

AKADEMIE DER WISSENSCHAFTEN UND DER LITERATUR

Abhandlungen der
Geistes- und sozialwissenschaftlichen Klasse
Einzelveröffentlichung • Nr. 9

**Thomas Burch, Johannes Fournier,
Kurt Gärtner und Andrea Rapp (Hrsg.)**

**Standards und Methoden
der Volltextdigitalisierung**

Beiträge des Internationalen Kolloquiums
an der Universität Trier, 8./9. Oktober 2001

AKADEMIE DER WISSENSCHAFTEN UND DER LITERATUR • MAINZ
FRANZ STEINER VERLAG • STUTTGART

2003

Zur geplanten Retrodigitalisierung des 'Ordbog over det danske Sprog'

Konzeption, Vorgehensweise, Perspektiven

VON JØRG ASMUSSEN (Kopenhagen)

Vorbemerkungen

Der Aufgabenbereich der 1911 gegründeten Gesellschaft für dänische Sprache und Literatur, 'Det Danske Sprog- og Litteraturselskab' (DSL), umfasst im Wesentlichen die Erarbeitung von Wörterbüchern der dänischen Sprache, die Ausarbeitung textkritischer Editionen der klassischen dänischen Literatur, die Bearbeitung und Herausgabe mittelalterlicher Rechtsdokumente, sogenannter Diplome, sowie das Aufbauen und Bereitstellen von Textkorpora in elektronischer Form.¹ Die DSL ist dem dänischen Kulturministerium unterstellt: die Aktivitäten der DSL werden teils durch das Kulturministerium und durch andere öffentliche Mittel, teils durch gemeinnützige private Stiftungen, insbesondere die Carlsberg-Stiftung, finanziert.

Im Wörterbuchbereich der DSL wird zur Zeit an drei Projekten gearbeitet: 'Gammeldansk Ordbog' ('Alddänisches Wörterbuch') ist ein Wörterbuch der mittelalterlichen dänischen Sprache und soll somit den Sprachstand des Zeitraums 1100–1515 beschreiben. Da die vom Kulturministerium bereitgestellten Mittel für dieses Projekt sehr begrenzt sind, steht die Vollendung dieses Wörterbuchs in weiter Ferne.

Die Gegenwartssprache wird in 'Den Danske Ordbog' ('Das dänische Wörterbuch', DDO) beschrieben; der Beschreibungszeitraum reicht von etwa 1950 bis in die unmittelbare Gegenwart. Das DDO ist ein weitgehend korpusbasiertes Wörterbuch, dessen Redaktion im Frühling 2002 abgeschlossen und das 2002–2003 als sechsbändiges Werk erscheinen wird. Das DDO wird genau wie das 'Gammeldansk Ordbog' auch als SGML-konformes elektronisches Dokument erfasst,² was das Bereitstellen einer elektronischen Version neben der herkömmlichen Buchausgabe erheblich erleichtern dürfte.

¹ Vgl. auch unter <http://www.dsl.dk>

² Zur Erfassung dient das SGML-Erfassungssystem GestorLEX der Kopenhagener Softwarefirma Textware A/S.

Schließlich wird an einer Reihe von Ergänzungsbänden zum 'Ordbog over det danske Sprog' (ODS), gearbeitet, dem 'ODS-Supplement': Der fünfte und letzte von diesen Ergänzungsbänden wird im Jahre 2003 abgeschlossen sein; das 'ODS-Supplement' wird ebenfalls digital erfasst.

Das 28-bändige 'Ordbog over det danske Sprog' ('Wörterbuch der dänischen Sprache', ODS) erschien im Zeitraum 1919–1956. Das Ziel dieses Werkes war die umfassende Beschreibung der dänischen Sprache des Zeitraums von 1700 bis 1950. Bedingt durch den langen Entstehungszeitraum sowie durch Änderungen des geplanten Umfangs im Laufe der Entstehungszeit – ursprünglich waren nur sechs Bände geplant – fällt die Beschreibung des Wortschatzes in den ersten Bänden relativ knapp aus; außerdem sind hier Wörter und Bedeutungen aus dem Entstehungszeitraum des ODS selbst kaum berücksichtigt. Das ODS-Supplement versucht, diesen Mangel auszugleichen und stellt somit auch einen nicht unwesentlichen Schlüssel zum Aufbau des ODS dar.

Das Ziel der Wörterbücher der DSL ist die umfassende Beschreibung des dänischen Wortschatzes vom Mittelalter bis in die Gegenwart hinein. Allerdings bleibt ein Wörterbuch des Frühneudänischen (etwa 1515–1700) dabei vorerst ein Desiderat.

Im Folgenden soll kurz die Entstehung und Entwicklung sowie die Konzeption und Struktur des ODS umrissen werden. Etwas breiteren Raum werden daraufhin Überlegungen zur Retrodigitalisierung eines Werkes wie dem ODS einnehmen. Schließlich soll noch auf eine Reihe von Perspektiven einer Digitalisierung hingewiesen werden. Im Zentrum der Erörterungen werden hierbei die Bedeutungsbeschreibungen des ODS stehen.

Entstehung und Entwicklung des ODS

Erste Vorarbeiten für ein umfassendes Wörterbuch der dänischen Sprache wurden ab 1882 vom Kopenhagener Professor für dänische Philologie Verner Dahlerup unternommen. 1901 geht er einen Vertrag mit dem damaligen Kopenhagener Verlag Nordisk Forlag über ein sechsbändiges Wörterbuch ein – und beginnt zusammen mit seiner Frau und einigen ehrenamtlichen Mitarbeitern das Sammeln von überwiegend literarischen sprachlichen Belegen. Die Arbeit geht sehr schleppend voran, Fristen und Umfangvorgaben können nicht eingehalten werden, das Projekt droht zu scheitern. Im Jahre 1915 wird es daher von der DSL übernommen und in einem festen redaktionellen Rahmen weitergeführt – nunmehr allerdings als wesentlich umfangreicheres Werk als ursprünglich geplant: 1954 erscheint der letzte von

insgesamt 27 Bänden, 1956 außerdem noch ein Quellenverzeichnis als 28. Band.

Als die DSL das Projekt übernahm, wurden als Grundlage für die weitere Arbeit äußerst detaillierte, schriftlich fixierte Redaktionsregeln erarbeitet. Aufgrund dieser Redaktionsregeln ist das ODS trotz seines Entstehungszeitraums von über 50 Jahren im Vergleich zu vielen anderen nationalsprachlichen Wörterbüchern dieses Umfangs strukturell ziemlich konsistent – vor allem sind die Redaktionsregeln erhalten, was sich als ein erheblicher Vorteil bei der Digitalisierung erweisen dürfte.

Konzeption und Struktur des ODS

Das ODS ist grundlegend ein historisches Bedeutungswörterbuch, das versucht, sämtliche Wörter und Wendungen des Zeitraums 1700 bis zur Mitte des zwanzigsten Jahrhunderts sowie ihre Bedeutungen möglichst detailliert zu erfassen. Dies führt in der Regel zu umfassenden und äußerst komplexen Artikeln, selbst bei scheinbar relativ überschaubaren Wörtern. So findet sich beispielsweise unter dem Stichwort *Mus* ('Maus') ein knapp vier Spalten umfassender Artikel; Abb. 1 zeigt einen kurzen Auszug hieraus.

Es dürfte unmittelbar nicht ganz leicht fallen, sich in einem dermaßen typographisch komprimierten Artikel zu orientieren: Eine schnelle, punktuelle Informationsfindung ist kaum möglich, eher muss der Benutzer den gesamten Artikel lesen. Man stelle sich vor, ein Deutschsprachiger möchte herausfinden, ob es im Dänischen Entsprechungen für Wendungen wie "mit Mann und Maus", "graue Maus" oder gar für Sprichwörter wie "Wenn die Maus satt ist, schmeckt das Mehl bitter" gibt. Es ist nicht unmittelbar zu ersehen, wo im Artikel man derlei Antworten finden kann.

Betrachtet man die Artikelstruktur des ODS einmal näher, so zeigt sich, dass ein Artikel zunächst aus dem Schlagwort besteht (dem bei homographen Einträgen eventuell eine römische Zahl vorausgeht), sodann bei Substantiven eine Angabe zum grammatischen Geschlecht ("en") enthält, worauf die Angabe der Aussprache folgt, allerdings in einer heute ungebräuchlichen phonetischen Notation. Der Platz bis zur ersten Bedeutungsnummer ist weiteren morphologischen sowie etymologischen Angaben – weitgehend mit Belegverweisen – gewidmet. Sodann folgt eine Gliederung in arabisch numerierten Grundbedeutungen (1–3), die wiederum in arabisch numerierte Unterbedeutungen unterteilt sein können. Diese – wie die Grundbedeutungen auch – können zudem in Bedeutungsabschnitte unterteilt sein, die durch zwei senkrechte Striche eingeleitet werden. Wendungen, Sprichwörter und andere

Wortgruppen stehen generell unter jeweils passenden Grundbedeutungen des Wortes aufgeführt. In Fällen, wo es keine passende Grundbedeutung gibt, werden Wortgruppen typisch einer zu diesem Zweck etablierten Grundbedeutung – z.B. “besondere Verwendungen” – zugeordnet, die dann häufig wieder weiter unterteilt wird, so im Artikel *Mus*. Dieses Prinzip erschwert das Auffinden von Wortgruppen erheblich und zwingt den Benutzer des ODS in der Regel dazu, einen gesamten Artikel zu lesen, um auszuschließen, dass er die gesuchte Information nicht übersehen hat. Die meisten Bedeutungsangaben werden mit überwiegend literarischen Belegen illustriert; bei einigen Wortgruppen kommt es allerdings vor, dass der redaktionelle Eintrag in der Form eines Belegs gegeben wird, was insbesondere dann orthographisch verwirrend sein kann, wenn der Beleg einer anderen orthographischen Konvention folgt als der redaktionelle Text.

Ausschließlich auf Bedeutungsangaben reduziert, ergibt sich die folgende Bedeutungsgliederung für den Artikel *Mus* – sämtliche Bedeutungserklärungen sind hier in etwas gekürzter Form ins Deutsche übersetzt.

- 1) Bezeichnung für kleine Nagetiere ...
 - || über die Maus als kleines, unscheinbares Tier
 - || manchmal auch über Tiere außerhalb der Familie der Muridae
- 2) Besondere Verwendungen der Bedeutung 1
 - 2.1) in Vergleichen
 - || *mit jemandem wie die Katze mit der Maus spielen* + Definition
 - || *hinter jemandem oder etwas wie die Katze hinter der Maus her sein* + Definition
 sowie sechs weitere Wortgruppen
 - 2.2) in anderen besonderen Verbindungen (Redewendungen u.ä.)
 - 2.3) in Sprichwörtern
- 3) über verschieden Dinge, die ... an eine Maus erinnern
 - 3.1) Muskel
 - 3.2) haariger Wuchs am Körper (insbes. an den Wangen)
 - 3.3) dasselbe wie *Gelenkmaus*
 - 3.4) die äußeren Geschlechtsteile der Frau
 - 3.5) dasselbe wie “*Kabelarmaus*”
 - 3.6) dasselbe wie *Fender*
 - 3.7) ein Stück Lunte, das als Zünder für Minen verwendet wird
 - 3.8) Bezeichnung für eine (maus)graue Porzellanschnecke
 - 3.9) dasselbe wie *Weidenkätzchen*

Wie die Übersicht zeigen dürfte, spiegelt die Bedeutungsgliederung den komplexen Aufbau des Artikels deutlich wieder. Ob die Gliederung an sich zweckmäßig ist, sei dahingestellt. Der Aufbau der Bedeutungsangaben selbst und seine Implikationen für eine digitalisierte Version des ODS sollen im folgenden Abschnitt kurz angeschnitten werden.

Überlegungen zur Retrodigitalisierung des ODS

Vom Umfang her gehört das ODS eher zu den bescheideneren historischen Bedeutungswörterbüchern. Trotz seiner 27 Bände beträgt die in ihm enthaltene Anzahl Zeichen – es sind etwa 100.000.000 – nur rund ein Drittel der Zeichenmenge des Grimmschen Wörterbuchs (DWB). Sowohl das 'Oxford English Dictionary' (OED), das 'Woordenboek der Nederlandsche Taal' (WNT) sowie das noch nicht vollendete 'Svenska Akademiens Ordbok' (SAOB) sind wesentlich umfangreicher als das ODS.

Gemein ist den obengenannten großen nationalsprachlichen Wörterbüchern des germanischen Sprachraums die Tatsache, dass sie alle zumindest zum Teil bereits in digitalisierter Form vorliegen, während der Plan einer Retrodigitalisierung des ODS erst kürzlich explizit formuliert wurde.³ Die im Vergleich zu den genannten Wörterbüchern recht späte Retrodigitalisierung des ODS hat somit den unschätzbaren Vorteil, dass die DSL sich bei der Planung und Durchführung eines derartigen Projektes an den Vorgängerprojekten orientieren kann.

Bei der Erfassung des Datenmaterials scheinen zwei prinzipielle Wege möglich – die sich z.T. auch kombinieren ließen, nämlich das Scannen des Materials mit anschließender OCR-Erkennung oder die Dateneingabe per Hand, also quasi ein Abtippen der Wörterbuchvorlage. Das schwedische SAOB z.B. wählte die Erfassung über Scanner und OCR. Dass dieses bei der oft komplexen Typographie historischer Wörterbücher vermutlich nicht das geeignetste Verfahren ist, davon zeugt allein die Tatsache, dass die optische Erfassung bereits 1983 begonnen, aber erst vor wenigen Jahren abgeschlossen wurde.⁴ WNT⁵ und DWB⁶ entschieden sich bei der Digitalisierung beide für die manuelle Erfassung – ein Verfahren offenbar, das sich relativ zügig durchführen lässt, wie insbesondere die Erfassung des DWB gezeigt hat.

Da das Schriftbild des ODS wie das der übrigen hier erwähnten Wörterbücher recht komplex und die Orthographie in den Zitaten darüber hinaus sehr uneinheitlich ist, dürften übliche OCR-Systeme mit der Umsetzung des Schriftbilds in eine digitale Zeichenrepräsentation überfordert sein und die

³ Ein Antrag auf Förderung liegt mittlerweile dem dänischen Kulturministerium vor. Eine Förderung durch das Kulturministerium ist jedoch seit dem dänischen Regierungswechsel vom November 2001 eher zweifelhaft, da die neue liberal-konservative Regierung drastische Kürzungen der öffentlichen Ausgaben im Kultur-, Forschungs- und Bildungsbereich angekündigt hat.

⁴ Vgl. <http://g3.spraakdata.gu.se/osa/index.html>.

⁵ Vgl. <http://www.inl.nl/WNT/>.

⁶ Vgl. <http://www.dwb.uni-trier.de>.

Fehlerquote bei der Erkennung entsprechend hoch liegen. Der Aufwand für die anschließende Korrektur wäre vermutlich immens und im voraus auch relativ unkalkulierbar. Da es jedoch das Anliegen der DSL ist, ein derartiges Projekt – einschließlich Bereitstellung auf CD-ROM und im Internet – mit einem festen, relativ straffen Budget von deutlich unter einer Million Euro innerhalb von drei Jahren durchzuführen, musste ein OCR-Verfahren wegen seiner Unwägbarkeiten ausscheiden. Die jetzige Planung beruht deshalb auf einer manuellen Erfassung durch das chinesische Erfassungsbüro TQY-DoubleKey.

Es sollen in diesem Rahmen keine technischen Einzelheiten bezüglich der Digitalisierung erörtert werden, zumal diese sich ohnehin an das DWB-Retrodigitalisierungsprojekt (vgl. BARTZ/CHRISTMANN/HILDENBRANDT/SCHARRES in diesem Band) anlehnen würden. Das Ziel der manuellen Erfassung wäre es, den Text weitgehend in seiner ursprünglichen typographischen Gestalt zu bewahren. Hiernach ließen sich verschiedene Wege der Umsetzung der typographischen Struktur in eine inhaltliche Struktur erproben. Ausgangspunkt für die Interpretation der typographischen Struktur würde diejenige Inhaltsstruktur bilden, die dem ODS-Supplement zugrunde liegt. Diese fußt auf den ursprünglich dem ODS zugrunde gelegten Redaktionsregeln, ist allerdings den strukturellen Bedürfnissen einer datenbankorientierten Manuskriptfassung mit dem Erfassungssystem *Compulexis*⁷ angepasst worden und daher etwas stringenter als die ursprüngliche ODS-Struktur. Anhand des Artikels *Bæver* ('Biber') lässt sich die Korrespondenz zwischen typographischer Gestalt und inhaltlicher Struktur aufzeigen. Die Typographie des Artikels *Bæver* wurde zu diesem Zweck von einem Redakteur des ODS-Supplements interpretiert und seine inhaltlichen Bestandteile – außer der Aussprache⁸ – wurden im Erfassungssystem des ODS-Supplements ihren entsprechenden Kategorien zugeordnet (vgl. Abb. 2). Ein Ausdruck eines dermaßen mit dem Erfassungssystem *Compulexis* erfassten Artikels ist mit wenigen für das ODS-Supplement generell geltenden Abweichungen identisch mit der Typographie des ODS.

Die Struktur sowie die Erkennung der einzelnen Strukturelemente anhand typographischer Merkmale soll hier nicht weiter erläutert werden; es geht hier lediglich um das Prinzip, dass sich durch weitere derartige exemplarische, manuelle Artikelumsetzungen ein Eindruck davon gewinnen ließe, welche Schwierigkeiten bei einer automatischen Umsetzung der typographi-

⁷ *Compulexis* wurde von der ehemaligen Oxforder Softwarefirma *Compluxis Ltd.* entwickelt.

⁸ Das ODS-Supplement gibt keine Angaben zur Aussprache.

schen in eine inhaltliche Struktur zu erwarten sind. Es wäre durchaus denkbar, dass die Struktur des ODS-Supplements als eine Zwischenstufe zu einer angestrebten XML-Struktur dienen könnte; diese Zwischenstufenstruktur würde das Einflechten der Ergänzungsbände in das ODS überdies vermutlich erheblich erleichtern.

Die in den Ergänzungsbänden verwendete Inhaltsstruktur ist also weitgehend kompatibel mit der Inhaltsstruktur des ODS, sie orientiert sich allerdings nicht an einem Standard wie z.B. SGML/XML, sondern wurde bereits in den achtziger Jahren speziell auf die Manuskripterfassung der Ergänzungsbände hin entwickelt. Verglichen mit dem SGML/XML-Standard zeichnet sie sich insbesondere durch die folgenden Merkmale aus: Es erfolgt nur eine Kennzeichnung terminaler Elemente, aber keiner Strukturen, und die Reihenfolge der einzelnen terminalen Elemente ist nicht durch eine strenge Grammatik festgelegt und von daher – innerhalb gewisser Grenzen – wesentlich freier als in SGML/XML. Überdies ist es möglich, die Typographie, in der Inhalte wiedergegeben werden sollen, elementunabhängig zu wechseln, obwohl gewisse Grundwerte vorgegeben sind. Aufgrund dieser Eigenschaften, insbesondere aber aufgrund einer nicht standardisierten Dokumentenbeschreibung, z.B. in Gestalt einer DTD, dürfte es methodisch problematisch sein, das ODS auf die Inhaltsstruktur des ODS-Supplements abzubilden. Deshalb wird diese lediglich als Zwischenstufe dienen können – teils als typographische Interpretationshilfe, teils als Hilfe beim Einflechten des ODS-Supplements ins retrodigitalisierte ODS. Als endgültige Strukturbeschreibung eines digitalisierten ODS einschließlich des ODS-Supplements wird allerdings eine geeignete, erweiterbare Struktur in der Form einer an den TEI-Guidelines orientierten XML-DTD dienen. Wesentlich hierbei ist, dass von vornherein die Möglichkeit vorgesehen sein sollte, das ODS zu einem späteren Zeitpunkt auch mit dem DDO zu kombinieren.

Die Phase der Umsetzung der typographischen Struktur in eine Inhaltsstruktur sowie die hierbei entstehenden Schwierigkeiten dürften weitgehend denen entsprechender Digitalisierungsprojekte ähneln und sollen hier nicht weiter erörtert werden.

Wesentlicher, aber auch wesentlich schwieriger, scheint die inhaltliche Interpretation derjenigen Aspekte der Artikelstruktur, die nicht durch die Typographie zum Ausdruck kommen, sondern implizit, häufig in den inhaltlichen Formulierungen im Artikel verankert sind. Dies sei anhand der Bedeutungserklärungen des Artikels *Mus* kurz dargelegt.

Wie sich später noch zeigen wird, können die Bedeutungserklärungen – selbst bei Wörtern, die relativ verwandte Dinge bezeichnen – im ODS recht unterschiedlich ausfallen, obwohl die formale Struktur der Artikel relativ

gleichförmig ist. Insbesondere ist zu beobachten, dass in einigen Bedeutungserklärungen das sprachliche Zeichen den Ausgangspunkt für die Erklärung bildet, in anderen Fällen jedoch das Bezeichnete als Bestandteil eines hierarchisch gegliederten ontologischen Systems.

Wird der Ausgangspunkt einer Bedeutungserklärung im sprachlichen Zeichen gelegt, ergibt sich das, was im Artikel *Mus* in den folgenden Definitionen zu sehen ist:

- 1) **Bezeichnung für kleine Nagetiere ...**
 || über die Maus als kleines, unscheinbares Tier
 || manchmal auch über Tiere außerhalb der Familie der Muridae
- 2) **Besondere Verwendungen der Bedeutung 1**
 - 2.1) **in Vergleichen**
 - 2.2) **in anderen besonderen Verbindungen** (Redewendungen u.ä.)
 - 2.3) **in Sprichwörtern**
- 3) **über verschiedene Dinge, die ... an eine Maus erinnern**
- 3.8) **Bezeichnung für eine (maus)graue Porzellanschnecke**

Die zoologische Grundbedeutung ist hier angegeben als "Bezeichnung für kleine Nagetiere". Man erhält somit eine Antwort auf die Frage, was das Wort "Maus" ist, nämlich: "Das Wort 'Maus' ist eine Bezeichnung für kleine Nagetiere". Nach diesem – bezeichnungsorientierten – Definitionsprinzip erhält man im Wesentlichen Beschreibungen der Verwendungsweisen des Wortes "Maus" – diese können auch konnotativer Art sein wie in "über die Maus als kleines, unscheinbares Tier".

Anders im neuen DDO, wo der Ausgangspunkt für die Bedeutungserklärungen stets der Inhalt des sprachlichen Zeichens bildet, und wo man folglich auf die Frage, was eine Maus ist, die folgende Antwort erhält: "Eine Maus ist ein Nagetier mit bestimmten Merkmalen." Nach diesem – inhaltsorientierten – Definitionsprinzip bilden die Definitionen selbst ontologische Hierarchien: Eine Maus ist ein Nagetier mit bestimmten spezifischen Merkmalen, ein Nagetier ist ein Säugetier mit bestimmten spezifischen Merkmalen, ein Säugetier wiederum ein Tier usw.

Auch das ODS verwendet das hierarchisch-ontologische Definitionsprinzip (es liegt auch vielen der bezeichnungsorientierten Definitionen implizit zugrunde), aber die Verwendung der Definitionsprinzipien ist nicht besonders konsequent im ODS: im Artikel *Mus* ist zwar die Hauptbedeutung 3 bezeichnungsorientiert definiert, ebenso wie die Unterbedeutung 3.8) "Bezeichnung für eine (maus)graue Porzellanschnecke"; die folgenden unter der Hauptbedeutung 3 verzeichneten Bedeutungsdefinitionen sind allerdings inhaltsorientiert definiert:

- 3.2) haariger Wuchs am Körper (insbes. an den Wangen)
- 3.4) die äußeren Geschlechtsteile der Frau
- 3.7) ein Stück Lunte, das als Zünder für Minen verwendet wird

Darüber hinaus finden sich die folgenden Bedeutungsdefinitionen, die Verweise auf andere Artikel sind:

- 3.3) dasselbe wie *Gelenkmaus*
- 3.5) dasselbe wie "*Kabelarmaus*"
- 3.6) dasselbe wie *Fender*
- 3.9) dasselbe wie *Weidenkätzchen*

Letztlich findet sich unter der Bedeutung 3 noch die synonymische Angabe

- 3.1) Muskel

Es stellt sich in diesem Zusammenhang die Frage, wie man sich bei einer Strukturierung des einmal erfassten Datenmaterials zu den unterschiedlichen Definitionsweisen verhält. Natürlich lassen sich anhand äußerer typographischer Kriterien relativ leicht z.B. die Haupt- und Unterbedeutungen in einem Artikel bestimmen und in entsprechende XML-Elemente setzen. Nehmen wir einmal an, wir hätten ein terminales Element `<definition>`, wobei wir der Einfachheit halber nicht zwischen den Haupt- und Unterbedeutungen, wie sie in der Typographie angelegt sind, unterscheiden; so würde dieses Element folgende, allein aus dem Artikel *Mus* abgeleitete, recht unterschiedliche Bedeutungstypen enthalten können:

Bedeutungstyp	Beispiel
Verwendungsangabe, Zeichen	<i>Bezeichnung für kleine Nagetiere</i>
Verwendungsangabe, Konnotation	<i>über die Maus als kleines, unscheinbares Tier</i>
Inhalt	<i>haariger Wuchs am Körper</i>
Synonym	<i>Muskel</i>
Verwendungsangabe mit artikelinternem Verweis	<i>Besondere Verwendungen der Bedeutung 1</i>
Verweis auf einen anderen Artikel	<i>dasselbe wie 'Fender'</i>

Es ist also auf jeden Fall von diesen sechs Typen von Bedeutungsdefinitionen auszugehen. Es wäre daher naheliegend, zukünftige Bedeutungselemente mit entsprechenden Typ-Attributen zu versehen. Ein Weg, diese teilweise automatisch zu ermitteln, wäre, sich an bestimmten, immer wiederkehrenden sprachlichen Formulierungen zu orientieren, wie "Bezeichnung für", "über ... als" usw.

Bedeutungstyp	Formale Kennzeichen
Verwendungsangabe, Zeichen	<i>Bezeichnung für ...</i>
Verwendungsangabe, Konnotation	<i>über [Schlagwort] als ...</i>
Inhalt	keines der übrigen Kennzeichen

Synonym	ein Wort
Verwendungsangabe mit artikelinternem Verweis	... <i>Bedeutung</i> [Ziffer]
Verweis auf einen anderen Artikel	<i>dasselbe wie</i> [ein Wort]

Perspektiven einer Digitalisierung

Anhand der Bedeutungserklärungen im Artikel *Mus* lässt sich exemplarisch zeigen, wie sich die Inhaltsstruktur eines Wörterbuchs nicht ausschließlich an der Typographie ablesen lässt, dass die Inhaltsstruktur häufig auch durch andere, nicht unbedingt formal erfassbare Zeichen zum Ausdruck kommt. Es dürfte davon auszugehen sein, dass dies nicht nur für Bedeutungsangaben, sondern auch für eine Reihe weiterer Bestandteile eines Wörterbuchartikels gilt. Man mag sich fragen, ob der Aufwand einer derartigen Kategorisierung von Bedeutungsangaben im Rahmen einer Retrodigitalisierung überhaupt lohnt und ob es nicht im Grunde genommen ausreichend und schon aufwändig genug wäre, sich mit der Auszeichnung des unmittelbar Sichtbaren, also der Interpretation der typographischen Struktur allein zu begnügen.

Die Digitalisierung eines Nachschlagewerks bedeutet unmittelbar die folgenden drei Vorteile gegenüber der Buchausgabe:

1. eine Erleichterung oder Beschleunigung der Recherche,
2. die Erweiterung der Recherchemöglichkeiten gegenüber der Buchversion, wobei dann insbesondere an Volltextrecherchen sowie auf direkten Zugriff auf bestimmte strukturelle Elemente gedacht wird,
3. einfache und allgemeine Zugänglichkeit, wenn ein digitales Wörterbuch über das Internet recherchierbar gemacht wird.

Man stelle sich vor, diese Ziele wären hinsichtlich des ODS bereits realisiert. Die Suche nach Schlagwörtern wie 'Mus' würde dann schnell zum gewünschten Artikel führen, der im Vergleich zur Buchvorlage sogar typographisch weit übersichtlicher gestaltet werden könnte. Auch die rasche Suche nach Wortgruppen wie dem dänischen Äquivalent für "mit Mann und Maus" ließe sich relativ leicht realisieren. Über die Möglichkeit einer Volltextsuche ließe sich 'querfeldein' durch sämtliche Artikel oder bestimmte Teile sämtlicher Artikel suchen.

Die Möglichkeit der Volltextsuche wäre allerdings auch die einzige unmittelbare Neuschöpfung im Vergleich zum herkömmlichen Wörterbuch. Es fragt sich daher, ob Geschwindigkeit und Breite bei der Recherche sowie der leichte Zugang tatsächlich die einzigen Vorzüge gegenüber einer Buchausgabe sein müssen, ob damit wirklich schon die Vorzüge des Mediums Com-

puter und Internet gegenüber dem Buch erschöpft sind. Technisch dürfte einiges mehr möglich sein; die Beschränkungen dürften hauptsächlich konzeptioneller Natur sein, denn die grundlegende Konzeption klassischer Wörterbücher ist verständlicherweise auf das Medium Buch ausgerichtet. Deshalb sollte man bei einer Retrodigitalisierung auch Möglichkeiten zur konzeptionellen Anpassung des digitalisierten Materials an das Medium Computer und Internet zumindest erwägen.

Im Hinblick auf die Bedeutungsbeschreibungen im ODS könnte eine solche Anpassung z.B. als Ziel haben, neben der herkömmlichen, wortbedeutungsorientierten Recherche ("Welche Bedeutungen hat das Wort Maus?") auch eine begriffsorientierte Recherche zu ermöglichen ("Welche Nagetiere gibt es?"). Dies würde aber eine Harmonisierung der Definitionsweisen erfordern, wobei u.a. auf die konsequente Verwendung geeigneter Oberbegriffe geachtet werden sollte, was gegenwärtig im ODS nicht der Fall ist, wie die folgenden Beispiele zeigen:⁹

Dänisch	Deutsch	Definition	Bd., ersch.
<i>bæver</i>	Biber	<i>Säugetier der Ordnung der Nagetiere.</i>	Bd. 3, 1921
<i>hamster</i>	Hamster	<i>(zool.) ausländisches Nagetier (aus der Familie der Mäuse), das im Herbst bedeutende Vorräte an Korn, Bohnen, Erbsen u. ä. von den Feldern sammelt; <i>Cricetus frumentarius</i> Pall.</i>	Bd. 7, 1925
<i>hare</i>	Hase	<i>Nagetier mit sehr langen Ohren, kurzen Vorderläufen und langen Hinterläufen, <i>Lepus</i> L.; insbesondere über die in Dänemark gew. Art: Grauhase, gew. Hase, <i>L. europæus</i>.</i>	Bd. 7, 1925
<i>kanin</i>	Kaninchen	<i>ein zur Hasengattung zählendes grabendes Nagetier, <i>Lepus cuniculus</i> L., in neuerer Zeit auch als Haustier gehalten (wegen seines Fleisches).</i>	Bd. 9, 1927
<i>mus</i>	Maus	<i>Bezeichnung für kleine Nagetiere der Gruppe Murinae (Gegensatz Ratte), insbes. für die Hausmaus; auch (zool.; insbesondere Mz.) m. weiteren Verw., teils als Gesamtbezeichnung für alle Tiere der Nagergruppe Murinae (echte Mäuse), teils für diese und verschiedene nahestehende Nagergruppen (Hamster u. a.), die zusammen die Familie der Muridae bilden.</i>	Bd. 14, 1933

⁹ Die Beispiele zeigen überdies, wie uneinheitlich die Bedeutungsdefinitionen hinsichtlich der Menge und der Art der in ihnen enthaltenen Informationen sind. Insbesondere die sehr knappe Definition des Stichworts *bæver* zeigt darüber hinaus, dass zur Zeit der Entstehung dieses Artikels lediglich ein sechsbändiges Wörterbuch geplant war.

rotte	Ratte	<i>größeres Nagetier der Gattung Maus (Familie der Murinae), insbes. Mus rattus (schwarze Ratte; vgl. Hausratte; bereits im Mittelalter eingewandert) od. M. decumanus od. norvegicus (braune Ratte; vgl. Wanderratte; hierzulande eingewandert Anfang des 18. Jahrhunderts), von welchen jetzt insbesondere die letzte Art ein sehr verbreiteter Schädling ist.</i>	Bd. 17, 1937
-------	-------	--	--------------

Die Oberbegriffe, die für eine begriffsorientierte Recherche im Datenmaterial ausgezeichnet sein müssten, sind nicht einheitlich "Nagetier", sondern "Nagetier", "Säugetier" und "Bezeichnung".

Wenn hier von Harmonisierung der Definitionsweisen die Rede ist, so bedeutet dies nicht notwendigerweise, dass die bisherigen Strukturen und Definitionen aufgelöst und von Grund auf neu bearbeitet und formuliert werden müssten, es heißt vielmehr, dass für erweiterte Recherchemöglichkeiten das digitalisierte ODS mit entsprechenden – für den Benutzer möglicherweise unsichtbaren – Auszeichnungen versehen werden könnte, z.B. Auszeichnungen, die bestimmte Bedeutungen bestimmten Stufen einer begrifflichen Hierarchie zuordnen. Das Erarbeiten derartiger Begriffshierarchien stellt dabei natürlich auch eine nicht unwesentliche Herausforderung dar.

Das Einbringen zusätzlicher Information in ein retrodigitalisiertes Werk, um es an das neue Medium Computer und Internet anzupassen, um so neue Recherchemöglichkeiten zu erschließen, sei hier lediglich als ein Beispiel genannt dafür, dass der Digitalisierungsprozess nicht nur die Umsetzung eines Buch-Datenbestands in einen Computer-Datenbestand bedeuten sollte, sondern auch die Umsetzung eines Buch-Konzepts in ein Computer und Internet-Konzept.

Die DSL wird bei einer eventuellen Retrodigitalisierung des ODS diesen Aspekt stets im Auge behalten. Das unmittelbare Ziel wird natürlich auch für die DSL die Umsetzung eines Buchdatenbestands in einen Computerdatenbestand sein, allein schon deshalb, weil es bei der Retrodigitalisierung um die Erhaltung eines historischen Werkes gehen sollte. Allerdings eröffnet ja gerade die Retrodigitalisierung einen erneuten redaktionellen Zugang zum Wörterbuch, den man nutzen sollte. Hand in Hand mit der rein konzeptionellen Anpassung des Wörterbuchs an ein neues Medium scheint eine redaktionelle Überarbeitung, die sowohl eine strukturelle als auch eine sprachliche Überarbeitung beinhalten sollte, durchaus lohnenswert. Durch eine derartige, behutsame Modernisierung dürfte das ODS sich nämlich einem weiteren Kreis potentieller Benutzer erschließen, insbesondere solchen, die eine eher schnelle, punktuelle Informationsfindung der allumfassenden Beschreibung

sämtlicher sprachlicher Aspekte eines Wortes bevorzugen. Das ultimative Ziel für die DSL wäre dabei ein lexikographisches Informationscenter, das sämtliche Wörterbücher der DSL in sich vereinen könnte und das der Allgemeinheit online zur Verfügung gestellt werden könnte. Auf diesem weiten Weg wäre die Retrodigitalisierung ein erster, kleiner Schritt.

I. Mus. en. [mu's] *Høysg. AG. 46.* (i bet. 3.1 ogs. † **Musc.** *Oecon. Journ. 1757. 52. v. Aph. Nath. I. 80.*) flt. d. s. ell. † -e (*LTid. 1740. 759. Reynike Fosz. (1747). 96.*) (glda. d. s. (flt. mōs, mys), ælda. d. s. (*Harp. Kr. 9. flt. mys. smst. 158.*), oldn. mūs (flt. mýss), eng. mouse, ty. maus, lat. mus, gr. mŷs; jf. Muskel, Musling, mosle, mysle, Mysling)

1) navn paa smaa gnavere af gruppen *Murinae* (mods. Rotte), spec. om husmusen; ogs. (zool.; især i flt.) m. videre anv., dels som fællesbetegnelse for alle dyr af gnavergruppen *Murinae* (de ægte mus), dels om disse og forsk. nærstaaende gnavergrupper (hamstere oft.), der tilsammen danner familien *Muridae*. de, som æde Svinekjød og vederstyggelige Ting og Muus: de skulle bortrives. *Es. 66. 17.* Rotter og Muus. *Holb. DR. I. 5.* *Forgangen Dag vor sluen Kat, | en ung og lekker Muus fik fat. *Thielos Arier. II. 36.* *Der sad to Muus . . | Sort Een, og hvid den Anden var. *Winth. D. 46.* *Rotter og Mus | Er' slemme i Hus. [*KH With. Børnebog. (1865). 67. Bøv P. I. 692.*] (jf. bet. 2.2) om musen som et lille, uanseligt dyr. *Ei den mindste Mus | Savner Tag og Hus. *Ing. RSE. VII. 239.* jf.: en lille Mus af en Kvinde. *V. Bergstrøm. HN. 92.* || undertiden ogs. om dyr uden for familien *Muridae* (navnlige gnavere), der minder om musene; især i ssgr. som Birke-, Blind-, Busk-, Hare-, Hassel-, Springmus, jf. ogs. Bælte-, Flager-, Spidsmus samt (om dyr uden for pattedyrene) Fjeder-, Guld-, Havmus. † den polske mus, europæisk steppeegern, *Spermophilus citillus. VSO. III. 170.*

2) særlige anvendelser af bet. 1. 2.1) i sammenligninger. *Paa Pølserne, som Kat paa Muus, han gloede. *Wess. 203.* De ligger . . og piber som en Mus i Barselsnød. *P. Møll. I. 388.* *Saa, før man saae sig om, | Var hun borte som en Muus. *Winth. HF. 44.* Fru K.s Øjne løb om i Stuen, sky undersøgende som en Mus's. *Schand. AE. 241.* *Her sidder du bag disse Vægge | og kryber sammen som en Mus. *Sødb. GD. 55.* jf. Flæsk

Abb. 1: Auszug aus dem Artikel *Mus*

Der Artikel *Bæver* wird im Folgenden typographisch annähernd so wiedergegeben, wie er im ODS steht. Eine Interpretation seiner inhaltlichen Elemente nach den für das ODS-Supplement geltenden Regeln ist in der unten stehenden Tabelle zu sehen. Die Kategoriebezeichnungen folgen dabei denen, die im Erfassungssystem Compulexis verwendet werden. Das Zeichen ¶ dient als Platzhalter für ein nachfolgend angegebenes Element: Das erste ¶ unter HST wird so z.B. ersetzt mit dem Inhalt aus dem ersten Element nach HST, das zweite mit dem zweiten usw.

Bæver, en. [...] béver. *Høysg.*
Anh.19. flt. -e ell. † d.s. (Holb.Berg.109).
 (ænyd. bever (vist i form paavirket af mnt.
 bever), glda. biæver, æda. (som tilnavn)
 biauer (*ValdJord.85; jf. stednavnet Bjevers-*
skov, æda. byaverscog. LundeDomkapitels
Gavebøger.(1884/89).71), oldn. bjórr, eng.
beaver, ty. biber, lat. fiber, jf. sanskr.
babhrú-, brun, ogs.: stor ichneumon; vist
besl. m. Bjørn, brun) 1) pattedyr af gna-
vernes orden, Castor L. Holb.MFbl.114.
*Epont.Atlas.I.610. *Er det kun Drift hos*
Bæveren, at bygge | Sig Hytte? Oebl.XII.
23. Hauch.II.313. vi see Bæveren bygge
sig Huse i Vandet og anlægge kunstige
Dæmninger for at sikre dem. Drejer.(Da
Folkekal.1841.141). inde mellem de sam-
menhobede Træer byggede hele Flokke
af Bævere Byer.JVJens.Br.83. 2) (især
fagl.) skind el. pelsværk af bæver (1);
bæverskind. en muffe af bæver|

Kategori	Beschreibung	Inhalt
HWD	Schlagwort	Bæver
GEN	Genus	en.
FLX	Flexionskategorie	flt.
BØJ	Flexiv	-e
FLX	Flexionskategorie	ell.
FRQ	Häufigkeitsangabe	† (Kreuzzeichen)
BØJ	Flexiv	d.s. 'dasselbe'
KIL	Quellenangabe	(Holb.Berg.109)
HST	Etymologische Angaben	(ænyd. [bever] (vist i form paavirket af mnt. [bever]), glda. [biæver], æda. (som tilnavn) biauer (¶; jf. stednavnet Bjeverskov, æda. by- averscog. ¶), oldn. bjórr, eng. beaver, ty. bi- ber, lat. fiber, jf. sanskr. babhrú-, brun, ogs.: stor ichneumon; vist besl. m. ¶, ¶)

KIL	Quellenangabe	ValdJord.85
KIL	Quellenangabe	LundeDomkapitels Gavebøger.(1884/89).71
HEN	Verweis	Bjørn
HEN	Verweis	brun
BNR	Bedeutungsnummer	1
DEF	Definition	pattedyr af gnavernes orden, Castor L.
KIL	Quellenangabe	Holb.MFbl.114.
KIL	Quellenangabe	EPont.Atlas.I.610.
CIT	Beleg	*Er det kun Drift hos Bæveren, at bygge Sig Hytte?
KIL	Quellenangabe	Oehl.XII.23.
KIL	Quellenangabe	Hauch.II.313.
CIT	Beleg	vi see Bæveren bygge sig Huse i Vandet og anlægge kunstige Dæmninger for at sikre dem.
KIL	Quellenangabe	Drejer.(DaFolkekal.1841.141)
CIT	Beleg	inde mellem de sammenhobede Træer byggede hele Flokke af Bævere Byer.
KIL	Quellenangabe	JVJens.Br.83.
BNR	Bedeutungsnummer	2
STL	Stilistische Angabe	(især fagl.)
DEF	Definition	skind el. pelsværk af bæver (1): bæverskind
CIT	Beleg	en muffe af bæver.
KIL	Quellenangabe	(redaktioneller Beleg)

Abb. 2: Korrespondenz zwischen typographischer Gestalt und inhaltlicher Struktur im ODS.