

Prof. Dr. Winfried Lenders

Vorlesung WS 2005/2006

Maschinelle Lexikographie und Wissensrepräsentation

# Gliederung

## I. Lexikographie

### 1. Aufgaben und Gliederung der Lexikographie

1.1 Wer hat Interesse an Wörterbüchern?

1.2 Wörterbuch - Lexikon - Gedächtnis

1.3 Typologie der Wörterbuchbenutzung und der Wörterbücher

1.4 Lexikographie und Lexikologie

### 2. Geschichte der Lexikographie/Lexikologie

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

### 3. Lexikalische Einheiten und Strukturen

3.1 Die lexikalische Makrostruktur

3.2 Die lexikalische Mikrostruktur

3.3 Typen lexikalischer Kategorien

3.4 Onomasiologische vs. semasiologische Wörterbücher

## II. Wissensrepräsentation

### 4. Das Wörterbuch als Netzwerk

4.1 Das Konzept des semantischen Netzwerks (WordNet)

4.2 Wörterbuch - Lexikalisches Wissen und Weltwissen

4.3 Mentale Modelle und Weltwissen

4.4 Repräsentation von Wissen in semantischen Netzen

4.5 Modell eines lexikalischen Netzes

4.6 Objektorientierte Darstellung lexikalischer Netze

4.7 Eigenschaftsvererbung in lexikalischen Netzen

4.8 Die Idee des redundanzfreien Lexikons

5. Darstellung von lexikalischem Wissen

5.1 Wörterbücher in maschinellen Systemen

5.2 Wortdatenbanken

5.3 Sprachtechnologie und maschinelle Lexikographie

6. Das Wörterbuch als Ressource maschineller Systeme

6.1 Wiederverwendung lexikalischer Information

6.2 Maschinelle Analyse lexikalischer Mikrostrukturen

6.3 Aufbau von Lexika aus Texten (Korpora)

## Literatur zur Einführung

Handbuch "Wörterbücher", hg. v. Hausmann/Reichmann/Wiegand/Zgusta. Vol. 1, Berlin (de Gruyter) 1989.

Chomsky, N., Aspects of the Theory of Syntax, Cambridge, Mass., 1965.

Handke, Jürgen: The Structure of the Lexicon. Human versus Machine. Berlin: Mouton-de Gruyter, 1995.

Lutzeier, Peter Rolf (1995): Lexikologie. Ein Arbeitsbuch. Tübingen: Stauffenburg.

Ruslan Mitkov (ed.): The Oxford Handbook of Computational Linguistics. Oxford 2003. Darin u.a. die Beiträge von Hanks, Matsumoto, Vossen.

# I. Lexikographie

## 1. Grundlegende Begriffe

### 1.1 Wozu braucht man Wörterbücher?

- ! Ein- und mehrsprachige Wörterbücher als Gegenstände des täglichen Gebrauchs
- ! Wörterbuch als **Dokumentation** des Wissens
- ! Wörterbücher und das **Streben** nach Wissen
- ! Wörterbücher und das Streben nach **Normierung**

## Wissenschaftliches Interesse an Wörterbüchern:

- Interesse des **Linguisten**: Wörterbuch ist Teil der sprachlichen Kompetenz des Menschen
- Interesse **des Sprachphilosophen**: In Wörterbüchern werden die Einheiten der Erkenntnis, die letzten Einheiten, die Kategorien des Denkens angegeben (vgl. *Leibniz' Alphabetum Omnium Cogitationum Humanarum*).
- Interesse **des Semiotikers**: Wörterbücher enthalten die Inventare der Zeichensysteme.

- Interesse **des Psychologen**:
  - a) unter kognitionspsychologischem Aspekt: Wörterbücher enthalten die Einheiten des kognitiven Vermögens des Menschen;
  - b) unter entwicklungspsychologischem Aspekt: Wörterbuch (Lexikon) als Speicher des Wissens bei der Entwicklung des Denkens.
- Interesse des **Neurophysiologen**: Lexikon enthält die potentiellen Einheiten des menschlichen Gedächtnisses, Lexikon ist u.U. Modell des Gedächtnisses.
- Interesse des **Computerlinguisten**: Computerinternes Lexikon als Repräsentation des sprachlichen Wissens.

## 1.2 Wörterbuch - Lexikon - Gedächtnis

### Definitionen

Was ist ein Wörterbuch?

Was ist ein Lexikon?

DUDEN- Deutsches Universalwörterbuch, 2. Aufl. Mannheim 1989:

„Wörterbuch, das: *Nachschlagewerk, in dem die Wörter einer Sprache nach bestimmten Gesichtspunkten ausgewählt, angeordnet u. erklärt sind*: ein ein-, zweisprachiges, etymologisches W.; ein W. der deutschen Umgangssprache.“

**verwandte Wörter:**

Wortregister = Wortindex:

„Wortindex, der: *alphabetisches Verzeichnis der in der betreffenden wissenschaftlichen Arbeit untersuchten, erwähnten Wörter, Begriffe.*“

Wörterverzeichnis = Wortverzeichnis:

„Wortverzeichnis, das: *Wortindex; Vokabular (2)*“

„Vokabular (2): *Wörterverzeichnis.*“

„**Lexikon, das;** -s ...ka u. ...ken [griech. *lexikón* (biblion) = Wörterbuch, zu: *lexikós* = das Wort betreffend, zu: *léxis* = Rede, Wort, zu: *légein* =

auflesen, sammeln; reden]: **1.** *nach Stichwörtern alphabetisch geordnetes Nachschlagewerk für alle Wissensgebiete od. für ein bestimmtes Sachgebiet*: ein L. in fünfzehn Bänden; ein L. der Kunstwissenschaft; L. für Wirtschaft und Handel; L. zur Sexualität; Ü er ist ein wandelndes/lebendes L. (ugs. scherh.; *weiß auf allen Gebieten Bescheid*). **2.** (veraltet) *Wörterbuch*. **3.** (Sprachw.) **a)** *Gesamtheit der selbständigen bedeutungstragenden Einheiten einer Sprache; Wortschatz im Unterschied zur Grammatik einer Sprache*; **b)** (in der generativen Grammatik) *Sammlung der Lexikoneinträge einer Sprache*;

Die Erklärung Lexikon (1) nach DUW legt eine Verbindung zwischen *Lexikon* und *Enzyklopädie* nahe:

Nach Brockhaus wird *Enzyklopädie* wie folgt definiert:

- 1) das Wissenschaftssystem. Die Frage nach dem systematischen Zusammenhang der Wissenschaften und Künste ...
- 2) ein umfassendes Werk, das den Zusammenhang des Wissens seiner Zeit oder ein Teilgebiet der Wissenschaft sammelt, einheitlich ordnet und übersichtlich darstellt, entweder systematisch, d.h. nach Staaten oder Wissenschaften gegliedert, oder in Stichwörtern nach dem Abc ...

Definition von Lexikon nach einem Sprachwissenschaftlichen Fachwörterbuch: Linguistisches Wörterbuch von Th. Lewandoski, Bd. 2, Heidelberg 1975 (UTB 201):

- 1) Die Gesamtheit der Wörter bzw. der Wortschatz einer (natürlichen) Sprache; das Zeichenrepertoire als das internalisierte Wissen des Sprachteilhabers von den lexikalischen Eigenschaften der Wörter/-Lexeme (phonologisch-phonetisch-orthographische, syntaktische und semantische Informationen).
- 2) Inventar bzw. Liste/Zusammenstellung der zu einer Zeit geltenden Wörter einer natürlichen Sprache, eines Dialekts, Soziolekts, einer Fachsprache oder Problembereichs (auch: Wörterbuch).
- 3a) In der generativen Transformationsgrammatik ... die Subkomponente der semantischen Komponente ...
- 3b) Die Subkomponente der Basis ...
- 3c) In der generativen Semantik ...

Aus diesen Definitionen lassen sich drei Hauptaufgaben der Lexika ableiten:

- 1) Das Wörterbuch als das sprachliche Wissen des Individuums in seiner Eigenschaft als Sprachteilhaber.
- 2) Das Wörterbuch als Inventar der Einheiten einer Sprache (strukturalistisch); Aspekt der Dokumentation, zugleich als Normbuch für richtiges Sprechen und Schreiben
  - Orthographie - Bedeutung - Stil
  - Übersetzung (Typen; vgl. Hausmann)
- 3) Das Wörterbuch in der linguistischen Theorie als Teil des Apparats, der vom Menschen zum Generieren und Verstehen von Äußerungen eingesetzt wird (generativ/generativistischer Aspekt).

Unterschied zur der Erklärung aus DUW:

Bei Lewandowski ist *Lexikon* auf **sprachliches Wissen** eingegrenzt.

Gemäß der Definition ist sprachliches Wissen

- phonetisch-phonologisch-orthographische
- syntaktische
- semantische Information.

## Zusammenhang zwischen Lexikon und Gedächtnis:

- Lexikon ist das internalisierte **Sprachwissen**.
- Gedächtnis ist der Ort dieses Wissens.
- Sprachwissen ist nur ein Teil des im Gedächtnis enthaltenen Wissens.
- Gedächtnis ist Teil eines übergeordneten Organismus, eines Systems:
  - künstliche Systeme - natürliche Systeme
  - andere Teilstrukturen: Perzeptoren (Wahrnehmung), Artikulation
  - Motivationale Strukturen, Problemlösungsstrukturen
- Gedächtnisse speichern langfristig, aber ‚dynamisch‘
- Zustand eines Systems - Prozess eines Systems

Mit dem Begriff ‚Gedächtnis‘ ist der **dynamische** Aspekt, mit dem Begriff des Lexikons / Wörterbuchs der **statische** Aspekt verbunden:

	dynamisch/ organismisch	statisch/ systemisch/ symbolisch
Kompetenzaspekt	sprachliches Gedächtnis	Lexikon
Dokumentationsaspekt	lexikalisches Wissen	Wörterbuch

## Wozu dienen Wörterbücher?

- Übersetzung
- Sprachdokumentation, Deskription:  
Erfassung des Wortschatzes, Beschreibung des Wortschatzes  
(Strukturalismus; Wörterbücher im Dienste der vergleichenden  
Sprachforschung)
- Normierung
- Terminologie
- Computersimulation von Sprachlichem Handeln: Maschinelle  
Sprachübersetzung; Natürlichsprachliche Systeme; Information  
Retrieval

## 1.3 Typologie der Wörterbuchbenutzung und der Wörterbücher

vgl. Abb. 13.1 aus

**Peter Kühn:** Typologie der Wörterbücher nach Benutzungsmöglichkeiten.  
In: Handbuch Lexikographie Bd. I, Berlin 1989, S. 111 ff.

# EIN WÖRTERBUCH BENUTZEN

als

## NACHSCHLAGEWERK (4.1.)

## LESEBUCH (4.2.)

zur

zur

Kompetenzkontrolle  
(4.1.1.)  
Rechtschreibwb.  
Aussprachewb.  
Flexionswb.  
Schwierigkeitenwb.  
—  
Textproduzenten  
Schüler  
Sekretär(in)

Textrezeptio  
(4.1.2.)  
zur

Textproduktion  
(4.1.3.)  
betr.

Übersetzung  
(4.1.4.)  
Übersetzungswb.  
—  
Schüler  
Studenten  
Übersetzer

Fach  
(Sprachen)  
arbeit  
(4.1.5.)  
Fachwb.  
—  
Laien  
Fachleute  
Übersetzer

Forschungs-  
arbeit  
(4.1.6.)  
Index  
Konkordanz  
Autorenwb.  
Morphemwb.  
Häufigkeitswb.  
wortgeschichtl. Wb.  
Namenwb.  
—  
Wissenschaftler

Verständigungs-  
sicherung  
(4.1.2.1.)  
Abkürzungswb.  
Fremdwb.  
Dialektwb.  
Sonderspr.wb.  
Zeitungswb.  
Wb. schwerer Wörter  
Namenwb.

Interpretationsver-  
stärkung  
(4.1.2.2.)  
Wortgeschichtl.wb.  
Neologismenwb.  
Archaismenwb.  
Fachwb.  
Begriffswb.  
Konversationslexikon  
—  
„Sprachgebildete“

Paradigmatik  
Begriffswb.  
Analogiewb.  
Synonymenwb.  
Antonymenwb.

Syntagmatik  
Konstruktionswb.  
Kollokationswb.

Reihen-  
bildung  
Reimwb.  
Häufigkeitswb.  
rückl. Wb.

Textproduzenten  
—  
Medner  
Werbetexter  
Wissenschaftler  
Journalisten  
Schriftsteller  
Schüler/Studenten

Erbauung u. Belehrung  
(4.2.1.)  
etymologisches Wb.  
wortgeschichtl. Wb.  
Redensartenwb.  
Sprichwb.  
Eponymenwb.  
Zitatenwb.  
Kreuzworträtselwb.  
Dialektwb.  
Sondersprachenwb.  
—  
„Sprachinteressierte“  
„Bildungsbürger“

Spracherlernung (4.2.2.)  
Synonymenwb.  
Homonymenwb.  
Schulwb.  
Grundwortschatzwb.  
Satzlexikon  
Bildwb.  
Lern(er)wb.  
—  
Schüler  
Studenten  
erwachsene  
Fremdsprachenlerner

Abb. 13.1: Möglichkeiten der Wörterbuchbenutzung

## 1.4 Lexikographie und Lexikologie

Lexikographie hat es in theoretischer Hinsicht mit Fragen zu tun wie

- welche Arten von Informationen gehören in ein Wb?
  - wie werden sie dargestellt? Z.B. Problem der Darstellung von morphologischen Informationen, Verweisen, Tiefe und Art der Lesartenangaben;
  - wie wird Bedeutung markiert?
  - was ist Pragmatik im Wörterbuch?
  - Wie werden pragmatische Angaben notiert? Markierung von Fachgebieten; Diachronie; Frequenzen; Dialekt; Normen etc.
  - Welche Wörter sind zu lexikalisieren?
  - Lemma und Lemmatypen
  - Wie ist das Wörterbuch anzuordnen (Makrostruktur)?
  - Inwieweit enzyklopädische Angaben?

Diese theoretischen Überlegungen nennt man  
**Theorie der Lexikographie.**

Davon zu unterscheiden ist die **Lexikologie** (Definition von Lewandowski):

Die Lehre von der Erforschung des Wortschatzes bzw. des Lexikons einer Sprache, die Beschreibung seiner Struktur; die Lehre vom Wort und vom Wortbestand, vom (offenen) lexikalischen Teilsystem der Sprache, seiner Gliederung und Veränderung. Hauptgegenstand der Lexikologie ist das Wort als Bestandteil des Wortschatzes.

Die Lexikologie untersucht (nach Schippan 1972, S. 7):

- den dialektischen Zusammenhang von gesellschaftlicher Entwicklung und Veränderungen im Wortschatz,
- den Aufbau des lexikalischen Teilsystems und die Beziehungen zwischen seinen Elementen,
- die Ursachen und Bedingungen für die Anpassung der Lexik an neue kommunikative Bedürfnisse.

Als wichtige Fragen ergeben sich (nach Agricola u.a., Hrsg., 1969, S. 58):

- die Bestimmung des Wortes und seiner Grenzen,
- das Wesen der lexikalischen Bedeutung,
- die innere Struktur der Wortbedeutung.

Mit diesen Aufgaben kann man Lexikologie auch als Teilgebiet der Semantik betrachten.