

# Natursprache versus Indexsprache in der Chemie-Dokumentation

Von Robert Fugmann\*

Professor Klaus Weissermel zum 60. Geburtstag gewidmet

Bei jeder gezielten Literatursuche muß man *vorhersehen* können, wie die fraglichen Begriffe im Speicher ausgedrückt sein könnten. Unter diesen Ausdrucksweisen schlägt man im Register nach, oder man benutzt sie als Suchbedingungen in einer mechanisierten Recherche. Bei *Allgemeinbegriffen* mangelt es der natursprachlichen, frei gewählten Ausdrucksweise an dieser Voraussehbarkeit, wie am Beispiel von Verbindungsklassen und Reaktionen gezeigt wird. Das Suchen in Speichern mit frei gewählter Ausdrucksweise ist demzufolge zwangsläufig mit Verlust verbunden, unabhängig davon, wie hoch die eigentliche Suchtechnik, z. B. das Computerprogramm, entwickelt sein mag. Durch eine Einschränkung der Ausdrucksmöglichkeiten, wie sie durch eine Indexsprache und ihre zuverlässige Anwendung erreicht wird, lassen sich die Voraussehbarkeit steigern und Informationsverluste entsprechend drastisch reduzieren. – Im Gegensatz hierzu sind die Ausdrucksweisen für *Individualbegriffe* (z. B. Einzelverbindungen, Autoren, Institutionen) in frei gewählter, natürlicher Sprache oft genügend gut voraussehbar. Der Nutzen einer Übersetzung in eine Indexsprache ist hier manchmal fraglich. Das gilt besonders dann, wenn die Indexsprache (noch) keinen treffenden Ausdruck zur Wiedergabe eines Fachbegriffes anbietet, sowie auch dann, wenn die natursprachliche Ausdrucksweise vieldeutig ist. Hier ist es ratsam, diese Ausdrucksweise in den Speicher zu übernehmen, zumindest zusätzlich zu einer versuchten indexsprachlichen Form. – Die Schwächen einer Einspeicherung in frei gewählter, natürlicher Sprache werden während des Wachstums eines Informationssystems erst relativ spät deutlich. Im Frühstadium dieses Wachstums kann Voraussehbarkeit zumindest teilweise noch durch Gedächtnisleistung ersetzt werden. Außerdem ist in einer so kurzen Zeitspanne die Ausdrucksweise der Autoren noch relativ einheitlich. Deswegen ist der Verlust an einschlägiger Information anfänglich gering, und er wird auch meistens nicht als eine immer stärker werdende Quelle von Recherchenfehlern erkannt. Auch tritt in diesem Stadium noch nicht der ständig wachsende Zeitbedarf für das Ersinnen und Formulieren von immer mehr alternativen Suchbedingungen in Erscheinung. Aus diesen Gründen sind natursprachliche Systeme oft ohne ausreichende Rechtfertigung bevorzugt worden. – Die charakteristischen Stärken und Schwächen beider Sprachenarten legen es nahe, daß man sie miteinander kombiniert, so daß ihre Vorzüge erhalten bleiben.

## 1. Einführung

Wenn man sich auf die Suche nach Fachliteratur begibt, in der ein bestimmtes Thema abgehandelt ist, dann muß man sich ein Schlagwort für dieses Thema überlegen oder mehrere von ihnen. Man hofft, unter einem solchen Schlagwort im Register zweckdienliche Literaturhinweise zu finden, oder man benutzt diese Schlagwörter als Suchauftrag für einen mechanisierten Speicher, in welchem Fachliteratur zugriffsbereit gesammelt ist.

Für den Literatursuchenden bedeutet es immer eine Erleichterung, wenn er diese Schlagwörter der vertrauten Fachsprache entnehmen kann, ohne daß er einschränkende Regelungen berücksichtigen muß. Dies ist bei den Registern und auch bei vielen computerisierten Speichern möglich. Eine kunstsprachliche – oder genauer gesagt „indexsprachliche“ – Ausdrucksweise ist beim Literatursuchenden hingegen nur wenig beliebt. Man kennt sie beispielsweise von einigen Bibliotheksklassifikationen und computerisierten Informationsspeichern her. Sie muß für den Suchzweck immer eigens erlernt und auch laufend geübt werden.

Natürlich sind Einfachheit und Bequemlichkeit allein noch nicht ausschlaggebend für die Brauchbarkeit eines Verfahrens; es muß auch die verlangte Leistung bringen. Ein Dokumentationsverfahren kann noch so einfach, bequem und kostensparend sein – wenn man mit ihm das Ziel einer ausreichend treffsicheren Literaturrecherche nicht erreichen kann, darf ihm trotzdem nicht der Vorzug gegeben werden.

Es sind schon viele vergleichende Untersuchungen über die Notwendigkeit oder Entbehrlichkeit jeder der beiden Sprachenarten in Informationssystemen unternommen worden; die Ergebnisse und Schlußfolgerungen sind sehr widersprüchlich (vgl. hierzu z. B. <sup>[1-23]</sup>). Was kann es Einfacheres und Wirksameres geben, so wird einerseits gefragt, als die Originaltexte direkt in den Speicher zu übernehmen? Jede denkbare Fragestellung müsse doch aus einem solchen Speicher beantwortbar sein, denn es ginge nichts verloren und es werde nichts hineininterpretiert, und oben-dre-in erspare man sich noch die kostspielige Übersetzung in eine Indexsprache. Dem wird andererseits die Vieldeutigkeit der natürlichen, vom Autor frei gewählten Ausdrucksweise entgegengehalten und die Kontextabhängigkeit ihrer Bedeutung. Auch unterliegen die Fachausdrücke einem ständigen Bedeutungswandel. All dies wirkt sich nachteilig auf die Treffsicherheit der Recherche in einem solchen natur- oder autorsprachlichen Speicher aus. Vor

[\*] Dr. R. Fugmann  
Wissenschaftliche Dokumentation, Hoechst AG  
D-6230 Frankfurt am Main 80

allem aber – und hierauf möchte ich mich im folgenden konzentrieren – läßt die natürliche, frei gewählte Sprache oft eine wichtige Eigenschaft vermissen, auf die es ankommt, wenn man sie nicht nur für den Zweck des Mitteilens und des Gedankenaustausches verwendet, sondern für den Zweck des gezielten Suchens: Die Voraussehbarkeit der sprachlichen Ausdrucksweise für einen gespeicherten Begriff<sup>[5, 6, 13, 24, 25]</sup>. Es liegt im Wesen der Indexsprachen, daß sie diese Voraussehbarkeit gewährleisten, wenn sie genügend zuverlässig angewendet werden.

## 2. Das semantische Dreieck

Wir müssen für diese Betrachtungen etwas weiter ausholen und gehen hierbei vom bekannten semantischen Dreieck<sup>[26]</sup> aus (vgl. auch <sup>[27–30]</sup>). Die Ecken dieses Dreiecks sind mit dem Gegenstand, dem Begriff und dem sprachlichen Ausdruck für den Begriff besetzt (Fig. 1).

Unter einem „Gegenstand“ wollen wir alles verstehen, worüber man eine sinnvolle Aussage machen kann<sup>[28]</sup>. Ein konkreter Gegenstand ist beispielsweise ein Molekül oder eine Klasse von ähnlichen Molekülen. Man kann z. B. sagen, aus welchen Atomen es besteht, in welcher Weise diese Atome miteinander verbunden sind und welche biologischen und chemischen Eigenschaften dieses Molekül hat. Ein abstrakter Gegenstand ist beispielsweise eine chemische Reaktion oder eine Stoffeigenschaft.

Als „Begriff“ betrachten wir die gedachte Zusammenfassung von sinnvollen und wesentlich erscheinenden Aussagen, die man über einen Gegenstand machen kann. Anders ausgedrückt umfaßt der Begriff die Summe der wesentlichen Merkmale eines Gegenstandes.

Wenn wir über Gedachtes reden oder schreiben wollen, dann müssen wir das Gedachte zum Ausdruck bringen: Wir müssen es mit Symbolen oder Wörtern unserer Fach- oder Umgangssprache bezeichnen. Durch Konventionen ist festgelegt, durch welches Symbol, durch welches Wort oder durch welche Wortgruppe ein bestimmter Begriff ausgedrückt werden soll. So bedeutet das Wort „Glaser-Reak-

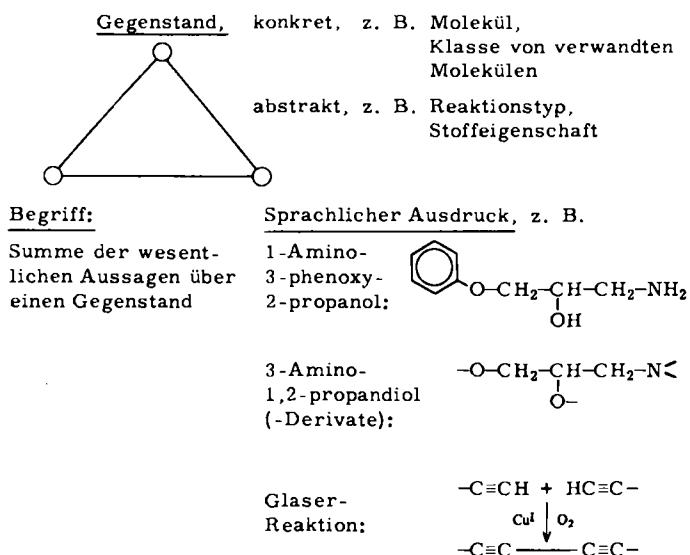


Fig. 1. Das semantische Dreieck.

tion“ eine bestimmte Art der Verknüpfung von acetylenischen Kohlenstoffatomen miteinander.

Anhand des semantischen Dreiecks kann man sich Gewißheit darüber verschaffen, worin das Suchziel besteht, wenn sich ein Informationssuchender beispielsweise mit der Frage nach Aminopropylenglykol an eine Dokumentationsstelle wendet. Sicherlich sucht er nicht nach dem Gegenstand selbst, also der Substanz Aminopropylenglykol, denn dann würde er sich gleich an den Lagerverwalter in einer Präparatensammlung wenden. Außerdem könnten abstrakte Gegenstände, z. B. eine Stoffeigenschaft oder ein Reaktionstyp, sowieso nicht in körperlicher Form herbeigeschafft werden.

Aber auch die Ausdrucksweise oder Schreibweise, die der Fragesteller zur Formulierung seines Problems benutzt hat, kann nicht das Ziel der Suche sein. Ein Fragesteller möge etwa an der Glaser-Reaktion interessiert sein. Dann wäre es ein großes Mißverständnis des Dokumentars, wenn er sich ausschließlich auf die Suche nach Texten begäbe, in denen genau das Wort „Glaser-Reaktion“ vor-

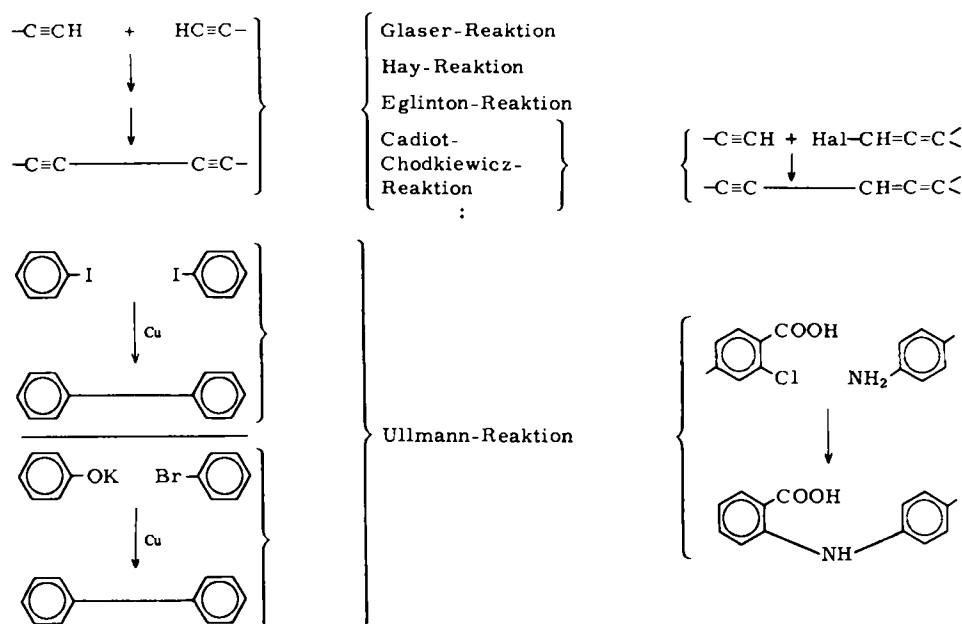


Fig. 2. Mehrdeutigkeit und Vielfalt der natursprachlichen Ausdrucksweise.

kommt. Vielmehr ist jeder Text von Interesse, in dem der fragliche Reaktionsbegriff auftritt, gleichgültig, durch welche Wörter oder Sätze oder Formelbilder er dort zum Ausdruck gebracht worden ist. Es kann dort beispielsweise von der Eglinton-, Hay- oder Cadiot-Chodkiewicz-Reaktion die Rede gewesen sein, wie in Figur 2 dargestellt.

Aber nicht nur kann ein und derselbe Begriff auf unterschiedliche Arten ausgedrückt sein, sondern ein und dieselbe Ausdrucksweise kann auch Grundverschiedenes bedeuten. Cadiot-Chodkiewicz-Reaktion wird auch für Kupplungsreaktionen von Acetylenen mit halogenierten Allenen verwendet, wobei konjugierte Enine entstehen (vgl. Fig. 2). Für viele Reaktionen kann man leicht ein Dutzend unterschiedliche Definitionen aus der Literatur zusammentragen.

Es ist also immer der Begriff, der vom Fragesteller gesucht wird und nicht der Gegenstand selbst und auch nicht die Ausdrucksweise des Fragestellers. Die Ausdrucksweise ist fast immer nur ein Hilfsmittel, um zum Ziel zu gelangen, nicht aber das Ziel selbst. Wir vervollständigen ent-

rechnet jede festgelegte, lineare Folge von alphanumerischen Zeichen, mit der man vereinbarungsgemäß einen bestimmten Begriff zum Ausdruck bringen will, z. B. ein Reaktionsname, ein Verbindungsname oder ein Personennamen. Wenn für einen Begriff *nichtlexikalische*, z. B. definitionsartige Ausdrucksweisen im Gebrauch sind, dann ist fast immer die Anzahl der denkbaren Ausdrucksweisen dieser Art sehr groß, oft sogar unbegrenzt groß. Dies ist sehr deutlich am Beispiel der Reaktionsbegriffe zu sehen. Für die Wiederauffindbarkeit der nichtlexikalisch ausgedrückten Begriffe hat dies weitreichende Konsequenzen (siehe Abschnitt 3.1 und 3.2).

### 3.1. Ausdrucksweisen für Reaktionstypen

Für die oxidative Kupplung von Acetylen werden beispielsweise etwa sechs Reaktionsnamen verwendet, von denen vier in Figur 3 eingetragen sind. Derselbe Fachbegriff kann aber auch nichtlexikalisch ausgedrückt sein, nämlich durch die Nennung der typischen Edukte und

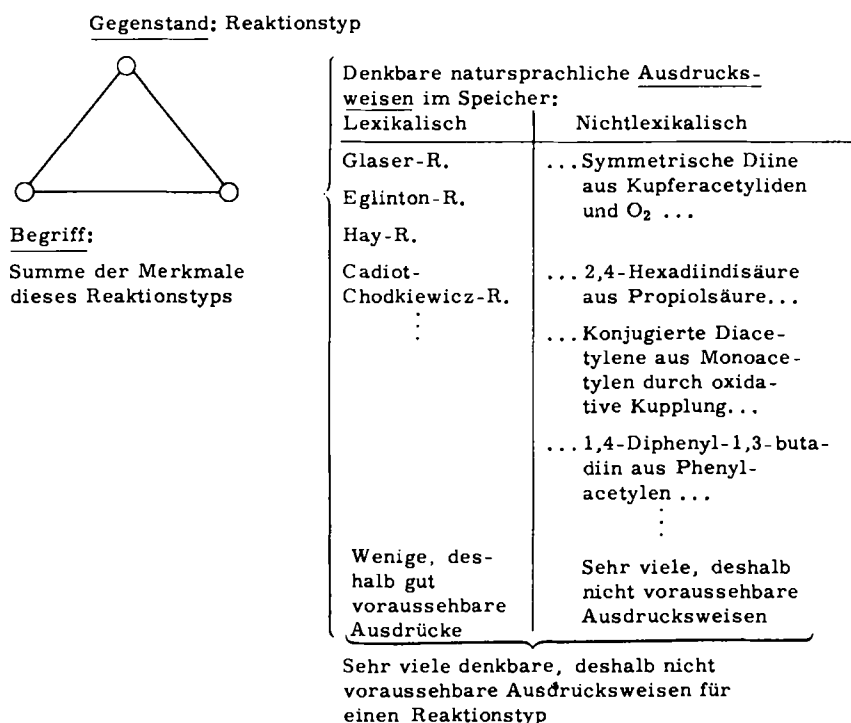


Fig. 3. Mangelnde Voraussehbarkeit von gespeicherten natursprachlichen Ausdrucksweisen für einen Reaktionstyp.

sprechend diesen Überlegungen unser Begriffsdreieck am Beispiel der Glaser-Reaktion und führen zugleich noch eine wichtige Unterscheidung zwischen zwei Arten von Ausdrucksweisen ein (siehe Fig. 3).

### 3. Lexikalische versus nichtlexikalische Ausdrucksweise

Es ist für unsere späteren Betrachtungen nützlich, unter den Ausdrucksweisen für Begriffe zwei Gruppen zu bilden und zwischen der „lexikalischen“ Bezeichnung einerseits und der „nichtlexikalischen“ andererseits zu unterscheiden. Als *lexikalisch* bezeichnen wir alle Ausdrucksweisen, die man eindeutig in ein Wörterbuch oder ein Lexikon einordnen und dort auch wieder nachschlagen kann. Hierzu

Produkte der Glaser-Reaktion. Diese Schreibweise für Reaktionen überwiegt sogar bei weitem in der Fachliteratur. Entscheidend für unsere Betrachtung ist es, daß die Zahl der denkbaren, nichtlexikalischen Schreibweisen für diesen Reaktionsbegriff unbegrenzt groß ist. Sie entspricht der unbegrenzt großen Zahl von denkbaren Edukten und Produkten dieser Reaktion.

Aus dieser Sachlage ergeben sich für das gezielte Wiederauffinden (Retrieval) weitreichende Konsequenzen. Wenn man sich nämlich auf die Suche nach einem Begriff begibt, dann muß man (im voraus!) wissen, wie dieser Fachbegriff im Speicher ausgedrückt ist, denn genau nach dieser Ausdrucksweise muß man suchen. Sind mehrere Ausdrucksweisen denkbar, dann muß man jede von ihnen beim Suchen berücksichtigen. Läßt man nur eine einzige

von diesen gespeicherten Ausdrucksweisen für den fraglichen Fachbegriff aus, dann führt dies zwangsläufig zu Informationsverlust. Alle diejenigen Texte können dann nicht mehr gefunden werden, in denen die vergessene Ausdrucksweise verwendet worden ist.

Wenn man nun die vielen Schreibweisen der Autoren für ein und denselben Reaktionsbegriff unbesehen in den Speicher übernimmt, dann müßte man beim Suchen nach einer Reaktion all die vielen Edukte und Produkte, durch die z. B. eine Glaser-Reaktion ausgedrückt sein kann, als Alternativmöglichkeiten in den Suchauftrag hineinnehmen. Dies ist aber unmöglich, und deswegen läßt sich bei einer solchen Einspeicherungsweise keine auch nur annähernd vollständige Recherche nach einer Reaktion durchführen.

Solange es wenigstens einen einzigen Namen als lexikalischen Ausdruck für einen Reaktionstyp gibt, fällt es meistens nicht sehr auf, wie lückenhaft die Suche nach Reakti-

kennen, wenn man die einschlägigen Texte bereits vor Augen hätte. Ihr Wortlaut kann nicht vorausgesehen werden. Anders ausgedrückt wird bei dieser Argumentation das als bekannt vorausgesetzt, was eigentlich erst in Erfahrung gebracht werden soll.

### 3.2. Ausdrucksweisen für Verbindungsklassen

Wir wollen noch ein anderes Beispiel mangelnder Voraussehbarkeit der natursprachlichen Ausdrucksweise betrachten. Wir nehmen an, daß ein Chemiker nach einer Verbindungsklasse sucht, sei es, daß er eines ihrer Mitglieder herstellen oder reagieren lassen will oder daß er sich für die gemeinsamen Eigenschaften der Mitglieder dieser Verbindungsklasse interessiert. Unser semantisches Dreieck ist in diesem Fall wie in Figur 4 dargestellt besetzt.

Eine solche Klasse dürfte nur in Ausnahmefällen mit genau dem Namen in den Originaltexten vorkommen, den

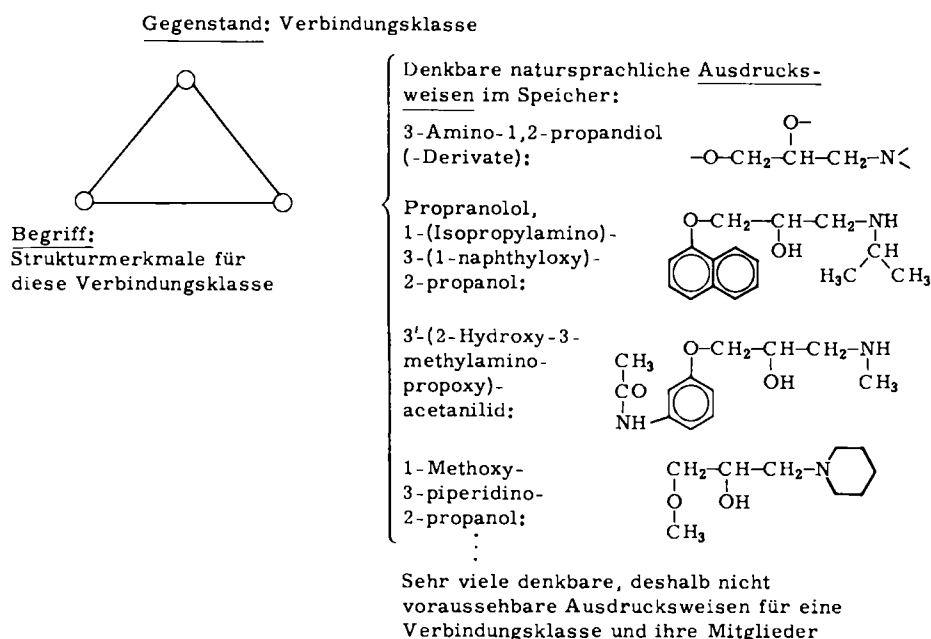


Fig. 4. Mangelnde Voraussehbarkeit von gespeicherten natursprachlichen Ausdrucksweisen für eine Verbindungsklasse.

onsnamen in Wirklichkeit ist. Man findet mit dem Reaktionsnamen immerhin einige Literaturstellen, aber was einem hierbei entgeht, und wieviel es ist, bleibt verborgen. – Es ist aber nur ein kleiner Bruchteil aller bekannt gewordenen Reaktionstypen, für die jemals ein Name geprägt wurde. Für die große Hauptmenge der Reaktionstypen gilt also, daß sie mit einem Reaktionsnamen auch nicht spurenweise auffindbar sind.

Gelegentlich wird entgegengehalten, daß man ja keineswegs sämtliche denkbaren autorsprachlichen Ausdrucksweisen zu berücksichtigen brauche, sondern nur die wenigen, die tatsächlich in den Speicher aufgenommen worden sind. Man könne sich schon ungefähr denken, an welchen Beispielen die interessierende Reaktion durchgeführt worden sei, und auf diese könne man sich in seiner Fragestellung beschränken. Dieser Ausweg ist aber nicht gangbar, denn in der Praxis und bei größeren Speichern kann man wirklich nicht im voraus wissen, durch welche Beispiele ein fraglicher Reaktionsbegriff tatsächlich im Speicher vertreten ist. Diese Ausdrucksweisen könnte man nur dann

der Fragesteller genannt hat, also etwa als „3-Amino-1,2-propandiol-Derivate“. Viel öfter werden statt dessen die Einzelverbindungen aus dieser Klasse in den Texten auftreten, und für all diese Einzelverbindungen würde sich der Fragesteller interessieren.

Nun ist aber wiederum die Zahl der denkbaren Einzelverbindungen, die zu einer Klasse gehören, unbegrenzt groß. Entsprechend groß ist die Anzahl der Verbindungs-namen, mit denen man im Suchspeicher rechnen müßte. Wiederum ist es unmöglich, all diese Namen als Suchbedingungen zu verwenden. Aus den gleichen Gründen wie vorher beim Reaktionstyp kann man dann nur noch mutmaßen, welche Verbindungs-namen nun tatsächlich im Speicher enthalten sein könnten. Vorhersehbar ist dies zum Zeitpunkt der Recherche nicht, und entsprechend lückenhaft wird eine Recherche ausfallen, bei der man sich auf solche Mutmaßungen stützen muß.

Die Schwächen der natursprachlichen Ausdrucksweise in Informationssystemen sind oft auf deren Vieldeutigkeit zurückgeführt worden. Daß dies nicht die tiefere Ursache

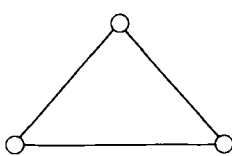
für die mangelnde Eignung zum Retrieval sein kann, ist am Beispiel der Verbindungsklassen deutlich zu sehen. Die chemische Nomenklatur ist sehr exakt und ermöglicht die eindeutige Benennung fast jeder Struktur (allerdings lassen sich oft mehrere systematische Namen für eine Verbindung bilden). Trotzdem ist sie für das Aufsuchen von Verbindungsklassen ungeeignet, wie wir am Beispiel der Aminopropandiol-Derivate gesehen haben. Die tiefere Ursache für diese Schwäche liegt wiederum darin, daß es zum Zeitpunkt der Recherche nicht voraussehbar oder rekonstruierbar ist, durch welche (zahlreichen) Namen für Einzelverbindungen eine Verbindungsklasse im Speicher vertreten ist.

Eines der am häufigsten vorgetragenen Argumente für die Einspeicherung der autorsprachlichen Ausdrucksweise besteht darin, daß hierbei keinerlei – möglicherweise verfälschende – Zusätze oder Auslassungen erfolgen. Daß dieses Argument für das Retrieval nicht stichhaltig ist, wurde zumindest für die bisher diskutierten Begriffsarten demonstriert. Die autorsprachliche Einspeicherung vermeidet hier jegliche Verfälschung, und trotzdem bleibt das Retrieval unbefriedigend.

### 3.3. Ausdrucksweisen für Einzelverbindungen

Als drittes aufschlußreiches Beispiel betrachten wir den Fall einer chemischen Einzelverbindung. Einzelverbindungen bilden aus informationswissenschaftlicher Perspektive insofern einen Sonderfall, als bei ihnen die nichtlexikalische Ausdrucksweise fast niemals vorkommt. Man braucht kaum damit zu rechnen, daß beispielsweise die in Figur 5 diskutierte Einzelverbindung durch ein natursprachliches Satzgefüge oder gar in mehreren ganzen Sätzen beschrieben wird. Immer wird im Text entweder der Verbindungsname oder die Strukturformel angegeben sein, die sich aber leicht durch ein geeignetes Programm in einen Verbindungsnamen oder eine kanonische Verknüpfungsmatrix umwandeln läßt (vgl. [39]), also in eine ebenfalls lexikalische Form. Dies bedeutet, daß die natursprachlichen Ausdrucksweisen für eine Einzelverbindung bei weitem nicht so mannigfaltig sind wie für Reaktionstypen oder Verbindungsklassen.

Wenn man die Namen einer Einzelverbindung für einen Suchauftrag zusammenstellen will, dann braucht man sich

Gegenstand: Einzelverbindung		
	Ausdrucksweisen im Speicher	
	Lexikalisch	Nichtlexikalisch
	HO-CH <sub>2</sub> -CH(OH)-CH <sub>2</sub> -NH <sub>2</sub>	
	3-Amino-1,2-propanediol	Fehlt
	1,2-Propandiol, 3-amino; Aminopropylenglykol; Nur wenige, nachschlagbare, deshalb gut voraussehbare Ausdrücke	

Begriff:  
Summe der (strukturellen) Merkmale der Einzelverbindung

Fig. 5. Gute Voraussehbarkeit von gespeicherten natursprachlichen Ausdrucksweisen für eine Einzelverbindung.

auch nicht den Kopf nach diesen Namen zu zerbrechen. Es liegt in ihrer lexikalischen Eigenart, daß man sie in Wörterbüchern oder Registern nachschlagen kann und dort auch bedeutungsgleiche zusätzliche Namen (auch Trivial- und Halbtrivialnamen) findet, einschließlich der Registry-Nummern vom Chemical Abstracts Service. Notfalls wird ein Nomenklaturexperte den systematischen Namen entwickeln können. Anders ausgedrückt ist es bei Einzelverbindungen gut voraussehbar, mit welchen natursprachlichen Schreibweisen sie im Speicher vertreten sein können. Entsprechend leicht fällt es demzufolge auch, einen Suchauftrag nach einer Einzelverbindung für einen Speicher zu formulieren, in den die natursprachlichen Ausdrücke unbesehen übernommen worden sind.

Beschränkt man sich also in einem Informationssystem darauf, ausschließlich Einzelverbindungen als Suchbedingungen zu verwenden, dann liegt es nahe, die natursprachliche Ausdrucksweise der Autoren direkt zu speichern. Zumindest bietet sich dieses einfache und billige Verfahren dann an, wenn die Autoren ihre Einzelverbindungen lückenlos benannt haben.

### 4. Indexsprachen in der Chemie

Wir haben gesehen, daß für die Fragestellung nach Reaktionen und Verbindungsklassen die bloße Übernahme der autorsprachlichen Ausdrucksweise in den Speicher nicht ausreicht. Vielmehr muß der wesentliche Teil einer jeden interessanten Publikation in eine andere, dem Wiederauffinden besser dienliche Sprache, eine „Indexsprache“, übersetzt werden – ein Vorgang, für den sich die Bezeichnung „Indexieren“ eingeführt hat<sup>[\*]</sup>.

Die natürlichste Art der indexsprachlichen Darstellung von Molekülstrukturen ist das topologische Verfahren. In einer Verknüpfungstafel wird jedes Atom in all seinen Verknüpfungen mit Nachbaratomen beschrieben, und zwar genauso exakt wie in der Strukturformel (vgl. [36]). Im Gegensatz zur etablierten chemischen Nomenklatur ist diese Darstellungsweise einer Teilstruktur unabhängig davon, wie diese in das Gesamtmolekül eingebaut ist. Im Falle der Aminopropylenglykole würde man in einer topologischen Fragestellung eine Kette von drei miteinander verbundenen Kohlenstoffatomen verlangen, von denen das erste und zweite mit Sauerstoff, das dritte mit Stickstoff verbunden ist, jeweils über Einfachbindungen. Diese Teilstruktur kann zuverlässig in jeder beliebigen Molekülstruktur erkannt werden, die dieser Verbindungsklasse angehört, wenn man ein entsprechend leistungsfähiges Maschinenprogramm zur Verfügung hat.

Allerdings läßt sich dieses Verfahren nur dann anwenden, wenn eine Struktur oder Teilstruktur Atom für Atom definiert ist. Schon ein einziger Alkyl- oder Alkylenrest von unbestimmter Länge und Verzweigung kann die topologische Darstellung im Speicher vereiteln. In solchen Fällen bewähren sich auch heute noch die traditionellen Fragmentcodes (vgl. [32-34]).

Prinzipiell läßt sich mit solchen Fragmentcodes jede beliebig verallgemeinerte Teilstruktur darstellen. Es kommt

[\*] Genauer kann man „Indexieren“ als das „Wiedergeben der Essenz eines wiederauffindbar zu machenden Dokuments in einer ausreichend voraussehbaren und wiedergabetreuen Form“ betrachten (vgl. [6, 40, 41]).

aber sehr darauf an, daß bei ihnen diejenigen Verallgemeinerungen berücksichtigt sind, die in den Texten und in den Fragestellungen auftreten. Streng formale Verallgemeinerungen erweisen sich in der Praxis als wenig nützlich, weil sie nur selten in Fragestellung und Einspeicherung in Anspruch genommen werden können. Beim Suchen nach einer Verbindungsklasse wird man z. B. kaum Gebrauch von Fragmenten wie der Anzahl der Stickstoff- und Sauerstoffatome oder der Ringe im Molekül machen.

Fragmentcodes dürfen auch nicht so umfangreich und unübersichtlich sein, daß die Zuverlässigkeit ihrer Anwendung in Frage gestellt ist. Dieser Gefahr sind vor allem diejenigen Fragmentcodes ausgesetzt, die sich mit einem Wortschatz begnügen und auf jegliche Grammatik verzichten.

An Indexsprachen für eine systematische Reaktionendokumentation sind bisher ausschließlich Fragmentcodes im Gebrauch (z. B. [33–35,37]). Beim Indexieren identifiziert man die Reaktionszentren und beschreibt sie in ihrem Zustand jeweils vor und nach der Reaktion. Treten mehrere Reaktionszentren in einer mehrstufigen Reaktion auf, z. B. mehrere Kohlenstoffatome, so muß im Interesse einer ausreichenden Wiedergabegenauigkeit zugleich ausgedrückt werden, wie diese verschiedenen Zustände gruppenweise einem gemeinsamen Reaktionszentrum zuzuordnen sind<sup>[37]</sup>.

## 5. Individualbegriff versus Allgemeinbegriff

Es fragt sich nun, inwieweit sich die bisherigen Betrachtungen über die Notwendigkeit oder Entbehrlichkeit einer Indexsprache verallgemeinern lassen. Es wäre sehr nützlich, wenn man es einem Begriff von vornherein ansehen könnte, ob er besser natursprachlich oder besser indexsprachlich zu behandeln ist.

Zur Beantwortung dieser Frage müssen wir nochmals einen kleinen Ausflug auf philosophisches Gebiet unternehmen und betrachten dazu Figur 6.

Wir knüpfen hierzu an unsere frühere Feststellung an, daß der Begriff die Summe aller wesentlichen Merkmale eines Gegenstandes ist. In dem Wort „wesentlich“ kommt zum Ausdruck, wie subjektiv die Begriffsbildung ist und wie stark sie von der Perspektive dessen abhängt, der einen Gegenstand näher betrachtet. Uns liegt die Perspektive des Chemikers am nächsten, der sich beispielsweise beim Anblick einer Flasche mit Aminopropylenglykol in erster Linie für den Inhalt dieser Flasche und speziell für die Molekülstruktur dieses Stoffes interessiert.

Als wesentlich wird der Chemiker beispielsweise die primäre Aminogruppe und die primäre alkoholische Hydroxygruppe ansehen. Schon diese beiden Merkmale reichen für den sinnvollen Begriff der aliphatischen Aminoalkohole aus. Ein weiteres Merkmal ist die sekundäre Hydro-

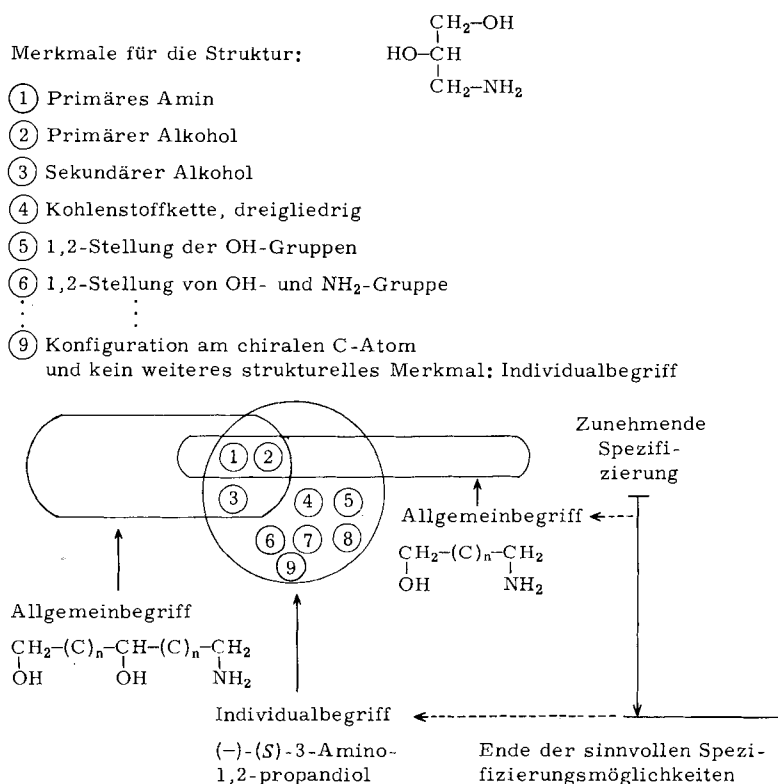


Fig. 6. Übergang vom Allgemeinbegriff zum Individualbegriff durch fortschreitende und abschließende Merkmalsvermehrung.

Die Wiedergabe nichtstrukturchemischer Begriffe im Speicher kann man durch Klassifikationen und Thesauri ausreichend voraussehbar machen. Beispiele hierfür sind verfahrenstechnische Begriffe, Stoffeigenschaften und Stoff-Verwendungsmöglichkeiten. Eine Vielzahl sehr unterschiedlicher Vokabularien wird in der Praxis für diesen Zweck verwendet.

xygruppe. Dies führt zum spezifischeren Allgemeinbegriff der aliphatischen Aminodiole. Merkmal auf Merkmal kann man so den immer spezifischer werdenden Allgemeinbegriffen hinzufügen.

Das letzte Merkmal wären vielleicht Angaben zur Konfiguration am chiralen Kohlenstoffatom. Eine weitere Spezifizierung durch Merkmalsvermehrung ist nicht möglich.

Man könnte zwar noch chemische, physikalische und biologische Eigenschaften nennen, auch die Summenformel ausrechnen, aber diese Merkmale tragen nicht mehr zur schärferen Charakterisierung und Definition der Molekülstruktur bei. Anders ausgedrückt sind wir beim Individualbegriff angelangt, bei einem Begriff also, zu dem es in Anlehnung an von Freytag-Löringhoff keinen sinnvollen Unterbegriff mehr gibt<sup>[31]</sup>.

Was hat dies alles mit der Frage nach natursprachlicher oder indexsprachlicher Ausdrucksweise zu tun? Die Erfahrung hat gezeigt, daß es die Allgemeinbegriffe sind, bei denen man immer mit einer uferlos großen Zahl von lexikalischen und nichtlexikalischen Ausdrucksweisen rechnen muß, mit Ausdrucksweisen also, die zur Anwendung einer Indexsprache zwingen. Individualbegriffe hingegen werden fast immer lexikalisch ausgedrückt, d. h. in einer Art, die schon in der unregelmäßigen natursprachlichen Ausdrucksweise recht gut voraussehbar ist. Da es die Hauptaufgabe einer Indexsprache ist, Voraussehbarkeit herzustellen, ist bei den Individualbegriffen die Notwendigkeit für die Anwendung von Indexsprachen viel geringer einzuschätzen<sup>[5, 6, 13]</sup>.

Reaktionstypen erscheinen aus dieser Perspektive ebenfalls als Allgemeinbegriffe, und es überrascht deswegen nicht, daß wir sie in der Literatur auch meistens in nichtlexikalischer Ausdrucksweise antreffen. Die Glaser-Reaktion erweist sich beispielsweise schon insofern als Allgemeinbegriff, als es bei ihr die speziellen Varianten der Eglington- und der Hay-Reaktion als Unterbegriffe gibt. Auch diese Reaktionen ließen sich durch die Art der Reaktionsbedingungen noch weitergehend spezifizieren. Es ist für diese Betrachtungen belanglos, daß für diese – noch spezifischeren – Reaktionsbegriffe noch keine Namen gebildet worden sind.

Als Allgemeinbegriff in diesem Sinne sind auch verfahrenstechnische Begriffe anzusehen, z. B. Destillieren, Filtrieren und Kristallisieren. Auch die zugehörigen Eigenschaftsbegriffe wie Destillierbarkeit, Kristallisationsfähigkeit und viele andere Stoffeigenschaften und -verwendungen rechnen zu den Allgemeinbegriffen.

Ein typischer Individualbegriff hingegen ist der Autor einer Publikation, eine bestimmte Institution, ein Land, eine bestimmte Patentschrift, ein bestimmtes Buch usw. All diesen Begriffen ist gemeinsam, daß sie fast immer mit Namen benannt sind, also praktisch nur in lexikalischer Form auftreten. Dies hat schon immer die Handhabung dieser Begriffe in Informationssystemen sehr erleichtert. Nicht umsonst sind die Autoren- und Summenformelregister schon immer die schnellsten und billigsten gewesen.

Strukturformeln von Einzelverbindungen und auch die zugehörigen topologischen Verknüpfungstabellen (vgl. Abschnitt 4) haben zwar nichtlexikalischen Charakter, können aber durch rein maschinelle Prozesse leicht in eine lexikalische Form gebracht werden, z. B. in die kanonische Form, die Morgan entwickelt hat<sup>[39]</sup>. Daß hier eine nichtlexikalische Ausdrucksweise in eine lexikalische Form maschinell übersetzt werden kann, beruht auf der sehr systematischen und einheitlichen Syntax, die in diesen Verknüpfungstabellen herrscht.

## 6. Verzicht auf Indexsprachen in der Chemie-Dokumentation

Bei vielen kommerziell zugänglichen Datenbanken mit on-line-Zugriff ist auf die Anwendung von speziellen Indexsprachen verzichtet worden. Zur Begründung wird oft betont, daß auf diese Weise die hohen Kosten für die fachkundige Aufbereitung oder für aufwendige Maschinenprogramme eingespart würden. Auch wird auf schlechte Erfahrungen hingewiesen, die mit bestimmten Indexsprachen gesammelt worden sind, sowie auf Retrieval-Experimente, aus denen die Überlegenheit oder zumindest Ebenbürtigkeit der nichtindexierenden Einspeicherung hervorginge. Auch sei im Zeitalter der Computertechnik das Festhalten an Indexsprachen überholt. Diese Argumente sind in so großer Häufigkeit und Einseitigkeit vorgetragen worden, daß eine Korrektur dieses Bildes zur Vermeidung von Fehlentwicklungen und Sackgassen (vgl. <sup>[42, 43]</sup>) angezeigt ist. Die Schnelligkeit des Zugriffs, auch über große Entfernungen hinweg, hat über die oftmals geringe Retrieval-Treffsicherheit hinweggetäuscht, die regelmäßig dort auftritt, wo auf Indexierung verzichtet worden ist.

Die Einsparung qualifizierter Arbeitskraft ist im Bereich der Individualbegriffe durchaus möglich. Sie sind im Text leicht erkennbar und auch durch die Eigennamen, mit denen sie benannt sind, scharf abgegrenzt. Bei Allgemeinbegriffen hingegen ist das Erkennen, Auswählen und richtige Bezeichnen schon bei bescheidenen Anforderungen an die Treffsicherheit eine Voraussetzung für das Retrieval. Diese Arbeit ist undenkbar ohne Fachkenntnisse. Wie soll später eine Frage nach der Isolierung von peptidartigen Bitterstoffen aus Pilzen beantwortet werden, wenn nicht bei der Einspeicherung die Zehntausende von Namen für Pilze und Peptide als solche erkannt und mit diesen Allgemeinbegriffen abrufbar gemacht werden? Sicherlich nicht auf die Art, daß man eine sehr große und ständig wachsende Zahl von Namen in die Fragestellung oder in das Maschinenprogramm als Suchalternativen hineinnimmt. Darüber hinaus müßte das Maschinenprogramm auch jeden zukünftigen Namen richtig verarbeiten – eine Aufgabe, die selbst dem Experten oftmals ausgedehnte Nachschlagearbeit abverlangt und die niemand im voraus leisten könnte, d. h. vor dem Bekanntwerden dieser neuen Namen. Hinzu kommt noch, daß die Bedeutung vieler Wörter erst aus dem Zusammenhang hervorgeht, in den sie eingebettet sind. Dieser Zusammenhang ist aber meistens in unvorhersehbarer Weise ausgedrückt und entzieht sich daher jeder vorprogrammierten Analyse.

Indexierungsarbeit ist auch erforderlich, um bei einer Reaktionendokumentation die jeweils reagierenden Zentren in den Edukten und Produkten zu erkennen und zueinander in Beziehung zu setzen, eine Arbeit, bei der wohl noch auf lange Zeit die Mitwirkung des Fachmannes unentbehrlich sein wird, von den einfachen Fällen einmal abgesehen. Der Verzicht auf jegliche Indexierung ist hier gleichbedeutend mit dem Verzicht auf eine brauchbare Reaktionendokumentation. Die Möglichkeiten und Grenzen der maschinellen Reaktionenindexierung haben Lynch und Willett et al.<sup>[44–47]</sup> untersucht. Im REACCS-System von Wipke<sup>[48]</sup> lassen sich die maschinelle und die intellektuelle Reaktionenindexierung miteinander kombinieren. – Wie wenig auch mit der unbesehenen, nichtindexierenden

Einspeicherung von Verbindungsnamen für eine Verbindungsklassenrecherche gewonnen ist, wurde bereits in Abschnitt 3.2 gezeigt.

Setzt man beim Verzicht auf Indexierung alle Hoffnungen auf zukünftige Maschinenprogramme, dann müßten diese auf geheimnisvolle Weise eine Aufgabe lösen, für die der Mensch, der diese Programme schreiben soll, gar keinen Lösungsweg kennt, weil er die Situationen nicht voraussehen kann, für die er die entsprechenden Instruktionen im Programm niederlegen muß. Wieviel Vertrauen würde man wohl einer Reparaturwerkstätte entgegenbringen, die damit wirbt, daß bei ihr ausschließlich ungelernte Kräfte und Automaten eingesetzt sind?

Das Urteil über die Treffsicherheit des Retrievals ist verbreitet noch durch Trugschlüsse geprägt (vgl. hierzu <sup>[38]</sup>). Ein Fragesteller, der heute mit dem Ergebnis seiner Recherche noch hochzufrieden ist, würde morgen vielleicht sehr verdrossen sein, wenn er aus einer anderen Quelle erfahren würde, was ihm vorher alles entgangen ist und wieviel dies gewesen ist. – Oft beruhigt man sich auch damit, daß es genüge, wenn man nur einige wenige „gute“ Fundstellen nachgewiesen erhielte. Aber genügt es wirklich, wenn die Auswahl dieser wenigen Fundstellen dem Zufall überlassen bleibt, wie es bei mangelhaft treffsicherem Retrieval fast immer der Fall ist? Meistens beruht das vom Systembenutzer abgegebene Urteil auf einem trügerischen Sicherheitsgefühl, denn er zieht zumeist nur die – weil offenkundigen – Ballastfehler in Betracht, nicht jedoch auch die – zunächst verborgen bleibenden – Verlustfälle. Eine genauere Untersuchung würde zeigen, daß jeglicher bewertende Vergleich von zwei Informationssystemen in vielerlei Hinsicht überraschend schwierig, zeitraubend und kostspielig ist, wenn man zu wirklich schlüssigen Ergebnissen gelangen will. Die Literatur ist angefüllt mit ausgedehnten Studien und jahrelangen Kontroversen auf diesem heiklen Gebiet.

Auch aus dem Verlauf von Retrievalexperimenten ist gelegentlich gefolgert worden, daß man auf eine spezielle Indexsprache verzichten und unbedenklich ohne jede Sprachenregelung und Indexierung einspeichern könne. Meistens hat es sich hierbei jedoch um unnatürlich kleine Sammlungen gehandelt, die sich obendrein noch auf die Literatur einer engen Zeitspanne beschränkt hatten. Unter diesen wirklichkeitsfremden Umständen kennen die Versuchspersonen den Inhalt und auch den Wortlaut der Versuchsdokumente unnatürlich gut. Auch wird die Erinnerung hieran durch die konzentrierte Beschäftigung mit diesen wenigen Dokumenten immer wieder aufgefrischt. Die Voraussehbarkeit der autorsprachlichen Ausdrucksweise wird also unter diesen Bedingungen gar nicht ernstlich auf die Probe gestellt. Unter diesen künstlichen Versuchsbedingungen tritt auch die große Vielfalt der autorsprachlichen Ausdrucksweise, mit der man es bei weit in die Vergangenheit hineinreichenden Sammlungen immer zu tun hat, noch nicht erschwerend in Erscheinung. Deshalb haben Systeme mit autorsprachlicher Einspeicherung oder anderweitig freier Wortwahl in solchen Versuchen immer unnatürlich gut abgeschnitten. Der Mangel an Voraussehbarkeit der Ausdrucksweise ist, anders ausgedrückt, in diesen Versuchen durch Gedächtnisleistung ersetzt worden. Aus dem gleichen Grund hat die freie Wortwahl auch immer dort unnatürlich gut abgeschnitten, wo die Einspei-

cherung und das Suchen einer einzigen Person überlassen ist.

Andererseits ist jede Übersetzung in eine andere Sprache problematisch, wenn der Ausgangstext vieldeutig ist und wenn die Zielsprache (noch) nicht über passende Ausdrucksmöglichkeiten verfügt. Die Übersetzung in eine Indexsprache macht von dieser Regel keine Ausnahme. In solchen Fällen ist es ratsam, auch die autorsprachliche Ausdrucksweise vorsorglich in den Speicher zu übernehmen, zumindest zusätzlich zu einer versuchten Übersetzung in die indexsprachliche Ausdrucksweise. Der autorsprachlich ausgedrückte Begriff wird auf diese Weise immerhin zugriffsbereit gehalten. Beispielsweise mag es in einem verbindlichen Vokabular für die Polymerchemie (vorerst) nur den Allgemeinbegriff des Verspinnens geben. Trotzdem lassen sich dann die spezifischeren Begriffe des Schmelzspinnens, Trockenspinnens, Naßspinnens usw. zusätzlich festhalten, wenn auch noch nicht mit der Zuverlässigkeit, die den Vokabularbegriffen innewohnt. Eine solche Kombination beider Sprachenarten hat sich in unserer Praxis und in der IDC<sup>[\*]</sup> als außerordentlich leistungsfähig erwiesen.

## 7. Schlußfolgerung

Wir hatten uns die Frage nach den Möglichkeiten und Grenzen von natursprachlichen und indexsprachlichen Informationssystemen gestellt. Immer dann, so lautet unsere Antwort, ist die Arbeitsweise mit frei formulierter Natursprache erwägenswert oder sogar überlegen, wenn einem Informationssystem ausschließlich das Aufsuchen von Individualbegriffen abverlangt wird, oder wenn die Übersetzung in eine Indexsprache problematisch ist. Dies gilt besonders dann, wenn Kosteneinsparungen beim Einspeichern vorrangig sind. Auf eine Indexsprache hingegen kann man nicht verzichten, wenn auch Fragen nach Allgemeinbegriffen mit passabler Treffsicherheit beantwortet werden sollen. Es ist unvermeidbar, daß diese Begriffe in eine indexsprachliche Ausdrucksweise übersetzt werden müssen. Der Grund hierfür liegt in der Unvorhersehbarkeit der natursprachlichen Ausdrucksweise für diese Begriffe.

Wenn also festgestellt wird, daß auch für die Dokumentation die frei formulierte natürliche Sprache das einzig Sinnvolle sei und daß Indexsprachen entbehrlich sind, oder wenn das genaue Gegenteil behauptet wird, dann müßte auch immer gesagt werden, ob diese Aussagen für Individualbegriffe oder für Allgemeinbegriffe gelten sollen. Es müßte auch dargelegt werden, welche Anforderung an die Treffsicherheit des Retrievals gestellt werden und insbesondere auch, wie groß die Sammlungen gewesen sind und über welche Zeiträume sie sich erstreckt haben.

Sind die Anforderungen an ein Informationssystem hoch, dann können sie nur durch sinnvolle Kombination der beiden Sprachenarten erfüllt werden. In der Chemie treten diese Zusammenhänge wohl am deutlichsten zutage. Sie lassen sich aber auch auf anderen Gebieten erkennen. Wenn man sich dies vor Augen hält, dann verflüchtigt sich

[\*] IDC: Internationale Dokumentationsgesellschaft für Chemie mbH, Hamburger Allee 26-28, D-6000 Frankfurt am Main 90.



auch eine alte Streitfrage, um die schon viele leidenschaftliche Diskussionen geführt worden sind.

Eingegangen am 17. März 1982 [A 421]

- [1] S. R. Ranganathan: *Prolegomena to Library Classification*, ASIA Publishing House, London 1967, S. 329 (Chapter MB).
- [2] F. W. Lancaster: *Vocabulary Control for Information Retrieval*, Information Resources Press, Washington, DC 1972, speziell S. 1, 139.
- [3] D. Soergel: *Indexing Languages and Thesauri*, Melville Publishing Company, Los Angeles, CA 1974.
- [4] B. C. Vickery: *On Retrieval System Theory*, Butterworths, London 1968, S. 65.
- [5] R. Fugmann, The Glamour and the Misery of the Thesaurus Approach, *Int. Classification 1* (1974) 76, speziell S. 78.
- [6] R. Fugmann, Toward a Theory of Information Supply and Indexing, *Int. Classification 6* (1979) 3, speziell S. 6, 14.
- [7] F. Gebhardt, I. Stellmacher, Design Criteria for Documentation Retrieval Languages, *J. Am. Soc. Inf. Sci.* 29 (1978) 191.
- [8] E. Svenonius, Current Issues in Subject Control of Information, *Libr. Q.* 47 (1977) No. 3, S. 326.
- [9] E. Svenonius, Natural Language vs. Controlled Vocabulary, *Proc. Forth Canadian Conference on Information Science*, London, Ontario, Mai 1976.
- [10] H. Wellisch, The Cybernetics of Bibliographic Control, *J. Am. Soc. Inf. Sci.* 31 (1980) 41, speziell S. 45.
- [11] R. A. Wall, Intelligent Indexing and Retrieval: A Man-Machine Partnership, *Inf. Process. Management* 16 (1980) 73.
- [12] H. Lehmann, A. Blaser: *Query Languages in Data Base Systems*, IBM Deutschland GmbH, Tiergartenstraße 15, D-6900 Heidelberg.
- [13] E. König: Verbindliches vs. freies Indexieren, in: *Numerische und Nicht-numerische Klassifikation zwischen Theorie und Praxis*. Proceedings der 5. Fachtagung der Gesellschaft für Klassifikation, Hofgeismar 1981, S. 263, speziell S. 266, INDEKS-Verlag Frankfurt/Main, ISBN 3-88672-009-8.
- [14] F. Wenzel, Lösung morphologischer Probleme im Frei-Text-Retrieval durch Segmentierung, *Nachr. Dok.* 30 (1979) 212.
- [15] H. Haendler, Selektionsgerechte Indikation von Sachgebieten und Sachverhalten, *Int. Classification 2* (1975) 25, speziell S. 28.
- [16] V. Cherniavsky, On Algorithmic Natural Language Analysis and Understanding, *Inf. Systems 3* (1978) 5.
- [17] R. G. Henzler, Free or Controlled Vocabularies, *Int. Classification 5* (1978) 21.
- [18] H. Wellisch, A Flow Chart for Indexing with a Thesaurus, *J. Am. Soc. Inf. Sci.* 23 (1972) 185.
- [19] J. O'Connor, Data Retrieval by Text Searching, *J. Chem. Inf. Comp. Sci.* 17 (1977) 181.
- [20] L. Rogalski, On-Line Searching of the American Petroleum Institute's Databases, *J. Chem. Inf. Comp. Sci.* 18 (1978) 9.
- [21] B. Sudarshan, Development of Reference Retrieval System with Simultaneous Building – up of Thesaurus for Industrial Information Centres, *Libr. Sci.* 16 (1979) Paper E.
- [22] J. N. Olsgaard, J. E. Evans, Improving Keyword Indexing, *J. Am. Soc. Inf. Sci.* 32 (1981) 71.
- [23] J. Rothmann, Online Searching and Paperless Publication, *J. Am. Soc. Inf. Sci.* 32 (1981) 71.
- [24] R. Fugmann, The Theoretical Foundation of the IDC System, *ASLIB Proc.* 24 (1972) 123.
- [25] J. Mills, Progress in Documentation, *J. Doc.* 26 (1970) 120, speziell S. 123.
- [26] C. K. Ogden, I. A. Richards: *The Meaning of Meaning*, A Harvest/HBJ Book, Harcourt Brace Jovanovich, London 1936, S. 11.
- [27] O. Sechser, Modi der Bedeutung von Einzelausdrücken in Retrievalsprachen, in: *Klassifikation und Erkenntnis II*. Proceedings der 3. Fachtagung der Gesellschaft für Klassifikation, Königstein 1979, S. 1, IN-DEKS-Verlag Frankfurt/Main.
- [28] I. Dahlberg, A Referent-Oriented, Analytical Concept Theory for INTERCONCEPT, *Int. Classification 5* (1978) 142, speziell S. 144.
- [29] I. Dahlberg, Zur Theorie des Begriffs, *Int. Classification 1* (1974) 12.
- [30] I. Dahlberg, Conceptual Definitions for INTERCONCEPT, *Int. Classification 8* (1981) 16.
- [31] B. von Freytag-Löringhoff: *Logik – Ihr System und Verhältnis zur Logistik*, Urban Taschenbuch 16, Kohlhammer, Stuttgart 1955, S. 27.
- [32] W. Nübling, W. Steidle, Dokumentationsring der chemisch-pharmazeutischen Industrie, *Angew. Chem.* 82 (1970) 618; *Angew. Chem. Int. Ed. Engl.* 9 (1970) 596.
- [33] R. Fugmann: The IDC-System, in J. E. Ash, E. Hyde: *Chemical Information Systems*, Ellis Horwood, Chichester 1975, S. 195 ff; R. Fugmann, W. Braun, W. Vaupel: GREMAS – ein Weg zur Klassifikation und Dokumentation in der organischen Chemie, *Nachr. Dok.* 14 (1963) 179.
- [34] Derwent Publications Ltd., Rochdale House 128, Theobalds Road, London WC1X 8RP (England).
- [35] J. Valls: Reaction Documentation, in W. T. Wipke, S. R. Heller, R. J. Feldmann, E. Hyde: *Computer Representation and Manipulation of Chemical Information*, Wiley, New York 1974.
- [36] E. Meyer, Vielseitige maschinelle Suchmöglichkeiten nach Strukturformeln, Teilstrukturen und Stoffklassen, *Angew. Chem.* 82 (1970) 605; *Angew. Chem. Int. Ed. Engl.* 9 (1970) 583.
- [37] R. Fugmann, G. Kusemann, J. H. Winter, The Supply of Information on Chemical Reactions in the IDC System, *Inf. Process. Management* 15 (1979) 303.
- [38] D. Soergel, Is User Satisfaction a Hobgoblin?, *J. Am. Soc. Inf. Sci.* 27 (1979) 256.
- [39] H. L. Morgan, The Generation of a Unique Machine Description for Chemical Structures, *J. Chem. Doc.* 5 (1965) 107.
- [40] R. Fugmann, Zur Frage der Vereinheitlichung des Indexierens, *Nachr. Dok.* 29 (1978) 121.
- [41] FID-Classification Research; *Int. Classification 8* (1981) 196.
- [42] R. Fugmann, Der Weg in die Sackgasse bei der mechanisierten Dokumentation, *Nachr. Dok.* 17 (1966) 79.
- [43] M. Woitschach, Automatische Dokumentation in der Sackgasse?, *Nachr. Dok.* 17 (1966) 74.
- [44] M. F. Lynch, P. Willett: The Production of Machine-Readable Descriptions of Chemical Reactions Using Wisswesser Line Notations, *J. Chem. Inf. Comp. Sci.* 18 (1978) 149.
- [45] M. F. Lynch, P. Willett: The Automatic Detection of Chemical Reaction Sites, *J. Chem. Inf. Comp. Sci.* 18 (1978) 154.
- [46] J. J. McGregor, P. Willett: Use of a Maximal Common Subgraph Algorithm in the Automatic Identification of the Ostensible Bond Changes Occurring in Chemical Reactions, *J. Chem. Inf. Comp. Sci.* 21 (1981) 137.
- [47] P. Willett: Computer Techniques for the Indexing of Chemical Reaction Information, *J. Chem. Inf. Comp. Sci.* 19 (1979) 156.
- [48] Computer System Searches Chemical Reactions, *Chem. Eng. News*, April 12, 1982.