

Schwarz, I. / Umstätter, W.: nfd Information - Wissenschaft und Praxis 50 (4) S. 197-203 (1999)

Die vernachlässigten Aspekte des Thesaurus: dokumentarische, pragmatische, semantische und syntaktische Einblicke.

Zusammenfassung - Summary

1. Einleitung

2. Historischer Exkurs und Thesaurusbegriff

3. Der Thesaurusbegriff in Information und Dokumentation

4. Die semiotische Auslegung des Thesaurusbegriffs

Schluss

Literatur

Zusammenfassung

Neuentwicklungen in der realen Welt ziehen Veränderungen in der Begriffswelt nach sich. So haben wir es derzeit beispielsweise nicht mehr nur mit Bibliotheken zu tun; vielmehr hat sich das Anwendungsspektrum des Begriffs in neuer Zeit so erweitert, daß wir von der elektronischen, der virtuellen und der digitalen Bibliothek sprechen, wobei diese Bezeichnungen nicht etwa identisch sind, sondern das jeweils Spezifische zu ihrer Entstehungszeit kennzeichnen. Die genaue Reflektion und Einordnung der modernen Thesaurusentwicklungen auf dem Gebiet der Information und Dokumentation zwingt ebenso zu eindeutigen begrifflichen Abgrenzungen und Benennungen. Bislang hat dies u.a. zu den Begriffen "Information Retrieval Thesaurus", "Metathesaurus", "Macrothesaurus" und nicht zuletzt "semantischer Thesaurus" geführt, die den Begriff des dokumentarischen Thesaurus jeweils spezifizieren. Mit dem Begriff des semiotischen Thesaurus, der den semantischen, den syntaktischen und insbesondere auch den pragmatischen Aspekt berücksichtigt, kennzeichnen wir nun eine neue Entwicklung, die den Thesaurus stärker auf das Begriffssystem des Endnutzers appliziert und damit einen bisher völlig vernachlässigten Bereich der Thesaurusforschung in das Zentrum der Betrachtung rückt. Angeregt von der Bezeichnung "semantischer Thesaurus" hat uns die Morris'sche Dreistrahligkeit des Zeichenbegriffs zu dieser Benennung geführt, und wir haben den Begriff des pragmatischen Thesaurus aus der semiotischen Perspektive hergeleitet.

Summary

Changing developments in the real existing world has to be accompanied by a convenient adaptation in our terminologic imaginations. Accordingly, we have not only classical libraries in this world - on the contrary, we could observe in the last years a diversification into electronic, virtual, and digital libraries. Sometimes used like synonyms these terms have shown evolutionary grown specifications which have to be distinguished. Also the modern thesaurus developments in the field of information and documentation compels us to have clear defined terms. Concepts like "information retrieval thesaurus", "metathesaurus", "macrothesaurus", and last but not least "semantic thesaurus" are established terms specifying the idea of the thesaurus in documentation. The new term semiotic thesaurus involves the semantic, the syntactic, and especially the pragmatic aspect, That makes the thesaurus especially applicable for the endusers interests. The pragmatic aspect of the thesaurus was totally neglected

till now. Stimulated by the term semantic thesaurus, at one hand, and on the other hand, by the semiotic theory of C.W. Morris, we found that it is consequent to create the idea of a pragmatic thesaurus.

1 Einleitung

Das Wort Thesaurus bedeutet, wie dokumentarisch allgemein bekannt, Schatz. Wobei wir im Wirtschaftsbereich unter *thesaurieren* das Horten von Geld oder Edelmetallen verstehen. Das Interesse eher vergeistigter Menschen, wie es Dokumentare nun einmal sind, konzentriert sich allerdings mehr auf das Horten von Wortschätzen. Solche Schätze gibt es zur Erschließung der Literatur für fast jedes Spezialgebiet und ebenso in Form von Universalthesauri für das sogenannte Weltwissen.

Damit gehört die Erarbeitung und Nutzung von Thesauri dieser Art nicht unbedingt zu den neuesten Errungenschaften der Dokumentation. Im Gegenteil: ihre Tage schienen bei zunehmendem Volltextretrieval schon gezählt. Nach Donna R. Dolan sollte bereits 1981 bei den "full text databases" kein kontrolliertes Vokabular mehr zwischen den Nutzern und effektivem Retrieval stehen.

Trotzdem beobachten wir in den letzten Jahren eine Renaissance der Thesaurusdiskussion, die allerdings zu mancher Irritation Anlaß gibt. Da findet man auf der einen Seite Bemühungen um die Wiederbelebung alter Thesauri und Klassifikationen, bis hin zur Dezimalklassifikation, während auf der anderen Seite von semantischen Thesauri die Rede ist, bei denen sich so mancher fragt, was er sich darunter vorstellen soll.

Der Begriff des "semantischen" Thesaurus hat insbesondere mit dem Unified Medical Language System (UMLS) der National Library of Medicine (NLM) seinen Einzug in die Welt der Information und Dokumentation gehalten. Während der Begriff des Thesaurus in der Dokumentation durch die DIN 1463 relativ eindeutig definiert ist, stellt sich die Frage nach der speziellen Bedeutung eines zusätzlich mit dem Attribut *semantisch* belegten Thesaurus. Diese Frage führte uns zur Semiotik bzw. Zeichentheorie, deren Untersuchungsgegenstand die drei Zeichendimensionen - die syntaktische, semantische und pragmatische - sind. Bemerkenswert ist, daß die pragmatische Dimension innerhalb der Semiotik, die nach C.W. Morris (1938) als "relation of signs to interpreters" definiert ist, bisher in so auffälliger Weise vernachlässigt wurde. Dagegen finden wir in letzter Zeit starke Bemühungen um die semantische Dimension, die sich mit den "relations of signs to the objects to which the signs are applicable" beschäftigt. Theoretisch am besten fundiert und ausgebaut ist die syntaktische Dimension, die auf der Informationstheorie von Shannon und Weaver basiert.

Umstätter (1998) hat den Begriff des *pragmatischen Thesaurus* im Kontext objektorientierter Inhaltserschließung und im Sinne von C.W. Morris eingeführt; gleichzeitig impliziert diese Bezeichnung die semiotische Auslegung des Thesaurusbegriffs, die in diesem Beitrag untersucht werden soll.

2 Historischer Exkurs und Thesaurusbegriff

Thesauri waren (und sind auch heute noch) vor allem Produkte lexikographischer Aktivitäten. Diese erreichten bereits im 4. Jh. v.Chr. bei den alexandrinischen Philologen eine erste Blüte. Zenodotus von Ephesus (~325 --234 v.Chr.), ein Schüler des Philetas und der erste Bibliothekar der alexandrinischen Bibliothek, sammelte beispielsweise in einem Glossar die schwierigen Wörter bei Homer und stellte damit ein sog. individual-sprachliches Wörterbuch zusammen (siehe Brockhaus zum Begriff des Wörterbuchs, 19.

Aufl., Bd. 24, S. 348-349). Aristophanes von Byzanz (~257 - ~180 v.Chr.), ein Schüler des Zenodotus und ebenfalls Bibliothekar in Alexandria, erstellte das sogenannte "Lexeis" (von λεγειν: reden, sagen; λεχω: lesen), von dem unser heutiges Wort Lexikon stammt. Die sogenannten Onomastica (vom griech. Ονομα: Name, Wort, bzw. Ονομαστοζ: nennbar) erschienen im 3. Jh. v.Chr. Callimachos von Cyrene (~310 - ~235 v.Chr.) sammelte Namen von Flüssen, Fischen, Winden und Monaten. Xenocritus und Philinus, beide aus Cos, stellten medizinische Begriffe zusammen.

Das "Onomasticon" von Julius Pollux von Naucratis, das auf den lexikographischen Arbeiten des Aristophanes, des Phrynichus Arabius (im 2. Jh. v.Chr.), des Pamphilos und Didymus von Chalcentera (~65 v.Chr. - ~10 n.Chr.) beruhte, enthielt die Wörter der griechischen Sprache einschließlich aller Synonyme. Es faßte zehn Klassen zusammen: Götter, Mensch (Medizin), Verwandtschaft, Wissenschaft und Kunst, Jagd, Nahrung, Handwerk, Gesetz, städtische Organisation, Gerätschaften. Darin ist ein grob gerastertes Bild der geistigen Welt von damals zu erkennen. Auch in späteren Jahrhunderten waren Klassifikationssysteme immer ein gewisses Abbild des Begrifflichen oder Ideologischen ihrer Zeit.

Laut Michajlov et al. (1970, S. 374) wurde der Begriff *Thesaurus* als Bezeichnung einer systematisierten Enzyklopädie erstmalig um 1265 von dem Florentiner Brunetto Latini (zw. 1210 und 1220-1294/95) verwendet. Er nannte seine Enzyklopädie "Li Livres dou Tresor" (etymologisch: *Thesaurus* - *treasure* - *Tresor*). Diese erste Laien-Enzyklopädie von Rang (Brockhaus, 19. Aufl., Bd. 6, S. 451) kennzeichnete den Wandel der mittelalterlichen Enzyklopädie vom 'Weltbuch' zum Thesaurus. Ein weiterer Beleg für die interessante Assoziation von weltlichem und geistigem Schatz ist die Inschrift über dem Rokokoportal der Bibliothek des ehemaligen Klosters Wiblingen bei Ulm, wo man 1762 die Worte aus dem Brief des Paulus an die Kolosser 2,3 angebracht hat: "In quo omnes thesauri sapientiae et scientiae" (In welchem alle Schätze der Weisheit und Wissenschaft enthalten sind).

Als *Thesaurus* wurden insbesondere im Mittelalter umfangreiche Werke betitelt, die möglichst vollständig ganze Wissensgebiete umfaßten. Am berühmtesten sind die diesen Titel führenden lexikographischen Werke der beiden Stephanus - Vater und Sohn Estienne - über die griechische und lateinische Sprache geworden (Bilder-Conversations-Lexikon, Bd. 4, S. 288, 413): 1531 veröffentlichte Robert Estienne (z.T. auch Etienne; 1503-1559), ein bedeutender französischer Buchdrucker seiner Zeit und "Vater der modernen Klassischen Lateinischen und Französischen Lexikographie", seinen "Latinae linguae Thesaurus" (Brockhaus, 19. Aufl., Bd. 6, S. 595; Estienne: *Dictionarium latinogallicum*, 1997). Dabei handelt es sich um ein etymologisches Wörterbuch der lateinischen Sprache, das die Stichwörter in alphabetischer Reihenfolge anordnet. Ein anderes Werk Estienne's - das "*Dictionarium latinogallicum*" - ist in der 3. Ausgabe von 1552 auch im Internet recherchierbar (ARTFL dictionary collection, 1997). Estienne's Sohn Henri II. (1528-1598) publizierte 1572 das 4-bändige etymologische Wörterbuch "*Thesaurus linguae graecae*" (Schatzkammer der griechischen Sprache), das seine Bedeutung als erstes umfassendes Wörterbuch des Altgriechischen bis heute nicht verloren hat und seit 1955 in neuer Form (hrsg. v. Snell et al.; Brockhaus, 19. Aufl., Bd. 22, S. 99) erscheint.

Eine völlig neue lexikographische Aktivität setzte im 18. Jh. bei den sogenannten Enzyklopädisten ein, die darunter ein umfassendes Bildungsinstrument und ein auf Vernunft gegründetes Kompendium verstanden, in dem das gesamte Wissen ihrer Zeit zusammengetragen sein sollte. Es war der groß angelegte aufklärerische Versuch, Unwissen, Torheit und Aberglauben in der Welt zu bekämpfen und damit auch den Beginn der Wissenschaftsgesellschaft von morgen vorzubereiten. Ein Versuch, der von der damaligen Kirche bekanntlich

mit großer Skepsis betrachtet wurde, dessen fundamentale Bedeutung aber kaum überschätzt werden kann. Berühmt wurde die zwischen 1751 und 1772 entstandene 28-bändige Enzyklopädie des Philosophen Denis Diderot, des Mathematikers Jean Le Rond D'Alembert und vieler anderer, die in der Société des Gens de lettres tätig wurden. Diese ursprüngliche Fassung enthält rd. 72.000 Artikel, wobei Mehrfacheinträge für viele Bezeichnungen enthalten sind, die aus ihrer Zuordnung zu mehreren Wissenskategorien - den sogenannten Siglen - resultieren. Darüber hinaus wird mittels eines Verweisungssystems verdeutlicht, wie die einzelnen Artikel miteinander verflochten sind, worauf Diderot besonderen Wert gelegt haben soll. (Brockhaus, 19. Aufl., Bd. 6, S. 453). Im Rahmen des "Project for American and French Research on the Treasury of the French Language" (ARTFL) wird die Online-Version der 1. Ausgabe dieser Enzyklopädie zur Zeit erarbeitet. Volume 1 ist bereits im Internet recherchierbar (ARTFL Project, University of Chicago, 1998).

Im 19. Jahrhundert folgten eine Vielzahl an populärwissenschaftlichen Konversationslexika, Nachschlagewerken und Lehr- bzw. Handbüchern auf den einzelnen sich ausweitenden Fachgebieten, die eine weitere Verbreitung der Wissenschaft - nun auch in der allgemeinen Bevölkerung der Industriestaaten - ermöglichten.

Als wohl bekanntester Thesaurus neuerer Zeit, der nach verwandten Begriffen geordnet ist, gilt "Roget's International Thesaurus", der als "Thesaurus of English Words and Phrases" 1852 von dem Arzt und Sekretär der Royal Society in London, Peter Mark Roget, publiziert wurde. Dieser Thesaurus ist, wie es im Vorwort heißt, "classified and arranged so as to facilitate the expression of ideas and assist in literature composition". Er ist seither mehr als 90mal von den Nachkommen des Verfassers neu aufgelegt worden und derzeit mit seinen rund tausend Kategorien auch im Internet recherchierbar (ROGET'S Thesaurus Search Form, 1997). In seinem eigenen Vorwort von 1852 schreibt Roget: "I had, in the year 1805, completed a classed catalog of words on a small scale, but on the same principle, and nearly in the same form, as the Thesaurus now published." Etwa zur gleichen Zeit erschien auch der Thesaurus des Sanskrit ("Amera Sinha, Cosha, or dictionary of the Sanskrit Language"), der 1808 veröffentlicht wurde und die Wörter zu folgenden Hauptklassen - von denen jede wiederum Unterklassen enthielt - zusammenfaßte: Himmel, Wetter, Zeit, Tugenden, Laster, Wissen, Gefühle, Sprache, Musik, Hölle, Wasser, Erde, Pflanzen, Tiere, Mensch, Familie, Medizin, Kleidung, Rasse, Stamm, Regierung, Armee, Berufe, Waren, Geräte (Sparck Jones, 1964, Anhang 1, S. 2).

Im deutschsprachigen Raum verbindet sich laut Scerba (1940, S. 69) der Begriff des Thesaurus Anfang des 20. Jh. zumeist mit dem "Thesaurus linguae latinae", einem Unternehmen, das bereits 1897 begründet wurde und seit 1900 herausgegeben wird. Es ist an der Bayerischen Akademie der Wissenschaften angesiedelt und besteht in der "Erarbeitung des ersten umfassenden wissenschaftlichen Wörterbuchs der antiken Latinität von den Anfängen bis 600 n.Chr. Die Grundlage dafür bildet ein Archiv (etwa 10 Millionen Zettel), das alle überlieferten Texte berücksichtigt - die älteren sind Wort für Wort verzettelt, aus den jüngeren ist in der Regel nur das lexikographisch Relevante exzerpiert" (Kommission, 1998). War man 1940 erst beim Buchstaben M angelangt (Scerba, 1940, S. 69), so liegen inzwischen zwei Drittel des Werkes vor (Kommission, 1998). Eine Besonderheit dieses Thesaurus besteht darin, daß er auch alle sogenannten Hapaxlegomena enthält, also Wörter, die in der betreffenden Sprache nur einmal aufgetreten sind, und daß unter jedem Wort sämtliche Zitate aus den in der betreffenden Sprache vorhandenen Texten angeführt werden.

Historisch betrachtet wurde der Begriff des Thesaurus also verwendet zur Benennung von Wortsammlungen, die

a. die Bedeutung von Wörtern bzw. deren Begriffswelten durch Beispiele aus ihrer konkreten Verwendung erklären (wie beispielsweise die Thesauri der

Estienne oder der "Thesaurus linguae latinae"), was unter semiotischen Gesichtspunkten ein pragmatisches Verfahren darstellt.

b. die Bedeutung von Wörtern bzw. Begriffswelten per definitionem oder über eine Wortfeldbegrenzung festlegen (wie beispielsweise das Onomastikon oder Roget's International Thesaurus); eine in semiotischem Sinne eher semantische Realisation.

Eine weitere Unterscheidung ergibt sich unter dem Aspekt der "Rekrutierung" des Vokabulars:

a. Thesauri für die natürliche Sprache, die mit einer großen Unschärfe und zeitweiligen Widersprüchlichkeit behaftet ist;

b. Thesauri für artifizielle oder quasi-artifizielle Sprachen (z.B. Recherchesysteme, Fachsprachen oder Wissensbasen), die eine möglichst klar definierte Terminologie zur Vermeidung von Widersprüchlichkeiten bzw. Ambiguitäten anstreben. Sie enthalten oft systematische Zeichensysteme bzw. Codierungen zur Relationierung und Hierarchiebildung.

Damit verstehen wir den Begriff *Thesaurus* im erweiterten Sinne als thematisch geordnetes Zeichensystem und im engeren Sinne als thematisch geordneten Wortschatz, aus dem sich die Bedeutung der jeweiligen Wortfelder syntaktisch ergibt. Er ist je nach Funktion ein Thesaurus:

- der natürlichen Sprache,
- der Fachtermini eines Spezialgebietes,
- der normierten Sprache einer Wissensbasis, der normierten Sprache eines Dokumentationssystems zur Literaturschließung,
- zur Organisation von materiellen und immateriellen Objekten (Semantik),
- zur Organisation von empfangenen Zeichenketten mit dem Ziel der Rekonstruktion semantischer Netze (Pragmatik).

3 Der Thesaurusbegriff in der Information und Dokumentation

Die Geschichte der im IuD-Bereich gebräuchlichen Thesauri läßt sich im Kern auf die Mitte unseres Jahrhunderts datieren, als sich zunehmend die Ansicht durchsetzte, daß konventionelle Retrievalsysteme die aufgrund des Wachstums von Wissenschaft und Technologie steigenden Informationsbedürfnisse nicht mehr befriedigen konnten (Roberts, 1984). Dieser Druck führte in den 50er und 60er Jahren sowohl dazu, verbesserte Retrievalsysteme zu schaffen, als auch dazu, neue Indexierungsverfahren - z.B. im Rahmen der Cranfield-Tests - zu erproben. Der erste "Information Retrieval Thesaurus" wurde laut Roberts (1984) am Engineering Department der E.I. Du Pont Nemours and Co., Inc. im Jahre 1959 praktisch angewendet.

Der Thesaurus war jedoch schon wesentlich früher, nämlich bereits 1878, als potentiell nützliche Hilfe bei der Informationssuche angesehen worden. W.F. Poole vertrat damals im Rahmen einer Diskussion zu Poole's Index die Ansicht, daß zusätzlich zum Index *cross references* als eigenständige Publikation erscheinen sollten, ähnlich denen in Roget's Thesaurus (Poole, 1878, S. 110). Eine solche separate Publikation würde einen breiteren Zugriff auf die Sachgebietsklassen des Indexes bieten, indem sie ein umfangreicheres, in Beziehung zueinander stehendes (*inter-connected*) Vokabular enthielte. Obwohl die Poole'sche Idee nicht praktisch umgesetzt wurde, nahm sie bestimmte Funktionen des ersten, am Du Pont erarbeiteten Thesaurus vorweg (Roberts, 1984, S. 282).

Im Februar 1947 wies der amerikanische Mathematiker Calvin Northrup Mooers

(1919-1994) auf die Nutzung eines Thesaurus in Verbindung mit einem mechanisierten Retrievalsystem hin (Roberts, 1984). Von ihm stammen auch die Bezeichnungen *descriptor* und *information retrieval*, wobei er letztere schon während seiner Studienzeit am Massachusetts Institute of Technology prägte (Garfield, 1997). Bereits bei Mooers war der Deskriptor ein Wort oder eine Wortverbindung zur Bezeichnung eines mehr oder weniger breiten Begriffs, wobei dessen Bedeutung nicht absolut der Bedeutung entsprechen mußte, in der das betreffende Wort im natürlichen Sprachgebrauch benutzt wird. In jedem Falle sollte jeder Deskriptor mit einer erläuternden Bemerkung versehen werden, aus der seine thesauruspezifische Bedeutung hervorging. Die Auswahl der Bedeutungen der Deskriptoren sollte eine genügend hohe Effektivität bei der Recherche sichern. Es sei hier am Rande bemerkt, daß Mooers damit von der irreführenden Vorstellung ausging, daß Indexierer mit Hilfe der Deskriptoren Dokumente beschreiben, obwohl sie diese im eigentlichen Sinne indizieren - d.h. sie weisen mit Hilfe der Thesaurusbegriffe darauf hin, in welchem gedanklichen Kontext die Dokumente relevant sein müßten.

Durch die vom DIN 1976 erstmals herausgegebene und 1987 novellierte Norm erhielt der Thesaurus im Bereich der Information und Dokumentation endgültig die Bedeutung einer "geordneten Zusammenstellung von Begriffen und ihren (vorwiegend natürlichsprachigen) Bezeichnungen, die in einem Dokumentationsgebiet zum Indexieren, Speichern und Wiederauffinden dient" (DIN 1463, Teil 1) - also das, was wir oben als Thesaurus im engeren Sinne bezeichnet haben.

Unter lexikographischen Gesichtspunkten ist der Thesaurus

- ein Synonymwörterbuch, indem er unterschiedliche Bezeichnungen mit übereinstimmenden Definitionen und Quasisynonyme mit sich ähnelnden Definitionen in Äquivalenzklassen zusammenfaßt und für diese Vorzugsbenennungen - die sog. Deskriptoren - festlegt. Damit ist er auch ein
- normatives bzw. präskriptives Wörterbuch, denn mit den Deskriptoren schreibt er das für die Indexierung und die Informationsrecherche zu verwendende Vokabular vor. Ebenso ist er
- ein zweisprachiges Wörterbuch, in dem die sogenannten Nichtdeskriptoren der natürlichen Sprache in Deskriptoren zusammengefaßt und "übersetzt" werden. Die Bedeutung der Deskriptoren kann im Thesaurus verändert sein, und zwar ist sie in der Regel breiter als die der entsprechenden Wörter und Wortverbindungen. Damit ist er zugleich
- eine normierte künstliche Sprache, insofern die Deskriptoren eine eigene, ihrer Funktion angemessene, Bedeutung gegenüber den entsprechenden Wörtern oder Wortverbindungen in der natürlichen Sprache besitzen, und er ist
- ein thematisch nach der Bedeutung der Deskriptoren geordneter Wortschatz. Dadurch ist er
- ein Bedeutungswörterbuch, indem er die geistigen Inhalte der Deskriptoren durch ihren Kontext bzw. ihr Wortfeld, den sog. Begriffssatz, wiedergibt. In diesem Sinne werden Deskriptoren von der Syntax in ihrer Bedeutung beschrieben. Mit anderen Worten: Thesauri definieren im Sinne einer Wortfeldbegrenzung die Begrifflichkeit, d.h. die Bedeutung einer Benennung, indem sie alle Nichtdeskriptoren ihren Deskriptoren zuordnen und letztere zueinander in eine syntaktische Relation setzen. Zur Ergänzung dieser Definition verdeutlichen sie den Inhalt der Deskriptoren durch sogenannte scope notes oder auch durch exaktere Begriffsbestimmungen.

Während der Zugang zum Wortmaterial in den frühen gedruckten Thesauri

aufgrund der Anordnung des Wortschatzes und der Anlage von Registern bestimmten Einschränkungen unterlag, ergeben sich heute, in Zeiten der invertierten Volltextspeicherung und -recherche wie auch der höchst differenzierten und "intelligenten" Visualisierung der Relationen in einem Thesaurus, derartige Beschränkungen nicht mehr. Dokumentarische Thesauri erlauben sowohl den Zugriff über Benennungen als auch über Begriffe, wie im nächsten Abschnitt zu erläutern sein wird. Ebenso unterliegt der Umfang der Wortsammlung nicht mehr den restriktiven Bedingungen, wie sie zu den Lochkartenzeiten Mooers oder zu Zeiten stark begrenzter elektronischer Speicherkapazitäten gegeben waren.

4 Die semiotische Auslegung des Thesaurusbegriffs

Die Semantik ist eine Teildisziplin der Zeichentheorie bzw. Semiotik, die als "Lehre von den Zeichen" insbesondere deren Funktion in der Kommunikation (bzw. im Kommunikationsprozeß) untersucht. Die heutige Fassung des Begriffs *Semantik* geht insofern auf C.W. Morris (1938) zurück, als er in seiner "Foundation of the theory of signs" drei unterschiedliche Dimensionen des Zeichenbegriffs begründet hat: die Semantik als Verhältnis zwischen Zeichen und dem Bezeichneten, die Syntax als Verhältnis der Zeichen zueinander, sowie die Pragmatik als Verhältnis zwischen Zeichen und Zeichenbenutzer. Zeichen werden in der Semantik für etwas gesetzt (für materielle bzw. immaterielle Objekte), in der Kommunikation in eine Syntax gebracht, die in gewisser Weise die Relation der Objekte widerspiegelt, und sie werden in der Pragmatik interpretiert, soweit Ambiguitäten bestehen. Morris entwickelte seine Theorie im Anschluß an C.S. Peirce, der in erster Linie die "Zeichensprache", genauer die "Sprache" schriftlich fixierter Zeichen, untersucht hatte. Unter dem wichtigen Gesichtspunkt, daß es nur möglich ist, in "Zeichen zu denken" (Peirce, 1967, S. 186), unterscheidet er hinsichtlich ihres Charakters, also hinsichtlich dessen, wofür sie stehen oder wie man sie interpretieren könnte, drei Arten von Zeichen nach ihrer Beziehung zu Objekten:

- 1) den Index, der auf etwas Übergeordnetes verweist (auf ein anderes Stichwort, Abzeichen; also auf vorgegebene Wörter bzw. Zeichen);
- 2) die Icons, die eine bildliche Entsprechung zum Objekt geben (Koffer und Schirm im Schild der Gepäckaufgabe; also etwas Assoziatives, wie es insbesondere bei den objektorientierten Benutzeroberflächen von Computerprogrammen der Fall ist);
- 3) das Symbol, welches einen Bereich ohne trivial anschauliche Entsprechung vertritt (eine Zifferncodierung, Rot für Liebe; Justitia für Gerechtigkeit; also etwas im übertragenen Sinne).

Zeitlich noch vor Peirce und Morris beschäftigten sich C.K. Ogden und I.A. Richards (1923) mit dem Begriff der Bedeutung. Das von ihnen eingeführte semantische oder besser semiotische Dreieck war ein früher Versuch, die Relation zwischen Symbol (das nicht in dem von C.S. Peirce präzisierten Sinne zu verstehen ist) und Bezugsobjekt, zwischen Signans (dem Bezeichnenden, das als Wort etwas symbolisiert) und dem Signatum (dem Bezeichneten, das sich auf ein Objekt bezieht) zu verdeutlichen. Dabei handelt es sich um eine indirekte Beziehung, die sich der Vermittlung eines Gedankens verdankt. Die Zuordnung von Bedeutung ist subjektiv und setzt die Fähigkeit zu sinngemäßem Verstehen voraus, wobei sich dieses über den "Merkmalsumfang" des Bezeichneten (Größe, Format, Rang, Wert usw.) vermittelt (Brockhaus, 19. Aufl., Bd. 3, S. 6).

Wenn in diesem Zusammenhang von einer gewissen Symmetrie (recipocal and reversible) (Ullmann, 1962, S. 55-57, 63ff) gesprochen wird, so bezieht sich diese darauf, daß man das semiotische Dreieck von beiden Seiten (Signatum -

Begriff - Signans bzw. Signans - Begriff - Signatum) her betrachten kann. Im ersten Fall sprechen wir von der Onomasiologie, die im eigentlichen Sinne eine Semantik ist, im zweiten von der Semasiologie, die der später von Morris verdeutlichten Pragmatik entspricht. Die hier angesprochene Reversibilität bezieht sich somit auf die Rekonstruktion auf der pragmatischen Seite, die den Vorgang der Semantik gewissermaßen rückgängig zu machen versucht.

Aus diesem Grunde kann die Suche einer Bedeutung entweder von der Benennung (Bezeichnung) oder vom Sinn bzw. Begriff her begonnen werden. Für die Suche von Bedeutungen von Benennungen werden gewöhnliche alphabetische Wörterbücher verwendet, während zur Suche von Bedeutungen nach deren Sinn sogenannte Begriffs- und auch Synonymwörterbücher benutzt werden. Die Leistung von Begriffswörterbüchern (wie zum Beispiel Roget's International Thesaurus) besteht darin, daß sie die Wahl der für die exakte Darstellung eines bestimmten Gedankens geeignetsten Wörter erleichtern (Michajlov et al., 1970, S. 373f.).

Das Unified Medical Language System wird von seinen Erarbeitern als semantischer Metathesaurus bezeichnet (vgl. Unified Medical Language System (UMLS), 1996). Es stellt kein vereinheitlichtes Begriffssystem her, wie es Macrothesauri wie etwa der Standard-Thesaurus Wirtschaft leisten (vgl. Stock, 1999), sondern bildet verschiedene Begriffssysteme bzw. -welten aufeinander ab. Diese Abbildung geht von den Bedeutungen der verwendeten Benennungen aus und wird über semantische Netze realisiert.

Semantische Thesauri schaffen eine Beziehung, man könnte auch von einer Syntax sprechen, zwischen den Objekten, die sie aus deren Bedeutung ableiten und in Form von Zeichen und ihren Relationen abbilden. Viele dieser Beziehungen sind hierarchisch eindeutig darstellbar. So besteht ein Buch aus einem Umschlag (Einband) und aus Seiten. Andere Betrachtungsweisen lassen weitere hierarchische Einteilungen zu, die nicht immer so eindeutig sein müssen. So kann ein Buch von einem oder mehreren Autoren, aus einem oder mehreren gegliederten Kapiteln, aus Seiten mit Text, Tabellen oder Bildern bestehen, es kann Informationen und Redundanzen enthalten oder auch farbig bzw. nichtfarbig sein. Neben mono- und polyhierarchischen Einteilungen sind heute auch logische oder funktionale Zusammenhänge in solchen semantischen Thesauri realisierbar. Die Relationen im Thesaurus sind auf verschiedene Art und Weise syntaktisch abbildbar. Gängig sind sogenannte semantischen Netze, Frame-Slot-Strukturen, Graphen oder auch neuronale Darstellungen. Sie sind für jeweils unterschiedliche Problemdarstellungen und Objektwelten geeignet.

Gerade weil jede dieser Beziehungen immer auch eine Syntax enthält, lassen sie sich durch die informationstheoretisch begründete Kommunikation in Bit übermitteln.

Schluß

Die Information und Dokumentation setzte die technischen Möglichkeiten ihrer Zeit schon immer stärker und schneller als das Bibliothekswesen um. Dies galt für die Dezimalklassifikation, die ein typisches Produkt der Lochkartenzeit am Beginn unseres Jahrhunderts war, für den Einsatz der Mikroverfilmung und der Reprografie, die ihre Blüte in der ersten Hälfte unseres Jahrhunderts hatte und ebenso für die sogenannte Online-Dokumentation, die ein Produkt der Informatik seit Mitte unseres Jahrhunderts ist. Wir gehen nun dazu über, dem Computer ein Verständnis für die Inhalte von Dokumenten zu vermitteln. Dies gelingt nur über moderne

semiotische Thesauri, in denen Worte und andere Bezeichnungen durch ihre Syntax und dabei insbesondere durch ihre hierarchische Organisation eine Begrifflichkeit erhalten, die auch für Computer verständlich ist.

Wir haben hier versucht zu zeigen, daß die scheinbare Renaissance von Thesauri in Wirklichkeit eine Weiterentwicklung in Richtung der semiotischen Thesauriist, weil sie mit Hilfe ihrer Relationen, die im Prinzip eine eigene Syntax darstellen, Bezeichnungen einer Begrifflichkeit zuordnen, die auf der Senderseite einen semantischen Aspekt haben und auf der des Empfängers den pragmatischen.

Damit haben wir es bei den Klassifikationen und Thesauri nicht mit einer Wiedergeburt sondern mit einer konsequenten Fortentwicklung der Organisation von Wissen in Bibliotheken und Dokumentationen zu tun. Aus den früheren auf Monohierarchie zielenden Klassifikationen des Bibliothekswesens (zur systematischen Aufstellung von Büchern) entwickelten sich zunächst die weitaus feiner gegliederten Thesauri der Dokumentation, zur Erschließung von nichtselbstständigen Publikationen. Diese Art von Publikationen enthielten bereits weitaus begrenztere Wissens Elemente als die Bücher. Insbesondere in den Naturwissenschaften erkennt man eine starke Vereinheitlichung, die daraus besteht, daß nach einer Einleitung mit kurzer Problemdarstellung, Material und Methode zur Problemlösung genannt werden, dann die Ergebnisse erscheinen, die in der Diskussion auch theoretisch begründet und mit Ergebnissen anderer Autoren vernetzt werden. Die Wissens Elemente bestehen damit im wesentlichen aus einer Information und ihrer Begründung. Die nun folgenden semiotischen Thesauri gehen daher noch einen Schritt weiter, indem sie noch schärfer umrissene Wissens Elemente in Wissensbanken auch den Computern "verständlich" machen. Diese Wissens Elemente können so knapp gehalten werden, weil sie durch den semiotischen Thesarus, in einer völlig neuen (intelligenten) Vernetzung in logische Relationen gebracht werden.

Auch wenn Thesauri und die meist gröber strukturierten Klassifikationssysteme, die zur Erschließung von Literatur dienen und dienen, in Zukunft weiterhin gebraucht werden, so ist die bibliothekarisch-dokumentarische Herausforderung unserer Zeit eindeutig im Bereich der semiotischen Thesauri zu sehn. In ihnen wird es insbesondere darauf ankommen, Wissens Elemente zu erfassen und sie einerseits den informationsverarbeitenden Maschinen "verständlich" zu machen, um sie andererseits in den Stand zu setzen, Endnutzern sinnvolle Antworten auf Ihre Fragen zu geben.

Literatur:

Amera Sinha, Cosha, or dictionary of the Sanskrit Language (1808). Serampoor.

ARTFL Project, University of Chicago (1998). Unter <http://humanities.uchicago.edu/ARTFL/>.

ARTFL dictionary collection (1997): a collection of French dictionaries from the 16th to the 19th centuries. Unter <http://humanities.uchicago.edu/ARTFL/projects/dicos.html>.

Bilder-Conversations-Lexikon, in 4 Bd. München [Nachdr.].

Brockhaus-Enzyklopädie: in 24 Bd. 19., völlig Neubearb. Aufl. Mannheim : Brockhaus.

Dolan, D. R. (1981). Database for the professor and the plebe. In Database 4, (3), 83-84.

Encyclopedie : Overview of the ARTFL Encyclopédie (1997). Unter <http://humanities.uchicago.edu/ARTFL/projects/encyc/overview.html>.

Estienne: Dictionarium latinogallicum (1997). Unter <http://humanities.uchicago.edu/ARTFL/projects/Estienne/>.

Garfield, E. (1997). A Tribute To Calvin N. Mooers, A Pioneer Of Information Retrieval. Unter http://www.the-scientist.lib.upenn.edu/yr1997/mar/comm_970317.html.

Kommission für die Herausgabe des Thesaurus linguae Latinae (1998). Unter <http://www.badw.de/akad4.htm>.

Michajlov, A.I., Cernyj, A.I. & Giljarevskij, R.S. (1970). Informatik : Grundlagen. Berlin.

Mooers, C.N. (1951). Zatocoding applied to mechanical organization of knowledge. In Amer. Doc. 1, 20-32.

Morris, C.W. (1938): Foundation of the theory of signs. Chicago.

Ogden, C.K. & Richards, I.A. (1923). The meaning of meaning.

Peirce, C. S. (1967). Einige Konsequenzen aus vier Unvermögen. In C. S. Peirce: Schriften I: Zur Entstehung des Pragmatismus. Frankfurt am Main.

Poole, W.F. (1878). The plan of the new "Poole's Index". In Library Journal 3, (3), 109-110.

Roberts, N. (1984). Historical studies in documentation. The pre-history of the information retrieval thesaurus. In Journal of documentation, 40, (4), 271-285.

Roget, P.M. (1852). Thesaurus of English Words and Phrases classified and arranged so as to facilitate the expression of ideas and to assist in literacy composition. London.

Roget, P.M. (1960). Thesaurus of English Words and Phrases. Abridged ed. Hamonsworth.

ROGET'S Thesaurus Search Form (1997). Unter http://humanities.uchicago.edu/forms_unrest/ROGET.html.

Scerba (1940). Erfahrungen der allgemeinen Theorie der Lexikographie (zit. nach Michajlov, S. 374).

Sparck Jones, K. (1964). Synonymy and semantic classification. Cambridge Language Research Unit. Cambridge.

Stock, M. (1999). Standard-Thesaurus Wirtschaft : Ein neuer Standard der Wirtschaftsinformation? In Password, (1), 22-29.

Umstätter, W. (1998). Das Digitale Lehr- und Handbuch der Bibliothekswissenschaft : (DHB)-Einleitung. Unter: <http://hub.ib.hu-berlin.de/~wumsta/dhb1.html>.

Unified Medical Language System (UMLS) (1996) : Current Bibliographies in Medicine 96-8. Unter <http://www.nlm.nih.gov/pubs/cbm/umlscbm.html>.

Wersig, G. (1985): Thesaurus-Leitfaden : Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. 2., erg. Aufl. unter Mitarbeit von P. Schuck-Wersig. München [u.a.].

Ullmann, S. (1962). Semantics. An introduction to the science of meaning. Oxford.

copyright: W. Umstätter

change: Ben Kaden 14.08.2002 [Ben Kaden \(contextur@aol.com\)](mailto:Ben.Kaden@aol.com)