahl sortiert werden, nicht lemmatisierte Belege bzw. Belege mit unsicherem Lemma sollen nicht angezeigt werden.

```

SQL: 1   select frage, plz, name, lemma from
      2   belege, orte, fragen
      3   where sachbereich = (
      4       select sachbereichnr
      5       from sachbereiche
      6       where sachbereich = 'Brot'
      7   ) and plz between '96190' and '96200'
      8   and sigle = ortssigle
      9   and (
     10       frage_nr = sbs
     11       or frage_nr = sno
     12       or frage_nr = snb
     13       or frage_nr = sob
     14       or frage_nr = smf
     15       or frage_nr = suf
     16   ) and lemma is not null
     17   and sicher = 'j'
     18   order by frage, plz

```

Kommentar: Aus der Tabelle *Fragen* werden diejenigen Fragen ausgewählt, deren Wert im Feld *Sachbereich* mit der Sachbereichsnummer des Sachbereichs *Brot* übereinstimmt (Z.3 – 7). Die Werte in der Spalte *Frage_Nr* der auszuwählenden Belege müssen den Werten in den Spalten *SBS*, *SNO*, *SNB*, *SOB*, *SMF* oder *SUF* der oben ausgewählten Fragen entsprechen (Z.10 – 15). Aus der Tabelle *Orte* werden diejenigen Orte ausgewählt, deren Postleitzahl zwischen 96190 und 96200 liegt (*plz between '96190' and '96200'*, Z.7). Der Wert im Feld *Ortssigle* der auszuwählenden Belege muss dem Wert *sigle* der gewählten Orte entsprechen (*sigle = ortssigle*, Z.8). Ist das Feld *Lemma* der so ausgewählten Belege nicht leer (*lemma is not null*, Z.16) und ist das Lemma sicher (*sicher = 'j'*, Z.17), werden Fragetext, Postleitzahl, Ortsname und Lemma als Ergebnis ausgegeben.

Ergebnis:

Frage	PLZ	Name	Lemma
Brötchen (allgemein) (Vorsicht: Sorten!)	96193	Wachenroth	WECKLEIN
Brötchen (allgemein) (Vorsicht: Sorten!)	96193	Wachenroth	BROETLEIN
Das letzte Stück	96193	Wachenroth	DAS HINTERE KIPFLEIN
"Morgen wollen wir backen (nicht auf 'wollen' beharren)	96193	Wachenroth	MORGEN BACKEN WIR
Runder harter Brotanschnitt	96193	Wachenroth	ANSCHNITTLEIN
Sehr großes Stück Brot	96193	Wachenroth	RANKEL
Sehr großes Stück Brot	96193	Weingartsgreuth	RANFT
Übliche Brotformen Rundbrot	96193	Wachenroth	LAIB BROT
Wenn das Brot zu lange liegen bleibt, dann tut es... (wenn es grün wird)	96193	Wachenroth	SCHIMMLET
Wenn das Brot zu lange liegen bleibt, dann tut es... (wenn es grün wird)	96193	Wachenroth	IST GESCHIMMELT

Tabelle 4.22: Ergebnisse der Abfrage 1 "Belege aus dem Sachbereich Brot"

b) Beispielabfrage 2: Belege mit der Zeichenfolge *Kalb*

Abfrage: Es sollen alle Belege (Fragetext, Postleitzahl, Ort, Belege in TEUTHOBD) aus dem Landkreis Kitzingen zu obligatorischen Fragen, die im Fragetext die Zeichenfolge *Kalb* enthalten, gefunden werden. Die Belege sollen nach Fragen sortiert werden. Belege, deren Feld *TeuthoBD* leer ist, sollen nicht angezeigt werden.

SQL:

```

1  select frage, plz, name, teuthobd
2  from fragen, orte, belege
3  where lk = (
4      select id
5      from lk
6      where landkreis = 'Kitzingen'
7  ) and sigle = ortssigle
8  and frage like '%Kalb%'
9  and frage_nr = suf
10 and teuthobd is not null
11 and obligatorisch = 'j'
12 order by frage, plz

```

Kommentar: Aus der Tabelle *Fragen* werden diejenigen Zeilen ausgewählt, deren Feld *Frage* die Zeichenfolge *Kalb* enthält (*frage like '%Kalb%'*, Z.8) und die obligatorisch erhoben wurden (*obligatorisch = 'j'*, Z.11). Der Wert im Feld *Suf* muss mit dem Wert im Feld *Frage_Nr* der Tabelle *Belege* übereinstimmen (Z.9), es werden also nur Belege zu den ausgewählten Fragen ausgegeben. Der Wert im Feld *Lk* der Tabelle *Orte* muss mit dem Wert des Feldes *Id* aus der Zeile der Tabelle *Lk* übereinstimmen, die im Feld *Landkreis* den Wert '*Kitzingen*' enthält (Z.3 – 7). Die Sigle der ausgewählten Orte muss mit der Ortssigle in der Tabelle *Belege* übereinstimmen (*sigle = ortssigle*, Z.7), es werden also nur Belege aus den ausgewählten Orten ausgegeben. Ist das Feld *TeuthoBD* der ausgewählten Belege nicht leer, werden Fragetext, Postleitzahl, Ortsname und Lautschrift als Ergebnis ausgegeben.

Ergebnis:

Frage	PLZ	Name	TeuthoBD
"Kalb"	97334	Sommerach	kʰɛl̩bl̩α
"Kalb"	97337	Dettelbach	k̩alb
"Kalb"	97346	Dornheim	kʰalb
"Kalb"	97348	Hüttenheim in Bayern	k̩alb̩l̩α
"Kalb"	97355	Castell	kʰɛl̩bl̩α
"Kalb"	97357	Altenschönbach	kh̩kʰalb
"Kalb"	97357	Altenschönbach	bul̩αkʰalb
"Kalb"	97359	Stadtschwarzach	kɛl̩bl̩α
"Kalben"	97334	Sommerach	kʰɛl̩bɛr̩n
"Kalben"	97337	Dettelbach	sɔdɛdamkɛl̩bɛrn

"Kalben"	97346	Dornheim	kəɫbɛrn
"Kalben"	97348	Hüttenheim in Bayern	kʌlb
"Kalben"	97355	Castell	kʰɛɫbɛr
"Kalben"	97357	Altenschönbach	kʰɛɫbɛrd
"Kalben"	97359	Stadtschwarzach	kɛɫbɛrd

Tabelle 4.23: Ergebnisse der Abfrage 2 "Belege mit der Zeichenfolge *Kalb*"**c) Beispielabfrage 3: Spontanbelege mit Zweckangabe *Lexik***

Abfrage: Es sollen alle Spontanbelege (Belege in TEUTHOBD, Lemma, Bedeutung) mit der Zweckangabe *Lexik* aus dem SUF-Erhebungsort Würzburg ermittelt werden.

SQL:

```

1  select teuthobd, lemma, bedeutung
2  from belege
3  where ortssigle = (
4      select sigle
5      from orte
6      where name = 'Würzburg'
7  ) and frage_nr like '%.S%'
8  and zweck like '%LEXIK%'

```

Kommentar: Aus der Tabelle *Belege* werden diejenigen Zeilen ausgewählt, deren Wert im Feld *Ortssigle* der Ortssigle des SUF-Erhebungsortes Würzburg in der Tabelle *Orte* entspricht (Z.3 – 7). Außerdem muss es sich um Spontanbelege (*frage_nr like '%.S%'*, Z.7) mit der Zweckangabe *Lexik* (*zweck like '%LEXIK%'*, Z.8) handeln. Für die so selektierten Belege werden Lautschrift, Lemma und Bedeutung als Ergebnis ausgegeben.

Ergebnis:

TeuthoBD	Lemma	Bedeutung
dö ^ə obsd	DOERROBST	OBERBEEEGRIFF FUER GETROCKNETES OBST
hadâosgəglɥgsd	AUSGLUCKSEN	ZU ENDE GAEREN
d̄is̄iwəne	DIE SIEBENE	GRENZSTEINSETZEN, OFFIZIELL
r̄äɐv	REIF	METALLRING, DER ALLES AM RAD ZUSAMMENHAELT
kholrābi	KOHLRABI	
šluxd ^h	SCHLUCHT	GANZ KLEINES TAL, WAGENBREITE
šdü ^{er} ds	STURZ	SCHRAEGSTELLUNG VOM RAD

Tabelle 4.24: Ergebnisse der Abfrage 3 "Spontanbelege mit Zweckangabe *Lexik*"**d) Beispielabfrage 4: Unterbelege im Erhebungsort Hammelburg**

Abfrage: Es sollen alle Unterbelege (Belege in TEUTHOBD, Lemma) des SUF-Erhebungsorts Hammelburg ermittelt werden. Zusätzlich sollen Frage- und Belegnummer des Ausgangsbelegs mit angegeben werden.

SQL:

```

1  select teuthobd, lemma, frage_quelle, beleg_quelle
2  from belege, verweise

```

```

3     where ortssigle = (
4         select sigle
5         from orte
6         where name = 'Hammelburg'
7     ) and frage_nr like '%.U%'
8     and sigle_ziel = ortssigle
9     and frage_nr = frage_ziel

```

Kommentar: Aus der Tabelle *Belege* werden diejenigen Zeilen ausgewählt, deren Wert im Feld *Ortssigle* der Ortssigle des SUF-Erhebungsortes Hammelburg aus der Tabelle *Orte* entspricht (Z.3 – 7). Bei den Belegen muss es sich außerdem um Unterbelege (*frage_nr like '%.U%'*, Z.7) handeln. Zur Ermittlung der Frage- und Belegnummer des Ausgangsbelegs muss der Wert im Feld *Frage_ziel* der Tabelle *Verweise* mit der Fragenummer der jeweils ermittelten Belege übereinstimmen (*frage_nr = frage_ziel*, Z.9), der Wert im Feld *Sigle_ziel* muss mit der Ortssigle der jeweils ermittelten Belege übereinstimmen (*sigle_ziel = ortssigle*, Z.8). Als Ergebnis werden Lautschrift, Lemma, Frage- und Belegnummer des Ausgangsbelegs ausgegeben.

Ergebnis:

TeuthoBD	Lemma	Frage_Quelle	Beleg_Quelle
nünle	?	106.011	002
vald	FELD	132.015	002
šdâuxdrôug ^h	STRAUCHTROG	024.005	002
gagş	GACKEL	400.005	003
gšmûesd	GESCHMUST	046.001	002
gêdrôüg†d	GETROCKELT	136.004	001

Tabelle 4.25: Ergebnisse der Abfrage 4 "Unterbelege im Erhebungsort Hammelburg"

5. Auszeichnung der Fragen unter sprachwissenschaftlichen Gesichtspunkten

Die Auszeichnung von Fragen unter sprachwissenschaftlichen Gesichtspunkten ermöglicht es, innerhalb der BAYDAT-Datenbank gezielt sprachwissenschaftliche Phänomene abzufragen. Anhand der Auszeichnungen lassen sich z.B. sämtliche Fragen ermitteln, die Flexionsformen eines bestimmten Verbs abfragen, beispielsweise alle Fragen, die flektierte Formen des Verbs *kaufen* erheben. Umgekehrt lassen sich Fragen finden, die Verben in einer bestimmten Flexionsform enthalten, z.B. alle Fragen, die Verben im Futur I erheben.

Als Basis der Auszeichnungen diene die BSA-Fragekonkordanz. Aus dieser wurden die Fragen ausgewählt, die für ein Phänomen, z.B. eine bestimmte Wortart, relevant erschienen. Dies waren insbesondere obligatorische Fragen zu einem entsprechenden Phänomen. Teilweise wurden auch nicht-obligatorische Fragen ausgewählt, wenn sie auf ein bestimmtes Phänomen abzielten, Frage 4369 "Was kommt oben auf den Topf?" erhebt z.B. ein Substantiv und wurde daher für die Auszeichnung von Substantiven ausgewählt. Waren bei nicht obligatorischen Fragen Suggestivformen angegeben, so wurden auch diese für die Auszeichnung mit ausgewählt, sie dienen in den jeweiligen Tabellen als Grundformen (vgl. 5.2). Während bei obligatorischen Fragen davon auszugehen ist, dass gegebene Antworten das abgefragte Phänomen zeigen, z.B. die dialektale Realisierung eines mittelhochdeutschen Bezugslautes, ist dies bei nicht obligatorischen Fragen nicht immer anzunehmen. Bei Frage 4369 "Was kommt oben auf den Topf?" ist zwar davon auszugehen, dass es sich bei den gegebenen Antworten um ein Substantiv im Nominativ Singular handelt, es ist jedoch nicht zwangsläufig davon auszugehen, dass die Gewährspersonen mit "Deckel" antworteten und die Antworten dadurch mit den Antworten auf Frage 3257 "Einen neuen Deckel", die obligatorisch den Akkusativ Singular des Substantivs *Deckel* erhebt, korrespondieren. Allerdings können nicht obligatorische Fragen von Abfragen ausgeschlossen werden, ebenso können Abfragen durch Beschränkungen auf einzelne Frageintentionen (Lautung, Morphologie usw.) genauer eingegrenzt werden.

Generell sollen die Auszeichnungen möglichst theorieneutral gehalten werden. Damit wird einerseits gewährleistet, dass die Auszeichnungen und deren Umsetzung in der BAYDAT-Datenbank einem möglichst großen Benutzerkreis offen stehen. Andererseits können durch spezifische Abfragen der Auszeichnungen bzw. der Datenbank sprachwissenschaftliche Theorien gewissermaßen "simuliert" werden. So können z.B. Fragen, die Verben in transitiver Verwendung erheben, ermittelt werden, indem durch Kreuzabfragen mehrerer Tabellen ermittelt wird, welche Fragen gleichzeitig finite Vollverben bzw. Kombinationen aus finiten

Hilfs-/Modalverben und infiniten Formen eines Vollverbs und Substantive im Akkusativ erheben (vgl. 5.2.2.3., Beispielabfrage 1).

5.1. Auszeichnung lautlicher Phänomene

5.1.1. Vokale

5.1.1.1. Generelles zur Auszeichnung der Vokale

Die Auszeichnung der Vokale beruht weitestgehend auf der von Werner König zusammengestellten MITTELHOCHDEUTSCHEN WORTLISTE²². Diese listet – nach Einzellauten gruppiert – die für den Vokalismus relevanten Fragen aus dem Fragebuch des SÜDWESTDEUTSCHEN SPRACHATLAS, auf dessen Basis das Fragebuch des SBS erarbeitet wurde, auf und gibt zu jeder aufgeführten Frage weitere Informationen, z.B. das mhd. Bezugswort. Als weitere Grundlage für die Auszeichnung dienten die für die Erstellung der Lautbände des SUF angefertigten Probekarten. Als Basis diente in beiden Fällen das standardisierte Mittelhochdeutsche. Anhand der Auszeichnung lassen sich damit Entwicklungen von mittelhochdeutscher Zeit hin zur Gegenwart beobachten, z.B. die Durchführung oder das Unterbleiben der frühneuhochdeutschen Monophthongierung bzw. Diphthongierung. Außerdem bietet die Auszeichnung auf Basis des standardisierten Mittelhochdeutschen den Vorteil, dass lautliche Entwicklungen nicht falsch interpretiert werden: Die lautliche Entwicklung der frühneuhochdeutschen Diphthongierung verlief z.B. vom mhd. Monophthong zum nhd. Diphthong, nicht umgekehrt.

Außerdem beschränkt sich die Auszeichnung auf Vokale in starktonigen Silben. Sie folgt damit ebenfalls Königs MITTELHOCHDEUTSCHER WORTLISTE und weitestgehend den im SUF erstellten Probekarten. Zur Untersuchung des nicht explizit ausgezeichneten Nebensilbenvokalismus können jedoch z.B. die innerhalb der Wortbildung ausgezeichneten Präfixableitungen zurate gezogen werden.

²² König, Werner, o.J.

Für die Auszeichnung wurde zunächst die BSA-Fragennummer der jeweiligen Frage notiert, darauf folgt der jeweilige mhd. Bezugsvokal in folgender Schreibung:

Kurzmonophthonge	Langmonophthonge	Diphthonge
a	â	ei1
e	ê	ei2
e:	î	ie
i	ô	ou
o	û	uo
u	ae	öu
ä	oe	üe
ö	iu/iü	
ü		

Abbildung 5.1: Schreibung der ausgezeichneten Vokale in der BAYDAT-Datenbank

Da die für die Auszeichnung verwendeten Zeichen in Hinblick auf die Übernahme in die BAYDAT-Datenbank auf den ASCII-Zeichensatz inklusive der Erweiterung für westeuropäische Alphabete (=ISO 8859-1) beschränkt bleiben, müssen Ligaturen sowie Diakritika innerhalb der Auszeichnung durch alternative Schreibungen ersetzt werden: Statt eines Längestrichs (z.B. \bar{a}) wird zur Kennzeichnung von Langmonophthongen der Zirkumflex verwendet (z.B. \hat{a}). Das auf Primärumlaut zurückzuführende Kurz-*e*, das in der Literatur häufig als ϵ dargestellt wird²³, wird nicht gesondert gekennzeichnet. Das auf das Germanische zurückgehende *e*, das in der Literatur oft als \ddot{e} dargestellt wird²⁴, wird als *e:* ausgezeichnet. Ligaturen, z.B. \ae zur Darstellung des umgelauteten Langmonophthongs²⁵, werden aufgelöst, für \ae wird also *oe* notiert.

Zusätzlich wird nach König in der Auszeichnung zwischen *ei*, das "durch den Schwund von inlautendem *g* in der ahd. Lautfolge *egi*"²⁶ entstand, und althochdeutschem *ei* unterschieden: Für ersteres wird *ei2*, für letzteres *ei1* notiert. Auch wird zwischen monophthongiertem ahd. *iu*, das als *iu* ausgezeichnet wird, und umgelautetem Langmonophthong , notiert als *iü*, unterschieden²⁷. Durch die explizite Kennzeichnung lässt sich die unterschiedliche Entwicklung der ursprünglichen Laute innerhalb des BSA-Erhebungsgebietes nachvollziehen.

Zu jedem ausgezeichneten Vokal wird des Weiteren dessen Art (*Lang-*, *Kurzmonophthong* oder *Diphthong*) notiert. Für Kurzmonophthonge wird nach König die Art der Silbe (*offen* oder *geschlossen*) notiert. Bei Langmonophthongen und Diphthongen unterbleibt die Auszeichnung der Silbenart, allerdings wird notiert, ob diese im Hiatus stehen.

²³ vgl. u.a. Paul/Wiehl/Grosse, S.90, Naumann, S.199

²⁴ vgl. Berner, S.250

²⁵ vgl. Paul/Wiehl/Grosse, S.29

²⁶ Wiesinger, S.91

²⁷ vgl. Renn/König, S.52 ff.

Für jeden Vokal wird das gesamte Vor- und Nachfeld im entsprechenden mhd. Bezugswort angegeben, für den Langvokal *â* im Bezugswort *straze*, abgefragt durch Frage 1046 der BSA-Fragekonkordanz "'Straße' (einen breiten, geteerten Weg nennt man eine...)" wird z.B. "str" als Vorfeld, "ze" als Nachfeld notiert. Durch die Auszeichnung des Vor- und Nachfeldes kann z.B. direkt auf Vokale vor bestimmten Endungen zugegriffen werden. Außerdem wird die Art des unmittelbar vorausgehenden sowie die Art des unmittelbar folgenden Lautes angegeben (*Plosiv, Frikativ, Nasal, Liquid, Affrikata, Vokal* oder *Halbvokal*), für genanntes Beispiel *straze* wird also für die Art des vorausgehenden Lautes *Liquid*, für die Art des Folgelautes *Frikativ* angegeben.

Für jeden ausgezeichneten Vokal werden schließlich dessen absolute Position innerhalb des mittelhochdeutschen Bezugswortes (*Anlaut, Inlaut, Auslaut*) und das bei König genannte mhd. Bezugswort notiert.

5.1.1.2. Datenbanktabellen zur Auszeichnung der Vokale

Die Auszeichnungen der Vokale finden sich innerhalb der BAYDAT-Datenbank in der Tabelle *Link_Vokale*:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer der Auszeichnung
Nr_Laut	Char(4)	laufende Nummer des ausgezeichneten Lautes
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Laut	Varchar2(5)	ausgezeichneter mhd. Bezugslaut
Art	Char(1)	Art des ausgezeichneten Vokals (Monophthong, Diphthong)
Silbe	Char(1)	Art der Silbe (bei Kurzvokalen)/Stellung im Hiatus (bei Langvokalen/Diphthongen)
Vorfeld	Varchar2(20)	Vorfeld des ausgezeichneten Vokals
Art_Vor	Char(1)	Art des vorausgehenden Lautes
Nachfeld	Varchar2(20)	Nachfeld des ausgezeichneten Vokals
Art_Nach	Char(1)	Art des nachfolgenden Lautes
Position	Char(1)	Position des ausgezeichneten Lautes innerhalb des Bezugsworts
Bezugswort	Varchar2(500)	Mhd. Bezugswort
Sugg_Form	Char(1)	ausgezeichneter Laut innerhalb einer Suggestierform oder nicht?

Tabelle 5.1: Datenfelder und -typen der Tabelle *Link_Vokale*

Zusätzlich zu den oben beschriebenen Auszeichnungen finden sich in der Tabelle im Feld *Nr* eine laufende Nummer der Auszeichnung sowie eine vierstellige laufende Nummer des ausgezeichneten Vokals im Feld *Nr_Laut*, die als Basis für die Verknüpfung der

ausgezeichneten Vokale innerhalb der grafischen Benutzeroberfläche der BAYDAT-Datenbank dient. Außerdem enthält die Tabelle ein Feld *Sugg_Form*, das anzeigt, ob der ausgezeichnete Vokal innerhalb einer Suggestierform oder innerhalb einer regulären Frage ausgezeichnet wurde.

Zur Redundanzvermeidung und aus Gründen der leichteren Sortierbarkeit werden die Werte für die Felder *Art*, *Silbe*, *Art_Vor*, *Art_Nach* und *Position* in folgender kodierter Form in die Tabelle *Link_Vokale* übernommen²⁸:

Feld	Kodierung	Auszeichnung
Art	1	Monophthong (kurz)
	2	Monophthong (lang)
	3	Diphthong
Silbe	1	offen
	2	geschlossen
	3	Hiat
Art_Vor/Art_Nach	1	Plosiv
	2	Frikativ
	3	Nasal
	4	Liquid
	5	Affrikata
	6	Vokal
	7	Halbvokal
Position	1	Anlaut
	2	Inlaut
	3	Auslaut

Tabelle 5.2: Kodierte Werte in der Tabelle *Link_Vokale*

5.1.1.3. Beziehungen der Tabelle *Link_Vokale* zu anderen Tabellen der BAYDAT-Datenbank

Die Tabelle *Link_Vokale* steht über das Feld *Fragenr* in Verbindung zur Tabelle *Fragen*:

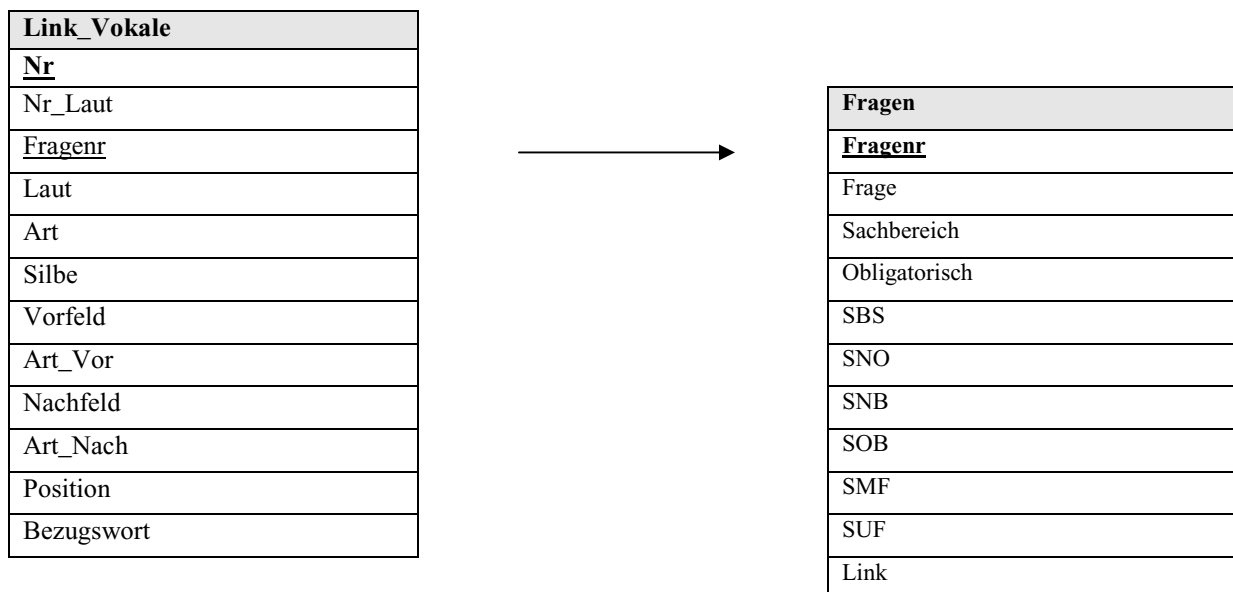


Abbildung 5.2: Schematische Darstellung der Beziehungen der Tabelle *Link_Vokale*

²⁸ für Details bezüglich Redundanzvermeidung und Sortierung vgl. 5.2.1.2.

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

5.1.1.4. Beispielabfragen

a) Beispielabfrage 1: Mhd. *ö* in offener Silbe vor stimmhaftem Plosiv

Abfrage: Es sollen alle Fragen (SUF-Fragennummer, Fragetext, mhd. Bezugswort) ermittelt werden, die im SUF obligatorisch mhd. *ö* in offener Silbe vor stimmhaftem Plosivlaut erheben.

SQL:

```

1      select suf, frage, bezugswort
2      from fragen, link_vokale
3      where fragen.fragenr = link_vokale.fragenr
4      and laut = 'ö'
5      and silbe = '1'
6      and (
7          nachfeld like 'b%'
8          or nachfeld like 'd%'
9          or nachfeld like 'g%'
10     ) and suf is not null
11     and obligatorisch = 'j'
```

Kommentar: Durch eine Kreuzabfrage der Tabellen *Fragen* und *Link_Vokale* werden die Fragennummern derjenigen Fragen ermittelt, die mhd. *ö* (*laut = 'ö', Z.4*) in offener Silbe (*silbe = '1', Z.5*) und vor stimmhaftem Plosiv (*nachfeld like 'b%' or nachfeld like 'd%' or nachfeld like 'g%', Z.7-9*) erheben. Werden die Fragen im SUF (*suf is not null, Z.10*) obligatorisch (*obligatorisch = 'j', Z.11*) erhoben, werden SUF-Fragennummer, Fragetext und mhd. Bezugswort als Ergebnis ausgegeben. Die Bedingung *fragen.fragenr = link_vokale.fragenr* (Z.3) stellt sicher, dass nur SUF-Fragennummer und Fragetexte der ermittelten Fragen ausgegeben werden. Fehlte die entsprechende Zeile, würden sämtliche SUF-Fragennummern und Fragetexte als Ergebnis ausgegeben.²⁹

Ergebnis:

SUF	Frage	Bezugswort
264.009	"Vögel"	vögel
264.010	"Vögelein"	vögellîn
328.001	"Bügeln"	bögelen
328.002	"Gebügelt"	gebögelet
388.014	"Böden"	böden
414.003	"Wir mögen es schon"	mögen
414.004	"Mögt ihr es?"	möget

Tabelle 5.3: Ergebnisse der Abfrage "Mhd. *ö* in offener Silbe vor stimmhaftem Plosiv"

²⁹ Die Abfrage hätte das kartesische Produkt aus den Tabellen *Fragen* und *Link_Verben* zum Ergebnis, vgl. Connolly/Begg, S.93 f.

b) Beispielabfrage 2: Mhd. *ie* vor Liquid

Abfrage: Es sollen alle obligatorischen Fragen (SBS-Fragennummer, Fragetext) ermittelt werden, die im SBS mhd. *ie* vor Liquid erheben.

SQL:

```

1  select sbs, frage
2  from fragen
3  where fragenr in (
4      select fragenr
5      from link_vokale
6      where laut = 'ie'
7      and art_nach = '4'
8  ) and sbs is not null
9  and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Vokale* werden die Fragennummern der Fragen ausgewählt, die mhd. *ie* (*laut* = 'ie', Z.6) vor Liquid (*art_nach* = '4', Z.7) erheben. Werden die entsprechenden Fragen im SBS (*sbs is not null*, Z.8) obligatorisch (*obligatorisch* = 'j', Z.9) erhoben, werden SBS-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SBS	Frage
272.003	"Es friert mich"
272.004	"(ge-)frieren"
424.001	"Verlieren" (Gegenteil von gewinnen)
424.003	"Verlier es nicht!"

Tabelle 5.4: Ergebnisse der Abfrage "Mhd. *ie* vor Liquid"

c) Beispielabfrage 3: Langmonophthonge im Hiatt ohne Hiatt-Tilger

Abfrage: Es sollen Fragen (SMF-Fragennummer, Fragetext, mhd. Bezugswort) gefunden werden, die im SMF obligatorisch Langmonophthonge im Hiatt erheben. Dabei soll das mhd. Bezugswort keinen Hiatt-Tilger enthalten.

SQL:

```

1  select smf, frage, bezugswort
2  from fragen, link_vokale
3  where fragen.fragenr = link_vokale.fragenr
4  and silbe = '3'
5  and art_nach <> '7'
6  and art = '2'
7  and smf is not null
8  and obligatorisch = 'j'
```

Kommentar: Durch eine Kreuzabfrage der Tabellen *Fragen* und *Link_Vokale* werden die Fragennummern derjenigen Fragen ermittelt, die Langmonophthonge (*art* = '2', Z.6) im Hiatt (*silbe* = '3', Z.4) erheben. Durch die Bedingung *art_nach* <> '7' (Z.5) wird ausgeschlossen, dass auf den Langvokal ein Halbvokal als Hiatt-Tilger folgt. Werden die ermittelten Fragen im SMF (*smf is not null*, Z.7) obligatorisch (*obligatorisch* = 'j', Z.8) erhoben, werden SMF-Fragennummer, Fragetext und mhd. Bezugswort als Ergebnis ausgegeben. Die Bedingung in

Z.3 verhindert wie in Beispielabfrage 1 die Bildung des karthesischen Produkts aus den abgefragten Tabellen *Fragen* und *Link_Vokale*.

Ergebnis:

SMF	Frage	Bezugswort
102.006	"Schreien"	schrîen
102.007	"Geschrien"	schrîen
141.008	"Es fängt an zu schneien"	snîen

Tabelle 5.5: Ergebnisse der Abfrage "Langmonophthonge im Hiatt ohne Hiatt-Tilger"

5.1.2. Konsonanten

5.1.2.1. Generelles zur Auszeichnung der Konsonanten

Die Auszeichnung der Konsonanten orientierte sich an den Probekarten, die für die Erstellung des Teilbands LAUTGEOGRAPHIE II (KONSONANTEN) des SUF angefertigt wurden. Auch hier diente, wie bei den Vokalen, der Lautstand des standardisierten Mittelhochdeutschen als Basis.

Zu jedem ausgezeichneten Konsonanten wurde die entsprechende Fragennummer der BSA-Fragekonkordanz notiert. Anschließend wurden die ausgezeichneten Konsonanten in folgender Schreibung angegeben:

Plosive	Nasale
b	m
d	n
g	
p	Liquide
t	r
k	l
Frikative	Affrikaten
ch	pf
ch2	z
f	
h	Halbvokale
s	j
s2	w
sch	
v	

Abbildung 5.3: Schreibung der ausgezeichneten Konsonanten in der BAYDAT-Datenbank

Unterschieden wird in der Auszeichnung zwischen ursprünglichem germanischem *s* (=s), z.B. im mhd. Bezugswort *mûs*, abgefragt in Frage 2429 der BSA-Fragekonkordanz, "'Maus' (Die Katze fängt eine...)" und dem aus der zweiten Lautverschiebung hervorgegangenen *s* (=s2), z.B. im mhd. Bezugswort *wazzar*, u.a. erhoben durch Frage 232 der BSA-Fragekonkordanz, "'Wasser' (Was gibt man dem Vieh zu trinken?)". Ebenso wird unterschieden zwischen *ch*, das auf das ursprüngliche germanische χ zurückgeht (=ch), z.B. im mhd. Bezugswort *kirche*, abgefragt in Frage 2316 der BSA-Fragekonkordanz, "'Kirche' (am Sonntag geht man in die...)", und dem auf die zweite Lautverschiebung zurückgehenden *ch* (=ch2), u.a. im mhd.

Bezugswort *bûch*, abgefragt durch Frage 2139 der BSA-Fragekonkordanz, "'Bauch' (der hat aber einen dicken...)"³⁰

Zu jedem ausgezeichneten Konsonanten wird zusätzlich dessen Art (*Plosiv*, *Frikativ*, *Liquid*, *Nasal*, *Affrikata* oder *Halbvokal*) notiert. Ebenso wie bei der Auszeichnung der Vokale wird das Vor- und Nachfeld des jeweiligen Konsonanten sowie die Art des unmittelbar vorausgehenden bzw. nachfolgenden Lautes notiert. Abschließend werden für jeden Konsonanten dessen absolute Position im Wortanlaut, -inlaut oder -auslaut sowie das mhd. Bezugswort angegeben.

5.1.2.2. Datenbanktabellen zur Auszeichnung der Konsonanten

Die Auszeichnungen der Konsonanten werden in die Tabelle *Link_Konsonanten* der BAYDAT-Datenbank übernommen:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer der Auszeichnung
Nr_Laut	Char(4)	laufende Nummer des ausgezeichneten Lautes
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Laut	Varchar2(5)	ausgezeichneter mhd. Bezugslaut
Art	Char(1)	Art des ausgezeichneten Konsonanten
Vorfeld	Varchar2(20)	Vorfeld des ausgezeichneten Konsonanten
Art_Vor	Char(1)	Art des vorausgehenden Lautes
Nachfeld	Varchar2(20)	Nachfeld des ausgezeichneten Konsonanten
Art_Nach	Char(1)	Art des nachfolgenden Lautes
Position	Char(1)	Position des ausgezeichneten Lautes innerhalb des Bezugsworts
Bezugswort	Varchar2(500)	Mhd. Bezugswort
Sugg_Form	Char(1)	Ausgezeichneter Laut innerhalb einer Suggestierform oder nicht?

Tabelle 5.6: Datenfelder und -typen der Tabelle *Link_Konsonanten*

Ebenso wie die Tabelle *Link_Vokale* enthält die Tabelle *Link_Konsonanten* ein Feld für die laufende Nummer der Verknüpfung (*Nr*) sowie ein Feld für eine laufende vierstellige Nummer des ausgezeichneten Konsonanten (*Nr_Laut*), die für die Verknüpfung der Auszeichnungen innerhalb der grafischen Benutzeroberfläche der BAYDAT-Datenbank genutzt wird. Die Tabelle enthält ebenfalls ein Feld *Sugg_Form*, um anzuzeigen, ob der betreffende Konsonant innerhalb einer Suggestierform oder innerhalb einer regulären Frage ausgezeichnet wurde.

³⁰ vgl. Naumann, S.205

Aus Gründen der Redundanzvermeidung und zur besseren Sortierbarkeit sind die Werte in den Feldern *Art*, *Art_Vor*, *Art_Nach* und *Position* in folgender Kodierung in die Tabelle *Link_Konsonanten* aufgenommen worden:

Feld	Kodierung	Auszeichnung
Art	1	Plosiv
	2	Frikativ
	3	Nasal
	4	Liquid
	5	Affrikata
	7	Halbvokal
	Art_Vor/Art_Nach	1
2		Frikativ
3		Nasal
4		Liquid
5		Affrikata
6		Vokal
7		Halbvokal
Position	1	Anlaut
	2	Inlaut
	3	Auslaut

Tabelle 5.7: Kodierte Werte in der Tabelle *Link_Konsonanten*

Unter den Kodierungen im Feld *Art* fehlt die Ziffer 6, die Vokale kennzeichnet. So bleiben die Kodierungen innerhalb der Felder *Art*, *Art_Vor* und *Art_Nach* untereinander konsistent.

5.1.2.3. Beziehungen der Tabelle *Link_Konsonanten* zu anderen Tabellen der BAYDAT-Datenbank

Die Tabelle *Link_Konsonanten* steht über das Feld *Fragenr* in Verbindung zur Tabelle *Fragen* der BAYDAT-Datenbank:

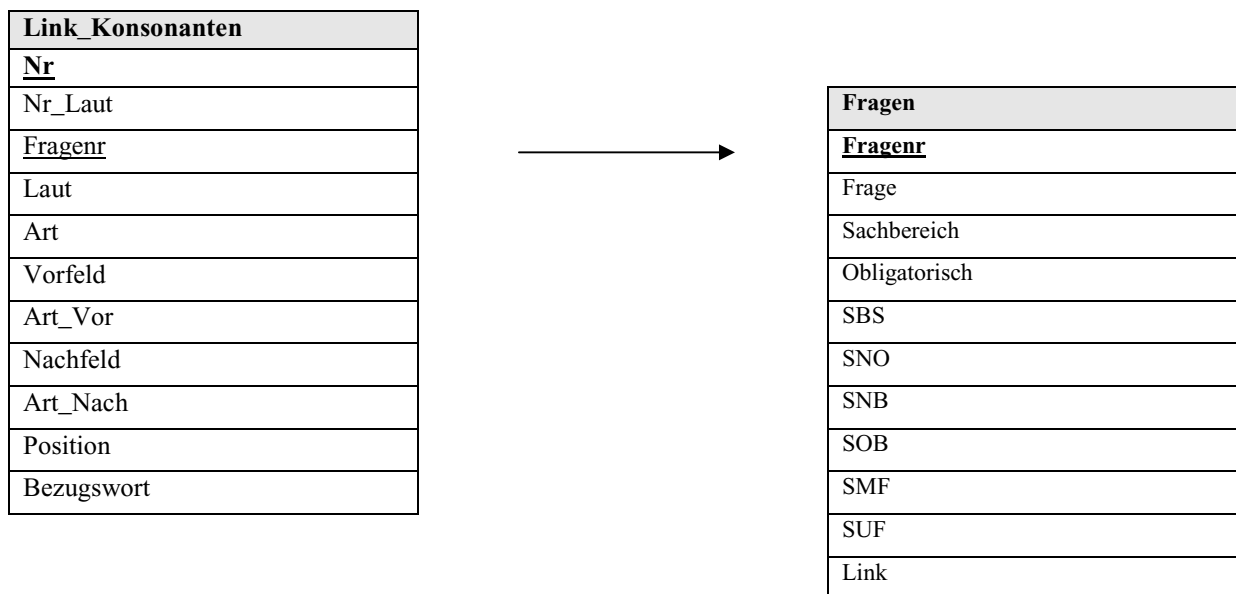


Abbildung 5.4: Schematische Darstellung der Beziehungen der Tabelle *Link_Konsonanten*

Fremdschlüssel sind in der Übersicht durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

5.1.2.4. Beispielabfragen

a) Beispielabfrage 1: Binnendeutsche Konsonantenschwächung

Abfrage: Durch die binnendeutsche Konsonantenschwächung werden aspirierte /p^h/ und /k^h/ im Anlaut "vor den Konsonanten /l, n, r, w/ zur stl. Lenis oder Halbfortis."³¹ Um die Auswirkung der binnendeutschen Konsonantenschwächung für Unterfranken zu überprüfen, sollen Fragen (SUF-Fragennummer und Fragetext) ermittelt werden, die im SUF obligatorisch *p* und *k* im Anlaut vor den Konsonanten *l, r, n, w* erheben.

SQL:

```

1   select suf, frage
2   from fragen
3   where fragenr in (
4       select fragenr from link_konsonanten
5       where (
6           laut like 'p%'
7           or laut like 'k%'
8       ) and (
9           nachfeld like 'l%'
10          or nachfeld like 'r%'
11          or nachfeld like 'n%'
12          or nachfeld like 'w%'
13       ) and position = '1'
14   ) and suf is not null
15   and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Konsonanten* werden die Fragennummern derjenigen Fragen ausgewählt, die *p* und *k* (*laut like 'p%' or laut like 'k%'*, Z.6 f.) vor *l, r, n, w* (*nachfeld like 'l%' or nachfeld like 'r%' or nachfeld like 'n%' or nachfeld like 'w%'*, Z.9 ff.) im Anlaut (*position = '1'*, Z.13) erheben. Werden die entsprechenden Fragen im SUF (*suf is not null*, Z.14) obligatorisch (*obligatorisch = 'j'*, Z.15) erhoben, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SUF	Frage
018.001	"Klauen" Plural
018.002	"Klaue" Singular (was man beschneiden muß)
080.001	"Pflug" (womit wird die Ackererde gewendet?)
080.002	"Pflüge"
082.008	"Klee"
172.005	"Pflaume"
172.006	"Pflaumen"
206.009	"Kropf"

³¹ Paul/Wiehl/Grosse, S.130

206.010	"Kröpfe"
214.001	"Knie"
214.002	"Knien" (in der Kirche muss man während der Wandlung...)
218.006	"Er hat's im Kreuz"
232.010	"Knecht" (Wer hilft dem Bauern bei der Arbeit?) (Rangordnung fixieren, wenn vorhanden)
254.005	"Krone" (Was hat der König auf dem Kopf?)
318.002	"Ein seidenes/seidiges Kleid"
354.008	"Kneten" (Wort nicht vorsagen?)

Tabelle 5.8: Ergebnisse der Abfrage "Binnendeutsche Konsonantenschwächung"

b) Beispielabfrage 2: Intervokalisches *h*

Abfrage: Es sollen Fragen (SNOB-Fragennummer, Fragetext) gefunden werden, die im SNOB obligatorisch intervokalisches *h* erheben.

SQL:

```

1   select sno, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_konsonanten
6       where laut = 'h'
7       and art_vor = '6'
8       and art_nach = '6'
9   ) and sno is not null
10  and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Konsonanten* werden die Fragennummern derjenigen Fragen ausgewählt, die intervokalisches (*art_vor = '6' and art_nach = '6', Z.7 f.) h* (*laut = 'h', Z.6*) erheben. Wurden die entsprechenden Fragen im SNOB (*sno is not null, Z.9*) obligatorisch (*obligatorisch = 'j', Z.10*) erhoben, werden SNOB-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SNO	Frage
090.002	"Ähre"
090.003	"Ähren"
114.001	"Ziehen" (wenn man den Wagen vorwärtsbewegen will, muss man hinten schieben und vorne...)
114.003	"Ich ziehe"
114.005	"Du ziehst"
114.007	"Er zieht"
114.008	"Wir ziehen"
114.009	"Ihr zieht"
114.010	"Sie ziehen"
140.011	"Ein hoher Baum"
216.001	"Zehe" (Genus)
216.002	"Zehen"
246.010	"Von einem hohen Turm (sieht man viel)"

Tabelle 5.9: Ergebnisse der Abfrage "Intervokalisches *h*"

b) Beispielabfrage 3: Mhd. *m* nach *e* im Auslaut

Abfrage: Es sollen Fragen (SNIB-Fragennummer, Fragetext, mhd. Bezugswort) gefunden werden, die im SNIB obligatorisch mhd. *m* nach *e* im Auslaut erheben.

SQL:

```

1  select snb, frage, bezugswort
2  from fragen, link_konsonanten
3  where fragen.fragenr = link_konsonanten.fragenr
4  and laut = 'm'
5  and vorfeld like '%e'
6  and position = '3'
7  and snb is not null
8  and obligatorisch = 'j'
```

Kommentar: Durch eine Kreuzabfrage der Tabellen *Fragen* und *Link_Konsonanten* werden Fragen ermittelt, die mittelhochdeutsch *m* (*laut = 'm'*, Z.4) nach *e* (*vorfeld like '%e'*, Z.5) im Auslaut (*position = '3'*, Z.6) enthalten. Wurden diese im SNIB (*snb is not null*, Z.7) obligatorisch (*obligatorisch = 'j'*, Z.8) erhoben, werden SNIB-Fragennummer, Fragetext und mhd. Bezugswort als Ergebnis ausgegeben. Die Bedingung in Z.3 verhindert die Bildung eines karthesischen Produkts aus den abgefragten Tabellen *Fragen* und *Link_Konsonanten*.

Ergebnis:

SNB	Frage	Bezugswort
160.004	"Einen Stecken in den Boden stecken"	bodem
316.006	"Faden"	vadem
388.012	"Boden" (in der Stube)	bodem

Tabelle 5.10: Ergebnisse der Abfrage "Mhd. *m* nach *e* im Auslaut"

5.2. Auszeichnung morphologischer Phänomene

Für die morphologische Auszeichnung der einzelnen Wortarten wurden zunächst die Fragen aus der BSA-Fragekonkordanz extrahiert, die Vertreter der entsprechenden Wortart als Einzelwort oder innerhalb eines Satzes obligatorisch erhoben bzw. die als Antwort einen Vertreter der entsprechenden Wortart erwarten ließen. Für die Auszeichnung von Verben wurden z.B. die obligatorische Frage 1047 "Da bauen sie wieder eine neue Straße." oder die nicht obligatorische Frage 1502 "Einen schweren Korb kann man nicht ...", die als Antwort ein Verb erwartet, extrahiert. Anschließend wurde zu jeder Frage die Grund- bzw. Nennform des für die Auszeichnung relevanten Wortes notiert, bei Verben z.B. der Infinitiv. Für Frage 1047 wurde also die Grundform *bauen* notiert, für Frage 1502 wurden die gegebenen Suggestierformen *lupfen* und *aufhucken* notiert. Bei obligatorischen Fragen wurde des Weiteren die in der Frage erhobene Form des jeweiligen Wortes notiert, für Frage 1047 z.B. *bauen*.

Im Anschluss wurden die erhobenen Formen weiter ausgezeichnet, bei Verben wurden u.a. Person, Numerus und Modus angegeben, für die Form *bauen* in Frage 1047 "3. Person Plural Indikativ". Ab diesem Punkt unterscheiden sich die Auszeichnungen der einzelnen Wortarten. Die spezifischen Auszeichnungen der einzelnen Wortarten werden in den folgenden Unterkapiteln näher beschrieben.

5.2.1. Verben

5.2.1.1. Generelles zur Auszeichnung der Verben

Zu sämtlichen Verben wurde zunächst eine Unterkategorie notiert. Als Unterkategorien werden für die Auszeichnung *Vollverb*, *Hilfsverb*, *Modalverb* und *Modalitätsverb* angenommen.

Zusätzlich wurde die Klasse jedes Verbs notiert. Als Klassen werden die beiden Hauptklassen³² der starken und schwachen Verben angenommen, gesondert gekennzeichnet werden Präterito-Präsentien (die im BSA erhobenen Präterito-Präsentien *dürfen*, *gönnen*, *können*, *mögen*, *müssen*, *sollen* und *wissen*) und Wurzelverben (*gehen*, *sein*, *stehen* und *tun*). Die Auszeichnung der Klassen basiert auf den Klassen des standardisierten Mittelhochdeutschen. Damit wird u.a. das Verb *bellen* (erhoben in den Fragen 543 "Bellen", 544 "Er bellt" und 545 "Gebellt"), das gegenwartssprachlich schwach flektiert³³, als starkes Verb ausgezeichnet. Die Auszeichnung nach dem standardisierten Mittelhochdeutschen bietet gegenüber der gegenwartssprachlichen Auszeichnung den Vorteil, dass sich Entwicklungen innerhalb der Flexion – eben vom mittelhochdeutsch stark flektierten zum gegenwartssprachlich schwach flektierten *bellen* – nachvollziehen lassen. Auch können damit falsche Schlussfolgerungen von Laienbenutzern der BAYDAT-Datenbank ausgeschlossen werden, die diese aufgrund der Beleglage ziehen könnten: Die Entwicklung ging im Beispiel *bellen* nicht von der schwachen zur starken Flexion, sondern umgekehrt. Problematisch erweist sich die Auszeichnung der Klasse beim Verb *wollen*, dessen mhd. Präsensindikativformen auf ehemaligen Konjunktivformen beruhen. Auf Basis der neuen Präsensformen wurden wiederum neue Konjunktiv- und (schwache) Präteritumformen gebildet. Damit steht das Verb *wollen* in der Nähe zu Präterito-Präsentien³⁴, deren Präsensformen auf die ursprünglichen Präteritumformen zurückzuführen sind³⁵. Da das Verb *wollen* allein dieses Phänomen zeigt, wird es nicht einer eigenen Klasse zugeordnet, sondern

³² vgl. Paul/Wiehl/Grosse, S.238

³³ vgl. Duden. Deutsches Universalwörterbuch, S.259, s.v. "bellen"

³⁴ vgl. Paul/Wiehl/Grosse, S.265

³⁵ Paul/Wiehl/Grosse, S.260

als schwaches Verb ausgezeichnet. Kontrahierte Verben³⁶ werden nicht genauer gekennzeichnet, sondern nach ihrer Flexion als starke oder schwache Verben markiert.

Zu starken Verben und Präterito-Präsentien wird weiterhin die Ablautreihe angegeben. Mögliche Werte für die Ablautreihen der starken Verben sind dabei Paul/Wiehl/Grosse³⁷ entsprechend *1a*, *1b*, *2a*, *2b*, *3a*, *3b*, *4*, *5*, *6* und *7*, Präterito-Präsentien werden den Ablautreihen 1 – 6 zugeordnet³⁸. Da für die Kennzeichnung der Klassen und Ablautreihen das System des Mittelhochdeutschen zugrunde gelegt wird, werden die ahd. Klassen der schwachen Verben nicht angegeben, da diese "im Mhd. nur noch wenig zur Geltung"³⁹ kommen.

Die weiteren Auszeichnungen unterscheiden sich für finite und infinite Verbformen: Für finite Verbformen werden Person (*1*, *2* oder *3*), Numerus (*Singular* oder *Plural*), Tempus (*Präsens* oder *Präteritum*) und Modus (*Indikativ*, *Konjunktiv I*, *Konjunktiv II* oder *Imperativ*) angegeben. Wie gezeigt, werden für die Auszeichnung des Tempus nur *Präsens* oder *Präteritum* vergeben, es werden also nur synthetische Tempusformen ausgezeichnet. Dies ist notwendig, da Hilfs- und Vollverben separat ausgezeichnet werden. Analytische Tempusformen lassen sich ermitteln, indem Fragen gesucht werden, die sowohl eine flektierte Form des für die analytische Tempusform notwendigen Hilfsverbs enthalten als auch infinite Formen eines Vollverbs. Perfektformen des Verbs *sagen* lassen sich z.B. ermitteln, indem Fragen gesucht werden, die sowohl Indikativ-Präsens-Formen des Hilfsverbs *haben* als auch das Partizip II des Vollverbs *sagen* enthalten (vgl. 5.2.1.4., Beispielabfrage 1). Das Genus verbi finiter Verben wird nicht gesondert ausgezeichnet, da dieses analytisch gebildet wird. Fragen, die das Vorgangspassiv erheben, lassen sich ermitteln, indem Fragen gesucht werden, die sowohl flektierte Formen der Hilfsverben *werden* oder *bekommen* als auch ein Partizip II eines Vollverbs enthalten. Zustandspassive lassen sich hingegen schwer ermitteln: Zustandspassive mit *haben* sind mit der entsprechenden Perfektform deckungsgleich und entsprechende Fragen damit bei fehlendem Kontext mehrdeutig, z.B. in Frage 2053 der BSA-Fragekonkordanz "'Er hat (sich) den/(der) Arm gebrochen' (Akk.Mask.) (reflexiv?)". Zustandspassive mit *sein* lassen sich vom Benutzer schnell ermitteln, da *sein* in der BSA-Fragekonkordanz nur insgesamt zwanzigmal als Hilfsverb erhoben wurde. Perfektformen intransitiver Verben lassen sich aufgrund der geringen Zahl an Belegen schnell von Zustandspassiven transitiver Verben unterscheiden.

³⁶ vgl. Paul/Wiehl/Grosse, S.272 ff.

³⁷ vgl. Paul/Wiehl/Grosse, S.243 ff.

³⁸ vgl. Paul/Wiehl/Grosse, S.261 ff.

³⁹ Paul/Wiehl/Grosse, S.254

Bei infiniten Verbformen wird die Form (*Infinitiv* oder *Partizip II*) angegeben, zusätzlich wird ihre Position ausgezeichnet (*nach Hilfs-/Modal-/Vollverb, nach zu oder freistehend*).

5.2.1.2. Datenbanktabellen zur Auszeichnung der Verben

Die Auszeichnungen der Verben finden sich in der BAYDAT-Datenbank in der Tabelle *Link_Verben* wieder:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	Infinitiv des ausgezeichneten Verbs
Grundform_Nr	Varchar2(4)	laufende Nummer der Grundform
Frform	Varchar2(500)	Form des ausgezeichneten Verbs in der entsprechenden Frage
Pos	Char(1)	Position des Verbs
Pers	Char(1)	Person des Verbs
Num	Char(1)	Numerus des Verbs
Mod	Char(1)	Modus des Verbs
Temp	Char(1)	Tempus des Verbs
Art	Char(1)	Unterkategorie des Verbs
Form	Char(1)	Art der infiniten Verbform
Verbart	Varchar2(10)	Klasse des Verbs
Ablautreihe	Varchar2(10)	Ablautreihe des Verbs
Sugg_Form	Char(1)	ausgezeichnetes Verb Suggestierform oder nicht?

Tabelle 5.11: Datenfelder und -typen der Tabelle *Link_Verben*

Zusätzlich zu oben beschriebenen Auszeichnungen finden sich in der Tabelle eine laufende Nummer im Feld *Nr* – dieses Feld dient innerhalb der Tabelle als Primärschlüssel – und im Feld *Grundform_Nr* eine laufende, vierstellige Nummer der jeweiligen Grundform (Nummern, die weniger als vier Stellen breit sind, werden führende Nullen vorangestellt). Diese dient als Hilfsmittel zur Verknüpfung der einzelnen Fragen innerhalb der grafischen Benutzeroberfläche der BAYDAT-Datenbank. Die Tabelle enthält außerdem ein Feld *Sugg_Form*, um anzuzeigen, ob es sich bei dem ausgezeichneten Verb um eine Suggestierform handelt oder um ein Verb innerhalb des Fragetextes.

In der Tabelle *Link_Verben* werden die Auszeichnungen aus Gründen der Redundanzvermeidung nicht im Klartext, sondern kodiert übernommen. Hierzu boten sich zunächst Abkürzungen an, z.B. *S* für 'Singular' oder *P* für 'Plural'. Allerdings würden in diesem Fall bei einer Sortierung der Formen nach dem Numerus Plural- vor Singularformen stehen. Deshalb wurden für die Kodierung Ziffern gewählt. Folgende Tabelle fasst die möglichen Werte für die einzelnen Felder zusammen:

Feld	Kodierung	Auszeichnung
Art	1	Vollverb
	2	Hilfsverb

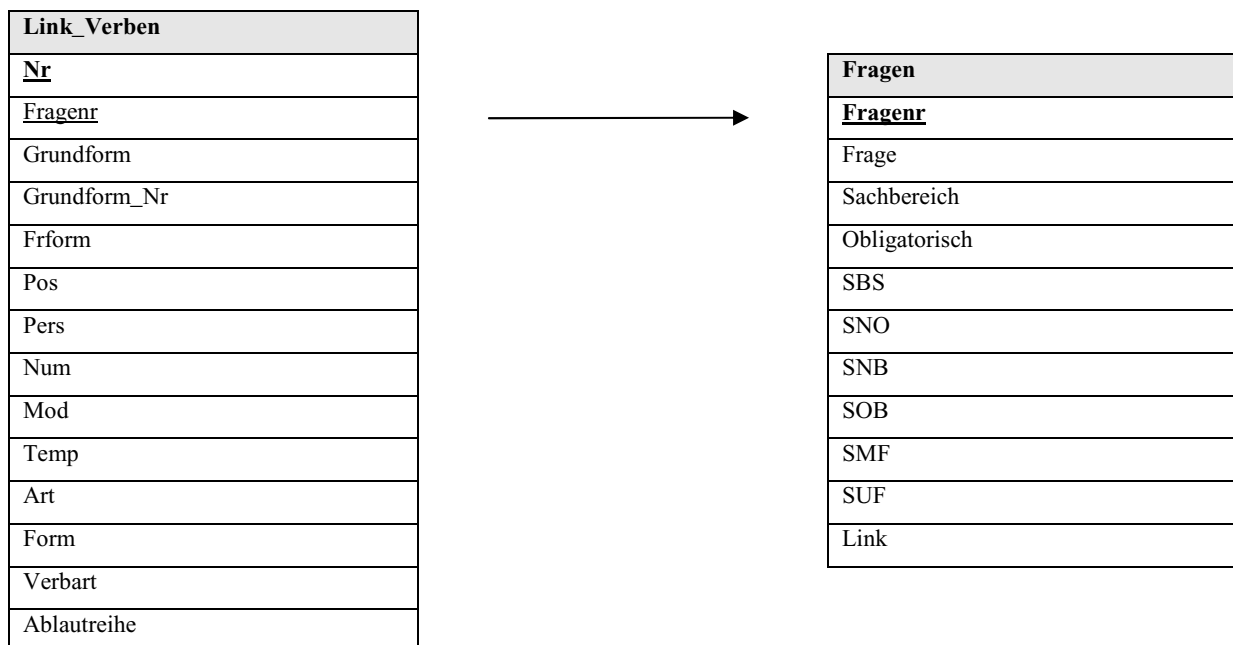
	3	Modalverb
	4	Modalitätsverb
Verbart	1	Starkes Verb
	2	Schwaches Verb
	3	Präteritopräsentium
	4	Wurzelverb
Pers	1	1. Person
	2	2. Person
	3	3. Person
Num	1	Singular
	2	Plural
Mod	1	Indikativ
	2	Konjunktiv I
	3	Konjunktiv II
	4	Imperativ
Temp	1	Präsens
	2	Präteritum
Form	1	Infinitiv
	2	Partizip II
Pos	1	freistehend
	2	nach <i>zu</i>
	3	nach Hilfsverb
	4	nach Modalverb
	5	nach Vollverb

Tabelle 5.12: Kodierte Werte in der Tabelle *Link_Verben*

Die Auszeichnungen der Ablautreihen werden nicht kodiert, mögliche Werte im entsprechenden Feld *Ablautreihe* sind *1a*, *1b*, *2a* usw.

5.2.1.3. Beziehungen der Tabelle *Link_Verben* zu anderen Tabellen der BAYDAT-Datenbank

Die Tabelle *Link_Verben* steht über das Feld *Fragenr* mit der Tabelle *Fragen* in Beziehung:

Abbildung 5.5: Schematische Darstellung der Beziehungen der Tabelle *Link_Verben*

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

Die Beziehung der Tabelle *Link_Verben* zur Tabelle *Fragen* kann als exemplarisch für alle Tabellen zur Auszeichnung morphologischer Phänomene gelten. Die Beziehungen der übrigen Tabellen zur Tabelle *Fragen* werden daher nicht genauer schematisch dargestellt.

5.2.1.4. Beispielabfragen

a) Beispielabfrage 1: Das Verb *sagen* im Perfekt

Abfrage: Es sollen alle obligatorischen Fragen (SMF-Fragennummer und Fragetext) aus dem Fragebuch des SMF gefunden werden, die das Verb *sagen* im Perfekt erheben.

SQL:

```

1      select smf, frage from fragen where fragenr in (
2          select fragenr from link_verben
3          where art = '2'
4          and mod = '1'
5          and temp = '1'
6          and grundform = 'haben'
7          and fragenr in (
8              select fragenr
9              from link_verben
10             where form = '2'
11             and grundform = 'sagen'
12         )
13     ) and obligatorisch = 'j'
14     and smf is not null
15     order by smf
```

Kommentar: Aus der Tabelle *Link_Verben* werden die Fragennummern (*fragenr*, Z.2) derjenigen Fragen ausgewählt, die das Hilfsverb (*art* = '2', Z.3) *haben*⁴⁰ im Indikativ (*mod* = '1', Z.4) Präsens (*temp* = '1', Z.5) enthalten. Gleichzeitig müssen die Fragen das Partizip II (*form* = '2', Z.10) des Verbs *sagen* enthalten. Die so ermittelten Fragennummern werden mit den Fragennummern der Tabelle *fragen* abgeglichen. Es werden nur die Fragen als Ergebnis ausgegeben, die im SMF (*smf is not null*, Z.14) als obligatorische Fragen (*obligatorisch* = 'j', Z.13) erhoben wurden.

⁴⁰ Für Verben, die das Perfekt mit dem Hilfsverb *sein* bilden, muss in Z.6 *grundform* = 'haben' durch *grundform* = 'sein' ersetzt werden. Durch eine Oder-Abfrage (*grundform* = 'haben' or *grundform* = 'sein') ließen sich Perfektformen von Verben, die ihr Perfekt alternativ mit *sein* oder *haben* bilden, ermitteln.

SMF	Frage
150.007	"Das hat unser Vater oft gesagt"
227.005	"Er hat es der Mutter gesagt"
227.006	"Wer hat das gesagt?" (Interr.Pron.betont)
227.007	"Wem hast du das gesagt?" (Interr.Pron.betont)
227.008	"Dem Sepp hab ich es gesagt" (oder auch andere Taufnamen, flektiert? Mit oder ohne Artikel? Spontan?)
228.002	"Ich hab's einem alten Mann gesagt"

Tabelle 5.13: Ergebnisse der Abfrage "Das Verb *sagen* im Perfekt"

b) Beispielabfrage 2: Verben im Futur I

Abfrage: Es sollen alle Fragen (Fragennummer und -text) aus der BSA-Fragekonkordanz gefunden werden, die Verben im Futur I abfragen.

SQL:

```

1      select fragenr, frage from fragen where fragenr in (
2          select fragenr from link_verben
3          where art = '1'
4          and form = '1'
5          and fragenr in (
6              select fragenr from link_verben
7              where art = '2'
8              and mod = '1'
9              and grundform = 'werden'
10         )
11     )

```

Kommentar: Aus der Tabelle *Link_Verben* werden diejenigen Fragennummern ausgewählt, deren zugehörige Fragen ein Vollverb (*art* = '1', Z.3) im Infinitiv (*form* = '1', Z.4) enthalten. Außerdem müssen die Fragen das Hilfsverb (*art* = '2', Z.7) *werden* im Indikativ (*mod* = '1', Z.8) enthalten. Die so ermittelten Fragennummern werden mit den Fragennummern aus der Tabelle *Fragen* abgeglichen, die entsprechenden Werte aus den Feldern *Fragenr* und *Frage* der Tabelle *Fragen* werden als Ergebnis ausgegeben.

Fragenr	Frage
2068	"Es wird schon werden"
2070	"Ich werde euch schon noch kriegen"
3565	"Denen werde ich es geben" (Demonstr.Pron.Dat.Plural, betont)

Tabelle 5.14: Ergebnisse der Abfrage "Verben im Futur I"

c) Beispielabfrage 3: Flexionsformen des Verbs *waschen*

Abfrage: Die Frage SUF 322.006 "'Waschen' (anderes Wort dafür? Ist zwahen bekannt?)" erhebt den Infinitiv des Verbs *waschen*. Ausgehend von dieser Frage sollen weitere Fragen (SUF-Fragennummer, Fragetext) ermittelt werden, mit welchen im SUF Formen des Verbs *waschen* erhoben wurden.

```

SQL:  1   select suf, frage
      2   from fragen
      3   where fragenr in (
      4       select fragenr
      5       from link_verben
      6       where grundform in (
      7           select grundform
      8           from link_verben
      9           where fragenr = (
     10               select fragenr
     11               from fragen
     12               where suf = '322.006'
     13           )
     14       )
     15   ) and suf is not null

```

Kommentar: Zunächst wird die Fragenummer (*fragenr*, Z.10) aus der Tabelle *Fragen* ermittelt, die der SUF-Fragennummer 322.006 entspricht. Anschließend wird die Grundform (*grundform*, Z.7) ermittelt, die in der Tabelle *Link_Verben* dieser Fragenummer entspricht. Es werden weitere Fragenummern aus den Zeilen der Tabelle *Link_verben* ermittelt, die im Feld *Grundform* den entsprechenden Wert – in diesem Fall *waschen* – enthalten. Diese Fragenummern werden mit den Fragenummern der Tabelle *Fragen* abgeglichen, die entsprechenden Werte aus den Feldern *Suf* und *Frage* werden nur ausgegeben, wenn die Fragen im SUF erhoben wurden (*suf is not null*, Z. 15).

SUF	Frage
322.006	"Waschen" (anderes Wort dafür? Ist zwahen bekannt?)
322.007	"Sie wäscht"
322.008	"Gewaschen"
322.010	"Ich bin gerade dabei zu waschen" (Übersetzen)

Tabelle 5.15: Ergebnisse der Abfrage "Flexionsformen des Verbs *waschen* "

d) Beispielabfrage 4: Partizip II von Verben der Ablautreihe 1a

Abfrage: Es sollen alle obligatorischen Fragen ausgewählt werden, die im SNIB das Partizip II von Verben der Ablautreihe 1a erheben (SNIB-Fragennummer und Fragetext).

```

SQL:  1   select snb, frage
      2   from fragen
      3   where fragenr in (
      4       select fragenr
      5       from link_verben
      6       where ablautreihe = '1a'
      7       and form = '2'
      8   ) and obligatorisch = 'j'
      9   and snb is not null

```

Kommentar: Es werden zunächst alle Fragenummern (*fragenr*, Z.4) aus der Tabelle *Link_Verben* ausgewählt, die das Partizip II (*form = '2'*, Z.7) von Verben der

Ablautreihe 1a (*ablautreihe* = '1a', Z.6) erheben. Diese werden mit den Frage-nummern aus der Tabelle *Fragen* abgeglichen. Als Ergebnis werden nur die-jenigen Fragen ausgegeben, die im SNIB (*snb is not null*, Z.9) obligatorisch (*obligatorisch* = 'j', Z.8) erhoben wurden.

Ergebnis:	SUF	Frage
	058.007	"Geritten"
	142.010	"Die Buben sind auf die Eiche gestiegen"
	190.010	"Die alten Häuser hat man abgerissen"
	248.011	"Gelitten"
	254.008	"Geblieden"
	280.011	"Geschienen"
	322.002	"Geschnitten"
	418.006	"Gebissen"
	434.005	"Ich habe dem Vater geschrieben"

Tabelle 5.16: Ergebnisse der Abfrage "Partizip II von Verben der Ablautreihe 1a"

5.2.2. Substantive

5.2.2.1. Generelles zur Auszeichnung der Substantive

Zu jedem Substantiv wurden Numerus (*Singular* oder *Plural*) und Kasus (*Nominativ*, *Genitiv*, *Dativ* oder *Akkusativ*) notiert. Das Genus (*Femininum*, *Maskulinum* oder *Neutrum*) wurde nur dann notiert, wenn es sich eindeutig ermitteln ließ. Bei Fragen, die explizit auf die Ermittlung des Genus eines Substantivs abzielen, z.B. 2097 "'Wade(n)' (Genus)" wurde das Genus prinzipiell nicht ausgezeichnet. Wenn das Genus ausgezeichnet ist, handelt es sich dabei um das nhd. Genus, da einerseits bei Substantiven, die ihr Genus seit dem Mhd. gewechselt haben, das Genus explizit erfragt wird, z.B. beim ehemals maskulinen Substantiv *Schnecke* (Frage 2409). Die ausgezeichneten nhd. Genera entsprechen damit den mhd. Genera. Andererseits richtet sich z.B. die Adjektivflexion, deren Auszeichnung unter 5.2.3. beschrieben wird, nach den nhd. Genera. Eine Auszeichnung der Genera nach dem nhd. Muster erscheint aus den gegebenen Gründen sinnvoll.

Soweit ermittelbar, wird die mhd. Flexionsklasse des Substantivs angegeben (*stark*, *schwach* oder *gemischt*).

Zusätzlich wird angegeben, ob Substantive nach Präpositionen erhoben wurden. So können Kasus, die von Präpositionen gefordert werden (z.B. Akkusativ in Frage 4203, "'Ich gehe in die Kirche.'"), von Objektkasus (z.B. Akkusativ in Frage 3398 "'Er holt eine Milch'") bzw. frei erhobenen Kasus (z.B. Akkusativ in Frage 3257 "'Einen neuen Deckel'") unterschieden werden.

5.2.2.2. Datenbanktabellen zur Auszeichnung der Substantive

Die Auszeichnungen der Substantive sind in der BAYDAT-Datenbank in der Tabelle *Link_Substantive* umgesetzt:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	Nominativ-Singular-Form des ausgezeichneten Substantivs
Grundform_Nr	Char(4)	laufende Nummer der Grundform
FrForm	Varchar2(500)	Form des ausgezeichneten Substantivs in der jeweiligen Frage
Kas	Char(1)	Kasus des ausgezeichneten Substantivs
Num	Char(1)	Numerus des ausgezeichneten Substantivs
Gen	Char(1)	Genus des ausgezeichneten Substantivs
Praep	Char(1)	Kasus nach Präposition
Flex	Char(1)	Flexionsklasse
Sugg_Form	Char(1)	ausgezeichnetes Substantiv Suggestierform oder nicht?

Tabelle 5.17: Datenfelder und -typen der Tabelle *Link_Substantive*

Wie die Tabelle *Link_Verben* enthält auch die Tabelle *Link_Substantive* zusätzlich zu den Feldern für oben genannte Auszeichnungen ein Feld für die laufende Nummer der Auszeichnung (*Nr*) und für eine laufende vierstellige Nummer der Grundform (*Grundform_Nr*). Auch diese Tabelle enthält ein Feld *Sugg_Form*, das anzeigt, ob es sich bei dem ausgezeichneten Substantiv um eine Suggestierform handelt oder nicht.

Die möglichen Werte für Kasus, Numerus und Genus sind aus den unter 5.2.1.2. genannten Gründen nicht im Klartext aufgenommen, sondern folgendermaßen als Ziffern kodiert in die BAYDAT-Datenbank übernommen worden:

Feld	Kodierung	Auszeichnung
Kas	1	Nominativ
	2	Genitiv
	3	Dativ
	4	Akkusativ
Num	1	Singular
	2	Plural
Gen	1	Femininum
	2	Maskulinum
	3	Neutrum
Flex	1	stark
	2	schwach
	3	gemischt

Tabelle 5.18: Kodierte Werte in der Tabelle *Link_Substantive*

Dem Feld *Praep* wird der Wert "j" zugewiesen, wenn das entsprechende Substantiv nach Präposition erhoben wurde, andernfalls erhält es den Wert "n".

5.2.2.3. Beispielabfragen

a) Beispielabfrage 1: Transitive Verben im Sachbereich *Essen und Trinken*

Abfrage: Es sollen alle obligatorischen Fragen (SUF-Fragennummer, Fragetext) ermittelt werden, die im SUF im Sachbereich *Essen und Trinken* erhoben wurden und die transitiv verwendete Verben abfragen, d.h. Fragen, die sowohl ein finites Verb als auch ein Substantiv im Akkusativ enthalten⁴¹.

SQL:

```

1   select suf, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_substantive
6       where kas = '4'
7       and praep = 'n'
8       and fragenr in (
9           select fragenr
10          from link_verben
11          where temp is not null
12      )and obligatorisch = 'j'
13   and suf is not null
14   and sachbereich = (
15       select sachbereichnr
16       from sachbereiche
17       where sachbereich = 'Essen und Trinken'
18   )
19   ) order by suf
```

Kommentar: Zunächst werden aus der Tabelle *Link_Fragen* diejenigen Fragennummern ausgewählt, die ein finites Verb (*temp is not null*, Z.11) enthalten. Die Fragen müssen außerdem ein Substantiv im Akkusativ (*kas = '4'*, Z.6) enthalten, der Akkusativ darf nicht von einer Präposition gefordert sein (*praep = 'n'*, Z.7). Anschließend wird überprüft, ob die Fragen im SUF (*suf is not null*, Z.13) als obligatorische Fragen (*obligatorisch = 'j'*, Z.12) erhoben wurden und ob ihre Sachbereichsnummer dem Sachbereich *Essen und Trinken* (*sachbereich = 'Essen und Trinken'*, Z.17) entspricht. Zutreffende Fragen werden nach ihrer SUF-Fragennummer (*order by suf*, Z.19) sortiert ausgegeben.

Ergebnis:

SUF	Frage
398.007	"Er hat eine dicke Birne gegessen"
402.001	"Sie haben unseren Vater vom Feld geholt"
402.003	"Er holt eine Milch"
402.005	"Ich habe (Hunger) gehabt" (einen?)
402.006	"Da hab` ich Hunger gekriegt"
418.001	"Ich habe kein Geld"

Tabelle 5.19: Ergebnisse der Abfrage "Transitive Verben im Sachbereich *Essen und Trinken*"

⁴¹ transitiv verwendete Verben meint hier transitive Verben im engeren Sinn, d.h. "Verben, die ein direktes Objekt regieren", vgl. Bußmann, S.711, s.v. "Transitivität"

a) Beispielabfrage 2: Substantive im Akkusativ nach Präpositionen

Abfrage: Es sollen alle obligatorischen Fragen (SOB-Fragennummer, Fragetext, Grundform des jeweiligen Substantivs) ermittelt werden, die im SOB Substantive im Akkusativ nach Präpositionen erheben.

SQL:

```

1   select sob, frage, grundform
2   from fragen, link_substantive
3   where kas = '4'
4   and praep = 'j'
5   and obligatorisch = 'j'
6   and sob is not null
7   and fragen.fragenr = link_substantive.fragenr

```

Kommentar: Aus der Tabelle *Link_Substantive* werden diejenigen Fragen ausgewählt, die Substantive im Akkusativ (*kas = '4'*, Z.3) nach Präpositionen (*praep = 'j'*, Z.4) erheben. Aus der Tabelle *Fragen* werden diejenigen Fragen ausgewählt, die im SOB (*sob is not null*, Z.6) obligatorisch (*obligatorisch = 'j'*, Z.5) erhoben wurden. Aus beiden Abfragen werden nur diejenigen Fragen ausgewählt, die die Anforderungen beider Anfragen erfüllen (*fragen.fragenr = link_substantive.fragenr*, Z.7). Als Ergebnis werden die SOB-Fragennummer, der Fragetext und die entsprechende Grundform des erhobenen Substantivs (*sob, frage, grundform*, Z.1) ausgegeben.

Ergebnis:

SOB	Frage	Grundform
128.010	"Er treibt das Vieh auf die Weide."	Weide
132.005	"Er geht auf den Berg (hinauf)"	Berg
142.007	"Die Buben sind auf die Eiche gestiegen"	Eiche
198.012	"Ich sehe noch gut ohne Brille (bzw. nicht mehr gut...)"	Brille
288.005	"In den April(en) schicken (am 1.April tut man jmd...)"	April
246.002	"Ich gehe in die Kirche."	Kirche
250.009	"Ich gehe in die Messe."	Messe
284.002	"Es kommt mir nicht in den Sinn"	Sinn

Tabelle 5.20: Ergebnisse der Abfrage "Substantive im Akkusativ nach Präpositionen"

c) Beispielabfrage 3: Kopula sein mit Substantiv

Abfrage: Es sollen alle im SBS erhobenen obligatorischen Fragen (SBS-Fragennummer und Fragetext) gefunden werden, die die Kopula *sein* in Verbindung mit einem Substantiv im Nominativ enthalten.

SQL:

```

1   select sbs, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_substantive
6       where kas = '1'
7       and fragenr in (
8           select fragenr
9           from link_verben

```



```

10         where art = '1'
11         and grundform = 'sein'
12     )
13 ) and sbs is not null
14     and obligatorisch = 'j'

```

Kommentar: Zunächst werden aus der Tabelle *Link_Verben* die Fragenummern zu Fragen ausgewählt, die das Verb *sein* (*grundform* = 'sein', Z.11) als Vollverb erheben. Aus diesen werden diejenigen Fragenummern ermittelt, deren Fragen ein Substantiv im Nominativ (*kas* = '1', Z.6) enthalten. Aus den ermittelten Fragen werden nur die SBS-Fragennummern und die Fragetexte der Fragen als Ergebnis ausgegeben, die im SBS (*sbs is not null* Z.13) obligatorisch (*obligatorisch* = 'j', Z.14) erhoben wurden.

Ergebnis:

SBS	Frage
016.003	"Das sind schöne Kühe"
058.008	"Wenn das Roß/der Gaul jünger wäre, könnte es/er besser ziehen"
140.001	"Das sind lange Äste"
178.006	"Die Flaschen sind gewiß voll"
198.007	"Aber seine Augen sind braun"
240.010	"Die Brille muß doch irgendwo sein"
252.003	"Heute war die Predigt langweilig"
270.009	"Es sind/hat Wolken am Himmel"
272.001	"Heute ist ... schön(es) Wetter"
272.003	"Heute ist ein schöner Tag"
290.008	"Das sind lauter heiße Tage gewesen" (Tage mit Umlaut?)
364.010	"Beide Sohlen sind hin" (sagt man beide?)
366.005	"Deine Hosen sind voller Dreck"
400.007	"Der Kaffee ist lau"
442.003	"Als ich noch ein kleiner Bub war"

Tabelle 5.21: Ergebnisse der Abfrage "Kopula *sein* mit Substantiv"

Anmerkung: Ursprünglich war hier eine Beispielabfrage intendiert, die Fragen mit als Substantiv realisierter Prädikativergänzung ermittelt. Diese Abfrage ist auf Grundlage der vorgenommenen Auszeichnungen nicht möglich. Um die Abfrage zu ermöglichen, müsste eine theoriegebundene Auszeichnung vorgenommen werden, für jedes Substantiv müsste die syntaktische Funktion angegeben werden. Mögliche theorieneutrale Auszeichnungen, z.B. die Angabe der Position eines Substantivs – vor bzw. nach dem finiten Verb – würden bei dieser speziellen Abfrage nicht zum gewünschten Ergebnis führen, in Frage SBS 252.003 "Heute war die Predigt langweilig" stehen sowohl die Nominativergänzung als auch die Prädikativergänzung nach dem finiten Verb. Eine Abfrage unter der Annahme, dass die Prädikativergänzung nach dem

finiten Verb stehen muss, würde durch diese Auszeichnung ebenso erfolglos bleiben.

Wie oben gezeigt, ist die Belegsituation für diese spezielle Abfrage recht überschaubar. Aus den 15 Fragen, die zu oben genannter Beispielabfrage als Ergebnis ausgegeben werden, können die Fragen mit substantivisch realisierter Prädikativergänzung vom Benutzer relativ leicht von Hand selektiert werden.

d) Beispielabfrage 4: Flexionsformen des Substantivs *Baum*

Abfrage: SUF-Frage 142.11 erhebt die Nominativ-Singular-Form des Substantivs *Baum*. Es sollen alle weiteren obligatorischen Fragen (SUF-Fragennummer, Fragetext) ermittelt werden, die im SUF Flexionsformen des Substantivs *Baum* erheben.

SQL:

```

1   select suf, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_substantive
6       where grundform = (
7           select grundform
8           from link_substantive
9           where fragenr = (
10              select fragenr
11              from fragen
12              where suf = '142.011'
13          )
14      )
15  )
16  and suf is not null
17  and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Fragen* wird die Fragennummer ermittelt (*fragenr*, Z.10), die der SUF-Frage 142.11 (*suf* = '142.011', Z.12) entspricht. Anschließend wird aus der Tabelle *Link_Substantive* die der Grundform (*grundform*, Z.7) entsprechende Fragennummer ermittelt. Weitere Fragen, die dieselbe Grundform enthalten (*fragenr*, Z.4) werden mit der Tabelle *Fragen* abgeglichen. Wenn die Fragen im SUF (*suf is not null*, Z.16) obligatorisch (*obligatorisch* = 'j', Z.17) erhoben wurden, werden die SUF-Fragennummer und der Fragetext als Ergebnis ausgegeben.

Ergebnis:

SUF	Frage
140.011	"Ein hoher Baum"
140.012	"Hohe Bäume"
142.011	"Baum"
142.012	"Baum" Plural
168.003	"Auf den Bäumen" (wo wachsen die Äpfel?)
170.001	"An den Ästen"/"an den Bäumen" (woran hängen die Äpfel? - eins von beiden)

Tabelle 5.22: Ergebnisse der Abfrage "Flexionsformen des Substantivs *Baum* "

5.2.3. Adjektive

5.2.3.1. Generelles zur Auszeichnung der Adjektive

Zu jedem Adjektiv wurde dessen Komparationsform (*Positiv*, *Komparativ* oder *Superlativ*) und Position (*attributiv*, *prädikativ*, *adverbial* oder *freistehend*) notiert. Als prädikative Position wird dabei nicht nur die Position nach Kopulaverben, z.B. in Frage 1428 "Der Baum ist dürr" nach der Kopula *sein* oder in Frage 4389 "Die Eier sind schlecht geworden, darum hat sie die Mutter weggeworfen" nach der Kopula *werden*, angesehen. Auch Adjektive, die in den jeweiligen Fragen die Rolle depictiver oder resultativer Prädikative⁴² übernehmen, werden als prädikativ ausgezeichnet, z.B. in Frage 0615 "Man hat das Heu nass hereingetan." Eine genauere Differenzierung der jeweiligen Prädikative wurde aufgrund der Beleglage – insgesamt finden sich in sämtlichen obligatorischen Fragen der BSA-Fragekonkordanz nur 42 prädikativ verwendete Adjektive – nicht vorgenommen, d.h. der Benutzer der BAYDAT-Datenbank kann aus der geringen Zahl prädikativ gebrauchter Adjektive depictive bzw. resultative Prädikative selbständig auswählen. Die Position eines Adjektivs wird als *adverbial* angegeben, wenn das Adjektiv in der jeweiligen Frage die Funktion einer Adverbiale übernimmt⁴³, *adverbial* bezeichnet hier also die Funktion, nicht die Wortart. Als *adverbial* wird z.B. das Adjektiv *schnell* in Frage 2725 "Geh schnell hinauf" markiert.

Bei attributiv verwendeten Adjektiven wird die Flexion (*stark* bzw. *schwach*) angegeben, in allen anderen Positionen wird mit *unflektiert* ausgezeichnet. Zusätzlich werden für flektierte Adjektive Genus (*Femininum*, *Maskulinum* oder *Neutrum*), Kasus (*Nominativ*, *Genitiv*, *Dativ* oder *Akkusativ*) und Numerus (*Singular* oder *Plural*) angegeben.

Die Auszeichnung der Zahlwörter folgt dem von Gallmann vorgeschlagenen Muster⁴⁴. Daher werden Kardinalzahlen, Ordinalzahlen und vereinzelt Mengenangaben (z.B. *der Andere* in Frage 4499 "Warst das du oder der andere?"; näheres zur Auszeichnung substantivierter Adjektive s.u.) ohne weitere Markierung als Adjektive ausgezeichnet. Die Kardinalzahl *Eins* in attributiver Verwendung wird aufgrund ihrer formalen Nichtunterscheidbarkeit vom unbestimmten Artikel als Artikelwort ausgezeichnet. Unbestimmte Ordinal- und Vervielfältigungszahlen, die Gallmann als Adjektive wertet, werden innerhalb des BSA nicht obligatorisch abgefragt.

Die im BSA obligatorisch erhobenen deadjektivischen Substantive, z.B. *der Größte* (Frage 3778), *der Kleinste* (Frage 3166) oder *das Schwere* (Frage 3223), wurden zwar als Substantive ausgezeichnet, da sie aber auch typische Flexionsmerkmale von Adjektiven

⁴² vgl. Gallmann: Der Satz, S.843 f.

⁴³ vgl. Gallmann: Der Satz, S.844 f.

⁴⁴ vgl. Gallmann: Die flektierbaren Wortarten, S.384 ff.

zeigen, insbesondere die Komparation bei den genannten konvertierten Superlativen, werden sie zusätzlich als Adjektive ausgezeichnet⁴⁵.

5.2.3.2. Datenbanktabellen zur Auszeichnung der Adjektive

Die Auszeichnungen der Adjektive finden sich in der BAYDAT-Datenbank in der Tabelle

Link_Adjektive:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	unflektierte Positivform des ausgezeichneten Adjektivs
Grundform_Nr	Char(4)	laufende Nummer der Grundform
FrForm	Varchar2(500)	Form des ausgezeichneten Adjektivs in der jeweiligen Frage
Komp	Char(1)	Komparationsform des ausgezeichneten Adjektivs
Pos	Char(1)	Position des ausgezeichneten Adjektivs
Flex	Char(1)	Flexion des ausgezeichneten Adjektivs
Num	Char(1)	Numerus des ausgezeichneten Adjektivs (wenn flektiert)
Kas	Char(1)	Kasus des ausgezeichneten Adjektivs (wenn flektiert)
Gen	Char(1)	Genus des ausgezeichneten Adjektivs (wenn flektiert)
Sugg_Form	Char(1)	ausgezeichnetes Adjektiv innerhalb Suggestierform oder nicht?

Tabelle 5.23: Datenfelder und -typen der Tabelle *Link_Adjektive*

Die Tabelle *Link_Adjektive* enthält, ebenso wie die Tabellen zur Auszeichnung von Verben und Substantiven, Felder für eine laufende Nummer der Auszeichnung (*Nr*) sowie für eine laufende vierstellige Nummer der Grundform (*Grundform_Nr*). Das Feld *Sugg_Form* zeigt an, ob es sich bei dem ausgezeichneten Adjektiv um eine Suggestierform handelt oder ob es im regulären Fragetext enthalten ist.

Die Werte für Komparationsstufe, Position, Flexion, Numerus, Kasus und Genus werden aus den unter 5.2.1.2. genannten Gründen in folgender kodierter Form in die BAYDAT-Datenbank übernommen:

Feld	Kodierung	Auszeichnung
Komp	1	Positiv
	2	Komparativ
	3	Superlativ
Pos	1	freistehend
	2	attributiv
	3	prädikativ
	4	adverbial

⁴⁵ zur deadjektivischen Konversion vgl. Fleischer/Barz, S.215 f.

Flex	1	unflektiert
	2	schwach
	3	stark
Num	1	Singular
	2	Plural
Kas	1	Nominativ
	2	Genitiv
	3	Dativ
	4	Akkusativ
Gen	1	Femininum
	2	Maskulinum
	3	Neutrum

Tabelle 5.24: Kodierte Werte in der Tabelle *Link_Adjektive*

5.2.3.3. Beispielabfragen

a) Beispielabfrage 1: Attributive Adjektive im Sachbereich *Das Vieh und seine Pflege*

Abfrage: Es sollen alle obligatorischen Fragen (Fragennummer und -text) aus dem Sachbereich *Das Vieh und seine Pflege* ermittelt werden, die Adjektive in attributiver Position enthalten.

```

SQL:  1      select fragenr, frage
      2      from fragen
      3      where fragenr in (
      4          select fragenr
      5          from link_adjektive
      6          where fragenr in (
      7              select fragenr
      8              from fragen
      9              where sachbereich = (
     10                  select sachbereichnr
     11                  from sachbereiche
     12                  where sachbereich = 'Das Vieh und seine
     13                  Pflege'
     14                  ) and pos = '2'
     15          )
     16      ) and obligatorisch = 'j'
```

Kommentar: Zunächst wird die Sachbereichnummer (*sachbereichnr*, Z.10) aus der Tabelle *Sachbereiche* ausgewählt, die dem Sachbereich *Das Vieh und seine Pflege* (*sachbereich* = 'Das Vieh und seine Pflege', Z.12 f.) entspricht. Anschließend werden alle Fragennummern (*fragenr*, Z.7) aus der Tabelle *Fragen* ausgewählt, deren Wert im Feld *Sachbereich* der zuvor ausgewählten Sachbereichnummer entspricht. Es wird überprüft, ob die Fragennummern in der Tabelle *Link_Adjektive* stehen, d.h. ob die jeweiligen Fragen ein Adjektiv enthalten, und ob das ausgezeichnete Adjektiv in attributiver Position (*pos* = '2', Z.14) steht. Die so ausgewählten Fragennummern werden mit der Tabelle *Fragen* abgeglichen, Fragennummern und -texte werden als Ergebnis ausgegeben, wenn die Frage obligatorisch (*obligatorisch* = 'j', Z.16) erhoben wurde.

Ergebnis:	Fragenr	Frage
	0103	"Das sind schöne Kühe"
	0104	"Das sind schöne Kälblein"
	0105	"Die alte Kuh"
	0114	"Die vorderen Füße"
	0115	"Die hinteren Füße"
	0245	"Ein blecherner Eimer"

Tabelle 5.25: Ergebnisse der Abfrage "Attributive Adjektive im Sachbereich *Das Vieh und seine Pflege*"

b) Beispielabfrage 2: Freistehende Adjektive im Komparativ

Abfrage: Es sollen alle obligatorischen Fragen (SBS-Fragennummer, Frage) ermittelt werden, die im SBS freistehende Komparative erheben.

SQL:

```

1   select sbs, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_adjektive
6       where komp = '2'
7       and pos = '1'
8   ) and obligatorisch = 'j'
9   and sbs is not null

```

Kommentar: Zunächst werden aus der Tabelle *Link_Adjektive* die Fragennummern (*fragenr*, Z.4) derjenigen Fragen ausgewählt, die freistehende (*pos* = '1', Z.7) Komparativformen (*komp* = '2', Z.6) erheben. Die so ausgewählten Fragennummern werden mit der Tabelle *Fragen* abgeglichen. Wenn die Fragen im SBS (*sbs is not null*, Z.9) obligatorisch (*obligatorisch* = 'j', Z.8) erhoben wurden, werden SBS-Fragennummer und Fragetext (*sbs, frage*, Z.1) als Ergebnis ausgegeben.

Ergebnis:	SBS	Frage
	196.005	"Kürzer"
	248.001	"Höher"
	248.003	"Tiefer"
	286.010	"Später"
	352.004	"Härter"
	370.002	"Leichter"
	370.004	"Weiter"
	370.006	"Breiter"
	370.008	"Schmäler"
	370.010	"Wärmer"
	370.012	"Heißer"
	370.014	"Kälter"
	370.016	"Größer"

372.002	"Stärker"
372.004	"Kleiner"
372.006	"Zäher"
466.008	"Näher"

Tabelle 5.26: Ergebnisse der Abfrage "Freistehende Adjektive im Komparativ"

c) Beispielabfrage 3: Stark flektierte Adjektive ohne Nominativ

Abfrage: Es sollen alle Fragen (SMF-Fragenummer und Fragetext) gefunden werden, die im SMF obligatorisch stark flektierte Adjektive erheben, die nicht im Nominativ stehen.

SQL:

```

1  select smf, frage
2  from fragen
3  where fragenr in (
4      select fragenr
5      from link_adjektive
6      where kas <> '1'
7      and flex = '3'
8  ) and obligatorisch = 'j'
9  and smf is not null

```

Kommentar: Aus der Tabelle *Link_Adjektive* werden zunächst die Fragenummern zu Fragen ausgewählt, die stark flektierte (*flex = '3', Z.7*) Adjektive enthalten. Diese dürfen nicht im Nominativ (*kas <> '1', Z.6*) stehen. Anschließend werden diese Fragenummern mit der Tabelle *Fragen* abgeglichen. Wenn die entsprechenden Fragen im SMF (*smf is not null, Z.9*) obligatorisch (*obligatorisch = 'j', Z.8*) erhoben wurden, werden die SMF-Fragenummer (*smf, Z.1*) und der Fragetext (*frage, Z.1*) als Ergebnis ausgegeben.

Ergebnis:

SMF	Frage
063.005	"Da bauen sie wieder eine neue Straße."
098.002	"In einer alten Hütte"
102.010	"Sie hat blaue Augen (Pers.Pron.unbet.)"
106.007	"Er hat einen dicken Backen"
125.003	"Sie hat lauter böse Kinder"
127.009	"Von einem hohen Turm (sieht man viel)"
175.005	"Er hat es einer alten Frau gebracht"
192.010	"Einen neuen Deckel"
201.007	"Er hat eine dicke Birne gegessen"
234.003	"Einen schönen Traum (Heute Nacht habe ich... gehabt)"
086.006	"Hinter unserem Haus steht ein Apfelbaum mit schönen roten Äpfeln."
101.002	"Er hat schwarze Haare"
101.005	"Er hat kurze Haare"
101.008	"Er hat weiße Haare"
183.001	"Der Mantel hat ihr nicht gefallen. Jetzt hat sie sich einen schöneren gekauft."
203.012	"Aus grünen Zwetschgen kann man keine Marmelade machen"
220.012	"Einen langen Brief haben wir ihr geschrieben"

Tabelle 5.27: Ergebnisse der Abfrage "Stark flektierte Adjektive ohne Nominativ"

d) Beispielabfrage 4: Prädikative Adjektive ohne Kopula sein

Abfrage: Es sollen alle Fragen (SNOB-Fragennummer und Fragetext) ermittelt werden, die im SNOB obligatorisch prädikativ verwendete Adjektive erheben. Diese sollen jedoch nicht die Kopula *sein* enthalten.

SQL:

```

1   select sno, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_adjektive
6       where pos = '3'
7       and fragenr in (
8           select fragenr
9           from link_verben
10          where art = '1'
11          and grundform <> 'sein'
12          and fragenr not in (
13              select fragenr
14              from link_verben
15              where art = '1'
16              and grundform = 'sein'
17          )
18      )
19   ) and obligatorisch = 'j'
20   and sno is not null

```

Kommentar: Aus der Tabelle *Link_Verben* werden die Fragennummern derjenigen Fragen ausgewählt, die Vollverben (*art* = '1', Z.10) enthalten. Bei diesen Vollverben darf es sich nicht um die Kopula *sein* (*grundform* <> 'sein', Z.11) handeln. Durch die innerste Abfrage (Z.13 – 16) wird außerdem ausgeschlossen, dass in den vorher ermittelten Fragen Fragen enthalten sind, die gleichzeitig ein von *sein* verschiedenes Vollverb und die Kopula *sein* enthalten. Frage 4191, "Das Mädchen ist schlecht in der Schule. Wenn es der Lehrer etwas fragt, dann weiß es nichts." enthält z.B. Formen der Vollverben *wissen* und *fragen*, aber auch eine Form der Kopula *sein*. Aus den gewählten Fragennummern werden anschließend durch Abfragen der Tabelle *Link_Adjektive* diejenigen ausgewählt, deren Fragen ein prädikativ verwendetes Adjektiv (*pos* = '3', Z.6) enthalten. Schließlich werden die ermittelten Fragennummern mit der Tabelle *Fragen* abgeglichen. Wurden die den Fragennummern entsprechenden Fragen im SNOB (*sno is not null*, Z.20) obligatorisch (*obligatorisch* = 'j', Z.19) erhoben, werden SNOB-Fragennummer (*sno*, Z.1) und Fragetext (*frage*, Z.1) als Ergebnis ausgegeben.

Durch die beiden innersten Abfragen würden Fragen, die sowohl prädikative Adjektive nach der Kopula *sein* als auch nach einer anderen Kopula erheben, als Ergebnis ausgeschlossen, als Beispiel wäre die Frage "Letzte Woche war

das Wetter schlecht. Seitdem wurde es wieder besser." denkbar. Da Fragen dieser Art, d.h. Fragen, die zwei prädikativ verwendete Adjektive nach unterschiedlichen Kopula erheben, in der BSA-Fragekonkordanz nicht auftauchen, führt die obige Abfrage zu dem gewünschten Ergebnis.

Ergebnis:

SNO	Frage
072.006	"Man hat das Heu nass hereingetan."
194.009	"Das geschieht dir recht"
210.004	"Das Wetter ist schön geworden"
210.010	"Da würde ich krank"
330.002	"Man muß fest reiben, sonst wird es nicht sauber"

Tabelle 5.28: Ergebnisse der Abfrage "Prädikative Adjektive ohne Kopula sein"

5.2.4. Pronomen

5.2.4.1. Generelles zur Auszeichnung der Pronomen

Als Pronomen werden innerhalb der BAYDAT-Datenbank Pronomen im engeren Sinn, d.h. "allein die substantiv[isch] gebrauchten P[ronomen]"⁴⁶ ausgezeichnet, z.B. *der* in Frage 3589 der BSA-Fragekonkordanz, "Jetzt kommt auch der dran". Attributiv gebrauchte Pronomen, z.B. *euer* in Frage 1872 der BSA-Fragekonkordanz, "Euer Haus", werden als Artikelwörter ausgezeichnet. Diese Trennung entspricht den u.a. von Gallmann⁴⁷ und Weinrich⁴⁸ gegebenen theoretischen Ansätzen, sie erfolgt hier aber auch aus praktischen Gründen: Vergleicht man die Felder der Tabelle *Link_Pronomen* (s. 5.2.4.2) mit den Feldern der Tabelle *Link_Artikel* (s. 5.2.5.2), so sind diese weitgehend deckungsgleich. Die Tabelle *Link_Pronomen* enthält jedoch die zusätzlichen Felder *PersB* zur Angabe der Person des Bezugswortes und *KasB* zur Angabe des Kasus des Bezugswortes, die Tabelle *Link_Artikel* enthält das zusätzliche Feld *FlexK* zur Angabe der Flexion des jeweiligen Artikelworts. Eine Aufteilung in zwei Datenbanktabellen wäre allein zur Auszeichnung definiter und indefiniter Artikel nötig gewesen, durch die Übernahme attributiv gebrauchter Pronomen in die Tabelle *Link_Artikel* werden jedoch leer stehende Felder in den beiden Tabellen *Link_Pronomen* und *Link_Artikel* vermieden, Speicherplatz wird eingespart. Durch die Trennung von Pronomen und Artikelwörtern ergeben sich auch Vorteile für den Nutzer der Datenbank: So wird im Feld *Flex* der Tabelle *Link_Adjektive* zwar die Flexion des jeweiligen Adjektivs angegeben, es kann jedoch nicht ermittelt werden, ob das Adjektiv aufgrund eines fehlenden Artikels oder aufgrund eines nicht flektierten Artikelworts stark flektiert. Durch die zusätzliche Abfrage der Tabelle *Link_Artikel* lässt sich ebendiese Information leicht ermitteln (vgl. 5.2.5.3, Beispielabfrage 4).

⁴⁶ Schaeder, Burkhard: Pronomen. In: Glück, S.552 f., s.v. "Pronomen"

⁴⁷ vgl. Gallmann: Die flektierbaren Wortarten, S.255 ff.

⁴⁸ vgl. Weinrich, S.406 ff.

Für jedes Pronomen wird dessen Art (*Demonstrativ-, Relativ-, Interrogativ-, Personal-, Indefinit- oder Reflexivpronomen*) notiert. Possessivpronomen in substantivischer Verwendung werden innerhalb der BSA-Fragekonkordanz nicht abgefragt. Für sämtliche Pronomen werden Kasus (*Nominativ, Genitiv, Dativ oder Akkusativ*) und Numerus (*Singular oder Plural*) angegeben. Das Genus (*Femininum, Maskulinum oder Neutrum*) wurde nur notiert, wenn es eindeutig zu ermitteln war, für Reflexivpronomen und Personalpronomen der ersten und zweiten Person sowie Personalpronomen der dritten Person Plural wird prinzipiell kein Genus angegeben, die Genusangaben der Indefinitpronomen folgen den Angaben bei Gallmann⁴⁹. Bei Reflexiv- und Personalpronomen wird zusätzlich die Person (*1, 2 oder 3*) notiert.

Da die Grundformen der Pronomen als Ansatzpunkt für die Verknüpfung der Fragen innerhalb der grafischen Benutzeroberfläche dienen, war es notwendig, für die Demonstrativpronomen *die* (dritte Person Singular Femininum) und *die* (dritte Person Plural) sowie für die Personalpronomen *sie* (dritte Person Singular Femininum) und *sie* (dritte Person Plural) unterschiedliche Grundformen anzusetzen (*die* bzw. *sie* für die Pronomen im Singular, *die_p* bzw. *sie_p* für die Pronomen im Plural). Auf diese Weise lassen sich innerhalb der Benutzeroberfläche Fragen zu den Singular- und Pluralformen jeweils getrennt anzeigen.

Lässt sich aus der jeweiligen Frage das Bezugswort des ausgezeichneten Pronomens ermitteln (z.B. *Mädchen* als Bezugswort zum Personalpronomen *es* in Frage 4191 der BSA-Fragekonkordanz "Das Mädchen ist schlecht in der Schule. Wenn es der Lehrer etwas fragt, dann weiß es nichts."), so wird dieses in seiner Grundform zusammen mit Kasus (*Nominativ, Genitiv, Dativ oder Akkusativ*), Numerus (*Singular oder Plural*), Genus (*Femininum, Maskulinum oder Neutrum*) und Person (*1, 2 oder 3*) notiert.

5.2.4.2. Datenbanktabellen zur Auszeichnung der Pronomen

Die Auszeichnungen der Pronomen finden sich in der BAYDAT-Datenbank in der Tabelle

Link_Pronomen:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	Nominativ-Singular-Form des ausgezeichneten Pronomens
Grundform_Nr	Char(4)	laufende Nummer der Grundform
FrForm	Varchar2(500)	Form des ausgezeichneten Pronomens in der jeweiligen Frage
GenB	Char(1)	Genus des Bezugsworts

⁴⁹ vgl. Gallmann: Die flektierbaren Wortarten, S.316 ff.

PersB	Char(1)	Person des Bezugsworts
KasB	Char(1)	Kasus des Bezugsworts
NumB	Char(1)	Numerus des Bezugsworts
Bezugswort	Varchar2(500)	Bezugswort
GenP	Char(1)	Genus des Pronomens
PersP	Char(1)	Person des Pronomens
KasP	Char(1)	Kasus des Pronomens
NumP	Char(1)	Numerus des Pronomens
Art	Char(1)	Art des Pronomens
Sugg_Form	Char(1)	ausgezeichnetes Pronomen innerhalb Suggestierform oder nicht?

Tabelle 5.29: Datenfelder und -typen der Tabelle *Link_Pronomen*

Die Tabelle *Link_Pronomen* enthält, ebenso wie die Tabellen zur Auszeichnung der übrigen Wortarten ein Feld für die laufende Nummer der Auszeichnung (*Nr*) sowie für eine laufende vierstellige Nummer der Grundform (*Grundform_Nr*). Die Tabelle enthält außerdem ein Feld *Sugg_Form*. Dieses zeigt an, ob sich das ausgezeichnete Pronomen innerhalb einer Suggestierform oder innerhalb des regulären Fragetextes befindet.

Die Werte für Genus, Person, Kasus und Numerus der Pronomen bzw. Bezugswörter sowie die Art der Pronomen werden aus den unter 5.2.1.2. genannten Gründen in folgender kodierter Form in die BAYDAT-Datenbank übernommen:

Feld	Kodierung	Auszeichnung
GenB/GenP	1	Femininum
	2	Maskulinum
	3	Neutrum
PersB/PersP	1	1. Person
	2	2. Person
	3	3. Person
KasB/KasP	1	Nominativ
	2	Genitiv
	3	Dativ
	4	Akkusativ
NumB/NumP	1	Singular
	2	Plural
Art	1	Demonstrativpronomen
	2	Relativpronomen
	3	Interrogativpronomen
	4	Personalpronomen
	5	Indefinitpronomen
	6	Reflexivpronomen

Tabelle 5.30: Kodierte Werte in der Tabelle *Link_Pronomen*

5.2.4.3. Beispielabfragen

a) Beispielabfrage 1: Personalpronomen mit vom Bezugswort verschiedenem Kasus

Abfrage: Es sollen alle Fragen (Fragennummer und -text, Bezugswort) ermittelt werden, die Personalpronomen erheben, die im Kasus nicht mit ihrem Bezugswort übereinstimmen.

SQL:

```

1   select fragen.fragenr, frage, bezugswort
2   from fragen, link_pronomen
3   where kasb <> kasp
4   and art = '4'
5   and fragen.fragenr = link_pronomen.fragenr

```

Kommentar: Aus den Tabellen *Fragen* und *Link_Pronomen* werden zunächst Fragenummern ausgewählt, die in beiden Tabellen vorkommen (*fragen.fragenr* = *link_pronomen.fragenr*, Z.5). Aus diesen werden die Fragen ausgewählt, die Personalpronomen (*art* = '4', Z.4) erheben, deren Kasus nicht mit dem Kasus ihres Bezugsworts übereinstimmt (*kasb* <> *kasp*, Z.3). Die Fragenummern, -texte und die in der Tabelle *Link_Pronomen* enthaltenen Bezugswörter werden als Ergebnis ausgegeben.

Ergebnis:

Fragenr	Frage	Bezugswort
4191	"Das Mädchen ist schlecht in der Schule. Wenn es der Lehrer etwas fragt, dann weiß es nichts."	Mädchen
4389	"Die Eier sind schlecht geworden, darum hat sie die Mutter weggeworfen"	Ei

Tabelle 5.31: Ergebnisse der Abfrage "Personalpronomen mit vom Bezugswort verschiedenem Kasus"

b) Beispielabfrage 2: Relativpronomen

Abfrage: Es sollen alle obligatorischen Fragen (SOB-Fragennummer und Fragetext) ermittelt werden, die im SOB Relativpronomen erheben.

SQL:

```

1   select sob, frage
2   from fragen
3   where fragenr in (
4   select fragenr
5   from link_pronomen
6   where art = '2'
7   ) and sob is not null
8   and obligatorisch = 'j'

```

Kommentar: Aus der Tabelle *Link_Pronomen* werden Fragenummern zu denjenigen Fragen ermittelt, die Relativpronomen (*art* = '2', Z.6) enthalten. Wenn die entsprechenden Fragen im SOB (*sob is not null*, Z.7) obligatorisch (*obligatorisch* = 'j', Z.8) erhoben wurden, werden die jeweilige SOB-Fragennummer und der Fragetext als Ergebnis ausgegeben.

Ergebnis:

SOB	Frage
208.010	"Ich weiß nicht, wer du bist"
208.017	"Ich weiß nicht, wer ihr seid"

Tabelle 5.32: Ergebnisse der Abfrage "Relativpronomen"

c) Beispielabfrage 3: Reflexivpronomen

Abfrage: Es sollen alle im SUF erhobenen obligatorischen Fragen (SUF-Fragennummer, Fragetext, Grundform des Verbs) ermittelt werden, die Reflexivpronomen enthalten. Zusätzlich sollen die Grundformen der in den Fragen vorkommenden Vollverben angegeben werden.

SQL:

```

1   select suf, frage, grundform from
2   link_verben, fragen
3   where link_verben.fragenr in (
4       select fragenr
5       from link_pronomen
6       where art = '6'
7   ) and art = '1'
8   and suf is not null
9   and obligatorisch = 'j'
10  and fragen.fragenr = link_verben.fragenr

```

Kommentar: Zunächst werden aus der Tabelle *Link_Pronomen* die Fragennummern derjenigen Fragen ermittelt, die Reflexivpronomen (*art = '6', Z.6*) erheben. Sind die entsprechenden Fragennummern auch in der Tabelle *Link_Verben* enthalten und enthalten diese außerdem Vollverben (*art = '1', Z.7*), wird überprüft, ob sie im SUF (*suf is not null, Z.8*) obligatorisch (*obligatorisch = 'j', Z.9*) erhoben wurden. Ist dies der Fall, so werden SUF-Fragennummer, Fragetext und die Grundform des in der jeweiligen Frage vorkommenden Vollverbs als Ergebnis ausgegeben. Die Anweisung in Z.10 verhindert die Bildung des kartesischen Produkts aus den Tabellen *Fragen* und *Link_Verben*.

Ergebnis:

SUF	Frage	Grundform
134.006	"Da würde ich mich fürchten"	fürchten
134.007	"Da habe ich mich gefürchtet"	fürchten
194.004	"Er hat (sich) den Kopf angeschlagen" (Akk.!)	anschlagen
212.007	"Sie hat sich die Hand verstaucht"	verstauchen
220.003	"Er hat sich eine Rippe gebrochen"	brechen
220.017	"Das kann ich mir schon denken" (Anlaut)	denken
222.001	"Das habe ich mir gleich gedacht"	denken
268.001	"Ihr müßt (euch) einen Schirm mitnehmen"	mitnehmen
310.006	"Es dreht sich im Kreis herum"	herumdrehen
390.008	"Setz dich hin!" (Richt.adv.)	hinsetzen
390.009	"Setz dich neben mich"	setzen
444.012	"Du solltest dich schämen"	schämen

Tabelle 5.33: Ergebnisse der Abfrage "Reflexivpronomen"

d) Beispielabfrage 4: Demonstrativpronomen in der 3. Person Singular Maskulinum

Abfrage: Es sollen alle Fragen (SMF-Fragennummer und Fragetext) gefunden werden, die im SMF obligatorisch Demonstrativpronomen, die in der 3. Person Singular Maskulinum stehen, erheben.

```

SQL:
1      select smf, frage
2      from fragen
3      where fragenr in (
4          select fragenr
5          from link_pronomen
6          where art = '1'
7          and persp = '3'
8          and nump = '1'
9          and genp = '2'
10     ) and obligatorisch = 'j'
11     and smf is not null

```

Kommentar: Aus der Tabelle *Link_Pronomen* werden die Fragenummern zu denjenigen Fragen ausgewählt, die Demonstrativpronomen (*art = '1'*, Z.6) in der 3. Person (*persp = '3'*, Z.7) Singular (*nump = '1'*, Z.8) Maskulinum (*genp = '2'*, Z.9) erheben. Wenn die entsprechenden Fragen im SMF (*smf is not null*, Z.11) obligatorisch (*obligatorisch = 'j'*, Z.10) erhoben wurden, werden die SMF-Fragennummer und der Fragetext als Ergebnis ausgegeben.

Ergebnis:	SMF	Frage
	167.006	"Den sollte man schon kennen"
	217.006	"Den erwische ich schon noch" (Dem.Pron.bet.)
	218.006	"Jetzt kommt auch der dran"
	224.005	"Der ist dazu zu jung"
	228.003	"Dem darf man es nicht sagen" (Demonstr. Pron.betont)
	236.006	"Ich habe es nicht diesem, sondern jenem gegeben" (Dem.Pron. des weiter entfernten Objekts)
	114.011	"Dem ist nicht zu helfen" (Dem.Pron.bet.)
	215.006	"Der verliert manchmal etwas."
	221.001	"Den haben sie ihr gestern erst gebracht"

Tabelle 5.34: Ergebnisse der Abfrage "Demonstrativpronomen in der 3. Person Singular Maskulinum"

5.2.5. Artikelwörter

5.2.5.1. Generelles zur Auszeichnung der Artikelwörter

Für jedes Artikelwort wird dessen Art (*definitiver* oder *indefinitiver Artikel*, *possessives*, *indefinites*, *demonstratives* oder *interrogatives Artikelwort*) sowie Kasus (*Nominativ*, *Genitiv*, *Dativ* oder *Akkusativ*), Numerus (*Singular* oder *Plural*), Genus (*Femininum*, *Maskulinum* oder *Neutrum*) und die Grundform des jeweiligen Bezugssubstantivs notiert. Bei possessiven Artikelwörtern wird zusätzlich deren Person (1, 2 oder 3) ausgezeichnet.

Prinzipiell stimmen "Artikelwörter [...] mit dem Substantiv in Genus, Numerus und Kasus überein."⁵⁰ Bei possessiven Artikelwörtern trägt jedoch der Stamm die Merkmale Person und Numerus sowie in der dritten Person Singular auch Genus, "die Endungen stimmen mit den Merkmalen des folgenden Substantivs überein."⁵¹ Stamm und Endung können dabei in

⁵⁰ Gallmann: Die flektierbaren Wortarten, S.263

⁵¹ Gallmann: Die flektierbaren Wortarten, S.285

Numerus und Genus voneinander abweichen. In Frage 1985 der BSA-Fragekonkordanz "Aber seine Augen sind braun" zeigt z.B. der Stamm des possessiven Artikelworts *seine* Singular Maskulinum an, das Bezugswort *Augen* trägt die Merkmale Neutrum Plural. Daher wird insbesondere bei possessiven Artikelwörtern zusätzlich das Numerus und Genus des jeweiligen Bezugswortes notiert.

Da die Adjektivflexion (stark oder schwach) von der Endung des vorangehenden Artikelworts (endungslos oder mit Endung) abhängt⁵², wird außerdem angegeben, ob es sich bei dem ausgezeichneten Artikelwort um ein Artikelwort mit bzw. ohne Flexionsendung handelt.

5.2.5.2. Datenbanktabellen zur Auszeichnung der Artikelwörter

Die Auszeichnung der Artikelwörter finden sich in der BAYDAT-Datenbank in der Tabelle

Link_Artikel:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	Nominativ-Singular-Form des ausgezeichneten Artikelworts
Grundform_Nr	Char(4)	laufende Nummer der Grundform
FrForm	Varchar2(500)	Form des ausgezeichneten Artikelworts in der jeweiligen Frage
FlexK	Char(1)	Flexionsendung oder nicht?
GenBez	Char(1)	Genus des Bezugsworts
KasBez	Char(1)	Kasus des Bezugsworts
NumBez	Char(1)	Numerus des Bezugsworts
Bezugswort	Varchar2(500)	Bezugswort
PersArt	Char(1)	Person des Artikelworts
GenArt	Char(1)	Genus des Artikelworts
KasArt	Char(1)	Kasus des Artikelworts
NumArt	Char(1)	Numerus des Artikelworts
Art	Char(1)	Art des Artikelworts
Sugg_Form	Char(1)	ausgezeichnetes Artikelwort innerhalb Suggestierform oder nicht?

Tabelle 5.35: Datenfelder und -typen der Tabelle *Link_Artikel*

Zusätzlich zu den unter 5.2.5.1. beschriebenen Feldern finden sich in der Tabelle ein Feld für die laufende Nummer der Auszeichnung (*Nr*) sowie ein Feld für die laufende Nummer der jeweils ausgezeichneten Grundform (*Grundform_Nr*). Die Tabelle enthält wie die übrigen Tabellen zur Auszeichnung sprachwissenschaftlicher Phänomene ein Feld *Sugg_Form*, das anzeigt, ob sich das ausgezeichnete Artikelwort in einer Suggestierform oder im regulären Fragetext befindet.

⁵² vgl. Gallmann: Die flektierbaren Wortarten, S.368

Die Werte für Genus und Numerus des Artikel- bzw. des Bezugswort sowie für Person, Kasus, Flexion und Art des Artikelworts werden aus den unter 5.2.1.2. genannten Gründen wie folgt in kodierter Form in die Tabelle *Link_Artikel* der BAYDAT-Datenbank übernommen:

Feld	Kodierung	Auszeichnung
GenArt/GenBez	1	Femininum
	2	Maskulinum
	3	Neutrum
PersArt	1	1. Person
	2	2. Person
	3	3. Person
KasArt/KasBez	1	Nominativ
	2	Genitiv
	3	Dativ
	4	Akkusativ
NumArt/NumBez	1	Singular
	2	Plural
FlexK	1	mit Endung
	2	ohne Endung
Art	1	definitiver Artikel
	2	indefinitiver Artikel
	3	possessives Artikelwort
	4	indefinites Artikelwort
	5	demonstratives Artikelwort
	6	interrogatives Artikelwort

Tabelle 5.36: Kodierte Werte in der Tabelle *Link_Artikel*

5.2.5.3. Beispielabfragen

a) Beispielabfrage 1: Artikelwörter mit vom Bezugswort verschiedenem Genus/Numerus

Abfrage: Es sollen alle im SUF als obligatorisch erhobenen Fragen (SUF-Fragennummer, Fragetext) ermittelt werden, die Artikelwörter enthalten, die in Genus oder Numerus vom Numerus oder Genus ihrer jeweiligen Bezugswörter abweichen.

SQL:

```

1      select suf, frage
2      from fragen
3      where fragenr in (
4          select fragenr
5          from link_artikel
6          where genbez <> genart
7          or numbez <> numart)
8      and suf is not null
9      and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Artikel* werden die Fragennummern derjenigen Fragen ermittelt, die Artikelwörter enthalten, die mit ihrem Bezugswort in Genus (*genbez <> genart*, Z.6) oder Numerus (*numbez <> numart*, Z.7) nicht übereinstimmen. Wenn die entsprechenden Fragen im SUF (*suf is not null*, Z.8) obligatorisch (*obligatorisch = 'j'*, Z.9) erhoben wurden, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:	SUF	Frage
	062.001	"Unsere Katze"
	182.004	"Euer Haus"
	182.005	"Unser Haus"
	198.009	"Aber seine Augen sind braun"
	228.011	"All meine Lebtag"
	244.005	"Man hat unserem Vater das Auto gestohlen"
	250.004	"Unser (Herr) Pfarrer"
	298.008	"Das hat unser Vater oft gesagt"
	366.005	"Deine Hosen sind voller Dreck"
	402.001	"Sie haben unseren Vater vom Feld geholt"
	434.005	"Sie hat ihren Vater angelogen"
	436.001	"Unsere Mutter"
	436.002	"Eure Mutter"
	436.008	"Ihr Bruder"
	436.009	"Meine Brüder"
	436.010	"Ihre Brüder"
	436.013	"Ihre Schwester"
	436.014	"Meine Schwestern"
	436.015	"Ihre Schwestern"
	472.004	"Sie hat ihren Vater geschimpft"

Tabelle 5.37: Ergebnisse der Abfrage "Artikelwörter mit vom Bezugswort verschiedenem Genus/Numerus"

b) Beispielabfrage 2: Artikellose Substantive in Sätzen

Abfrage: Es sollen Fragen (SUF-Fragennummer und Fragetext) ermittelt werden, die im SUF obligatorisch artikellose Substantive in Sätzen erheben.

SQL:

```

1  select suf, frage
2  from fragen
3  where fragenr in (
4      select fragenr
5      from link_substantive
6      where fragenr in (
7          select fragenr
8          from link_verben
9          where art = '1'
10     ) and fragenr not in (
11         select fragenr
12         from link_artikel
13     )
14 ) and suf is not null
15 and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Substantive* werden Fragenummern zu denjenigen Fragen ausgewählt, die die jeweiligen Substantive im Satz erheben, d.h. die Fragen müssen neben dem Substantiv ein Vollverb enthalten. Dies wird durch die Abfrage in den Zeilen 7 – 9 sichergestellt. Außerdem darf die Frage keine Artikelwörter enthalten: Fragenummern, die in der Tabelle *Link_Artikel* stehen, werden durch die Abfrage zwischen Zeile 10 und 13 vom Ergebnis

ausgenommen. Wurden die so ermittelten Fragen im SUF (*suf is not null*, Z.14) obligatorisch (*obligatorisch = 'j'*, Z.15) erhoben, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SUF	Frage
016.002	"Das sind schöne Kühe"
016.003	"Das sind schöne Kälblein"
140.001	"Das sind lange Äste"
198.008	"Sie hat blaue Augen" (Pers.Pron.unbet.)
198.013	"Ich sehe noch gut ohne Brille" (bzw. nicht mehr gut...)
200.010	"Er liest ..."
200.011	"Wir lesen ..."
200.012	"Ihr lest ..."
214.013	"Er ist noch gut zu Fuß"
240.009	"Sie hat lauter böse Kinder"
272.001	"Heute ist ... schön(es) Wetter"
272.002	"Heute ist... schlecht(es) Wetter"
284.007	"Er ist vor zehn Jahren gekommen"
290.008	"Das sind lauter heiße Tage gewesen" (Tage mit Umlaut?)
298.005	"Behüte dich Gott"
298.006	"Behüte euch Gott"
308.002	"(Herr) Nachbar, komm mal herüber"
350.003	"Soll ich morgen auch wieder Brot bringen?"
402.005	"Ich habe (Hunger) gehabt" (einen?)
402.006	"Da hab` ich Hunger gekriegt"
418.003	"Wenn ich Geld hätte"
418.004	"Wenn du Geld hättest"
436.011	"Das sind denen ihre Brüder"
468.011	"Ich bekomme noch fünfundzwanzig Pfennig"

Tabelle 5.38: Ergebnisse der Abfrage "Artikellose Substantive in Sätzen"

c) Beispielabfrage 3: Possessive Artikelwörter in der 1. Person Singular

Abfrage: Es sollen Fragen ermittelt werden, die im SMF obligatorisch possessive Artikelwörter in der ersten Person Singular (*mein*) erheben.

SQL:

```

1   select smf, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_artikel
6       where art = '3'
7       and persart = '1'
8       and numart = '1'
9   ) and smf is not null
10  and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Artikel* werden Fragennummern zu denjenigen Fragen ausgewählt, die possessive Artikelwörter (*art = '3'*, Z.6) in der ersten Person (*persart = '1'*, Z.7) Singular (*numart = '1'*, Z.8) enthalten. Wurden diese im

SMF (*smf is not null*, Z.9) obligatorisch (*obligatorisch = 'j'*, Z.10) erhoben, werden SMF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SMF	Frage
119.002	"All meine Lebtag"
123.010	"Das hat mich/mir meine Mutter gelehrt/gelernt"
222.003	"Mein Bruder"
222.005	"Meine Brüder"
222.008	"Meine Schwester"
222.010	"Meine Schwestern"
238.003	"Er hat meinen Bruder gemalt"

Tabelle 5.39: Ergebnisse der Abfrage "Possessive Artikelwörter in der 1. Person Singular"

d) Beispielabfrage 4: Endungslose Artikelwörter mit Adjektiven

Abfrage: Es sollen alle im SOB als obligatorisch erhobenen Fragen (SOB-Fragennummer und Fragetext) ermittelt werden, die sowohl endungslose Artikelwörter als auch Adjektive enthalten.

SQL:

```

1      select sob, frage
2      from fragen
3      where fragenr in (
4          select fragenr
5          from link_artikel
6          where flexk = '2'
7          and fragenr in (
8              select fragenr from link_adjektive
9          )
10     ) and obligatorisch = 'j'
11     and sob is not null

```

Kommentar: Zunächst werden sämtliche Fragennummern aus der Tabelle *Link_Adjektive* ausgewählt, d.h. die Fragennummern derjenigen Fragen, die Adjektive enthalten (Z.8). Enthalten die Fragen gleichzeitig endungslose Artikelwörter (*flexk = '2'*, Z.6) und sind die entsprechenden Fragen im SOB (*sob is not null*, Z.10) obligatorisch (*obligatorisch = 'j'*) erhoben, werden SOB-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SOB	Frage
100.008	"Das ist ein altes Maß" (Genus!)
132.012	"Ein ebenes Feld"
140.010	"Ein hoher Baum"
168.005	"Ein großer Apfel"
170.012	"Ein gutes Jahr (das Gegenteil von ein schlechtes Jahr)"
240.007	"Ein gesundes Kind"
272.004	"Heute ist ein schöner Tag"
286.006	"Ein früher Winter (wenn es schon im Oktober schneit)"
286.007	"Ein später Winter"
318.002	"Ein seidenes/seidiges Kleid"

368.002	"Ein enger Ärmel (das Gegenteil von einem weiten Ärmel ist...)"
442.003	"Als ich noch ein kleiner Bub war"

Tabelle 5.40: Ergebnisse der Abfrage "Endungslose Artikelwörter mit Adjektiven"

5.2.6. Junktionen

5.2.6.1. Generelles zur Auszeichnung von Junktionen

Zu jeder Junktion wird deren Art (*Subjunktion* oder *Konjunktion*) notiert. Zusätzlich wird die Semantik der jeweiligen Junktion angegeben (*additiv*, *adversativ*, *alternativ*, *temporal*, *kausal*, *konditional*, *modal* oder *neutral*). Die semantische Einteilung folgt dabei dem von Nübling vorgeschlagenen Modell⁵³.

5.2.6.2. Datenbanktabellen zur Auszeichnung der Junktionen

Die Auszeichnungen der Junktionen finden sich in der BAYDAT-Datenbank in der Tabelle *Link_Subkon*:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	ausgezeichnete Junktion
Grundform_Nr	Char(4)	laufende Nummer der Junktion
Art	Char(1)	Art (Konjunktion/Subjunktion)
Sem	Char(1)	Semantik der Junktion
Sugg_Form	Char(1)	ausgezeichnete Junktion innerhalb Suggestierform oder nicht?

Tabelle 5.41: Datenfelder und -typen der Tabelle *Link_Subkon*

Die Tabelle *Link_Subkon* enthält, ebenso wie die Tabellen zur Auszeichnung der übrigen Wortarten eine laufende Nummer der jeweiligen Auszeichnung (im Feld *Nr*) sowie eine laufende vierstellige Nummer der Grundform (im Feld *Grundform_Nr*). Das Feld *Sugg_Form* gibt an, ob es sich bei der ausgezeichneten Junktion um eine Suggestierform oder um eine Junktion im Fragetext handelt.

⁵³ vgl. Nübling, S.628 ff. (Konjunktionen) bzw. 633 ff. (Subjunktionen)

Die Werte für Art und Semantik der Junktoren werden aus den unter 5.2.1.2. genannten Gründen in folgender kodierter Form in die BAYDAT-Datenbank übernommen:

Feld	Kodierung	Auszeichnung
Art	1	Subjunktion
	2	Konjunktion
Sem	1	Additiv
	2	Adversativ
	3	Alternativ
	4	Temporal
	5	Kausal
	6	Konditional
	7	Modal
	8	Neutral

Tabelle 5.42: Kodierte Werte in der Tabelle *Link_Subkon*

5.2.6.3. Beispielabfragen

a) Beispielabfrage 1: Kausale Subjunktionen

Abfrage: Es sollen alle Fragen (SMF-Fragennummer, Fragetext) ermittelt werden, die im SMF obligatorisch kausale Subjunktionen erheben.

SQL:

```

1   select smf, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_subkon
6       where art = '1'
7       and sem = '5'
8   ) and obligatorisch = 'j'
9   and smf is not null

```

Kommentar: Zunächst werden aus der Tabelle *Link_Subkon* die Fragennummern derjenigen Fragen ausgewählt, die kausale (*sem = '5'*, Z.7) Subjunktionen (*art = '1'*, Z.6) enthalten. Werden diese Fragen im SMF (*smf is not null*, Z.9) obligatorisch (*obligatorisch = 'j'*, Z.8) erhoben, werden SMF-Fragennummer und Fragetext als Ergebnis ausgegeben.

SMF	Frage
210.009	"Sie ißt nichts, weil sie krank ist"
212.006	"Die langweilen sich, weil sie nichts zu tun haben"

Tabelle 5.43: Ergebnisse der Abfrage "Kausale Subjunktionen"

b) Beispielabfrage 2: Konjunktionen im SBS und SUF

Abfrage: Es sollen alle obligatorischen Fragen (SUF- und SBS-Fragennummer, Fragetext) ermittelt werden, die sowohl im SUF als auch im SBS Konjunktionen erheben.

SQL:

```

1   select sbs, suf, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_subkon

```

```

6           where art = '2'
7     ) and suf is not null
8     and sbs is not null
9     and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Subkon* werden die Fragenummern aller Fragen ausgewählt, die Konjunktionen (*art = '2'*, Z.6) enthalten. Wurden diese Fragen sowohl im SUF (*suf is not null*, Z.7) als auch im SBS (*sbs is not null*, Z.8) obligatorisch (*obligatorisch = 'j'*, Z.9) erhoben, werden SBS-Fragennummer, SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SBS	SUF	Frage
198.007	198.009	"Geh hinunter und hol mir die Wolle"
304.005	304.005	"Aber seine Augen sind braun"
308.007	308.007	"Er steht nicht hier, sondern dort"
444.002	444.001	"Ich tue das, ob du willst oder nicht"
464.004	464.004	"Ich habe es nicht diesem, sondern jenem gegeben" (Dem.Pron. des weiter entfernten Objekts)

Tabelle 5.44: Ergebnisse der Abfrage "Konjunktionen im SBS und SUF"

5.2.7. Präpositionen

5.2.7.1. Generelles zur Auszeichnung der Präpositionen

Zu jeder Präposition wird der regierte Kasus (*Genitiv*, *Dativ* oder *Akkusativ*) notiert. Die Auszeichnung entspricht dabei dem Fragetext, wenn der regierte Kasus in der Frage vom regierten Kasus in der Standardsprache abweicht, werden beide Kasus ausgezeichnet. Z.B. wird in Frage 4463 der BSA-Fragekonkordanz, "'Wegen dir haben wir verloren'", die Präposition *wegen* mit Dativ abgefragt, hier wird zusätzlich der Genitiv als geregierter Kasus notiert (vgl. 5.2.7.3., Beispielabfrage 3).

Des Weiteren wird die Bedeutung jeder Präposition (*lokal*, *temporal*, *modal*, *kausal* oder *neutral*) ausgezeichnet. Die Auszeichnung orientiert sich dabei an dem von Nübling⁵⁴ vorgeschlagenen Modell.

Ist die ausgezeichnete Präposition mit einer Artikelform verschmolzen (*am*, *im* usw.), so wird dies entsprechend gekennzeichnet.

⁵⁴ Nübling, S.610 ff.

5.2.7.2. Datenbanktabellen zur Auszeichnung der Präpositionen

Die Auszeichnung der Präpositionen finden sich in der BAYDAT-Datenbank in der Tabelle *Link_Praepositionen*:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	ausgezeichnete Präposition
Grundform_Nr	Char(4)	laufende Nummer der Präposition
Kasus	Char(1)	regierter Kasus der ausgezeichneten Präposition
Semantik	Char(1)	durch ausgezeichnete Präposition ausgedrücktes semantisches Verhältnis
Verschmelzungsform	Char(1)	liegt Verschmelzungsform vor oder nicht
Sugg_Form	Char(1)	ausgezeichnete Präposition innerhalb Suggestierform oder nicht?

Tabelle 5.45: Datenfelder und -typen der Tabelle *Link_Praepositionen*

Die Tabelle *Link_Praepositionen* enthält, genau wie die Tabellen zur Auszeichnung der übrigen Wortarten eine laufende Nummer der jeweiligen Auszeichnung im Feld *Nr* sowie eine laufende vierstellige Nummer der Grundform im Feld *Grundform_Nr*. Auch diese Tabelle enthält ein Feld *Sugg_Form*, das anzeigt, ob sich die ausgezeichnete Präposition innerhalb einer Suggestierform oder innerhalb des regulären Fragetexts befindet.

Die Werte für Kasus und Semantik der Präpositionen werden aus den unter 5.2.1.2. genannten Gründen in folgender kodierter Form in die BAYDAT-Datenbank übernommen:

Feld	Kodierung	Auszeichnung
Kasus	2	Genitiv
	3	Dativ
	4	Akkusativ
Semantik	1	Lokal
	2	Temporal
	3	Modal
	4	Kausal
	5	Neutral

Tabelle 5.46: Kodierte Werte in der Tabelle *Link_Praepositionen*

Das Feld *Verschmelzungsform* erhält den Wert 'j' bei Präpositionen mit Verschmelzung, den Wert 'n' bei Präpositionen ohne Verschmelzung.

5.2.7.3. Beispielabfragen

a) Beispielabfrage 1: Modale Präpositionen

Abfrage: Es sollen alle Fragen (SUF-Fragennummer, Fragetext) ermittelt werden, die im SUF obligatorisch modale Präpositionen erheben.

```

SQL:      1      select suf, frage
          2      from fragen
          3      where fragenr in (
          4          select fragenr
          5          from link_praepositionen
          6          where semantik = '3'
          7      ) and suf is not null
          8      and obligatorisch = 'j'

```

Kommentar: Aus der Tabelle *Link_Praepositionen* werden die Fragenummern derjenigen Fragen ausgewählt, die modale Präpositionen (*semantik = '3'*, Z.6) enthalten. Wurden die entsprechenden Fragen im SUF (*suf is not null*, Z.7) obligatorisch (*obligatorisch = 'j'*, Z.8) erhoben, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SUF	Frage
144.006	"Mit Ketten"
198.013	"Ich sehe noch gut ohne Brille" (bzw. nicht mehr gut...)
206.007	"Er hat Schwierigkeiten mit den Zähnen"
212.009	"Er faßt es mit beiden Händen an" (Dat. Plural)
212.013	"Er schreibt alles mit der linken Hand"
214.013	"Er ist noch gut zu Fuß"
248.009	"Man tut mit allen Glocken läuten"
328.012	"Mit dem Besen tut man kehren"
364.001	"Die Schuhe sind aus Leder"
374.001	"Er ist im Stehen eingeschlafen"

Tabelle 5.47: Ergebnisse der Abfrage "Modale Präpositionen"

b) Beispielabfrage 2: Präposition *in* mit Verschmelzung

Abfrage: Es sollen alle obligatorischen Fragen aus der BSA-Fragekonkordanz (Fragennummer und -text) ermittelt werden, die die Präposition *in* mit Verschmelzung erheben.

```

SQL:      1      select fragenr, frage
          2      from fragen
          3      where fragenr in (
          4          select fragenr
          5          from link_praepositionen
          6          where grundform = 'in'
          7          and verschmelzung = 'j'
          8      ) and obligatorisch = 'j'

```

Kommentar: Aus der Tabelle *Link_Praepositionen* werden die Fragenummern derjenigen Fragen ausgewählt, die die Präposition *in* (*grundform = 'in'*, Z.6) mit Verschmelzung (*verschmelzung = 'j'*, Z.7) enthalten. Werden die Fragen innerhalb der BSA-Fragekonkordanz obligatorisch (*obligatorisch = 'j'*, Z.8) erhoben, werden Fragennummer und -text als Ergebnis ausgegeben.

Ergebnis:	Fragenr	Frage
	0204	An zugiger Stelle lagern, "sie sind im Schatten", "sie tun schattnen"
	2065	"Ihr seid im Garten"
	2066	"Sie sind im Bett"
	2127	"Er hat's im Kreuz"
	2622	"Im März" (im Märzen?)
	2749	"Es dreht sich im Kreis herum"
	2804	"Ich habe gehört, er sei im Bett"
	3194	"Er ist im Stehen eingeschlafen"
	3527	"Die Kinder spielen im Garten"

Tabelle 5.48: Ergebnisse der Abfrage "Präposition *in* mit Verschmelzung"

c) Beispielabfrage 3: Präpositionen mit Genitiv

Abfrage: Es sollen alle obligatorischen Fragen aus der BSA-Fragekonkordanz (Fragennummer und -text) ermittelt werden, die Präpositionen, die den Genitiv regieren, erheben.

SQL:

```

1   select fragenr, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_praepositionen
6       where kasus = '2'
7   ) and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Praepositionen* werden die Fragennummern derjenigen Fragen ausgewählt, die Präpositionen, die den Genitiv regieren (*kasus* = '2', Z.6), enthalten. Werden die Fragen innerhalb der BSA-Fragekonkordanz obligatorisch (*obligatorisch* = 'j', Z.7) erhoben, werden Fragennummer und -text als Ergebnis ausgegeben.

Ergebnis:	Fragenr	Frage
	4463	"Wegen dir haben wir verloren"

Tabelle 5.49: Ergebnisse der Abfrage "Präpositionen mit Genitiv"

5.2.8. Adverbien

5.2.8.1. Generelles zur Auszeichnung der Adverbien

Zu jedem Adverb wird dessen Klasse angegeben (*Situierungs-*, *Konjunktion-*, *Präpositional-/Pronominal-* oder *Interrogativadverb*). Die Klassenunterteilung entspricht dabei dem von Nübling vorgeschlagenen Modell⁵⁵. Auf eine Unterteilung der Situierungsadverbien in absolute Adverbien und Pro-Adverbien wird dabei aufgrund der geringen Beleglage verzichtet: Innerhalb der gesamten BSA-Fragekonkordanz wurden 244 Situierungsadverbien mit 72 unterschiedlichen Grundformen ausgezeichnet, darunter zahlreiche Adverbien, die aufgrund ihrer unklaren Semantik, z.B. lokales/temporales *da*, doppelt ausgezeichnet wurden.

⁵⁵ vgl. Nübling, S.578 ff.

Entsprechend bleibt es dem Nutzer der Datenbank vorbehalten, aus den gegebenen Adverbien die gewünschten absoluten Adverbien bzw. Pro-Adverbien auszuwählen, eventuell nach einer vorherigen Einschränkung der Belege nach der Semantik der Adverbien.

Darüber hinaus wurde für Situierungs-, Präpositional-/Pronominal-, Interrogativ- und Konjunkionaladverbien deren Semantik notiert (*temporal, modal, kausal, konzessiv, konsekutiv, adversativ* oder *kopulativ*), bei Lokaladverbien wurde weiter zwischen statischen Ortsadverbien und direktionalen Richtungsadverbien unterschieden⁵⁶. Konnte aufgrund des fehlenden Kontexts die Semantik einzelner Adverbien nicht klar ermittelt werden, wurden diese mehrfach ausgezeichnet. So wurde z.B. das Adverb *da* in Frage 3401 der BSA-Fragekonkordanz, " 'Da hab' ich Hunger gekriegt' ", sowohl als Orts- als auch als Temporaladverb ausgezeichnet.

Nübling macht auf Abgrenzungsprobleme zwischen Adverbien und anderen Wortarten aufmerksam⁵⁷. Bei der Auszeichnung der Fragen aus der BSA-Fragekonkordanz fiel eine klare Trennung zwischen Adverbien und homonymen Partikeln aufgrund des fehlenden Kontexts oft schwer. In Zweifelsfällen wurden die entsprechenden Wörter sowohl als Adverb als auch als Partikel ausgezeichnet.

5.2.8.2. Datenbanktabellen zur Auszeichnung der Adverbien

Die Auszeichnungen der Adverbien werden in die Tabelle *Link_Adverbien* der BAYDAT-Datenbank übernommen:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	ausgezeichnetes Adverb
Grundform_Nr	Char(4)	laufende Nummer des Adverbs
Semantik	Char(1)	Semantik des ausgezeichneten Adverbs
Art	Char(1)	Art des ausgezeichneten Adverbs
Sugg_Form	Char(1)	ausgezeichnetes Adverb innerhalb Suggestierform oder nicht?

Tabelle 5.50: Datenfelder und -typen der Tabelle *Link_Adverbien*

Die Tabelle enthält, wie die anderen Tabellen zur Auszeichnung morphologischer Phänomene, ein Feld für die laufende Nummer der Auszeichnung (*Nr*) sowie ein Feld für eine laufende vierstellige Nummer der jeweiligen Grundform (*Grundform_Nr*). Die Tabelle enthält ebenfalls ein Feld *Sugg_Form*. Dieses gibt an, ob sich das ausgezeichnete Adverb in einer Suggestierform oder im Fragetext befindet.

⁵⁶ vgl. Nübling, S.580 f.

⁵⁷ vgl. Nübling, S.576

Aus Gründen der besseren Sortierbarkeit und zur Redundanzvermeidung werden die Werte der Felder *Semantik* und *Art* in folgender numerischer Kodierung in die Tabelle *Link_Adverbien* übernommen:

Feld	Kodierung	Auszeichnung
Semantik	1	Ort
	2	Richtung
	3	Temporal
	4	Modal
	5	Kausal
	6	Konzessiv
	7	Konsekutiv
	8	Adversativ
	9	Kopulativ
Art	1	Situierungsadverb
	2	Interrogativadverb
	3	Präpositionaladverb
	4	Konjunkionaladverb
	5	Kommentaradverb

Tabelle 5.51: Kodierte Werte in der Tabelle *Link_Adverbien*

5.2.8.3. Beispielabfragen

a) Beispielabfrage 1: Lokale Interrogativadverbien

Abfrage: Es sollen Fragen (SUF-Fragennummer, Fragetext) ermittelt werden, die im SUF obligatorisch lokale Interrogativadverbien erheben.

SQL:

```

1  select suf, frage
2  from fragen
3  where fragenr in (
4      select fragenr
5      from link_adverbien
6      where art = '2'
7      and semantik = '1'
8  ) and obligatorisch = 'j'
9  and suf is not null

```

Kommentar: Aus der Tabelle *Link_Adverbien* werden die Fragennummern derjenigen Fragen ausgewählt, die lokale (*semantik* = '1', Z.7) Interrogativadverbien (*art* = '2', Z.6) erheben. Wurden die Fragen im SUF (*suf is not null*, Z.9) obligatorisch (*obligatorisch* = 'j', Z.8) abgefragt, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SUF	Frage
423.001	"Ich weiß nicht, wo ich bin"
423.002	"..., wo du bist"
423.003	"..., wo er ist"
423.004	"..., wo wir sind"
423.005	"..., wo ihr seid"
423.006	"..., wo sie sind"
476.007	"Wo lauft denn ihr hin?"

Tabelle 5.52: Ergebnisse der Abfrage "Lokale Interrogativadverbien"

b) Beispielabfrage 2: Richtungsadverbien auf -wärts

Abfrage: Es sollen Fragen (SBS-Fragennummer, Fragetext) gefunden werden, die im SBS obligatorisch Richtungsadverbien auf *-wärts* erheben.

SQL:

```

1   select sbs, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_adverbien
6       where grundform like '%wärts'
7       and semantik = '2'
8   ) and sbs is not null
9   and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Adverbien* werden die Fragennummern derjenigen Fragen ausgewählt, die Richtungsadverbien (*semantik = '2', Z.7*), die auf *-wärts* (*grundform like '%wärts', Z.6*) enden, enthalten. Werden die entsprechenden Fragen im SBS (*sbs is not null, Z.8*) obligatorisch (*obligatorisch = 'j', Z.9*) erhoben, werden SBS-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SUF	Frage
306.009	"Geh rückwärts (mit dem Hinterteil voraus)"
310.005	"Aus-/ein-/auf-/abwärts" mit dem Geschäft geht's auf-/abwärts. Krankheit: es geht aufwärts

Tabelle 5.53: Ergebnisse der Abfrage "Richtungsadverbien auf -wärts"

5.2.9. Partikeln**5.2.9.1. Generelles zur Auszeichnung der Partikeln**

Zu den Partikeln wurde lediglich deren jeweilige Klasse (*Grad-, Fokus-, Negations-, Modal- oder Gesprächspartikel*) notiert. Die Auszeichnung folgt dabei der von Nübling vorgeschlagenen Unterteilung.⁵⁸

5.2.9.2. Datenbanktabellen zur Auszeichnung der Partikeln

Die Auszeichnung der Partikeln findet sich innerhalb der BAYDAT-Datenbank in der Tabelle *Link_Partikeln*:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Frage
Grundform	Varchar2(500)	ausgezeichnete Partikel
Grundform_Nr	Char(4)	laufende Nummer der Partikel
Art	Char(1)	Art der ausgezeichneten Partikel
Sugg_Form	Char(1)	ausgezeichnete Partikel in Suggestierform oder nicht?

Tabelle 5.54: Datenfelder und -typen der Tabelle *Link_Partikeln*

⁵⁸ vgl. Nübling, S.595 ff.

Wie die übrigen Tabellen zur Auszeichnung morphologischer Phänomene enthält die Tabelle *Link_Partikeln* ein Feld für die laufende Nummer der Auszeichnung (*Nr*) sowie ein Feld für eine laufende vierstellige Nummer der entsprechenden Grundform (*Grundform_Nr*), die der Verknüpfung der Fragen innerhalb der grafischen Benutzeroberfläche der BAYDAT-Datenbank dient. Die Tabelle enthält außerdem ein Feld *Sugg_Form*, das angibt, ob sich die ausgezeichnete Partikel innerhalb einer Suggestierform oder innerhalb des Fragetextes befindet.

Die Werte im Feld *Art* wurden aus Gründen der Redundanzvermeidung und besseren Sortierbarkeit in folgender kodierter Form in die Tabelle *Link_Partikeln* aufgenommen:

Feld	Kodierung	Auszeichnung
Art	1	Modalpartikel
	2	Fokuspartikel
	3	Gradpartikel
	4	Gesprächspartikel
	5	Negationspartikel

Tabelle 5.55: Kodierte Werte in der Tabelle *Link_Partikeln*

5.2.9.3. Beispielabfragen

a) Beispielabfrage 1: Negationspartikeln mit finitem Vollverb und Substantiv

Abfrage: Es sollen Fragen (SUF-Fragennummer, Fragetext) gefunden werden, die Negationspartikeln enthalten. Außerdem sollen die Fragen mindestens ein finites Vollverb sowie mindestens ein Substantiv enthalten und im SUF obligatorisch erhoben worden sein.

SQL:

```

1   select suf, frage
2   from fragen
3   where fragenr in (
4       select fragenr
5       from link_verben
6       where form is null
7       and art = '1'
8       and fragenr in (
9           select fragenr
10          from link_partikeln
11          where art = '5'
12      ) and fragenr in (
13          select fragenr
14          from link_substantive
15      )
16   ) and suf is not null
17   and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Verben* werden die Fragennummern derjenigen Fragen ausgewählt, die finite (*form is null*, Z.6) Vollverben (*art = '1'*, Z.7) enthalten. Außerdem müssen die Fragen eine Negationspartikel – überprüft durch die Unterabfrage Z.8 – 11 – und ein Substantiv – überprüft durch die Unterabfrage

Z.12 – 14 – enthalten. Wurden die entsprechenden Fragen im SUF (*suf is not null*, Z.16) obligatorisch (*obligatorisch = 'j'*, Z.17) erhoben, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

SUF	Frage
164.009	"Schlag mir einen Nagel ein, sonst hält es nicht!"
282.002	"Der Lehrer weiß auch nicht alles"
284.002	"Das paßt mir nicht in den Kram"
455.012	"Die Kühe (unbetont!) kauf ich nicht!"
455.013	"Die Häuser (betont!) gefallen uns nicht!"

Tabelle 5.56: Ergebnisse der Abfrage "Negationspartikeln mit finitem Vollverb und Substantiv"

a) Beispielabfrage 2: Gradpartikeln bei Adjektiven

Abfrage: Es sollen Fragen (SBS-Fragennummer, Fragetext) gefunden werden, die im SBS obligatorisch Gradpartikeln bei Adjektiven erheben.

SQL:

```

1      select sbs, frage
2      from fragen
3      where fragenr in (
4          select fragenr
5          from link_partikeln
6          where art = '3'
7          and fragenr in (
8              select fragenr
9              from link_adjektive
10         )
11     ) and sbs is not null
12     and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Partikeln* werden die Fragennummern derjenigen Fragen ausgewählt, die Gradpartikeln (*art = '3'*, Z.6) erheben. Gleichzeitig müssen die Fragen mindestens ein Adjektiv enthalten. Dies wird durch die Unterabfrage Z.7 – Z.9 überprüft. Wurden die entsprechenden Fragen im SBS (*sbs is not null*, Z.11) obligatorisch (*obligatorisch = 'j'*, Z.12) erhoben, werden SBS-Fragennummer und Fragetext als Ergebnis ausgegeben.

SBS	Frage
304.006	"Geh aber ganz leise hinein"

Tabelle 5.57: Ergebnisse der Abfrage "Gradpartikeln bei Adjektiven"

5.3. Auszeichnung von Wortbildungen

5.3.1. Generelles zur Auszeichnung der Wortbildungen

Zur Auszeichnung von Wortbildungen wurden aus der BSA-Fragekonkordanz zunächst die Fragen ausgewählt, die obligatorisch Wortbildungen erheben bzw. solche Fragen, die als Antwort eine Wortbildung erwarten ließen, z.B. weil eine Wortbildung als Suggestierform vorgegeben war. So erhebt z.B. Frage 1014 der BSA-Fragekonkordanz obligatorisch das Determinativkompositum *Drahtzaun*, diese Wortbildung wurde entsprechend ausgezeichnet.

Zu Frage 1054 der BSA-Fragekonkordanz – "Feld, auf dem Heu geerntet wird" – wird z.B. das Determinativkompositum *Heufeld* als Suggestierform angegeben, auch diese Wortbildung wird entsprechend ausgezeichnet. Fragen, die obligatorisch Wortbildungen erheben, können über das Feld *Obligatorisch* der Tabelle *Fragen* leicht von Fragen unterschieden werden, die aufgrund einer Suggestierform eine Wortbildung als Antwort erwarten lassen (vgl. die Beispielabfragen unter 5.3.4.).

Als Wortbildungen werden hier nur solche Formen ausgezeichnet, die synchron als Wortbildungen erkennbar sind. Die Auszeichnung folgt damit dem von Fleischer/Barz vorgeschlagenen Modell⁵⁹. Damit wird z.B. das Substantiv *Schwein*, das durch Frage 370 der BSA-Fragekonkordanz – "Sau (Allg. Bezeichnung für das Schwein)" – abgefragt wird, nicht als Wortbildung ausgezeichnet, obwohl es sich historisch gesehen um eine Wortbildung, eine "Erweiterung (Zugehörigkeitsbildung) zu *Sau*"⁶⁰ handelt. Damit fehlen dem sprachwissenschaftlichen Nutzer der BAYDAT-Datenbank zwar explizite Informationen zur historischen Wortbildung, diese Informationen sind dennoch implizit in der Datenbank enthalten und können durch entsprechende Suchabfragen, für o.g. Beispiel z.B. durch eine Abfrage aller Fragen, die das Substantiv *Schwein* als Bestandteil enthalten, ermittelt werden. Durch die Beschränkung der Auszeichnungen auf synchron erkennbare Wortbildungen wird einerseits der Laienbenutzer nicht überfordert, andererseits können die Auszeichnungen fundierter vorgenommen werden, da in Zweifelsfällen nicht auf unsichere Etyma zurückgegriffen werden muss.

Im Gegensatz zu Fleischer/Barz⁶¹ werden innerhalb der BAYDAT-Datenbank nicht nur standardsprachliche Wortbildungen ausgezeichnet, sondern auch dialektale Wortbildungen, wenn diese eindeutig als solche zu erkennen sind. So wird z.B. die Suggestierform *Heuet*, die zu Frage 1054 der BSA-Fragekonkordanz – "Wie heißt das Feld, auf dem Gras wächst und auf dem Heu geerntet wird?" – angegeben wird, als Wortbildung ausgezeichnet, obwohl das DEUTSCHE UNIVERSALWÖRTERBUCH das Substantiv *Heuet* als süddeutsch bzw. schweizerisch kennzeichnet⁶². *Heuet* ist jedoch als Wortbildung mit der Konstituente *Heu* erkennbar und wird entsprechend ausgezeichnet.

Die weitere Auszeichnung der Wortbildungen erfolgte auf Basis der jeweiligen Wortbildungskonstituenten, d.h. für jede Wortbildungskonstituente einer Wortbildung ist jeweils eine Zeile in der Tabelle *Link_Wortbildungen* der BAYDAT-Datenbank vorgesehen.

⁵⁹ vgl. Fleischer/Barz, S.9

⁶⁰ Kluge/Seebold, S.832, s.v. "Schwein"

⁶¹ vgl. Fleischer/Barz, S.8

⁶² vgl. Duden. Deutsches Universalwörterbuch, S.764, s.v. "Heuet" (hier mit der Bedeutung *Heuernte*)

Die Auszeichnung der Wortbildungen beschränkt sich dabei auf die erste Ebene: Für die Wortbildung *Sumpfdotterblume*, erhoben in Frage 2792 der BSA-Fragekonkordanz, werden z.B. die Konstituenten *Sumpf* und *Dotterblume* notiert, *Dotterblume* wird nicht weiter ausgezeichnet. Als Wortbildungskonstituente werden hier im Gegensatz zu Fleischer/Barz⁶³ auch Bestandteile von Wortbildungen ohne unmittelbare Konstituenten (Konversionen, implizite Ableitungen) angesehen. Diese terminologische Ungenauigkeit ist jedoch notwendig, um Redundanzen innerhalb der Datenbank zu vermeiden: So finden sich im Feld *Konst* der Tabelle *Link_Wortbildungen* z.B. die unmittelbaren Konstituenten *Kirsche* und *Baum* des in Frage 1677 – "Die Kirschbäume blühen heuer spät" – erhobenen Determinativkompositums *Kirschbaum*, aber auch *gehen* als Basis der impliziten Ableitung *Gang*, die in Frage 1908 der BSA-Fragekonkordanz abgefragt wird. Die Gleichbehandlung von zugrunde liegenden Wortbildungsbasen in Wortbildungen ohne UK-Struktur und Wortbildungskonstituenten in Wortbildungen mit UK-Struktur spart ein Feld in der Tabelle *Link_Wortbildungen* ein, was einerseits zu einer Einsparung von Speicherplatz, andererseits zu einer Vereinfachung von Abfragen der Tabelle *Link_Wortbildungen* führt.

Als Konstituente wird dabei außerdem die Nennform des jeweiligen Wortes angenommen (für Substantive der Nominativ Singular, für Verben der Infinitiv, für Adjektive die unflektierte Positivform), die nicht immer mit der in der Wortbildung enthaltenen Form übereinstimmt. Frage 1049 der BSA-Fragekonkordanz – "Sträßlein" – erhebt z.B. den Diminutiv zum Substantiv *Straße*. Als Konstituente wird dabei nicht *Sträß-* sondern *Straße* notiert. Dies hat den Vorteil, dass die Konstituenten mit den Grundformen in den Tabellen zur Auszeichnung morphologischer Phänomene übereinstimmen. Über eine Kreuzabfrage der Tabellen *Link_Wortbildungen* und *Link_Substantive* lassen sich so für genanntes Beispiel sämtliche Fragen ermitteln, die Wortbildungen zum Substantiv *Straße* erheben (vgl. hierzu auch Beispielabfrage 2 unter 5.3.4.).

Affixe wurden jeweils in der Form notiert, die in der jeweiligen Wortbildung vorlagen, für Frage 2542 der BSA-Fragekonkordanz – "'Krankheit' (evtl. Wahr-, Gewohnheit)" – z.B. *-heit*, für Frage 3013 – "Feuchtigkeit" – dagegen *-igkeit*. Aufgrund der Beleglage erschien eine Mehrfachauszeichnung nach den jeweiligen kombinatorischen Varianten, für vorheriges Beispiel *-heit*, *-keit* und *-igkeit*⁶⁴, insgesamt nicht sinnvoll. Doppelt ausgezeichnet wurden lediglich Diminutiva, jeweils mit dem Suffix *-chen* und *-lein*. Hier rechtfertigte die Beleglage – insgesamt wurden in der BSA-Fragekonkordanz 76 Diminutivbildungen erhoben – die

⁶³ vgl. Fleischer/Barz, S.45 ff.

⁶⁴ vgl. Fleischer/Barz, S.158 ff.

doppelte Auszeichnung. Generell wurden Präfixe mit nachgestelltem (*be-*, *ent-* usw.), Suffixe mit vorangestelltem Bindestrich (*-chen*, *-ig* usw.) notiert.

Zu jeder Wortbildungskonstituente wurde die entsprechende Wortart (*Verb*, *Substantiv*, *Adjektiv*, *Adverb*, *Präposition*, *Interjektion* oder *Eigennamen*) notiert. Um Redundanzen innerhalb der Datenbank zu vermeiden und entsprechend ein separates Feld in der Tabelle *Link_Wortbildungen* einzusparen, wurden *Partizip*, *Präfix* und *Suffix* ebenfalls als Wortart notiert.

Um den Zugriff auf die Ausgangswortart (*Verb*, *Substantiv* oder *Adjektiv*) von Derivationen und Konversionen zu erleichtern, wurde diese zusätzlich notiert, wenn die Wortart der Konstituente von der Wortart des Wortbildungsproduktes abwich. Zwar ist diese Information in der Datenbank durch die Auszeichnung der Wortart der jeweiligen Konstituenten enthalten und könnte über die Abfrage der Wortart derjenigen Konstituente einer Derivation, die nicht Präfix oder Suffix ist, ermittelt werden, allerdings müssten bei deverbale Ableitungen bzw. Konversionen sowohl Konstituenten mit Wortart *Verb* als auch Konstituenten mit Wortart *Partizip* abgefragt werden, eine separate Auszeichnung erschien daher sinnvoll.

Enthält die Wortbildung ein Fugenelement, so wird dieses bei jeder ausgezeichneten Konstituente notiert. Die Auszeichnung orientiert sich dabei an der Wortbildung in der jeweiligen Frage, nicht an der Standardsprache.

Zu jeder Wortbildung wird außerdem die jeweilige Wortbildungsart notiert. Die Wortbildungsarten entsprechen dabei den Wortbildungsarten bei Fleischer/Barz⁶⁵. In der BSA-Fragekonkordanz finden sich dabei Wortbildungen der folgenden Wortbildungsarten: *Determinativkompositum*, *Kopulativkompositum*, *Possessivkompositum*, *explizite Präfix-/Suffixderivation*, *implizite Derivation*, *kombinierte Präfixkonversion* und *Konversion*. Zusätzlich zu den bei Fleischer/Barz gelisteten Wortbildungsarten wird hier die Zusammenbildung als Wortbildungsart ausgezeichnet. Fleischer/Barz verzichten auf den Begriff *Zusammenbildung* und analysieren explizite Suffixderivationen mit einer Wortgruppe als Basis nicht gesondert von anderen expliziten Suffixderivationen⁶⁶. Barz nimmt jedoch die Zusammenbildung als "Spezialfall der Derivation bei Substantiv und Adjektiv"⁶⁷ an, weshalb sie für die Auszeichnung der Wortbildungen als eigene Wortbildungsart angenommen wird.

Schließlich wird zu jeder Konstituente das jeweilige Wortbildungsprodukt und dessen Wortart (*Verb*, *Substantiv*, *Adjektiv* oder *Adverb*) notiert.

⁶⁵ vgl. Fleischer/Barz, S.44 ff.

⁶⁶ vgl. Fleischer/Barz, S.47

⁶⁷ Barz, S.674

Folgende Abbildung zeigt die Auszeichnung der Wortbildungen am Beispiel der Frage 430 der BSA-Fragekonkordanz, "'Hühnlein' (die kleinen Hühner, Junghühner)":

```
0430;Hühnlein;Substantiv;explizite Suffixderivation;Huhn;Substantiv;-;-
0430;Hühnlein;Substantiv;explizite Suffixderivation;-lein;Suffix;-;-
0430;Hühnlein;Substantiv;explizite Suffixderivation;-chen;Suffix;-;-
```

Abbildung 5.6: Auszeichnung von Wortbildungen

5.3.2. Datenbanktabellen zur Auszeichnung der Wortbildungen

Die Auszeichnungen von Wortbildungen finden sich in der BAYDAT-Datenbank in der Tabelle *Link_Wortbildungen*:

Feld	Datentyp	Inhalt
Nr	Varchar2(4)	laufende Nummer
Fragenr	Char(4)	Nummer der ausgezeichneten Wortbildungskonstituente
Wb_Produkt	Varchar2(500)	Wortbildungsprodukt
Wa_Produkt	Char(1)	Wortart des Wortbildungsprodukts
Wb_Art	Char(1)	Art der Wortbildung
Konst	Varchar2(500)	ausgezeichnete Wortbildungskonstituente
Wa_Konst	Char(1)	Wortart der ausgezeichneten Wortbildungskonstituente
Wa_De	Char(1)	Ausgangswortart bei Ableitungen und Konversionen
Konst_Nr	Char(1)	laufende Nummer der Wortbildungskonstituente
F_Element	Varchar2(5)	Fugenelement in der jeweiligen Wortbildung
Sugg_Form	Char(1)	ausgezeichnete Wortbildungskonstituente innerhalb Suggestierform oder nicht?

Tabelle 5.58: Datenfelder und -typen der Tabelle *Link_Wortbildungen*

Die Tabelle *Link_Wortbildungen* enthält neben der laufenden Nummer der jeweiligen Auszeichnung (*Nr*) eine laufende vierstellige Nummer der jeweiligen Wortbildungskonstituente (*Konst_Nr*). Die laufende Nummer der Wortbildungskonstituente erlaubt einen leichteren Zugriff auf Wortbildungen mit identischen Konstituenten innerhalb der grafischen Benutzeroberfläche der BAYDAT-Datenbank. Das Feld *Sugg_Form* gibt an, ob die ausgezeichnete Wortbildungskonstituente in einer Suggestierform oder im regulären Fragetext enthalten ist.

Obwohl für die Wortbildungen die Sortierreihenfolge bestimmter Feldwerte keine Rolle spielt (anders als bei der Sortierung z.B. der Kasus bei der Auszeichnung der Substantive), werden die Werte für Ausgangswortart bei Ableitungen und Konversionen, Wortarten der Wortbildungskonstituenten und des Wortbildungsprodukts sowie der Wortbildungsart als

Ziffern kodiert. Durch die Reduktion der Ausgangswerte auf Kodierungen mit einer Breite von lediglich einem Zeichen werden Redundanzen innerhalb der Datenbank vermieden:

Feld	Kodierung	Auszeichnung
Wa_De	1	Deverbal
	2	Desubstantivisch
	3	Deadjektivisch
Wa_Produkt	1	Verb
	2	Substantiv
	3	Adjektiv
	4	Adverb
Wb_Art	1	Determinativkompositum
	2	Kopulativkompositum
	3	Possessivkompositum
	4	implizite Ableitung
	5	explizite Präfixableitung
	6	explizite Suffixableitung
	7	Zusammenbildung
	8	Konversion
	9	kombinierte Präfixkonversion
Wa_Konst	1	Verb
	2	Substantiv
	3	Adjektiv
	4	Adverb
	5	Präposition
	6	Interjektion
	7	Eigennamen
	8	Partizip
	9	Präfix
	0	Suffix

Tabelle 5.59: Kodierte Werte in der Tabelle *Link_Wortbildungen*

5.3.3. Beziehungen der Tabelle *Link_Wortbildungen* zu anderen Tabellen der BAYDAT-Datenbank

Die Tabelle *Link_Wortbildungen* ist über das Feld *Fragenr* mit der Tabelle *Fragen* verknüpft, über das Feld *Konst* steht die Tabelle in Verbindung zu den Tabellen zur Auszeichnung morphologischer Phänomene:

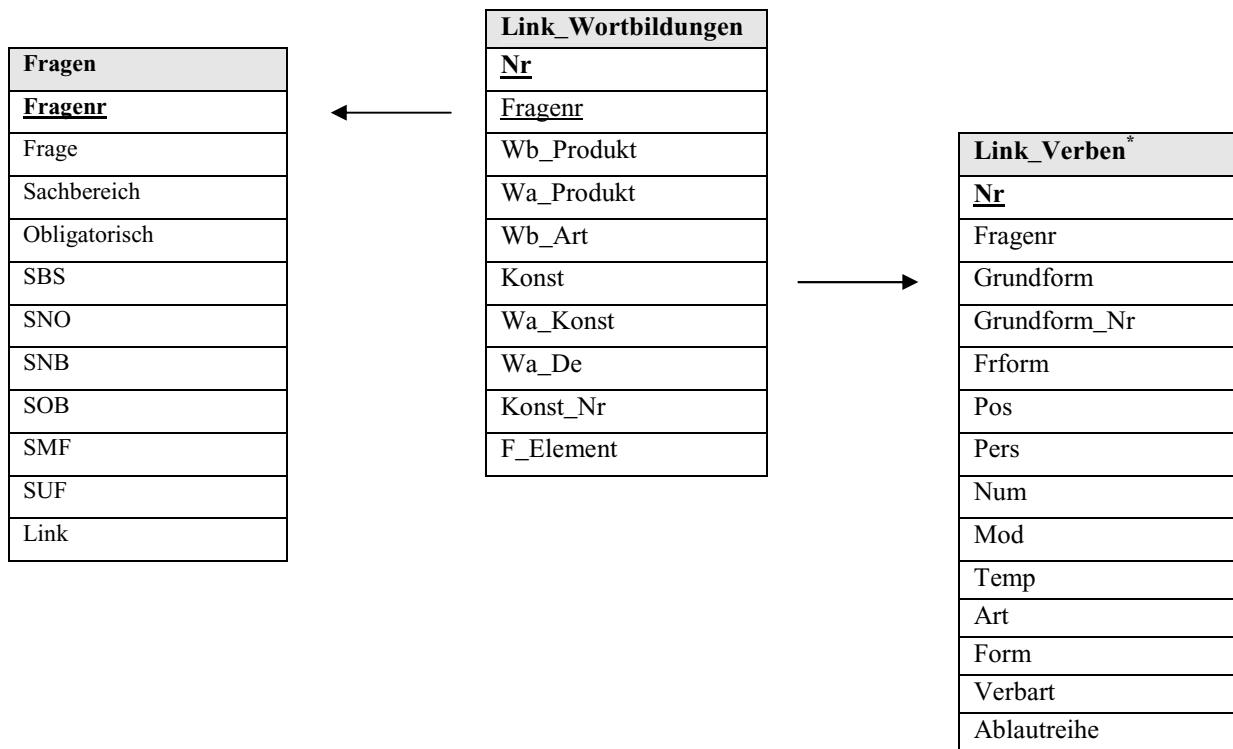


Abbildung 5.7: Schematische Darstellung der Beziehungen der Tabelle *Link_Wortbildungen*

* Die Tabelle *Link_Verben* steht hier exemplarisch für die Beziehung der Tabelle *Link_Wortbildungen* zu den Tabellen zur Auszeichnung morphologischer Phänomene. Über das Feld *Konst* lassen sich Beziehungen zum jeweiligen Feld *Grundform*, das sich in jeder Tabelle zur Auszeichnung morphologischer Phänomene findet, herstellen.

In der Übersicht sind Fremdschlüssel durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

5.3.4. Beispielabfragen

a) Beispielabfrage 1: Diminutivbildungen

Abfrage: Es sollen alle obligatorischen Fragen (SBS-Fragennummer, Fragetext) gefunden werden, die im SBS Diminutiva erheben.

SQL:

```

1   select sbs, frage
2   from fragen
3   where fragenr in (
4         select fragenr
5         from link_wortbildungen
6         where konst in ('-lein', '-chen')
7   ) and obligatorisch = 'j'
8   and sbs is not null
```

Kommentar: Aus der Tabelle *Link_Wortbildungen* werden die Fragennummern derjenigen Fragen ausgewählt, die Diminutiva erheben, also explizite Ableitungen mit dem Suffix *-chen* bzw. *-lein* (*konst in ('-lein', '-chen')*, Z.6; da bei Diminutivbildungen beide Suffixe *-chen* und *-lein* notiert werden, würde die Angabe eines Suffixes genügen: *konst = '-chen'* bzw. *konst = '-lein'*). Werden die entsprechenden Fragen im SBS (*sbs is not null*, Z.8) obligatorisch (*obligatorisch = 'j'*, Z.7) erhoben, werden SBS-Fragennummer und Fragetext als Ergebnis ausgegeben.

SBS	Frage
010.005	"Kälblein"
036.004	"Das Kitzlein"
048.005	"Hühnlein" (die kleinen Hühner, Junghühner)
050.006	"Flügelein"
060.003	"Hündlein"
102.007	"Wägelein" Singular
102.008	"Wägelein" Plural
106.006	"Rädlein" Singular
128.007	"Sträßlein"
164.010	"Fäßlein" Singular
164.011	"Fäßlein" Plural
168.004	"Äpfelchen" Singular
180.005	"Träublein"
198.008	"Äuglein"
204.006	"Bärtlein"
212.004	"Äderchen" Sing.
212.005	"Äderchen" Pl.
264.008	"Vögelein"
312.008	"Blümlein"
340.003	"Gärtlein"
382.011	"Gläslein"
390.011	"Bänklein"
394.001	"Döchtlein"

Tabelle 5.60: Ergebnisse der Abfrage "Diminutivbildungen"

b) Beispielabfrage 2: Determinativkomposita mit der Konstituente *Baum*

Abfrage: Frage 1363 der BSA-Fragekonkordanz erhebt das Substantiv *Baum*. Ausgehend von dieser Frage sollen alle Fragen (SMF-Fragennummer und Fragetext) ermittelt werden, die im SMF obligatorisch Determinativkomposita mit der Konstituente *Baum* erheben.

SQL:

```

1  select smf, frage
2  from fragen
3  where fragenr in (
4      select fragenr
5      from link_wortbildungen

```

```

6         where konst = (
7             select grundform
8             from link_substantive
9             where fragenr = '1363'
10        ) and wb_art = '1'
11    ) and obligatorisch = 'j'
12    and smf is not null

```

Kommentar: Die innerste *select*-Anweisung (Z.7 – 9) ermittelt zunächst die Grundform, die in der Tabelle *Link_Substantive* für Frage 1363 notiert ist (=Baum). Anschließend werden aus der Tabelle *Link_Wortbildungen* die Fragenummern derjenigen Fragen ausgewählt, die Determinativkomposita (*wb_art* = '1', Z.10) mit der entsprechenden Wortbildungskonstituente *Baum* enthalten. Wurden die jeweiligen Fragen im SMF (*smf is not null*, Z.12) obligatorisch (*obligatorisch* = 'j', Z.11) erhoben, werden SMF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:

SMF	Frage
054.008	"Leiterbaum"
069.008	"Weidenbaum"
160.003	"Baumwolle/"baumwollen (Substantiv oder Adjektiv)
086.006	"Hinter unserem Haus steht ein Apfelbaum mit schönen roten Äpfeln."

Tabelle 5.61: Ergebnisse der Abfrage "Determinativkomposita mit der Konstituente *Baum*"

c) Beispielabfrage 3: Komposita im Sachbereich *Gemüse*

Abfrage: Es sollen alle Fragen (SUF-Fragennummer und Fragetext) ermittelt werden, die im SUF Komposita (Determinativ-, Kopulativ- oder Possessivkomposita) aus dem Sachbereich *Gemüse* erheben.

SQL:

```

1     select suf, frage from fragen
2     where fragenr in (
3         select fragenr
4         from link_wortbildungen
5         where wb_art in ('1', '2', '3')
6     ) and sachbereich = (
7         select sachbereichnr
8         from sachbereiche
9         where sachbereich = 'Gemüse'
10    ) and suf is not null
11    and obligatorisch = 'j'

```

Kommentar: Aus der Tabelle *Link_Wortbildungen* werden zunächst die Fragenummern aller Fragen ermittelt, die Komposita erheben (*wb_art in ('1', '2', '3')*, Z.5). Wurden die Fragen im Sachbereich *Gemüse* (die entsprechende Sachbereichnummer wird über die Abfrage Z.7-9 ermittelt) im SUF (*suf is not null*, Z.10) obligatorisch (*obligatorisch* = 'j', Z.11) erhoben, werden SUF-Fragennummer und Fragetext als Ergebnis ausgegeben.

Ergebnis:	SUF	Frage
	342.012	"Sauerkraut"
	346.013	"Ein Kopf Blaukraut"
	346.014	"Ein Kopf Weißkraut"

Tabelle 5.62: Ergebnisse der Abfrage "Komposita im Sachbereich Gemüse"

d) Beispielabfrage 4: Deverbale Suffixableitungen mit Suffix -er

Abfrage: Es sollen alle Fragen ermittelt werden (Fragennummer und -text), die obligatorisch deverbale Suffixableitungen mit dem Suffix *-er* erheben.

SQL:

```

1  select fragenr, frage
2  from fragen
3  where fragenr in (
4      select fragenr
5      from link_wortbildungen
6      where konst = '-er'
7      and wa_de = '1'
8      and wb_art = '6'
9  ) and obligatorisch = 'j'
```

Kommentar: Aus der Tabelle *Link_Wortbildungen* werden die Fragennummern der Fragen ausgewählt, die deverbale (*wa_de = '1', Z.7*) Suffixableitungen (*wb_art = '6', Z.8*) mit dem Suffix *-er* (*konst = '-er', Z.6*) erheben. Wurden die entsprechenden Fragen in der BSA-Fragekonkordanz obligatorisch (*obligatorisch = 'j', Z.9*) erhoben, werden Fragennummer und -text als Ergebnis ausgegeben.

Ergebnis:	Fragenr	Frage
	0601	"Mäher" Woher kamen die Mäher?
	2268	"Lehrer" (Wer unterrichtet die Kinder in der Schule?)
	2572	"Der Lehrer weiß auch nicht alles"
	3697	"Das sag ich dem Lehrer"
	4191	"Das Mädchen ist schlecht in der Schule. Wenn es der Lehrer etwas fragt, dann weiß es nichts."
	4192	"Pächter"

Tabelle 5.63: Ergebnisse der Abfrage "Deverbale Suffixableitungen mit Suffix -er"

6. Zusätzliche Elemente

Neben den unmittelbaren Erhebungsdaten der BSA-Teilprojekte bietet sich die BAYDAT-Datenbank als Plattform für die Speicherung weiterer Daten, die sich mittelbar aus den Dialekterhebungen und der Arbeit an den Atlasbänden ergaben, an. Zum jetzigen Zeitpunkt sind in der BAYDAT-Datenbank Scans der Originalfragebücher aus dem SUF sowie Fotografien, die bei den Erhebungen für den SUF vor Ort entstanden, gespeichert. Im Folgenden wird die Integration dieser zusätzlichen Elemente in die BAYDAT-Datenbank beschrieben. Die BAYDAT-Datenbank soll auch in Zukunft für die Speicherung weiterer zusätzlicher Elemente offen stehen.

6.1. Scans der Originalfragebücher

Die Fragebücher mit den Erhebungen des SUF wurden zum Zweck der zukunftsicheren Archivierung eingescannt und liegen am UNTERFRÄNKISCHEN DIALEKTINSTITUT im TIFF- sowie GIF-Format auf DVD-ROM vor. Um diese Bilddateien einerseits besser recherchierbar zu machen und sie andererseits der Öffentlichkeit als unmittelbaren Beleg für die elektronische Erfassung zugänglich zu machen, wurden sie in die BAYDAT-Datenbank integriert.

6.1.1. Beschreibung der Datenbanktabellen

Um die Bilddateien mit ihren jeweiligen Ausgangsbelegen zu verknüpfen, sind zwei Tabellen notwendig: die Tabelle *Fragebuch_Scans*, die die gescannten Fragebuchseiten enthält und die Tabelle *map_scans*, die die Bilddateien mit den Ausgangsbelegen verknüpft.

a) Die Tabelle *Map_Scans*

Feld	Datentyp	Inhalt
Fragenr	Char(7)	SUF-Fragennummer
Auflage	Char(1)	Auflage des SUF-Fragebuchs
Seite	Char(3)	laufende Seitennummer im SUF-Fragebuch

Tabelle 6.1: Datenfelder und -typen der Tabelle *Map_Scans*

Da das Fragebuch des SUF in zwei in Satz und Umfang variierenden Auflagen vorliegt, wird die Tabelle *Map_Scans* benötigt, um einem gegebenen Ort und einer gegebenen Frage die entsprechende Fragebuchseite zuzuordnen. Das Feld *Auflage* aus der Tabelle *Orte* korrespondiert zu diesem Zweck mit dem Feld *Auflage* in *Map_Scans*, das Feld *Suf* aus der Tabelle *Fragen* entspricht dem Feld *Fragenr* in der Tabelle *Map_Scans*, die entsprechende Fragebuchseite ist im Feld *Seite* der Tabelle *Map_Scans* festgehalten.

b) Die Tabelle *Fragebuch_Scans*

Feld	Datentyp	Inhalt
Nr	Char(12)	Kombination aus Ortssigle und laufender Seitennummer im SUF-Fragebuch
Scan	Blob	Bilddatei der jeweiligen Fragebuchseite
Type	Varchar2(20)	Dateityp der jeweiligen Bilddatei

Tabelle 6.2: Datenfelder und -typen der Tabelle *Fragebuch_Scans*

Das Feld *Nr* enthält eine Kombination aus Ortssigle und der laufenden Seitennummer aus dem Fragebuch des SUF. Nachdem über das Feld *Auflage* aus der Tabelle *Orte* über *Map_Scans* die Seitennummer ermittelt wurde, kann über eine Kombination der Seitennummer und der aus der Tabelle *Orte* zu ermittelnden Ortssigle – diese Kombination entspricht dem Wert des Feldes *Nr* in der Tabelle *Fragebuch_Scans* – die Bilddatei, die im Feld *Scan* der Tabelle *Fragebuch_Scans* abgelegt ist, zu einer gegebenen Frage ausgegeben werden. Das Feld *Type* gibt den Dateityp der im Feld *Scan* gespeicherten Datei an und ist für eine Anzeige im Webbrowser notwendig.

6.1.2. Beziehungen der Tabellen zu den Scans der Originalfragebücher

Die Beziehung zwischen den Tabellen *Map_Scans* und *Fragebuch_Scans* lässt sich schematisch folgendermaßen darstellen:



Abbildung 6.1: Schematische Darstellung der Beziehung der Tabellen *Map_Scans* und *Fragebuch_Scans*

Primärschlüssel sind in der Abbildung durch Fettdruck und Unterstreichung gekennzeichnet.

6.1.3. Beziehungen der Tabellen zu den Scans der Originalfragebücher zu anderen Tabellen der BAYDAT-Datenbank

Über eine Kreuzabfrage der Tabellen *Map_Scans*, *Fragebuch_Scans*, *Fragen* und *Orte* lassen sich die Scans der Fragebuchseite zu einer gegebenen Frage an einem gegebenen Erhebungsort ermitteln bzw. anzeigen. Die Beziehungen der Tabellen *Orte* und *Fragen* zu den Tabellen *Fragebuch_Scans* und *Map_Scans* lassen sich schematisch darstellen:

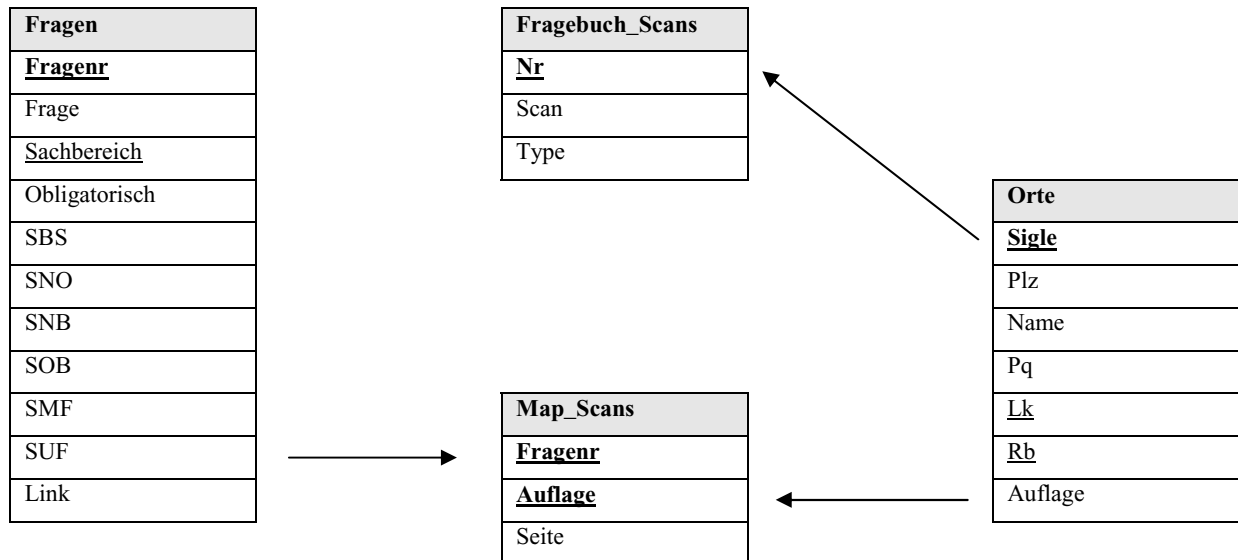


Abbildung 6.2: Schematische Darstellung der Tabellen *Map_Scans* und *Fragebuch_Scans* zu anderen Tabellen der BAYDAT-Datenbank

Primärschlüssel sind in der Abbildung durch Fettdruck und Unterstreichung gekennzeichnet, Fremdschlüssel durch Unterstreichung.

6.1.4. Beispielabfrage zu den Scans der Originalfragebücher

Abfrage: Es sollen Belege zu Frage SUF 2.3 "Orte der Umgebung" aus dem SUF-Erhebungsort Leubach (Belege in TEUTHOBD, Lemmata) gefunden werden. Zusätzlich soll der entsprechende Wert aus dem Feld *Nr* der Tabelle *Fragebuch_Scans* angezeigt werden.

SQL:

```

1  select teuthobd, lemma, nr
2  from fragebuch_scans, belege
3  where ortssigle = (
4      select sigle
5      from orte
6      where name = 'Leubach'
7  ) and frage_nr = (
8      select suf
9      from fragen
10     where frage = 'Orte der Umgebung'
11  ) and nr = (
12     select sigle
13     from orte
14     where name = 'Leubach'
15  ) || (
16     select seite
17     from map_scans
18     where auflage = (
19         select auflage
20         from orte
21         where name = 'Leubach'
22     ) and fragenr = (
23         select suf
24         from fragen
25         where frage = 'Orte der Umgebung'
26     ))

```

Kommentar: Zunächst werden aus der Tabelle *Belege* die Zeilen ausgewählt, die in der Spalte *Ortssigle* die Ortssigle des SUF-Erhebungsortes Leubach enthalten (Z.3 – 6). Außerdem müssen die ausgegebenen Zeilen im Feld *Frage_Nr* die SUF-Fragennummer zur Frage "Orte der Umgebung" enthalten (Z.7 – 10). Die Abfrage zwischen Z.11 und 27 wählt aus der Tabelle *Fragebuch_Scans* die Zeile aus, die die Nummer der gescannten Fragebuch-Seite enthält. Diese setzt sich zusammen aus der Ortssigle für den SUF-Erhebungsort Leubach (Z.11 – 14) und der Seitennummer aus der Tabelle *Map_Scans* (Z.15 – 26), die der für Leubach verwendeten Auflage des SUF-Fragebuchs (Z.18 – 21) und der entsprechenden Fragennummer (Z.22 – 25) entspricht. Zusammengesetzt werden die Werte für *Sigle* (Z.12) und *Seite* (Z.16) durch den Operator *'/'* (Z.15). Als Ergebnis werden die Belege in TEUTHOBD, die entsprechenden Lemmata und die Nummer des Scans ausgegeben (Z.1).

Ergebnis:

TeuthoBD	Lemma	Nr
vr̄āngem̄ə	FRANKENHEIM	suf001leu001
rūdšwin̄ə	RUEDENSCHWINDEN	suf001leu001
ōw̄er̄vl̄ōdinḡə	OBERFLADUNGEN	suf001leu001
dišd̄ə ^(a) d	DIE STADT	suf001leu001
h̄ovl̄əš	HUFLAR	suf001leu001

Tabelle 6.3: Ergebnisse der Beispielabfrage zu den Scans der Originalfragebücher

Bei Verwendung entsprechender Software, z.B. einem Webbrowser, lässt sich die Bilddatei der entsprechenden Seite darstellen, ein Ausdruck des Scans aus o.g. Abfrage findet sich in Anhang A.

6.2. Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS

Während der Erhebungen zum SUF wurden von den Exploratorinnen und Exploratoren zahlreiche Fotografien aufgenommen, die u.a. landwirtschaftliche Geräte, althergebrachte Arbeitsabläufe und Einrichtungsgegenstände zeigen. Diese Aufnahmen wurden am UNTERFRÄNKISCHEN DIALEKTINSTITUT als Kontaktabzüge archiviert. Für eine zukunftssichere Archivierung in digitaler Form werden die Kontaktabzüge gescannt, die Fotografien können anschließend über die Online-Schnittstelle der BAYDAT-Datenbank der Allgemeinheit zugänglich gemacht werden, indem sie mit den zugehörigen Fragen aus der BSA-Fragekonkordanz verknüpft werden. Die ersten Bilddateien im Format JPEG sind bereits in die BAYDAT-Datenbank integriert.

6.2.1. Beschreibung der Datenbanktabellen

Die Bilddateien sind mit jeweils relevanten Fragen aus der BSA-Fragekonkordanz verknüpft, die Aufnahme eines Taufsteins ist z.B. mit Frage 3673 der BSA-Fragekonkordanz, "Taufe" verknüpft. Um die Verknüpfung zwischen BSA-Fragekonkordanz und den Bilddateien zu ermöglichen, sind zwei Tabellen notwendig: eine Tabelle *Link_Bilder*, die die BSA-Fragekonkordanz mit den jeweiligen Bilddateien in Verbindung setzt, und eine Tabelle *Bildarchiv*, die die Bilddateien enthält.

a) Die Tabelle *Link_Bilder*

Feld	Datentyp	Inhalt
Fragenr	Char(4)	Fragennummer
Name	Varchar2(500)	Name der Bilddatei

Tabelle 6.4: Datenfelder und -typen der Tabelle *Link_Bilder*

Die Tabelle *link_bilder* enthält im Feld *fragenr* die Fragennummer aus der BSA-Fragekonkordanz, im Feld *Name* den Namen der mit der jeweiligen Frage verknüpften Bilddatei. Durch die Verknüpfung über die Tabelle *link_bilder* können mehrere Bilddateien mit einer Frage verknüpft werden, ohne dass dabei Redundanzen durch mehrfache Aufnahme einzelner Bilddateien entstehen. So enthält das Bildarchiv zurzeit zwei Bilddateien zur BSA-Frage 2878 "Wäscheklammern", *Waescheklammer.jpg* sowie *Waschpfloecke.jpg*. Umgekehrt können einzelne Bilder mit mehreren Fragen verknüpft werden.

b) Die Tabelle *Bildarchiv*

Feld	Datentyp	Inhalt
Name	Varchar2(500)	Name der Bilddatei
Type	Varchar2(20)	Dateityp der Bilddatei
Breite	Varchar2(6)	Breite der Bilddatei in Pixeln
Hoehe	Varchar2(6)	Höhe der Bilddatei in Pixeln
Thumb	Blob	verkleinertes Vorschaubild für die jeweilige Bilddatei
Bild	Blob	Bilddatei im Vollformat

Tabelle 6.5: Datenfelder und -typen der Tabelle *Bildarchiv*

Die Tabelle *Bildarchiv* enthält Felder für den Namen der Bilddatei im Feld *Name*, den Dateityp im Feld *Type* und die tatsächliche Bilddatei im Feld *Bild*. Zusätzlich enthält sie Felder zur Angabe der Pixelmaße der jeweiligen Datei in den Feldern *Breite* und *Höhe* sowie ein Feld *Thumb* zur Ablage eines verkleinerten Vorschaubildes. Diese drei Felder ermöglichen es, in einer Vorschau den Benutzer der grafischen Benutzeroberfläche der BAYDAT-Datenbank vor dem Download der tatsächlichen Bilddatei über deren Größe und Inhalt zu informieren.

6.2.2. Beziehungen der Tabellen zu den Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS

Die Beziehung zwischen den Tabellen *Link_Bilder* und *Bildarchiv* lassen sich schematisch folgendermaßen darstellen:

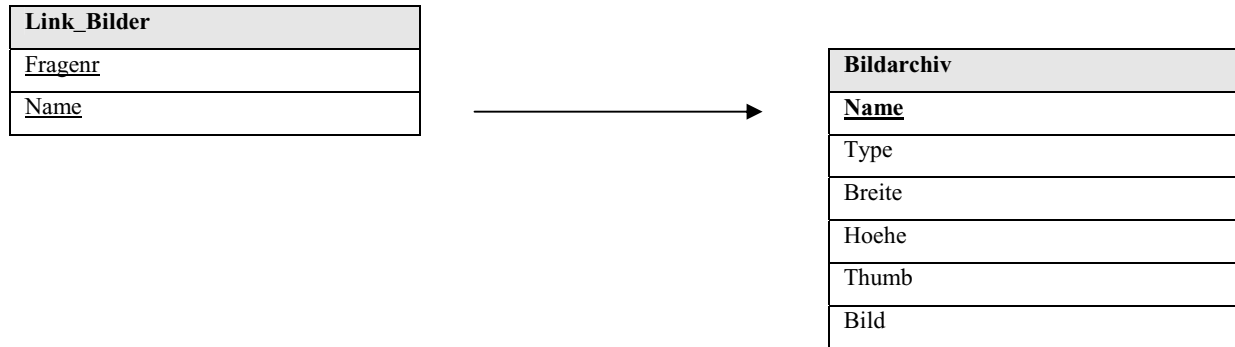


Abbildung 6.3: Schematische Darstellung der Beziehung der Tabellen *Link_Bilder* und *Bildarchiv*

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Fettdruck und Unterstreichung.

6.2.3. Beziehungen der Tabellen zu den Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS zu anderen Tabellen der BAYDAT-Datenbank

Über eine Kreuzabfrage der Tabellen *Fragen*, *Link_Bilder* und *Bildarchiv* lassen sich die mit einer Frage verknüpften Bilddateien ermitteln bzw. – eine geeignete Software vorausgesetzt – anzeigen. Die Beziehungen der Tabellen *Link_Bilder* und *Bildarchiv* zur Tabelle *Fragen* lassen sich schematisch darstellen:

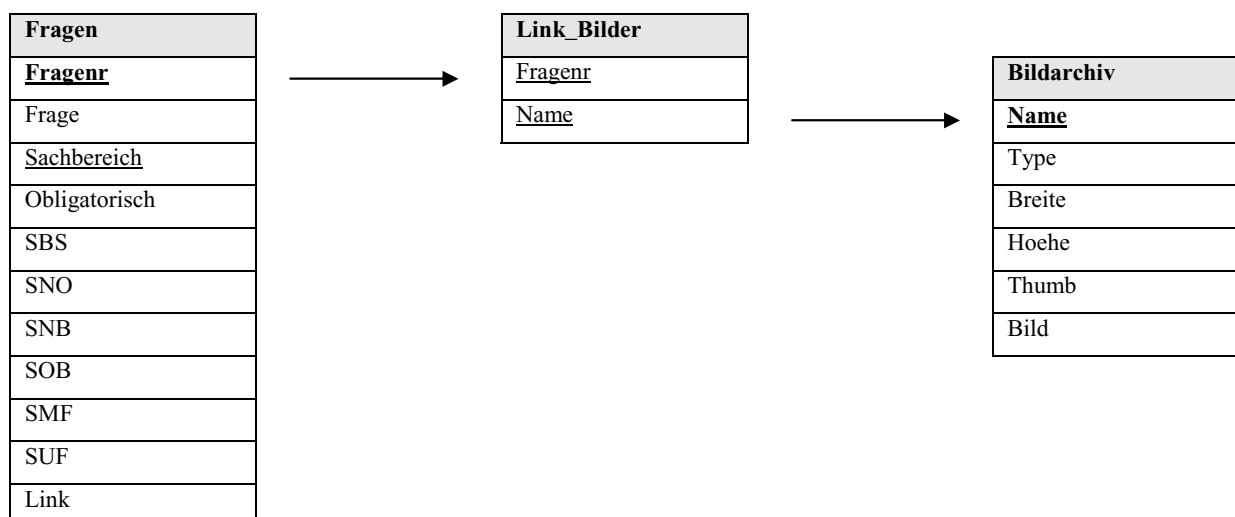


Abbildung 6.4: Schematische Darstellung der Tabellen *Link_Bilder* und *Bildarchiv* zu anderen Tabellen der BAYDAT-Datenbank

Fremdschlüssel sind in der Abbildung durch Unterstreichung gekennzeichnet, Primärschlüssel durch Unterstreichung und Fettdruck.

6.2.4. Beispielabfrage zu den Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS

Abfrage: Es sollen Fragen aus dem Sachbereich *Pferd* (Fragennummer, -text, Name der Bilddatei) gefunden werden, die im SUF obligatorisch erhoben wurden und mit Bilddateien verknüpft sind.

SQL:

```

1      select fragen.fragenr, frage, bildarchiv.name
2      from fragen, link_bilder, bildarchiv
3      where sachbereich in (
4          select sachbereichnr
5          from sachbereiche
6          where sachbereich = 'Pferd'
7      ) and fragen.fragenr in (
8          select fragenr
9          from link_bilder
10     ) and link_bilder.name = bildarchiv.name
11     and link_bilder.fragenr = fragen.fragenr
12     and suf is not null
13     and obligatorisch = 'j'
```

Kommentar: Durch eine Kreuzabfrage der Tabellen *Fragen*, *Link_Bilder* und *Bildarchiv* werden Fragen ermittelt, die im SUF (*suf is not null*, Z.13) im Sachbereich *Pferd* (Z.3 – 6) obligatorisch (*obligatorisch = 'j'*, Z.13) erhoben wurden. Sind zu den so ermittelten Fragen Bilddateien vorhanden (Z. 7 – 9), werden Fragennummer, -text und Name der Bilddatei als Ergebnis ausgegeben. Die Anweisungen in Z.10 bzw. Z.11 verhindern, dass als Ergebnis das kartesische Produkt aus den Tabellen *Fragen*, *Link_Bilder* und *Bildarchiv* ausgegeben wird.

Ergebnis:

Fragenr	Frage	Name
0503	"Kummet" (Genus)	Kummet1
0503	"Kummet" (Genus)	Kummet2
0503	"Kummet" (Genus)	Pferdefuhrwerk
0507	"Joch"	Ankenjoch

Tabelle 6.6: Ergebnisse der Beispielabfrage zu den Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS

7. Zusammenfassendes zum Aufbau der Datenbank

Beim Aufbau der Datenbank wurde darauf geachtet, je nach Kontext zusammengehörende Informationen auf mehrere Tabellen zu verteilen. Dies geschah einerseits aus Gründen der Redundanzvermeidung. Wären beispielsweise die Regierungsbezirke im Klartext in die Spalte *Rb* der Tabelle *Orte* (vgl. 4.2.1.) aufgenommen worden, hätte dies zu Redundanzen geführt: Statt z.B. des Kürzels *uf* für den Regierungsbezirk Unterfranken hätte für jeden unterfränkischen Erhebungsort in der Spalte *Rb* der Wert *Unterfranken* angegeben werden müssen. Pro Erhebungsort wären also zehn zusätzliche Zeichen nötig gewesen.

Außerdem konnten durch die Aufteilung von Informationen auf mehrere Tabellen Fehleingaben vermieden werden. Da z.B. die Spalte *Rb* der Tabelle *Orte* als Fremdschlüssel auf die Spalte *Id* der Tabelle *Regbez* verweist (vgl. 4.2.2.), kann sie lediglich die Werte *uf* (=Unterfranken), *mf* (=Mittelfranken), *ob* (=Oberbayern), *nb* (=Niederbayern) und *no* (=SNOB, Oberfranken und Oberpfalz) enthalten. Sollten die Quelldaten beim Einlesen der Datenbank abweichende Werte enthalten, würde die entsprechende Zeile unter Ausgabe einer Fehlermeldung nicht in die Datenbank übernommen, der Fehler könnte anschließend korrigiert werden.

Der hier gewählte Ansatz entspricht damit dem von Edgar Frank Codd vorgeschlagenen Normalisierungsprozess, der einerseits der Redundanzvermeidung, andererseits dem Vermeiden von Fehleingaben innerhalb von Datenbanken dient.⁶⁸

Bei der Auszeichnung der Fragen unter sprachwissenschaftlichen Gesichtspunkten erschienen zunächst zwei Möglichkeiten des Tabellendesigns möglich: Eine zusammenfassende Tabelle, die Auszeichnungen sämtlicher sprachwissenschaftlicher Phänomene zusammenfasst sowie die Aufteilung der Auszeichnungen auf mehrere Tabellen. Ersterer Ansatz erwies sich aus mehreren Gründen als nicht praktikabel:

Redundanz

Wären die Tabellen zu den Auszeichnungen unter sprachwissenschaftlichen Gesichtspunkten zu einer Tabelle zusammengefasst worden, hätte dies zu Redundanzen innerhalb der Datenbank geführt. Z.B. enthält die Tabelle *Link_Verben* Felder für Tempus und Modus (vgl. 5.2.1.2.), die Tabelle *Link_Substantive* enthält Felder für Kasus und Genus (vgl. 5.2.2.2.). Bei einer Zusammenfassung beider Tabellen blieben bei der Auszeichnung von Substantiven die Felder für Tempus und Modus leer, umgekehrt

⁶⁸ vgl. Connolly/Begg, S.375 ff., Warrander, S.43 ff.

können für Verben keine Werte für die für Kasus und Genus vorgesehen Felder angegeben werden.

Inhaltliche Gründe

Bei der Zusammenfassung sämtlicher Tabellen zur Auszeichnung sprachwissenschaftlicher Phänomene würde eine einzige Tabelle entstehen, die sämtliche in Kapitel 5 aufgeführten Felder enthielte. Eine Zeile, die die Auszeichnungen eines Lautes enthält, würde Felder zur Morphologie enthalten, eine Zeile, die Auszeichnungen eines Verbs enthält, würde, wie oben gezeigt, Felder zu Kasus und Genus enthalten usw. Zwar würde die Funktionalität der Datenbank hierdurch nicht beeinträchtigt, dennoch könnten sich durch eine Tabellenstruktur, die nicht den tatsächlichen, inhaltlichen Gegebenheiten entspricht, Probleme ergeben. Würde die Tabelle in ihrer Gesamtheit oder in Teilen unkommentiert, z.B. im Rahmen einer XML-Ausgabe, ausgegeben werden, könnte dies zu Unklarheiten auf Nutzerseite führen.

Daneben ergibt sich durch die Aufteilung der Auszeichnungen auf mehrere Tabellen ein entscheidender Vorteil: Sollen die Auszeichnungen innerhalb der BAYDAT-Datenbank in Zukunft erweitert werden, können ohne weiteres entsprechende Tabellen hinzugefügt werden, ohne dass bestehende Tabellen geändert werden müssten. Denkbar wäre z.B. eine Auszeichnung der Fragen unter kulturwissenschaftlichen Gesichtspunkten. Diese könnten in einer separaten Tabelle erfasst werden, andernfalls müssten einer gemeinsamen Tabelle für sämtliche Auszeichnungen Spalten hinzugefügt werden. Dies würde auch zu technischen Problemen führen, da Skripten, die zur Überführung der Daten in die BAYDAT-Datenbank erstellt wurden, entsprechend angepasst werden müssten.

8. Die Online-Anbindung der BAYDAT-Datenbank

Um die Erhebungsdaten des BSA einem möglichst großen Benutzerkreis – von Fachwissenschaftlern bis hin zu interessierten Laien – zugänglich zu machen, galt der Entwicklung einer Onlineschnittstelle für die BAYDAT-Datenbank mit einer einfach zu bedienenden grafischen Benutzeroberfläche von Projektanfang an ein großes Interesse. Folgendes Kapitel beschreibt die technische Realisierung der Onlineschnittstelle mithilfe der Middleware COCOON, die Umsetzung der TEUTHONISTA-Lautschrift für die Anzeige im Webbrowser und den Aufbau sowie die Funktionalität der grafischen Benutzeroberfläche der BAYDAT-Datenbank.

8.1. COCOON – Eine Middleware auf XML-Basis

Klassische Datenbank Anwendungen bauen normalerweise auf einer zweischichtigen Client-Server-Architektur auf⁶⁹: Über einen so genannten Client, in der Regel ein Desktoprechner, wird auf einen Datenbankserver zugegriffen. Der Client ist dabei für die Darstellung der Benutzeroberfläche und die Ausführung der Programmlogik, z.B. zur Erstellung und Ausführung benutzergenerierter SQL-Abfragen, zuständig. Benutzeroberfläche und Programmlogik sind dabei in den meisten Fällen in proprietäre Softwarelösungen integriert, die vom Datenbanknutzer beschafft, installiert und im Zweifelsfall gewartet werden müssen. Beispiel für eine entsprechende Software ist ORACLE SQLPLUS, ein rudimentärer Texteditor, der es erlaubt, SQL-Abfragen zu erstellen und entsprechende Ergebnisse anzuzeigen.

Onlineanwendungen benutzen normalerweise eine dreischichtige Architektur: Zwischen dem Datenbankserver und dem Client wird eine mittlere Schicht, ein so genannter Application Server⁷⁰, auch Middleware⁷¹ genannt, eingefügt. Der Application Server übernimmt dabei die Programmlogik, er stellt z.B. auf Nutzereingaben hin SQL-Abfragen zusammen, die anschließend an den Datenbankserver gesendet werden und bereitet die entsprechenden Ergebnisse für die Anzeige auf. Der Zugriff erfolgt auf Nutzerseite üblicherweise über einen Webbrowser, der zur Anzeige der Benutzeroberfläche genutzt wird. In der Regel wird zur Erstellung der Benutzeroberfläche die Auszeichnungssprache HTML benutzt. Diese erleichtert einerseits aufgrund ihrer leichten Erlernbarkeit die Programmierung der grafischen Benutzeroberfläche, andererseits ist aufgrund der Tatsache, dass HTML-Seiten in jedem Webbrowser anzeigbar sind, keine Installation spezieller Software auf dem Client-Rechner notwendig.

⁶⁹ *Two-Tier client-server architecture*, vgl. Connolly/Begg S.955 f.

⁷⁰ vgl. Connolly/Begg, S.956

⁷¹ vgl. Langham/Ziegeler, S.14

Die Online-Anbindung der BAYDAT-Datenbank erfolgt über die von der APACHE SOFTWARE FOUNDATION⁷² veröffentlichte Middleware COCOON. COCOON bietet gegenüber anderen Lösungen folgende Vorteile:

a) COCOON ist ein OPEN-SOURCE-Projekt

COCOON ist ein OPEN-SOURCE-Projekt, für den Einsatz von COCOON müssen keine Lizenzgebühren entrichtet werden.⁷³ Außerdem ist der COCOON-Quellcode offen zugänglich und kann bei Bedarf unter Einhaltung der APACHE LICENSE⁷⁴ den eigenen Bedürfnissen angepasst werden.

b) COCOON ist plattformunabhängig

COCOON wurde in der von SUN MICROSYSTEMS⁷⁵ veröffentlichten Programmiersprache JAVA entwickelt. Im Gegensatz zu anderen Programmiersprachen ist JAVA nicht systemabhängig, d.h. in JAVA entwickelte Programme laufen ohne Veränderungen des Quellcodes auf unterschiedlichen Betriebssystemen, z.B. MICROSOFT WINDOWS XP oder LINUX. Zum Einsatz von COCOON muss auf dem jeweiligen Webserver lediglich eine so genannte SERVLET ENGINE installiert sein, die es ermöglicht, JAVA-Code auf dem Webserver auszuführen.

c) COCOON ermöglicht die Integration unterschiedlicher Datenquellen

COCOON kann Daten aus unterschiedlichen Quellen beziehen und diese für die Anzeige auf dem jeweiligen Client-Rechner aufbereiten. Als Datenquellen können dabei u.a. auf dem Webserver gespeicherte XML-Dokumente dienen, über die JDBC-Schnittstelle⁷⁶ kann jedoch auch auf beliebige Datenbanken, die die Schnittstelle unterstützen (u.a. ORACLE) zugegriffen werden.

d) COCOON ermöglicht die Ausgabe von Ergebnissen in unterschiedlichen Formaten

COCOON basiert auf der Auszeichnungssprache XML. Diese ähnelt in ihrer Struktur der für die Erstellung von Webseiten verwendeten Auszeichnungssprache HTML, erlaubt jedoch das Erstellen eigener Vokabulare zur inhaltlichen Gliederung von Daten. Während sich aus einer HTML-Tabelle keinerlei Informationen über den Inhalt der einzelnen Felder entnehmen lassen

⁷² APACHE SOFTWARE FOUNDATION. <<http://www.apache.org>> (29.5.2006)

⁷³ vgl. Rosen

⁷⁴ Apache License. Version 2.0. <<http://www.apache.org/licenses/LICENSE-2.0>>

⁷⁵ SUN MICROSYSTEMS. <<http://www.sun.com>> (29.5.2006)

⁷⁶ JAVA DATABASE CONNECTIVITY, vgl. Connolly/Bogg, S.978 ff.

(vgl. Abbildung 8.1), können entsprechende Informationen durch die Verwendung von XML sichtbar gemacht werden (vgl. Abbildung 8.2):

<pre><table> <tr> <td>suf149wue</td> <td>97074</td> <td>Würzburg</td> </tr> </table></pre>	<pre><ort> <sigle>suf149wue</sigle> <plz>97074</plz> <name>Würzburg</name> </ort></pre>
--	---

Abbildung 8.1: Ausgabe eines BSA-Erhebungsortes in Form einer HTML-Tabelle

Abbildung 8.2: XML-formatierte Ausgabe eines BSA-Erhebungsortes

Beim Zugriff auf die BAYDAT-Datenbank erzeugt COCOON zunächst ein XML-Dokument, das in seiner Struktur dem Beispiel aus Abbildung 8.2 entspricht. Der Zugriff erfolgt dabei über XSP⁷⁷, ein XML-Vokabular, das die Verwendung der vollständigen JAVA-Logik innerhalb COCOONS erlaubt. Dies ermöglicht einerseits die Erstellung komplexer SQL-Abfragen auf die BAYDAT-Datenbank, andererseits kann XSP in der Zukunft für die Realisierung weiterer Anwendungen, die auf den Daten der BAYDAT-Datenbank basieren, genutzt werden.

Für die Anzeige auf dem Client-Rechner wird das entsprechende Dokument weiter aufbereitet. Hierzu steht u.a. XSLT⁷⁸ zur Verfügung. XSLT ist ein vom WORLD WIDE WEB CONSORTIUM⁷⁹ empfohlenes XML-Vokabular⁸⁰, das es ermöglicht, XML-Dokumente in andere Formate zu transformieren. Innerhalb der BAYDAT-Online-Anbindung wird XSLT genutzt, um HTML-Seiten zu erzeugen, prinzipiell kann XSLT jedoch zur Erzeugung zahlreicher Formate benutzt werden, z.B. ließen sich RTF-Dokumente oder JAVA-Code erzeugen. Außerdem lassen sich mithilfe des XML-Vokabulars XSL-FO, ebenfalls eine Empfehlung des WORLD WIDE WEB CONSORTIUMS⁸¹, XML-Dokumente in PDF-Dateien umformen. Innerhalb der Online-Lösung der BAYDAT-Datenbank wird XSL-FO verwendet, um Beleglisten im PDF-Format zu erzeugen. Die zur Weiterverarbeitung von XSLT- bzw. XSL-FO-Stylesheets benötigte Software ist bereits in COCOON integriert, u.a. verwendet COCOON den von der APACHE SOFTWARE FOUNDATION veröffentlichten XSLT-Prozessor XALAN⁸². Allerdings können die mitgelieferten Komponenten bei Bedarf durch eigene Lösungen ersetzt werden.

⁷⁷ EXTENSIBLE SERVER PAGES, vgl. Langham/Ziegeler, S.296 ff.

⁷⁸ EXTENSIBLE STYLE LANGUAGE FOR TRANSFORMATIONS, vgl. Tidwell, S.1 ff.

⁷⁹ World Wide Web Consortium. <<http://www.w3.org>>

⁸⁰ vgl. Clark

⁸¹ EXTENSIBLE STYLESHEET LANGUAGE FORMATTING OBJECTS, vgl. Adler et al.

⁸² vgl. Langham/Ziegeler, S.39

Vor allem die Nutzung von XML-Technologien macht COCOON zu einer sinnvollen Softwarelösung für die Realisierung der BAYDAT-Onlineanbindung: XML hat sich mittlerweile als Standard für den Datenaustausch im Internet entwickelt, außerdem ist XML im Vergleich zu anderen Technologien recht einfach zu erlernen, ebenso die damit verbundenen Vokabulare XSP, XSLT und XSL-FO.

8.2. Umsetzung der kodierten Belege in TEUTHOBD/TEUTHOUC

Die in TEUTHONISTA-Lautschrift transkribierten Belege wurden, wie unter 2.2.2. gezeigt, für die elektronische Weiterverarbeitung in eine ASCII-konforme Kodierung umgesetzt. Der in Leubach auf die Frage SUF 2.2 "Aufnahmeort" erhobene Beleg wurde im Fragebuch als $\text{l} \ddot{o} \text{b} \text{i} \text{X}$ transkribiert und für die elektronische Weiterverarbeitung in der ASCII-Ortsdatei für Leubach den Kodierungsrichtlinien entsprechend als $LO=5BI5.X7$ umgesetzt.

Um die Lautschrift auf WINDOWS-Systemen darstellen zu können, wurde für die BAYDAT-Datenbank die Schriftart TEUTHOBD entwickelt. Für TEUTHOBD wird zwar, wie auch für die Kodierungen in den Ortsdateien, eine ASCII-Kodierung benutzt, die Zuordnung der einzelnen Zeichen entspricht dabei jedoch nicht den Kodierrichtlinien: Innerhalb der Kodierrichtlinien werden z.B. Ziffern verwendet, um Diakritika anzuzeigen. Die Ziffern können dabei jedoch, abhängig vom vorausgehenden Grundzeichen, für unterschiedliche Diakritika stehen. So zeigt die Ziffer 2 nach vokalischen Grundzeichen (z.B. $A2$, $E2$) den Öffnungsgrad *geschlossen* an (z.B. ä , ë), nach konsonantischen Grundzeichen (z.B. $B2$) zeigt sie eine Fortisierung an (z.B. ę). Den Kodierungen entsprechen in TEUTHOBD die Zeichenfolgen $a4$ und $e4$ bzw. $b5$ ⁸³.

Für die Übernahme der TEUTHONISTA-Lautschrift in das Feld *TeuthoBD* der Tabelle *Belege* wurden die Kodierungen automatisch in die entsprechenden Zeichenfolgen der Schriftart TEUTHOBD umgesetzt, aus der Kodierung $A2$ wurde entsprechend $a4$, aus der Kodierung $B2$ $b5$ usw. Häufig auftretende Kodierfehler – statt des *Et*-Zeichens (&), das Reduktionen anzeigt, wurde z.B. häufig die Ziffer 6 eingegeben – konnten dabei automatisch korrigiert werden, seltenere Kodierfehler hingegen nicht. Im Zweifelsfall können zur Überprüfung der Belege die Originalkodierung bzw. für die Belege aus dem SUF die Scans der Originalfragebücher herangezogen werden.

Da sich bei der Anzeige der ASCII-kodierten Schriftart TEUTHOBD Probleme mit der Anzeige in MICROSOFT INTERNET EXPLORER 6 ergeben (s.u.), wurde die Kodierung der

⁸³ Für eine vollständige Übersicht der Zeichenkodierungen in TEUTHOBD bzw. TEUTHOUC vgl. Tabelle 8.1

Schriftart TEUTHOBD in der Schriftart TEUTHOUC von ASCII zu UNICODE verändert, die Zeichen beider Schriftarten bleiben dabei identisch.

UNICODE ist ein von der ISO verabschiedeter Standard⁸⁴, der durch eine Verbreiterung der Zeichenbreite auf bis zu 32 Bit (gegenüber der sieben bzw. acht Bit ASCII-Zeichenbreite) die Darstellung aller weltweit gebräuchlichen Alphabete mittels eines einzelnen Zeichensatzes ermöglicht⁸⁵. Dabei wird jedem bekannten Zeichen exakt ein Codepunkt zugeordnet, dem Großbuchstaben *A* aus dem lateinischen Alphabet ist z.B. der Codepunkt *0041* zugeordnet⁸⁶, der Ligatur *ß* der Codepunkt *00DF*⁸⁷. Die Schriftart TEUTHOUC besteht einerseits aus Zeichen, die vorgegebenen Codepunkten aus dem UNICODE-Zeichensatz zugeordnet werden können; so erhalten die Kleinbuchstaben und die öffnende bzw. schließende Klammer die entsprechenden UNICODE-Kodierungen, der Kleinbuchstabe *b* z.B. die Kodierung *0062*. Andererseits enthält die Schriftart TEUTHOUC Zeichen, die keinem UNICODE-Codepunkt zugeordnet werden können, z.B. einzelne Diakritika oder Grenzwerte zwischen zwei Vokalen.

Für Zeichen, die keinem UNICODE-Codepunkt zugeordnet werden können, steht innerhalb des UNICODE-Coderaums die so genannte *Private Use Area* zur Verfügung, die von Spezialanwendungen wie z.B. der BAYDAT-Datenbank zur Darstellung von innerhalb UNICODE nicht vorhandener Zeichen frei genutzt werden kann⁸⁸. Allerdings werden die Codepunkte innerhalb der *Private Use Area* im MICROSOFT INTERNET EXPLORER 6 als Grundzeichen interpretiert, ihnen wird, genau wie Buchstabenzeichen u.ä., ein eigener Leerraum zugewiesen. Werden Diakritika Codepunkten aus der *Private Use Area* zugeordnet, führt dies daher zu Anzeigeproblemen. So wird z.B. der Lautschriftbeleg aus dem SUF-Erhebungsort Leubach auf Frage 0002 der BSA-Fragekonkordanz, "Bewohner desselben [=Bewohner des Erhebungsortes]", der regulär als $l\ddot{o}, b_i, \chi\theta^{\wedge}$ dargestellt werden sollte, im MICROSOFT INTERNET EXPLORER wie folgt dargestellt:

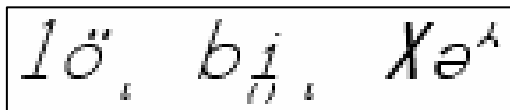


Abbildung 8.3: Darstellung von Belegen bei Nutzung der *Private Use Area* im MICROSOFT INTERNET EXPLORER

⁸⁴ ISO 10646; vgl.: Aliprand et al., S.1

⁸⁵ vgl. Aliprand et al., S.1

⁸⁶ vgl. Unicode Inc. (Hrsg.): C0 Controls and Basic Latin.

⁸⁷ vgl. Unicode Inc. (Hrsg.): C1 Controls and Latin-1 Supplement.

⁸⁸ Die *Private Use Area* umfasst die Codepunkte *E000 – F8FF*. Vgl.: Unicode Org. (Hrsg.): Private Use Area.

Es wird jeweils nur das erste Diakritikum korrekt über dem jeweiligen Grundzeichen dargestellt. Stehen weitere Diakritika bei einem Grundzeichen, so werden diese als Grundzeichen interpretiert.

Identische Anzeigeprobleme treten bei Verwendung der ASCII-kodierten Schriftart TEUTHOBD auf, da hier teilweise Diakritika auf Codepunkte gelegt werden, die vom Browser als Grundzeichen interpretiert werden, z.B. liegt das Diakritikum zur Anzeige der Fortisierung (z.B. $\dot{\text{d}}$) auf dem ASCII-Codepunkt, der innerhalb der ASCII-Codetabelle zur Darstellung der Ziffer 5 vorgesehen ist.

Prinzipiell könnten für diejenigen Diakritika aus der TEUTHONISTA-Lautschrift, für die UNICODE-Codepunkte vorgesehen sind, die entsprechenden Codepunkte verwendet werden. Es ergeben sich jedoch Anzeigeprobleme beim Übereinanderstellen mehrerer Diakritika. Der UNICODE-Standard erlaubt zwar, einem Grundzeichen mehrere Diakritika zuzuordnen⁸⁹, im Browser werden die Diakritika jedoch auf einer Linie oberhalb des Grundzeichens gruppiert. Der Lautschriftbeleg aus dem SUF-Erhebungsort Weimarschmieden auf Frage 0015 der BSA-Fragekonkordanz, "Nächste größere Stadt", der dem UNICODE-Standard entsprechend als $\mathring{\text{n}}\mathring{\text{i}}\mathring{\text{s}}\text{dad}$ dargestellt werden müsste, wird in gängigen Webbrowsern wie folgt dargestellt:



Abbildung 8.4: Darstellung mehrerer Diakritika über einem Grundzeichen in MICROSOFT INTERNET EXPLORER 6 (links) und MOZILLA FIREFOX 1.5 (rechts)

Um Diakritika aus der TEUTHONISTA-Lautschrift auf mehreren Ebenen korrekt im Browser anzuzeigen, wurden diesen Codepunkte aus der UNICODE-Codetabelle zugewiesen, die bereits für Diakritika vorgesehen waren. Dabei wurde darauf geachtet, dass die Codepunkte zumindest für Diakritika erster Ebene mit den UNICODE-Richtlinien übereinstimmen. Allerdings stimmen einige Zeichen nicht mit der eigentlichen UNICODE-Definition überein, die vorgeschlagene Lösung ist nicht vollständig konform mit dem UNICODE-Standard. Allerdings bleibt sie bis zur besseren Umsetzung der UNICODE-Richtlinien in zukünftigen Webbrowser-Generationen die einzig gangbare, um die Anzeige der TEUTHONISTA-Lautschrift in sämtlichen aktuellen Webbrowsern zu ermöglichen.

Für die nicht darstellbaren Grundzeichen, z.B. Grenzwerte zwischen zwei Vokalen, wurden Codepunkte aus der *Private Use Area* vergeben, die gewählte Lösung entspricht damit dem UNICODE-Standard.

⁸⁹ vgl.: Aliprand et al., S.44 ff.

Die folgende Übersicht zeigt sämtliche Zeichen der Schriftarten TEUTHOBD und TEUTHOUC mit ihren jeweiligen ASCII- bzw. UNICODE-Codepunkten:

	Ebene	Beispiel	ASCII (=TEUTHOBD) ⁹⁰	UNICODE (=TEUTHOUC)
VOKALE				
Grundwerte ⁹¹		a	61	0061
		l	49	e003
		ö	6F 3F	0075 0308
		ü	75 3F	0075 0308
Öffnungsgrad ⁹²		ä	61 24	0061 0324
		â	61 34	0061 0323
		ã	61 2E	0061 0345
		ä	61 3A	0061 031D
Öffnungsgrad (Halbwerte)		ä	61 34 39	0061 0323 0309
		ä	61 24 29	0061 0324 030b
Rundung	1	ë	65 3F	0065 0308
	1	ë	65 3F 28	0065 0308 030f
	1	ë	65 5C	0065 0310
	1	ë	65 5C 38	0065 0310 030e
Zentralisierung ⁹³		ä	E2 61	0351 0061
		ä	C2 61	0352 0061
		ä	F4 65	031f 0065
		ä	D4 65	031e 0065
		ä	EE 69	0354 0069
		ä	CE 69	0355 0069

⁹⁰ Für die ASCII-Codepunkte wird in der Tabelle jeweils deren Hexadezimalwert angegeben. Eine Umrechnung in die jeweiligen Dezimalwerte ist z.B. über den mit MICROSOFT WINDOWS ausgelieferten Rechner möglich.

⁹¹ Der Grundwert a ist hier beispielhaft für die vokalischen Grundwerte angegeben. Die Codepunkte der übrigen vokalischen Grundwerte entsprechen den Codepunkten der jeweiligen Kleinbuchstaben in der ASCII- bzw. UNICODE-Codetabelle.

⁹² Der Grundwert a ist hier ebenfalls als Beispiel angegeben, für Diakritika über/unter anderen Grundwerten muss der Codepunkt 61 durch die Codepunkte der entsprechenden Grundwerte ersetzt werden. Codepunkte von Diakritika, die zu einem Grundzeichen neu hinzutreten bzw. Diakritika und Grundzeichen, die das erste Mal in der Tabelle aufgeführt werden, sind fett gedruckt.

⁹³ Die Codepunkte des (doppelt) zentralisierten o entsprechen den Codepunkten des (doppelt) zentralisierten e; zur Darstellung des (doppelt) zentralisierten o muss der Codepunkt des Grundwerts entsprechend ersetzt werden. Für die Darstellung der Zentralisierung gilt anders als bei den übrigen Diakritika die Reihenfolge *Klammer – Zentralisierung – Grundwert*.

		ʌ	FB 75	0356 0075
		ʌ	DB 75	0357 0075
		ʌ	7E FB 75	0307 0356 0075
Nasalierung	1	ã	61 2B	0061 032a
	1	ã	61 2B 28	0061 032a 030f
	1	ã	61 2A	0061 033a
	1	ã	61 2A 38	0061 033a 030e
	2	ã	61 36	0061 0303
	2	ã	61 26	0061 031a
Quantität	1	ā	61 32	0061 0304
	1	ā	61 33	0061 0321
	1	ā	61 3C	0061 0302
	1	ā	61 32 3E	0061 0304 0317
	2	ā	61 22	0061 0305
	2	ā	61 A7	0061 0322
	2	ā	61 3E	0061 0317
	2	ā	61 22 7C	0061 0305 0316
	3	ā	61 B2	0061 0312
	3	ā	61 7C	0061 0316
Silbengrenze		ā	61 40	0061 0325
Akzent	1	á	61 37	0061 0301
	1	á	61 30	0061 0300
	2	á	61 2F	0061 0341
	2	á	61 3D	0061 0340
	3	á	61 7B	0061 0314
	3	á	61 7D	0061 0315
Halbvokal		ä	61 5E	0061 031c
		j	6A	006a
Reduktionsvokale		ə	45	e001
		ɔ	41	e000
Grenzwerte		ɔ̃	E4	e010
		ɔ̃	C4	e011

		Ⓒ	C6	e013
		Ⓔ	E5	e012
		Ⓕ	C5	e020
		Ⓖ	EB	e030
		Ⓒ	CB	e031
		Ⓔ	EA	e032
		Ⓕ	CA	e033
		Ⓖ	EF	e050
		Ⓒ	F6	e060
		Ⓔ	D6	e061
		Ⓕ	F2	e062
		Ⓖ	D2	e063
		Ⓒ	F3	e064
		Ⓔ	FC	e070
		Ⓕ	DC	e071
		Ⓖ	FA	e072
		Ⓒ	DA	e073
KONSONANTEN				
Grundwerte ⁹⁴		b	62	0062
Fortisierung		ḃ	62 35	0062 0329 ⁹⁵
		Ḅ	62 25	0062 0348 ⁹⁶
Lenisierung		ḅ	62 2C	0062 0353
		Ḇ	62 3B	0062 0321
Stimmhaftigkeit		ḇ	62 34	0062 0323 ⁹⁷
Implosion		Ḉ	62 2D	0062 0350
Glottisschlag		ʔ	71	e119
Silbische Funktion		ɳ	6E 40	006e 0325
Spirantisierung		ḥ	42	e121

⁹⁴ Der Grundwert b ist hier exemplarisch angegeben; die Codepunkte weiterer konsonantischer Grundwerte entsprechen den ASCII- bzw. UNICODE-Codepunkten der entsprechenden Kleinbuchstaben. Abweichungen, z.B. die Neubelegung des ASCII-Codepunkts 63 durch den Grundwert λ werden an der entsprechenden Stelle angegeben.

⁹⁵ bzw. F0 (ASCII) und 0363 (UNICODE) für Grundzeichen mit Unterlänge

⁹⁶ bzw. E1 (ASCII) und 0364 (UNICODE) für Grundzeichen mit Unterlänge

⁹⁷ bzw. E0 (ASCII) und 0365 (UNICODE) für Grundzeichen mit Unterlänge

		đ	44	e122
		g	47	e101
Bilabialer Reibelaut		β	57	e100
s-Laute		š	73 23	0073 030c
		ʃ	53	e116
		ṣ̌	53 27	e116 0313
		ś	73 37	0073 0301
ch-Laute		χ	63 5B	0063 0358
		χ	63 31	0063 035a
		x	78	0078
		χ	58	e118
		χ	58 5D	e118 0359
		χ	58 21	e118 035b
		x	78 31	0078 035a
Nasale		ŋ	4E	e110
		ŋ	A3	e123
l-Laute		l̥	4B	e103
		l̥	4D	e105
		l̥	4F	e106
		l̥	6C 40 39	006c 0325 0309
		l̥	6C B0	006c 031b
		ʈ	FD	E125
		ʈ	FF	E126
		ʈ	DD	E127
		ʈ	DE	E128
		ʈ	EE 6C	0354 006c
		l̥	6C 5E	006c 031c
		l̥	56	e107
		l̥	FE	e124
		l̥	4A	e120
		l̥	6C 2E	006c 0345
		l̥	51	e104

		Ł	4C	004c
r-Laute		ŕ	7A	e114
		ŕ	5A	e115
		ł	79	e112
		ř	72 23	0072 030c
		ř	72 23 34	0072 030c 0323
		Ŕ	52	e111
		ŕ	54	e113
		ř	52 23	e111 030c
Grenzwerte		ŕ m	F1	e074
		m ŕ	D1	e075
Sonstiges		ħ	48	e102
		h	5F	036a
		(AB	0028
)	BB	0029

Tabelle 8.1: Übersicht über Zeichen aus der TEUTHONISTA-Lautschrift mit UNICODE- (=TEUTHOUC) und ASCII-Codepunkten (=TEUTHOBD)

UNICODE ist der Standard-Zeichensatz von z.B. XML und der Programmiersprache JAVA und kann von gängigen Webbrowsern problemlos verarbeitet werden, UNICODE-kodierte Zeichen können entsprechend dargestellt werden. Um die UNICODE-kodierte Schriftart TEUTHOUC anzeigen zu können, muss diese jedoch auf dem Rechner des BAYDAT-Nutzers installiert werden. Verweise auf die entsprechende Schriftartendatei sowie eine Installationsanleitung finden sich auf der Startseite des Internetauftritts der BAYDAT-Datenbank.

Zum jetzigen Zeitpunkt ist es nicht möglich, UNICODE-kodierte Zeichen in PDF-Dokumenten darzustellen. Daher wird zur Erstellung von Beleglisten (vgl. 8.3.2.2.) die ASCII-kodierte Schriftart TEUTHOBD benutzt. Diese muss jedoch nicht auf dem Rechner des BAYDAT-Benutzers installiert werden. Die entsprechende Schriftartendatei liegt zentral auf dem Webserver, die Schriftart wird von hier in die PDF-Dokumente eingebunden. Zur eigenen Verwendung wird jedoch auch die ASCII-kodierte Schriftart TEUTHOBD auf der BAYDAT-Startseite zum Download angeboten. Außerdem können beide Schriftarten eigenen Bedürfnissen angepasst und z.B. fehlende Zeichen ergänzt oder bestehende Zeichen verändert werden.

8.3. Die grafische Benutzeroberfläche

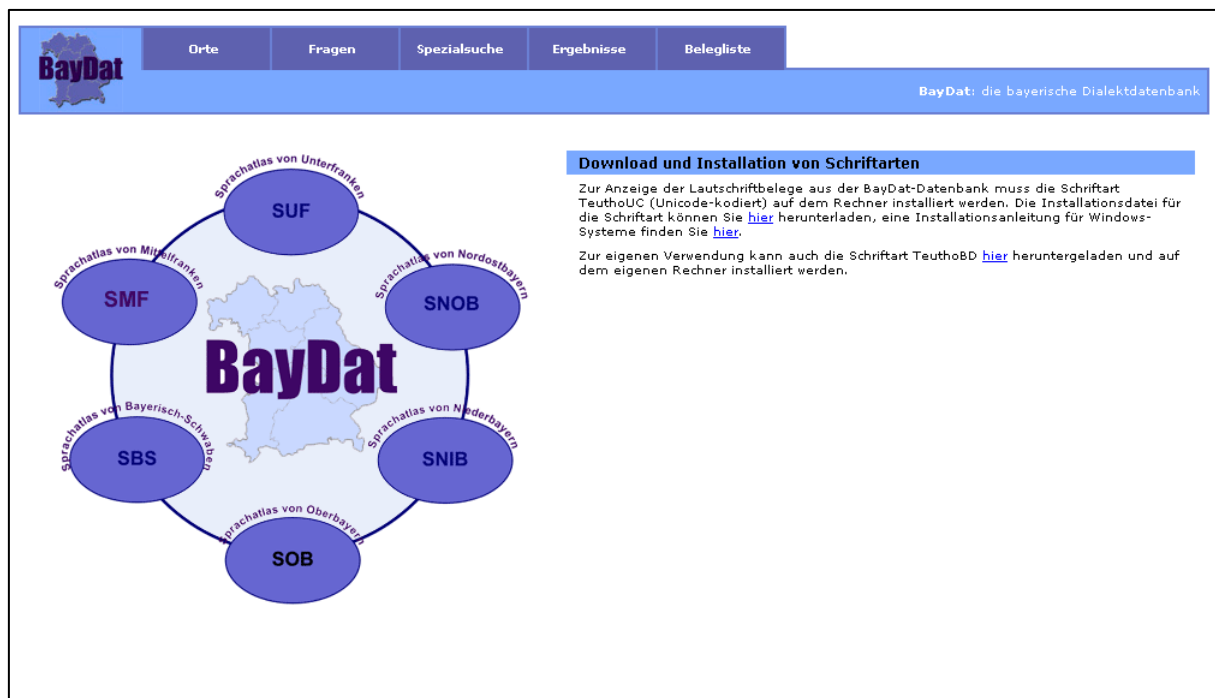


Abbildung 8.5: Startseite des BAYDAT-Internetauftritts

Abbildung 8.5 zeigt die Startseite des Internetauftritts der BAYDAT-Datenbank. Diese gliedert sich in einen oberen Bereich und einen unteren Bereich. Im oberen Bereich befindet sich eine Navigationsleiste, die den Wechsel zwischen den Menüpunkten *Orte*, *Fragen*, *Spezielsuche*, *Ergebnisse* und *Belegliste* ermöglicht. Die Navigationsleiste bleibt über den gesamten Internetauftritt der BAYDAT-Datenbank erhalten und die einzelnen Menüpunkte werden – je nach gewählter Unterseite – grafisch hervorgehoben. Hierdurch wird dem Nutzer die Orientierung innerhalb des BAYDAT-Internetauftritts erleichtert⁹⁸. Durch die Realisierung der Navigation in Form von Karteikartenreitern wird dem Nutzer die Orientierung ebenfalls erleichtert: Einerseits sind Karteikartenreiter zur Navigation selbsterklärend und im Vergleich zu anderen Navigationsmöglichkeiten schwer zu übersehen⁹⁹, andererseits wird diese Form der Navigation mittlerweile von zahlreichen anderen Internetseiten genutzt, sie dürfte dem Nutzer deshalb bekannt sein und sich ihm intuitiv erschließen.

Im unteren Bereich finden sich die wichtigsten Informationen zum Projekt BAYDAT. Der untere Bereich könnte in Zukunft auch zur Veröffentlichung von aktuellen Entwicklungen rund um die BAYDAT-Datenbank o.ä. genutzt werden.

⁹⁸ Diese Art der *persistent* bzw. *global navigation* wird u.a. von Steve Krug empfohlen; vgl. Krug, S.62

⁹⁹ vgl. Krug, S.80

8.3.1. Auswahl von Orten und Fragen

Um Belege anzeigen zu können, müssen zunächst über den Menüpunkt *Orte* die relevanten BSA-Erhebungsorte und über die Menüpunkte *Fragen* und/oder *Spezielsuche* die gewünschten Fragen aus der BSA-Fragekonkordanz ausgewählt werden:

8.3.1.1. Der Menüpunkt *Orte*

The screenshot shows the BayDat website interface. At the top, there is a navigation bar with tabs for 'Orte', 'Fragen', 'Spezielsuche', 'Ergebnisse', and 'Belegliste'. The 'Orte' tab is selected. Below the navigation bar, there is a search box labeled 'Schnellsuche' with a 'Los!' button. To the left of the search results, there is a sidebar for 'Regierungsbezirke' with options for 'Mittelfranken', 'Niederbayern', 'Oberbayern', 'Unterfranken', 'Landkreise', and 'Altlandkreise'. The main content area displays a table of search results with columns for 'PLZ', 'Ort', 'LK', and 'ALK'. The table is titled 'Auswählen | Alle | Keine'. The results are as follows:

PLZ	Ort	LK	ALK
<input checked="" type="checkbox"/> 97355	Abtswind	KT	GEO
<input type="checkbox"/> 97357	Altenshönbach	KT	GEO
<input type="checkbox"/> 97355	Castell	KT	GEO
<input type="checkbox"/> 97337	Dettelbach	KT	KT
<input type="checkbox"/> 97346	Dornheim	KT	SEF
<input type="checkbox"/> 96160	Geiselwind	KT	SEF
<input type="checkbox"/> 97348	Hüttenheim in Bayern	KT	KT
<input checked="" type="checkbox"/> 97318	Kitzingen	KT	KT
<input type="checkbox"/> 97355	Rüdenhausen	KT	GEO
<input type="checkbox"/> 97334	Sommerach	KT	GEO
<input checked="" type="checkbox"/> 97359	Stadtschwarzach	KT	KT

At the bottom of the table, there is a link 'Auswählen | Alle | Keine'. To the right of the table, there is a sidebar for 'Ausgewählte Orte' with a list of selected locations: 'Castell', 'Dornheim', and 'Hüttenheim in Bayern'. Below this list, there is a link 'Abwählen | Alle | Keine' and a section for 'Ausgewählte Fragen' with the text 'Keine Frage(n) gewählt' and 'Gespeicherte Abfragen' with the text 'Keine Abfragen gespeichert'.

Abbildung 8.6: Der Menüpunkt *Orte*¹⁰⁰

Abbildung 8.6 zeigt die Seite, die nach Auswahl des Menüpunkts *Orte* erscheint. Im oberen Teil der Seite befindet sich die bereits beschriebene Navigationsleiste zur Auswahl der einzelnen Menüpunkte (1), der Reiter *Orte* ist grafisch hervorgehoben. Links befindet sich ein Feld für die Suche nach BSA-Erhebungsorten (2), in der Mitte werden die Suchergebnisse angezeigt (3). Beim erstmaligen Aufruf der Seite bleibt der mittlere Teil der Seite leer, in der Abbildung werden der Anschaulichkeit wegen bereits Suchergebnisse – die BSA-Erhebungsorte aus dem unterfränkischen Landkreis Kitzingen – angezeigt. Der rechte Teil der Seite zeigt die vom Benutzer ausgewählten Orte und Fragen (4).

¹⁰⁰ Die Erhebungsdaten aus dem SBS und dem SNOB liegen momentan noch nicht vollständig vor. Aus diesem Grund fehlen im Screenshot die Regierungsbezirke Bayerisch-Schwaben, Oberfranken und Oberpfalz.

a) Das Suchfeld (2)

Der Benutzer hat zwei Möglichkeiten, nach BSA-Erhebungsorten zu suchen: über das Textfeld oder über die darunter angeordneten Felder *Regierungsbezirke* und *Landkreise*.

Das Textfeld erlaubt die Suche nach Ortsnamen oder Postleitzahlen. Dabei reicht die Eingabe eines Teils des Ortsnamens aus, um BSA-Erhebungsorte, deren Ortsnamen die entsprechende Zeichenfolge enthalten, zu finden. Eine Eingabe von *main* führt zu den Suchergebnissen *Bayerisch Gmain, Frickenhausen am Main, Homburg am Main, Karlstadt am Main, Klingenberg am Main, Lohr am Main, Mainberg, Stockstadt am Main* und *Sulzbach am Main*. Die Groß- und Kleinschreibung spielt bei der Suche keine Rolle: Die Eingaben *Main, main* oder *MAIN* führen allesamt zu den oben genannten Suchergebnissen. Bei der Eingabe von Postleitzahlen kann die gesamte, fünfstellige Postleitzahl angegeben werden, die Eingabe der Anfangsziffern reicht jedoch aus, um BSA-Erhebungsorte zu finden, deren Postleitzahlen mit den entsprechenden Ziffernfolgen beginnen. Die Eingabe *8244* findet beispielsweise die BSA-Erhebungsorte *82444 Schlehdorf, 82445 Grafenaschau, 82447 Spatenhausen* und *82449 Schoeffau*.

Abhängig von der Eingabe des Benutzers wird die Tabelle *Orte* der BAYDAT-Datenbank mit verschiedenen Bedingungen abgefragt: *select * from orte where name like '%BENUTZEREINGABE%'* bzw. *select * from orte where plz like 'BENUTZEREINGABE%'*¹⁰¹.

Eine zweite Möglichkeit, BSA-Erhebungsorte zu finden, bieten die unter dem Suchfeld gruppierten Felder *Regierungsbezirke* und *Landkreise*. Nach dem Aktivieren des jeweiligen Feldes durch Linksklick erscheint eine Auflistung der Regierungsbezirke – in Abbildung 8.6. dargestellt – bzw. eine Liste der bayerischen Landkreise. Durch Linksklick auf die Namen der Regierungsbezirke bzw. Landkreise werden die BSA-Erhebungsorte aus den entsprechenden Regierungsbezirken bzw. Landkreisen als Suchergebnisse angezeigt. Die Listen der Regierungsbezirke und Landkreise können durch erneute Aktivierung der Felder durch Linksklick wieder ausgeblendet werden.

Auch hier wird, abhängig von der Aktivierung einzelner Links, die Tabelle *Orte* der BAYDAT-Benutzerdatenbank unter verschiedenen Bedingungen abgefragt. Für die Anzeige der BSA-Erhebungsorte aus dem Landkreis Kitzingen kommt z.B. folgende Abfrage zur Anwendung: *select * from orte where lk = 'KT'*. Um sämtliche BSA-Erhebungsorte aus dem Regierungsbezirk Mittelfranken anzuzeigen, wird folgende Abfrage generiert: *select * from orte where rb = 'mf'*.

¹⁰¹ Durch Angabe des Asterisks (*) werden sämtliche Spalten der entsprechenden Tabelle ausgewählt; vgl. Connolly/Begg, S.117

b) Die Anzeige der Suchergebnisse (3)

Die Ergebnisse einer unter a) getätigten Suchanfrage werden in der Mitte der Seite angezeigt. Abbildung 8.6 zeigt die Ergebnisse einer Suchanfrage nach BSA-Erhebungsorten aus dem Landkreis Kitzingen. Die Ergebnisse werden zunächst nach Ortsnamen in alphabetischer Reihenfolge sortiert angezeigt. Durch Linksklick auf den jeweiligen Spaltenkopf lassen sich die Ergebnisse nach Postleitzahl oder Landkreis sortieren, ein weiterer Linksklick auf den jeweiligen Tabellenkopf kehrt die Suchreihenfolge um, es wird gegen das Alphabet (*Ort, Landkreis*) bzw. in absteigender Folge (*PLZ*) sortiert. Dies ermöglicht dem Benutzer der Oberfläche einen schnelleren Zugriff auf die auszuwählenden Orte und kann vor allem dann nützlich sein, wenn die Anfrage eine größere Ergebnismenge zurückgibt, z.B. bei der Suche nach Orten aus gesamten Regierungsbezirken.

Durch Aktivierung einzelner Auswahlkästchen in der linken Tabellenspalte können einzelne Orte vorselektiert werden, durch Linksklick auf den Link *Alle* ober- bzw. unterhalb der Suchergebnisse werden alle Auswahlkästchen aktiviert. Um die aktivierten Auswahlkästchen zu deaktivieren, genügt ein Linksklick auf den Link *Keine* ober- bzw. unterhalb der Anzeige der Suchergebnisse. Ein Linksklick auf den Link *Auswählen* übernimmt die vorselektierten Orte in die endgültige Auswahl.

c) Das Feld *Ausgewählte Orte* (4)

Die unter b) ausgewählten Orte werden in das Feld *Ausgewählte Orte* übernommen. Die Auswahl bleibt dabei auch nach einem Wechsel zwischen den einzelnen Menüpunkten erhalten. Um einzelne Orte aus der Auswahl zu entfernen, muss das jeweilige Auswahlkästchen deaktiviert und durch Linksklick auf den Link *Abwählen* unterhalb des Feldes *Ausgewählte Orte* bestätigt werden. Ein Linksklick auf den Link *Alle* deaktiviert alle Auswahlkästchen, ein Linksklick auf *Keine* macht die Deaktivierung rückgängig.

8.3.1.2. Der Menüpunkt *Fragen*

The screenshot shows the BAYDAT website interface. At the top, there is a navigation bar with tabs for 'Orte', 'Fragen', 'Spezielsuche', 'Ergebnisse', and 'Belegliste'. The 'Fragen' tab is currently selected. Below the navigation bar, there is a search area on the left with a 'Schnellsuche' field and a 'Sachbereiche' dropdown. The main content area displays a table of questions. The table has columns for 'Auswählen', 'Link', 'Nr.', 'Frage', and 'Suggestierformen'. The questions listed are related to cats. On the right side, there are two summary panels: 'Ausgewählte Orte' and 'Ausgewählte Fragen', both showing selected items with checkboxes.

Auswählen	Link	Nr.	Frage	Suggestierformen
<input type="checkbox"/>	↻	0552	"Unsere Katze"	-
<input type="checkbox"/>	↻	0553	"Unsere Katzen"	-
<input type="checkbox"/>	↻	0554	"Mit unseren (unsernen) Katzen sind wir zufrieden"	-
<input type="checkbox"/>	-	0555	Männliche Katze	-
<input type="checkbox"/>	-	0556	Weibliche Katze	-
<input type="checkbox"/>	↻	0557	Wenn die Katzen im Februar - März so verrückt sind, sind sie/tun sie	-
<input type="checkbox"/>	-	0558	Während der Paarungszeit tut der Kater ...	-
<input type="checkbox"/>	-	0559	Sich spielend balgen (von jungen Katzen)	-
<input type="checkbox"/>	-	0560	Wenn die Katze sich wohl fühlt, tut sie ...	-
<input type="checkbox"/>	↻	0561	"Tape" (Lautung und Bed.) (Pfote, Spur u.ä.)"	-

Abbildung 8.7: Der Menüpunkt *Fragen*

Abbildung 8.7 zeigt die Seite, die nach der Auswahl des Menüpunktes *Fragen* erscheint. Sie gliedert sich wie die Seite zum Menüpunkt *Orte* in vier Teile: Im oberen Teil befindet sich die bereits beschriebene Navigationsleiste zur Auswahl der einzelnen Menüpunkte (1), der Reiter *Fragen* ist grafisch hervorgehoben. Auf der linken Seite befindet sich ein Feld für die Suche nach Fragen aus der BSA-Fragekonkordanz (2), die Suchergebnisse werden im mittleren Teil der Seite angezeigt (3). Beim erstmaligen Aufruf des Menüpunktes *Fragen* bleibt der mittlere Teil der Seite leer, zur besseren Anschaulichkeit werden in der Abbildung bereits Suchergebnisse – hier alle Fragen aus dem Sachbereich *Katze* – angezeigt. Der rechte Teil der Seite zeigt die vom Benutzer ausgewählten Orte und Fragen (4).

a) Das Suchfeld (2)

Dem Nutzer werden zwei Möglichkeiten geboten, nach Fragen aus der BSA-Fragekonkordanz zu suchen: das Textfeld und das darunter liegende Feld *Sachbereiche*.

In das Textfeld können beliebige Zeichenfolgen eingegeben werden, anschließend werden aus der Tabelle *Fragen* der BAYDAT-Datenbank diejenigen Fragen ermittelt, die die entsprechende Zeichenfolge im Fragetext enthalten. Dabei kommt folgende SQL-Abfrage zur

Anwendung, der Platzhalter *BENUTZEREINGABE* entspricht dabei dem jeweils eingegebenen Suchbegriff: *select fragenr, frage from fragen where frage like '%BENUTZEREINGABE%'*. Außerdem wird durch folgende SQL-Abfrage im Feld *Suggestierform* der Tabelle *Suggestierformen* nach der jeweiligen Zeichenfolge gesucht, es werden anschließend die zugehörigen Fragen als Ergebnis ausgegeben: *select fragenr, frage from fragen where fragenr in (select fragenr from suggestierformen where suggestierform like '%BENUTZEREINGABE%')*. Die Groß- und Kleinschreibung spielt bei der Eingabe des Suchbegriffs keine Rolle: Die Suchbegriffe *maus* und *Maus* haben beide die Fragen 2429 "'Maus' (Die Katze fängt eine...)" und 2436 "Maulwurf (nicht: Wühlmaus!)" (aufgrund der Zeichenfolge *maus* im Fragetext, aber auch aufgrund der Suggestierformen *Mauswurf* und *Schermaus*) zum Ergebnis.

Eine zweite Suchmöglichkeit bietet das Feld *Sachbereiche* unterhalb des Suchfeldes. Nach Aktivierung des Feldes durch Linksklick erscheint eine Liste der 59 Sachbereiche aus der BSA-Fragekonkordanz. Nach einem weiteren Linksklick auf den Namen eines Sachbereichs werden die zum jeweiligen Sachbereich gehörigen Fragen als Ergebnis angezeigt. Die Liste der Sachbereiche kann durch erneuten Linksklick auf das Feld *Sachbereiche* ausgeblendet werden.

Um die Fragen der einzelnen Sachbereiche anzuzeigen, wird die Tabelle *Fragen* der BAYDAT-Datenbank abgefragt, als Suchkriterium wird die jeweilige Sachbereichsnummer angegeben, für Fragen aus dem Sachbereich *Katze* z.B. *select fragenr, frage from fragen where sachbereich = '15'*.

b) Die Anzeige der Suchergebnisse (3)

Die Ergebnisse der unter a) getätigten Suchanfrage werden in der Mitte der Seite angezeigt. Abbildung 8.7 zeigt dabei die Ergebnisse einer Suchanfrage nach Fragen aus dem Sachbereich *Katze*. Als Ergebnisse werden dabei die BSA-Fragennummern, die Fragetexte sowie die zur jeweiligen Frage angegebenen Suggestierformen angezeigt. Durch Aktivierung der Auswahlkästchen in der linken Tabellenspalte und anschließende Bestätigung durch Linksklick auf den Link *Auswählen* ober- bzw. unterhalb der Suchergebnisse werden die jeweiligen Fragen in die endgültige Auswahl übernommen. Eine Aktivierung des Links *Alle* aktiviert alle Auswahlkästchen, ein Linksklick auf den Link *Keine* deaktiviert sämtliche Auswahlkästchen.

c) Das Feld *Ausgewählte Fragen* (4)

Die unter b) ausgewählten Fragen werden in das Feld *Ausgewählte Fragen* übernommen. Auch diese Auswahl bleibt bei einem Wechsel zwischen den einzelnen Menüpunkten erhalten. Um Fragen abzuwählen, muss zunächst das Auswahlkästchen vor der jeweiligen Frage deaktiviert werden. Nach Bestätigung durch Linksklick auf den Link *Abwählen* ist die entsprechende Frage abgewählt. Durch Aktivierung des Links *Alle* werden sämtliche Auswahlkästchen deaktiviert, durch *Keine* werden die deaktivierten Auswahlkästchen wieder aktiviert.


d) Zusatzinformationen

Durch Linksklick auf den jeweiligen Fragetext lassen sich weitere Informationen zu jeder Frage anzeigen. In der Spalte *Link* wird dabei angezeigt, welche Zusatzinformationen zur jeweiligen Frage vorhanden sind:

Enthält die Spalte *Link* einen Bindestrich – in der Abbildung oben ist dies z.B. bei Frage 558 "Während der Paarungszeit tut der Kater..." der Fall – sind als Zusatzinformationen lediglich die Fragenummern aus den einzelnen BSA-Projekten hinterlegt:

Frage 0558:	Während der Paarungszeit tut der Kater ...
Details	
Sprachatlas von Bayerisch Schwaben:	<i>Frage in diesem Projekt nicht erhoben</i>
Sprachatlas von Niederbayern:	<i>Frage in diesem Projekt nicht erhoben</i>
Sprachatlas von Oberbayern:	<i>Frage in diesem Projekt nicht erhoben</i>
Sprachatlas von Nordostbayern:	<i>Frage in diesem Projekt nicht erhoben</i>
Sprachatlas von Mittelfranken:	035.002
Sprachatlas von Unterfranken:	<i>Frage in diesem Projekt nicht erhoben</i>

Abbildung 8.8: Zusatzinformationen zu Frage 558 "Während der Paarungszeit tut der Kater..."

Das Symbol  in der Spalte *Link* zeigt an, dass zur jeweiligen Frage zusätzlich zu den Fragenummern der Einzelprojekte Verknüpfungen auf Basis der Auszeichnungen sprachwissenschaftlicher Phänomene vorliegen. Die entsprechenden Verknüpfungen können durch Linksklick auf den Fragetext der jeweiligen Frage angezeigt werden. Ein Linksklick auf den Fragetext der Frage 553 "Unsere Katzen" führt z.B. zu folgender Anzeige:

Verknüpfungen	
Frage 0553:	"Unsere Katzen"
Substantive:	
Katze (3 Verknüpfungen)	
Artikelwörter:	
unser (12 Verknüpfungen)	

Abbildung 8.9: Verknüpfungen zu Frage 553 "'Unsere Katzen'"


Es wird angezeigt, dass zu Frage 553 eine Grundform in der Tabelle *Link_Substantive* – nämlich *Katze* – angegeben ist. Die Grundform *Katze* kommt in der Tabelle *Link_Substantive* dreimal vor, ein Linksklick auf die Grundform *Katze* zeigt die entsprechenden Fragen (Frage 552 "'Unsere Katze'", Frage 553 "'Unsere Katzen'" und Frage 554 "'Mit unseren (unsernen) Katzen sind wir zufrieden'") als Suchergebnis an, außerdem werden für Substantive Genus, Kasus und Numerus in der jeweiligen Frage angegeben. Die Fragen werden dabei – aufgrund der Ziffernkodierung in den Feldern *Kas* und *Num* der Tabelle *Link_Substantive* – in der korrekten Reihenfolge sortiert: Fragen zu Singularformen stehen vor Fragen zu Pluralformen, Fragen, die den Nominativ erheben, stehen vor Fragen, die den Dativ erheben:

Auswählen Alle Keine				
	Link	Nr.	Frage	Suggestierformen
Katze (Femininum), Nominativ Singular				
<input type="checkbox"/>	L	0552	"Unsere Katze"	-
Katze (Femininum), Nominativ Plural				
<input type="checkbox"/>	L	0553	"Unsere Katzen"	-
Katze (Femininum), Dativ Plural				
<input type="checkbox"/>	L	0554	"Mit unseren (unsernen) Katzen sind wir zufrieden"	-
Auswählen Alle Keine				

Abbildung 8.10: Fragen mit dem Substantiv *Katze*

Außerdem ist zu Frage 553 in der Tabelle *Link_Artikel* eine Grundform – nämlich *unser* – verzeichnet. Die Grundform kommt innerhalb der Tabelle *Link_Artikel* zwölfmal vor. Ein Linksklick auf das Grundwort *unser* würde die entsprechenden Fragen anzeigen.

Je nach Frage und Auszeichnung können Verknüpfungen zu weiteren Wortarten, aber auch zu Wortbildungen und zur Lautung angezeigt werden. Durch die Verknüpfungen lassen sich schnell sämtliche Fragen, die Formen eines bestimmten Wortes erheben, ermitteln, z.B. sämtliche Fragen zu Formen des Verbs *kaufen*, des Substantivs *Katze* oder der Präposition *in*.

Enthält die Spalte *Link* das Symbol , ist die jeweilige Frage mit Bilddateien aus dem Fotoarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS verknüpft. Dies ist z.B. der Fall bei Frage 1890 "Türen":

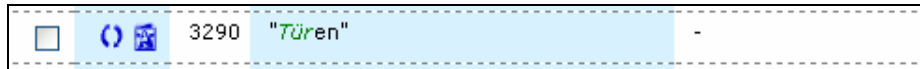


Abbildung 8.11: Anzeige von Verknüpfungen mit Bilddateien

Ein Linksklick auf den Fragetext "Türen" zeigt neben den bereits genannten Informationen (Frageummern in den BSA-Teilprojekten, vorhandene Verknüpfungen) die in der Tabelle *Bildarchiv* gespeicherte Bilddatei in verkleinerter Form als so genannten Thumbnail an, außerdem werden Dateiname, -größe und Pixelmaße der Bilddatei angegeben:

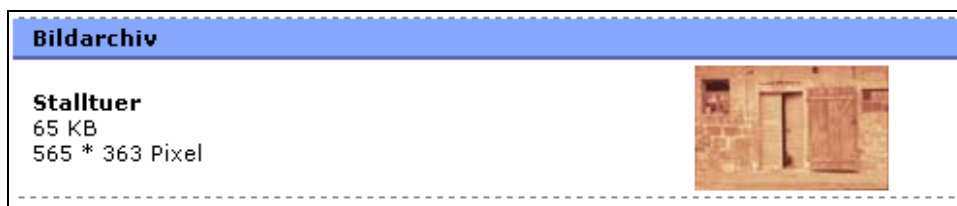


Abbildung 8.12: Thumbnail und Detailinformationen zur Bilddatei *Stalltuer*

Nach Linksklick auf den Thumbnail wird die Bilddatei in Originalgröße in einem neuen Fenster angezeigt.

8.3.1.3. Der Menüpunkt *Spezielsuche*

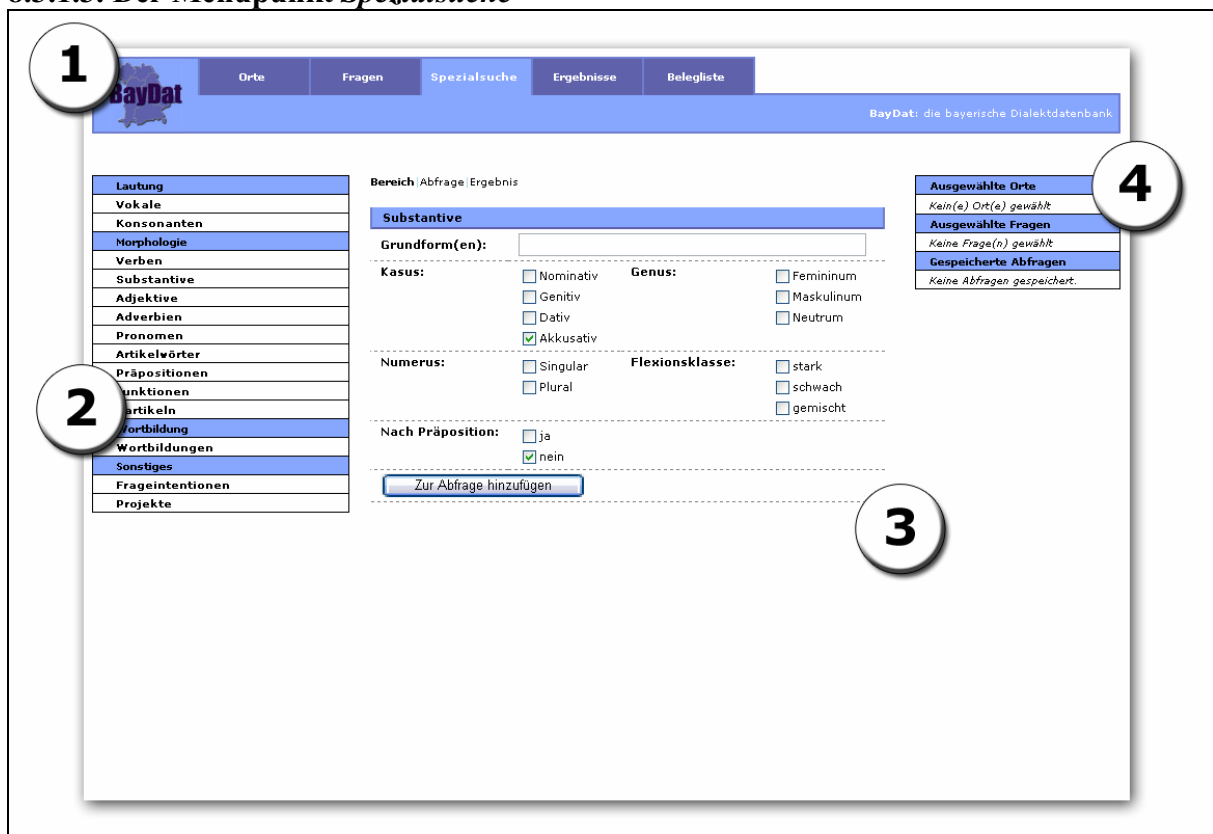


Abbildung 8.13: Der Menüpunkt *Spezielsuche*

Der Menüpunkt *Spezielsuche* gibt dem Nutzer der BAYDAT-Datenbank die Möglichkeit, komplexe Datenbankabfragen zu erstellen und so gezielt nach Fragen zu suchen, die bestimmte linguistische Phänomene, z.B. transitiv verwendete Verben oder Verben im Perfekt, erheben.

Abbildung 8.13 zeigt die Seite, die nach der Auswahl des Menüpunkt *Spezielsuche* erscheint, aus der Kategorienleiste (2) wurde die Kategorie *Substantive* ausgewählt. Die Seite gliedert sich in vier Teile. In der bereits beschriebenen Navigationsleiste (1) wird zur Orientierung der Menüpunkt *Spezielsuche* grafisch hervorgehoben. Die Kategorienleiste (2) erlaubt dem Nutzer, Suchmasken für sprachwissenschaftliche Phänomene aus den Bereichen Lautung, Morphologie und Wortbildung aufzurufen, außerdem kann er die Suchergebnisse nach Frageintentionen und Erhebung in den BSA-Teilprojekten eingrenzen. Im mittleren Teil der Seite (3) werden je nach Auswahl (*Bereich*, *Abfrage* oder *Ergebnis*) Suchmasken für die einzelnen Kategorien, die vom Benutzer formulierte Datenbankabfrage oder das Suchergebnis angezeigt. Im rechten Teil der Seite (4) werden bereits ausgewählte Orte und Fragen angezeigt.

a) Die Kategorienleiste (2)

Die in der Kategorienleiste angegebenen Kategorien entsprechen den in Kapitel 5 beschriebenen Auszeichnungen. Durch Linksklick auf die jeweilige Kategorie wird im mittleren Teil der Seite (3) die entsprechende Suchmaske angezeigt. Nach Linksklick auf *Frageintentionen* bzw. *Projekte* erscheinen Suchmasken, die eine Eingrenzung der Suchergebnisse nach Frageintention (Laut-, Lexik-, Morphologiefragen usw.) bzw. nach Erhebung in den BSA-Teilprojekten erlauben.

b) Die Suchmasken (3)

Abbildung 8.13 zeigt die Suchmaske, die nach Linksklick auf die Kategorie *Substantive* erscheint. Sie bietet dem Nutzer zunächst die Möglichkeit, eine bzw. mehrere Grundformen einzugeben, die in den gesuchten Fragen enthalten sein sollen. Die hier einzugebende Grundform entspricht dabei der Nennform der jeweiligen Wortart, für Substantive also der Nominativ Singular, für Verben der Infinitiv usw. Bei Eingabe mehrerer Grundformen sind diese lediglich durch ein Leerzeichen zu trennen, z.B. "Hund Katze Maus". Eine entsprechende Anfrage findet Fragen, die entweder das Substantiv "Hund" *oder* das Substantiv "Katze" *oder* das Substantiv "Maus" enthalten. Um Fragen zu finden, die das Substantiv "Hund" *und* das Substantiv "Katze" *und* das Substantiv "Maus" enthalten, ist

zunächst lediglich eine der Grundformen einzugeben und die Übernahme zur Abfrage durch Linksklick auf den Button "Zur Abfrage hinzufügen" zu bestätigen. Anschließend kann durch erneutes Auswählen der Kategorie *Substantive* eine weitere Grundform, die in der Frage enthalten sein soll, eingegeben werden.

Statt der Grundform können für die Kategorien *Vokale* und *Konsonanten* mittelhochdeutsche Bezugslaute angegeben werden. Die Schreibung entspricht dabei der unter 5.1.1.1. bzw. 5.2.1.1. beschriebenen Schreibweise. Innerhalb der Kategorie *Wortbildungen* können anstelle der Grundform einzelne Wortbildungskonstituenten angegeben werden. Präfixe sind dabei mit schließendem ("ver-", "ent-" usw.), Suffixe mit einem führenden Bindestrich einzugeben ("-ung", "-er" usw.). Die Suchmasken für Artikelwörter sowie Pronomen erlauben zusätzlich die Eingabe eines Bezugswortes, anzugeben ist auch hier die entsprechende Nennform, für Substantive also der Nominativ Singular.

Die weiteren Auswahlmöglichkeiten innerhalb der Suchmasken entsprechen weitestgehend den Spalten und Werten der in Kapitel 5 beschriebenen Tabellen zur Auszeichnung linguistischer Phänomene, für die Kategorie *Substantive* also *Kasus*, *Genus*, *Numerus*, *Flexionsklasse* und *Nach Präposition* mit den jeweiligen Werten, für *Genus* z.B. *Femininum*, *Maskulinum* und *Neutrum*. Bei der Aktivierung einzelner Auswahlkästchen ist zu beachten, dass die einzelnen Werte mit *oder* verknüpft werden. Werden z.B. die Auswahlkästchen für die Kasus *Nominativ* und *Genitiv* aktiviert, so findet die entsprechende Datenbankabfrage sowohl Fragen, die Substantive im Nominativ erheben, als auch Fragen, die Substantive im Genitiv enthalten. Sollen Fragen gefunden werden, die Substantive im Nominativ *und* Substantive im Genitiv enthalten, muss zunächst eines der Auswahlkästchen – *Nominativ* oder *Genitiv* – aktiviert und der entsprechende Kasus durch Linksklick auf *Zur Abfrage hinzufügen* in die Abfrage übernommen werden. Anschließend können durch neuerliche Auswahl der Kategorie *Substantive* weitere Auswahlkästchen aktiviert und die jeweiligen Werte zur Abfrage hinzugefügt werden.

c) Anzeige der formulierten Datenbankabfrage

Nach Linksklick auf den Button *Zur Abfrage hinzufügen* wird im mittleren Teil der Seite die vom Nutzer formulierte Datenbankabfrage angezeigt, für die Werte in Abbildung 8.13 z.B.:

Bereich Abfrage Ergebnis	
Substantive	
Kasus:	Akkusativ
Nach Präposition:	Nein
<input type="button" value="Abschicken!"/> <input type="button" value="Abfrage zurücksetzen"/>	

Abbildung 8.14: Vom Nutzer zusammengestellte Datenbankabfrage/Suche nach Substantiven im Akkusativ

Durch Linksklick auf *Abschicken* kann nun nach Fragen gesucht werden, die Substantive im Akkusativ erheben, wobei es sich bei dem erhobenen Akkusativ nicht um einen Präpositionalkasus handeln soll. Sollte die Abfrage falsch formuliert worden sein, kann diese durch Linksklick auf *Abfrage zurücksetzen* gelöscht werden. Daneben können weitere Bedingungen zur Abfrage hinzugefügt werden, z.B. kann durch Auswahl der Kategorie *Verben* und Aktivierung des Auswahlkästchens *Vollverb* festgelegt werden, dass Fragen, die als Ergebnis zurückgegeben werden, sowohl ein Substantiv im Akkusativ als auch ein Vollverb enthalten sollen. Auf diese Weise können Fragen, die Vollverben in transitiver Verwendung erheben, gefunden werden:

Bereich Abfrage Ergebnis	
Verben:	
Kategorie:	Vollverb
Substantive	
Kasus:	Akkusativ
Nach Präposition:	Nein
<input type="button" value="Abschicken!"/> <input type="button" value="Abfrage zurücksetzen"/>	

Abbildung 8.15: Vom Nutzer zusammengestellte Datenbankabfrage/Suche nach transitiven Verben

Durch Linksklick auf den Link *Abfrage* oberhalb der Suchanfrage kann jederzeit von der Anzeige der Suchkategorien bzw. der Suchergebnisse zur Anzeige der Abfrage zurückgewechselt werden. Die Abfrage bleibt solange erhalten, wie der Nutzer nicht auf andere Menüpunkte, z.B. *Orte* oder *Fragen* wechselt oder die Abfrage durch Linksklick auf *Abfrage zurücksetzen* löscht.

Durch Linksklick auf *Abschicken!* wird die Datenbankabfrage gestartet, die Ansicht wechselt zur Anzeige der Abfrageergebnisse.

d) Anzeige der Abfrageergebnisse

Bereich Abfrage Ergebnis			
Auswählen Alle Keine			
	Nr.	Frage	Suggestierformen
<input checked="" type="checkbox"/>	0578	"Der Bauer mäht das Gras."	-
<input type="checkbox"/>	0579	"Ihr mäht das Gras."	-
<input type="checkbox"/>	0589	"Ich habe keinen Dangel mehr"	-
<input type="checkbox"/>	0615	"Man hat das Heu nass hereingetan."	-
<input checked="" type="checkbox"/>	0927	"Einen Vorteil geben"	-
<input type="checkbox"/>	1047	"Da bauen sie wieder eine neue Straße."	-
<input type="checkbox"/>	1053	"Er treibt das Vieh auf die Weide."	-
<input checked="" type="checkbox"/>	1115	"Heute sieht man die Berge."	-
<input type="checkbox"/>	1134	"Da haben sie einen Graben ausgegraben"	-
<input type="checkbox"/>	1135	"Wir graben einen Graben"	-
<input type="checkbox"/>	1159	"Er hat einen Hecht gefangen"	-
<input type="checkbox"/>	1426	"Man muß den Baum fällen"	-
<input type="checkbox"/>	1528	"Einen Stecken in den Boden stecken"	-

Abbildung 8.16: Ergebnis der Abfrage aus Abb. 8.15

Abbildung 8.16 zeigt einen Teil der Ergebnisse, die der Nutzer zu der in Abbildung 8.15 formulierten Abfrage erhält. Als Ergebnis werden dabei Fragenummer, -text sowie Suggestierformen derjenigen Fragen ausgegeben, auf die die Suchbedingungen zutreffen. Die einzelnen Fragen können wie unter 8.3.1.2. b) beschrieben zu den ausgewählten Fragen hinzugefügt werden.

e) Ausgewählte Fragen (4)

Die unter Punkt d) ausgewählten Fragen erscheinen anschließend im rechten Teil der Seite unter dem Punkt *Ausgewählte Fragen*. Um einzelne Fragen abzuwählen, ist wie unter 8.3.1.2. c) beschrieben vorzugehen, die ausgewählten Fragen bleiben auch bei einem Wechsel zwischen den einzelnen Menüpunkten (*Orte, Fragen usw.*) erhalten.

8.3.2. Die Anzeige der Suchergebnisse

Die Belege, die zu den zuvor ausgewählten Fragen in den gewählten Orten erhoben wurden, können alternativ im HTML-Format über den Menüpunkt *Ergebnisse* oder im PDF-Format über den Menüpunkt *Belegliste* angezeigt werden:

8.3.2.1. Der Menüpunkt *Ergebnisse*

The screenshot shows the BayDat website interface. At the top, there is a navigation bar with tabs for 'Orte', 'Fragen', 'Spezielsuche', 'Ergebnisse', and 'Belegliste'. The 'Ergebnisse' tab is selected. Below the navigation bar, there is a header section with 'Vollanzeige | Alle | Keine' and the title 'Frage 0552: "Unsere Katze"'. Below this, there is a list of language atlases: 'Sprachatlas von Bayerisch Schwaben: 062.001', 'Sprachatlas von Mittelfranken: 034.006', 'Sprachatlas von Niederbayern: 062.001', 'Sprachatlas von Oberbayern: 062.001', and 'Sprachatlas von Unterfranken: 062.001'. Below the list, there is a table with columns 'Ort', 'Beleg', 'Lautschrift', 'Lemma', and 'Scan'. The table contains three rows of results for 'Unsere Katze':

Ort	Beleg	Lautschrift	Lemma	Scan
<input type="checkbox"/> 97355 Castell	001	unə ^r kʰɑds	-	
<input type="checkbox"/> 97346 Dornheim	001	unərkʰɑds	-	
<input type="checkbox"/> 97348 Hüttenheim in Bayern	001	unərkʰɑds	-	

Below this table, there is another section for 'Frage 0553: "Unsere Katzen"'. It lists the same language atlases with ID 062.002. Below this, there is another table with columns 'Ort', 'Beleg', 'Lautschrift', 'Lemma', and 'Scan'. The table contains three rows of results for 'Unsere Katzen':

Ort	Beleg	Lautschrift	Lemma	Scan
<input type="checkbox"/> 97355 Castell	001	unə ^r kʰɑdsn	-	
<input type="checkbox"/> 97346 Dornheim	001	unərkʰɑdsn	-	
<input type="checkbox"/> 97348 Hüttenheim in Bayern	001	unərkʰɑdsn	-	

Abbildung 8.17: Der Menüpunkt *Ergebnisse*

Abbildung 8.17 zeigt die Seite, die nach Aufruf des Menüpunkts *Ergebnisse* erscheint. Sie gliedert sich in zwei Teile: der obere Teil (1) enthält die bekannte Navigationsleiste, zur Orientierung ist der Menüpunkt *Ergebnisse* grafisch hervorgehoben.

Im unteren Teil (2) werden die Belege für die zuvor ausgewählten Fragen und Orte im HTML-Format angezeigt – in der Abbildung die Belege zu den Fragen 552 und 553 der BSA-Fragekonkordanz aus den BSA-Erhebungsorten Castell, Dornheim und Hüttenheim in Bayern. Die Fragen und Orte entsprechen der Auswahl aus Abbildung 8.7. Wurden zuvor keine Fragen bzw. Orte ausgewählt und können deshalb keine Belege angezeigt werden, so wird der Benutzer durch die Meldung "Keine Frage(n) gewählt" bzw. "Kein(e) Ort(e) gewählt" darauf

hingewiesen und kann anschließend zu den Menüpunkten *Fragen*, *Spezialsuche* oder *Orte* wechseln.

Zu jeder gewählten Frage werden die BSA-Fragennummer und der Fragetext angezeigt, darunter folgen die Fragennummern aus den Fragebüchern des SBS, SMF, SNIB, SOB sowie SUF. Die Informationen werden hierfür aus den entsprechenden Feldern der Tabelle *Fragen* bezogen.

Für jeden Beleg werden der Erhebungsort mit Postleitzahl, die Belegnummer, der Beleg in Lautschrift und – soweit vorhanden – das Lemma angezeigt. Die Datenbankabfrage für diese Anzeige entspricht weitestgehend Beispielabfrage 2 unter 4.3.4. (Orte und Fragen werden je nach Benutzereingabe durch die ausgewählten Orte und Fragen ersetzt), allerdings wird der Inhalt des Feldes *Obligatorisch* der Tabelle *Frage* hier nicht überprüft, zusätzlich zur Abfrage unter 4.3.4. werden die Werte der Felder *Beleg_Nr* und *Lemma* aus der Tabelle *Belege* ausgegeben. Ist das Feld *TeuthoUC* für einen Beleg zu den zuvor ausgewählten Fragen und Orten leer, wird stattdessen die Begründung für den Nullbeleg, d.h. der entsprechende Wert aus dem Feld *Nullbeleg* der Tabelle *Belege* ausgegeben. Wurde eine ausgewählte Frage in den selektierten Orten nicht erhoben, wird dies durch die Meldung "Frage an diesem Ort nicht erhoben" entsprechend angezeigt.

Ein grüner Haken in der rechten Tabellenspalte zeigt an, dass für den entsprechenden Beleg ein Scan der originalen Fragebuchseite vorhanden ist. Eine GIF-Datei des jeweiligen Scans wird nach Linksklick in einem separaten Fenster angezeigt.

Durch Aktivierung der Auswahlkästchen in der linken Spalte der Tabellen können einzelne Belege für die Vollanzeige, d.h. eine Anzeige sämtlicher Kommentarfelder zum jeweiligen Beleg, selektiert werden. Durch Linksklick auf den Link *Alle* ober- bzw. unterhalb der Tabellen werden sämtliche Auswahlkästchen aktiviert, ein Linksklick auf *Keine* macht die Aktivierung rückgängig. Der Link *Vollanzeige* führt schließlich zur Vollanzeige der selektierten Belege:

Ort: 97659 Burgwallbach
Frage 0727: "Quer/"zwer(ch) wie wird die Anwand gepflügt?
Beleg: 001
Lautschrift:
-
Antwort Spontanfrage:
-
Lemma:
-
Bedeutung:
-
Grammatische Angaben:
-
Kommentar der Gewährsperson:
DIE ANWENDE WIRD QUER <i>BEIACKERN</i>
Kommentar des Explorators:
-
Kommentar zur Geltung der Antwort:
-
Angaben zur Exploration:
-
Angaben zur Gewährsperson:
-
Zweck:
-
Geographische Angaben:
Planquadrat: -
Ortsname: -
Grund für Nullbeleg:
-
Verweise:
Verweise auf Unterbelege:
084.U01

Abbildung 8.18: Beispiel für eine Vollanzeige

Abbildung 8.18 zeigt die Vollanzeige des ersten Belegs, der im SUF-Erhebungsort Burgwallbach zu Frage 727 aus der BSA-Fragekonkordanz, "'Quer'/'zwer(ch)' Wie wird die Anwand gepflügt?" erhoben wurde. Von der Vollanzeige aus kann durch die unter Verweise angegebenen Links zur Vollanzeige von Belegen gewechselt werden, auf die vom angezeigten Beleg aus verwiesen wird (Spontanbelege oder – wie in der Abbildung – Unterbelege). Zur Erstellung der Links wird auf die Tabelle *Verweise* der BAYDAT-Datenbank zurückgegriffen. In der Abbildung findet sich ein Verweis auf einen Unterbeleg: Die Vollanzeige des Unterbelegs kann auch durch Linksklick auf das jeweilige Lemma des Unterbelegs – in der Abbildung *Beiackern* im Kommentar der Gewährsperson – aufgerufen werden. Unterbelege werden in der grafischen Benutzerfläche also entsprechend der Beschreibung unter 4.3.2. umgesetzt.

8.3.2.2. Der Menüpunkt *Belegliste*

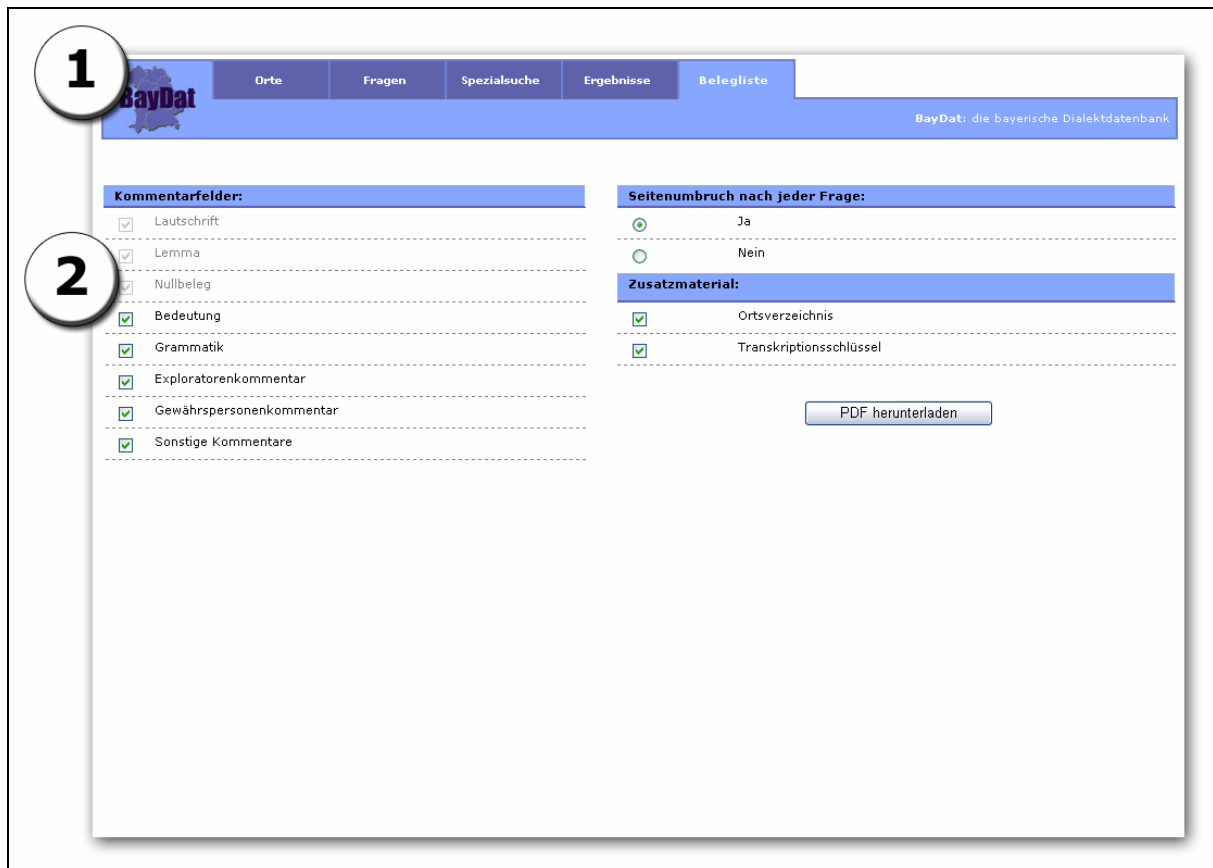


Abbildung 8.19 Der Menüpunkt *Belegliste*

Abbildung 8.19 zeigt die Seite, die nach dem Aufruf des Menüpunkts *Belegliste* erscheint. Der obere Teil der Seite (1) enthält die bereits beschriebene Navigationsleiste, der Menüpunkt *Belegliste* ist grafisch hervorgehoben.

Im unteren Teil der Seite (2) kann der Benutzer diejenigen Kommentarfelder auswählen, die in der Belegliste angezeigt werden sollen. Der Beleg in TEUTHOBD-Lautschrift bzw. die Begründung für Nullbelege sowie eventuell vorhandene Lemmata werden in jedem Fall in die Belegliste übernommen, die Punkte sind deshalb in der Auswahl grau dargestellt. Die Auswahlkästchen für die übrigen Kommentarfelder sind nach dem Aufruf des Menüpunkts *Belegliste* zunächst aktiviert. Sollen einzelne Kommentarfelder nicht in der Belegliste erscheinen, müssen die Auswahlkästchen entsprechend deaktiviert werden.

Soll eine Belegliste für mehrere Fragen erstellt werden, kann der Benutzer durch Aktivierung des Radio-Buttons *Ja* unter dem Punkt *Seitenumbruch nach jeder Frage* einen Seitenumbruch nach jeder Frage erzwingen. Da sich die Seitenumbrüche innerhalb der Frage nicht kontrollieren lassen, können bei Beleglisten für mehrere Fragen Überschriften am Seitenende bzw. der letzte Beleg für eine Einzelfrage auf einer neuen Seite stehen. Derartige Formatierungsfehler werden durch einen Seitenumbruch nach jeder Frage minimiert.

Schließlich können mit der Belegliste ein Ortsverzeichnis der ausgewählten Orte sowie ein Transkriptionsschlüssel für die TEUTHONISTA-Lautschrift ausgegeben werden (vgl. Anhang B). Das Ortsverzeichnis listet, nach BSA-Teilprojekten gruppiert, Ortssigle, Postleitzahl, Ortsname und Landkreis der in die jeweilige Belegliste aufgenommenen Orte auf. Der Transkriptionsschlüssel gibt, nach Vokalen und Konsonanten gruppiert, die Aussprache der einzelnen Zeichen der TEUTHONISTA-Lautschrift an.

Nach Aktivierung des Buttons *PDF herunterladen* wird die Belegliste im PDF-Format erzeugt und im ACROBAT READER oder der jeweils auf dem Rechner des Benutzers installierten Software zur Anzeige von PDF-Dokumenten angezeigt.

Jeder Belegliste wird zusätzlich zu den auswählbaren Elementen (Ortsverzeichnis und Transkriptionsschlüssel) ein Inhaltsverzeichnis vorangestellt. Die einzelnen Punkte des Inhaltsverzeichnis sind mit den tatsächlichen Unterpunkten (Fragen, Ortsverzeichnis, Transkriptionsschlüssel) verknüpft, durch Linksklick auf den jeweiligen Punkt im Inhaltsverzeichnis kann direkt zum entsprechenden Unterpunkt gewechselt werden.

Die Ausgabe der Ergebnisse als Beleglisten im PDF-Format hat gegenüber der Ausgabe der Suchergebnisse im HTML-Format den Vorteil der besseren Archivierbarkeit. Außerdem können aufgrund des Dateiformats bessere Ergebnisse beim Ausdruck erzielt werden. Allerdings nimmt die Erzeugung einer PDF-Belegliste softwarebedingt mehr Zeit in Anspruch als die Anzeige der Suchergebnisse im HTML-Format. Die Erzeugung von PDF-Beleglisten wird daher vor allem für die Archivierung von Suchergebnissen empfohlen, nicht jedoch für die schnelle Anzeige der Suchergebnisse von spontanen Suchabfragen der BAYDAT-Datenbank.

Der Ausdruck einer Belegliste der in Abb. 8.17 gezeigten Belege findet sich zur Veranschaulichung im Anhang B.

9. Abschluss und Ausblick

Ein Ziel des Projekts BAYDAT war es, die Erhebungsdaten des BSA an einer zentralen Stelle zusammenzuführen, um so ihre zukunftssichere Speicherung zu gewährleisten und die Daten außerdem für weitere Recherchen zu erschließen. Dies konnte durch die Aufbereitung der Erhebungsdaten – von der BSA-Fragekonkordanz über die BSA-Ortsliste hin zu den ASCII-Ortsdateien der sechs Teilprojekte – und ihrer strukturierten Speicherung in Form der BAYDAT-Datenbank gewährleistet werden. Durch die Auszeichnung der Fragen unter sprachwissenschaftlichen Gesichtspunkten ist es nun außerdem möglich, direkt auf sprachwissenschaftliche Phänomene zuzugreifen. Während die Suche nach relevanten Fragen und Belegen für die einzelnen Atlasbände ein recht zeitaufwendiges Unterfangen darstellte – die Projektfragebücher mussten für jedes Einzelphänomen durchgesehen werden, Probekarten und Beleglisten mussten erstellt und auf ihre Aussagekraft hin untersucht werden – kann der Nutzer der BAYDAT-Datenbank in Sekundenschnelle auf Fragen, die das gewünschte Phänomen beinhalten, zugreifen und sich entsprechende Beleglisten ausgeben lassen. Durch die theorieneutralen Auszeichnungen sind dabei einerseits relativ einfache Anfragen möglich, z.B. nach bestimmten Flexionsformen eines Verbs, andererseits können – wie in den zahlreichen Beispielabfragen gezeigt – durch die kombinierte Abfrage mehrerer Datenbanktabellen sehr komplexe Anfragen erstellt werden.

Die zentrale elektronische Speicherung bietet einen weiteren, bisher nur am Rande erwähnten Vorteil: Fehler, z.B. in den Kodierungen der ASCII-Ortsdateien können, anders als in den Originalfragebüchern, schnell und ohne großen Aufwand korrigiert werden. Durch die zentrale Speicherung können außerdem regelmäßig auftauchende Fehler über das gesamte BSA-Material gefunden und verbessert werden. Daneben können die Daten nach Bedarf ergänzt und erweitert werden, z.B. können in Zukunft weitere Auszeichnungen hinzugefügt werden. Außerdem könnten z.B. die Tonbandaufzeichnungen, die sämtliche Erhebungen des SUF dokumentieren, digitalisiert und mit den Belegdaten verknüpft werden, die transkribierten Belege könnten somit hörbar gemacht werden. Andere Ergänzungen zum ursprünglichen BSA-Material – die Scans der Originalfragebücher des SUF und des Fotoarchivs des UNTERFRÄNKISCHEN DIALEKTINSTITUTS – wurden in Kapitel 6 bereits vorgestellt.

Ein weiteres Ziel des Projekts BAYDAT war es, die Erhebungsdaten des BSA öffentlich zugänglich zu machen. Dieses Ziel konnte durch die Online-Schnittstelle der BAYDAT-Datenbank realisiert werden. Diese erlaubt den weltweiten Zugriff auf das BSA-Material, aufgrund der einfach zu bedienenden grafischen Benutzeroberfläche ermöglicht die Online-Schnittstelle außerdem einem breiten Benutzerkreis die Erschließung der BSA-Daten:

Laienbenutzer können sich z.B. schnell und unkompliziert über den dialektalen Wortschatz ihrer Heimatregion informieren, daneben können z.B. Sprachwissenschaftler recht komplexe Datenbankabfragen zusammenstellen und das Material so für ihre wissenschaftlichen Recherchen nutzen.

Die bisherigen Möglichkeiten, die die Online-Schnittstelle bietet – die Anzeige von Suchergebnissen im HTML-Format bzw. die Erstellung von Beleglisten im PDF-Format – können dabei durch weitere Anwendungen ergänzt werden, eine Einbindung dieser Anwendungen ist aufgrund der verwendeten Middleware COCOON einfach zu realisieren. So entsteht zurzeit im Rahmen des Projekts BAYDAT eine Diplomarbeit im Fach Informatik, die das Material mithilfe von *Data Mining* erschließt und die Ergebnisse in Form von automatisch erstellten Karten präsentiert. Eine Einbindung des dabei entstehenden Programms in die Onlineschnittstelle der BAYDAT-Datenbank wäre denkbar. Außerdem wäre die Hinterlegung des gesamten Datenmaterials in einem softwareunabhängigen Datenformat, z.B. XML wünschenswert, die Auswertung des Materials wäre somit unabhängig von der Onlineoberfläche und der verwendeten Software. Entsprechende Ausgaben lassen sich über die verwendete Middleware leicht erstellen. Die hier vorgestellten, bisherigen Möglichkeiten der Online-Schnittstelle können daher als erster Schritt gesehen werden, dem in Form von Anwendungen, die das Material unter weiteren Gesichtspunkten oder in anderer Form erschließen, weitere folgen sollen.

Literaturverzeichnis

- ABBEY, MICHAEL/COREY, MIKE/ABRAMSON, IAN: Oracle 9i für Einsteiger. München, 2002.
- ADLER, SHARON et al.: Extensible Stylesheet Language (XSL). Version 1.0. W3C Recommendation 15 October 2001. 2001. <<http://www.w3.org/TR/xsl/>> (29.5.2006)
- ALIPRAND, JOAN et al. (Hrsg.): The Unicode Standard. Version 4.0. Harlow et al., 2003.
- BARZ, IRMHILD: Die Wortbildung. In: Duden. Die Grammatik. Hrsg. von der Dudenredaktion. 7., völlig neu erarbeitete und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich, 2005, S.641 – 772. (=Duden, Bd.4)
- BERNER, ELISABETH: Mittelhochdeutsch. In: Schmidt, Wilhelm: Geschichte der deutschen Sprache. 9., verbesserte Auflage, erarbeitet unter der Leitung von Helmut Langner und Norbert Richard Wolf. Stuttgart, 2004, S.240 – 297.
- BUßMANN, HADUMOD: Lexikon der Sprachwissenschaft. 3., aktualisierte und erweiterte Auflage. Stuttgart, 1990.
- CABARGA, LESLIE: Logo, Font and Lettering Bible. A Comprehensive Guide to the Design, Construction and Usage of Alphabets and Symbols. Cincinnati, 2004.
- CLARK, JAMES (Hrsg.): XSL Transformations (XSLT). Version 1.0. W3C Recommendation 16 November 1999. 1999. <<http://www.w3.org/TR/xslt>> (29.5.2006)
- CODD, EDGAR FRANK: A Relational Model of Data for Large Shared Databanks. 1970. <<http://www.acm.org/classics/nov95/toc.html>> (29.5.2006)
- CONNOLLY, THOMAS/BEGG, CAROLYN: Database Systems. A Practical Approach to Design, Implementation and Management. 3. Auflage. Harlow et al., 2002.
- DUDEN. Deutsches Universalwörterbuch. Hrsg. vom Wissenschaftlichen Rat der Dudenredaktion. 4., neu bearbeitete und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich, 2001.
- EICHINGER, LUDWIG M. (Hrsg.): Sprachatlas von Oberbayern. 6 Bde. In Bearbeitung. (=Bayerischer Sprachatlas Regionalteil 6)
- EROMS, HANS-WERNER (Hrsg.): Sprachatlas von Niederbayern. 7 Bde. Heidelberg, 2005 ff. (=Bayerischer Sprachatlas Regionalteil 5)

- FLANAGAN, DAVID: JavaScript. Das umfassende Referenzwerk. 2. Auflage. Cambridge et al., 1997.
- FLEISCHER, WOLFGANG/BARZ, IRMHILD: Wortbildung der deutschen Gegenwartssprache. 2., durchgesehene und ergänzte Auflage. Tübingen, 1995.
- FRITZ-SCHEUPLEIN, MONIKA/KÖNIG, ALMUT/KRÄMER-NEUBERT, SABINE/WOLF, NORBERT RICHARD: Wörterbuch von Unterfranken. Eine lexikographische Bestandsaufnahme. 2. Auflage. Würzburg, 1997.
- GALLMANN, PETER: Der Satz. In: Duden. Die Grammatik. Hg. von der Dudenredaktion. 7., völlig neu erarbeitete und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich, 2005, S.773 – 1066. (=Duden, Bd.4)
- GALLMANN, PETER: Die flektierbaren Wortarten. In: Duden. Die Grammatik. Hg. von der Dudenredaktion. 7., völlig neu erarbeitete und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich, 2005, S.146 – 572. (=Duden, Bd.4)
- GLÜCK, HELMUT: Metzler Lexikon Sprache. 2., überarbeitete und erweiterte Auflage. Stuttgart/Weimar, 2000.
- HINDERLING, ROBERT/ROWLEY, ANTHONY/SCHEUERER, FRANZ XAVER (Hrsg.): Sprachatlas von Nordostbayern. 7 Bde. Heidelberg, 2004 ff. (=Bayerischer Sprachatlas Regionalteil 4)
- HORTON, IVOR: Java 2. JDK 5 Edition. 5. Auflage. Birmingham, 2005.
- HUNTER, DAVID et al.: Beginning XML. Birmingham, 2000.
- KLEPSCH, ALFRED: Sprachatlas von Mittelfranken. Bd.1. Einführung. Heidelberg, 2003. (=Bayerischer Sprachatlas. Regionalteil II. Sprachatlas von Mittelfranken, Bd.1)
- KLUGE, FRIEDRICH: Etymologisches Wörterbuch der deutschen Sprache. Bearbeitet von Elmar Seebold. 24., durchgesehene und erweiterte Auflage. Berlin/New York, 2002.
- KÖNIG, WERNER: Mittelhochdeutsche Wortliste. Vervielfältigtes Typoskript. Augsburg, o.J.
- KÖNIG, WERNER/WELLMANN, HANS (Hrsg.): Sprachatlas von Bayerisch-Schwaben. 13 Bde. Heidelberg, 1996 ff. (=Bayerischer Sprachatlas Regionalteil 1)
- KÖNIG, WERNER: Sprachatlas von Bayerisch-Schwaben. Bd.1. Einführung. Heidelberg, 1997. (=Bayerischer Sprachatlas. Regionalteil I. Sprachatlas von Bayerisch-Schwaben, Bd.1)
- KRUG, STEVE: Don't Make Me Think. A Common Sense Approach to Web Usability. Indianapolis, 2000.

- LANGHAM, MATTHEW/ZIEGLER, CARSTEN: Cocoon. Building XML Applications. Indianapolis et al., 2002.
- MOCZAR, LAJOS/ASTON, JEREMY: Cocoon. Developer's Handbook. Indianapolis, 2002.
- MUNSKE, HORST HAIDER/KLEPSCH, ALFRED (Hrsg.): Sprachatlas von Mittelfranken. 6 Bde. Heidelberg, 2003 ff. (=Bayerischer Sprachatlas Regionalteil 2)
- NAUMANN, HORST: Althochdeutsch. In: Schmidt, Wilhelm: Geschichte der deutschen Sprache. 9., verbesserte Auflage, erarbeitet unter der Leitung von Helmut Langner und Norbert Richard Wolf. Stuttgart, 2004, S.190 – 239.
- NIEDERMEIER, STEPHAN: Cocoon 2 und Tomcat. Bonn, 2004.
- NÜBLING, DAMARIS: Die nicht flektierbaren Wortarten. In: Duden. Die Grammatik. Hrsg. von der Dudenredaktion. 7., völlig neu erarbeitete und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich, 2005, S.573 – 640. (=Duden, Bd.4)
- PAUL, HERMANN: Mittelhochdeutsche Grammatik. Überarbeitet von Peter Wiehl und Siegfried Grosse. 24. Auflage. Tübingen, 1998. (=Sammlung kurzer Grammatiken germanischer Dialekte, A, Hauptreihe, Bd. 2)
- PAWSON, DAVE: XSL-FO. Sebastopol et al., 2002.
- RENN, MANFRED/KÖNIG, WERNER: Kleiner Bayerischer Sprachatlas. München, 2006.
- ROSEN, LAWRENCE E.: Open Software License V. 2.1. 2004. <<http://opensource.org/licenses/osl-2.1.php>> (29.5.2006)
- SCHWARTZ, RANDAL L./CHRISTIANSEN, TOM: Learning Perl. 2. Auflage. Sebastopol et al., 1997.
- THAU, DAVE: The Book of JavaScript. A Practical Guide to Interactive Web Pages. San Francisco, 2000.
- TIDWELL, DOUG: XSLT. Sebastopol et al., 2001.
- UNICODE INC. (Hrsg.): C0 Controls and Basic Latin. Range 0000 – 007F. 2005. <<http://www.unicode.org/charts/PDF/U0000.pdf>> (29.5.2006)
- UNICODE INC. (Hrsg.): C1 Controls and Latin-1 Supplement. Range 0080 – 00FF. 2005. <<http://www.unicode.org/charts/PDF/U0080.pdf>> (29.5.2006)
- UNICODE INC. (Hrsg.): Private Use Area. Range E000 – F8FF. 2005. <<http://www.unicode.org/charts/PDF/UE000.pdf>> (29.5.2006)

- VEEN, JEFFREY: *The Art & Science of Web Design*. Indianapolis et al., 2001.
- WALL, LARRY/CHRISTIANSEN, TOM/ORWANT, JON: *Programming Perl*. 3., durchgesehene und erweiterte Auflage. Sebastopol, 2000.
- WARRENDER, ROBERT L.: *Databases*. Exeter, 2003.
- WEINRICH, HARALD: *Textgrammatik der deutschen Sprache*. 2. revidierte Auflage. Hildesheim/Zürich/New York, 2003.
- WIESINGER, PETER: *Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten*. Band 2. Berlin, 1970. (=Studia Linguistica Germanica Bd.2/2)
- WOLF, NORBERT RICHARD/KRÄMER-NEUBERT, SABINE (Hrsg.): *Sprachatlas von Unterfranken*. 7 Bde. Heidelberg, 2005 ff. (=Bayerischer Sprachatlas Regionalteil 3)
- ZELDMAN, JEFFREY: *Designing with Web Standards*. Indianapolis et al., 2003.

Tabellenverzeichnis

2.1	Auszug aus der BSA-Fragekonkordanz	9
2.2	Kommentartrennzeichen in den ASCII-Ortsdateien	14
3.1	BSA-Erhebungsorte mit Postleitzahlen	17
3.2	BSA-Erhebungsorte mit Postleitzahl, Landkreis und Regierungsbezirk	17
3.3	Datentypen innerhalb der BAYDAT-Datenbank	18
3.4	BSA-Erhebungsorte mit Postleitzahl, Regierungsbezirk und Kfz-Kennzeichen ...	18
3.5	Landkreise und zugehörige Kfz-Kennzeichen	19
3.6	Ergebnisse der SQL-Beispielabfrage 1	21
3.7	Ergebnisse der SQL-Beispielabfrage 2	21
3.8	Ergebnisse der SQL-Beispielabfrage 3	21
4.1	Datenfelder und -typen der Tabelle <i>Fragen</i>	23
4.2	Datenfelder und -typen der Tabelle <i>Suggestierformen</i>	24
4.3	Datenfelder und -typen der Tabelle <i>Frageintentionen</i>	25
4.4	Datenfelder und -typen der Tabelle <i>Frageintentionen_Klartext</i>	25
4.5	Datenfelder und -typen der Tabelle <i>Sachbereiche</i>	25
4.6	Ergebnisse der Abfrage 1 "Obligatorische Fragen mit der Zeichenfolge <i>Kalb</i> " ...	26
4.7	Ergebnisse der Abfrage 2 "Morphologiefragen aus dem Sachbereich <i>Gemüse</i> " ...	27
4.8	Ergebnisse der Abfrage 3 "Suggestierformen mit der Zeichenfolge <i>Haus</i> "	28
4.9	Ergebnisse der Abfrage 4 "Obligatorische Fragen im Sachbereich <i>Katze</i> "	28
4.10	Datenfelder und -typen der Tabelle <i>Orte</i>	29
4.11	Datenfelder und -typen der Tabelle <i>Lk</i>	30
4.12	Datenfelder und -typen der Tabelle <i>Regbez</i>	30
4.13	Ergebnisse der Abfrage 1 "Erhebungsorte mit Postleitzahlen zwischen 97000 und 97200"	31
4.14	Ergebnisse der Abfrage 2 "Erhebungsorte im Landkreis Fürth"	32
4.15	Ergebnisse der Abfrage 3 "Unterfränkische Erhebungsorte mit Anfangsbuch- staben <i>P</i> "	32
4.16	Datenfelder und -typen der Tabelle <i>Belege</i>	33
4.17	Abkürzungen für Angaben zur Gewährsperson	36
4.18	Abkürzungen für Angaben zum Explorator und zur Exploration	36
4.19	Abkürzungen für Angaben zur Geltung der Antwort	37
4.20	Abkürzungen für Begründungen von Nullbelegen	37
4.21	Datenfelder und -typen der Tabelle <i>Verweise</i>	38

4.22	Ergebnisse der Abfrage 1 "Belege aus dem Sachbereich <i>Brot</i> "	42
4.23	Ergebnisse der Abfrage 2 "Belege mit der Zeichenfolge <i>Kalb</i> "	43
4.24	Ergebnisse der Abfrage 3 "Spontanbelege mit Zweckangabe <i>Lexik</i> "	44
4.25	Ergebnisse der Abfrage 4 "Unterbelege im Erhebungsort Hammelburg"	45
5.1	Datenfelder und -typen der Tabelle <i>Link_Vokale</i>	49
5.2	Kodierte Werte in der Tabelle <i>Link_Vokale</i>	50
5.3	Ergebnisse der Abfrage "Mhd. <i>ö</i> in offener Silbe vor stimmhaftem Plosiv"	51
5.4	Ergebnis der Abfrage "Mhd. <i>ie</i> vor Liquid"	52
5.5	Ergebnisse der Abfrage "Langmonophthonge im Hiatt ohne Hiatt-Tilger"	53
5.6	Datenfelder und -typen der Tabelle <i>Link_Konsonanten</i>	54
5.7	Kodierte Werte in der Tabelle <i>Link_Konsonanten</i>	55
5.8	Ergebnisse der Abfrage "Binnendeutsche Konsonantenschwächung"	56
5.9	Ergebnisse der Abfrage "Intervokalisches <i>h</i> "	57
5.10	Ergebnisse der Abfrage "Mhd. <i>m</i> nach <i>e</i> im Auslaut"	58
5.11	Datenfelder und -typen der Tabelle <i>Link_Verben</i>	61
5.12	Kodierte Werte in der Tabelle <i>Link_Verben</i>	62
5.13	Ergebnisse der Abfrage "Das Verb <i>sagen</i> im Perfekt"	64
5.14	Ergebnisse der Abfrage "Verben im Futur I"	64
5.15	Ergebnisse der Abfrage "Flexionsformen des Verbs <i>waschen</i> "	65
5.16	Ergebnisse der Abfrage "Partizip II von Verben der Ablautreihe 1a"	66
5.17	Datenfelder und -typen der Tabelle <i>Link_Substantive</i>	67
5.18	Kodierte Werte in der Tabelle <i>Link_Substantive</i>	67
5.19	Ergebnisse der Abfrage "Transitive Verben im Sachbereich <i>Essen und Trinken</i> " ..	68
5.20	Ergebnisse der Abfrage "Substantive im Akkusativ nach Präpositionen"	69
5.21	Ergebnisse der Abfrage "Kopula <i>sein</i> mit Substantiv"	70
5.22	Ergebnisse der Abfrage "Flexionsformen des Substantivs <i>Baum</i> "	71
5.23	Datenfelder und -typen der Tabelle <i>Link_Adjektive</i>	73
5.24	Kodierte Werte in der Tabelle <i>Link_Adjektive</i>	73
5.25	Ergebnisse der Abfrage "Attributive Adjektive im Sachbereich <i>Das Vieh und seine Pflege</i> "	75
5.26	Ergebnisse der Abfrage "Freistehende Adjektive im Komparativ"	75
5.27	Ergebnisse der Abfrage "Stark flektierte Adjektive ohne Nominativ"	76
5.28	Ergebnisse der Abfrage "Prädikative Adjektive ohne Kopula <i>sein</i> "	78
5.29	Datenfelder und -typen der Tabelle <i>Link_Pronomen</i>	79

5.30	Kodierte Werte in der Tabelle <i>Link_Pronomen</i>	80
5.31	Ergebnisse der Abfrage "Personalpronomen mit vom Bezugswort verschiedenem Kasus"	81
5.32	Ergebnisse der Abfrage "Relativpronomen"	81
5.33	Ergebnisse der Abfrage "Reflexivpronomen"	82
5.34	Ergebnisse der Abfrage "Demonstrativpronomen in der 3. Person Singular Maskulinum"	83
5.35	Datenfelder und -typen der Tabelle <i>Link_Artikel</i>	84
5.36	Kodierte Werte in der Tabelle <i>Link_Artikel</i>	84
5.37	Ergebnisse der Abfrage "Artikelwörter mit vom Bezugswort verschiedenem Genus/Numerus"	85
5.38	Ergebnisse der Abfrage "Artikellose Substantive in Sätzen"	87
5.39	Ergebnisse der Abfrage "Possessive Artikelwörter in der 1. Person Singular"	88
5.40	Ergebnisse der Abfrage "Endungslose Artikelwörter mit Adjektiven"	88
5.41	Datenfelder und -typen der Tabelle <i>Link_Subkon</i>	89
5.42	Kodierte Werte in der Tabelle <i>Link_Subkon</i>	89
5.43	Ergebnisse der Abfrage "Kausale Subjunktionen"	90
5.44	Ergebnisse der Abfrage "Konjunktionen im SBS und SUF"	91
5.45	Datenfelder und -typen der Tabelle <i>Link_Praepositionen</i>	91
5.46	Kodierte Werte in der Tabelle <i>Link_Praepositionen</i>	92
5.47	Ergebnisse der Abfrage "Modale Präpositionen"	92
5.48	Ergebnisse der Abfrage "Präposition <i>in</i> mit Verschmelzung"	93
5.49	Ergebnisse der Abfrage "Präpositionen mit Genitiv"	94
5.50	Datenfelder und -typen der Tabelle <i>Link_Adverbien</i>	95
5.51	Kodierte Werte in der Tabelle <i>Link_Adverbien</i>	95
5.52	Ergebnisse der Abfrage "Lokale Interrogativadverbien"	96
5.53	Ergebnisse der Abfrage "Richtungsadverbien auf <i>-wärts</i> "	97
5.54	Datenfelder und -typen der Tabelle <i>Link_Partikeln</i>	97
5.55	Kodierte Werte in der Tabelle <i>Link_Partikeln</i>	97
5.56	Ergebnisse der Abfrage "Negationspartikeln mit finitem Vollverb und Substantiv"	98
5.57	Ergebnisse der Abfrage "Gradpartikeln bei Adjektiven"	99
5.58	Datenfelder und -typen der Tabelle <i>Link_Wortbildungen</i>	103
5.59	Kodierte Werte in der Tabelle <i>Link_Wortbildungen</i>	103

5.60	Ergebnisse der Abfrage "Diminutivbildungen"	105
5.61	Ergebnisse der Abfrage "Determinativkomposita mit der Konstituente <i>Baum</i> " ...	106
5.62	Ergebnisse der Abfrage "Komposita im Sachbereich <i>Gemüse</i> "	107
5.63	Ergebnisse der Abfrage "Deverbale Suffixableitungen mit Suffix <i>-er</i> "	107
6.1	Datenfelder und -typen der Tabelle <i>Map_Scans</i>	108
6.2	Datenfelder und -typen der Tabelle <i>Fragebuch_Scans</i>	109
6.3	Ergebnisse der Beispielabfrage zu den Scans der Originalfragebücher	111
6.4	Datenfelder und -typen der Tabelle <i>Link_Bilder</i>	112
6.5	Datenfelder und -typen der Tabelle <i>Bildarchiv</i>	112
6.6	Ergebnisse der Beispielabfrage zu den Fotografien aus dem Bildarchiv des UNTERFRÄNKISCHEN DIALEKTINSTITUTS	114
8.1	Übersicht über Zeichen aus der TEUTHONISTA-Lautschrift mit UNICODE- (=TEUTHOUC) und ASCII-Codepunkten (=TEUTHOBD)	121

Abbildungsverzeichnis

2.1	Auszug aus der BSA-Ortsliste	10
2.2	Auszug aus der ASCII-Ortsdatei für den SUF-Erhebungsort Würzburg, <i>Wuerzbu</i>	11
4.1	Beziehungen zwischen den Tabellen zur BSA-Fragekonkordanz	26
4.2	Beziehungen zwischen den Tabellen zur BSA-Ortsliste	30
4.3	Beziehungen zwischen den Tabellen zu den ASCII-Ortsdateien	40
4.4	Beziehungen der Tabelle <i>Belege</i> zu anderen Tabellen	41
5.1	Schreibung der ausgezeichneten Vokale in der BAYDAT-Datenbank	48
5.2	Schematische Darstellung der Beziehungen der Tabelle <i>Link_Vokale</i>	50
5.3	Schreibung der ausgezeichneten Konsonanten in der BAYDAT-Datenbank	53
5.4	Schematische Darstellung der Beziehungen der Tabelle <i>Link_Konsonanten</i>	55
5.5	Schematische Darstellung der Beziehungen der Tabelle <i>Link_Verben</i>	62
5.6	Auszeichnung von Wortbildungen	102
5.7	Schematische Darstellung der Beziehungen der Tabelle <i>Link_Wortbildungen</i> ...	104
6.1	Schematische Darstellung der Beziehung der Tabellen <i>Map_Scans</i> und <i>Fragebuch_Scans</i>	109
6.2	Schematische Darstellung der Tabellen <i>Map_Scans</i> und <i>Fragebuch_Scans</i> zu anderen Tabellen der BAYDAT-Datenbank	110
6.3	Schematische Darstellung der Beziehung der Tabellen <i>Link_Bilder</i> und <i>Bild-</i> <i>rchiv</i>	113
6.4	Schematische Darstellung der Tabellen <i>Link_Bilder</i> und <i>Bildarchiv</i> zu anderen Tabellen der BAYDAT-Datenbank	113
8.1	Ausgabe eines BSA-Erhebungsortes in Form einer HTML-Tabelle	117
8.2	XML-formatierte Ausgabe eines BSA-Erhebungsortes	117
8.3	Darstellung von Belegen bei Nutzung der <i>Private Use Area</i> im MICROSOFT INTERNET EXPLORER	119
8.4	Darstellung mehrerer Diakritika über einem Grundzeichen in MICROSOFT INTERNET EXPLORER 6 (links) und MOZILLA FIREFOX 1.5 (rechts)	120
8.5	Startseite des BAYDAT-Internetauftritts	126
8.6	Der Menüpunkt <i>Orte</i>	127
8.7	Der Menüpunkt <i>Fragen</i>	130
8.8	Zusatzinformationen zu Frage 558 "Während der Paarungszeit tut der Kater..."	132
8.9	Verknüpfungen zu Frage 553 "'Unsere Katzen'"	133

8.10	Fragen mit dem Substantiv <i>Katze</i>	133
8.11	Anzeige von Verknüpfungen mit Bilddateien	134
8.12	Thumbnail und Detailinformationen zur Bilddatei <i>Stalltuer</i>	134
8.13	Der Menüpunkt <i>Spezialsuche</i>	134
8.14	Vom Nutzer zusammengestellte Datenbankabfrage/Suche nach Substantiven im Akkusativ	137
8.15	Vom Nutzer zusammengestellte Datenbankabfrage/Suche nach transitiven Verben	137
8.16	Ergebnis der Abfrage aus Abb. 7.15	138
8.17	Der Menüpunkt <i>Ergebnisse</i>	139
8.18	Beispiel für eine Vollanzeige	141
8.19	Der Menüpunkt <i>Belegliste</i>	142

Anhang A:

Ausdruck einer gescannten Fragebuchseite

(Seite 1 aus dem Fragebuch des SUF-Erhebungsortes Leubach)

8. OKT. 1991

DER AUFNAHMEORT

<p>2.1 Aufnahmeort (mundartl. Form) Na</p>	<p>lëbix (Leubach)</p>
<p>2.2 Bewohner desselben Mo</p>	<p>lëbix^h (Leubacher)</p>
<p>2.3 Die Orte der nächsten Umgebung (wichtig sind die Aufnahmeorte) mit Richtungsadverbien, die sich auf die Ortschaften beziehen Na, Mo ④ (die Stadt) *Fladungen*</p>	<p>① vrā[̄]ggəm? (Frankenheim) ② rüdšwinz (Rüdenswinden) ③ öwərvlōding? (Oberfladungen) ④ di šdōd ⑤ hovlōš (Huflos)</p>
<p>2.3Ba Spottnamen der Ortseinwohner und der Leute der Nachbarortschaften</p>	<p>① šdōlšāis¹ (Stallscheifer) "Leubach" ② dsixāina (Zigener) "Rüdenswinden" ③ gāsdāēsl "Oberfladungen" (bes. Stenese) ④ māolava (Maulaffen) "Fladungen"</p>
<p>2.3Bb Name des Flusses/Baches im Ort (mundartlich)</p>	<p>lëbixgrōwə (Leubachgraben)</p>
<p>2.3Bc Name des nächsten Hauptflusses (mundartlich) /Main/</p>	<p>šdrō (Streu)</p>

Anhang B:

Belegliste (entsprechend Abfrage Kp. 7)

Belegliste

Erstellt am 11.09.2006 um 11:13 Uhr.

Inhalt

0552: "Unsere Katze"	2
0553: "Unsere Katzen"	3
Ortsverzeichnis	4
Transkriptionsschlüssel	5

0552: "Unsere Katze"*Sprachatlas von Bayerisch-Schwaben: 062.001**Sprachatlas von Mittelfranken: 034.006**Sprachatlas von Niederbayern: 062.001**Sprachatlas von Nordostbayern: 062.001**Sprachatlas von Oberbayern: 062.001**Sprachatlas von Unterfranken: 062.001***97355 Castell (suf167cas):** ʏnə^rkʰɑ̃ds**97346 Dornheim (suf176dor):** unərkʰɑ̃ds**97348 Hüttenheim in Bayern (suf175hue):** unərkʰɑ̃ds

0553: "Unsere Katzen"*Sprachatlas von Bayerisch-Schwaben: 062.002**Sprachatlas von Mittelfranken: 034.007**Sprachatlas von Niederbayern: 062.002**Sprachatlas von Nordostbayern: 062.002**Sprachatlas von Oberbayern: 062.002**Sprachatlas von Unterfranken: 062.002***97355 Castell (suf167cas):** ʏnə^rkʰɑ̃d̥sn**97346 Dornheim (suf176dor):** unərkʰɑ̃d̥sn**97348 Hüttenheim in Bayern (suf175hue):** unərkʰɑ̃d̥sn

Ortsverzeichnis

Sprachatlas von Unterfranken (SUF)

Signle	PLZ/Ort	Landkreis
suf167cas	97355 Castell	Kitzingen
suf175hue	97348 Hüttenheim in Bayern	Kitzingen
suf176dor	97346 Dornheim	Kitzingen

Transkriptionsschlüssel

Die dialektalen Belege sind in der Teuthonista-Lautschrift transkribiert. Diese erlaubt eine genaue Darstellung unterschiedlicher Lautformen durch die Kombination von Zeichen des lateinischen Alphabets mit diakritischen Zusatzzeichen.

Eine detailliertere Darstellung der Teuthonista-Lautschrift findet sich im Einleitungsband des SBS (SBS, Bd. 1, 162f.).

Vokale

a, e, i, o, u, ö, ü neutrale, standardnahe Lautung

Schwa-Laute:

ə e-Schwa
 ɐ a-Schwa

Vokalquantität:

e Vokallänge ist unmarkiert
 ē lange Artikulation
 ê halblange Artikulation
 ě sehr kurze Artikulation

Vokalqualität:

ɛ geschlossene Artikulation
 ɛ̄ sehr geschlossene Artikulation
 ɛ̆ offene Artikulation
 ɛ̇ sehr offene Artikulation
 ɛ̈ zentralisierte Artikulation
 ẽ Artikulation mit leichter Lippenrundung
 ẽ̃ nasalierte Artikulation
 ĭ halvokalischer Gleitlaut (ähnlich *j*)
 ũ Gleitlaut zwischen *u* und *w*

KonsonantenVerschlusslaute:

b, d, g	Lenes
p, t, k	Fortes
ph, th, kh	behauchte Fortes
b̥, d̥, g̥	fortisierte Lenes
p̥, t̥, k̥	lenisierte Fortes
b̥, d̥, g̥	zwischen Verschluss- und Reibelaut

Reibelaute:

v	labiodental
s	dental-alveolar (<i>s</i> -Laut)
š	palatal (<i>sch</i> -Laut)
x	velar (<i>ach</i> -Laut)
χ	palatal (<i>ich</i> -Laut)
χ	palatovelar (<i>ech</i> -Laut)
ħ	schwach gerieben

Nasale und Liquide:

m	labial
n	dental-alveolar
ŋ	palatal/velar
r	"Zungenspitzen"- <i>r</i>
ɹ	einmal geschlagenes vorderes <i>r</i>
ʀ	"Zäpfchen"- <i>r</i>
l	standardnahe Artikulation
ɫ	velarisiert, retroflex
ɭ	halbvokalisch