

Zitate per Mausklick? Das Textkorpus zum WÖRTERBUCH DER BAIRISCHEN MUNDARTEN IN ÖSTERREICH (WBÖ) als leistungsstarkes Werkzeug für die lexikographische Praxis

Eveline Wendl-Vogt

Institut für Österreichische Dialekt- und Namenlexika,
Österreichische Akademie der Wissenschaften, Wohllebengasse
12-14/2, A-1040 Wien, Österreich

Abstract

The DATABASE OF ELECTRONIC TEXTS of the LEXICON OF BAVARIAN DIALECTS IN AUSTRIA (WBÖ) (TEXTKORPUS zum WBÖ)—The Lexicon of Bavarian Dialects in Austria (Wörterbuch der bairischen Mundarten in Österreich [WBÖ]) is based on a collection of about 4 million single records. They represent the variety of the regional, social and historical Bavarian dialects. About 10% of all entries are excerpts from texts of various types. The lexicographer has to draw the illustrative quotations from the original texts. The digitized full-texts have been dated and localized, thus each quotation can be placed in time and area enabling the lexicographer to choose a representative number of examples. For the definitions, the most appropriate and illustrative quotations have to be found, in order to actively support the lexicographer's work, the Institute of Lexicography of Austrian Dialects and Names / Institut für Österreichische Dialekt- und Namenlexika (<http://www.oeaw.ac.at/dinamlex>) started a new project: the so-called DBÖ (Database of the Bavarian Dialects in Austria / Datenbank der bairischen Mundarten in Österreich), financed by the Österreichische Akademie der Wissenschaften / Austrian Academy of Sciences. One main database includes the so-called *Hauptkatalog*, the main archive; there are additional databases which are necessary to retrieve correct information about the questionnaire, the localization of the word and the date of its recording. One component of the DBÖ is the database of electronic texts of the WBÖ, including some 90 Austrian texts spanning several centuries, representing the Austrian dialects. The original texts are scanned using the Austrian OCR-programme proLector V1.20 (A.1) which allows the training of various fonts. The machine-readable texts are converted in TUSTEP (Tübinger System von Textverarbeitungs-Programmen / Tübingen System of Text Processing Programmes). The TUSTEP-files get an alphanumeric key, which allows one to

retrieve each quotation from the database in a chronological order and to sort it according to its localization. Finally the texts are broken down into the requisite sized pieces for quoting in the WBÖ-entries. Using a special programme the quotations can easily be reconnected and replaced in the proper context from which they were drawn. The digital texts are a valuable source for the lexicographer freeing him from the monotony of checking over and over again, thus leaving him more time for his proper work, namely the writing of entries for the WBÖ. Furthermore, the digital texts are important for speeding up the dictionary's publication in accordance with the guidelines of the *Straffungskon-*

E-mail:

eveline.wandl-vogt@oeaw.ac.at zept 1993 and 1998.

0 Datenbankunterstütztes Wörterbuch versus digitales Wörterbuch: Ausgangspunkt

Die Nutzung der neuen Medientechnik für die Lexikographie geht heute weit über die digitale Erfassung des Textes und das Recherchieren in digitalen Quellenwerken hinaus. Das Schlagwort *Hypermedia* birgt gerade für Wörterbuchprojekte ungeahnte Möglichkeiten.

Die Überführung eines Jahrhundertprojekts vom Zettelkatalog auf neue Medien funktioniert weder so schnell noch so problemlos, wie dies bei einem im 21. Jahrhundert konzipierten Wörterbuch vorausgesetzt werden kann.¹

Im folgenden Beitrag beschränke ich mich auf die Konzeption und Nutzung des Textkorpus (TK) zum WBÖ für das Erstellen der Wörterbuchartikel, ohne dabei auf eine theoretische Weiterführung des TK für die Nutzung im Hyperraum jenseits der wörterbuchinternen Arbeit einzugehen.

1 Vom handschriftlichen Exzerpt zum Datenbankeintrag: Auszüge aus der Wörterbuchgeschichte²

1911 wurde in Wien die Kommission zur Schaffung des Österreichisch-Bayerischen Wörterbuches der damaligen Kaiserlichen Akademie der Wissenschaften gegründet, ebenso eine entsprechende Arbeitsstelle in München, mit dem Bestreben, den Wortschatz aller deutschen Dialektlandschaften systematisch zu erfassen und in Form von

großlandschaftlichen Mundartwörterbüchern lexikographisch aufzuarbeiten und für den bairischsprachigen Raum an die Stelle einer Neubearbeitung von Schmellers *Bayerischem Wörterbuch*³ ein gesamtbairisches Wörterbuch treten zu lassen.

1913 wurde mit der systematischen Materialerhebung begonnen, die über Jahrzehnte andauern sollte.

Das Bearbeitungsgebiet des *Wörterbuchs der bairischen Mundarten in Österreich (WBÖ)*, das v.a. die Basisdialekte, vereinzelt auch höherschichtige Sprachebenen, lexikographisch bearbeitet und dokumentiert, umfasst das heutige Österreich ohne das alemannische Vorarlberg, die von Altösterreich aus besiedelten Sprachinseln und jene bairischsprachigen Gebiete, die zur Zeit der Monarchie österreichisch waren, z.B. Südtirol und Südmähren⁴.

1963 wurde in Wien die erste Lieferung des WBÖ publiziert und das Vorhaben der redaktionellen Zusammenführung der Wörterbuchunternehmungen von Wien und München aufgegeben.⁵

Der so genannte Hauptkatalog (HK), ein rd. 4 Mio. Unikate umfassendes Zettelarchiv, stellt das Basismaterial für das WBÖ dar, dessen 'einzige, sprachwissenschaftlich, volkskundlich und kulturhistorisch bedeutsamen Belegsammlungen des Bairischen der Gegenwart und der Vergangenheit /.../ eine breite Dokumentation und eine umfassende Interpretation des regionalen Wortschatzes ermöglichen.'⁶

Die Diskussion um die hochgerechnete Bearbeitungszeit des Wörterbuchs führte 1993 bzw. 1998 zur Verabschiedung eines Straffungskonzeptes bzw. eines Neuen Straffungskonzeptes, das seit der 33. Lieferung (= 1. Lieferung des 5. Bandes) umgesetzt wird. In diesem Konzept wird unter Punkt III die

Beschleunigung der Publikation durch den Einsatz einer Datenbank beschrieben und in der Einrichtungssitzung vom 24. November desselben Jahres die Weiterführung der bestehenden TUSTEP-Datenbank (Datenbank der bairischen Mundarten in Österreich [DBÖ]) beschlossen.

W. Bauer und E. Kühn (1998, 371f.) skizzieren die Ziele der Digitalisierung und die Anwendungsmöglichkeiten der DBÖ, wobei ich mich in diesem Beitrag auf die Hilfestellung bei der Artikellarbeit am konkreten Beispiel eines historischen Literaturexzerptes beschränken werde.

Die Digitalisierung der Belege wird voraussichtlich 2008 abgeschlossen sein und dem WBÖ zu einer rascheren Publikation verhelfen. Der Abschluss der Publikation wurde mit 2020 festgesetzt.

2 Individualität und Standardisierung: Die unterschiedlichen Materialquellen des WBÖ

Im Wesentlichen können die Sammlungen des HK folgendermaßen gegliedert werden⁷:

- a) rd. 55% von Dialektolog(inn)en in direkter Methode erhobene und in Lautschrift notierte Mundartbelege
- b) rd. 35% von Gewährspersonen auf systematische Fragebogaussendungen (indirekte Methode) eingeschickte Antworten und von Laien zusammengestellte Wortschatzsammlungen
- c) rd. 10% Exzerpte aus Regionalwörterbüchern, Mundartdichtungen, einschlägigen dialektologischen Dissertationen und wissenschaftlichen Abhandlungen, Sprachatlanten, historischen Quellen u. dgl.⁸

Für diesen Beitrag von besonderem Interesse ist das Belegmaterial der Kategorie c), weshalb darauf näher eingegangen werden soll.

Obwohl beim ambitionierten Vorhaben, die *'literarischen Denkmäler der Mundart von der althochdeutschen Zeit ab bis heute'*⁹ aufzuarbeiten, bald Abstriche gemacht werden mussten, ist der Anteil der Literaturexzerpte im WBÖ beachtlich.

Im Hinblick auf eine datenbankunterstützte Belegbearbeitung sollen beispielhaft folgende Exzerpttypen hervorgehoben werden:

1. Exzerpte aus über Glossare erschlossenen historischen Quellen (B1)
2. Exzerpte aus Regionalwörterbüchern, die entweder ohne Seitenzahlen exzerpiert sind (B2a) oder auf Grund ihres Alters schützenswert erscheinen (B2b)
3. Exzerpte aus über Glossare erschlossener Mundartliteratur (B3)

3 Datenbankkonstruktion: Das Textkorpus und seine Stellung im Datenbankverbund

Im März 1993 begann man am Institut für Österreichische Dialekt- und Namenlexika (I DINAMLEX) nach intensiven Vorüberlegungen und Planungen mit der Dateneingabe in die DBÖ, wobei man sich für die Software des *Tübinger Systems von Textverarbeitungs-Programmen* (TUSTEP) entschied.¹⁰ Konzept und Realisierung der DBÖ sowie die Programmierung der für die Er- und Bearbeitung nötigen Anwendungen in TUSTEP liegen im Aufgabenbereich der wissenschaftlichen Mitarbeiter(innen) des Instituts.

3.1 Belegdatenbanken

Die Belegdatenbanken dienen als Beleg- und Informationsarchiv für unterschiedlichste Suchanfragen.

Im Zentrum des Datenbankverbundes steht die Hauptkatalogdatenbank (HKDb), die Belegdatenbank im engeren Sinne, die dem eins-zu-eins digitalisierten HK entspricht. Rund 1,5 Mio. Stichworteinträge, von denen viele Mehrfacheinträge mit phonetischen, semantischen oder kontextbedingten Sonderformen zu einem Stichwort darstellen, stehen derzeit¹¹ für die digitale Abfrage zur Verfügung.

Die angegliederte Pflanzennamendatenbank (PfNDb) entspricht dem digitalisierten Pflanzennamenskatalog, einer onomasiologisch angelegten Sammlung mundartlicher Pflanzenbezeichnungen.

Rund 31.000 Mundartbezeichnungen dokumentieren die reiche Fülle des Materials.

Die Bilddatenbank (BDb) ist eine ergänzende Belegdatenbank, in der rd. 1.800 Abbildungen abgespeichert sind.

Im Textkorpus (TK) zum WBÖ wird unterschiedliches Quellenmaterial im Volltext gespeichert. Näh. s.u.

3.2 Quelldatenbanken

In fünf Quelldatenbanken wird quellenkritische Arbeit archiviert, die bislang nicht systematisch zugänglich bzw. großteils nicht erarbeitet ist/war. Außerdem dienen diese Datenbanken als Basisdatenbanken für die programmgestützten Erweiterungen der HKDb.

In der Belegortdatenbank (ODb) sind die Namen sämtlicher im Quellenmaterial aufscheinender Belegorte gespeichert und über die Gemeindatenbank (GemDB) einer aktuell gültigen administrativen Einheit (Gemeinde) bzw. in der Gebietsdatenbank (GDb) einem für das WBÖ definierten Mundartgebiet zugeordnet.¹² Diese Datenbanken sind Basis einer angedachten kartographischen Verortung.¹³ Rund 4.000 Ortsangaben lassen auf die Größe und Dichte des Bearbeitungsgebietes schließen; rund 670 Mundartregionen sind ausgewiesen und ortsgenau abgegrenzt. Belegorte bzw. Mundartgebiete, die im WBÖ mit Gültigkeit der Straffungsrichtlinien nicht mehr berücksichtigt werden, sind systematisch gekennzeichnet.¹⁴

Im WBÖ zitierte bzw. für dessen Erstellung relevante Literatur wird in einer Quellen- und Literaturdatenbank (LDb) zusammengefasst, mit einem WBÖ-Kurzzitat versehen und – wenn möglich und sinnvoll – georeferenziert. Rund 2.200 Literaturzitate dokumentieren die reiche Belegstellen- und Sekundärliteratur.

Ausgangspunkt einer ersten Synonymensuche ist die Fragebogendatenbank (FbDb), in welcher die onomasiologisch angelegten Fragesammlungen für die systematischen Abfragen bei der WBÖ-Materialerhebung gespeichert sind. Derzeit stehen rund 17.000 Detailfragen aus 109 Fragebogen und über 700 Einzelfragen aus Ergänzungsfragebogen zur Abfrage zur Verfügung und sind mit der HKDb verknüpft.

Informationen über Personen, die in unterschiedlicher Weise an der WBÖ-Erstellung, von der Materialsammlung bis zur Artikelverfassung, beteiligt sind/waren, werden in der Mitarbeiter(innen)-datenbank (MDb) archiviert. Derzeit bekunden rund 3.900 Namen, z.T. samt kurzen Angaben zu Leben und Beitrag zum WBÖ (vereinzelt samt Datierung), die Vielfalt und Buntheit des HK-Bestandes. Weitere Angaben (z.B. Qualität der Sammlung, Besonderheiten der Umschrift, Interpretationshilfen für die Transkription etc.) sind nur sporadisch erfasst und harren einer erschöpfenderen Darstellung.

3.3 Aufbau des Einzeleintrags in der HKDb

Primäres Ziel der HKDb ist es, die Informationen des Originalbelegs digital abzubilden und einzelnen Feldern zuzuordnen. Dadurch wird der Zugriff auf einzelne Informationseinheiten des Originalbelegs ermöglicht. Die mit den Abb.1–4 korrespondierenden Einträge in die HKDb sind unter den Abb.6–9 einzusehen.

Anhand des HKDb-Eintrags *trischacken* soll die Datenbankmaske kurz erläutert werden (vgl. Tabelle 1).

3.4 Das Textkorpus im Besonderen

1995 wurde mit dem Aufbau eines TK zum WBÖ begonnen, um die Arbeit mit literarischen Exzerpten zu beschleunigen und somit den ambitionierten Zielen, die im Neuen Straffungskonzept von 1998 gesteckt worden waren, durch Datenbankunterstützung näher zu kommen. Dabei spielten einerseits Überlegungen bzgl. der generellen Verfügbarmachung (und Archivierung) von Texten eine Rolle (vgl. I. F. Castelli [1847]) als auch die Beschleunigung des Zugriffs (z.B. in über Glossare erschlossener Literatur, vgl. Sa.WeistGl., A. Matosch [1884]). Im Gegensatz zum Deutschen Rechtswörterbuch¹⁵ besteht das TK zum WBÖ nur aus maschinenlesbaren Texten; Bilddateien der jeweiligen Texte und/oder weiterer Quellen sind mit Ausnahme des digitalen WBÖ nicht vorhanden.

Das TK wird aus folgenden Korpora aufgebaut:

1. PrimärTK

Das PrimärTK fungiert als Belegdatenbank für das WBÖ; Näh. s.u.

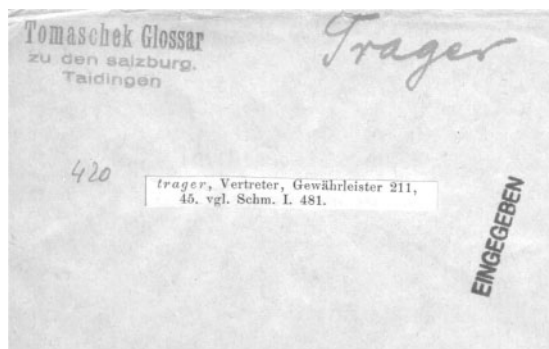


Abbildung 1 Beispiel 1: Exzerpt aus einem historischen Glossar

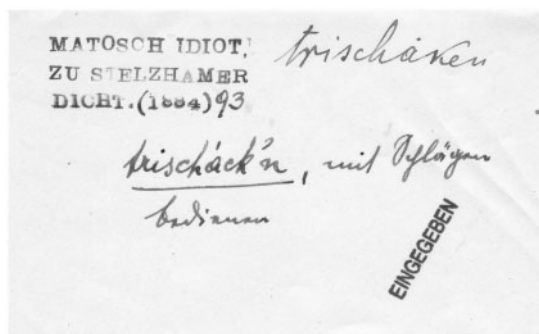


Abbildung 4 Beispiel 3: Exzerpt aus einem Glossar zu einer Mundartdichtung

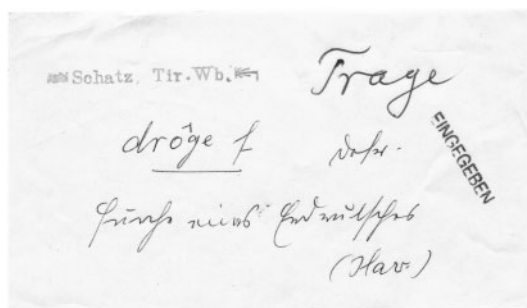


Abbildung 2 Beispiel 2a: Exzerpt aus einem Regionalwörterbuch

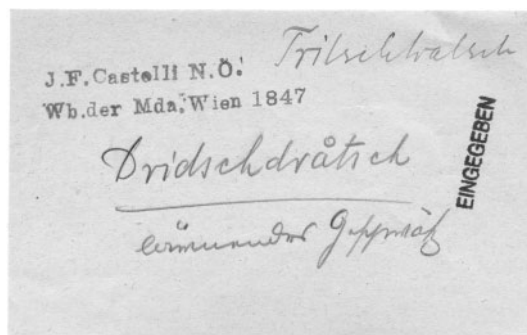


Abbildung 3 Beispiel 2b: Exzerpt aus einem historischen Wörterbuch

2. SekundärTK

Das SekundärTK soll relevante Sekundärliteratur zusammenstellen und der Lexikographin / dem Lexikographen als Informationsdatenbank

dienen. Es befindet sich derzeit in Planungs- und Testphase. Da das PrimärTK für die Artikelarbeit vordringlich anzusehen ist, wird dem SekundärTK auch bei der Bearbeitung sekundäre Priorität beigemessen.

3. Digitales WBÖ

Das digitale WBÖ ist in Bearbeitung. Sämtliche bisher erschienen WBÖ-Bände und Lieferungen stehen institutsintern als PDF-Dokumente mit hinterlegtem Volltext für Recherchen zur Verfügung. Im Jahr 2008 soll das WBÖ als elektronische Publikation im Rahmen des Forschungsportals des Verlages der Österreichischen Akademie der Wissenschaften¹⁶ angeboten werden.

In Konzeption – und vereinzelt im Test – sind Möglichkeiten diffizilerer Suchalgorithmen und komplexerer Abfragen, die einer späteren Ausbauphase der elektronischen Edition vorbehalten sein sollen.

4 Das Primärtextkorpus¹⁷: Aufbau und Voraussetzungen für die Abfrage

Im PrimärTK stehen der Lexikographin / dem Lexikographen derzeit 87 digitale Einzeltexte zur Verfügung, die in den meisten Fällen Einzelwerken entsprechen. Dabei wird ein weiter zeitlicher, geographischer und textsortenspezifischer Bogen gespannt. 64 Texte sind bereits für ein entsprechendes Artikelmakro¹⁸ aufbereitet, das eine Suche

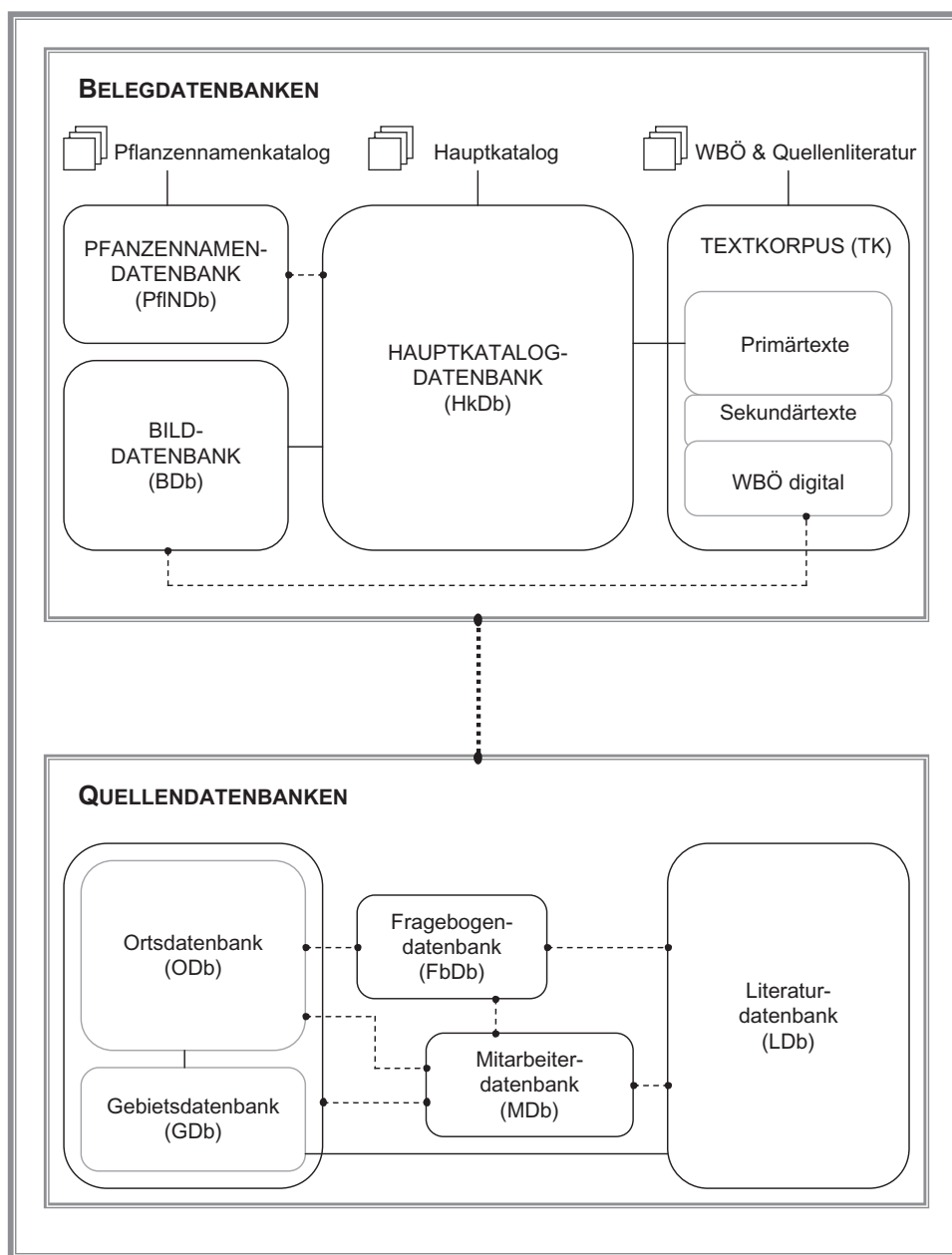


Abbildung 5 Aufbau der Datenbank der bairischen Mundarten in Österreich (DBÖ) [Ausschnitt]

via Mausfunktion in Volltexten ermöglicht. 23 weitere Texte stehen in Bearbeitung. Eine programmgestützte Recherche in diesen Texten ist nicht möglich und funktioniert über eine Eingabe eines TUSTEP-internen Kommandos.

Das PrimärTK ist mit den 87 angesprochenen Textdateien nicht abgeschlossen. Nach Maßgabe der finanziellen Mittel und je nach Verfügbarkeit der WBÖ-Mitarbeiter(innen) sollen weitere Texte digital verfügbar gemacht werden.

```
*A* HK 170, t1700723.sch#46
*HL* Trager:1
*QU* Tomaschek Glossar zu den salzburg. Taidingen *S* 420
*QN* {j1xxx:4} Sa. @ SaWeistGl. 420 [H-Weist.1; Wb:UrkWs]
===
*LT1* trager:1
*BD/LT1* Vertreter; Gewährleister
*****
```

Abbildung 6 Mit Abb. 1 korrespondierender Eintrag in der HKDb

```
*A* HK 168, t1680630.pir#2
*HL* Drage:1
*QU* Tir.Wb. *O* Defr.
*QN* {1} Tir. @ TirWb.(1955-1956) Bd., S. [HA-3602/1-2; MdaWb]
===
*LT1* dröge
*BD/LT1* Furche eines Erdrutsches
*ETO* slav.
```

Abbildung 7 Mit Abb. 2 korrespondierender Eintrag in der HKDb

```
*A* HK 182, t182#215.37 = t1821203.ned#132
*HL* Tritsch_tratsch:1
*QU* Castelli NÖ Wb. der Mda. Wien (1847)
*QN* {6} NÖ @ NöWb.(1847) S. [HA-3628; MdaWb]
===
*LT1* Dridschdrätsch
*BD/LT1* lärmendes Geschwätz
*****
```

Abbildung 8 Mit Abb. 3 korrespondierender Eintrag in der HKDb

```
*A* HK 182, t182#80.39 = t1821113.ned#119
*HL* trisch.äcken:6
*NL* trisch.äggen:6
*QU* Matosch Idiot. zu Stelzhammer Dicht. (1884) *S* 93
*QN* {5.1c,5.1d*j1874} minnv.:lnnv.:OÖ (v.1874) @ Stelzhld./M. S. [M-3707/IV; Wb:MdaDichtg]
===
*LT1* trischäck'n
*BD/LT1* mit Schlägen bedienen
*****
```

Abbildung 9 Mit Abb. 4 korrespondierender Eintrag in der HKDb

4.1 Arbeitsschritte vom Buch zum digitalen Text

Die Digitalisierung einer Textvorlage wird am I DINAMLEX im Wesentlichen in folgenden drei Arbeitsschritten bewältigt:

1. Vom gedruckten Text zum (Ausgangs-)File
2. Vom (Ausgangs-)File zum fehlerminimierten File
3. Vom fehlerminimierten File zum spezifischen WBÖ-Artikelfile

4.1.1 Vom gedruckten Text zum (Ausgangs-)File

Der gedruckte Text wird mittels des Programms proLector V1.20 (A.1) eingescannt. Die österreichische OCR-Software wurde gewählt, da das Programm über keine vordefinierte Wissensbasis verfügt. Jeder Zeichensatz¹⁹ kann genau trainiert werden. Auf diese Weise wird die Fehlerquote in Abhängigkeit von der Druckqualität möglichst niedrig gehalten.

Die in den wörterbuchspezifischen Belegtexten häufig vorkommenden Sonderzeichen können mit bis zu vier Zeichen codiert werden, wobei am I DINAMLEX ein vereinfachter TUSTEP-Code verwendet wird, der jenem der DBÖ entspricht und somit einerseits optimale Kompatibilität gewährleistet, andererseits das digitale Erfassen und Ausdrucken jedes bisher vorkommenden Sonderzeichens ermöglicht hat.

Der eingescannte Text, seitenweise als Dokument abgelegt, wird zusammengehängt und programm-unterstützt in TUSTEP konvertiert. Dafür steht eine Programmfolge zur Verfügung, die in Abhängigkeit von den Besonderheiten des jeweiligen Textes adaptiert wird.

Der Einzeltext ist originalgetreu wiedergegeben. Die Einheit der Zeile ist beibehalten, wird jedoch beim späteren Weiterverarbeiten für das Artikelschreiben nicht berücksichtigt, ausgenommen hiervon ist Mundartliteratur, bei welcher Schrägstriche als Trennungsstriche für Reim-(Zeilen-)Grenzen fungieren. Sonderzeichen werden codiert. Das Seitenende ist markiert.

4.1.2 Vom (Ausgangs-)File zum fehlerminimierten File

Das TUSTEP-File wird von studentischen Hilfskräften überarbeitet. Besonderes Augenmerk wird

Tabelle 1 Die Datenbankfelder der HKDb

Kennung	Erläuterung	Beispiel
A	Archiv-Feld:Lokalisierung des Einzelbelegs im Belegmaterial des HK und in der DBÖ.	*A* HK 182, t1821113.ned#119 Lade im HK: 182; Name der Datei, in der der Beleg gespeichert ist: t1821113.ned [Kürzel als Kennung für den bearbeitenden Datatypisten]; Belegnummer: 119
HL	Hauptlemma-Feld:Lokalisierung im alphabetischen Gefüge: Hauptstichwortansatz samt Wortartbestimmung.	*HL* trischacken:6(6 = Wortart = Verb)
NL	Nebenlemma-Feld:Dem Hauptstichwort (HL) untergeordneter Stichwortansatz samt Wortartbestimmung.	*NL* trisch.ággen:6 (codiertes Sonderzeichen)zur Lautung: <i>trischáckn</i>
QU	Quellenfeld (Original):Beleglokalisierung und Datierung lt. Originalbeleg des HK.	*QU* Matosch Idiot. zu Stelzhamer Dicht. (1884) *S* 93
QN	Quellenfeld (systematisiert):Einheitliche Beleglokalisierung und Datierung nach automatischer Verknüpfung mit den Quelldatenbanken. Siglenkennung.	*QN* {5.1c,5.1d*}1874} mInnv.:OÖ (v.1874) @ Stelzh.Id. S. [M-3707/IV; Wb:MdaDichtg]
===	Trennmerkmal zwischen Kopf und Belegteil der Maske.	
LT/*KT*	Belegfeld:Lautung/Belegzitat lt. Originalbeleg des HK.	*LT* trischáckn
BD/LT/*BD/KT*	Bedeutungsfeld (zum angegebenen Beleg): Bedeutungsangabe zur Lautung <i>n</i> (hier 1) lt. Originalbeleg des HK.	*BD/LT* mit Schlägen bedienen
ETO/*ETB*	Eymologiefeld:Etymologische Angaben lt. Originalbeleg (*ETO*) bzw. lt. Bearbeiter(in) (*ETB*)	[keine Angabe]

Tabelle 2 Anteil der unterschiedlichen Textsorten im PrimärTK zum WBÖ

Textsorte	Anzahl aufbereitet ²⁹	digitalisiert ³⁰
Historische Gebrauchstexte	19	8
Historisch-literarische Texte ³¹	2	1
Historische Wörterbücher, Idiotika	14	1
Regionalwörterbücher ³²	12	6
Texte zur (hist.) Umgangssprache	1	3
(Historische) Mundartliteratur	16	3
Texte zu Volkskunde und Brauchtum	–	1
Gesamt	64	23
	87	

auf die richtige Codierung und Umsetzung der Sonderzeichen gelegt.

4.1.3 Vom fehlerminimierten File zum spezifischen WBÖ-Artikelfile

In einem dritten Schritt weisen studentische Hilfskräfte in Absprache mit den Wissenschaftler(inne)n

dem Text (oder Einzelabschnitten des Textes) eine gültige Belegstellenangabe zu. Die Belegstellenangabe enthält zusätzlich zum eigentlichen Kurzzitat einer Textquelle lt. Literaturverzeichnis, Angaben zur Lokalisierung und Datierung eines Belegs und gibt Hinweise zu dessen qualitativer Einstufung. Mittels einer alphanumerischen Sigle wird die WBÖ-analoge Sortierung²⁰ des Einzelbelegs in der DBÖ sichergestellt.

Abschließend wird der auf diese Weise bereinigte und bearbeitete Text programmunterstützt in Einzelzitate zerlegt. Die Zitatsgrenzen sind textsortenabhängig durch Absätze (Wörterbücher), Satzzeichen wie Punkte (z.B. Mundartliteratur) u.ä. definiert.

Einen Einblick in den Aufbau eines Einzeltextes des PrimärTK gibt Abb.10 (s.u.).

4.2 Die Sigle

Die Sigle, ein alphanumerischer Code, ermöglicht einerseits die programmgestützte Verwaltung der DBÖ-Einträge z.B. nach räumlichen, zeitlichen,

** [1565:U\5.2b] versatzung, verpfantung, beständen verlassen, procureien, rechtfertigung hindergengn, verträgen und verschreibungen mit unvogtpern kinden handln soll, dergleichen mit leuten, die mit rêt, gehörn, gesicht und ierer vernunft halben geprüchlich und manghaft sint etc., und die sach etwas ansechlich, namhaft und trëfflich ist, so soll man die durch freund und ordenliche herschaft mit gewaltigen gerhaben oder mit versorgern, tragern, vormun- den, beistand und verantwurtern notturflich versehen, wie sich zu sölli-
^# Seitenende
chem gebürt, sonst ön daz wären sollich handlung unbestantig, unpündig und krafftlos aus den gnaden und freiheitn, damit sölich leut durch ge- maine recht begnat und befreit sint. ÖWeist. 1,212 (öUPinzg. 1565)@@

** [1699X:U\5.5a] 36. Articul: Hausierer, kräxentrager betreffent. ÖWeist. 1,127 (swFlachg. 17.Jh.)@@

** [1699X:U\5.5a] Die unangesessnen, umbschweifenden hausierer, kroxentrager sein wie in a. 1629 publicirter ord- nung bei dem 13. articul zu sechen obgeschofft, mit auftrag, das an maut- steten und die uberreiter auf sie guete achtung geben, ire kräxen durch- suecht und da sich waß ungleiches dorin bfindt eingezogen werden. ÖWeist. 1,127 (swFlachg. 17.Jh.)@@

Abbildung 10 Mit Abb. 1 korrespondierender Eintrag im TK zum WBÖ

Tabelle 3 Form der Sigle in Abhängigkeit von der Textsorte

Beispiel	Quelle	Textsorte	Sigle
B1	SaWeistGl.	Historischer Text	{j1xxx:4} Sa. @ SaWeistGl. [H-Weist.1; Wb:UrkWs]
B2a	TirWb.(1955–1956)	Regionalwörterbuch	{1} Tir. @ TirWb.(1955–1956) [HA-3602/1-2; MdaWb]
B2b	NöWb.(1847)	historisches Regionalwörterbuch	{6} NÖ @ NöWb.(1847) [HA-3628; MdaWb]
B3	StelzhId./M.	Glossar zu Mundartliteratur	{5.1c,5.1d*j}1874} mInnv.:Inn::OÖ (v.1874) @ StelzhId./M. [M-3707/IV; Wb:MdaDichtg]

sammler(in)- und textsortenspezifischen Bezügen, als auch die für das Artikelschreiben wesentliche Sortierung der Einzeleinträge nach WBÖ-relevanten Kriterien.

Die Sigle ist zweigeteilt. Im ersten Teil, in geschwungenen Klammern, erfolgt mittels sogenanntem Gebietskey eine geographische und – gegebenenfalls – zeitliche Zuordnung. Im zweiten Teil, der in eckigen Klammern steht, eine Systematisierung.

Konkret können drei Formen der Sigle unterschieden werden, die der Lexikographin / dem Lexikographen einen Hinweis über die Relevanz des einzelnen Belegs für die Klassifizierung einer Lautungsangabe geben:

1. {jJahreszahl:Gebietskey} Gebietsangabe @ Kurzzitat [Systematik] Der Beleg ist nicht direkt für die Lautungsbearbeitung benützbar, z.B. ein Exzerpt aus einer historischen Quelle, vgl. B1 u. Abb.6.

2. {Gebietskey*jJahreszahl} Gebietsangabe @ Kurzzitat [Systematik] Der Beleg ist (bedingt) auch für die Lautungsbearbeitung relevant, z.B. ein Exzerpt aus einem Werk der Mundartliteratur, vgl. B3 u. Abb.9.

3. {Gebietskey} Gebietsangabe @ Kurzzitat [Systematik] Der Beleg ist definitiv für die Lautungsbearbeitung von Bedeutung, z.B. Exzerpt aus einem Regionalwörterbuch, vgl. B2a u. Abb.7.

4.2.1 Der Gebietskey

Die Gebietsangabe richtet sich generell nach der kleinsten erschließbaren Raumeinheit, welcher ein Einzeleintrag zuzuordnen ist. Während im WBÖ außer in Ausnahmefällen nur noch Kleingebiete zitiert werden, sind die kleinsten in der Sigle der DBÖ erfassten Einheiten Gemeinden. Im Text des *QU*- bzw. *QN*- Feldes der DBÖ sind jedoch

Tabelle 4 Der Gebietskey

Gebietskey	Auflösung	Sonstiges
1C.1j02	Gm. Ischgl	1(Tirol)C(Nord) ³³ .1(Westtirol)j(Oberpaznaun)02(Ischgl)
6.2j15	Gm. Pottendorf	6(Niederösterreich).2(Wiener Becken)j(südlich)15(Pottendorf)
6	Gm. Wien	
0.1a01	Gm. Enego [Jeneve]/IT	0(sbair. Außenmundart).1(Sprachinsel)a(Sieben Gemeinden/Sette Comuni)01(Enego)
0.5a01	Gm. Samnaun/CH	0(sbair. Außenmundart)5(an sbair. Bearbeitungsgebiet anschließend)a(Samnaun)01(Samnaun)
8.1a08	Gm. Nová Pec [Neuofen]/CZ	8(mbair. Außenmundart).1(Böhmerwald)a(unterer Böhmerwald)08(Nová Pec)
wb0	im WBÖ nicht zu berücksichtigende Orte/Gebiete	Beispiel: unteres Laßnitztal - Verweis auf WBÖ-Gebietsbezeichnung: <i>sMSt.</i> (südliche Mittelsteiermark)
ALT	mit Gültigkeit der Neuen Straffungsrichtlinien von 1998 nicht mehr bearbeitete Orte/Gebiete	Beispiel: Nordbairische Sprachinsel Egerland/CZ

auch kleinere lokale Einheiten (z.B. Ortschaften, Einzelhöfe) lt. Originalbeleg erfasst.

So bleibt gewährleistet, dass in der DBÖ keine Informationen des HK-Originalzettels verloren gehen.

Der Gebietskey setzt sich wie folgt zusammen:

1. Die 1. Zahl ist die Großregionkennzahl. Diese entspricht in Österreich (1-7) weitgehend dem Bundesland. Ausnahmen bilden das in Nord und Ost geteilte Tirol, das mit 1B und 1C definiert ist und dem auf Grund sprachlicher Kriterien auch das ehemalige Südtirol mit 1A zugeteilt ist; ebenso Niederösterreich und Wien, die mit der Kennzahl 7 zusammengefasst sind. Die Kennzahlen 0, 8 und 9 sind ausschließlich außerösterreichischen Beleggebieten vorbehalten. Die Kennzahl 0 fasst südbairische Sprachinseln zusammen, wie z.B. Sappada [Pladen] (I), aber auch die an südbairisches Sprachgebiet angrenzende Gemeinde Samnaun (CH). Die Kennzahl 8 steht für an österreichisches Staatsgebiet angrenzende mittelbairische Dialektgebiete, das ehemalige Böhmen und Mähren, Teile der westlichen Slowakei, Westungarns und des nordöstlichen Slowenien. Die Kennzahl 9 definiert mittelbairische Sprachinseln, wie z.B. Brno [Brünn] (CZ).
2. Die alphanumerische Kombination nach dem Punkt ergibt die Kleinregionensigle; sie steht

für ein Klein-Gebiet, eine bestimmte Gegend oder Talschaft.

3. Abgeschlossen wird der Gebietskey durch die zweistellige Gemeindekennzahl, die den Einzeleintrag der kleinsten administrativen Einheit zuordnet.
4. Belege außerhalb des (aktuellen) WBÖ-Untersuchungsgebiets²¹ werden mit wb0* eingeleitet.
5. In den Lokalisationsdatenbanken sind darüber hinaus Gebiets- und Siedlungsbezeichnungen festgehalten, die in der bisherigen Publikation verwendet worden sind, die jedoch nicht der neuen Definition entsprechen. Diese sind mit ALT ausgewiesen.

4.2.2 Die Systematik

Um der Lexikographin /dem Lexikographen und potentiellen Zweitnutzer(inne)n der DBÖ wertvolle Informationen zwecks Belegqualifizierung zukommen zu lassen, werden diese in einer zweiten Kennung verpackt, dem zweiten Teil der Sigle.

Der Systematisierung des Einzeleintrags vorangestellt ist die institutsinterne Bibliothekssignatur, die das rasche Auffinden des Originals gewährleisten soll. Die eigentliche Systematisierung des Einzeleintrags erfolgt über Kürzel, die Auskunft über die Art des zu Grunde liegenden Materials geben sollen, z.B.: Dichtg = Dichtung; Mda = Mundart; Urk = Urkunde; Wb = Wörterbuch; Ws = Wortschatz

5 Der Einsatz des TK für die praktische Artikelarbeit: Das Beispiel *TRAGER*

E.-M. Jakobs (1999, 327) beurteilt den ‘Rückgriff auf andere Texte nicht nur [als] typisch, sondern [als] eine wesentliche Voraussetzung’ wissenschaftlichen Textproduzierens. Zweifelsohne ist die umfangreiche Dokumentation historischer Belegstellen und literarischer Beleg-Ausschnitte zu einem Lemma, bzw. konkreter zu einer Bedeutungsangabe, auch ein wesentliches Qualitätsmerkmal des WBÖ.

Am Beispiel des Glossarbelegs *trager* soll die Nutzung des Textkorpus für die praktische Artikelarbeit zusammenfassend dargestellt werden.

5.1 Der Originalbeleg

Ausgangsbasis ist der Originalzettel aus dem HK, hier zum Beispiel ein Literaturexzerpt aus dem Tomaschek-Glossar zu den salzburgischen Taidingen (Sa.WeistGl.; vgl. B1 u. Abb.1) und dessen Entsprechung in der HKDb (vgl. Abb.6).

5.2 Die Zitat-Recherche

Eine erste Recherche im TK verhilft zur Klärung, ob es sich um einen Beleg handelt, der in das WBÖ aufgenommen werden soll. Im Falle einer positiven Beurteilung wird zur Überprüfung der buchstabengetreuen Wiedergabe bzw. zur Auswahl eines aussagekräftigen Vollzitats ebenso wie zur Erfassung der Bedeutung in den digitalisierten Text Einschau gehalten.

Die am I DINAMLEX erarbeitete Programmsequenz *SuchLit* wird über die Mausleiste am Arbeitsbildschirm der Lexikographin / des Lexikographen gestartet.

Nach Auswahl des entsprechenden Textes aus dem TK können bis zu 9 Suchbegriffe in die *Eingabemaske*²² eingetragen werden.

An Hand des Exzerpts *trager* im HK und dem dementsprechenden Beleg in der HKDb lässt sich zeigen, wie die Suche angelegt werden muss, um erfolgreich zu sein. Die key-sensitive Suche nach *_Trager_*²³, der dem Exzerpt entspricht, weist keinen Beleg in der angegebenen Quelle aus. Nur wenn die Suche weiter gefasst wird (z.B. mit *trager*²⁴),

Trager ... — †d) Vorsteher, Sprecher, Bevollmächtigter f. best. Personen(gruppen): *mit versorgern, tragnern, vormunden* öUPinzg. (1565) Ö.Weist. 1,212; weitere Komp. s. DBÖ: †(Ge-mēins-pflicht)-, †(Treus)-, †(Klag)-, †(Lēhen[s])-, †(Leib)-, †(Ge-mēins)-, †(Pērg-rēcht)-, †(Schērm)-, †(Ob-sicht[s])-, (Ge-walt)-.

Abbildung 11 Mit Abb. 1 korrespondierender Ausschnitt aus dem WBÖ-Artikel *Trager*

wird der gewünschte Textausschnitt gefunden, der im Original in der Form *tragnern* belegt ist. Außerdem werden auf diese Weise zwei Komposita erfasst, die in manchen Fällen für die Bedeutungszuordnung relevant sein können: *kräxentrager* bzw. *kroxentrager*. Das Belegwort *tragnern* kann samt Textumgebung und Siglenkennzeichnung eingesehen und für die Bedeutungsbeschreibung verwendet werden. Es ist im herauskopierten Textauszug im Sinne der Benutzerfreundlichkeit farbig unterlegt.

5.3 Einbau in die Bedeutungsparaphrase

Innerhalb der jeweiligen Bedeutungsangabe wird der aussagekräftigste Beleg bestimmt, wobei als Auswahlkriterien Schreibung, Kontext, Zuverlässigkeit, sowie räumliche und zeitliche Zuordnung fungieren.

Dem Sprachgefühl der Lexikographin / des Lexikographen obliegt unter Beachtung der WBÖ-Richtlinien die Auswahl eines Beleg-Ausschnitts als lexikographisches Beispiel für das Lemma oder eine bestimmte Bedeutungsangabe.

6 Erwartungshaltungen an das TEXTKORPUS und deren Erfüllung bzw. Erfüllbarkeit

Bereits im 3. Band des WBÖ finden sich im Vorwort Überlegungen zur Straffung der Artikel, um ‘das Voranschreiten des Wörterbuches dadurch zu beschleunigen’. Ebenda wird unter anderem angeregt, ‘wo es ohne Minderung der für einen Wörterbuchartikel relevanten Aussage möglich ist, /.../ bei der

Zitierung mehrfacher historischer Belege /.../ eine etwas knappere Artikelgestaltung' anzustreben.

Diese Überlegungen wurden im Straffungskonzept von 1993 und später im Neuen Straffungskonzept von 1998 weitergeführt und konkretisiert. Die erschöpfende Behandlung historischer und mundartlicher Belegstellenliteratur ist zeitaufwändig und druckseitenintensiv, weshalb sie als idealer Anknüpfungspunkt für eine Straffung angesehen werden kann. Andererseits ist die Dokumentation schriftlicher Quellen für ein Belegwort ein entscheidendes Qualitätsmerkmal des WBÖ.

6.1 Artikelwertigkeit

'1.2. Bei der Artikelausarbeitung nicht mehr zu berücksichtigen, aber gegebenenfalls mit Verweis auf die Datenbank anzuführen sind:

*1.2.1. Isolierte und historische Wörter einschließlich Ableitungen und Komposita vor 1800.'*²⁵

Vorab ist zu entscheiden, ob ein Beleg überhaupt für die Artikelbearbeitung relevant ist und im WBÖ behandelt werden soll oder – falls nicht – ob er (zumindest) durch einen im WBÖ aufgenommenen Verweis auf die DBÖ erfasst werden soll.

Da sich die Lexikograph(inn)en beim WBÖ nur auf das im I DINAMLEX vorhandene Belegmaterial stützen, ist eine Recherche in der DBÖ oftmals ausschlaggebend für die Auslese der sogenannten artikelwertigen Belege bzw. Stichwörter. Wesentlich unterstützt wird die Auswahl durch die Lemmaliste, die jeder Bearbeiterin/jedem Bearbeiter in digitaler und ausgedruckter Form vorliegt.

Darüber hinaus kann durch die rasche elektronische Recherche im TK der Begriff des isolierten Vorkommens häufig verifiziert oder falsifiziert werden und damit eine, zumindest relative, Quantifizierung des vorhandenen Belegmaterials vorgenommen werden, die ebenfalls entscheidend dazu beiträgt, ein Wort als artikelwertig oder nicht artikelwertig einzustufen.

6.2 Belegauswahl

'1.3.1. Von einer ausführlichen Darbietung der vorhandenen historischen Belege ist abzusehen;

*im allgemeinen wird nur ein historischer Beleg aus jedem Jahrhundert pro Bedeutung angeführt.'*²⁶

Innerhalb eines Artikels muss das zitierfähige Material nach unterschiedlichen Kriterien sortiert werden:

1. Bedeutungsverifizierung der Belegstelle und Zuordnung zu der im Artikel erarbeiteten Bedeutungsparaphrase: Die Formulierung einer Bedeutungsangabe und Zuordnung der Belegstellen zu einzelnen Bedeutungsparaphrasen ist die wesentliche Arbeit der Lexikographin/des Lexikographen. Sie kann durch Datenbankunterstützung beschleunigt, jedoch nicht automatisiert vorgenommen werden.
2. Datierung und Lokalisierung der Belegstelle: Datierung und Lokalisierung einer Belegstelle gestalten sich im besonderen bei den historischen Quellen häufig als schwierig, z.B. bei den für das WBÖ relevanten österreichischen Weistümern. Deren Urkundensammlungen weisen mehrfach eine komplexe Überlieferungsgeschichte auf; Herrschaften, die Dokumente ausstellten, sind abgekommen. Der fehleranfällige und zeitaufwändige Vorgang dieser Datierungs- und Lokisierungsarbeiten, der früher von jeder Lexikographin / jedem Lexikographen bei jedem ausgewählten Zitat vollzogen werden musste, um festzustellen, ob ein Beleg überhaupt für das WBÖ relevant ist (räumlicher und zeitlicher Rahmen) wird jetzt in einem einzigen Arbeitsgang durch die oben besprochene Auszeichnung mittels Sigle vorgenommen.

6.3 Belegdarbietung

*'2.5. Kürzungen des Kontexts bei Textbeispielen: Beim Zitieren von Textpassagen aus historischen Quellen und aus der Mundartliteratur soll der Kontext so kurz wie möglich gehalten werden.'*²⁷

Der damalige Redaktor des WBÖ, W. Bauer, bedauert im Vorwort zum WBÖ 5,II diese *'für den historisch orientierten und interessierten Lexikographen schmerzhaft Änderung.'* Die Darbietung des

Beleg-Ausschnitts, die für das Verständnis der Textstelle entscheidende Länge sowie erläuternde Erklärungen, obliegen der Lexikographin / dem Lexikographen.

Die Auswahl jenes Zitats, das in der Zusammenschau aller gefundenen Textstellen die geringste Kontextumgebung benötigt, um aussagekräftig zu sein, ermöglicht, die gewünschte Straffung im Darstellungsstil bei Textziten einzuhalten.

Durch das TK wird die Materialbasis zum WBÖ ausgeweitet und durch die Auszeichnung mittels Sigle dennoch eine rasche und effiziente Sortierung gemäß den Sortierrichtlinien des WBÖ²⁸ sichergestellt. Die schnelle Zusammenschau über eine Vielzahl von Einzelbelegen gewährleistet darüber hinaus die Einhaltung der hohen Qualitätsansprüche beim WBÖ.

Da Abschreibfehler ebenso wie Datierungs- und Lokalisierungsfehler künftig einerseits durch die programmgestützte Übernahme der Textstelle in den jeweiligen Artikel sowie andererseits durch einheitliche Siglenkennzeichnung minimiert werden können, führt der Einsatz des TK diesbezüglich zu einer Qualitätssteigerung des WBÖ.

6.4 Lexikographie versus (?) Informationstechnologie

Das TK zum WBÖ erweist sich, wie die Zusammenschau gezeigt hat, als leistungsstarkes Hilfsmittel für die Lexikographin / den Lexikographen beim Schreiben der Wörterbuchartikel. Darüber hinaus kann das TK eine Quelle für weitere (wissenschaftliche) Arbeiten darstellen.

Die Leistungen, die das TK heute für das WBÖ-Team erbringt und die hier kurz diskutiert worden sind, sollen nicht darüber hinwegtäuschen, dass mit dem Aufbau eines derartigen TK aufwändige und interdisziplinäre Arbeiten verbunden sind. Dabei sind nicht nur konzeptionelle Überlegungen anzusprechen, die in der Hand der Lexikographin / des Lexikographen liegen, die/der mit dem Text vertraut und mit der Artikelerstellung befasst ist. Die intensive Nutzung digitaler Texte erfordert neben problembezogenen, lexikographischen Sachkenntnissen entweder persönliche Programmierfähigkeiten (wie dies am I DINAMLEX der Fall war und ist) und/oder gute Kenntnisse der aktuellen Medientechnologien, um in Zusammenarbeit mit

Informationstechnolog(inn)en die Möglichkeiten auszuschöpfen, welche die neuen Medien für die Lexikographie bereitstellen.

Jenseits des Tellerrandes des eigenen lexikographischen Projektes stellt ein derartiges TK eine Quelle kulturhistorischer Forschung dar, die die Kulturtradition eines bestimmten Gebiets dokumentiert. Moser (1990, 234) spricht sich in seiner Rezension des WBÖ positiv über die Wortbelege aus, *‘die hier vom Mittelalter bis zur Zeitungssprache der Gegenwart eingearbeitet erscheinen [und] ... dem Kulturhistoriker und nicht zuletzt dem Volkskundler zugleich wichtige Anhaltspunkte an die Hand geben’*.

Mittels digitaler Vernetzung könnte dieser Kulturraum weiter definiert werden und nicht an den administrativen Grenzen des jeweiligen Wörterbuch-Bearbeitungsgebietes enden. Als ein denkbare Projekt für die Lexikographie zukünftiger Jahre spricht sich Speer (2002, 107) beispielsweise für ein *‘Wörterbuch der ländlichen Rechtsquellen des bairisch-österreichischen Sprachraumes’* aus, für welches die historischen Rechtstexte des TK zum WBÖ einen wesentlichen Beitrag darstellen könnten.

DEUTSCHE Zusammenfassung: Das Textkorpus zum WÖRTERBUCH DER BAIRISCHEN MUNDARTEN IN ÖSTERREICH (WBÖ) und seine Bedeutung für die praktische Artikellarbeit

Das WBÖ basiert auf einer umfangreichen Sammlung (ca. 4 Mio. Unikate), die seit 1913 zusammengetragen worden ist. Es versteht sich als sprachwissenschaftliches, aber auch als historisches und volkskundliches Informationsmedium. Deshalb wurden neben umfangreichen Sammlungen alle wesentlichen in Betracht kommenden regionalen Wörterbücher und Dissertationen exzerpiert, darüber hinaus jedoch auch entsprechende literarische Quellen, z.B. die Österreichischen Weistümer. Schätzungsweise 10% des Gesamtmaterials sind auf Textzitate unterschiedlichster Art zurückzuführen. Die zum Teil ohne Seitenangabe bzw. mit

Seitenangabe einer nicht mehr zugänglichen Ausgabe oder nur über Glossare exzerpierten Einzelbelege müssen von der Lexikographin / vom Lexikographen in mühevoller Kleinarbeit auf das Vollzitat rückerschlossen, ausgewählt, datiert und lokalisiert werden. Diese Arbeiten sind zeitaufwändig und fehleranfällig.

1995 wurde daher am Institut für Österreichische Dialekt- und Namenlexika (I DINAMLEX) der Österreichischen Akademie der Wissenschaften mit der Errichtung eines Textkorpus (TK) zum WBÖ begonnen. Dieses versteht sich als ergänzende Belegdatenbank zur "(Hauptkatalog)Datenbank der bairischen Mundarten in Österreich (DBÖ)", einer seit 1993 im Aufbau befindlichen Belegdatenbank. Ziel des TK ist es, die wesentlichsten im WBÖ zitierten Quellen (Primärtextkorpus [PrimärTK]) sowie wichtige Sekundärliteratur (Sekundärtextkorpus, Testphase) für die digitale Abfrage zur Verfügung zu stellen. Ebenso ist das für die Artikellarbeit zur Verfügung gestellte (retro)digitalisierte WBÖ dem TK zuzurechnen.

Das PrimärTK, das sich nach Maßgabe der finanziellen Mittel und je nach Verfügbarkeit der WBÖ-Mitarbeiter im Aufbau befindet, umfasst derzeit 87 digitalisierte Texte; 64 sind für die Artikellarbeit aufbereitet, 23 befinden sich in Bearbeitung.

Das Einlesen der Texte und deren Aufbereitung erfolgt institutsintern in drei Arbeitsschritten:

1. Die Texte werden eingescannt (proLector V1.20 [A.1]). Da das Programm über keine vordefinierte Wissensbasis verfügt, kann jeder Zeichensatz sehr genau trainiert werden (für jedes Dokument wird mindestens eine eigene Schriftbasis angelegt; Sonderzeichen können mit bis zu vier Einzelzeichen codiert werden), wodurch die Fehlerquote niedrig ist. Abschließend werden die digitalen Files konvertiert (TUSTEP).
2. In einem weiteren Arbeitsschritt werden sie von studentischen Hilfskräften bearbeitet (Überprüfung auf Fehler, Einfügen von Seitennummern und Zitation, Lokalisierung und Datierung etc.). Eine alphanumerische Sigle wird eingefügt, mittels derer die räumliche und zeitliche Einordnung des Textes erfolgt.

Der Aufbau der Sigle:

- a) {jJahreszahl:Gebietskey} Gebietsangabe @ Kurzzitat [Systematik] Der Beleg ist nicht direkt für die Lautungsbearbeitung relevant.
- b) {Gebietskey*jJahreszahl} Gebietsangabe @ Kurzzitat [Systematik] Der Beleg ist (bedingt) für die Lautungsbearbeitung von Bedeutung.
- c) {Gebietskey} Gebietsangabe @ Kurzzitat [Systematik] Der Beleg ist definitiv für die Lautungsbearbeitung zu verwenden.

Gebietskey: alphanumerische Kennzahl zur Gebietskennzeichnung und TUSTEP-internen Belegsartierung gemäß WBÖ-Zitierrichtlinien

Systematik: alphabetische Kennung zur Belegqualifizierung einer Quellenangabe.

3. Abschließend werden die Textfiles mittels einer 1998 entwickelten TUSTEP-Programmsequenz für die programmgestützte Recherche durch die Lexikographin / den Lexikographen aufbereitet.

Der Einbau in die Bedeutungsparaphrase obliegt der Lexikographin / dem Lexikographen.

Durch den Einsatz des TK für die praktische Artikellarbeit am WBÖ ergeben sich mehrere Vorteile, die sowohl eine Qualitätssicherung (in bestimmten Bereichen sogar eine Qualitätssteigerung) gewährleisten sollen, als auch trotz erweiterter Materialbasis die Einhaltung des Neuen Straffungskonzepts von 1998 unterstützen. Häufig aus dem Blickwinkel der Betrachtung fällt in diesem Zusammenhang jedoch die intensive Auseinandersetzung des Lexikographen mit informationstechnologischen Belangen, von der Konzeption eines Projektes bis zu seiner praktischen Durchführung, die arbeitsintensiv sind und interdisziplinäres Agieren voraussetzen, jedoch auch die Archivierung und Wiederverwertbarkeit des Grundlagenmaterials sicherstellen.

7 Referenzen

Arbeitsplan und Geschäftsordnung für das bayerisch-österreichische Wörterbuch (1912). Maschinschriftl. Manuskript, Wien. (1912) – online: Institut

- für Österreichische Dialekt- und Namenlexika (09.10.2003). Arbeitsplan 1912. <<http://www.oeaw.ac.at/dinamlex/WBOE.html>>. (28.01.2008).
- Bauer, W.** (1998). *Historische Quellen des WBÖ zu den bairisch-österreichischen Sprachvarietäten des 14.-19. Jahrhunderts*. In **R. Bergmann** (Hg.), *Probleme der Textauswahl für einen elektronischen Thesaurus, Beiträge zum ersten Göttinger Arbeitsgespräch zur historischen deutschen Wortforschung, 1. und 2. November 1996*, Stuttgart, Leipzig, pp. 81–103.
- Bauer, W. and Kühn, E.** (1998). *Vom Zettelkatalog zur Datenbank. Neue Wege der Datenverwaltung und Datenbearbeitung im "Wörterbuch der bairischen Mundarten in Österreich"*. In **C. J. Hutterer and G. Pauritsch** (Hgg.), *Beiträge zur Dialektologie des ostoberdeutschen Raumes*, Göttingen, pp. 369–382. (= **U. Müller, F. Hundsnurscher and C. Sommer** [Hgg.], *Göppinger Arbeiten zur Germanistik*: 636).
- Bergmann, H.** (01.04.2002). *Das Wörterbuch der bairischen Mundarten in Österreich – ein Jahrhundertprojekt*. <<http://science.orf.at/science/news/48167>>. In: Science ORF.at. <<http://science.orf.at/science/main?tmp=9076>>. (28.01.2008).
- Bergmann, H.** (10.11.2003). Streiflichter aus der Geschichte des Instituts für Österreichische Dialekt- und Namenlexika. PowerPoint-Präsentation, dargeboten beim Symposium "Deutsche Wortforschung und Kulturgeschichte" (24.-27. September 2003, Wien). <<http://www.oeaw.ac.at/dinamlex/Symposium2003.html>>. (28.01.2008).
- NöWb.** (1847) = **Castelli, I. F.** (1847). *Wörterbuch der Mundart in Oesterreich unter der Enns*, Wien.
- Datenbank der bairischen Mundarten in Österreich (DBÖ) zum Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*, Institut für Österreichische Dialekt- und Namenlexika der Österreichischen Akademie der Wissenschaften. Wien. Bearbeitungsstand 07. 2006. – Beispiele online: Institut für Österreichische Dialekt- und Namenlexika (23.07.2003). Trage. <<http://www.oeaw.ac.at/dinamlex/dbo/d168%20abb.html>>. (28.01.2008). Fahn(en)trager. <<http://www.oeaw.ac.at/dinamlex/fahntrager.htm>>. (28.01.2008). Hochzeit. <<http://www.oeaw.ac.at/dinamlex/hochzeit.htm>>. (28.01.2008).
- Geyer, I.** (05.02.2000). *Die digitale Dialektdatenbank Österreichs (DBÖ) und das Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*. In: Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing 16/3. Oxford, pp. 319–330.
- Online: In: *Protokoll des 78. Kolloquiums über die Anwendung der Elektronischen Datenverarbeitung in den Geisteswissenschaften an der Universität Tübingen*. Hg. v. Zentrum für linguistische Datenverarbeitung der Universität Tübingen. (05.02. 2000). <<http://www.zdv.uni-tuebingen.de/static/skripte/tustep/prot/prot781-diadb.html>>. (06.07.2006).
- Hauptkatalog (HK) zum Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*, Institut für Österreichische Dialekt- und Namenlexika der Österreichischen Akademie der Wissenschaften. Wien. Bearbeitungsstand 12. 2005.
- Jakobs, E.-M.** (1999). *Textvernetzung in den Wissenschaften: Zitat und Verweis als Ergebnis rezeptiven, reproduktiven und produktiven Handelns*. Tübingen. (= Reihe Germanistische Linguistik 210).
- Kühn, E.** (1998). *Intensive Nutzung der EDV im Bereich Verwaltung und Bearbeitung des Sprachdatenmaterials beim WBÖ*. In **R. Grosse** (Hg.), *Bedeutungserfassung und Bedeutungsbeschreibung in historischen und dialektologischen Wörterbüchern*, Stuttgart, Leipzig, pp. 221–232. (= *Abhandlungen der sächsischen Akademie der Wissenschaften zu Leipzig, Philosophisch-historische Klasse*: 71/1).
- StelzhId./M. = Matosch, A.** (1884). *Idiotikon zu Franz Stelzhamer's Dichtungen in obderennsscher Mundart*. In **P. Rosegger** (Hg.), *Franz Stelzhamer's Ausgewählte Dichtungen* 4, Wien, Pest, Leipzig.
- Moser, O.** (1990). *Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*. Rezension der 27. Lieferung. In *Österreichische Zeitschrift für Volkskunde* XLIV, pp. 234–235.
- Neues Straffungskonzept für das Wörterbuch der bairischen Mundarten in Österreich (WBÖ)* (1998). Maschinschriftlich. Wien. – online: Institut für Österreichische Dialekt- und Namenlexika (09.10.2003). Straffungskonzept 1998. <<http://www.oeaw.ac.at/dinamlex/WBOE.html>>. (28.01.2008).
- Reiffenstein, I.** (1997-1999). *Das neue Straffungskonzept für das Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*, In: *Österreichische Akademie der Wissenschaften* (Hg.), *Anzeiger der Philosophisch-historischen Klasse der Österreichischen Akademie der Wissenschaften*: pp. 134/1: 113–126.

- SaWeistGl.** = **Tomaschek, K.** (1870). *Glossar zu den salzburgischen Taidingen*. In **Kaiserliche/Österreichische Akademie der Wissenschaften** (Hg.), *Österreichische Weist(h)ümer* 1: 349f1432.
- TirWb.**(1955–1956) = **Schatz, J.** (1955f.). *Wörterbuch der Tiroler Mundarten*. Innsbruck 1955/1956; unveränderter Nachdruck 1993. (= *Schlern-Schriften*, 119, 120).
- Speer, H.** (2002). *Rechtssprachlexikographie und neue Medien*. In **V. Ägel, A. Gardt, U. Haß-Zumkehr and Th. Roelcke** (Hgg.), *Das Wort. Seine strukturelle und kulturelle Dimension*, Tübingen, pp. 89f110.
- Straffungskonzept für das Wörterbuch der bairischen Mundarten in Österreich (WBÖ)* (1993). Maschinschriftliche Unterlage zu TOP 3 der Einrichtungssitzung am 24. November 1993. Wien. – online: Institut für Österreichische Dialekt- und Namenlexika (09.10.2003). *Straffungskonzept* 1993. <<http://www.oeaw.ac.at/dinamlex/WBOE.html>>. (28.01.2008).
- Tübinger System von Textverarbeitungs-Programmen (TUSTEP)*.<<http://www.uni-tuebingen.de/zdv/tustep/index.html>>. (09.02. 2004).
- Wandl-Vogt, E.** (2002). *Digitale Volltexte als Arbeitsbehelf für die Dialektlexikographie am Beispiel des Textkorpus zum "Wörterbuch der bairischen Mundarten in Österreich (WBÖ)"*. In **Th. Burch, J. Fournier, K. Gärtner, A. Rapp** (Hgg.), *Standards und Methoden der Volltextdigitalisierung*, Trier, pp. 177f185; 338f. (= *Abhandlungen der Akademie der Wissenschaften und der Literatur, Mainz; Geistes- und sozialwissenschaftliche Klasse*; Einzelveröffentlichung 9).
- Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*. (1963lfd.), **E. Kranzmayer, Kommission für Mundartkunde und Namenforschung** (ab Bd. 3), **Institut für Österreichische Dialekt- und Namenlexika** (ab Bd. 4lfd.) (Hgg.), Wien. (= *Bayerisch-Österreichisches Wörterbuch I. Österreich*).
- Wörterbuch der bairischen Mundarten in Österreich (WBÖ)*, Beiheft Nr. 2, *Erläuterungen zum Wörterbuch, Abkürzungsverzeichnis, Lauttabelle, Artikelstruktur, Literaturverzeichnis, Gemeindenamenregister, Verzeichnis der Gebietsnamen (mit einer Übersichtskarte und 6 Detailkarten)* (2005), **Institut für Österreichische Dialekt- und Namenlexika** (Hg.), Wien. (= *Bayerisch-Österreichisches Wörterbuch I. Österreich*).
- Weitere Literaturangaben zu Wörterbuch und Datenbank: Institut für Österreichische Dialekt- und Namenlexika (23.07.2003). Publikationen. <<http://www.oeaw.ac.at/dinamlex/Publikationen.html>>. (28.01.2008).

Notes

- 1 Zum Einsatz neuer Technologien bei einem Jahrhundertprojekt vgl. H. Speer (2002) 101f.
- 2 Näh. s. H. Bergmann (2003).
- 3 **Schmeller, J. A.** (1872-1877). *Bayerisches Wörterbuch*. 2 Bände. München (2.Auflage).
- 4 Vgl. dazu die Übersichtskarte im WBÖ Beiheft Nr. 2; zu den Gebietseinschränkungen durch das Straffungskonzept vgl. Straffungskonzept (1998) § 1.1; s.a. I. Reiffenstein (1997-1999) 114 u. WBÖ Beiheft Nr. 2 11-13.
- 5 Das Werk erscheint heute in zwei getrennten Publikationsreihen, als WBÖ und als *Bayerisches Wörterbuch*. (1995 lfd.), **Kommission für Mundartforschung der Bayerischen Akademie der Wissenschaften** (Hg.), München (= *Bayerisch-österreichisches Wörterbuch II. Bayern*).
- 6 W. Bauer/E. Kühn (1998) 370.
- 7 Auf die Artenvielfalt einer Quellengattung kann hier aus Platzgründen nicht näher eingegangen werden.
- 8 Vgl. WBÖ Beiheft Nr. 2 35ff.
- 9 Arbeitsplan (1912) 1.
- 10 Näh. zu Entscheidungsfindung und Konzeption der DBÖ s. E. Kühn (1998).
- 11 Bearbeitungsstand 07.2006.
- 12 Verbreitungsangaben werden im WBÖ seit dem 5.Band nur noch in Form von Gebietsangaben realisiert, die auf Grund historisch-dialektaler sowie unter Berücksichtigung administrativer Kriterien gegeneinander abgegrenzt worden sind; vgl. Straffungskonzept (1993) § 2.2.2.
- 13 Unter <http://www.wboe.at> kann ein diebezügliches Projekt eingesehen werden, das vom FWF unterstützt wird und von einer internationalen, transdisziplinären Forscher(innen)gruppe in den Jahren 2007–2008 ausgeführt wird.
- 14 Rund 600 Orte und 420 Mundartgebiete liegen in vom WBÖ ab Band 5 nicht mehr berücksichtigten Regionen. Näheres s. Straffungskonzept (1993) § 1.1, Straffungskonzept (1998) § 1.1; s.a. I. Reiffenstein (1997-1999) 114 u. WBÖ Beiheft Nr. 2 11.
- 15 Vgl. H. Speer (2003).
- 16 <http://verlag.oeaw.ac.at>.
- 17 Im Folgenden werden nur Texte im TUSTEP-Format berücksichtigt. Einige wenige Texte liegen als.doc- oder.txt- Files vor. Es ist daran gedacht, diese Texte zu konvertieren und in das TK zum WBÖ zu integrieren.
- 18 Artikelmakros sind im I DINAMLEX erstellte TUSTEP-Makros, die für die konkrete Artikelbearbeitung mittels der DBÖ eingesetzt werden können.
- 19 Jede Schrift, die auf klar separierten Einzelbuchstaben basiert, kann gelesen werden, was insofern

- auswahlentscheidend wirkte, als viele Quellentexte in Frakturschrift gesetzt sind.
- 20 Generell folgt die Verbreitungsangabe im WBÖ strengen Grundsätzen, wobei im Wesentlichen vom frühesten zum jüngsten Beleg, von Westen nach Osten, vom Südbairischen zum Mittelbairischen zitiert wird.
 - 21 Vgl. Neues Straffungskonzept (1998) § 1.1; s.a. I. Reiffenstein (1997-1999) 114 u. WBÖ Beiheft Nr. 2 11.
 - 22 Die Artikelmakros laufen auf der TUSTEP-Kommandoebene und nicht über eine eigens definierte Maske mit Fenstern. Makros regeln ebenso den Dialog mit den Benutzer(inne)n.
 - 23 Leerzeichen-Großbuchstabe *T* Kleinbuchstabenkette *rager* - Leerzeichen
 - 24 Alle Buchstabenkombinationen, in denen die Groß- und/oder Kleinbuchstabenkette *trager* vorkommt, ohne Angabe von Leerzeichen vor dem Wort oder daran anschließend, werden erfasst.
 - 25 Neues Straffungskonzept (1998) ebd.; s.a. I. Reiffenstein (1997-1999) 114; WBÖ Beiheft Nr.2 11.
 - 26 Neues Straffungskonzept (1998) ebd.; s.a. I. Reiffenstein (1997-1999) 115 u. WBÖ Beiheft Nr. 2 11.
 - 27 Neues Straffungskonzept (1998) ebd.; s.a. I. Reiffenstein (1997-1999) 116 u. WBÖ Beiheft Nr. 2 13.
 - 28 Die Belegsartierung und Zitierung im WBÖ erfolgt im Wesentlichen von Westen nach Osten, vom Südbairischen zum Mittelbairischen und vom frühesten zum jüngsten Beleg.
 - 29 D.h. für das Artikelmakro *SuchLit* [= *Suche in Literatur*] aufbereitet, das eine Recherche via Mausclick ermöglicht.
 - 30 D.h. digitalisiert zur Verfügung, aber nicht für das Artikelmakro *SuchLit* aufbereitet.
 - 31 Ausgenommen Mundartliteratur.
 - 32 Inkl. WBÖ.
 - 33 Im Gegensatz zum italienischen Südtirol (11.) und dem von Nordtirol räumlich getrennten Osttirol (12.)