

Ein Massiv wird zur Scheibe

Der größte deutsche Sprach-Speicher wird bald durchkämbar sein wie nie: Eine elektronische Version des Grimmschen Wörterbuches steht kurz vor der Vollendung.

Ganz ruhig ging es zu an den beiden Berliner Schreibtischen. „Nur das Kritzen der Feder war zu hören“, allenfalls „manchmal ein leises Hüstel“ von Jacob Grimm. „Die Brauen hoben oder senkten sich“ während des Arbeitens auch bei Wilhelm, dem jüngeren der Brüder, wie dessen Sohn Herman bemerkte. Beraten mussten beide sich kaum. „Zuweilen blickten sie in die leere Luft. Manchmal standen sie auf, nahmen

geschätzt, sondern weit über 100. Erst im Januar 1961 war das gigantische, seither unerreichte Werk endlich bis zum 32. Band und dem finalen „Zypressenzweig“ komplett, ein Quellenverzeichnis folgte gar noch 1971. 67 744 eng bedruckte Textspalten hatten die Grimms und ihre vielen Nachfolger gefüllt; in der Originalversion wiegen die 3000 Euro teuren Bände satte 84 Kilo – kaum etwas für enge Studierstuben und kärgliche Gelehrtengehälter.



BERT BOSTELMANN / BILDPICTO

Trierer Digitalisierungsexperten*: Ein Puzzle aus 300 Millionen Schriftzeichen

ein Buch heraus... und blättern darin.“ Mehr nicht.

Das Ziel war ja auch klar. In der Klausur der beiden unzertrennlichen Forscher lief in den Jahrzehnten um 1850 der Rohstoff für ein Projekt zusammen, das bis heute einzigartig geblieben ist: eine Übersicht des deutschen Wortschatzes, umfassend und historisch transparent wie nie zuvor, mit Belegen, die von den Grimms selbst und mehr als 80 Helfern auf etwa 600 000 „kleinen Duodezblättchen“, also Kartezetteln, gesammelt worden waren.

Nicht 4 Jahre freilich, wie der optimistische Wilhelm anfangs vermutet hatte, dauerte die Arbeit, auch nicht 15, wie später

Selbst eine Taschenbuchausgabe des Monumentalwerks, seit 1984 erhältlich, schlägt mit 499 Euro ins Kontor.

„Wir möchten deutlich darunter bleiben“, sagt Kurt Gärtner, seit kurzem pensionierter Altgermanist an der Universität Trier. Es wäre nur im Sinne der Grimms, wenn die elektronische Version des Wörter-Massivs nicht wieder nur für große Bibliotheken und ein paar Spezialisten zugänglich würde: Träumten sie doch nationalromantisch davon, dass ihr „Deutsches Wörterbuch“ als Sprachschatz „zum Hausbedarf“ zählen werde, aus dem im Familienkreis „mit Verlangen, oft mit Andacht“ vorgelesen werden sollte.

Weniger Andachtsstimmung als emsiger Betrieb herrscht momentan in einigen Räumen des „Kompetenzzentrums für

elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften“, einem nüchternen Bau auf dem idyllisch am Berghang gelegenen Campus der Trierer Universität. Im Frühjahr soll der elektronische Grimm fertig sein; für Thomas Burch, den technischen Leiter, läuft alles nach Plan. Bald, sagt er lächelnd, kann jeder, der Lust hat, am heimischen PC sein „Deutsches Wörterbuch“ durchstöbern – raffinierter und schneller, als die Brüder es je hätten träumen können.

Einen Vorgeschmack gibt es schon: Unter der Internet-Adresse www.DWB.uni-trier.de/index.html lässt sich der gesamte Textbestand des Werks aufblättern. „Bachenspeck“ oder „nachgültig“, „verfumfeien“ oder gar „Yoga“ – die Daten sind zur Stelle. Wie in der späteren CD-Rom-Version hängt ganz links eine Alphabetleiste, aus der man in die Liste der 331 056 Wörter gelangt. Ein weiterer Klick und der Volltext des Wort-Artikels erscheint in der Mitte des Bildschirms; ein Zusatzfenster rechts hilft bei längeren Eintragungen, sich im Aufbau zurechtzufinden.

Natürlich wird die CD-Rom erheblich mehr bearbeiten können: verschachtelte und spezielle Anfragen, auch allein nach Lyrikbeispielen, nach Quellenautoren – bei denen an Platz zwei nach Goethe nicht Schiller, sondern der frühe Wörtersammler Kaspar Stieler steht – oder grammatischen Sonderfällen; Kopiermöglichkeiten, Lesezeichen und obendrein eine möglichst getreue Nachbildung der ursprünglichen Buchseiten sind im Paket inbegriffen.

Um die 300 Millionen gedruckten Zeichen zu solcher Hochleistung zu trimmen, hat das junge Team um Kurt Gärtner, der in seiner Zunft einer der wenigen konsequenten Elektronikpioniere ist, seit fünf Jahren ein gigantisches Puzzle in mehreren, klar abgegrenzten Schritten durchgespielt: Erst mussten seine Mitarbeiter die Daten kontrollieren und abgleichen, um sie dann in eine abfragbare Form zu übertragen; parallel dazu entstand die übersichtliche Programmoberfläche.

Eingelesen wurde der Text nicht mit Scannern, sondern per Tastatur – in China. Gleich zweimal tippten des Deutschen nicht kundige, aber schriftzeichenbewusste Hilfskräfte in Nanjing die etwa 35 000 Seiten ab, so dass ein automatischer Abgleich beider Versionen schon die meisten Fehler bereinigte. Exotische Zeichen mussten dennoch von Hand in Trier nachgetragen werden, nicht nur Hebräisches oder Astrologiekürzel: Das Wort „Zickzack“ etwa illustriert eine kleine Sägezahnlinie.

Nach dieser Ausputzarbeit ist freilich erst eine Rohfassung erreicht. „Woher soll die Maschine wissen, dass ‚H. Sachs‘ derselbe Autor ist wie ‚Hans Sachs‘ oder bloß ‚Sachs‘?“, erklärt Hans-Werner Bartz, der sich über der Arbeit am Grimmschen Da-

* Thomas Burch, Hans-Werner Bartz, Vera Hildenbrandt, Kurt Gärtner.

Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm im Internet - Mozilla

DWB Vorworte | Quellenverzeichnis | Datenbanksuche | Benutzerhinweise | Enden

T G GRIMM bis GRINDBÜSZERIN (Band 9, Spalten 346 bis 372) ARTIKELGLIEDERUNG

A
B
C
D
E
F
G
H
I
J
K
L
M
N

GRIMM, *m.*
GRIMMBART, *m.*
GRIMMBÄRTIG
GRIMMBEBEND
-SCHNAUBEND
-WALLEND:
-BESCHÄUMT
-ENTBRANNT
-ERHITZT
-GERÖTET
-VERFLUCHT:
-VERSTÖRT
-VERZERRT
GRIMMBITTER
GRIMMBÖSE, *a.*
GRIMMBRUMM
GRIMMCHEN

GRIMM, *m.*
A. *form.* substantivbildung zum adj. grimm, aber anscheinend erst sekundär aus dem älteren subst. grimmī (s. grimme) entwickelt (vgl. PAUL *dtsh.* *gramm.* 2, 112); *ahd.* erst ein vereinzelt gagrim (s. igrimm A), auch *mhd.* noch selten, vgl. LEXER 1, 1084; dazu von dem grimme *Pol.* 226, 9; erst *nhd.* voll entfaltet als ersatz für das absterbende fem. grimme (s. d.); im älteren *nhd.* erscheint nicht selten ein *swm.* grimme, besonders *obd.*, zumal *alem.*: die ... wandelbaren ... sind ... on allen grimmen
HEROLD-FORER *Gesners thierb.* (1563) 15a: dasz sy (die bären) ired grimmen und wilde vergäzen 15b: ausz

GRIMM
A. *form.* substantivbildung zum adj. grimm, aber anscheinend
B. bedeutung u. gebrauch.

GRIMME
die älteste, im *ahd.* fast allein bezeugte (grimmī, grimmīn C)
1) von lebenden wesen: 'wut, wütender zorn': grymme furc
2) auch wo das subst. nicht auf lebende wesen geht, zumal
3) von besonderem interesse sind einige nur beim passional

GRIMMEN
urspr. starkes vb. aus derselben wurzel, die im adj. grimm
1) urspr. 'mit den zähnen knirschen', *freudere Dief.* 246c; d
2) sehr gern als subst. infinitiv: das grimmen primo est cor

Stichwort: _____

Internet-Version des digitalen „Deutschen Wörterbuchs“: Automatisch zerpfückt, automatisch wieder lesbar gemacht

tengebirge zu einem wahren Jäger der Sonderfälle entwickelt hat. Sogar für die scheinbar narrensichere Gliederung der Artikel in Teil I, Abschnitt 1, Fall a) und so weiter ließen sich die Wörterbuchredakteure im Laufe der Jahrzehnte Dutzende verschiedener Reihenfolgen einfallen, die nun möglichst rasch, also programmgestützt, vereinheitlicht werden müssen – ganz zu schweigen von der nicht weniger trickreichen Unterscheidung zwischen Trenn-, Binde- und anderen Strichen.

Das Ergebnis dieser langwierigen Mühen formt eine Übersetzungsroutine dann blitzschnell in den so genannten SGML-Code um, eine mit Markierungen durchsetzte Darstellung ähnlich der von Webseiten, und prüft das Output auch gleich nach losen Sinn-Enden durch. Hier wird erkennbar, warum die Trierer das lexikalische Rohmaterial nicht nach eigenem Standard, sondern nach den Regeln der weltweit wissenschaftlich anerkannten „Text Encoding Initiative“ aufarbeiten: Sowohl zur Prüfung wie zur automatischen Abfrage der bereinigten Daten können sie vorhandene Werkzeuge einsetzen.

Die elegante Darstellung im Bildschirmfenster etwa, aber auch das Zerpfücken der Artikel in rasch zu verarbeitende Datenbankbestände ist Sache von CoST (Copenhagen SGML Tool), einer Skriptsprache, für die im Grimm-Team Vera Hildenbrandt zuständig ist. Jeder Bildschirmklick, auch im Internet, setzt sozusagen unter der Motorhaube des Systems einen Mechanismus in Gang, der den gewünschten Abschnitt hervorholt, lesbar macht und, wenn nötig, aufgliedert.

Gerade diese Übersichtsfunktion ermöglicht nun, wovon Jacob Grimm schon 1854 in der Vorrede zum ersten Band des

Werks träumte, dass man sich „nach Bienenweise“ einfach mal „in die Kräuter und Blumen“ der bunten Muttersprachweise „niederlassen“ kann – ganz zu schweigen von den Recherchen, zu denen das gewaltige Sprachmaterial findige Forscher anregen dürfte. Meist kommen sie dabei sogar ein gutes Stück über den Kenntnisstand des Wörterbuchs hinaus.

Doch das ist kein Wunder bei einem Opus, dessen erste, noch recht schmale Bände von zwei Menschen allein verfasst wurden – Wilhelm Grimm übernahm den Buchstaben D, Jacob bewältigte den Rest von A bis zum Wort „Frucht“ –, während spätere Bearbeiter immer ausuferndere Artikel lieferten. Für „Amtmännin“ etwa notierte Jacob Grimm schlicht, dass man



Grimm-Brüder Jacob, Wilhelm
Bienenfleiß im Sprach-Kräutergarten

seine Mutter in der hessischen Heimat einst so genannt habe, während der längste Artikel, das Wort „stehen“, horrende 328 beispielgespickte Spalten umfasst.

Als der Weisheit letzten Schluss sehen Gärtners Leute ihr Werk ohnehin nicht. Umso mehr belegt der digitale Grimm, wie gute Planung mit Augenmaß die drohende Verewigung mancher großer Nachschlagewerke überwinden könnte. Bislang sind die Briten mit ihrem legendären, Grimm vielfach überflügelnden „Oxford English Dictionary“, das schon in zweiter Auflage auf CD verfügbar ist und laufend ergänzt wird, weit voraus. Aber vielleicht bekommt durch die neue CD-Rom auch die zweite „Grimm“-Auflage, seit Jahrzehnten in Arbeit, neuen Schub.

Die Trierer jedenfalls haben längst andere Riesenwerke aufgespürt, die sie wie den Grimm „retrodigitalisieren“ möchten. Wohl das spannendste Projekt für viele, denen historisches Wissen am Herzen liegt, ist der „Krünitz“. Diese „Oekonomische Enzyklopädie oder allgemeines System der Staats- Stadt- Haus- und Landwirtschaft“ ist mit ihren 242 Bänden, die von 1773 bis 1858 erschienen, eine ungeheure, gewöhnlich aber nur mühsam verfügbare Fundgrube der Fachkenntnis. Zudem findet kaum noch ein Historiker Zeit, sich durch die oft ellenlangen Artikel zu kämpfen. Da kommt eine Volltextaufbereitung à la Grimm gerade recht.

Getreu seiner Offenheitsmaxime hat das „Kompetenzzentrum“ schon die ersten Proben ins Netz gestellt; unter www.kruenitz.uni-trier.de sind sie zu besichtigen. Und Kurt Gärtners freundlich-gewitztes Lächeln signalisiert: Auch das wird nicht die letzte Tat der fleißigen Digitalisierer gewesen sein. JOHANNES SALTZWEDEL