

Prof. Dr. Winfried Lenders

Vorlesung

Maschinelle Lexikographie und Wissenrepräsentation

Teil I: Lexikographie  
Kap. 2

## Kap. 2 Geschichte der Lexikographie

2.1 Motive zur Herstellung von Wörterbüchern

2.2 Phasen der Lexikologie und Lexikographie in Deutschland bis zur Gegenwart

2.3 Neue Impulse durch die Informationslinguistik

2.4 Begriffe: *machine readable* vs. *machine tractable dictionaries*

2.5 Einige weitere Begriffe

## 2. Geschichte der Lexikographie/Lexikologie

Drei Motive zur Herstellung von Wörterbüchern:

- a) Übersetzung von einer Sprache in eine andere (vgl. die frühen Glossare zur lat.-gr. Übersetzung; in altdt. Zeit zur lat.-dt. Übersetzung, hervorgegangen aus den malichen Interlineartexten (vgl. ahd. Abrogans)).
- b) Bewahrung, Sammlung, Dokumentation des Wortschatzes einer Sprache, einer Sprachepoche, eines Autors, einer soz. Schicht etc.
- c) Normierung der Wortbedeutung, Aussprache, Orthographie, des syntaktischen Gebrauchs etc. Entstanden nach 1750

In Deutschland große Zeit der Lexikographie nach 1500, im Zeitalter des Humanismus, der Reformation und der Renaissance.



Tontafel mit einer **lexikalischen Liste** von Schweinebezeichnungen.

Uruk, W 121139. OeXVI4, Zeit der Schicht Iva (um 3200 v. Chr.)

Ton, L. 7,6 cm, B. 5,1 cm , H. 1,5 cm.

Berlin, Staatliches Museum zu Berlin, Vorderasiatisches Museum.

Abb. aus: Der Turmbau zu Babel. Katalog zur Ausstellung des Kunsthistorischen Museums Wien in Graz. Wien 2003, Bd. III B, S. 32.

## 5 große Bestrebungen dieser Zeit:

1. Neuentdeckung des Griechischen und Lateinischen (Ducange: *Thesaurus linguae latinae*)
2. Entstehung der Philologien
3. Entdeckung des Zusammenhang der indoeuropäischen Sprachen
4. Entdeckung neuer Sprachen in neuen Erdteilen (Motiv der Übersetzung)
5. Enzyklopädischer Gedanke. (frz. Enzyklopädie: ähnliche Unternehmungen in anderen Ländern). Umfassende Dokumentation der Nationalsprachen. Es entstand der Begriff des Wörterbuch. (s. die Einleitung zum Wörterbuch der Gebr. Grimm, IX: Gedanke des Wörterbuchs als umfassende Sammlung. (Anno 1854).

Vier Phasen der **Lexikologie** in Deutschland (nach Stötzel), d.h. der

Wissenschaft von Struktur und Aufbau der Wörterbuch, und sich daraus ergebende Wörterbücher:

### 1. Phase: 17. Jahrhundert

Methodendiskussion zwischen Schottelius, Harsdörffer, Gneisitz und Fürst Ludwig von Anhalt-Köthen innerhalb der "Fruchtbringenden Gesellschaft" (Henne 1968, 80-114)

Ergebnis: Gerüst für ein großes dt. Wörterbuch (1663). Wörterbuch von Schottelius:

**J.G. Schottelius:** Ausführliche Arbeit von der Teutschen Haubt-Sprache, Braunschweig 1663, hg.v.W. Hecht (= Deutsche Neudrucke, Reihe Barock 11), Tübingen 1967, S. 1270-1450 (=5. Buch, 6. Tractat).

Schottelius stellt ca. 5000 alphabetisch geordnete Stammwörter auf, meist endungslos. Die Stammwörter sollten Wurzeln der germanischen Sprache sein.

Idee des Wortfamilienbuchs

## 2. Phase: von 1700 bis Adelung (1747)

Übergang vom wortbildungsbezogenen Stammwortprinzip (Schottelius) zum Einzelwortprinzip

Grund: Bequemlichkeit des Benutzers

erstes großes Wörterbuch:

**Joh. Christoph Adelung:** Versuch eines vollständigen grammatisch kritischen Wörterbuch der hochdeutschen Mundart...". Leipzig 1774-1786.

## Wörterbuchtypologie von G.W. Leibniz:

- a) nach Gebrauchsbereichen: WB der häufigsten Wörter des Hochdeutschen und der Mundarten WB der Termini
- b) nach Anordnungskriterien: Bedeutungs- und Benennungswörterbücher

### 3. Phase: seit 1774

Adelungs Wörterbuch enthielt eventuell stilist. Wertungen, insofern die Schriftsprache von Klopstock, Herder, Wieland, Schiller und Goethe als Norm angenommen wurde. Das 'meißnersche' Deutsche galt als vorbildlich.

Im übrigen: Wortschatz wird als nicht fixierbar angesehen.  
nicht normative, sondern deskriptive Vorgehensweise.

In Zweifelsfällen gilt der Sprachgebrauch als Maßstab für Sprachrichtigkeit.

bei Adelung:

Synchronischer Ansatz, keine Ethymologie. Aber: diachronische Anordnung der Wortbedeutung.

Adelung war kein Purist. Fremdwörter werden zugelassen.

Kritiker Adelungs: August Hermann Campe: Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke. 1801.

Wörterbuch als Mittel der Sprachreinigung.

## 4. Phase: seit Mitte des 19. Jahrhunderts

Etymologie als neue lexikologische Methode  
Textphilologie als neues Interessengebiet

Idee des panchronischen Belegwörterbuch der **Gebrüder Grimm.**

2 Gesichtspunkte:

- a) Philologie
- b) Sprachwissenschaft

## Grimmsches Wörterbuch:

philologisches Wörterbuch, sprachgeschichtliche Ableitungen  
Wortschatz der Vergangenheit, historische Belegsammlung, Archiv





## **Wörterbücher in Zuge der strukturalistischen Sprachwissenschaft**

Wörterbücher als Normbücher, als Hilfsmittel, als Auskunftsbücher über Sprache (DUDEN) aufgefasst:

nach 1945 mehrere große deutsche **Wörterbücher**:

Ⓒ Mackensen (4. Aufl. 1962)

Ⓒ Bertelsmann-Wahrig (Gütersloh 1966, Neuausgabe 1975)

Ⓒ DUDEN, 6-Bändig

Ⓒ Brockhaus-Wahrig - Deutsches Wörterbuch, in 6-Bänden. Hrsg. v. G. Wahrig, H. Krämer, H. Zimmermann.

- C dtv-Wahrig (1978; Neuausg. Hg. v. R. Burfeind-Wahrig 1997; weitere Auflage 2000)
- C Duden-Universalwörterbuch (2. völlig neu bearb. und erw. Auflage 1989; 4., neu bearbeitete und erweiterte Auflage Mannheim 2001)
- C zahlreiche Spezialwörterbücher

## 2.4 Neue Impulse durch die Informationstechnik

Entwicklung der maschinellen Lexikographie seit ca. 1960.

Computer als Hilfsmittel bei der Verwaltung und Aufbereitung großer Datenmengen

Index- und Konkordanzherstellung

Der Einsatz des Computers führte dazu, dass die Probleme der Lexikographie schärfer gesehen und die Aufgaben exakter formuliert wurden, vor allem hinsichtlich der Darstellung der Wortartikel.

1960: Kolloquium über maschinelle Methoden der literarischen Analyse und der Lexikographie in Tübingen.

Zunächst vielfache Arbeit mit Lochkartentechnik.

Verwendbarkeit von Datenbanken. G. Wahrigs Schrift "Neue Wege in der

Wörterbucharbeit" von 1967 brachten den Durchbruch.

Heute: Zahlreiche Wörterbücher auf CD-ROM verfügbar und integrierbar in Schreibsysteme wie MSWord, WordPerfect etc.

Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf 2 CD-ROMs, Verlag Zweitausendeins

## **Gründe für die heutige Bedeutung des Wörterbuch in der CL:**

1. Umgang mit Wörterbüchern als neu Form der technisch gestützten Informationserschließung.

Wörterbücher auf Personalcomputern, Computerterminals, Btx, Bildplatten, CD-ROM etc. (Oxford English Dictionary, van Dale Lexitron, Webster, Longman, Collins, DUDEN, Langenscheid etc.)

2. Neue sprachverarbeitende Systeme:  
Maschinelle Sprachübersetzung,  
Schnittstellen zu Expertensystemen,  
Frage-Antwort-Systemen,  
maschinelles Information Retrieval
  
3. Neue linguistische Modelle, die dem Wörterbuch (Lexikon) eine wichtige Rolle beimessen.

Generative Transformationsgrammatik von Noam Chomsky: Lexikon als Teil der Basis einer Generativen Grammatik

Fillmore: "The Case for Case", 1968

Lexical Functional Grammar und andere funktionale Grammatik

Diskurstheorie

Verstärkte Projektstätigkeit auf dem Gebiet der maschinellen Lexikographie nach 1970:

- C Gerhard Wahrigs Projekte mit der DFG, Bertelsmann und Brockhaus
- C Japan: "Japan Electronic Dictionary Research Institute Ltd." seit ca. 1987
- C Arbeitsstellen in Schweden, Italien, Niederlande, und Frankreich
- C New Oxford English Dictionary, Cambridge Univ. Press, Univ. of Waterloo; Collins COBUILD

Für das Deutsche: Keine zentrale öffentlich geförderte Einrichtung, deren Aufgabe die Zusammenstellung eines maschinenlesbaren Korpus der deutschen Gegenwartssprache und dessen lexikographische Auswertung wäre.

Zum Teil Kooperation zwischen Verlagen, Computerfirmen und Forschern.

Programme der Europäischen Union: "language industries"

Zur Zeit sind mehrere große Vorhaben im Gespräch:

- Ⓒ Das digitale Wörterbuch des Deutschen (Berlin-Brandenburgische Akademie der Wissenschaften) ([www.bbaw.de/iag/dig\\_woerterbuch/index.html](http://www.bbaw.de/iag/dig_woerterbuch/index.html))
  - Ⓒ Großes Deutsche Wörterbuch am IDS
  - Ⓒ Neubearbeitung des Grimmschen Wörterbuchs
- daneben verschiedene Arbeiten auf dem Feld der maschinellen Lexikographie im weitesten Sinne (z.B. GermaNet in Tübingen).

Perspektiven:

Wörterbücher als Netzwerke: WordNet, FrameNet

Verwendung im Semantic Web

## 2.5 Grundlegende Begriffe

maschinenlesbare Wörterbücher und maschinenverarbeitbare Wörterbücher

### a) maschinenlesbare Wörterbücher (*machine readable dictionaries*)

MRDs sind

- ! elektronische Versionen der gedruckten Gebrauchswörterbücher, die überwiegend auf Datenträgern (früher: Satzbändern) vorliegen
- ! speziell mit EDV und für die maschinelle Informationsverarbeitung erstellte Gebrauchswörterbücher

weitere Perspektiven des maschinenlesbaren Wörterbuchs:

Hat das Wörterbuch in Buchform auf längere Sicht noch eine Chance oder wird es durch andere Formen der Präsentation und Verbreitung lexikalischer Information abgelöst?

Im Bereich der Fachsprachen:

integrierte lexikalische Gesamtsysteme für den Lexikographen, den Übersetzer, in Büros, Redaktionen, für Wissenschaftler und Schriftstellern usw.  
setzen sich durch.

Es gibt für Wörterbücher neue Präsentationsformen und Nutzungsumgebungen:

- WB auf PC oder Laptop, auf CD-ROM oder DVD
- WBer im Internet
- "Dictionary Server" (WB auf dem Terminal) als Teil eines
  - integrierten WB-Systems, erstellt aus vorliegenden Wbern mit Hilfe der Satzdaten
  - Durchgriff auf lexikalische Erfahrungssammlungen und Texte

## **Wörterbücher im Datennetz**

Das einsprachige alphabetische Bedeutungswörterbuch ist nur eine Komponente eines umfassenden maschinell zugänglichen lexikalischen Systems.

Zugriff auf vernetzte lexikalische Datenbanken

Hypertextartige Verknüpfung der lexikalischen und textualen Ressourcen.

## Beispiele:

### elektronische Gebrauchswörterbücher auf CD-ROM

- DUDEN-Universalwörterbuch
- dtv-Wahrig

### Wörterbuchserver:

- z.B. EURODICAUTOM

### Web-Seite:

<http://www.phil.uni-erlangen.de/~p2gerlw/ressourc/lex.html>

Komponenten, die in ein lexikalisches Netz integriert werden können:

- Rechtschreib- und Aussprachewörterbücher, auch für Fremdwörter;
- alphabetische Bedeutungswörterbücher und konzeptionell geordnete Wörterbücher (Synonymenwörterbücher etc.);
- zweisprachige Übersetzungswörterbücher, jeweils für Hin- und Her-Übersetzung;
- Bedeutungswörterbücher, Rechtschreib- und Aussprachewörterbücher in anderen Sprache;
- idiomatische Wörterbücher, mehrsprachig und in mehreren Sprache;
- Wörterbücher schwieriger Übersetzungsfälle (z.B. Erfahrungssammlungen literarischer Übersetzer);
- terminologische Wörterbücher;
- Enzyklopädien;
- Texte mit Belegen etc.

Arbeitsplätze z.B. für Übersetzer, Terminologen, Lektoren mit Anschlüssen an Informationssysteme, Übersetzungssysteme etc.

Bisherigen Ordnungsstrukturen (z.B. die alphabetische) bedeutungslos.

Vermeidung von Redundanzen

## b) maschinenverarbeitbare Wörterbücher (*machine tractable dictionaries* (MTD))

Wörterbücher für Computer zur grammatischen, d.h. z.B. morphologischen, syntaktischen, semantischen, pragmatischen Analyse in MÜ, Information Retrieval, Natürlichsprachlichen Systemen

Anfangs nur Angaben zur Wortklasse und Flexion der Wörter

heute auch **semantische** Information, z.B. Selektionsbeschränkungen

Semantik im traditionellen Wörterbuch: natürlichsprachliche Bedeutungsparaphrasenangaben

Semantik im maschinenverarbeitbaren Wörterbuch: Konstrukt sprachliche repräsentationssprachliche Angaben

## **2.6 Hauptunterschied zwischen *machine readable* und *machine tractable dictionaries*:**

MRDs sind für die Benutzung durch den Menschen bestimmt, daher:

Alle Information möglichst in natürlicher Sprache!

MTDs sind für die Benutzung durch den Computer bestimmt, daher:

Die Angaben sind in künstlicher, formaler, kanonischer Sprache.

## **Was ist eine kanonische Angabe?**

Die Angaben sind einem Kanon von Beschreibungsmöglichkeiten entnommen. Vgl. die Gebrauchsanleitung eines Wörterbuchs.

Morphologische Angaben in MRDs sind in der Regel in kanonischer Form, z.B. Wortklasse, Flexionsklasse etc.

Bedeutungsangaben (=semantische Angaben) sind in MRDs in natürlicher Sprache:

## Beispiele für Wörterbücher mit natürlichsprachlicher Bedeutungsangabe:

DTV-Wahrig (Wahrig 1978, 421)

- horchen**
- 1 aufmerksam zuhören; horch:
  - 2 etwas zu hören versuchen, auf ein Geräusch warten; ich horche schon die ganze zeit, ob ich nicht seine Schritte höre; auf die Atemzüge eines Kranken horchen.
  - 3 <umg.> lauschen, heimlich mithören an der Tür; an der Tür lauschen

## COLLINS COBUILD

- listen**
- 1 If you **listen** to someone who is talking or to a sound that you can hear, you give your attention to them or to it.
  - 2 If you **listen** for something that you are expecting to hear or **listen out** for it, you keep alert and make an effort to be ready to hear it if it occurs.
  - 3 You say **listen** when you want someone to pay attention to you because you are going to say something important.

In dieser Form kann ein Computer die Bedeutungsangaben nicht verarbeiten!

Daher: Bedeutungsangaben in formaler Sprache:

Beispiele: aus dem Wb eines MÜ-Systems (CAT-2)\_

lex={ lu=zelt,lex='Zelt',head={ cat=n,ehhead={ case=(nom;acc;dat),agr={ num=sing } } } }.[].

lex={ lu=zelt,lex='Zelte',head={ cat=n,ehhead={ case=(nom;gen;acc),agr={ num=plu } } } }.[].

lex={ lu=zelt,lex='Zelten',head={ cat=n,ehhead={ case=dat,agr={ num=plu } } } }.[].

lex={ lu=zelt,lex='Zelts',head={ cat=n,ehhead={ case=gen,agr={ num=sing } } } }.[].

lex={ lu=zentraleinheit,lex='Zentraleinheit',head={ cat=n,ehhead={ agr={ num=sing } } } }.[].

lex={ lu=zentraleinheit,lex='Zentraleinheiten',head={ cat=n,ehhead={ agr={ num=plu } } } }.[].

lex={ lu=zentrum,lex='Zentren',head={ cat=n,ehhead={ agr={ num=plu } } } }.[].

lex={ lu=zentrum,lex='Zentrum',head={ cat=n,ehhead={ case=(nom;acc;dat),agr={ num=sing } } } }.[].

lex={ lu=zentrum,lex='Zentrums',head={ cat=n,ehhead={ case=gen,agr={ num=sing } } } }.[].

lex={ lu=zerlegen,lex='Zerlegung',head={ cat=n,ehhead={ agr={ num=sing,gen=fem } } } }.[].

lex={ lu=zerlegen,lex='Zerlegungen',head={ cat=n,ehhead={ agr={ num=plu,gen=fem } } } }.[].

lex={ lu=ziegelstein,lex='Ziegelstein',head={ cat=n,ehhead={ case=(nom;acc;dat),agr={ num=sing } } } } }.[].

Weitere Beispiele in

Heß/Brustkern/Lenders: Maschinenlesbare Deutsche Wörterbücher.  
Tübingen 1983.

Wie kann man Bedeutungsangaben in einer kanonischen oder formalen Sprache formulieren?

Wie kann man natürlich-sprachliche Bedeutungsangaben in eine kanonische Form bringen?

Dies sind zentrale Probleme der Computerlinguistik!

Drei Gruppen von Verfahren zur kanonischen semantischen Bedeutungsbeschreibung:

- a) durch semantische Merkmale, z.B. +/- human; +/- zählbar
- b) durch semantische Komponenten und Primitive, z.B. primitive Akte
- c) durch semantische Relationen in semantischen Netzen, z.B. Synonymie, Hyponymie, Antonymie

Wichtige Unterscheidung:

Wörterbuch der geschlossenen Wortklassen

Wörterbuch der offenen Wortklassen

## Zusammenfassung

Die Informationstechnik brachte neue Gesichtspunkte in die Lexikographie:

- Wörterbuch und lexikalische Datenbank
- Wörterbuch und lexikalisches Wissen
- sprachliches Wissen vs. Weltwissen, Allgemeinwissen
- Maschinelle Lexikographie als neue Methode der Herstellung von (gedruckten) Wörterbüchern
  - allgemeinsprachliche Wörterbücher
  - fachsprachliche Wörterbücher (vgl. hierzu die DIN-Bemühungen)

Probleme und Forderungen an die maschinelle Lexikographie:

- Korpusbasiertheit
- Widerspruchsfreiheit
- optimale Verweisstrukturen