

Theoretische Aspekte zur Kommunikation in der Chemie

Von Robert Fugmann^[*]

Nichts ist praktischer als Theorie

Ludwig Boltzmann

Das erfolgreiche Suchen nach Publikationen unter thematischen Gesichtspunkten wird von Gesetzmäßigkeiten beherrscht, deren Beziehungen zur Thermodynamik unverkennbar sind. Je stärker eine Literatursammlung angewachsen ist, desto größere Ordnung muß in ihr herrschen, weil Geduld, Zeitaufwand und Gedächtnisleistung, mit denen der Mensch mangelnde Ordnung zu kompensieren pflegt, nicht proportional mitwachsen können. Mit den traditionellen Dokumentationsmethoden lassen sich derartig hohe Ordnungen, wie sie heute in der Fachliteratur erforderlich sind, nicht herstellen. Setzt man aber moderne technische Methoden zur Verbesserung der Kommunikation ein, so muß auch die begriffliche Grundlage eines solchen Dokumentationsverfahrens revidiert und den neuen Möglichkeiten angepaßt werden. In der Chemie sind die Voraussetzungen für den wirkungsvollen Einsatz moderner technischer und begrifflicher Methoden vergleichsweise günstig, weil hier die Terminologie weiter entwickelt ist als in den meisten anderen Wissenschaften.

1. Einführung

Für das Voranschreiten einer jeden Forschungsarbeit spielt es eine wesentliche Rolle, wie intensiv die Kommunikation mit der wissenschaftlichen Umwelt der Gegenwart und Vergangenheit ist. Ein unmittelbarer praktischer Nutzen erwächst jedem Wissenschaftler und Techniker aus dieser Kommunikation, wenn sie ihm Versuchsergebnisse zugänglich macht, die er nun nicht mehr in eigenen zeitraubenden und kostspieligen Versuchen zu ermitteln braucht. Oftmals erweisen sich spezielle Forschungen überhaupt als überflüssig, da das fragliche Problem bereits früher gelöst oder einer Lösung nahegebracht worden ist.

Ebenso wichtig ist die heuristische Rolle, die die Kommunikation schon immer in der Wissenschaft gespielt hat. Wenn man sich eine Hypothese über unerforschte naturgesetzliche Zusammenhänge bilden will, so sind die einschlägigen Beobachtungen von Fachkollegen von gleichem Wert wie die eigenen Erfahrungen, und eine Hypothese steht auf um so sichereren Füßen, je größer die Zahl der Beobachtungen ist, die in Betracht gezogen worden sind. Je vollständiger weiterhin alle späteren Versuchsergebnisse auf ihre Vereinbarkeit mit dieser Hypothese geprüft werden – wiederum im wesentlichen eine Frage wirksamer Kommunikation – desto rascher und sicherer entwickelt sich die Hypothese zur festgefügtten Theorie, und desto rascher kann es zur Bildung einer neuen Hypothese mit höherem Erklärungswert kommen.

Hätte man stattdessen das erkenntnistheoretische Rohmaterial ausschließlich durch eigene Experimente und unter Verzicht auf Kommunikation erarbeiten wollen, so hätten oft für ein solches Vorhaben viele Menschenalter nicht ausgereicht. Es muß auch dahingestellt

bleiben, ob es nach der Versenkung in ein derartiges Übermaß von Detailarbeit überhaupt noch zu der schöpferischen Idee gekommen wäre.

Kommunikation ist auch immer im Spiele, wenn man – mehr oder minder zufällig – auf neue Methoden und Erkenntnisse auf anderen Arbeitsgebieten stößt, wonach diese sich vielfach als äußerst fruchtbar auf dem eigenen Spezialgebiet erweisen. Viele Grenzgebiete zwischen Physik, Biologie und Chemie verdanken ihre relativ raschen Fortschritte weitgehend einer interdisziplinären Kommunikation. Jedem forschend Tätigen gereicht es auf längere Sicht zum Nachteil, wenn er sich von diesem wechselseitigen Geben und Nehmen zurückhält.

Zuweilen liegt die Ursache hierfür im persönlichen Arbeitsstil und in einer wenig ausgeprägten Kommunikationsbereitschaft. Stets aber sind außerdem andere Umstände mit im Spiele, die paradoxerweise erst durch die wissenschaftliche Tätigkeit selbst und durch die Bekanntgabe ihrer Ergebnisse entstehen, und zwar um so stärker, je intensiver die wissenschaftliche Arbeit betrieben wird. Es genügt offenbar nicht, daß Erfahrungen und Erkenntnisse lediglich *existieren*, abgedruckt in einer wenig verbreiteten Zeitschrift oder dargeboten auf einer exklusiven Vortragstagung. Sie müssen vielmehr auch für denjenigen leicht zugänglich und auffindbar sein, der den Erkenntnisprogreß vollziehen will oder der die publizierten Versuchsergebnisse für seine eigenen Studien braucht. Nur dann können sie im heuristischen oder praktischen Sinne nützlich werden.

Die Menschen unterscheiden sich nun, wie man weiß, beträchtlich in ihrer Neigung, sich auf die Suche nach den Erfahrungen anderer zu begeben, um sie für ihre eigene Situation, wenn auch mit aller gebotenen Kritik, in Betracht zu ziehen. So werden auch die Wissenschaftler individuell sehr unterschiedliche Anstrengungen unternehmen, um fremde Erfahrungen ausfindig zu

[*] Dr. R. Fugmann
Farbwerke Hoechst AG
623 Frankfurt/M. 80, Postfach 800320

machen. Publiziertes Wissen wird daher um so weniger genutzt, je mühsamer die Suche danach ist. Dies gilt um so mehr, als die Erfolgsaussicht beim Suchen oft ungewiß ist, während man durch das eigene Experiment in abschätzbarer Zeit mit Sicherheit zu einem Ergebnis kommt.

Bereits heute ist eine beträchtliche Diskrepanz zwischen dem tatsächlichen Aufwand in Forschung und Entwicklung einerseits und andererseits dem Aufwand zu verzeichnen, welcher ausreichen würde, wenn nur die Kommunikation funktionierte. Bezeichnend für die Resignation, die hier gelegentlich schon angetroffen werden kann, ist der Ratschlag, man möge auf jede orientierende Literaturstudie verzichten, wenn ein Forschungsprojekt die Kosten von 100000 Dollar voraussichtlich nicht überschreitet^[1].

Der Umfang an unbeabsichtigter Doppelarbeit durch mangelnde Informationen über das vorhandene Wissen ist mehrmals abgeschätzt worden. Die verlässlichsten Werte für die Chemie dürften bei ca. 30% liegen^[2,3]. Völlig ungewiß bleibt es aber, in wie vielen Fällen es bisher nur deshalb noch nicht zu neuer naturwissenschaftlicher Erkenntnis gekommen ist, weil einem Genius gerade die entscheidenden Steine im Mosaik der Einzelfakten — obwohl publiziert — gefehlt haben, welche das Naturgesetz hätten ahnen lassen können.

Die Situation in der Wissenschaft ist heute mit derjenigen eines Produktionsbetriebes vergleichbar geworden, der seine Erzeugnisse auf Vorrat produziert und aufbewahrt, bei dem aber bei einer Kundenanfrage das Wiederfinden derartig viel Mühe macht, daß man, anstatt im Lager nachzusuchen, lieber neu produziert. Wenn ein solches Unternehmen letztlich scheitert, so ist dies gewiß nicht auf Mängel bei den Herstellungsverfahren zurückzuführen. Gerade dort kann man sogar eine enorme Produktivität erkennen. Der Fehlschlag ist vielmehr Mängeln zuzuschreiben, die auf einem ganz anderen Gebiet liegen, das aber für den Gesamterfolg der Produktionsarbeit von nicht geringerer Bedeutung ist.

Dieser Aufsatz soll zeigen, welcher Art die Hindernisse sind, die bisher einer flüssigen Kommunikation in der Wissenschaft und insbesondere in der Chemie im Wege gestanden haben und auf welche Weise diese Hindernisse mit der größtmöglichen Aussicht auf Dauererfolg überwunden werden können.

Für diese Betrachtungen ist es von sekundärer Bedeutung, ob das publizierte Wissen der Zeitschriften- oder Patentliteratur entstammt, ob es in Büchern erschienen ist oder — zunächst — in mündlicher Form übermittelt worden ist. Einen entscheidenden Unterschied macht es hingegen aus, ob die Interessenlage des Informationensuchenden definierbar ist oder nicht. Nur im ersten Fall kann er sich nämlich *gezielt* auf die Literatursuche begeben. Andernfalls bleibt ihm nur der Weg der *ungezielten* Literaturstudie offen.

2. Die ungezielte Literaturstudie

Es liegt im Wesen des wissenschaftlichen Fortschrittes, daß er von niemandem im Detail vorausgesagt werden kann. Demzufolge kann auch niemand im voraus

spezifizieren, welche künftigen Entdeckungen ihn besonders interessieren werden. Immer erst *nach* dem Kennenlernen einer Neuerung, also *a posteriori*, kann er von Fall zu Fall und rein subjektiv entscheiden, ob diese für ihn interessant ist oder nicht. Die Entdeckung des Penicillins oder der Edelgasverbindungen, die Bindung von molekularem Stickstoff an organische Moleküle unter physiologischen Bedingungen, die überraschende Protonierungsmöglichkeit gesättigter Kohlenwasserstoffe, die Anhäufung reiner Neutronenmaterie in bestimmten Himmelskörpern sind nur einige Themen als Beispiele, die wohl jeden Chemiker interessieren haben. Aber kaum einer hätte *a priori*, d.h. *vor* ihrem Bekanntwerden, sein Interesse in dieser Richtung definieren und etwa eine Dokumentationsstelle beauftragen können, ihm entsprechende Publikationen zuzuleiten, falls sie erscheinen sollten. Nur durch ungezieltes regelmäßiges Lesen der Fachliteratur und durch kollegiale Kontakte kann er sich hier auf dem Laufenden halten oder durch Stöbern in einer Bibliothek.

„Welches sind die günstigsten Reaktionsbedingungen für die Reaktion A?“ „Welche wirtschaftlichen Aussichten sind dem Produkt B einzuräumen?“ „Durch welche Substituenten wird die Toxizität des Stoffes C herabgesetzt?“ Auch dies sind Fragestellungen, auf die eine jede Dokumentationsmethode die Antwort schuldig bleiben muß. Der Fragesteller wäre nicht imstande, im *voraus* festzulegen, welcher von zwei Reaktionsabläufen günstiger ist, wenn nicht nur die Ausbeute, sondern auch der apparative Aufwand, die Zugänglichkeit der Hilfsstoffe, die gewünschte Reinheit des Produktes usw. in Betracht zu ziehen sind. Die Dokumentation kann lediglich Publikationen zu diesen Themen herbeischaffen helfen, an denen der Fragesteller sich sein subjektives Urteil selbst bilden muß.

3. Die gezielte Literaturstudie

Die ungezielte Literaturstudie liefert Beiträge zum Arbeitsthema bestenfalls von einem bestimmten Zeitpunkt an. Die vorher erschienene Literatur hat man unbeachtet vorüberziehen lassen müssen. Um auch diese Literatur wieder für sich zu mobilisieren, muß man die Lektüre zwangsläufig auf einen sehr kleinen Teil der insgesamt bisher erschienenen Literatur beschränken, und diesen kleinen Teil gilt es, *gezielt* aufzusuchen.

Der Urzustand der Literatur, in seiner rein chronologischen Folge, ist hierfür ungeeignet, denn das thematisch Interessierende ist hier fast völlig regellos verstreut. Man muß also auf eine aufbereitete Form der Literatur zurückgreifen, in welcher das Interessierende mehr oder minder stark *lokalisiert* ist, entweder selbst oder in Gestalt seiner Stellvertreter, der Zitate. Diese aufbereitete Form der Literatur kann in Monographien oder Handbüchern oder einem gut geordneten Bibliothekskatalog bestehen oder auch in einem Register. Es genügt aber auch, wenn das Interessierende immer erst im Bedarfsfalle lokalisiert wird, z.B. durch Selektionsvorgänge an einer Lochkartei, an einem

[1] G. M. Corney, *Steel* 152, 20 (1963).

[2] L. H. Flett, *Chem. Week* 1960, 75.

[3] A. Neelameghan, *Library Sci.* 5, 264 (1968).

Magnetbandspeicher usw. Eine solche Literaturlaufbereitung durchzuführen, damit thematisch Einschlägiges lokalisiert oder lokalisierbar wird, ist Aufgabe der Wissenschaftlichen Dokumentation.

3.1. Terminologische Voraussetzungen

In welcher Form eine solche Literaturlaufbereitung auch immer ihren Niederschlag gefunden haben mag, zumeist ist sie von *anderen* Wissenschaftlern durchgeführt worden als von denen, die diese Hilfsmittel später für ihre gezielten Literaturstudien benutzen wollen. Auf diesen entscheidenden Umstand ist es zurückzuführen, daß die Erfolgsaussicht einer jeden gezielten Literaturstudie sehr wesentlich davon abhängt, wie weit die Terminologie eines Fachgebietes entwickelt ist. Es kommt entscheidend darauf an, wieweit es einem Autor terminologisch möglich ist, seine Vorstellungen, Ideen und Erfahrungen treffend in Worten der Fachsprache auszudrücken, und wieweit diese sprachliche Ausdrucksform von einem zweiten Fachmann so verstanden werden kann, wie es vom Autor gemeint war. Terminologische Unklarheiten können durch keinen noch so großen Einsatz an wissenschaftlichem Personal und technischen Hilfsmitteln kompensiert werden.

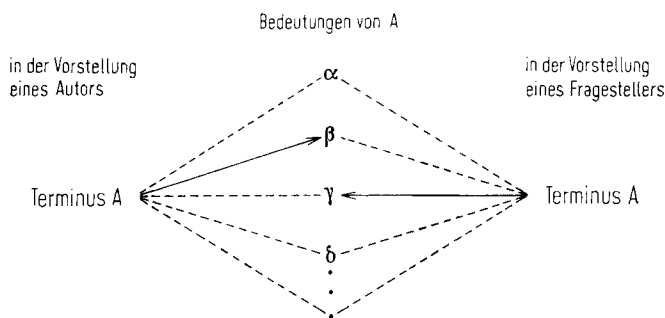


Abb. 1. Vielfalt der Bedeutungen eines Terminus A.

Als Beispiel betrachten wir den ungünstigen Fall, daß für den Terminus A viele unterschiedliche Bedeutungen α , β , γ , δ usw. verbreitet sind. Zwei Fachleute, Autor und Fragesteller, werden also unterschiedliche Vorstellungen mit dem Terminus A verbinden (Abb. 1).

Es ist leicht, beispielsweise für den Terminus „Korrosion“ ein Dutzend höchst unterschiedliche Definitionen zusammenzutragen, selbst, wenn man von der Bedeutung dieses Terminus auf medizinischem und geologischem Gebiet absieht.

Der Fachmann mit der Vorstellung γ als Fragesteller erwartet nun, daß die Literatur zum Thema γ mit dem Terminus A auffindbar ist, im einfachsten Falle also etwa unter dem Schlagwort A in einem Register. Es darf hierbei keine Rolle spielen, welche Worte ein Autor zufällig zur Beschreibung dieses Themas gewählt hat. Dieselbe Forderung wird aber auch von den Fachleuten mit den Vorstellungen α , β , δ usw. erhoben. Sie alle erwarten, daß die Literatur zu ihren — unterschiedlichen — Auffassungen unter dem Suchterminus A lokalisiert ist.

Das äußerste, was man bei dieser Sachlage vom Dokumentar verlangen kann, ist, daß er *sämtliche* Auffassungen α , β , γ , δ usw. von einem Terminus A berücksichtigt, wenn sie im Text einer Publikation erkennbar sind, und daß er sie alle mit Hilfe des gleichen Suchterminus A auffindbar macht. Dort stößt nun jeder Fragesteller, gleichgültig, welcher Auffassung er huldigt, auf Literaturhinweise, von denen er einen mehr

oder minder großen Anteil als Ballast betrachtet, denn sie entsprechen nicht seiner individuellen Auffassung vom Terminus A (z.B. γ), sondern auch den abweichenden Auffassungen α , β , δ usw. Dieser Ballast ist unausweichlich um so größer, je heterogener die Auffassungen jetzt sind und früher waren.

In der Praxis sind die Verhältnisse jedoch wesentlich ungünstiger. Es ist nämlich unmöglich, sämtliche Auslegungen für die zahlreichen Termini überhaupt nur zu ermitteln und nebeneinander ständig bei der Literaturlaufbereitung zu berücksichtigen. Es werden also etwa nur die Bedeutungen α und β zugrundegelegt, und alle anderen Auslegungen γ , δ usw. bleiben unberücksichtigt. Sie werden also auch nicht, wenn sie in einer Publikation auftreten, mit dem Terminus A gekennzeichnet. Hierdurch tritt nun für einen Teil der Literatursuchenden eine wesentlich bedenklichere, wenn auch weniger offenkundige Folge ein als nur Ballast: Verlust an einschlägiger Information.

Wenn auf einem Gebiet die *Mehrzahl* der Fachausdrücke terminologisch unklar ist, so ist es eine höchst undankbare, niemals befriedigend zu lösende Aufgabe, wenn man dort eine intersubjektive, auf Gemeinschaftsarbeit basierende Dokumentation der Fachliteratur unterhalten will. Es bleibt nur der Weg offen, daß jeder Fachmann sich eine eigene Kartei anlegt, in welcher er seine individuellen Auslegungen der Termini zugrundelegen kann. Von einer ausgedehnten Kommunikation mit der Literatur seines Fachgebietes bleibt er dann zwangsläufig ausgeschlossen, weil er bestenfalls Publikationen festhalten kann, die er selbst gelesen hat.

Das Kernstück jeder Chemiesdokumentation, das Gebiet der Stoffbegriffe, ist im wesentlichen frei von terminologischen Schwierigkeiten der geschilderten Art. Erhebliche Unklarheiten hingegen existieren in der Terminologie von Reaktionstypen. Noch größer sind die Schwierigkeiten auf den Grenzgebieten mit verfahrenstechnischen, biologischen und anwendungstechnischen Eigenschaftsbegriffen. Wenn die Kommunikation auch auf diesen Gebieten funktionieren soll, so ist noch viel terminologische Arbeit zu leisten^[*].

3.2. Der zweite Hauptsatz der Thermodynamik in der Dokumentation

Bei jeder gezielten Literaturstudie vertraut man darauf, daß im angesprochenen Teilbereich der Sammlung von Publikationen oder ihrer Stellvertreter das thematisch Interessierende möglichst gut *lokalisiert* ist, zumindest soweit es überhaupt definierbar ist. Wenn beispielsweise die Aufgabe gestellt ist, die Originalliteratur auf dem Gebiet der herbizid wirksamen Harnstoffe zusammenzustellen, so soll die Sammlung gleichsam in zwei Teile zerfallen, und in einem der beiden Teile, dem eigentlich Interessierenden, soll das Einschlägige möglichst vollständig lokalisiert sein. Außerdem soll es dort möglichst sauber befreit sein von allem Nichteinschlägigen. Je vollkommener eine solche Lokalisierung erreicht ist, desto weniger Suchzeit muß der Literatursuchende aufwenden, um in diesem Teilbereich das tatsächlich Relevante zu er-

[*] Zur größeren Klarheit sind in Abschnitt 5.1 einige Termini aus dem Dokumentationsgebiet mit den Bedeutungen zusammengestellt, in denen sie hier benutzt werden.

mitteln, und desto sicherer ist er, daß ihm nichts Relevantes entgeht, wenn er sich auf die Durchsicht des lokalisierten Teilbereiches beschränkt. Je unsicherer er in dieser Hinsicht ist, desto mehr muß er – soweit es seine Zeit erlaubt – sein Gedächtnis mobilisieren, um wenigstens diejenigen Fälle noch wiederzufinden, die ihm früher einmal an unvorhersehbaren Plätzen begegnet sind (vgl. Schema 3 in Abschnitt 5.1).

Eine solche Lokalisierung herzustellen ist gleichbedeutend mit Ordnungschaffen, wenn man unter Ordnung das „sinnvolle Beieinandersein von Teilen eines Ganzen“ versteht. Die Anzahl der Dokumentationssysteme, die für diese Zwecke entwickelt worden sind, ist verwirrend groß. Wenn man aber die thermodynamisch orientierte Definition des Ordnungsbegriffes zugrundelegt, so kann man sich relativ leicht einen Überblick über diese Systeme verschaffen und vor allem ihre höchst unterschiedliche Leistungsfähigkeit und ihre Zukunftsaussichten besser beurteilen. *All diese Systeme lassen sich treffend durch den Grad an Ordnung charakterisieren, den sie in der Literatur herzustellen vermögen, sowie dadurch, in welchem Umfang sie zu noch höheren Ordnungsgraden fortentwickelt werden können.*

Eine Sammlung von Literaturstellen kann sich in Zuständen sehr unterschiedlicher Entropie^[*] befinden, wobei der Zustand der geringsten Entropie und der größten Ordnung sozusagen die höchste Arbeitsfähigkeit aufweist^[4]. In einem sehr realen Sinne ist eine Literatursammlung produktiv und leistet gleichsam Arbeit, wenn diese Ordnung einem Fragesteller in Gestalt einer thematisch geordneten Literaturzusammenstellung in praktischer oder heuristischer Weise zugute kommt. Ein Dokumentationssystem von hoher Arbeitsfähigkeit erspart somit viel Suchzeit und Gedächtnisleistung, also Beiträge von der Art, wie sie immer nur in sehr beschränktem Umfang bei der Literatursuche beigesteuert werden können. Darüber hinaus regt es gerade durch diese Eigenschaften viele Wissenschaftler dazu an, ausgedehnte Literaturstudien zu unternehmen.

3.2.1. Die Anforderungen an den Ordnungsgrad

Wir betrachten den Fall, daß in einer Literatursammlung eine Ordnung mittleren Grades herrscht oder herstellbar ist. Es mögen sich unter den Literaturstellen, die zu einem Thema lokalisiert oder lokalisierbar sind, 20% befinden, die auf das Thema des Fragestellers einschlägig sind. Dann macht es einen wesentlichen Unterschied aus, ob sich unter zehn lokalisierten Hinweisen zwei einschlägige befinden oder ob es sich um 200 unter 1000 handelt. Aus reinem Mangel an Suchzeit würde man sich im zweiten Falle mit einigen

willkürlichen Stichproben begnügen und einen Verlust an Information in Kauf nehmen müssen.

Ist es der Literaturfachmann, der sich der Mühe unterzieht, stellvertretend für den Fragesteller, nun die lokalisierten Literaturstellen auf Einschlägiges hin zu durchsuchen, so wird er bei der Vielzahl der Aufträge, die bei ihm zusammenlaufen, bald überlastet sein. Dies gilt insbesondere dann, wenn die Sammlung unentwegt weiter wächst. Er kann Rechenaufträge nur noch in ständig abnehmender Zahl bearbeiten und ist froh über jeden Auftrag, der ihm *nicht* erteilt wird. Bald muß er darüber hinaus auch Abstriche an der Qualität machen und sich immer öfter darauf beschränken, das Material aus *Teilen* der Sammlung durchzuprüfen („die letzten Jahrgänge sind am wichtigsten“).

Bleibt diese Mühe hingegen dem Fragesteller selbst überlassen, so werden allmählich im Laufe der Zeit die Rechenaufträge versiegen, denn dies ist nach dem Mooersschen Gesetz^[6] für den Fragesteller die einfachste Art, sich einer Überlastung durch Sucharbeit zu entziehen.

Das Ende einer solchen Entwicklung besteht in beiden Fällen darin, daß das System schließlich außer Dienst gestellt wird, weil der intellektuelle und technische Aufwand, der in die Fortführung des Systems immerfort und unvermindert hineingesteckt werden müßte, nicht mehr gerechtfertigt werden kann. Bei dieser Entwicklung hat die Arbeitsfähigkeit des Systems in einem sehr realen Sinne laufend abgenommen und ist schließlich auf Null abgesunken.

Würde hingegen in der oben erwähnten größeren Sammlung eine *Ordnung höheren Grades* herstellbar sein, derart, daß die 200 einschlägigen Literaturstellen nicht mehr unter 1000, sondern vielleicht nur unter 300 lokalisiert sind, so würde man die dargebotene Information viel eher nutzen können, denn diese 300 Literaturstellen würde man im Regelfalle noch lesen und gedanklich verarbeiten können.

Hieraus folgt, daß die Anforderungen an die Ordnung in einer Literatursammlung mit deren Wachstum ständig zunehmen. Dies ist bedingt durch die begrenzte und konstant bleibende Zeitreserve des Menschen, der das System benutzt. Wenn man also das Wachstum einer Literatursammlung als einen natürlichen und sogar höchst erwünschten Vorgang ansieht, so muß man auch die Notwendigkeit anerkennen, laufend die Ordnung zu steigern, die in dieser Sammlung herrscht oder die dort hergestellt werden kann.

Besser noch stellt man möglichst frühzeitig eine Ordnung derart hohen Grades her, daß diese auch für eine möglichst lange Zeitspanne in der Zukunft noch ausreicht, um die Arbeitsfähigkeit der Sammlung zu erhalten. Es entspricht nämlich einem jeden speziellen Ordnungsgrad eine bestimmte Methode der Literaturaufbereitung und des Lokalisierens in dieser Sammlung (s. Abschnitt 4). Würde man in kurzen Abständen immer wieder die Methode ändern, um eine höhere Ordnung zu erreichen und die Arbeitsfähigkeit anzuhoben, so würde im Laufe der Jahre die Sammlung derart heterogen werden, daß sie als Ganzes kaum noch verlässlich recherchiert werden kann, denn für jeden Zeitabschnitt der Literaturaufbereitung wäre eine besondere Variante des Lokalisierens erforderlich.

Eine weitere Einflußgröße, welche neben dem Wachstum der Sammlung die Anforderungen an die Ordnung emportreibt, ist die *Häufigkeit, mit der die Sammlung zur Auskunftserteilung herangezogen wird*. Natur-

[*] „Entropie“ ist im folgenden immer im thermodynamischen, nicht im informationstheoretischen Sinn zu verstehen [5].

[4] Erich Meyer: Grundfragen der Patentedokumentation. VDI-Verlag, Düsseldorf 1960, S. 207.

[5] Vgl. W. Meyer-Eppler: Grundlagen und Anwendung der Informationstheorie. Springer-Verlag, Berlin 1969, S. 78.

[6] C. N. Mooers, Amer. Documentation 11, 204 (1960).

gemäß wird eine geordnete Literatursammlung während ihres Wachstums für den Fragesteller laufend attraktiver. Die Zeit aber, die dem Menschen zum Suchen zur Verfügung steht, kann nicht nur dadurch erschöpft werden, daß für jeden Suchgang im Laufe der Jahre immer mehr Zeit beansprucht wird. Die zunehmende Recherchenfrequenz im Laufe der Jahre, selbst bei unverändertem Zeitaufwand für jede Anfrage, kann zur gleichen Überlastung führen. Je häufiger also gesucht werden muß, desto eher lohnt es sich, dieser zunehmenden Sucharbeit durch eine entsprechend gesteigerte Ordnung entgegenzuwirken.

In welchem Umfang eine Literaturdokumentation zukunftsicher ist, erweist sich, wenn die Spezialisierungstendenz in der Wissenschaft anfängt, sich auf die Art der Fragestellungen an diese Sammlung auszuwirken. Es ist oben ein Beispiel betrachtet worden, in welchem 200 Literaturstellen, die für ein bestimmtes Thema einschlägig sind, zu ermitteln waren. Wenn man nach 20 Jahren diese Frage in unveränderter Form wiederholen würde, so wären vielleicht 2000 einschlägige Literaturstellen das Ergebnis. Ein Thema, zu welchem die Literatur heute noch durchaus übersehbar ist, kann schon in naher Zukunft einen kaum noch zu übersehenden Umfang angenommen haben. Für die Anforderungen der Steroid-Chemiker früherer Zeiten hätte es beispielsweise durchaus genügt, wenn man bei der Literaturaufbereitung jede einschlägige Publikation mit dem Schlagwort „Steroid“ charakterisiert hätte. Heute hingegen zielen die Fragestellungen auf dem Steroidgebiet spezifisch auf die exakten Positionen bestimmter Substituenten am Ringgerüst ab, auf die genaue Lage von Doppelbindungen, auf die Konfigurationen und Konformationen an Kohlenstoffatomen usw. Diese Fragen können heute nur dann beantwortet werden, wenn bereits vor Jahren vorsorglich begonnen worden ist, Methoden auszuarbeiten, zu erproben und laufend bei der Literaturaufbereitung anzuwenden, welche es heute gestatten, die Literatur nach diesen Details zu lokalisieren, ohne daß man sie erneut von Anfang bis Ende auf diese Details hin durchlesen muß. Analog ist es schon heute notwendig, *Vorsorge für solche Spezialisierungen* zu treffen und heute schon solche Details festzuhalten, die mit Sicherheit erst in der fernerer Zukunft als Suchbedingungen vom Fragesteller formuliert werden.

So besteht ein wesentliches Merkmal einer zukunftsicheren Literatur- und Patentdokumentation darin, daß sie zu jeder Zeit über erhebliche, vom Fragesteller noch nicht ausgeschöpfte Reserven an Spezifizierungsmöglichkeiten verfügt. Es wird, anders ausgedrückt, ständig *Vorsorge für Ordnungsgrade* getroffen, die mit Sicherheit erst in der Zukunft in Anspruch genommen werden und die, wenn man sie nur an den gegenwärtigen Bedürfnissen mißt, stets übertrieben erscheinen müßten.

3.2.2. Der Preis der Ordnung

Jeder Übergang zu höheren Ordnungsgraden, wie er durch die drei genannten Einflußgrößen (Speicherwachstum, zunehmende Recherchenfrequenz, Spezi-

alisierungstendenz) erforderlich wird, ist nun – thermodynamisch gesehen – untrennbar mit einem gesteigerten Aufwand verbunden. Kein Übergang zu höherer Ordnung tritt spontan ein, er muß vielmehr durch eine Energiezufuhr erzwungen werden. Andernfalls hätte man es mit einem Perpetuum mobile zweiter Art zu tun. Unausweichlich ist dieser Energieaufwand um so größer, je höher die Ordnungsgrade sind, die erreicht werden sollen.

Auf dem literarischen Gebiet besteht dieser Energieaufwand zum Ordnungschaften im wesentlichen in intellektueller Arbeitsleistung. Aus den Publikationen müssen diejenigen Begriffe und Begriffsrelationen herausgezogen werden, nach denen die Literatur später geordnet werden (können) soll. Diese Fakten müssen in eine Form gebracht werden, die bei späteren Recherchen voraussehbar ist, damit der Recherchauftrag treffsicher formuliert werden kann.

Oft soll neben dieser Arbeit noch ein Referat entstehen, welches die wesentlichen Fakten einer Publikation in übersichtlicher Form wiedergibt. — Hinzu kommen oft noch Programmierungskosten und der Aufwand für die benutzten Maschinen.

Je höher die Ordnung ist, die man in einer Literatursammlung herstellen will, desto mehr muß man für diese Arbeitsgänge aufwenden. Will man beispielsweise nicht nur Literatur zu bestimmten Verbindungen lokalisieren können, sondern auch zu bestimmten Herstellungsmethoden für diese Verbindungen, vielleicht sogar gleichzeitig noch unter bestimmten verfahrenstechnischen Bedingungen, so müssen zusätzlich auch alle reaktionschemischen und verfahrenstechnischen Begriffe aus der Literatur herausgezogen, in eine entsprechende Form gekleidet und gespeichert werden.

Liegt es schon im Wesen einer jeden Literatursammlung, die in der Zukunft noch fortgeführt werden und arbeitsfähig bleiben soll, daß sie heute bereits mit einem Ordnungsgrad arbeitet, welcher ihr erst in der Zukunft abverlangt wird, so gilt das Entsprechende auch für den zu treibenden Aufwand. Bei wenig weitblickender Betrachtungsweise wird es daher immer den Anschein haben, als ob man durch eine kleine Herabsetzung der Ordnung – zwangsläufig verbunden mit einem Ansteigen der Sucharbeit – eine sehr drastische Reduktion des Aufwandes bei der Literaturaufbereitung erzielen könne. Tatsächlich wäre zwar durchaus auf diese Weise eine momentane Kostensenkung zu erreichen. Jeder derartige Schritt würde jedoch die Zukunftsaussichten dieses Systems schmälern, da er zu niedrigeren Ordnungsgraden führt und damit die Reserven abbaut, auf die sich das System sehr bald wird stützen müssen. Es ist verfehlt, über den überraschend hohen Aufwand für eine zukunftsichere Dokumentation Klage zu führen, denn durch den ständig steigenden Nutzen eines wachsenden Systems, sofern es nur arbeitsfähig bleibt, ist der hohe Aufwand mehr als gerechtfertigt.

Es gehört zu den Besonderheiten der Arbeit auf dem Dokumentationsgebiet, daß es außerordentlich leicht zum Absinken der Arbeitsfähigkeit einer Literatursammlung kommen kann und daß die Konsequenzen einer solchen Entwicklung außerordentlich gravierend sind. Die bloße, unveränderte Fortsetzung einer Arbeitsweise, die sich anfänglich und im Kleinen durchaus bewährt hat, kann dazu führen, daß die Ergebnisse dieser Arbeit im Laufe der Zeit immer unbrauchbarer werden. Diese Entwertung betrifft rückwirkend auch diejenigen Beiträge, die man nach dieser Methode in früheren Jahren geleistet hat, und die damals noch voll befriedigt haben, als die Sammlung noch viel kleiner gewesen ist.

Kommt es schließlich wegen allzu starken Absinkens der Arbeitsfähigkeit zur Stillegung eines Dokumentationssystems, so müssen praktisch die Früchte der gesamten Dokumentationsarbeit der vergangenen Jahre verloren gegeben werden. Abgesehen von den Leistungen, die die Sammlung — wenn auch in ständig absinkendem Maße — bis zur Stillegung abgegeben hat, hat sich etwa der gleiche Zustand wieder eingestellt, als wäre in den vergangenen Jahren überhaupt keine Literaturdokumentation getrieben worden. Entsprechend groß ist dann der Verlust an Zeit, Arbeitskraft und Arbeitsergebnissen einzuschätzen.

Weitaus geringere Anforderungen an die Ordnung und damit an den erforderlichen Mindestaufwand sind bei Systemen zu stellen, die nicht wachsen und immer auf einen kleinen Umfang beschränkt bleiben werden. Diese Situation ist auf solchen Arbeitsgebieten anzutreffen, auf denen die Literatur rasch veraltet oder aus anderen Gründen uninteressant wird.

Ähnlich liegen die Verhältnisse, wenn die Literaturaufbereitung nur dazu dienen soll, dem Leser die laufenden Neuerscheinungen nicht völlig ungeordnet zuzuleiten, sondern ihm wenigstens einen annähernd gezielten Zugriff zu demjenigen Literaturbereich zu ermöglichen, in welchem er mit höherer Wahrscheinlichkeit Interessierendes erwarten darf. Derartige Dienste gibt es seit längerer Zeit für die Zeitschriften- und Patentliteratur. Für diesen Zweck reichen grobe Gliederungen aus, denn der jeweils durchzuprüfende Teil der Sammlung ist vergleichsweise sehr klein. Er umfaßt immer nur die Neuerscheinungen der jeweils letzten Ausgabe eines solchen Schnellinformationsorgans.

Es wäre aber ein Trugschluß zu glauben, daß man derartige Stoffgliederungen oder Heftregister nur mechanisch zu kumulieren brauchte, um auf kostensparende Weise zu einer brauchbar aufbereiteten großen Literatursammlung zu gelangen. Da eine solche Kumulierung nämlich nicht mit einer nochmaligen (und zwar vertieften!) Literaturaufbereitung verbunden ist, sind in der kumulierten Sammlung auch immer nur Ordnungen des unverändert niedrigen Grades möglich. Eine solche Sammlung würde sehr bald ihre Arbeitsfähigkeit einbüßen und damit für die Recherche wertlos werden.

Ebenfalls kumulierend ist auch ein Weg, der zuweilen für das weite Feld der Patentedokumentation in Betracht gezogen wird. Für die einzelnen Teilgebiete sollen funktionsfähige Spezial-Dokumentationen entwickelt werden, was nicht allzu schwierig und kostspielig ist. Das Zusammenfassen dieser Teilsysteme, rein organisatorisch und räumlich, soll dann eine umfassende Dokumentation des Gesamtgebietes ergeben. Nun kann aber die Ordnung in diesem kumulierten System niemals größer sein als es von den einzelnen Teilen her vorbestimmt ist. Für jeden dieser Teile waren aber weit aus geringere Anforderungen an die Ordnung zu erfüllen. Entsprechend gering wird auch die Ordnung im kumulierten System sein. Dort aber sind die Anforderungen um vieles höher. Als Ergebnis einer solchen Kumulierung würde kein integriertes System für das Gesamtgebiet entstehen, sondern ebenfalls ein Chaos.

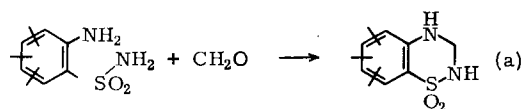
3.3. Die Art der verlangbaren Ordnung

Wir haben bisher in abstrakter Weise die Höhe der Ordnung betrachtet, welche wünschenswert oder unerlässlich ist, wenn eine Literatursammlung arbeitsfähig bleiben soll. Der Ordnungsgrad war ein Maßstab dafür, in welchem Umfang im speziell interessierenden Teil der geordneten Literatursammlung Ballast an nicht einschlägigen Publikationen anzutreffen ist und wie viele relevante Publikationen, welche in der Sammlung durchaus enthalten sind, in diesem Teil noch fehlen (Verlust). Hierbei ist zunächst noch die Frage offengeblieben, welche besonderen Vorausset-

zungen eine Publikation erfüllen soll, um als relevant für das Thema des Fragestellers bewertet zu werden. Zunächst ist nochmals festzustellen, daß ein Fragesteller nur unter einer Voraussetzung erwarten darf, die Literatur im geordneten Zustand, wie er ihm vorschwebt, anzutreffen: Er muß in der Lage sein, diese Art der Ordnung zu definieren. Ein Auftrag, der nicht definierbar ist, ist auch nicht delegierbar, nicht an einen Literaturfachmann und noch viel weniger an irgendeinen Suchmechanismus, der entsprechend programmiert werden müßte. Daß ein Teil der mangelhaft definierbaren Literaturprobleme gänzlich in den Bereich der ungezielten Literaturstudie gehört, ist bereits in Abschnitt 2 dargelegt worden.

Die Mehrzahl der chemischen Rechenaufträge aus der Praxis bewegt sich aber erfahrungsgemäß in einem ausgedehnten Grenzbereich zwischen der ungezielten und gezielten Literaturstudie. Dem Fragesteller wird, wenn er sich selbst auf die Suche begibt, diese Problematik meistens überhaupt nicht bewußt. Er kann bei den bescheidenen Hilfsmitteln, die ihm hierfür zur Verfügung stehen — meistens sind es herkömmliche Registerwerke — den interessierenden Literaturbereich ohnehin nicht eng genug abgrenzen. Er muß also im Regelfalle immer viel mehr Literatur intellektuell durchprüfen als für sein Problem relevant ist. Nach Einblicknahme in diese Publikationen ist es ihm ein leichtes, intuitiv und subjektiv zu entscheiden, ob er auch die begleitenden Suchbedingungen erfüllt sieht, die er noch nicht zur ersten Abgrenzung des Literaturbereiches benutzt hat. Hierzu bedarf es keiner besonderen vorherigen Festlegung über die Art dieser begleitenden Bedingungen.

Wird nun eine solche Literaturstudie delegiert, so besteht eine wesentliche Aufgabe des Dokumentars darin, durch Rücksprache mit dem Fragesteller zu klären, was in seinem Auftrag als Definition des Literaturproblems aufzufassen ist und was nur als Beispiel gemeint war. Bei einer Fragestellung nach Literatur zur Reaktion (a) könnte es sich beispielsweise



bei einer solchen Rücksprache herausstellen, daß nicht nur der Formaldehyd selbst, sondern auch seine Derivate (Acetale, 1,1-Halogenäther usw.) sowie Formaldehydgeneratoren (Urotropin, Methanol und Dichromat, Ameisensäure unter Disproportionierung, desgl. Oxalsäure usw.) interessieren. Auch Umsetzungen mit „niedrigen aliphatischen Aldehyden“ ganz generell und ihren Generatoren und Derivaten (Enolkörper, Epoxide usw.) kommen evtl. in Betracht. Selbst bezüglich des angegebenen Ausgangsmaterials könnte die getroffene Festlegung bedenklich sein. Nicht gefunden würden beispielsweise die Ringschlußreaktionen der acylierten Aminoverbindung oder von Acylamino-sulfonsäurechloriden, bei denen in einem einzigen Reaktionsschritt der Austausch des Halogens gegen die Aminogruppe und der Ringschluß gelingen. Als definierbares Suchziel würde hier nach nur noch der Ringschluß zum abgebildeten 1,2,4-Thiadiazinderivat verbleiben, und die so ermittelten Literaturstellen müssen a posteriori auf das Erfülltsein der begleitenden Suchbedingungen geprüft werden, am besten durch den Fragesteller selbst.

Aber selbst dann, wenn aus einem Rechenauftrag alles eliminiert worden ist, was nicht a priori festlegbar

ist, bleibt dennoch das Urteil über die Relevanz von lokalisierten Literaturstellen zuweilen problematisch. Trotzdem ist irgendeine Grenzziehung dieser Art unerlässlich. Als Relevanzkriterien sind mehrere Maßstäbe vorgeschlagen worden, z.T. sogar mit einer abgestuften Bewertung^[7]. Für unsere Betrachtungen genügt es aber, eine grobe Einteilung in relevante und nichtrelevante Publikationen vorzunehmen.

Hierzu betrachten wir als Beispiel die Fragestellung „Gewinnung von α -Aminocarbonsäuren durch fermentative Prozesse mit Mikroorganismen“.

Ein unbefangener Fragesteller erwartet von einem Dokumentationssystem, daß es ihm möglichst alle Publikationen lokalisiert (soweit sie überhaupt erfaßt sind), in welchen

- a) die gewünschten Begriffe in gleicher oder in höherer Spezifität vorkommen,
- b) die Begriffe in derselben sinngebenden (oder in noch engerer) Verknüpfung vorkommen.

Es darf hierbei keine Rolle spielen, welche Worte ein Autor zufällig gewählt hat, um seine Ergebnisse oder Ideen zum Ausdruck zu bringen. Auch braucht der Autor das Interessierende nicht unbedingt expressis verbis ausgesagt zu haben. Beispielsweise erübrigt sich bei einem „aerob“ verlaufenden Prozeß der Hinweis, daß Mikroorganismen beteiligt sind. Ist von einer Gewinnung der Aminosäuren im „Fermenter“ die Rede, einem weitverbreiteten Apparatetyp für diese Prozesse, so ist ebenfalls die Mitwirkung von Mikroorganismen genügend deutlich impliziert. Die Bildung von Aminosäuren ist bereits ausgedrückt, wenn über eine vollständig verlaufene Hydrolyse von Eiweißkörpern berichtet wird, oder über die Bestimmung der Spaltprodukte mit typischen Aminosäurereagentien usw.

Nach all diesen Kriterien würden die folgenden Themen als relevant anzusehen sein:

1. „Aufarbeitung der Kulturfiltrate von *Acetobacter suboxydans* zur Isolierung von L-Glutaminsäure als Natriumsalz“.
2. „Man züchtet einen Zucker zersetzenden Mikroorganismus in einem Nährmedium mit einer verwertbaren Stickstoffquelle und isoliert die gebildeten Aminosäuren mit Hilfe von Ionenaustauschern“.
3. „Biotechnische Gewinnung von Diaminopimelinsäure durch *E. coli* in submersen und aeroben Kulturen“.
4. „Verstärkte Bildung von freien Aminosäuren, insbesondere von Glutaminsäure, beim aeroben Abbau von Weizenstroh“.
5. „Herstellung von L-Aminosäuren aus Ketosäuren in Gegenwart von katalytisch wirkender L-Glutaminsäure mit Hilfe eines Multienzymsystems aus *Cl. butylicum* unter Zusatz von Extrakten aus *E. coli*“.

Nicht relevant im Sinne der obigen Kriterien wären die folgenden Themen:

6. „Bildung von α -Ketoglutarinsäure durch *Aerobacter aerogenes* in einem glutaminsäurehaltigen Nährmedium“.
7. „Säurebildung in Wurstmassen nach Infektion mit dem Mikroorganismus *Lactobacillus plantarum*“.
8. „Bestimmung von Aminocarbonsäuren, insbesondere von Glutaminsäure, durch Züchtung einer glutaminsäurebenötigenden *Pseudomonas*-Mutante auf einem Mangelnährboden“.

[7] Vgl. S. E. Robertson, J. Documentation 25, 93 (1969).

Sollte doch ein allgemeines Thema vom Typ 7, zusammen mit den spezifisch zutreffenden, von Interesse sein, so wäre vorsorglich die Fragestellung allgemeiner zu formulieren. Aber irgendeine Abschirmung des Fragestellers gegenüber dem allzu Allgemeinen ist unerlässlich, und am nächstliegenden ist die Grenzziehung dort, wo die gespeicherte Publikation allgemeiner wird als die Fragestellung. Andernfalls ließe sich die Lage einer solchen Grenze überhaupt nicht mehr verlässlich definieren.

Bei speziellen Recherchen auf dem Patentgebiet können die Verhältnisse allerdings gerade umgekehrt sein. Bei einer Patentlagenrecherche interessieren alle allgemeiner formulierten Ansprüche, die einen neu anzumeldenden Gegenstand umfassen. Hier sind gerade diejenigen Gegenstände von geringerem Interesse, die in spezifischerer Weise den Begriffsumfang des neu anzumeldenden Gegenstandes erfüllen.

3.4. Die Deformation einer Fragestellung

Jeder Benutzer eines Dokumentationssystems sollte sich vergegenwärtigen, daß die annähernd verlustlose und ballastfreie Zusammenstellung aller relevanten Publikationen in einer Sammlung einen Idealfall extrem hoher Ordnung darstellt. Ein Dokumentationssystem kann diesen Idealfall stets nur insoweit approximieren, als die Mittel in Gestalt von genügend qualifiziertem Personal und von technischen Hilfsmitteln hierfür verfügbar sind. Darüberhinaus muß der Benutzer sich aber auch dareinfinden, daß ein Teil seiner Erwartungen sich als generell unerfüllbar erweisen könnte, da er außerstande ist, diese Erwartungen zu definieren.

Andererseits ist der Dokumentar gut beraten, wenn er den erfüllbaren Anforderungen an sein Dokumentationssystem die denkbar größte Beachtung schenkt. Das sicherste Mittel, den Anforderungen der Praxis Rechnung zu tragen, besteht in der ausgedehnten Bedarfsanalyse im Kreise der zukünftigen Benutzer und im Bemühen, für die hierbei ermittelten Typen von Fragestellungen die Lokalisierung oder die Lokalisierbarkeit der Fachliteratur zu gewährleisten, und zwar im Sinne der obengenannten Kriterien.

Insbesondere wäre es ein schweres Mißverständnis der Aufgabe der Dokumentation, wenn die Erwartungen des Fragestellers bezüglich seiner allgemeinen Fragebegriffe nicht erfüllt werden. Um Literatur zu „Aminocarbonsäuren“ in bestimmten Zusammenhängen zu finden, sollte es nicht notwendig sein, daß der Fragesteller sämtliche denkbaren spezifischen Vertreter dieser Gruppe aufzählt oder daß er etwa sämtliche (!) denkbaren Arten von Mikroorganismen nennen muß, die im gewünschten Zusammenhang in Frage kommen können. Es ist Aufgabe des Dokumentars, dafür zu sorgen, daß die Fragestellung nach einem Allgemeinbegriff nicht nur von einem Speicherbegriff im genau gleichen Spezifitätsniveau erfüllt wird, sondern auch von allen gespeicherten Unterbegriffen zu diesem Allgemeinbegriff. Sieht sich aber der Fragesteller gezwungen, alle in Frage kommenden spezifischen Begriffe als Alternativen zu seinem Allgemeinbegriff aufzuzählen, so ist hiermit meistens eine Verfälschung oder Deformation des ursprünglichen Fragethemas verbunden, denn die Aus-

wahl der spezifischeren Begriffe, auf welche sich der Fragesteller dann notgedrungen festlegen muß, wird meistens lückenhaft sein. Hiernach kann er unter keinen Umständen mehr eine befriedigende Lokalisierung der ihn interessierenden Literatur erwarten. Das beste Ergebnis, das nach der Deformation des Fragethemas durch die erzwungene Spezifizierung überhaupt noch erzielt werden kann, ist die exakte Lokalisierung nach dem deformierten Thema, und diese muß zwangsläufig verlustbehaftet sein, verglichen mit der ursprünglichen Form der Fragestellung.

Natürlich wird man bei dem erzwungenen Ausweichen auf spezifischere Begriffe in erster Linie diejenigen berücksichtigen, bei denen die Wahrscheinlichkeit am größten ist, daß sie in der Literatur auch tatsächlich anzutreffen sind. Beispielsweise hat man sie von früheren Literaturstudien her noch in Erinnerung. Der Glutaminsäure wird man also den Vorzug geben gegenüber einer Aminostearinsäure usw. Der häufig erteilte Ratschlag aber, man müsse eben nur die „richtigen“ Unterbegriffe anfragen, ist in dieser allgemeinen Form wertlos. Er setzt bereits die Kenntnis dessen voraus, was eigentlich erst in Erfahrung gebracht werden soll: Die Kenntnis der zu diesem Thema vorhandenen Literatur und der Zusammensetzung des Speichers. A posteriori, d. h. nach Einblicknahme in den Gesamtspeicher, ließe sich eine solche Auswahl leicht treffen. Hier wird aber dasselbe Urteil a priori verlangt.

Steht umgekehrt für einen Begriff der Fragestellung wie „Aminocarbonsäure“ oder „Mikroorganismen“ im System nur ein allgemeiner Begriff zur Verfügung („aliphatische Carbonsäuren“, „primäre aliphatische Amine“ bzw. „pflanzliche Lebewesen“, „Organismen“), so tritt wiederum eine Deformation des ursprünglichen Fragethemas ein, diesmal durch eine erzwungene Verallgemeinerung. Die unausweichliche Folge ist Ballast unter den lokalisierten Publikationen, je nach dem Ausmaß der Deformation. Auch hier sind die Ursachen rein begrifflicher Natur, und keine noch so leistungsfähige Lokalisierungstechnik kann hieran etwas ändern.

Hieraus muß man die Forderung ableiten, daß es für möglichst jedes wissenschaftliche Fragethema, soweit es definierbar ist, eine möglichst genau treffende Formulierung geben sollte, unter welcher die einschlägige Literatur lokalisiert ist (Register) oder lokalisierbar ist (mechanisierte Systeme). Es muß mit anderen Worten eine möglichst große Zahl von „Fragetermini“ geben, wenn man hierunter jede Formulierung der vorgenannten Art versteht. Je größer die Zahl der verfügbaren Fragetermini ist und je differenzierter diese sind, desto größer ist die Chance, für ein ausgewähltes Fragethema einen genau treffenden Frageterminus vorzufinden, eine Grundvoraussetzung für eine verlust- und ballastarme Lokalisierung.

Zuweilen sieht man sich auch gezwungen, auf Begriffe auszuweichen, die in keiner systematischen Beziehung zum Thema des Fragestellers mehr stehen. Werden beispielsweise Herstellungswege für eine bestimmte Stoffklasse gesucht, so bedeutet die Fragestellung nach mutmaßlichen Autoren eine beträchtliche Verfälschung. Entsprechend groß sind Ballast und Verlust. — Ähnlich liegen die Verhältnisse, wenn man auf die von einem Autor gebrachten Zitate angewiesen ist. Sie verweisen naturgemäß auf begrifflich sehr heterogene Publikationen, weil auch die Arbeit des zitierenden Autors selbst aus verschiedenen Aussagen besteht, zu denen zitiert wird. Auch kann kein Autor nur annähernd vollständig die-

jenigen Vorarbeiten von Fachkollegen zitieren, an denen ein Fragesteller aus seiner Perspektive interessiert sein könnte. Was die Anzahl von verfügbaren Fragetermini anbelangt sowie ganz generell das Ausmaß, in welchem eine Deformation des Fragethemas vermeidbar ist, so sind bei den gegenwärtigen Verfahren der praktischen Chemiedokumentation die größten Unterschiede anzutreffen. Entsprechend stark unterscheiden sie sich im Ordnungsgrad, welcher mit ihrer Hilfe in der Fachliteratur erreichbar ist.

4. Praktische Verfahren zur gezielten Literaturstudie

Die Vielzahl dieser Dokumentations-Methoden läßt sich am leichtesten überblicken, wenn man sie in drei Hauptgruppen einteilt:

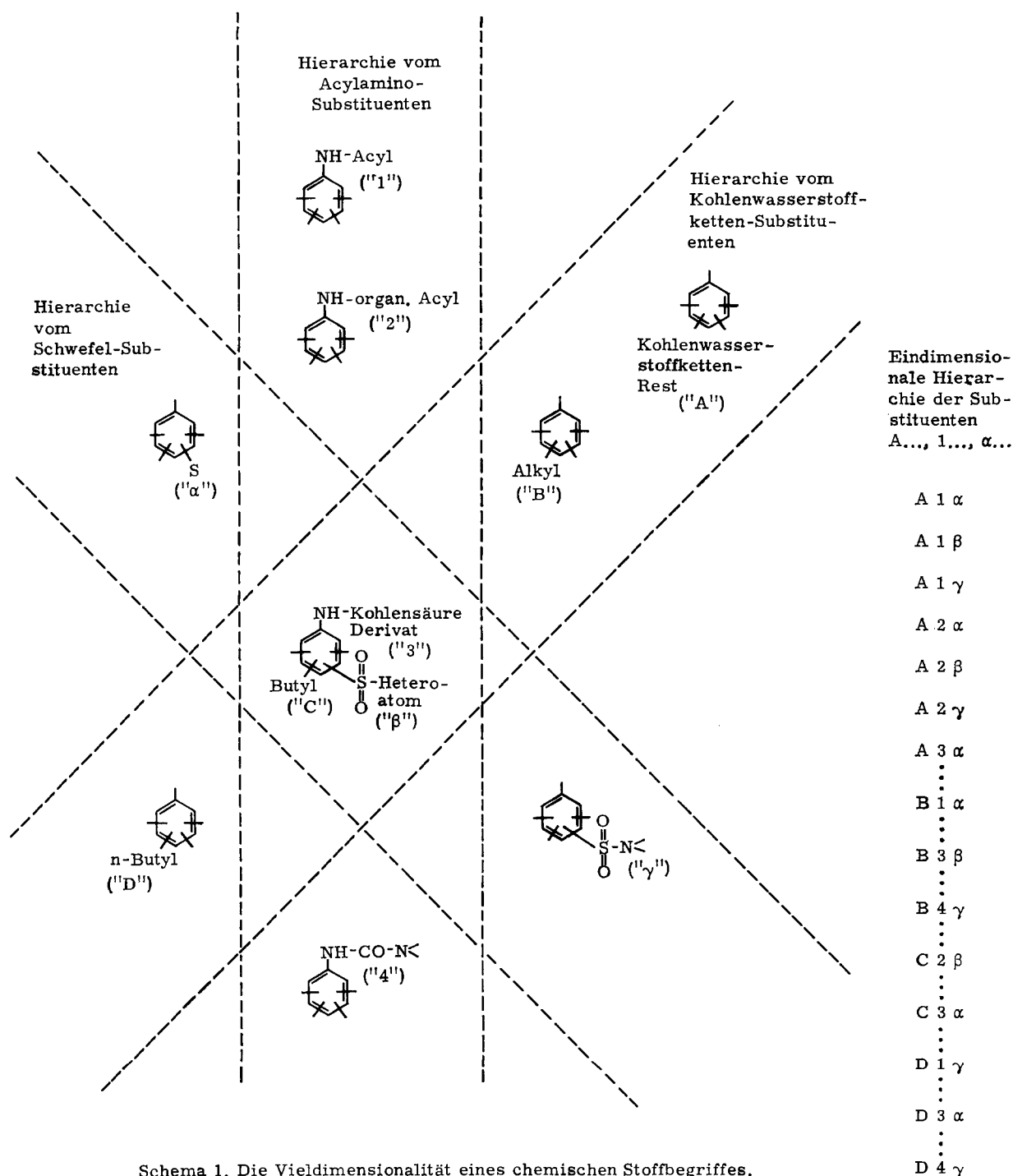
1. Die rein intellektuelle Lokalisierung,
2. die rein mechanisierte Lokalisierung,
3. die kombiniert intellektuell-mechanisierte Lokalisierung.

Die intellektuelle und die rein mechanisierte Lokalisierung sind ausgeprägte Gegenpole in der wissenschaftlichen Dokumentation mit charakteristischen Vorzügen und Mängeln. Keine dieser beiden Methoden kann zu Ordnungen höheren Grades führen, wenn sie isoliert angewendet wird, wenn auch ihre Unzulänglichkeiten ganz unterschiedlicher Art sind. Erst durch eine sinnvolle Kombination dieser beiden Methoden lassen sich ihre spezifischen Mängel überwinden; die leistungsfähigsten Dokumentationssysteme basieren heute auf einer solchen kombinierten Konzeption.

4.1. Die rein intellektuelle Lokalisierung

Bei der rein intellektuellen Lokalisierung der Literatur wird eine jede Publikation, die in den festgelegten Erfassungsbereich fällt, von einem Fachmann daraufhin geprüft, welche wesentlichen Fachbegriffe oder Themen in ihr ausgedrückt oder angedeutet sind. Hierfür braucht eine Publikation keinerlei besondere Normen zu erfüllen. Es bedeutet kein Hindernis, wenn die Publikation in einer geläufigen Fremdsprache abgefaßt ist oder wenn Handelsnamen, wissenschaftliche Nomenklatur, Bilder oder gar selbstgeschaffene, keiner Norm entsprechende Bezeichnungen für Strukturen oder Reaktionen genannt sind. Gegebenenfalls zieht der Dokumentar Nachschlagewerke zu Rate, berät sich mit seinen Kollegen oder wendet sich an eine Auskunftsstelle. So entsteht weitgehend unabhängig von den zufällig vom Autor gewählten Formulierungen ein treffendes Bild von dem, was ein Autor gemeint hat. Hiermit ist – im Gegensatz zur später zu erörternden rein mechanisierten Lokalisierung – eine Grundvoraussetzung dafür erfüllt, daß später die Literatur zu dem Gemeinten verläßlich wiedergefunden werden kann.

Daneben vergegenwärtigt sich der Fachdokumentar laufend ein weitgehend vorgegebenes Vokabular von Termini für Allgemeinbegriffe und Themen, von denen er mit einiger Wahrscheinlichkeit annehmen kann,



Schema 1. Die Vieldimensionalität eines chemischen Stoffbegriffes.

daß sie Gegenstand einer sinnvollen Fragestellung werden. Im einfachsten Fall handelt es sich um das für Sachregister charakteristische Vokabular für Allgemeinbegriffe.

Stellt der Dokumentar fest, daß eine Publikation für eines dieser Schlagwörter einschlägig ist oder für deren mehrere, so wird diese Publikation mit den entsprechenden Schlagwörtern gekennzeichnet^[8]. Damit ist der wesentlichste Schritt für die intellektuelle Lokalisierung einer Publikation vollzogen, z.B. unter den Schlagwörtern in einem Register oder einer Steilkartei oder auf den Begriffskarten einer Sichtlochkartei.

[8] C. L. Bernier, Amer. Documentation 16, 323 (1965); vgl. auch J. Koblitz: Schlagwortgebung in Information und Dokumentation. VEB Bibliographisches Institut, Leipzig 1968.

4.1.1. Mangelnde Ordnung im Vokabular der Fragetermini

Mit welchen Schwierigkeiten man bei dieser Arbeitsweise zu kämpfen hat, sei am Beispiel einer Struktur aus dem Schema 1 gezeigt. Wir betrachten den Fall, daß Literatur zu Kohlensäurederivaten von Amino-butyl-benzolsulfonsäurekörpern gesucht wird. Um in einem Sachregister das treffende Schlagwort oder wenigstens die annähernd treffenden Schlagwörter zu finden, müßte man das gesamte Vokabular sicher im Kopf haben. Fast an jeder beliebigen Stelle im Vokabular kann nämlich ein einschlägiges Schlagwort auftreten. Oftmals wird man überhaupt keinen annähernd treffenden Terminus finden. So sicher, wie es hier verlangt wird, läßt sich ein Vokabular aber nur dann

memorieren, wenn es nicht mehr als ein- bis zweitausend Schlagwörter umfaßt. Übersteigt ein Vokabular diese arbeitspsychologisch vorgegebene Grenze, so wird die Chance gering, daß man sich an die treffendsten Termini erinnert, und daß man sie auch tatsächlich benutzt, erst bei der Einspeicherung und später bei der Recherche.

Viele Sachregister imponieren heute mit einem außerordentlich großen und differenzierten Vokabular für Allgemeinbegriffe. Aber wenn man beispielsweise unter dem Schlagwort „Dien-Synthese“ in einem Jahresregister zur Weltliteratur nicht mehr als ein Dutzend zitierter Arbeiten findet, so kann man abschätzen, wie wenig verläßlich die intellektuelle Lokalisierung mit einem solchen Vokabular in Wirklichkeit ist.

Das Problem der Ordnung begegnet uns hier auf einer anderen Ebene: Soll in einer Sammlung von Publikationen Ordnung geschaffen werden, so muß zuvor im Vokabular der Schlagwörter Ordnung herrschen. Nicht aufs Memorieren und geduldige Suchen darf man sich verlassen müssen, wenn man unter den Schlagwörtern eine optimale Auswahl treffen will. Vielmehr müssen an einem voraussehbaren Platz in diesem Vokabular die für diese Auswahl in Frage kommenden Schlagwörter anzutreffen sein, und zwar möglichst lückenlos und möglichst wenig durchsetzt mit nichteinschlägigen Schlagwörtern. Wird diese Ordnung nicht hergestellt, so muß das Vokabular immer auf einen viel zu kleinen Umfang beschränkt werden, wenn es arbeitsfähig sein soll. Dies wiederum wirkt sich nachteilig auf die Lokalisierung der Literatur mit Hilfe dieses Vokabulars aus, denn mit einem allzu kleinen Vokabular ist kaum eine unverfälschte Wiedergabe eines Fragethemas möglich. So verbleibt wegen der mangelnden Ordnung im Vokabular nach der rein intellektuellen Lokalisierung noch hochgradige Unordnung in der Literatur, d.h. Verlust und Ballast.

Die tiefere Ursache für die mangelnde Ordnung in der Literatur nach der rein intellektuellen Lokalisierung liegt darin, daß es mit Hilfe eines herkömmlichen Registers, in welchem diese Art der Lokalisierung zu meist ihren Niederschlag findet, nicht möglich ist, eine ausreichend hohe Ordnung der Fachtermini abzubilden. Will man die verwandtschaftlichen Beziehungen zwischen den Fachtermini dadurch zum Ausdruck bringen, daß man sie in räumlicher Nähe zueinander anordnet, so ist dies immer nur für eine einzige Dimension von verwandtschaftlichen Beziehungen möglich. Fachbegriffe können aber über viele Dimensionen hinweg in verwandtschaftlichen Beziehungen zueinander stehen. Beispielsweise kann man von dem zentral gezeichneten Stoffbegriff im Schema 1 in verschiedenen Richtungen zu verwandten Begriffen kommen, je nachdem, ob man sich in der Teilhierarchie der Kohlenwasserstoffkette, der Acylaminogruppe oder des Schwefelsubstituenten bewegt. Allein an diesem einfachen Beispiel ergeben sich $3 \times 4 \times 4 = 48$ Stoffbegriffe, wenn man sich die freien Plätze im Schema mit weiteren Schnittpunkten der Dimensionen ausgefüllt denkt.

Wollte man nun diese miteinander verwandten Stoffbegriffe in linearer Folge anordnen, wie es in abgekürz-

ter Schreibweise am rechten Rand des Schemas angedeutet ist, so könnte man einer solchen Anordnung immer nur eine einzige Dimension als primäres Ordnungsprinzip zugrunde legen, etwa die Dimension der Kohlenwasserstoffkette. Am Anfang einer solchen Liste würden dann alle Vertreter mit dem allgemeinen Rest „Kohlenwasserstoffkette“ („A“) stehen, der Reihe nach kombiniert mit den Varianten aus den anderen Dimensionen. Es folgen alle Begriffe mit dem Substituenten „Alkyl“ („B“). Am Schluß wären alle Stoffbegriffe mit dem n-Butyl-Substituenten aufgeführt. An der Ordnung bezüglich der Dimension der Kohlenwasserstoffkette wäre nichts zu bemängeln. Dafür sind nun aber die Stoffbegriffe aus der Acylamino- (z.B. „2“) und der Schwefeldimension über die ganze Aufzählung verstreut. In der Praxis sind sie viel regelloser verstreut, als es in dem Musterbeispiel mit nur drei, gleichmäßig kurzen und jeweils vollbesetzten Dimensionen der Fall ist.

Welche Dimension unter vielen möglichen man auch immer zugrunde legt, stets werden die verwandtschaftlichen Beziehungen auseinandergerissen, die in den anderen Dimensionen herrschen. Zielt nun eine Frage auf einen Begriff aus einer der vernachlässigten Dimension ab, so steht man einem fast völlig ungeordneten Schlagwortvokabular gegenüber.

Die eindimensionale Anordnung der Begriffe beeinträchtigt nicht nur den Überblick über das Vokabular. Sie ist auch die Ursache dafür, daß es für eine Vielzahl von Begriffen überhaupt keine ungezwungene Einordnungsmöglichkeit gibt. Immer dann nämlich, wenn in der bevorzugten Dimension vom Autor oder Fragesteller überhaupt keine Aussage gemacht wird („Kernsubstituierte Carbamylaminosulfonsäureamide“ — es fehlt die Aussage über den Alkylsubstituenten —) muß man irgendeine Aussage in der bevorzugten Dimension unterstellen, damit man diesen Begriff in dem eindimensionalen Vokabular wiedergeben kann. Dies ist gleichbedeutend mit einer Verfälschung des ursprünglichen Begriffs.

4.1.2. Lokalisierung in nur einer einzigen Dimension

Die unmittelbare Folge hiervon ist, daß man im Sachregister nun auch die Literatur selbst nur nach einer einzigen bevorzugten Dimension geordnet vorfindet. Dies bedeutet keinen Nachteil, solange man nach nur einem einzigen Schlagwort recherchiert, also beispielsweise nur entweder nach „Harnstoffkörpern“ oder nur nach „Sulfonsäureamiden“ oder nur nach „Herbiziden“. Bei einer Fragestellung nach einer Kombination von Schlagwörtern, wie z.B. „Harnstoffkörper mit Sulfonsäureamidgruppen als Herbizide“, muß man unter den Harnstoffkörpern alle aufgeführten spezifischeren Schlagwörter prüfen, ob sich unter ihnen welche befinden, die die herbizide Wirkung einschließen, oder man begibt sich bei den „Herbiziden“ auf die Suche usw.

Erweist sich dies als allzu zeitraubend, so kann der Dokumentar ein neues Schlagwort in das Vokabular aufnehmen, welches dem „zusammengesetzten“ Begriff entspricht („Harnstoffherbizide“ o.ä.). In einem solchen Schlagwort sind die konstituierenden Einfachbegriffe in sozusagen ganzheitlicher, „präkoordinier-

ter“ Form wiedergegeben. Bei der Literaturanalyse muß der Dokumentar diese zusammengesetzten Schlagwörter immer zusätzlich memorieren und die einschlägige Literatur, auf welche er trifft, mit ihnen kennzeichnen.

Dieser Möglichkeit sind von zwei Seiten natürliche Grenzen gezogen. Zunächst hat die Bildung eines hochgradig präkoordinierten Schlagwortes nur dann einen Sinn, wenn es im Sprachgebrauch auch geläufig ist. Andernfalls ist es dem Fragesteller gar nicht gegenwärtig, so daß er dieses Schlagwort bei seinen Literaturstudien niemals benutzen würde. Es wäre aus diesem Grunde kaum sinnvoll, einen Terminus „Harnstoffherbizid“ künstlich zu schaffen und in das Vokabular aufzunehmen. Auch wäre der Dokumentar überfordert, wenn er bei der Literaturanalyse fortgesetzt derartige ungewöhnliche Termini im Auge behalten müßte.

Aber selbst dann, wenn es durchaus geläufige Schlagwörter für derartig hochgradig zusammengesetzte Begriffe gibt, verbleibt immer noch der unvermeidbare Nachteil eines jeden präkoordinierten Schlagwortvokabulars: Selbst bei dem äußersten arbeitspsychologisch noch vertretbaren Umfang kann es nicht entfernt die erwünschte Vollständigkeit erreichen. Wollte man beispielsweise alleine für alle praktisch wichtigen binären Koordinationen vom Typ: Stoffklasse-Eigenschaft präkoordinierte Schlagwörter bilden, so wäre ein solches Vokabular viel zu umfangreich, um bei der laufenden Literaturanalyse noch verlässlich memorierbar zu sein. Darüberhinaus bleiben die verwandtschaftlichen Beziehungen zwischen den Schlagwörtern latent, ein nicht minder schwerwiegender Nachteil. Es gäbe nämlich dann auch „Pyrimidinherbizide“, „Carbaminatherbizide“ usw., und bei einer Fragestellung nach „Herbiziden allgemein“, in bestimmtem Zusammenhang, steht man vor einem fast völlig ungeordneten Schlagwortvokabular.

Der eindimensionale Charakter des Sachregisters ist aber nicht nur die Ursache für ein lückenhaftes Vokabular von Fragetermini, sondern sogar, wenn auch in recht versteckter Weise, die Ursache für eine lückenhafte Lokalisierung der Literatur unter einem jeden Schlagwort. Dies führt zu Verlust. Es ist nämlich rein technisch nicht möglich und auch nicht sinnvoll, unter einem Schlagwort im Register *sämtliche* einschlägigen Publikationen zu zitieren. Beispielsweise wären unter „Harnstoffe“ derartig viele Strukturen aufzuführen, daß eine solche (lineare!) Zusammenstellung wegen ihres Umfangs und wegen der ihr gleichzeitig innewohnenden Unordnung praktisch keine spezifischere Fragestellung mehr erlauben würde, etwa nach „sulfurierten Phenylharnstoffen“. Damit ein solches stark belegtes Schlagwort überhaupt noch einen praktischen Zweck erfüllt, müssen die hierzu gehörigen einschlägigen Zitate gewaltsam reduziert werden. Es ist daher bei den Sachregistern üblich, daß man unter den Zitaten bei den meisten Schlagwörtern eine Auswahl trifft, und nur die typischsten Publikationen zu einem Schlagwort werden zitiert. Beispielsweise hat es sich bei einem weit verbreiteten Jahresregister zur chemischen Weltliteratur herausgestellt, daß volle 30% aller Publikationen über Phenothiazine nicht unter dem betreffenden Schlagwort zitiert waren^[9]. Diese Art des Verlustes beim Sachregister springt allerdings nicht ins Auge und ist daher weitgehend unbekannt.

So hat der Dokumentar, welcher das Schlagwortvokabular für ein Register der herkömmlichen Art zusammenzustellen hat, die Wahl, stärker mit einfachen Begriffen oder aber bevorzugt mit zusammengesetzten Begriffen zu arbeiten. Mit Schlagwörtern zu einfachen Begriffen kann er zwar die Lokalisierung für eine größere Zahl von sehr allgemeinen Themen bewerkstelligen, für spezifischere Themen aber nur mit starkem Ballastanteil. Andererseits wird es bei einem

Übergewicht der spezifischeren, präkoordinierten Fragetermini wegen der zwangsläufigen Lückenhaftigkeit eines solchen Vokabulars für viele Themen überhaupt keine Lokalisierung geben.

4.1.3. Unterteilung der Begriffe in Abhängigkeit von der Art des Registers

Charakteristisch für das Sachregister sind seine Schlagwörter für Allgemeinbegriffe, d. h. für solche Begriffe, die durch Hinzufügen weiterer begrifflicher Merkmale noch stärker spezifizierbar sind. „Carbamylaminoalkyl-benzolsulfonsäuren“ lassen sich z. B. nach der Länge der Alkylkette, nach deren Verzweigungen, nach deren Konfiguration usw. weiter spezifizieren. Je nachdem, welche Merkmale man einem solchen Allgemeinbegriff hinzufügt, in welcher Reihenfolge man sie ihm hinzufügt und um wieviele es sich hierbei handelt, fungiert ein solcher Allgemeinbegriff als Oberbegriff für viele verschieden lange Hierarchien von spezifischeren Begriffen.

Was diese Merkmale zur weiteren Unterteilung anbetrifft, so muß ihre Auswahl mit großer Sorgfalt getroffen werden. Beispielsweise wäre es aus chemischer Perspektive wenig sinnvoll, die oben genannten Phenylharnstoffe nach ihrer Verbrennungswärme, nach ihrer Diffusionsgeschwindigkeit unter Standardbedingungen oder nach der Struktur ihrer Kristallgitter zu unterteilen. Hierdurch würden für das Dokumentationssystem eines synthetisch arbeitenden Chemikers gänzlich unbrauchbare Stoffklassenbegriffe entstehen. Auch Molekulargewicht und elementare Zusammensetzung spielen als Klassenmerkmale aus dieser Perspektive eine untergeordnete Rolle, denn diese Merkmale können innerhalb einer sinnvollen Stoffklasse höchst unterschiedliche Werte annehmen. Kein Fragesteller aus der synthetischen organischen Chemie wäre geneigt, sich bei einer Recherche nach einem Allgemeinbegriff bezüglich solcher unwesentlicher Merkmale festzulegen, und kaum ein Autor würde diesbezügliche Aussagen machen. Entsprechend unbrauchbar wären für ihn solche Einteilungen. Will man also Allgemeinbegriffe (insbesondere Stoffklassen) einspeichern und mit ihnen recherchieren, so müssen zuvor die jeweils wesentlich erscheinenden Merkmale der Gegenstände ermittelt und der Unterteilung zugrunde gelegt werden (canon of relevance^[10]).

Einer völlig anderen Sachlage begegnen wir bei den Individualbegriffen. Bei ihnen hat die fortschreitende Spezifizierung der Allgemeinbegriffe eine nicht mehr überschreitbare Grenze erreicht. Sämtliche nur denkbaren Merkmale des Begriffes sind herausgearbeitet und festgelegt worden^[11]. Dies betrifft sowohl die wesentlichen als auch die unwesentlichen Merkmale. 2-Carbamylamino-5-n-butyl-benzolsulfonsäureamid läßt sich durch kein stoffliches Merkmal noch weiter

[10] S. R. Ranganathan: Prolegomena to Library Classification. Asia Publishing House, London 1967, S. 146.

[11] v. Freytag-Löringhoff: Logik – ihr System und ihr Verhältnis zur Logistik. Kohlhammer-Verlag, Stuttgart 1955, S. 27.

[9] B. Loev, J. chem. Documentation 1, 27 (1961).

spezifizieren. Auch Verbrennungswärme, Molekulargewicht und elementare Zusammensetzung liegen fest, sowohl für die Einspeicherung als auch für die Fragestellung. Unter diesen Umständen hat man freie Hand, für die Anordnung der Individualbegriffe beliebige Merkmale auszuwählen, mögen sie nun wesentlich sein oder nicht. Bei dieser Wahlfreiheit wird man möglichst einfach zu handhabende Merkmale auswählen, z. B. die elementare Zusammensetzung. Wenn man diese Merkmale schrittweise in einer beliebig festgelegten Reihenfolge anwendet (Kohlenstoff-, Wasserstoff-, Stickstoffgehalt usw.), entsteht immer nur eine einzige Hierarchie, in welcher jeder Individualbegriff zwangsläufig seinen Platz findet.

Es bereitet keinerlei Schwierigkeiten, eine rein lineare Begriffssystematik, die aus nur einer einzigen Hierarchie besteht, mit Hilfe eines (ebenfalls eindimensionalen!) Registers der herkömmlichen Art darzustellen. Aus diesem Grunde läßt sich die Literatur zu einer individuellen Verbindung mit Hilfe eines Formelregisters ohne große Mühe lokalisieren. Es wäre verfehlt, wollte man allein für diesen Zweck die erheblich aufwendigeren mehrdimensionalen, mechanisierten Techniken anwenden.

Auch bei Autoren handelt es sich um Individualbegriffe. Der Name, eventuell noch der Zeitpunkt der Geburt, ist stets gegeben bzw. feststellbar, und auf dieser Grundlage läßt sich wieder eine rein formale, eindimensionale Hierarchie aufbauen, für welche das Register die naturgegebene Darstellungsform ist.

Nach unseren Erfahrungen spielt allerdings der Individualbegriff in den literarischen Interessen der Wissenschaftler eine viel geringere Rolle als es nach der Benutzung der Autoren- und Formelregister den Anschein hat. In den weitestens Fällen nämlich, in denen eine Literaturrecherche mit einer individuellen Strukturformel oder mit einem Autorennamen aufgenommen wird oder in Auftrag gegeben wird, ist bereits eine Verfälschung des eigentlichen Fragethemas vorausgegangen. Was den Fragesteller in Wirklichkeit interessiert, ist weniger die Einzelverbindung, sondern eine ganze Stoffklasse, in welcher beispielsweise die Länge von Alkylketten schwanken kann. Die niedergelegte Strukturformel, mit welcher die Literaturrecherche begonnen wird, ist meistens mehr als Beispiel, weniger als Definition des Fragethemas gedacht. Es ist die einfache Handhabung der Formel- und Autorenregister, die den Fragesteller zur Benutzung des Individualbegriffes verführt, sowie die Schwierigkeit und Ungewißheit, im Sachregister ein treffendes allgemeines Schlagwort für ein Fragethema zu finden. So ist die praktische Bedeutung von Dokumentationssystemen und Notationen, die ausschließlich für Einzelverbindungen geeignet sind, oft überschätzt worden.

Wenn auch die Register wegen ihrer Eindimensionalität die Literatur immer nur in recht mäßigen Ordnungsgraden präsentieren können, so sind sie trotzdem von großem Wert für den Chemiker. Sie ermöglichen ihm ungezielte Literaturstudien, bei welchen er sich nicht allzu detailliert auf ein Suchziel festlegen muß. Er entscheidet vielmehr subjektiv und a posteriori, was für ihn von Interesse ist. Benutzt er die Register für diesen Zweck, so bedeutet die geringere Ordnung in ihnen keinen großen Mangel. Immerhin verfügt man auf diese Weise über ein wesentlich größeres Einzugsgebiet an fachlichen Informationen, als wenn man nur einige wenige Fachzeitschriften liest.

So werden die Register auch in Zukunft ihren Platz neben den großen mechanisierten Speichern behaupten können.

4.2. Die rein mechanisierte Lokalisierung

Die mangelhafte Lokalisierung der Fachliteratur bei der rein intellektuellen Methode führt dazu, daß das Einschlägige durch (erneute!) sachverständige Einblicknahme von der Menge des Ballastes abgetrennt werden muß. Der Arbeitsaufwand hierfür ist mit dem Wachstum der Literatur proportional mitgestiegen. Was aber nicht mitgewachsen ist, ist die Zeitreserve, die jedem Wissenschaftler für die Literaturarbeit zur Verfügung steht. Die Möglichkeit, die nach der rein intellektuellen Lokalisierung noch verbliebene Unordnung durch intellektuelle Beiträge zu kompensieren, ist demzufolge ständig geringer geworden, relativ gesehen. Aus dieser Perspektive verdienen alle mechanisierten Methoden hohes Interesse, welche die Lokalisierung der Literatur auch ohne intellektuelle Einblicknahme in die betreffenden Publikationen gestatten.

Es gehört heute zum Stand der Technik auf allen Gebieten, daß man Maschinen einzusetzen versucht, wenn der Mensch bei einer Arbeit durch die Ansprüche an das Tempo oder die Zuverlässigkeit überfordert ist. Elektronische programmgesteuerte Rechenanlagen haben auch für solche Prozesse große Verbreitung gefunden, die vordem gänzlich menschlicher Gedankenarbeit überlassen waren. Bei den zuweilen erstaunlichen Ergebnissen hat sich die Bewunderung fast ausschließlich auf die Maschine konzentriert, obwohl doch der Mensch, welcher der Maschine die Instruktionen gegeben hat, einen mindestens ebenso großen Erfolg für sich verbuchen kann. Denn der Programmierende muß vorher die Abläufe in einer sonst nur selten erforderlichen Vollkommenheit analysieren, was zuweilen eine wissenschaftliche Arbeit hohen Ranges ist. Gerät trotzdem die Maschine in eine Situation, die der Mensch nicht vorausgesehen hat, so findet sie keine Instruktion vor und kann die betreffende Aufgabe nicht lösen. *Es ist für unsere Betrachtungen wichtig, wenn auch scheinbar trivial festzuhalten, daß nur im voraus festlegbare, vom Menschen voraussehbare Abläufe programmierbar sind.*

4.2.1. Die mangelnde Voraussehbarkeit der Speichertermini

Auch in der Dokumentation hat man es beim Wiederfinden von Publikationen in einem Speicher mit einer großen Zahl von Einzeloperationen zu tun, von welchen man sich wünscht, daß sie möglichst schnell und zuverlässig ablaufen. Hat es sich bei der rein intellektuellen Lokalisierung als empfindlicher Engpaß erwiesen, daß jede einzuspeichernde Publikation vom Menschen gelesen werden muß, so liegt es nahe, auch für diesen Zweck Maschinen einzusetzen zu versuchen. Bei einem Musterbeispiel von rein mechanisierter Lokalisierung werden die Texte Buchstabe für

Buchstabe in den Speicher einer elektronischen Rechenanlage übertragen, ohne daß hierbei ein Mensch normierend oder interpretierend eingreift. Rein technisch leistet die Maschine auch das Wiederfinden eines jeden Wortes oder Wortteiles an jeder Stelle des Textes.

Die Aufgabe der wissenschaftlichen Dokumentation besteht aber nicht darin, Publikationen ausfindig zu machen, in denen die *verbale Formulierung* des Fragestellers für sein Thema wiederkehrt; es sollen vielmehr die Publikationen lokalisiert werden, die in genügend enger *thematischer Beziehung* zum Fragethema stehen^[12]. Soll nun eine Rechenanlage diese Sucharbeit übernehmen, so müssen ihr genaue Instruktionen gegeben werden, welche speziellen Formulierungen in den eingespeicherten Publikationen sie als relevant für das Fragethema betrachten soll. Diese Instruktionen müssen fortgesetzt wirksam sein, wenn die Maschine den Speicher durchforscht und hierbei die eingespeicherten Formulierungen mit dem Frageterminus vergleicht. Wird hierbei eine genaue oder genügend genaue Übereinstimmung zwischen dem „Speicherterminus“ einer Publikation und dem Frageterminus festgestellt, so wird die betreffende Publikation als relevant betrachtet. Ihre Texte oder bibliographischen Daten werden dann in einer Liste zusammengestellt oder anderweitig optisch fixiert. Damit ist die rein mechanisierte Lokalisierung der einschlägigen Literatur im Speicher vollzogen.

Als „Speicherterminus“ bezeichnen wir jede Formulierung eines gespeicherten Themas, sofern diese Formulierung zu einer mechanisierten Lokalisierung von Publikationen dient. Hierbei spielt es keine Rolle, ob es sich um Klartextwörter oder um Codierungen irgendwelcher Art handelt.

Für einen Fachmann ist es nun leicht, aus einem vorgelegten Text zu erkennen, welche Begriffe und Begriffsverknüpfungen ausgedrückt oder angedeutet sind. Dies ist auch der Weg, welcher bei der rein intellektuellen Lokalisierung immer beschritten wird. Das Urteil, welches hier verlangt wird, hat gleichsam *aposteriorischen* Charakter: Immer erst *nach* Einblicknahme in eine fragliche Formulierung braucht entschieden zu werden, welche begriffliche Bedeutung diese Formulierung hat.

Unsere Fachsprache hat in erster Linie die Aufgabe zu erfüllen, wissenschaftliche Fakten möglichst präzise und lebendig darzustellen. Jedem einzelnen beläßt sie einen weiten Spielraum, welche Ausdrucksweise er hierbei wählt und welchen persönlichen Stil er entwickelt. Gerade eine solche Sprache ist aber hochgradig ungeeignet, wenn man auf ihrer Grundlage relevante Literatur auf rein mechanisiertem Wege lokalisieren will. Das Urteil über die Beziehung zwischen einer speziellen Ausdrucksweise auf der einen Seite und andererseits dem, was mit diesem Ausdruck gemeint ist, muß nämlich nun in genau umgekehrter Richtung vollzogen werden als dies üblicherweise und *aposteriorisch* in der wissenschaftlichen Kommunikation geschieht. Gegeben ist ein Begriff, und *a priori*, ohne eine Einblicknahme in relevante Publikationen, muß beurteilt werden, welche Ausdrucksweisen für diesen Begriff denkbar sind.

All diese Formulierungen müssen im Speicher vermutet werden und müssen demzufolge im Frageterminus im voraus berücksichtigt werden, denn dies gehört zu den Instruktionen, die die Maschine für die Relevanzprüfung an den eingespeicherten Publikationen benötigt.

Nun weiß man, daß es sehr viel leichter ist, ein eingetretenes Ereignis nachträglich zu beschreiben, als dieses Ereignis vorausszusagen. Was die rationale Voraussagbarkeit eines Ereignisses anbetrifft, z. B. das Formulieren eines Begriffes mit den Wörtern der natürlichen und der fachlichen Sprache, so ist sie überhaupt nur dann gegeben, wenn die Gesetzmäßigkeiten im Ablauf dieses Ereignisses bekannt sind. Bis heute ist aber noch keine Gesetzmäßigkeit erkannt worden, welche eine Voraussage gestatten würde, in welchen Wörtern ein Mensch eine Idee oder ein experimentelles Ergebnis in seiner natürlichen Sprache zum Ausdruck bringen wird. Sollte es tatsächlich eine solche Gesetzmäßigkeit geben, so sind wir jedenfalls von ihrer Kenntnis heute noch sehr weit entfernt. Rein statistische Regelmäßigkeiten, die man im Sprachgebrauch von Autoren gelegentlich feststellen kann, helfen hier nicht weiter^[13]. Sie besitzen keine Gültigkeit für den Einzelfall, lassen also keine Voraussage darüber zu, welche Wörter ein Autor beispielsweise bei seiner ersten Patentanmeldung zu diesem Thema gewählt haben könnte. Gerade auf den Einzelfall kommt es aber in der wissenschaftlichen Dokumentation an.

Es ändert sich nichts am apriorischen Charakter solcher Festlegungen für den Frageterminus, daß diejenigen Publikationen bereits existieren, die Auskunft darüber geben könnten, welche Speichertermini für den Begriff der Fragestellung bereits vorgekommen sind und demzufolge in der Festlegung des Frageterminus in Betracht zu ziehen sind. Entscheidend allein ist, daß diese Publikationen dem Fragesteller unbekannt sind und daß er niemand kennt, den er danach fragen kann oder fragen will.

Wenn allerdings ein Fragesteller die Publikationen, die er sucht, früher selbst gelesen hat und wenn er sich auch noch genau an die Einzelheiten der sprachlichen Formulierung erinnert, dann steht ihm durchaus der *aposteriorische* Weg zur Formulierung des Frageterminus offen. Man muß aber bedenken, wie klein heute der Anteil an Literatur geworden ist, der dem Wissenschaftler durch eigenes Lesen noch zu Gesicht kommt. Noch viel kleiner ist der Bruchteil hiervon, den er sich so detailliert gemerkt hat, wie es bei der rein mechanisierten Lokalisierung verlangt wird. Auf diesem Wege läßt sich also immer nur ein verschwindend geringer Bruchteil von Literatur wiederfinden, welcher für ein Fragethema relevant ist.

Vor welche begrifflichen Probleme man bei der rein mechanisierten Lokalisierung gestellt ist, kann man sich leicht an der obigen Zusammenstellung (s. Abschnitt 3.3) von Themen zur Frage nach „Gewinnung von α -Aminocarbonsäuren durch fermentative Prozesse mit Mikroorganismen“ vergegenwärtigen. Man wird sich vergeblich bemühen, irgendeine allgemeine Regel zu finden, nach welcher ein formal arbeitender Suchmechanismus erkennen könnte, welche der aufgeführten Speicherthemen für das Fragethema relevant sind und welche nicht. Immer wird man müheelos Beispiele konstruieren oder in der Literatur finden können, bei denen eine solche Regel versagt, selbst wenn man noch so viele Verfeinerungen an ihr vorgenommen hat.

[12] C. L. Bernier, J. chem. Documentation 2, 117 (1962).

[13] M. E. Lesk, Amer. Documentation 20, 27 (1969).

Auch bei Fragestellungen nach Strukturen allein liegen die Verhältnisse nicht wesentlich günstiger, obwohl hier durch die Vorschriften der Nomenklatur die Benennungen besser überschaubar sind. Es ist beispielsweise der Versuch unternommen worden, Literatur zu Schiffchen Basen, die in einem bestimmten Zusammenhang genannt sein sollen, zu ermitteln^[14]. Für den strukturellen Teil der Fragestellung ist es als ausreichend erachtet worden, nach der Zeichenfolge „imin“ oder „Schiff base“ (es handelte sich um Speichertexte in Englisch) oder „azomethin“ zu suchen. Tatsächlich fällt es schwer, durch bloßes Überdenken weiterer Benennungsmöglichkeiten einen solchen Suchauftrag noch nennenswert zu vervollständigen.

Unvermeidbar bleiben bei einer solchen apriorischen Festlegung des Frageterminus eine Reihe von Verbindungsamen unberücksichtigt, teils weil sie — obwohl durchaus geläufig — durch bloßes Nachdenken allein niemals vollständig zusammengestellt werden können, teils weil sie gänzlich unvorhersehbar sind (z. B. „Anil“, „Hydrobenzamid“, „Benzyliden...amin“, „Azallylverbindungen“ usw.). Als Folge hiervon tritt Verlust an relevanter Information ein. Dieser Verlust ist allerdings kaum anders aufzudecken als durch Parallelrecherchen mit leistungsfähigeren Systemen und durch mühsames Vergleichen der Ergebnisse. Hierzu kommt es jedoch nur in den seltensten Fällen. So bleibt dieser Mangel längere Zeit unerkannt, und das erste Urteil über die Leistungsfähigkeit der rein mechanisierten Lokalisierung fällt regelmäßig viel zu günstig aus.

Wesentlich leichter zu beurteilen ist der Ballast bei diesem Verfahren, denn er tritt bei jeder Lokalisierung offen zutage^[15]. In diesem Falle wird er durch Wörter wie „Iminoäther“, „Kalimineralien“, „Eliminierungsreaktion“, „Ketenimine“, „Äthylenimin“ ausgelöst sowie durch viele Trivialbezeichnungen wie „Capaurimin“ usw. Es werden unerwünschterweise z. B. auch sämtliche Imine erfaßt, die am Stickstoff nicht durch Kohlenstoff substituiert sind.

In vielen Fällen würde eine Literaturrecherche, ausschließlich basierend auf *Namen* für Strukturen und Partialstrukturen, schon allein deswegen ins Leere stoßen, weil für viele publizierte Strukturformeln überhaupt keine Verbindungsamen angegeben werden. Dies gilt z. B. auch für alle allgemein gehaltenen Patentansprüche.

Die tiefere Ursache für diese Fehlschläge liegt darin, daß man bei der rein mechanisierten Lokalisierung keinen Unterschied macht zwischen dem Wort und seiner Bedeutung. Man begnügt sich damit, bei der Fragestellung nach einem *Begriff* lediglich diejenigen *Wörter* in den gespeicherten Publikationen aufzusuchen, mit denen dieser Begriff rein zufällig ausgedrückt worden ist. Hier holt man auf dem Dokumentationsgebiet die ganze Verwirrung nach, die in der früheren Literatur zur Logik in analoger Weise zu verzeichnen war, als man noch nicht zwischen Objektsprache und Metasprache unterschieden hat^[16]. Wird aber — um hier Abhilfe zu schaffen — die Forderung gestellt, auch alle anderen denkbaren Formulierungen in Betracht zu ziehen, nicht nur die vom Fragesteller gewählte, so setzt man wiederum dasjenige als bereits bekannt voraus, was eigentlich erst in Erfahrung gebracht werden soll: Die Kenntnis der einschlägigen Publikationen im Speicher mit allen Details ihrer Formulierung. Das stillschweigende Postulat von der Voraussehbarkeit der Speichertermini, auf welchem die rein me-

chanisierte Lokalisierung letztlich basiert, steht somit auf schwachen Füßen. Es ist nicht verwunderlich, daß sich auf diese Weise bisher immer nur Ordnungen von bescheidenen Graden erreichen ließen, d. h. Lokalisierungen mit beträchtlichen Anteilen von Verlust und Ballast. — Allerdings kann selbst dies schon einen Fortschritt bedeuten. Auch läßt es sich nicht leugnen, daß zuweilen die intellektuelle Lokalisierung recht unvollkommen durchgeführt wird, insbesondere dann, wenn sie sich nach den beschränkten Möglichkeiten einer eindimensionalen Technik orientieren muß. In solchen Fällen kann die rein mechanisierte Methode zuweilen durchaus konkurrieren. Dies scheint — zumindest im Experiment und in kleinem Maßstab — insbesondere dann der Fall sein, wenn die Maschinenarbeit auf einer extrem sorgfältigen Sprachanalyse basiert^[17, 18].

Wie hoch sind die Aussichten einzuschätzen, daß bei entsprechender Weiterentwicklung der mechanisierten Methode eines Tages die Rechenanlage den Menschen vollwertig bei der Literaturanalyse ersetzen kann? Die obigen Überlegungen haben gezeigt, welche geradezu phantastischen Anforderungen in diesem Fall an die Maschine (und an die Menschen, die sie programmieren müßten!) zu stellen wären. Die Rechenanlage müßte einen jeden Text, auf den sie beim Durchforschen ihres Speichers trifft, in sachkundiger Weise daraufhin beurteilen können, ob er für das Fragethema relevant ist oder nicht. Aber lassen sich wirklich all die vielen Situationen *voraussehen* und in einen Maschinenalgorithmus fassen, in den der Mensch die Implikationen eines Autors interpretieren muß? Nur so werden doch beispielsweise die Ausgangsstoffe auffindbar, wenn ein Autor sich mit der Angabe des Endproduktes und der eindeutig verlaufenden Herstellungsreaktion begnügt hat. Nur dann kann man eine Bekämpfungsmethode für einen menschenpathogenen Parasiten auffinden, wenn der Autor lediglich die Ausrottung eines typischen Zwischenwirtes für diesen Parasiten beschrieben hat, weil er es dem Leser überlassen konnte, hieraus die sachkundigen Schlußfolgerungen zu ziehen.

Der Wissenschaft wäre wenig damit gedient, wenn man durch den kritiklosen Übergang zu rein mechanisierten Dokumentationsmethoden das Leistungsniveau der wissenschaftlichen Dokumentation gewaltsam senkte. Überall ist man bestrebt, der Forschung immer leistungsfähigere Werkzeuge zur Verfügung zu stellen, und zwar aus guten Gründen. Sollten ausgerechnet in der wissenschaftlichen Dokumentation die entgegengesetzten Maximen gelten?

Ist man sich aber sehr wohl der Schwächen der rein mechanisierten Lokalisierung bewußt und hält trotzdem ihre Anwendung in der praktischen Großdokumentation der chemischen Zeitschriften- oder gar Patentliteratur für vertretbar, so dürfte eine Unterschätzung des Ordnungsgrades vorliegen, welcher in einem solchen Speicher jetzt oder in der nahen Zukunft erreichbar sein muß, wenn er arbeitsfähig bleiben soll. Ein solcher Weg führt unweigerlich in die Sackgasse^[19, 20].

Werden also hohe Anforderungen an die Ordnung in einer Literatursammlung gestellt, so müssen besondere

[14] Kallner, Vortrag am 21. Mai 1969 an der Universität Köln.

[15] D. S. Colombo u. J. E. Rush, J. chem. Documentation 9, 47 (1969).

[16] Vgl. W. Stegmüller: Wissenschaftliche Erklärung und Begründung. Springer-Verlag, Berlin 1969, S. 30.

[17] G. Salton, Proc. 19th Annual Conference of the Assoc. for Computing Machinery, Philadelphia 1964.

[18] J. O'Connor, J. Assoc. Computing Machinery 12, 490 (1965).

[19] M. Woitschach, Nachr. Dokumentation 17, 74 (1966).

[20] R. Fugmann, Nachr. Dokumentation 17, 79 (1966).

Anstrengungen unternommen werden, damit die Speichertermini voraussehbar werden. Welcher Art diese Maßnahmen im einzelnen sein können, wird in Abschnitt 4.3.1 behandelt.

4.2.2. Lokalisierung in vielen Dimensionen

Ein Mangel der rein intellektuellen Lokalisierung war darin erblickt worden, daß im Vokabular der Schlagwörter Ordnung immer nur in einer Dimension herzustellen ist und daß demzufolge auch die Literatur immer nur in einer einzigen Dimension lokalisiert werden kann. Will man nun die vieldimensionalen Begriffsbeziehungen in Schema 1 ohne Verzerrungen wiedergeben, so muß man sich nur einer ebenfalls vieldimensionalen Wiedergabeweise bedienen.

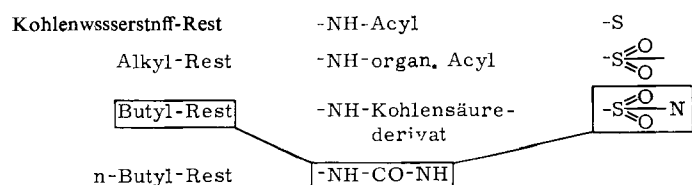
In den meisten chemischen Stoffbegriffen kann man – allein schon aufgrund ihrer Definition – eine Reihe von begrifflichen Bausteinen erkennen, nach denen man *isoliert* fragen können möchte, oder auch *in Kombination* mit beliebigen anderen Bausteinen. Häufig treten diese Bausteine auch isoliert als Allgemeinbegriffe in der einzuspeichernden Literatur auf („Harnstoffe“, „Sulfonamide“ usw.). Viele dieser Bausteine sind so eng miteinander verwandt, daß man sie in eine gemeinsame Hierarchie von wechselseitig über- und untergeordneten Begriffen einordnen kann. Jede Hierarchie entspricht nun einer der Dimensionen, in welche man einen zusammengesetzten Stoffbegriff zu zerlegen hat, wenn man ihn in all seinen Dimensionen darstellen will. Für das Beispiel vom Schema 1 würde sich etwa das folgende Bild ergeben (Schema 2).

Zum bestpassenden Frageterminus für ein Fragethema gelangt man nun sehr verläßlich und auf einfache

Trotz des großen Umfanges ist der Überblick über ein so großes Vokabular keinen Augenblick in Frage gestellt. Die Zahl der Einfachbegriffe, aus denen die kompliziertesten Fragetermini zusammengesetzt werden können, bleibt klein und ist leicht memorierbar.

Die Zerlegung von zusammengesetzten Begriffen in einfachere ist hier am Spezialfall des chemischen Stoffbegriffs aufgezeigt worden. Es handelt sich hier um ein wohl bekanntes und bewährtes philosophisches Prinzip, welches generell in der Dokumentation eine große Rolle spielt. Letztlich beruht eine jede Systematik in der Wissenschaft auf dieser analytischen Betrachtungsweise. Man kann ein solches Verfahren noch auf Axiomen und auf klassifikationstheoretischen Prinzipien gründen und auf Begriffsverknüpfungen höherer Stufen anwenden. Dann spricht man von einer analytisch-synthetischen Klassifikation, um deren Entwicklung sich vor allem *Ranganathan* und seine indische Schule verdient gemacht haben [21, 22].

Durch diese Zerlegung eines zusammengesetzten Begriffes in seine Einfachbegriffe sind die Voraussetzungen dafür geschaffen worden, daß die Fachliteratur nun nach jedem Einfachbegriff lokalisiert werden kann, insbesondere aber auch nach einer jeden beliebigen *Koordination* von solchen Einfachbegriffen. Wenn man einen jeden dieser Einfachbegriffe auf einem geeigneten Träger darstellt (durch eine Begriffskarte bei der Sichtlochkartei, durch eine Lochposition auf einer Lochkarte, durch ein typisches Magnetisierungsmuster auf einem Magnetband usw.), so kann man durch eine geeignete Selektionstechnik alle Publikationen lokalisieren, in denen *sämtliche* Dimensionen der Fragestellung erfüllt sind, und nicht etwa nur eine einzige. Es werden also nur noch solche Publikationen als Ergebnis der Lokalisierung präsentiert, in denen etwa die Begriffe „Harnstoffkörper“ und „Alkylbenzol“ und „Halogenbenzol“ und „herbizide Wirkung“ vorkommen, nicht mehr aber auch die zahl-



Schema 2. Hierarchisch geordnete dreidimensionale Begriffsliste für das Beispiel aus Schema 1.

Weise dadurch, daß man aus jeder Einzelhierarchie für eine Dimension den treffenden Einfachbegriff auswählt und ihn mit einem ebenso frei wählbaren Einfachbegriff aus einer anderen Dimension nachträglich koordiniert. In Schema 2 ist beispielsweise das Carbamylamino-butyl-benzol-sulfonsäureamid durch eine solche „Postkoordination“ der Einfachbegriffe (durch Verbindungsstriche) dargestellt. Durch diese Postkoordination von Einfachbegriffen, die man zuvor durch gedankliche Analyse gebildet hat, werden Fragetermini in sehr großer Zahl verfügbar.

Aus diesem Schema kann man auch die Fragetermini für eine jede Kombination entnehmen, die aus einer *geringeren* Zahl von Dimensionen zusammengesetzt ist. – Analoge Erleichterungen ergeben sich für die Einspeicherung, wenn man auch dort eine Zerlegung in die gleichen Dimensionen durchführt.

reichen anderen Publikationen, in denen nur ein einziger Begriff aus der genannten Gruppe vorkommt, wie man dies bei der eindimensionalen Lokalisierung in Kauf nehmen müßte. Hierdurch sind bei der mechanisierten vieldimensionalen Lokalisierung ungleich höhere Ordnungsgrade erreichbar.

4.3. Die kombiniert intellektuell-mechanisierte Lokalisierung

Zieht man eine Bilanz über die spezifischen Vorzüge und Schwächen der beiden Lokalisierungsverfahren, so ergibt sich das in Tabelle 1 in den beiden ersten Zeilen gezeigte Bild.

[21] S. R. Ranganathan, Library Sci. 6, 193 (1969).

[22] A. Neelameghan, Herald Library Sci. 8, 96 (1969).

Tabelle 1. Leistungsfähigkeit von drei Lokalisierungsmethoden. Die jeweilige Überlegenheit einer Methode ist durch ein + angedeutet.

Leistung in Bezug auf Methode der Lokalisierung	Erkennung des Gemeintem	Chance für treffenden Frage-terminus	Aufwand an Lesezeit nach Lokali-sierung	Vorausseh-barkeit der Speicher-termini
rein intellektuell	+	—	— [a]	+ [b]
rein mechanisiert	—	+ [c]	+	—
kombiniert	+	+ [c]	+	+

[a] Hoher Aufwand zum Abtrennen des Relevanten vom Ballast.

[b] Die Voraussehbarkeit ist durch die Übersetzung in das vorgegebene Vokabular der Schlagwörter gegeben.

[c] Durch Postkoordination.

Offenkundig können sich die beiden ersten Methoden sehr wirksam gegenseitig ergänzen, so daß ihre Vorzüge erhalten bleiben, ihre spezifischen Schwächen jedoch überwunden werden. Bei einer solchen kombinierten Konzeption (Zeile 3 in Tabelle 1) hätte die rein intellektuelle Methode bei der gedanklichen Analyse der zu speichernden Publikationen einzusetzen. Das vom Autor Gemeinte wird erkannt, gleichgültig, ob es nun in Worten oder in Bildern explizit ausgedrückt worden ist oder ob es nur impliziert wurde. Alles, was Gegenstand einer sinnvollen späteren Fragestellung sein kann, wird für die Einspeicherung ausgewählt, soweit es die technischen und wirtschaftlichen Möglichkeiten erlauben.

Ein weiterer wesentlicher intellektueller Beitrag besteht darin, daß die Form dieser Aussagen im Speicher voraussehbar gemacht wird. Hierzu können Klartextschlagwörter für Begriffe und Begriffsverknüpfungen dienen, entweder eigens zu diesem Zweck gebildete oder solche, die unter bestimmten Bedingungen aus dem Originaltext übernommen werden. Je nach den verfügbaren technischen Hilfsmitteln werden diese Schlagwörter gespeichert. Wichtig ist lediglich, daß die Speichertermini in dieser Form vom betreffenden Hilfsmittel erkennbar sind und daß sie dadurch zu einer mechanisierten Lokalisierung der so gekennzeichneten Publikationen dienen können, einzeln oder in Postkoordination. So kommen nach der intellektuellen Vorbereitung die Vorzüge der mechanisierten Methoden voll zum Tragen.

Es ist verschiedentlich der Versuch unternommen worden, auch dadurch zu einer kombiniert intellektuell-mechanisierten Lokalisierungsmethode zu gelangen, daß man nachträglich die Schlagwörter eines herkömmlichen Registers auf mechanisch abfragbare Träger übernommen hat. Man muß aber immer bedenken, wie große Konzessionen an die begriffliche Sauberkeit der Schlagwörter gemacht werden mußten, wenn sie für ein Register bestimmt waren, und wie unzuverlässig auch ihre Zuordnung zu den Publikationen ist. Dies gilt insbesondere durch die hochgradige Präkoordination, die in diesen Schlagwörtern verwirklicht ist. Übernimmt man solche Schlagwörter unbeesehen in eine Technik, welche die begriffliche Postkoordination durchaus leistet, so wäre die Leistungs-

fähigkeit der mechanisierten Technik nicht entfernt ausgenützt. *Es gibt kein Schlagwortvokabular, welches gleichermaßen sowohl für die Arbeit mit einem Register als auch mit einem mechanisierten System geeignet ist.*

4.3.1. Wege zur Voraussehbarkeit der Speichertermini

Dokumentationsmethodisch gesehen sind es zwei verschiedene Aufgaben, je nachdem, ob man die Voraussehbarkeit der Ausdrucksweisen im Speicher für *Begriffe selbst* anstrebt oder aber die Voraussehbarkeit der besonderen *Art der Verknüpfungen*, in denen die Begriffe in den Publikationen vorkommen.

Zwar lassen sich oft verschiedene Standpunkte dafür rechtfertigen, ob man einen bestimmten Vorstellungsinhalt als einen einzigen Begriff auffassen soll („Harnstoffherbizid“) oder besser als eine Verknüpfungen von gedanklich separaten Begriffen („Harnstoffkörper“; „Herbizide Wirksamkeit“). Hat man aber einmal für ein bestimmtes Fachgebiet eine Festlegung getroffen, nach welchen Begriffskategorien die Fachliteratur lokalisierbar gemacht werden soll („Stoffe“, „Reaktionen“, „Eigenschaften“ usw.), so ist damit auch gleichzeitig eine Richtlinie dafür geschaffen, was überhaupt als ein zusammengesetzter Begriff zu betrachten ist und wie weit die Zerlegung zusammengesetzter Begriffe vorangetrieben werden muß. Es dürfen nämlich hiernach nur noch solche Begriffe als Einheiten bestehen bleiben, welche Vorstellungsinhalte aus nur noch einer einzigen Kategorie enthalten („kategorieine Begriffe“). Andernfalls würde man dulden, daß in der Systematik ein bestimmter Begriff aus einer Kategorie regellos verstreut bleibt, je nachdem, mit welchem anderen Begriff er zufällig vergesellschaftet ist. Das Zusammenfügen der durch diese Zerlegung gebildeten Einfachbegriffe zu den Begriffen, wie sie vom Fragesteller gewünscht werden, muß die Postkoordination leisten.

4.3.2. Speichertermini für Begriffe

Der erste Schritt zur Voraussehbarkeit des Speicherterminus für einen Einfachbegriff besteht darin, daß man – im Gegensatz zur voll mechanisierten Methode – den Originaltext der Publikationen nicht mehr völlig unkontrolliert in den Speicher einfließen läßt. Man führt vielmehr in einem „Thesaurus“^[23] laufend Buch darüber, welche verschiedenen Ausdrucksweisen für einen bestimmten Begriff bisher schon vorgekommen sind und in den Speicher Eingang gefunden haben. Im Falle einer Fragestellung vergewissert man sich zuvor, welche Ausdrucksweisen für die Begriffe des Fragethemas und für die verwandten Begriffe bisher im Speicher vorgekommen sind. All diese zählt man auf und formuliert sie als alternative Suchbedingungen in einem entsprechend umfangreichen Frageterminus („manipulierte Gruppenbildung“)^[24].

Die Wirksamkeit dieser Methode ist am größten, solange der Speicher noch relativ klein ist, d. h. so lange noch nicht allzu viele verschiedene Termini für einen Begriff und seine Verwandten vorgekommen sind. Im Laufe der Zeit aber müssen immer mehr Speichertermini für jeweils einen Begriff berücksichtigt werden. Außerdem wächst wegen der bald einsetzenden Tendenz zu stärker spezifizierten Fragestellungen die Zahl der Begriffe, für welche in jeder einzelnen Fragestellung

[23] D. Soergel: Klassifikationssysteme und Thesauri. Deutsche Gesellschaft für Dokumentation, Frankfurt/M. 1969.

[24] R. Fugmann, Nachr. Dokumentation 13, 120 (1962).

diese Maßnahme durchzuführen ist. Leicht kann dann die Arbeit, alle diese Alternativen festzulegen und zu handhaben, für Mensch und Maschine untragbare Ausmaße annehmen.

Es bedeutet nämlich keinen echten Ausweg, all die vielen Alternativen für einen Fragebegriff ein für alle Male festzulegen und der Maschine als fest vorgegebene Instruktion mitzuteilen, so daß sie für einen aufgegebenen Fragebegriff automatisch alle Alternativen in den Frageterminus einbezieht. Das Berücksichtigen und Aufsuchen von sehr vielen Alternativen kann einen nicht vertretbaren Aufwand an Maschinenzeit bedeuten, insbesondere dann, wenn man mit der Anzahl der Alternativtermini je Fragebegriff in die Größenordnung oberhalb von tausend gerät.

Vorteilhaft bei diesem Verfahren ist hingegen, daß man sehr frühzeitig über einen recherchierbaren Speicher verfügt, ohne daß man erst den langwierigen Aufbau einer Begriffssystematik abwarten muß. Diese Systematik kann vielmehr unabhängig und parallel zur Einspeicherung entwickelt werden. Hierbei kann man sich auch laufend nach den praktischen Bedürfnissen orientieren, denen man beim Einspeichern und Recherchieren begegnet.

Ein weiterer Schritt zu verbesserter Voraussehbarkeit der Speichertermini besteht darin, daß man für einen Begriff nur noch einen einzigen Terminus im Speicher zuläßt und alle anderen bedeutungsgleichen Ausdrücke (Synonyma) durch diesen einen Standardterminus ersetzt^[25].

Zu diesem Zweck kann man alle Texte vor der endgültigen Einspeicherung maschinell daraufhin durchforschen lassen, ob in ihnen derartige zu eliminierende Synonyma vorkommen. Die Maschine hat den Auftrag, solche Termini durch die Standardform zu ersetzen. Stößt sie hierbei auf Termini, für welche sie noch keine Umsetzungsvorschrift hat, so wird eine entsprechende Nachricht herausgegeben, wonach der Mensch die Instruktionen entsprechend vervollständigen kann. Schwierigkeiten entstehen bei diesem Verfahren dadurch, daß es Termini gibt, die je nach dem Zusammenhang, in dem sie vorkommen, gänzlich verschiedene Bedeutungen annehmen können, z.B. Zelle, Öl, Plasma („Homonyma“). Sie müßten also in unterschiedliche Standardtermini umgewandelt werden. Außerdem ist noch immer die sichere Erkennung des vom Autor Gemeinten infragegestellt, denn die intellektuelle Textanalyse wird nicht generell durchgeführt, sondern nur in den von der Maschine gemeldeten Bedarfsfällen. Manche Implikation, die der Mensch leicht zwischen den Zeilen herauslesen könnte, bleibt hierbei unerkannt.

Diesem Mangel kann nur dadurch abgeholfen werden, daß man davon abgeht, einen Originaltext Wort für Wort und ohne jede weitere Bearbeitung in den Speicher zu übernehmen. Vielmehr müßte sich ein Fachmann an dieser Stelle in die Dokumentationsabläufe einschalten. Nur er kann erkennen, was mit den Worten und Bildern des Autors alles ausgesagt und angedeutet ist, so daß eine Publikation auch bezüglich des Angedeuteten ergänzt und damit auffindbar gemacht werden kann. Liegt ihm aber bereits eine Publikation zur Interpretation vor, so steht – außer dem hiermit eventuell verbundenen Arbeitsaufwand – nichts im Wege, daß er bei dieser Gelegenheit auch gleich die Normierung der Ausdrucksweise für Begriffe für den

Speicher vornimmt. Eine solche Übersetzung von angetroffenen Ausdrucksweisen in eine vorgegebene, festvereinbarte und damit voraussehbare Form ist nichts anderes als die traditionelle „Verschlüsselung“, die auch heute noch die weitestverbreitete Art der Vorbereitung von wissenschaftlichen Publikationen für die Einspeicherung ist, keineswegs ein antiquiertes Verfahren, sondern eine überaus sinnvolle Maßnahme, wenn man hohe Ordnungsgrade anstrebt^[26].

Der Arbeitsaufwand beim Verschlüsseln läßt sich durch Maschineneinsatz erheblich reduzieren. Es genügt, ein Klartextwort niederzuschreiben, und die Maschine ersetzt dieses (oder die bedeutungsgleichen anderen Wörter) aufgrund vorgegebener Instruktionen durch eine Notation^[25, 27, 28].

Der Fortschritt ist unverkennbar, welcher immer mit einer solchen Normierung der Ausdrucksweise verbunden ist, gleichgültig, ob diese nun rein maschinell oder intellektuell vollzogen worden ist. Er besteht darin, daß nunmehr für einen jeden Begriff der Fragestellung nur noch ein einziger Speicherterminus berücksichtigt zu werden braucht, zumindest solange es sich um *Individualbegriffe* handelt, also etwa um Einzelverbindungen oder um eine ganz bestimmte Reaktion.

Die Sonderstellung der Individualbegriffe wird deutlich, wenn man sich die Relevanzregel aus Abschnitt 3.3 vergegenwärtigt. Für eine Fragestellung mit dem Begriff α sind alle Publikationen relevant, in denen α selbst oder irgendein spezifischerer Begriff vorkommt, für welchen also α ein Oberbegriff ist. Bei den Individualbegriffen liegen die Verhältnisse nun deswegen so einfach, weil es zu ihnen überhaupt keine spezifischeren Begriffe gibt. Zielt nun aber eine Fragestellung auf einen Allgemeinbegriff ab, z.B. auf „Herbizid“, so müssen Vorkehrungen zum Auffinden von Literatur zu spezifischeren Begriffen wie „Selektivherbizid“, „Kontaktherbizid“, „Bodenherbizid“ usw. getroffen werden. Es müssen nun sämtliche spezifischeren Begriffe, die unter diesen Allgemeinbegriff fallen, der Reihe nach als alternativ Suchbedingungen im Frageterminus aufgezählt werden. Daß nach dem zuvor beschriebenen Ordnungsschritt für jeden dieser spezifischeren Begriffe nur noch ein einziger Speicherterminus berücksichtigt zu werden braucht (also nur etwa „Bodenherbizid“ und nicht auch noch zusätzlich das Synonym „Wurzelherbizid“), bedeutet bereits eine wesentliche Erleichterung für das Zusammenstellen des Frageterminus. Trotzdem können sich hierbei noch erhebliche Probleme ergeben, wenn die Zahl der zugehörigen spezifischeren Begriffe groß ist.

Wenn die Verhältnisse so gelagert sind wie in der Chemie, daß nämlich zu einem Allgemeinbegriff („Harnstoffkörper“) eine unübersehbare große Zahl von zu-

[25] M. Scheele, Naturwissenschaften 52, 1 (1965).

[26] Vgl. hierzu: National Acad. of Sciences, National Research Council, Washington, D. C., Publication 1150: Survey of Chemical Notation Systems; Publication 1278; D. E. H. Frear: European Nonconventional Notation Systems; F. A. Tate, Annu. Rev. Information Sci. Technol. 1967, 285.

[27] Ernst Meyer, J. chem. Documentation 9, 112 (1969).

[28] H. Behmenburg u. M. Glos, Nachr. Dokumentation 12, 146 (1961).

gehörigen spezifischeren Begriffe existiert („Halogen-phenylharnstoff“, „Halogen-alkyl-phenylharnstoff“, „Polyhalogen-dialkyl-sulfamino-carbanilid“ usw.), so ist jeder Versuch zum Scheitern verurteilt, alle derartigen Begriffe auch nur annähernd vollständig aufzuzählen und von einer Maschine mit tragbarem Aufwand als Suchbedingung handhaben zu lassen. Hier gibt es nur einen sicheren Weg zu der Ordnung der Art und des Grades, wie sie vom Fragesteller erwartet wird: Die verwandtschaftlichen Beziehungen zwischen den Begriffen müssen in ihren fest vereinbarten Speichertermini mit einer solchen Deutlichkeit zum Ausdruck kommen, daß sie auch für eine Maschine erkennbar sind, wenn sie diesen Begriffen beim Durchforschen des Speichers begegnet. Dies erreicht man auf einfache Weise etwa dadurch, daß man sämtliche Harnstoffkörper mit einem Terminus einspeichert, in welchem stets eine bestimmte Zeichenfolge vorkommt, etwa „PD“, gleichgültig, wie das Molekül anderweitig noch strukturiert ist. Anders ausgedrückt, vereinbart man „PD“ als verbindliche Kennzeichnung *sämtlicher* Harnstoffkörper.

Hat man alle Strukturen in Gestalt solcher „Systematischer Notationen“ eingespeichert, so ist es ein leichtes, einen Frageterminus nach „Harnstoffkörpern“ (etwa in Verbindung mit einer bestimmten herbiziden Wirkung) zu formulieren. Die Maschine erhält den Auftrag, im Speicher nach der Zeichenfolge „PD“ zu suchen. Ballastfrei und verlustlos werden die einschlägigen Publikationen lokalisiert. Bei dieser Methode braucht man nicht mehr zu wissen oder zu vermuten, welche spezifischen Harnstoffstrukturen sich im Speicher befinden.

Dieses Prinzip kann man zu noch höherer Leistungsfähigkeit entwickeln, wenn man dafür sorgt, daß alle Kohlensäurekörper, als Oberbegriff für die Harnstoffe, das festvereinbarte Zeichen „P“ in den systematischen Notationen erhalten. Auf eine Frage mit diesem Allgemeinbegriff werden dann zwangsläufig alle eingespeicherten Guanidine, Carbodiimide, Isocyanate, Harnstoffe, Isoharnstoffe usw. aufgefunden, wenn für diese Körper die Zeichenfolgen „PA“, „PB“, „PC“, „PD“, „PE“ usw. festgelegt worden sind („mechanisierte Gruppenbildung“)^[24, 29].

Auch für biologische Begriffe ist der Übergang zu systematischen Notationen unerlässlich, wenn die Anzahl dieser Begriffe im Speicher nur einigen Umfang angenommen hat^[25, 30].

4.3.3. Speichertermini für Begriffsverknüpfungen

All diese Maßnahmen reichen jedoch noch nicht aus, um den Anforderungen zu genügen, die bei sehr großen Literatursammlungen, in denen auch häufig recherchiert wird, gestellt werden müssen. Zwar ge-

lingt es mit ihrer Hilfe, wenigstens die Formulierungen für *einzelne Fachbegriffe* und für ihre Verwandten im Speicher vorausschauen zu machen, so daß es nun auf verlässliche Weise möglich ist, für diese Begriffe einen Suchauftrag zu formulieren.

Aber die vielfältigen Ausdrucksweisen für die *Verknüpfungen und Zusammenhänge*, in denen die Fachbegriffe in den gespeicherten Publikationen vorkommen können, bleiben auch nach diesen Maßnahmen noch unübersehbar. Sie lassen sich demzufolge nicht oder nur recht unvollkommen in die mechanisierte Lokalisierung einbeziehen. Wenn eine Fragestellung gerade auf eine typische Verknüpfung solcher Fachbegriffe abzielt, so muß man viel Ballast und Verlust in Kauf nehmen, weil sich das Relevante nicht scharf genug lokalisieren läßt. Dieses Problem ist in neuerer Zeit intensiv bearbeitet worden^[31].

Die Techniken der mechanisierten Lokalisierung unterscheiden sich außerordentlich stark darin, in wieweit sie die Besonderheiten in den Verknüpfungen der Begriffe bei der Einspeicherung zum Ausdruck und bei der Lokalisierung zur Wirkung bringen können. Bei den einfachen Handlochkartensystemen kann man nicht viel mehr erwarten als ein einziges Niveau der Koordination und eine einzige Art der Verknüpfungslogik. Elektronische Rechenanlagen ermöglichen es hingegen in Verbindung mit einer hochentwickelten Programmierung, viele Feinheiten der Koordination und der Verknüpfungslogik bei der Einspeicherung darzustellen und beim Vergleich zwischen Frageterminus und Speicherterminus zu verwerten (vgl. ^[32–37]).

4.3.4. Unterscheidung von verschiedenen Niveaus der Begriffsverknüpfung

Es ist aufschlußreich, die Publikationen näher in Augenschein zu nehmen, welche man durch die bloße Postkoordination der Fragebegriffe „Harnstoffkörper“ und „Alkylbenzol“ und „Halogenbenzol“ und „herbizide Wirksamkeit“ lokalisieren könnte. Zwar würden alle lokalisierten Publikationen die Bedingung erfüllen, daß in ihnen die genannten Begriffe vorkommen. Aber es ist noch keine Gewähr dafür gegeben, daß es tatsächlich der Harnstoffkörper ist, welcher die verlangte herbizide Wirksamkeit besitzt. Es kann in einer Publikation von herbizid wirksamen Pyrimidinen

[31] Vgl. J. M. Perreault, Proceedings of the International Symposium on Relational Factors in Classification. Information Storage and Retrieval 3, 177 (1967).

[32] E. Marden: Haystaq: A Mechanized System for Searching Chemical Information. National Bureau of Standards, Washington, D. C., Technical Note 264, Sept. 1965.

[33] J. Frome u. P. T. O'Day, J. chem. Documentation 2, 248 (1962).

[34] R. Fugmann, Classification Research, Munksgaard, Copenhagen, 1965, 341.

[35] P. A. D. De Maine, K. Kloss u. B. A. Marron: The Solid System. National Bureau of Standards, Washington, D. C., Technical Note 413, Aug. 1967.

[36] G. Salton, Commun. Amer. Assoc. Computing Machinery 1962, 103.

[37] R. Fugmann, H. Nickelsen, I. Nickelsen u. J. H. Winter, Angew. Chem. 82, 611 (1970); Angew. Chem. internat. Edit. 9, Nr. 8 (1970).

[29] Vgl. hierzu R. Fugmann, W. Braun u. W. Vaupel, Angew. Chem. 73, 745 (1961).

[30] Chemical-Biological Coordination Center, National Research Council, Washington, D. C. (1950): A Method of Coding Chemicals for Correlation and Classification, Biology Code.

die Rede sein, welche ausgehend von Harnstoffkörpern synthetisiert worden sind, und die begleitenden strukturellen Bedingungen können in anderen Zusammenhängen erfüllt sein, z.B. im Lösungsmittel, in den Strukturen von Vorprodukten, in den Stoffen peripherer Reaktionsfolgen usw. Dann erfüllen diese Publikationen rein formal durchaus die Suchbedingungen im Frageterminus, ohne im eigentlichen Sinne relevant zu sein. Dem kann nur dadurch vorgebeugt werden, daß schon im Speicherterminus die *Niveaus einer engeren und weiteren Verknüpfung der Begriffe*, wie sie in der Publikation angetroffen werden, zum Ausdruck gebracht werden. Zumindest müssen diejenigen strukturellchemischen Begriffe, die zu ein und derselben Verbindung gehören, in besonders enger Koordination dargestellt werden. Andernfalls kommt man über sehr bescheidene Ordnungsgrade im lokalisierten Literaturangebot nicht hinaus. — Aus dem gleichen Grunde ist es wünschenswert, daß die *Stoffeigenschaften* nicht lediglich genannt, sondern auch den betreffenden Stoffen richtig zugeordnet werden.

Zuweilen sind für einen Fragesteller nur solche Publikationen von Interesse, in denen ausdrücklich die Verwendung von zwei Stoffen *im Gemisch* beschrieben ist, etwa in einer pharmazeutischen Zubereitung. Dann sind alle anderen Publikationen ohne Interesse, in denen diese Stoffe in anderem Zusammenhang erwähnt sind, z.B. zum Vergleich ihrer Wirklichkeiten. Oftmals interessiert ein Stoff nur in Verbindung mit einer bestimmten anderen Komponente als Stabilisator, Lösungsvermittler, als wirkungssteigernder oder nebenwirkungsreduzierender Zusatz usw.

Interessiert man sich dafür, ob die Herstellung einer Stoffklasse aus einer bestimmten anderen Klasse durch aufeinanderfolgende Reaktionsschritte beschrieben worden ist, so besteht das Bedürfnis, alle diejenigen Stoffe enger miteinander zu koordinieren, welche tatsächlich auch *der gleichen Stufenfolge von Reaktionen angehören*, nicht aber verschiedenen, voneinander unabhängigen Reaktionen, die alle in der gleichen Publikation beschrieben sein können. — Darüber hinaus ist es wünschenswert, die *zeitliche Reihenfolge* der Reaktionsschritte und Stoffe darzustellen und im Frageterminus zum Ausdruck zu bringen. Andernfalls muß man Ballast durch solche Publikationen in Kauf nehmen, in welchen die umgekehrte Reihenfolge beschrieben ist.

Wird die Herstellung einer bestimmten *Stoffklasse* durch einen bestimmten *Reaktionstyp* gesucht, so kann es bei bloßer Koordination der beiden Begriffe dadurch zu Ballast kommen, daß in den lokalisierten Publikationen zwar die gesuchte Stoffklasse und auch der Reaktionstyp beschrieben ist, jedoch unabhängig voneinander und in verschiedenen Zusammenhängen.

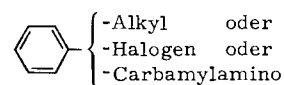
Selbst innerhalb einer Molekülstruktur kann man noch differenzierte Niveaus der Koordination erkennen. Sie spiegeln sich in der Forderung wider, daß die Alkyl-, Halogen- und Carbamylamino-Reste den *gleichen Benzolring substituieren* sollen und nicht etwa über verschiedene Benzolringe verstreut sein dürfen, die in einer gespeicherten Molekülstruktur vorkommen können. Auch sollen diese Substituenten diesen einen Benzolring jeweils nur ein einziges Mal substituieren, und es sollen in diesem Ring keine weiteren Substituenten vorkommen usw.

4.3.5. Logik der Begriffsverknüpfung

Wenn ein Fragesteller Literatur zur therapeutischen Anwendung eines speziellen Kombinationspräparates sucht, so verlangt er beispielsweise die Begriffe „Penicillin“ und „Tetracyclin“ in der logischen Verknüp-

fung der Konjunktion. Dann sind ihm zweifellos all diejenigen Publikationen unerwünscht, in denen nur *entweder* Penicilline *oder* aber Tetracycline, d.h. beide in Disjunktion, beschrieben sind.

Analog soll ein Körper vom Typ



wie er in der Patentliteratur häufig vorkommt, nicht als Antwort auf eine Fragestellung lokalisiert werden, in welcher *alle drei* Substituenten am gleichen Benzolring gefordert werden.

Daneben gibt es noch viel verwickeltere Fälle einer mehrstufigen Logik, insbesondere in der Patentliteratur des Makromolekulargebietes. Ein Stoff A wird entweder mit dem Stoff B oder mit einem Gemisch copolymerisiert, welches seinerseits entweder aus den Komponenten C und D besteht oder aber aus dem Stoff E, welchem wahlweise entweder F oder G beigegeben sein kann. Würde man hier die Details der logischen Verknüpfung außer acht lassen, so müßte man mit einem derartig großen Ballastanteil im lokalisierten Literaturangebot rechnen, daß dieser auf die Dauer untragbar wäre. Denn die Literatur auf diesem Gebiet ist bezüglich der Art der verwendeten Stoffe A, B usw. relativ monoton. Durch die bloße Forderung des Erwähntseins dieser Stoffe in einer Publikation ließe sich also noch keine ausreichend scharfe Lokalisierung erzielen.

Es ist in Abschnitt 3.2.1 ausgeführt worden, daß mit zunehmender Speichergröße und mit der fortgesetzten Proliferation des Wissens die Fragestellungen an einen Literaturspeicher immer spezifischer werden. Diese Spezifizierung der Fragestellung kommt durch die Hinzunahme weiterer, einschränkender Suchbegriffe zustande, welche positiver oder *negativer Art* sein können.

Eine Fragestellung nach Herbiziden bestimmter Wirkungsweise, welche *nicht* dem Harnstoff- oder Pyrimidintyp angehören sollen, ist durchaus sinnvoll. Auch können im Einzelfall solche herbizid wirksamen arylierten Harnstoffe besonders interessant sein, welche am Phenylrest beliebige Substituenten tragen dürfen, nur nicht den Sulfonsäure- oder Alkylsubstituenten, etwa weil diese Gruppe bereits erschöpfend untersucht worden ist.

Natürlich dürfen bei einer solchen Fragestellung nicht solche Publikationen zurückgewiesen werden, bei denen die „verbotenen“ Begriffe in irgendeinem anderen, möglicherweise gänzlich peripheren Zusammenhang erwähnt sind. Eine Publikation kann durchaus für die erstgenannte Fragestellung relevant sein, obwohl die verbotenen Harnstoff- oder Pyrimidinstrukturen in ihr vorkommen, nämlich beiläufig erwähnt, etwa als Standardpräparate für den Wirkungsvergleich.

Analog würde es eine weitgehende Deformation der Fragestellung bedeuten, wenn man das Verbot des Alkylrestes *an einem ganz bestimmten* Phenylrest generell in das Verbot einer jeglichen kerngebundenen Alkylgruppe in der Gesamtstruktur ummünzen würde. Erst recht dürfen durch eine Negation in der Fragestellung nicht solche gespeicherten Publikationen zurückgewiesen werden, in denen ein verbotener Begriff zwar vorkommt, aber einer Gruppe von Alternativen

angehört. Einen solchen disjunktiv genannten Begriff kann man sich nämlich vollwertig durch einen anderen, „unschädlichen“ ersetzt denken, falls er zu den verbotenen Begriffen zählen sollte.

Negationen in einem Frageterminus auszusprechen ist demnach überhaupt nur dann zulässig, wenn es möglich ist, ihre Wirkung exakt auf die Teile einer Publikation zu beschränken, auf welche die Fragestellung tatsächlich abzielt oder abzielen darf. Auch hierzu ist es notwendig, schon bei der Einspeicherung streng zwischen verschiedenen koordinativen Niveaus zu unterscheiden sowie die disjunktiven Begriffsverknüpfungen entsprechend zu kennzeichnen. Andernfalls käme es, wenn auch versteckt, zu einer Verfälschung des zu speichernden Themas. Werden diese Vorkehrungen nicht frühzeitig getroffen, so kann es später empfindlich an Möglichkeiten zur Formulierung von stärker spezifizierten Fragestellungen mangeln.

Die begriffliche Postkoordination hat mit dem Einzug der vieldimensionalen Techniken (Sicht-, Rand-, Schlitz-, Maschinenlochkarten, Magnetbandspeicherung usw.) starke Verbreitung gefunden. Trotz all ihrer Vorzüge (siehe Abschnitt 4.2.2) ist jedoch der Erfolg zumeist weit hinter den Erwartungen zurückgeblieben. Bei der Zerlegung von *mehreren* zusammengesetzten Begriffen einer Publikation oder einer Fragestellung in die konstituierenden Einfachbegriffe entsteht nämlich zunächst ein völlig unstrukturiertes Konglomerat von vielen Einfachbegriffen. Aus einem solchen Konglomerat kann nun jede beliebige Koordination von Einfachbegriffen herausgelesen werden und nicht etwa nur die eine, welche genau den ursprünglich zusammengesetzten Begriffen entspricht. Falls also nicht durch ein besonderes Verfahren dargestellt wird, in welcher Weise die Einfachbegriffe zusammenhängen, kommt es zu einer erheblichen Verfälschung der ursprünglichen Aussage.

Streng genommen läßt sich die volle Vorausssehbarkeit der Speichertermini für Begriffe und Begriffsverknüpfungen nur dadurch erreichen, daß man das vom Autor Gemeinte in eine neue, *künstliche Sprache* übersetzt. Diese Sprache muß gänzlich für den Zweck des Wiederfindens zugeschnitten sein, nicht etwa auf den Zweck des Mitteilens, wie die natürlichen Sprachen. Hierzu muß diese Kunstsprache über einen großen Wortschatz verfügen, der überdies noch äußerst systematisch strukturiert sein muß, so daß die Wahl eines treffenden Terminus nicht ein Übermaß von Gedächtnisleistung und Geduld beim Suchen erfordern. Diese Kunstsprache muß darüberhinaus – ebenfalls im Interesse einer großen Wiedergabetreue – eine sehr ausgereichte „Grammatik“ besitzen, damit die Verknüpfungen von Begriffen, wie sie im Dokument vorkommen oder in der Fragestellung verlangt werden, ausdrückbar sind [38].

Betrachtet man die wissenschaftliche Dokumentation aus dieser Perspektive, so kann man leicht abschätzen, ob ein Dokumentationssystem den Anforderungen in einer Großdokumentation gewachsen sein kann. Wird

dort auf jede Art von Kunstsprache verzichtet und werden auch sonst keinerlei Anstrengungen in Richtung auf die Vorausssehbarkeit der Speichertermini unternommen, so muß man die Zukunftsaussichten eines solchen Systems gering einschätzen. Ähnlich liegen die Verhältnisse, wenn ein Dokumentationssystem zwar auf einer Kunstsprache basiert, wenn aber deren Aussagekraft allzu gering ist. Ein Wortschatz für nur wenige tausend Fachbegriffe, vielleicht noch ohne jede höher entwickelte Möglichkeit zur Wiedergabe von Begriffsverknüpfungen, ist keine ausreichende Grundlage. Viele Projekte für eine Großdokumentation in der Wissenschaft werden heute diskutiert und sogar in Angriff genommen, ohne daß geklärt ist, ob die vorgesehene Methode von der begrifflichen Seite her den bald sich entwickelnden Anforderungen gewachsen sein wird. Die Bereitstellung von elektronischen Rechenanlagen allein ist noch keine Garantie für die Leistungsfähigkeit einer Methode.

4.4. Topologische Dokumentationsmethoden

Wir knüpfen an die Feststellung an, daß bei konsequenter Postkoordination außerordentlich hohe Anforderungen an die Wiedergabetreue von Begriffsverknüpfungen gestellt werden. Dies ist die tiefere Ursache dafür, daß eine so überaus naheliegende Handhabung der chemischen Strukturformel sich erst in der heutigen Zeit zögernd einzuführen beginnt. Der chemische Stoffbegriff bietet sich nämlich für die analytische Betrachtungsweise an wie kein zweiter wissenschaftlicher Begriff. Er ist nicht nur bildhaft, sondern sogar gleichsam materiell aus Einfachbegriffen aufgebaut, nämlich aus den Atomen selbst.

Was bisher einer konsequenten Zerlegung in die atomaren Einfachbegriffe im Wege gestanden hat, war das Fehlen einer Methode zur Wiedergabe der besonderen Verknüpfungen der Atome in einer Struktur, und zwar eine Wiedergabeweise, die auch in den Frageterminus für eine Stoffklasse einbezogen werden kann und die demzufolge bei der Lokalisierung wirksam ist. Ohne dieses Werkzeug wäre aus den in Abschnitt 4.3 erörterten Gründen eine derartig weitgehende Begriffsanalyse nutzlos gewesen. Erforderlich ist ein Frageterminus, in welchem nicht nur die Atome, sondern auch ihre charakteristischen Verknüpfungen miteinander festgelegt sind. In Ermangelung genügend leistungsfähiger Methoden war die Dokumentation bisher immer gezwungen, mit der gedanklichen Analyse der Strukturbegriffe bei stärker präkoordinierten Begriffen wie „Harnstoffe“, „Pyrimidine“ usw. Halt zu machen.

Nun kann man eine Strukturformel im mathematischen Sinne als einen „Graphen“ auffassen, d.h. als ein abstraktes System von „Knoten“ und „Strecken“. Den Knoten ordnet man die Bedeutung von Atomen zu, die Strecken werden als Bindungen zwischen den Atomen interpretiert. Hiernach kann man sich die mathematischen Gesetzmäßigkeiten, die in solchen topologischen Systemen herrschen, zunutze machen, um Strukturformeln als Graphen vollständig zu be-

[38] Vgl. S. R. Ranganathan, Library Sci. 5, 296 (1968).

schreiben und einzuspeichern, und auch – worauf es uns besonders ankommt – mit einem Graphen als Frageterminus wieder aufzufinden. Insbesondere ist es auf diesem Wege möglich, jede beliebige, aus Atomen und ihren Verknüpfungen bestehende Teilstruktur, z.B. die Harnstoff- oder die Pyrimidingruppierung, in einer topologisch gespeicherten Struktur auf mechanisierte Weise wiederzuerkennen^[39–41].

Dadurch, daß bei der topologischen Einspeicherung grundsätzlich alle Details einer Strukturformel festgehalten werden, lassen sich gegebenenfalls auch solche Teilstrukturen anfragen, denen man nie zuvor eine Bedeutung zugebilligt hätte, für die in keinem andersartigen System ein passender Frageterminus verfügbar wäre und für die man noch nicht einmal einen passenden Namen angeben könnte. Beispielsweise bereitet es in einem topologischen System keinerlei Schwierigkeiten, Biuret- oder Triuretstrukturen oder jeden anderen beliebig höher organisierten Harnstoffkörper treffsicher zu lokalisieren, unabhängig davon, auf welche Weise eine solche Struktur in ein größeres Molekül eingebaut ist.

Nun trifft man aber in der Fachliteratur und in den Fragestellungen häufig auf Stoffbegriffe, die sich nicht oder höchstens auf verfälschende Weise durch eine bestimmte Folge von Atomen topologisch darstellen lassen (z.B. „ungesättigte aliphatische Aminocarbonsäure-Derivate“, „alicyclische Diketone“). Die Lösung dieses Problems besteht darin, daß man parallel ein zweites, nichttopologisches System benützt, welches ganz besonders auf die Einspeicherung und Anfrage von Allgemeinbegriffen zugeschnitten ist. Die beiden Systeme stehen in der Weise in Wechselwirkung, daß alle topologisch eingespeicherten Strukturformeln maschinell in die verallgemeinerte Wiedergabeform des zweiten Systems umgewandelt werden^[39]. Dort sind sie gemeinsam mit den Allgemeinbegriffen, welche auf direktem Wege im zweiten System eingespeichert werden mußten, für eine nichttopologische Recherche zugänglich, und zwar in einer äußerst wirtschaftlichen Weise. Der (aufwendigen) topologischen Recherche brauchen dann nur noch diejenigen wenigen Publikationen unterworfen zu werden, welche bereits auf die vorangegangene Recherche im zweiten System angesprochen haben.

Ein solches kombiniertes System arbeitet bei Recherchen nach Molekülstrukturen praktisch ballast- und verlustfrei. Es ist somit weitgehend unempfindlich gegenüber dem Wachstum des Speichers, der Zunahme der Recherchenhäufigkeit und gegenüber einer steigenden Spezifizierung der Fragestellungen. Es kann viele Jahre hindurch unverändert beibehalten werden und führt daher zu einem homogenen, leicht recherchierbaren Speicher. Allerdings nimmt es nach den oben angestellten thermodynamischen Überlegungen nicht wunder, daß der Aufwand für derartig vollkommene Ordnungsmöglichkeiten beträchtlich ist.

[39] Ernst Meyer, *Angew. Chem.* 82, 605 (1970); *Angew. Chem. internat. Edit.* 9, Nr. 8 (1970).

[40] D. P. Leiter, H. L. Morgan u. R. E. Stobough, *J. chem. Documentation* 5, 238 (1965).

[41] D. J. Gluck, *J. chem. Documentation* 5, 43 (1965).

In der topologischen Dokumentation von Strukturformeln begegnet uns das analytisch-synthetische Prinzip in einer weit entwickelten Form. Die Begriffsanalyse wurde bis hin zu den letzten noch sinnvollen Bausteinen, den atomaren Einfachbegriffen, durchgeführt. In welcher Weise diese Einfachbegriffe – buchstäblich „synthetisch“ – im Verband des Moleküls miteinander verknüpft sind, wird exakt dargestellt. Mit einem extrem kleinen und übersichtlichen und demzufolge leicht memorisierbaren Vokabular von rund hundert Einfachbegriffen, den chemischen Elementen selbst, steht eine unbegrenzt große Zahl von (voraussehbaren!) Speichertermini und eine ebenso große Zahl von Fragetermini zur Verfügung. Trotz dieser großen Zahl bereitet es keinerlei Schwierigkeiten, stets den treffenden Terminus zu finden, denn man braucht ihn nicht in einem großen Verzeichnis nachzuschlagen. Er wird einfach durch Postkoordination gebildet, angepaßt an jeden Einzelfall der Einspeicherung und Anfrage.

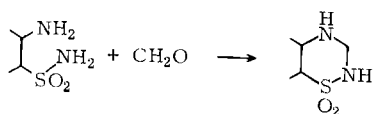
4.5. Dokumentation von Reaktionstypen

Zu den Begriffen, bei welchen der Boden für das analytisch-synthetische Dokumentationsprinzip ebenfalls gut vorbereitet ist, gehören auch die Reaktionstypen. Die geläufigen Bezeichnungsweisen für Reaktionstypen („Hofmannscher Abbau“, „Dien-Synthese“, „Beta-Oxidation“, „Krebs-Cyclus“ usw.) erscheinen aus dieser Perspektive als Termini für hochgradig zusammengesetzte Begriffe. So lange hier als Dokumentationstechniken immer nur gedruckte Register und Nachschlagewerke von eindimensionaler Struktur zur Verfügung gestanden haben, hat es auch gar keinen anderen Weg gegeben als den über hochgradig präkoordinierte Termini. Innerhalb ihrer naturgegebenen Grenzen ist der Nutzen solcher Hilfsmittel unbestreitbar, ebenso wie bei den ebenfalls eindimensionalen Sach- und Formelregistern.

Analog zu den rein strukturechemischen Begriffen mangelt es aber bei einer solchen nichtanalysierenden, ganzheitlichen Betrachtungsweise an Transparenz der Begriffsbeziehungen. Eine allgemeine Fragestellung nach Reaktionstypen ist dadurch empfindlich behindert. Ringschlüsse beispielsweise, die mit einer bestimmten Komponente durchgeführt worden sind, sind völlig regellos über ein solches Vokabular von präkoordinierten Reaktionsbegriffen verstreut und lassen sich selbst bei noch so großer Gedächtnisleistung und Geduld beim Suchen kaum lückenlos zusammenstellen. Auch muß die Einspeicherung nach einem solchen Verfahren stets lückenhaft bleiben. Nicht nur ist ein solches Vokabular zu groß und zu schlecht memorisierbar, es ist obendrein insofern zwangsläufig lückenhaft, als für die überwiegende Mehrzahl der Reaktionstypen, insbesondere für sämtliche neuen Reaktionstypen, überhaupt keine eingebürgerten Kurzbezeichnungen existieren. Hinzu kommt noch, daß derartige Bezeichnungen meistens ganz ungenügend definiert sind.

Die Vorstellungen darüber, welche präparativen Varianten z.B. der Friedel-Crafts-Reaktion zuzuordnen sind, gehen weit auseinander. Für die einen genügt bereits die Verwendung der typischen Katalysatoren bei einer Reaktion, um sie dieser Klasse zuzuordnen, andere hingegen stellen Bedingungen bezüglich der reagierenden Stoffe usw.

Zerlegt man hingegen auch Reaktionsbegriffe in analoger Weise wie bei den strukturellen Begriffen, so kann man diesen Mängeln abhelfen^[29, 34, 42, 43]. Hierzu prüft man nach gewissen Kriterien, welche Teile eines Moleküls überhaupt an der Reaktion beteiligt sind. Jeden dieser Teile beschreibt man in seinem Zustand vor und nach der Reaktion. Auf diese Weise kommt man zu reaktionschemischen Einfachbegriffen. Handelt es sich beispielsweise um die Reaktion



so wäre der Kohlenstoff vom Formaldehyd als Reaktionsteilnehmer zu betrachten. Ein weiteres Merkmal dieser Reaktion wäre der entstandene Thiadiazinring. Hiernach würde sich die folgende Schreibweise für diese Reaktion ergeben:

1. Formaldehyd → CH₂-Glieder von Heterocyclen
2. — → Thiadiazinring, entstanden
3. C-NH₂ → C-NH als Glieder von Heterocyclen

Um die reaktionsbeteiligten Molekülteile zu benennen, kann man das gleiche Vokabular wie bei der Strukturedokumentation verwenden. Neu ist hierbei lediglich, daß die Einfachbegriffe als reaktionsbeteiligte Partialstrukturen zu interpretieren sind und daß die besondere Art der Verknüpfung zwischen diesen Einfachbegriffen darzustellen ist.

Durch eine solche Aufzählung der konstituierenden reaktionschemischen Einfachbegriffe und durch die Wiedergabe ihrer Zusammengehörigkeit lassen sich alle Reaktionstypen für die Einspeicherung und Anfrage vollständig beschreiben. Analog zur Strukturedokumentation genügt hierzu ein relativ kleines und übersichtliches Vokabular von Einfachbegriffen.

Man kann sich leicht vergegenwärtigen, daß die bloße vollständige Erfassung aller Ausgangs-, Zwischen- und Endprodukte keineswegs eine Reaktionsdokumentation ersetzen kann. Sucht man beispielsweise (in einem bestimmten Zusammenhang) die Alkylierung von Aromaten durch ungesättigte Ketten, so würde es nicht zum Ziel führen, wenn man lediglich die Bedingung stellen könnte, daß in einer Publikation beide Komponenten als reaktionsbeteiligte Stoffe beschrieben sein sollen. Sie könnten nämlich dort überhaupt nicht *miteinander*, sondern *mit anderen Stoffen* in Reaktion getreten sein. Haben sie aber tatsächlich miteinander reagiert, so muß keineswegs die gesuchte Anknüpfung der Kohlenstoffkette an den Aromaten eingetreten

sein. Gemäß der Fragestellung dürfen die Komponenten noch andere reaktionsfähige Gruppen enthalten, und diese können in der unterschiedlichsten Weise miteinander reagiert haben.

Will man auch auf anderen Gebieten in der Dokumentation ähnlich erfolgreich sein wie in der reinen Chemie, so wird dies nur durch eine Abkehr vom bisherigen ganzheitlichen, präkoordinierenden Denken möglich sein und durch den Übergang zur analytisch-synthetischen Betrachtungsweise von Fachbegriffen, verbunden mit einer hochentwickelten Technik zur Wiedergabe der Begriffsverknüpfungen in Einspeicherung und Anfrage.

Ist schon auf reaktionschemischem Gebiet diese Betrachtungsweise relativ wenig verbreitet, so ist dies noch viel weniger der Fall bei Begriffen aus Nachbargebieten wie „Korrosion“, „Zonenschmelzen“, „Wirkelschichttrocknen“ usw. Bei den Fachleuten muß sich erst noch zuvor die Einsicht durchsetzen, daß mit dem Schaffen eines neuen Fachausdruckes auch gleichzeitig die Verpflichtung verbunden ist, eine Definition, zumindest aber eine ausführliche Beschreibung dieses Begriffes zu geben.

5. Schlußbetrachtung

Was die Vervollkommnung unserer experimentellen Möglichkeiten in der Chemie und in den Naturwissenschaften generell betrifft, so hat man bisher keine Anstrengungen und Kosten gescheut. Die denkbar wirksamsten Techniken sind entwickelt und eingesetzt worden, teilweise sogar in einem Umfang, welcher die Kraft einzelner Länder übersteigt und zu internationalen Zusammenschlüssen geführt hat. Auch ist es seit Jahrhunderten eine Selbstverständlichkeit, daß sich der Mensch nicht mit der Wirksamkeit seiner angeborenen Sinnesorgane begnügt, wenn es ans *Beobachten* der Phänomene geht.

Aber bei all unseren Versuchen, diese Phänomene zu *deuten*, mangelt es uns heute noch empfindlich an Methoden, die zu vergleichbarer Vollkommenheit entwickelt wären und die in ähnlichem Ausmaß über das hinausgingen, was der Mensch an natürlichen *gedanklichen* Hilfsmitteln für die Forschungsarbeit mitbringt.

Naturwissenschaftliche Erkenntnis ist gleichbedeutend mit genial geordneter Erfahrung. Um einer bisher verborgen gebliebenen Naturgesetzlichkeit auf die Spur zu kommen, trägt man möglichst viel publiziertes Material zusammen, das hiermit in Verbindung stehen könnte. In hohem Maße sind wir hierbei heute noch auf unsere Geduld und Gedächtnisleistung angewiesen und auf glückliche Zufälle, die uns relevantes Material in die Hände spielen. Dieses Erfahrungsgut, ergänzt durch eigene Experimente, durchforscht man nach Analogien, um Hinweise auf die gesuchten naturgesetzlichen Zusammenhänge zu gewinnen. Was überhaupt als Analogie, als Übereinstimmung im wirklich Wesentlichen, hypothetisch postuliert werden kann, müssen uns Phantasie und wissenschaftliche Intuition

[42] H. J. Ziegler, J. chem. Documentation 6, 81 (1966).

[43] G. E. Vleduts, Information Storage and Retrieval 1, 117 (1963).

sagen. Aber hiernach müssen nach dem neuartigen Gesichtspunkt der Hypothese die Fakten neu geordnet werden, um die bestätigenden von den widerlegenden zu trennen und um sie gegeneinander abwägen zu können. Jede Falsifikation der Hypothese führt zu neuen Modifikationen mit höherem Erklärungswert, wonach wiederum das Zusammentragen neuen Materials und das Ordnen nach neuen Gesichtspunkten beginnt. Da die Möglichkeiten zur Neuordnung wissenschaftlicher Fakten nach jedem beliebigen Gesichtspunkt von potentiell heuristischem Wert heute noch so wenig entwickelt sind, stoßen wir auch hier fortgesetzt an die Grenzen unserer Geduld und Gedächtnisleistung. So sind wir gegenwärtig noch weit davon entfernt, die Möglichkeiten auszuschöpfen, die in der wirksamen Kommunikation mit dem zeitgenössischen und dem tradierten Wissen liegen.

Es ist wichtig sich zu vergegenwärtigen, daß der bloße Umfang des publizierten Wissens und sein fortgesetztes Wachstum *nicht* die tiefere Ursache für das Einfrieren und Erstarren der Kommunikation ist. Wäre dies so, so könnte man keinerlei Gegenmaßnahmen ergreifen. Die eigentliche Ursache besteht vielmehr in der mangelnden Ordnung, welche die Wissenschaft von Anbeginn unter ihren eigenen Ergebnissen geduldet hat. Allerdings hat sich diese Unordnung mit dem ständigen Anwachsen unseres Erfahrungsschatzes immer stärker ausgewirkt, weil ihr nicht durch eine fortgesetzte Steigerung von Geduld und Gedächtnisleistung beim Suchen entgegengewirkt werden kann.

Unter diesen Umständen ist jede Resignation vor dem Kommunikationsproblem, wie sie heute zuweilen anzutreffen ist, fehl am Platze. Vielmehr erwächst hieraus für die Wissenschaft selbst eine echte Aufgabe von hohem Rang^[44]. Sie muß verhindern, daß die Effektivität von Forschungs- und Entwicklungsarbeit immer stärker hinter den Erwartungen zurückbleibt, welche man in Anbetracht des Aufwandes an Arbeitskraft und experimentellen Hilfsmitteln hegen darf.

Dies ist allerdings eine Aufgabe, für deren Lösung rein fachwissenschaftliches Denken nicht ausreicht. Will man mehr erzielen als nur ein kurzlebiges Provisorium von begrenzter Arbeitsfähigkeit, so müssen wesentliche Prinzipien aus geisteswissenschaftlichen Bereichen befolgt werden. Insbesondere ergeben sich enge Berührungen mit Mathematik, Logik und Erkenntnislehre, und hierüber sind in den letzten Jahren ausgedehnte Studien angestellt worden. Es trifft sich günstig, daß gerade in neuerer Zeit in Gestalt von programmgesteuerten elektronischen Rechenanlagen auch die technischen Hilfsmittel zur Verfügung stehen, die zur Erfüllung der Forderungen der Theorie notwendig sind. Somit erscheint sowohl von der theoretischen als auch von der technischen Seite her die Zeit reif für einen Durchbruch zu grundlegend neuen, wirksamen Kommunikationsmethoden.

[44] Report of The Presidents' Science Advisory Committee USA, The White House, Jan. 1963 (Weinberg-Bericht): Science, Government and Information. Beiheft Nr. 12 zu den Nachr. Dokumentation.

Wir sehen der wissenschaftstheoretisch hochinteressanten Entwicklung entgegen, daß die Ergebnisse der Philosophie erstmalig unmittelbare praktische Auswirkungen auf die Effektivität der naturwissenschaftlichen Forschung auszuüben beginnen, und zwar als erstes auf dem Gebiet der Chemie, weil hier die Grundvoraussetzungen für eine flüssige Kommunikation vonseiten der Terminologie und der Begriffsbildung am weitestgehenden erfüllt sind.

5.1. Bedeutung der verwendeten Dokumentationstermini

Ordnung:

Zustand, welcher durch das sinnvolle Beieinandersein der Teile eines Ganzen gekennzeichnet ist.

Handelt es sich bei dem „Ganzen“ um eine Sammlung von Publikationen oder deren Stellvertreter (Referate, Zitate, zusammenfassend im folgenden immer „Dokumente“ genannt) oder um eine Sammlung von Termini für Fachbegriffe, so hat der Benutzer einer solchen Sammlung zu spezifizieren, welche Art von Beieinandersein dieser Gegenstände für sein gegenwärtiges spezielles Suchziel oder für seine künftigen Suchziele „sinnvoll“ ist.

In diesem Fall bedeutet „Beieinandersein“, daß die dem Suchziel entsprechenden Gegenstände

- a) möglichst vollständig an einem voraussehbaren Platz zusammengeführt sein sollen und
- b) daß an diesem Platz möglichst wenige Gegenstände anzutreffen sein sollen, die dem Suchziel nicht entsprechen.

Ballast:

Diejenigen Dokumente oder Termini, welche nach der Lokalisierung am voraussehbaren Platz noch angetroffen werden, obwohl sie nicht dem Suchziel entsprechen.

Verlust:

Diejenigen Dokumente oder Termini, welche nach der Lokalisierung nicht am voraussehbaren Platz angetroffen werden, obwohl sie dem Suchziel entsprechen.

Ordnungsgrad:

Maßstab dafür, in welchem Umfang durch die Lokalisierung ein sinnvolles Beieinandersein erreicht worden ist, d.h. in welchem Umfang Ballast und Verlust am voraussehbaren Platz vermieden worden sind (Schema 3).

Relevanz:

Ein Dokument heißt relevant (oder „einschlägig“, „dem Suchziel entsprechend“),

- a) wenn es alle Begriffe des Suchzieles in gleicher oder in höherer Spezifität aufweist und
- b) wenn in ihm die Begriffe des Suchzieles in der gewünschten sinngebenden Verknüpfung vorkommen, mit der bei a) angegebenen Toleranz.

Das Vorkommen der verlangten Begriffe und der Begriffsverknüpfungen in dem Dokument muß für einen Fachmann dieses Gebietes zwingend zu konstatieren sein. Dieses Urteil ist unabhängig von der Ausdrucksweise, die der Autor gewählt hat.

Frageterminus:

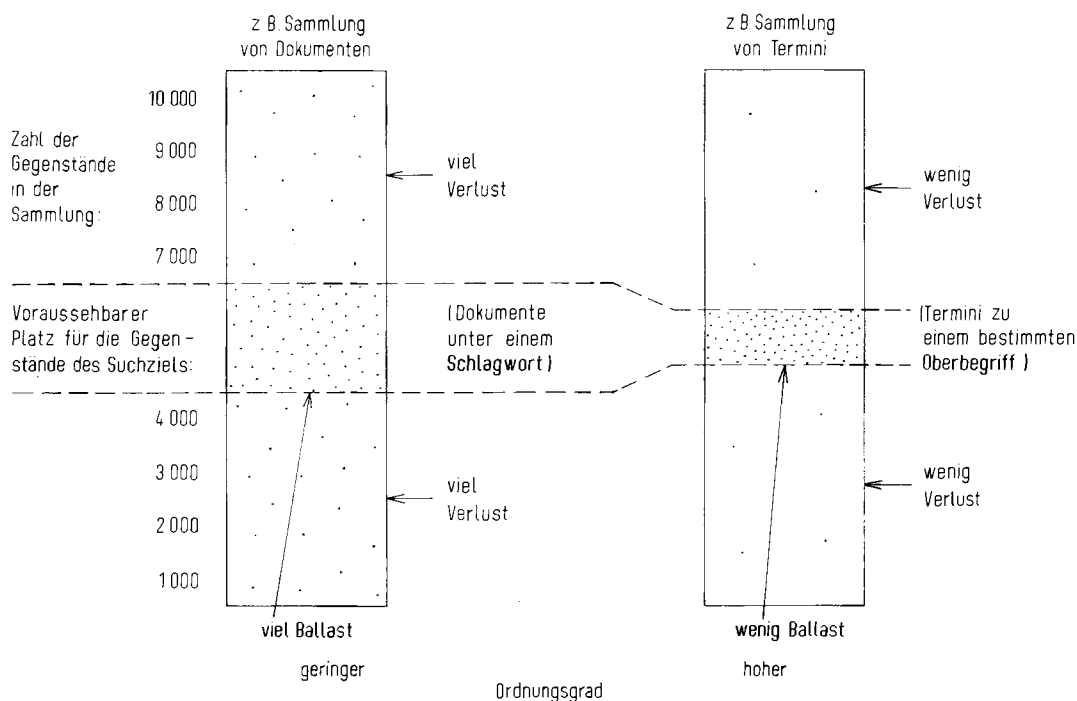
Diejenige Formulierung eines Suchzieles, welche zum Lokalisieren von Dokumenten dient.

Speicherterminus:

Diejenige Formulierung eines eingespeicherten Dokuments, mit welcher der Frageterminus auf mechanisiertem Wege verglichen wird, und welche zum mechanisierten Lokalisieren dieses Dokuments dient.

Schlagwort:

Frageterminus, welcher bei der intellektuellen Relevanzprüfung einem zu speichernden Dokument gedanklich oder physisch zugeteilt wird, gleichgültig, ob er im Dokument



Schema 3. Die Rechtecke symbolisieren die Größe einer Sammlung von Gegenständen. Jeder Punkt bedeutet einen dem Suchziel entsprechenden Gegenstand. Die Zwischenräume zwischen den Punkten bedeuten die komplementäre Anzahl von nicht dem Suchziel entsprechenden Gegenständen.

selbst vorkommt oder nicht. — Ein Schlagwort kennzeichnet den Platz, an welchem die relevanten Dokumente lokalisiert sind.

Lokalisieren:

Herstellen von Ordnung in einer Sammlung von Gegenständen. — Der voraussehbare Platz, an welchem z.B. die dem Suchziel entsprechenden („relevanten“, „einschlägigen“) Dokumente zusammengeführt werden, kann ein Platz zur körperlichen Ablage der Dokumente selbst sein, oder ein Platz zur Aufzählung der Publikationen-Stellvertreter (z.B. unter dem Schlagwort in einem Register) oder eine eigens für das Suchziel des Benutzers angefertigte Zusammenstellung von Dokumenten.

Intellektuelles Lokalisieren von Dokumenten:

Lokalisieren von Dokumenten ausschließlich durch sachverständige Einblicknahme in diese Dokumente.

Die intellektuelle Lokalisierung findet ihren Niederschlag in der Zuteilung von kennzeichnenden Schlagwörtern zu den Dokumenten oder in der Zuordnung von Dokumenten zu ihren jeweils kennzeichnenden Schlagwörtern.

Rein mechanisiertes Lokalisieren von Dokumenten:

Lokalisieren von Dokumenten ohne jede sachverständige Einblicknahme in diese Dokumente, allein aufgrund von im voraus festgelegten Suchinstruktionen.

Kombiniert intellektuell-mechanisiertes Lokalisieren von Dokumenten:

Intellektuelles Lokalisieren, gefolgt von einer anschließenden Lokalisierung allein aufgrund von im voraus festgelegten Suchinstruktionen, ohne nochmalige sachverständige Einblicknahme in die bereits intellektuell lokalisierten Dokumente.

Recherche:

Lokalisieren von Dokumenten zu einem a priori festgelegten Suchziel oder zu einem Suchziel, das der Fragesteller a

posteriori, d.h. nach Einblicknahme in gesammelte Dokumente, abgewandelt hat.

Kategorie:

Allgemeinster Begriff, für den es in dem betreffenden Sachgebiet keinen Oberbegriff mehr gibt, der als Suchziel sinnvoll wäre.

Einfacher Begriff:

Begriff, der durch die logische Unterteilung einer einzigen Kategorie entsteht.

Zusammengesetzter Begriff:

Begriff, der durch Synthese aus mehreren Einfachbegriffen entsteht. Er enthält die Summe der Merkmale der Einfachbegriffe.

Präkoordination:

Koordination von mindestens zwei einfachen Begriffen zu einem zusammengesetzten Begriff, *bevor* die Lokalisierung vollzogen wird.

Nur für diejenigen zusammengesetzten Begriffe wird die Lokalisierung vollzogen, für welche im voraus diese Koordinationen gebildet worden sind.

Postkoordination:

Koordination von mindestens zwei einfachen Begriffen zu einem zusammengesetzten Begriff, *nachdem* die Lokalisierung für die einfachen Begriffe vollzogen ist.

Für all diejenigen Begriffe kann durch Postkoordination die Lokalisierung vollzogen werden, welche durch Koordination von Einfachbegriffen gebildet werden können, und zwar ohne (nochmalige) sachverständige Einblicknahme in die betreffenden Dokumente.

Ein Teil der Arbeiten, durch welche die oben geschilderten Erfahrungen gesammelt worden sind, ist durch die Internationale Dokumentationsgesellschaft für Chemie mbH gefördert worden.

Eingegangen am 14. August 1969 [A 767]