

Zeitschrift: Bericht über das Jahr / Schweizerdeutsches Wörterbuch : Schweizerisches Idiotikon
Herausgeber: Schweizerisches Idiotikon
Band: - (2006)

Artikel: Idiotikon digital : Überlegungen zu einer elektronischen Ausgabe des Schweizerdeutschen Wörterbuchs
Autor: Bickel, Hans
DOI: <https://doi.org/10.5169/seals-1091504>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 26.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Idiotikon digital: Überlegungen zu einer elektronischen Ausgabe des Schweizerdeutschen Wörterbuchs

von Hans Bickel

Meine Damen und Herren, liebe Kolleginnen und Kollegen, ich darf Ihnen in meinem Referat in einem stark verkürzten Überblick zeigen, wie eine Digitalisierung des Idiotikons aussehen könnte, welchen Nutzen die Benutzer daraus ziehen könnten und mit welchem Aufwand eine solche Digitalisierung verbunden ist. Wie Sie dem Untertitel des Referats entnehmen können, handelt es sich vorläufig nur um Überlegungen zur Digitalisierung und nicht um ein bereits konkretes Projekt. Der hier vertretene Standpunkt entspricht ausschliesslich meiner persönlichen Sicht und ist nicht mit der Redaktion abgesprochen.

1. Grundsätzliche Vorbemerkungen und Problemstellung

Was ist der erste Eindruck, den man als frischgebackener Redaktor am Idiotikon und vom Idiotikon hat? Mein erster Eindruck war: Es gibt hier unglaublich grosse Wissensbestände, eine unheimliche Substanz, sowohl auf der Redaktion, in der Bibliothek und natürlich auch besonders in den publizierten Artikeln des Idiotikons. Der zweite Eindruck ist allerdings: Das Idiotikon macht es einem nicht immer ganz leicht, an diese Wissensbestände, an diese Substanz heranzukommen. Manchmal hat man sogar den Eindruck, diese unheimliche Fülle an Wissen verstecke sich fast ein wenig in den langen, dichten, manchmal auch etwas unübersichtlichen Artikeln.

Wenn man dagegen Projekte aus dem Informatikbereich, z.B. Wörterbuchprojekte von Computerlinguisten anschaut, hat man häufig genau den gegenteiligen Eindruck. Man hat dort zwar wunderbare Möglichkeiten, an Informationen heranzukommen, schöne «Tools», wie man in der Computerwelt sagt, dafür aber häufig wenig substantielle Informationen, die überhaupt gezeigt werden können.

So gesehen, müsste die Erschliessung der substantiellen Wissensbestände des Idiotikons mit den Mitteln der Informatik zu einer idealen Verbindung führen. Was bringt eine Digitalisierung?

- Die Digitalisierung erlaubt einen Zugriff, der schnell, ortsunabhängig, zeitunabhängig und kostengünstig ist;
- nach der Digitalisierung können die Informationen auf ganz unterschiedliche Weise präsentiert werden;
- man hat dank der Digitalisierung ganz andere Suchmöglichkeiten, d.h. man kann nach Dingen suchen, die bisher mit vernünftigem Aufwand nicht auffindbar waren, z.B. nach einzelnen Wörtern oder Wortformen im Innern eines Artikels oder nach zusätzlichen Belegen für ein Wort. Damit wird das Wörterbuch auch der Wörterbuchforschung besser zugänglich;
- die Digitalisierung bringt ganz generell eine grössere Breitenwirkung, eine stärkere Rezeption und eine grössere Bekanntheit des Werks (die Internetversion des Grimmschen Wörterbuchs wurde zwischen September 2004 und Mai 2006, also während 600 Tagen, mehr als 2 Mio. Mal besucht oder, auf Tage umgerechnet, pro Tag ungefähr 3350 Mal. Auch wenn ein guter Teil dieser Besuche eine zufällige Komponente haben dürfte, ist das doch eine eindruckliche Zahl).

Allerdings ist auch mir klar, dass man von einer Digitalisierung keine Wunder erwarten darf, die Digitalisierung kann die redaktionelle Arbeit nicht ersetzen und höchstens geringfügig beschleunigen. Eine Digitalisierung ist also in erster Linie benutzerorientiert.

Bevor wir uns mit der Digitalisierung beschäftigen, werfen wir einen Blick auf die gedruckte Ausgabe: Erschienen sind bisher, wie Sie alle wissen, 15½ Bände, das Gesamtwerk wird nach derzeitiger Planung 17 Bände umfassen. Dies entspricht ca. 32400 Spalten oder 17 Mio. Wörtern oder 94 Mio. Buchstaben.

Für ein gedrucktes Werk sind das eindruckliche Zahlen. So beeindruckend der Umfang des gedruckten Werkes auch ist, er stellt für die Digitalisierung kein unlösbares Problem dar. Wenn der gesamte Text inkl. aller Formatangaben in den Computer eingegeben wird, ist mit einem Umfang von etwa 300 Megabytes zu rechnen. D.h., auf einer herkömmlichen Compact Disc, könnte man gar ein Werk von doppeltem Umfang unterbringen. Etwas anders sieht es aus, wenn alle Seiten des Idiotikons als digitale Bilder erfasst und gespeichert werden. Dann braucht man auch für heutige Begriffe ziemlich leistungsstarke Verarbeitungs- und Speichersysteme.

2. Zugriffsstrukturen

Wörterbücher werden bekanntlich nicht linear von A bis Z gelesen, sondern gewöhnlich suchen die Benutzer eine bestimmte Information. Um die Informationen, die in einem Wörterbuch stecken, zu finden, hat jedes Wörterbuch eine oder mehrere so genannte *Zugriffsstrukturen*. Unter *Zugriffsstruktur* versteht man in der Lexikografie einerseits das System, nach dem die Artikel in einem Wörterbuch geordnet sind, und andererseits weitere Hilfsmittel, die helfen, eine bestimmte Information zu finden.

Das Idiotikon verfügt im Wesentlichen über zwei Zugriffsstrukturen, nämlich über die Anordnung der Stichwörter nach der Haupt-oder Stammsilbe und über das bzw. die Register in glattalphabetischer Ordnung. Diese beiden Zugriffsstrukturen ergänzen sich in idealer Weise. Wenn man das System der Lemmaanordnung einmal begriffen hat, findet man wesentlich mehr Informationen als in einem «normalen», alphabetisch geordneten Wörterbuch. Denn man findet zu einem Grundwort immer auch gleich alle Ableitungen und Zusammensetzungen mit dem entsprechenden Stichwort als Grundwort.

Wenn ich also etwa das Stichwort *Bröt* nachschlage, erhalte ich nicht nur den Artikel *Bröt*, sondern erfahre auch, welche Ableitungen und Zusammensetzungen mit Brot als Grundwort gebildet werden, es sind über 260, von *Abendbröt* über *Ölbröt* bis zu *Züribröt*.

Was wir im Idiotikon anders als in einem normalalphabetischen Wörterbuch auf Anhieb nicht finden, sind die Zusammensetzungen mit Brot als Bestimmungswort. Dazu braucht es das alphabetische Register. Zusammensetzungen mit *Bröt* als Bestimmungswort sind auch sehr zahlreich, es gibt weit über 100 Stück von *Brötbank* bis *Brötwib*, und sie sind natürlich über alle Bände verteilt. Dank des Registers findet man sie an sich problemlos, einziger Nachteil ist, dass man, da es bis nach Abschluss des letzten Bandes kein Gesamtregister geben kann, an fünf verschiedenen Orten nachschlagen muss, um die Übersicht über alle Zusammensetzungen zu gewinnen.

Was man aber nicht findet, und das ist bei anderen gedruckten Wörterbüchern auch so, sind Wortbestandteile, also etwa Ableitungssilben, die nicht am Wortanfang stehen. Ebenfalls nicht gezielt auffindbar sind weitere Informationen innerhalb der Artikel, beispielsweise zusätzliche Belegstellen zu anderen Lemmata oder alle Belegstellen eines bestimmten Autors.

Brot als Grundwort

Äbeⁿdröt, Acherbröt, Agatheⁿbröt, Allerheiligeⁿbröt,
 Almoseⁿbröt, Alpbröt, Ammannbröt, Änggelibröt,
 Angsterbröt, Ankeⁿbröt, Äreⁿbröt, Armeⁿbröt,
 Armeⁿsünderbröt, Eggeⁿbröt, Eichleⁿbröt, Eierbröt,
 Engelsbröt, Erdepfelbröt, ... Chälblerbröt, Candisbröt,
 Chaufröt, Kempheⁿbröt, Chërneⁿbröt, Chesteneⁿbröt,
 Chilcheⁿbröt, Chilchwihbröt, Chindbettibröt, Kindeⁿbröt,
 Chirsibröt, Chleiⁿbröt, Klïenbröt, Chnebelbröt, ... Ölbröt,
 Öreⁿbröt, Österbröt, Bärbröt, Bäreⁿbröt, Beckeⁿbröt,
 Perchisbröt, Berlinerbröt, Bëttelbröt, Bëtteⁿbröt, Bëttibröt,
 Bëttlerbröt, Pfaffeⁿbröt, Pfannbröt, Pfileⁿbröt, Pfingstbröt,
 Pfisterbröt, Pfreundbröt, Bierbröt, Bireⁿbröt, Bodeⁿbröt,
 Bollbröt, Bölleⁿbröt, Böneⁿbröt, Botteⁿbröt, Bröd,
 Brückeⁿbröt, Büreⁿbröt, Pürlibröt, Bütelbröt, Butterbröt,
 Randbröt, ... Zettelbröt, Zimmetbröt, Zinsbröt, Zuckerbröt,
 Züri^{ch}bröt

Brot als Bestimmungswort

Brötbank, Brötbeck, Brötbeileⁿ, Brötcharreⁿ, Brötchasteⁿ,
 Brötchorb, Brötchranz, Brötgott, Brötgrämpler, Bröthab,
 Bröthangeⁿ, Brötholi, Bröthurd, Bröthüs, Brötmüllli,
 Brötmurpf, Brötmurreⁿ, Brötmütscheⁿ, Brötnapf, Brötsämeⁿ,
 Brötschäl, Brötschänder, Brötschatz, Brötschätzer,
 Brötschërben, Brötschlag, Brötschnitten, Brötschragen,
 Brötschüflen, Brötschüssel, Brötschüßleⁿ,
 Brötschweißbadstubeⁿ, Brötschweizi, Brötsparer, Brötspënd,
 Brötstand, Brötstellen, Brötstübli, Brötsuppeⁿ, Brötvater,
 Brötdarreⁿ, Brötdili, Brötg^ewicht, Brötsunntag, Bröttäfer,
 Bröttäscheⁿ, Bröttäßeⁿ, Bröttax, Brötteig, Bröttisch,
 Brötttrageⁿ, Brötttrageriⁿ, Bröttrog, Brötttruckeⁿ, Bröttschifereⁿ,
 Bröttschilli, Bröttuech, Bröttuggler, Bröttünkli, Bröttürli,
 Brötturteⁿ, Bröttütscher, Brötwageⁿ, Brötwëgeⁿ, Brötwëger,
 Brötweggeⁿ, Brötwīb

Soviel zum Idiotikon als gedrucktem Werk. Ich komme nun zum Kern meines Referats, zu den Überlegungen, wie die gedruckten Teile in eine digitale Fassung überführt werden könnten.

3. Möglichkeiten zu einer schrittweisen Digitalisierung

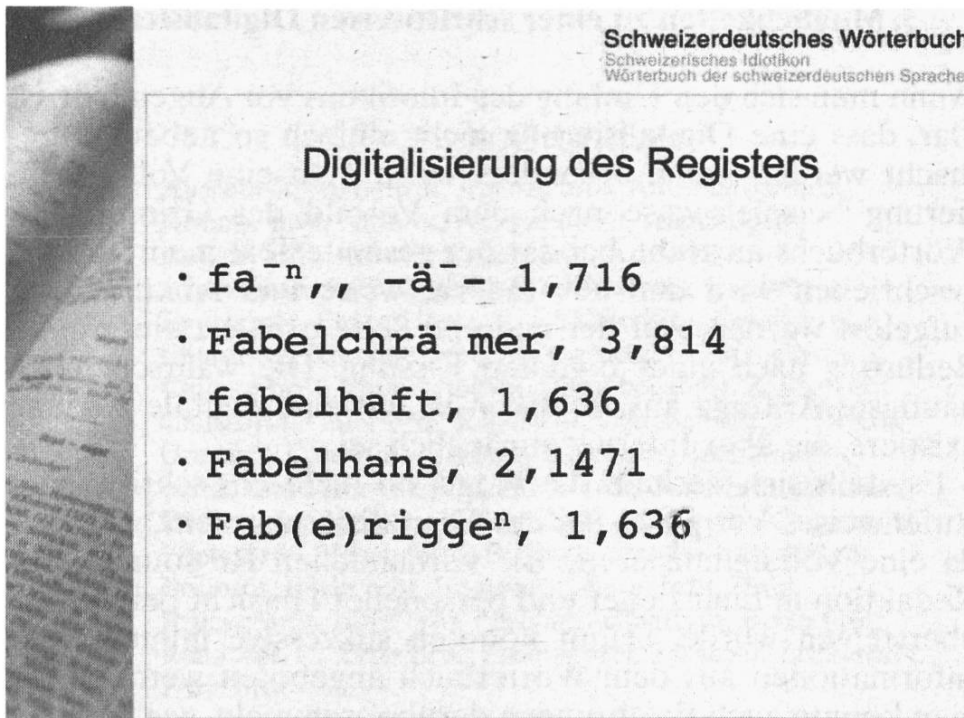
Wenn man sich den Umfang des Idiotikons vor Augen hält, ist klar, dass eine Digitalisierung nicht einfach so nebenher gemacht werden kann, besonders wenn man eine Volldigitalisierung beispielsweise nach dem Vorbild des Grimmschen Wörterbuchs anstrebt, bei der der gesamte Text mehrfach abgeschrieben wird und alle Querverweise und Abkürzungen aufgelöst werden. Auf der anderen Seite existiert ein grosses Bedürfnis nach einer digitalen Fassung. Die wahrscheinlich häufigste Anfrage ans Idiotikon ist, ob eine digitale Fassung existiere, die über Internet zugänglich sei.

Es stellt sich deshalb die Frage, ob nicht ein schritt- oder stufenweises Vorgehen bei der Digitalisierung denkbar wäre, da eine Volldigitalisierung die vorhandenen Ressourcen der Redaktion in finanzieller und personeller Hinsicht bei weitem übersteigen würde. Damit könnten sukzessive immer mehr Informationen aus dem Wörterbuch angeboten werden, und man könnte auch Erfahrungen darüber sammeln, wie die digitale Information genutzt wird. Ich könnte mir vorstellen, dass die Digitalisierung in fünf Schritte aufgeteilt wird, wobei jeder Schritt auf der Vorarbeit des anderen aufbauen kann und auch auf der nächsten Stufe noch Sinn macht. Die fünf Schritte oder Stufen sind:

1. Digitalisierung des alphabetischen Registers;
2. Verknüpfung des Registers mit digitalisierten Bildern der gedruckten Bände;
3. unkorrigierte Volltextdigitalisierung als reine Suchfunktion;
4. korrigierte Volltextdigitalisierung;
5. korrigierte und annotierte Volltextdigitalisierung mit spezialisiertem Abfragesystem und automatisierten Querverweisen.

Als ersten Schritt könnte ich mir die Digitalisierung des Registers vorstellen. Dieses liegt heute bereits in elektronischer, also digitaler Form vor, jedoch in mehreren Textdateien, die als Vorlage für den Druck gedient haben. Es ist in der vorliegenden Form nicht abfragbar, da es nicht für die Maschinenverarbeitung konzipiert wurde.

Sie sehen auf der Folie einen Ausschnitt aus dem Buchstaben F.



Man sieht, die Längen sind als separate Zeichen gesetzt, Varianten sind so dargestellt, dass man sie nicht maschinell auflösen kann. Dass der erste Eintrag die Lautung $fā^n$ und $fä^n$ haben kann, erschliesst sich einer Maschine nicht, auch *Fabrigge* und *Faberigge* ist so nicht elektronisch verarbeitbar.

Digitalisierung des Registers hiesse, die Einträge in einer Datenbank abzuspeichern, die Varianten so aufzulösen, dass sie als vollständige Einträge nebeneinander stehen, so dass sich ein Benutzer kaum um lautliche und graphematische Varianten kümmern muss und das gesuchte Wort so eintippen kann, wie er es kennt und dann möglichst den korrekten Registereintrag erhält. Was gewinnt man mit einem digitalen Register:

- Das Nachschlagen im Register geht schnell, man muss nur noch an einer Stelle nachschlagen;
- Die Einträge können aktuell sein und auch die Hefte des noch unvollständigen Bandes 16 enthalten;
- Laien gelangen schneller ans Ziel, weil alternative Schreibungen angeboten werden können;
- Eine solche Datenbank kann mit kleinem Aufwand ins Internet gestellt werden und hilft, die Benutzung des Idiotikons zu erleichtern;
- Und schliesslich kann man nach Dingen suchen, die man im gedruckten Register und im Wörterbuch so nicht findet, nämlich nach einzelnen Wortbildungselementen.

Ein Resultat der Suche nach Wortbildungselementen könnte z.B. so aussehen wie auf der unten stehenden Folie, die einen kleinen Ausschnitt aus der Abfrage nach der Endung *-igi* enthält.

Schweizerdeutsches Wörterbuch Schweizerisches Idiotikon Wörterbuch der schweizerdeutschen Sprache	
Digitalisierung des Registers	
Äßigi, 1,500	Äßigi
Eigenrichtigi, 6,471	Eigenrichtigi
Eige(n)sinnigi, 7,1070	Eigensinnigi Eigesinnigi
Einigi, 1,280	Einigi
Ei(n)richtigi, 6,471	Einrichtigi Eirichtigi
Er(n)stigi, 1,466	Ernstigi Erstigi
Ertligi, 1,479	Ertligi
Ē(w)haftigi, 1,8	Ēwhaftigi Ēhaftigi
Fälligi, 1,762	Fälligi
Vergëßligi, 2,454	Vergëßligi
Vergurti(gi), 2,445	Vergurtigi Vergurti
Verhebigi, 2,941	Verhebigi
Verirrtigi, 1,411	Verirrtigi

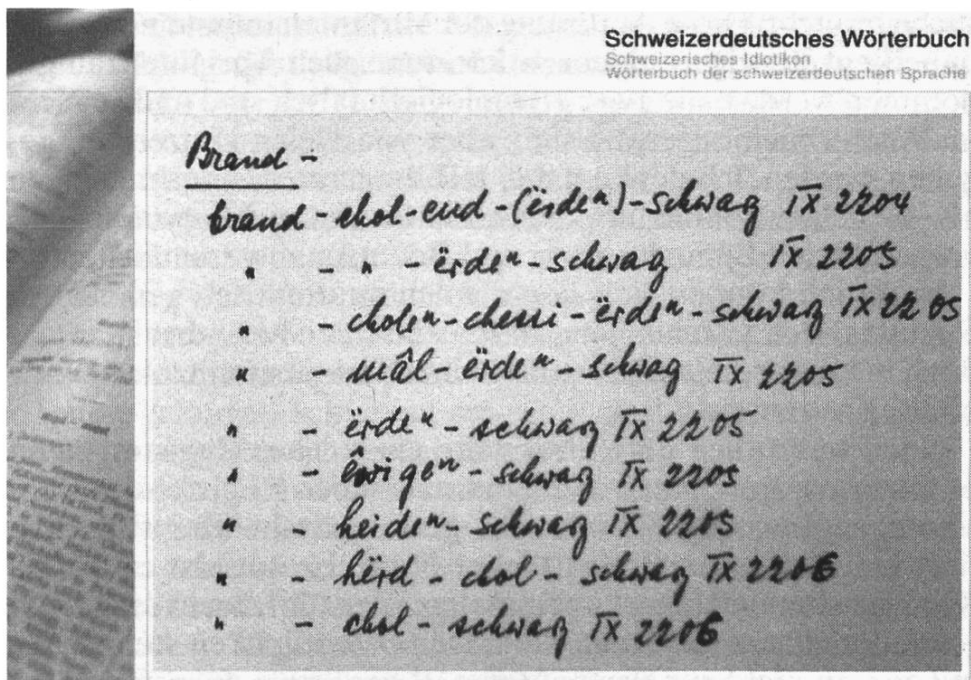
Links steht der Eintrag so, wie er für das gedruckte Register erfasst wurde, rechts stehen die Auflösungen, die man für die Absuche braucht. Diese Auflösung der Varianten müsste man nicht unbedingt anzeigen, dadurch könnten auch Varianten aufgenommen werden, die zwar etymologisch falsch sind und so nicht im Wörterbuch angesetzt sind, aber von vielen Nutzern eingegeben werden. Ich denke dabei z.B. an germ. *-ë-*, das man auch als *-ä-* geben könnte, um die Suche für die nicht etymologisch vorgebildeten Benutzerinnen und Benutzer zu vereinfachen.

Natürlich können sich in ein solch automatisch generiertes Resultat auch falsche Kandidaten einschleichen, aber man hat doch in kurzer Zeit einen guten Überblick über einzelne Wortbildungsvarianten.

Noch wesentlich nützlicher wäre ein solches Register, wenn es um die Angaben aus dem grammatischen Register des Idiotikons angereichert würde. Das grammatische Register existiert lediglich in handschriftlicher Form. Es umfasst ca. 11 000 Karteikarten und lagert in einmaliger Ausführung in mehreren Karteikastenauf der Redaktion. Der Öffentlichkeit steht es somit zurzeit nicht zur Verfügung.



Auf der Abbildung oben sehen Sie ein paar Beispielkategorien aus dem grammatischen Register. Zu jeder Kategorie existieren Karteikarten mit Hinweisen auf einzelne Artikel. Eine Karte aus der 5. Kategorie *Verstärkende 1. Glieder* sieht zum Beispiel so aus:



Hier als erstes, verstärkendes Glied das Wort *Brand-*, das in Wörtern wie *brand-chol-end-schwarz*, *brand-chol-erden-schwarz* usw. vorkommt. Das ist natürlich nur ein kleiner Ausschnitt aus dieser Kategorie und auch nur ein Ausschnitt aus den Zusammensetzungen mit *Brand-* als verstärkendem Vorderglied.

Wenn man nun diese Angaben im elektronischen Register bei den einzelnen Stichwörtern dazugeben würde, hätte man mit relativ wenig Aufwand ein grammatisches Register, das allen Interessierten ohne grosse Kosten zur Verfügung gestellt werden könnte. Ich schätze, dass die Digitalisierung des grammatischen Registers in ungefähr 200 Stunden zu schaffen wäre.

Damit sind aber die Möglichkeiten eines digitalen Registers noch nicht erschöpft. Es gibt zwei weitere Möglichkeiten, die ich hier noch ganz kurz vorstellen will, die ich aber noch nicht wirklich zu Ende gedacht und am Material ausprobiert habe, und von denen ich daher auch noch nicht weiss, wie viel Zeit sie in Anspruch nehmen würden.

Die erste der beiden weiteren Möglichkeiten ist ein Register verhochdeutscher Stichwörter oder Lemmata. Dies könnte daher sinnvoll sein, weil die verschiedenen Mundartwörterbücher ihre Stichwörter ganz unterschiedlich angesetzt haben, die einen näher an der Mundart, die anderen näher bei der nhd. Standardsprache. Wer ein Stichwort in mehreren Wörterbüchern nachschlagen will, muss sich zuerst mit der Lemmatisierung der verschiedenen Wörterbücher beschäftigen. So ist beispielsweise das Verb *speien* im Idiotikon unter *spüwen* angesetzt, im Grimmschen Wörterbuch unter *speien* und unter *speuen*, im Elsässischen Wörterbuch unter *spauen*.

Dies wird spätestens dann zum Problem, wenn Wörterbücher elektronisch vernetzt werden sollen, wie es die Uni Trier mit dem Projekt eines digitalen Verbunds von Dialektwörterbüchern plant.

Ein Register verhochdeutscher Lemmata würde also bedeuten, dass man alle Ansätze im Idiotikon mit der standarddeutschen Entsprechung oder, bei Wörtern, die es nur in der Mundart gibt, einer pseudostandarddeutschen Entsprechung ins Register aufnimmt. In vielen Fällen wäre eine solche Verhochdeutschung einfach, indem man lediglich die mundartliche Lautform ins Hochdeutsche transferiert, also beispielsweise aus *Hūs* *Haus* macht. In anderen Fällen ist es etwas schwieriger, so, wenn Wörter, die es im Standard in dieser Ableitung nicht gibt wie z.B. *Bsüffni* (schwere Betrunkenheit)

oder *Lismete*, verhochdeutsch werden sollen. Hier müsste man ein Umsetzungssystem entwickeln, das natürlich auch mit den anderen Wörterbüchern abgesprochen ist, so dass eine Verknüpfung verschiedener Wörterbücher über eine solche Lemmaliste dann auch tatsächlich möglich wird.

Eine weitere Möglichkeit, die ich nur der Vollständigkeit halber anfügen will, ist die Erstellung eines Umkehrwörterbuches oder eines hochdeutschen Bedeutungs- und Kategorienregisters. Dies würde bedeuten, alle Bedeutungsangaben in einem Register zu erfassen, um das Wörterbuch gezielt nach inhaltlichen Kriterien absuchen zu können. Man könnte sich also beispielsweise gezielt eine Liste aller im Idiotikon behandelten Personennamen geben lassen oder eine Liste aller Lemmata, die z. B. mit dem Sprechen und Artikulieren zu tun haben oder alle Bezeichnungen für Kochgeschirr usw. Damit würde das Idiotikon einer systematischen Wort- und Kulturgeschichtsforschung zugänglich, indem auch eine onomasologische Sicht auf den schweizerdeutschen Wortschatz möglich würde. Die Herstellung eines derartigen Registers wäre aber ein grösseres Projekt, das nicht ohne zusätzliche Ressourcen realisiert werden kann.

Ich komme damit zu einem nächsten Digitalisierungsschritt, nämlich zur Digitalisierung des Wörterbuchteils.

Jede Digitalisierung des Gesamtwerkes beginnt mit dem fotografischen Einlesen der Buchseiten, in der Regel auch dann, wenn das Werk, wie es beim Grimmschen Wörterbuch geschehen ist, in China abgetippt wird. Denn so muss man nur elektronische Dateien verschicken und nicht 20 Kilo Bücher, die für das Abtippen auch eher sperrig sind. Da man das gesamte Werk ohnehin scannen muss, ist es auch denkbar, dass man die Bilder der einzelnen Buchseiten in einer Datenbank speichert und sie mit dem Register verknüpft, so dass die Benutzer mit Hilfe des elektronischen Registers direkt auf den Artikel zugreifen können, dort allerdings keinen Text herauskopieren können und auch keine Recherche im Volltext der Artikel machen können, sondern lediglich das digitale Bild der Buchseite sehen. Wenn man also etwa im Register nach *Wīn* ‚Wein‘ sucht, würde man das entsprechende Bild des Artikels erhalten.

Damit wäre bereits viel gewonnen, man könnte mit jedem Computer die Idiotikonartikel lesen und Ausdrucken, die Herstellung der entsprechenden DVD oder die Bereitstellung über Internet könnte ziemlich kostengünstig geschehen. Allerdings wären bei einer solchen Bereitstellung des gesamten

Werkes über Internet einige verlegerische und urheberrechtliche Knacknüsse zu lösen.

Nächste Stufe könnte eine unkorrigierte Volltextdigitalisierung des gesamten Werkes sein. Das heisst, die zuvor eingeleseenen bzw. gescannten Bilder werden mit einem so genannten OCR-Programm gelesen und in Text umgewandelt, jedoch nicht korrekturgelesen.

Eine unkorrigierte Volltextdigitalisierung ist nur dann möglich, wenn die Benutzer dieser Version immer auch das Bild der Originalseite vor Augen haben, damit nicht fehlerhafte Informationen vermittelt werden. Auf diese Weise ist die digitalisierte Version genau so zuverlässig wie die gedruckten Bände, man hätte einzig zusätzlich die Möglichkeit, im Text des Werkes zu suchen, wobei man einige wenige falsche und einige verpasste Treffer in Kauf nehmen müsste.

Als Format für eine solche kombinierte Bild-Textfassung käme das PDF-Format in Frage. Das PDF-Format gilt heute als Standard auf allen Computer-Plattformen. Programme für das Lesen von PDF-Dateien sind gratis erhältlich und damit praktisch auf jedem Computer installiert. Es würde sich daher anbieten, eine solche Digitalisierung im PDF-Format abzuspeichern, und zwar als PDF-Datei mit Original-Bild der eingescannten Seite und darunter liegendem unkorrigiertem digitalem Text. Damit erhielte man einen absuchbaren Volltext mit einer Korrektheit von mehr als 99 %. Mit Fehlern hätte man vor allem bei der Mundarttranskription und bei diakritischen Zeichen zu rechnen.

Weit aufwändiger ist die Herstellung einer korrigierten Volltext-Fassung des Wörterbuchs. Dies würde bedeuten, entweder:

- OCR mit nachfolgender Korrektur durch geschultes Personal oder
- abtippen (vorzugsweise in China) nach der Double- oder Tripple-Keying-Methode.

Letzteres bedeutet, dass der gesamte Text zwei- oder besser dreimal abgeschrieben wird und die Fassungen anschliessend miteinander verglichen werden, so dass möglichst wenig Tippfehler unentdeckt bleiben. Beide genannten Verfahren sind ziemlich arbeitsintensiv.

Wie sieht ein zu korrigierender Text aus, der mit OCR eingeleseenen wurde? Ich zeige Ihnen dazu ein Stück Text aus dem Artikel *Wĩⁿ*, der für das OCR-Programm schwierig zu lesen ist.



Korrigierte Volltextdigitalisierung

Am gewöhnlichsten mit Bez. auf **a)** Wert-
schätzung, qualitative und geschmackliche Eigenschaf-
ten. *Di Wina, wā maⁿ hitu chöufft, sind nimma^r, was
š' friejer sind g'siⁿ. JULEN. Föifendsechzg [Öchslegrade]
brüchi 's, händ die alteⁿ Rëbpüreⁿ g'seit, das^s es en
rëchteⁿ W. gäb ZrS. Mich g'lustet eⁿmäl umheⁿ es Glas*

Am gewöhnlichsten mit Bez. auf a) Wert-
schätzung, qualitative und geschmackliche Eigenschaf-
ten. *Di Wina, wā maⁿ hitu chöufft, sind nimma^r, was
š' friejer sind g'siⁿ. JuLEN. Föifendsechzg [Öchslegrade]
brüchi 's, händ die alteⁿ Rëbpüreⁿ g'seit, das^s es en
rechteⁿ W gä b ZrS. Mich g'lustet eⁿmäl umheⁿ es Glas*

Oben ist das originale Bild, unten das mit dem Programm Fine-Reader in Text umgesetzte Resultat. Man sieht, dass das Programm besonders mit Diakritika kämpft, auch z. T. mit dem häufigen Schriftwechsel wie bei *Öchslegrade*. Dann aber auch und besonders mit den Kapitälchen (z. B. bei *Julen*) und mit hochgestellten Buchstaben. Insgesamt enthält die Stelle 10 Fehler, man kommt hier auf eine Korrektheit von nur 96 % (bei insgesamt 270 Zeichen ohne Lehrzeichen). Auf den gesamten Text ist die Fehlerquote jedoch tiefer. Das Lesen der hochgestellten Buchstaben könnte man mit Training vielleicht noch etwas verbessern, aber es bleibt in jedem Fall ein bedeutender Korrekturaufwand zu bewältigen und man müsste genau berechnen, ob elektronisches Einlesen oder Abtippen das bessere und kostengünstigere Verfahren ist.


Eine solche digitalisierte und korrigierte Volltextfassung macht den gesamten Text zuverlässig absuchbar. Allerdings kann man damit noch nicht gezielt nach einzelnen Elementen suchen wie z. B. nach bestimmten Lemmata oder nach Wörtern in Mundartbelegen, nach Literaturnachweisen usw. So etwas ist nur möglich, wenn der Text annotiert ist, d. h., wenn im Text computerlesbare Strukturelemente erfasst sind.

Für eine solche Annotierung steht heute ein fast allgemein akzeptierter Standard zur Verfügung, der garantiert, dass die

Daten von verschiedensten Systemen weiterverarbeitet werden können. Dieser Standard ist ein speziell für Philologen entwickeltes Auszeichnungssystem namens *Text Encoding Initiative (TEI)*. Für Wörterbücher gibt es innerhalb von TEI spezialisierte Markierungen, die unter dem Namen *TEI.dictionaries* zusammengefasst werden. Diese Markierungen sind nur noch computerlesbar und benötigen ein Abfrage- oder Darstellungssystem, mit dem die im Wörterbuch enthaltene Information in verschiedenster Weise ausgegeben werden kann.

Eine Annotierung mit TEI verlangt einen ziemlich grossen Aufwand an Handarbeit, ermöglicht aber auf der anderen Seite eine absolut flexible Textsuche und Ergebnisdarstellung. Je nach Tiefe der Annotierung und mit entsprechendem Programmieraufwand wäre es möglich, quasi auf Knopfdruck neue Wörterbücher herzustellen, indem man auswählt, welche Elemente ein- und welche ausgeblendet werden.

Ich kann das nur noch kurz am Artikel *Sabel* demonstrieren. Das Lesen und Erfassen von Struktur und Inhalt eines Idiotikon-Artikels braucht sehr viel Konzentration und auch ein Stück Übung. Schon nur die Möglichkeit, einen Artikel anders darstellen zu können, kann beim Lesen hilfreich sein. Aus Platzgründen sind die Artikel im gedruckten Werk sehr dicht, man hat auch mit einiger Übung Mühe, die Artikelstruktur



Schweizerdeutsches Wörterbuch
 Schweizerisches Idiotikon
 Wörterbuch der schweizerdeutschen Sprache

Siglenauflösung

GRh., Wil; SCH; TH, Z tw.: 1. a) eig., wie nhd. Säbel, dann übh. von jedem (militärischen) Seitengewehr (doch zT. nur scherzh.). wohl allg. *Du hest eⁿ churzeⁿ* S. ... einsoldat, auf da ... Uli, der Schlosser. Eine Erzählung aus dem Emmenthal; von Arthur von Almen. Soldat und ha ... Bern 1897 (Gute Schriften 24). ... vorwärts [aus der ... steck heraus] oder ich hol den Sabel., VALMEN 1897. [Die entwaffneten Aufständischen] *sīⁿ gäng choⁿ chäreⁿ, mīⁿ söll-neⁿ doch 's Füsi oder der Sabel umegēⁿ*. EGÜNTER 1908. *Wo d' Französen Alles 'plünderet g'haⁿ heiⁿ, het-mī^{ch} Nūt so g'rou^weⁿ wie im Atti sīⁿs neu^w Ordinanzg'wer und de^r gröss Sabel, wo im Bürechrieg zwē Schueh olts Mäss ussen ab'brocheⁿ worden isch und doch no^{ch}-n-es guets Halbchlöfter g'macht het mit samt dem Hefti*. BWYSS 1863. S. auch

und einzelne Bedeutungsziffern zu erkennen. In einer elektronischen Version gibt es diese Platzbeschränkung nicht mehr und man kann die Artikel je nach Bedürfnis ganz unterschiedlich darstellen.

Eine digitale Version erlaubt es beispielsweise, die Siglen und Abkürzungen im Hypertextverfahren so zu verknüpfen, dass mit einem Klick jeweils die entsprechende Information eingeblendet werden kann.

Möglich wäre auch, beim Klicken auf die im Text enthaltenen Ortsangaben jeweils eine Karte mit der entsprechenden Lokalisierung einzublenden. Man könnte sich auch überlegen, ob jeweils alle Orte zu einer spezifischen Bedeutung auf der Karte dynamisch angezeigt werden könnten, so dass jeweils automatisch eine Art Sprachkarte entstünde.

Ebenso wäre es möglich, durch gezieltes Ein- und Ausblenden von verschiedenen Strukturteilen eines Artikels diesen leichter lesbar zu machen, etwa indem ausser dem Lemma nur die Bedeutungsangaben gegeben werden, so dass automatisch eine Artikeldisposition erzeugt würde, die die Übersicht über längere Artikel erleichtert.

Man könnte sich auch überlegen, ob es sinnvoll ist, zwischen rezenten mundartlichen und historischen Belegen zu unterscheiden, so dass man sich wahlweise die eine oder die andere Gruppe der Belege zeigen lassen könnte. Weiter könnte man alle Querverweise innerhalb des Wörterbuchs wie Synonyme und Verweise auf weitere Belege auflösen.

Diese kurzen Schlaglichter haben Ihnen hoffentlich gezeigt, dass eine Digitalisierung auch mit relativ kleinen, überschaubaren Projekten beginnen kann und bereits neue Zugänge zum Werk eröffnet. Wenn ich die Entwicklung der letzten Jahre im Bereich der Lexikografie und noch mehr im Bereich der Nachschlagewerke überhaupt anschau, dann glaube ich nicht, dass ein Unternehmen, das von einer breiteren Öffentlichkeit wahrgenommen werden will, um eine Digitalisierung herumkommt. Daher denke ich, dass sich für das Idiotikon nicht die Frage stellt, ob eine digitale Fassung hergestellt werden soll, sondern nur wann und wie man eine solche angeht.