

## EIN NEUER ANSATZ ZUR FORMALBEGRIFFSANALYTISCHEN OBJEKTKLASSIFIKATION UND SEINE ANWENDUNG AUF AROMATISCHE HETEROCYCLISCHE VERBINDUNGEN

*Hans-Georg Bartel*

Institut für Chemie der Humboldt-Universität zu Berlin, Hessische Straße 1-2, D-10115 Berlin

*Prof. Dr. RUDOLF WILLE zu seinem 60. Geburtstag im November 1997 gewidmet*

**Abstract** Es werden die Grundlagen eines neuen Verfahrens der Gegenstandsklassifikation vorgestellt, das auf der Basis der Formalen Begriffsanalyse deren Vorteile (kein Verlust an Information, Aussagemöglichkeiten über die Gründe der Klassenzugehörigkeit, Bewertung des Klassenzusammenhangs u. a.) nutzt. Als Novum ist das Erkennen und Beschreiben einer eventuell existenten Sonderklasse zu nennen. Hinsichtlich einer breiteren Anwendung werden im Sinne der Allgemeingültigkeit Fälle von Abhängigkeit bzw. Unabhängigkeit von Beziehungen zwischen den Gegenständen diskutiert. Aus aktuell-historischem Anlaß, aber auch aus Gründen der Anschaulichkeit hinsichtlich der chemischen Erfahrungen werden Daten als Demonstrationsbeispiel untersucht, in denen heterocyclische aromatische Verbindungen mit verschiedenen Aromatizitätskriterien verknüpft sind.

### 1. Historischer Anlaß ...

Es war im Jahre 1996 zweier Wissenschaftler zu gedenken, deren Wirken neben anderem eng mit der Theorie der aromatischen chemischen Verbindungen verbunden ist. Anlaß dazu gaben der 100. Todestag des Chemikers FRIEDRICH AUGUST KEKULÉ VON STRADONITZ (7.9.1829/Darmstadt – 13.7.1896/Bonn) und der 100. Geburtstag des Physikers ERICH HÜCKEL (9.8.1896/[Berlin-]Charlottenburg – 16.2.1980/Marburg).

Die Bedeutung KEKULÉS läßt sich treffend aus den Bemerkungen ERNST VON MEYERS (1847–1916) ablesen:

*Besonders glücklich hat sich die Strukturtheorie in der Hand Kekulé's gezeigt, als er an die Aufgabe herantrat, die Konstitution der sogen. aromatischen Verbindungen zu entziffern. Diese letzteren waren von ihm als Abkömmlinge des Benzols definiert; die*

*erste Aufgabe mußte demnach darin bestehen, die Struktur dieses schon seit langer Zeit bekannten Kohlenwasserstoffs ... zu ermitteln.* [1, S. 307–308]

KEKULÉ hatte seine „Vision“ der Formel des Benzens schon einige Jahre zuvor gehabt, ehe er seinen klassischen Aufsatz *«Sur la Constitution des Substances Aromatique»* am 27. Januar 1865 von CHARLES ADOLPHE WURTZ (1817–1884) vor der Société Chimique de Paris vorlesen und dann veröffentlichen ließ [2]. Dieser Tag kann somit als der eigentliche Geburtstag der Theorie des Benzens und damit zugleich der Aromatizität angesehen werden.

Auch ERICH HÜCKEL hatte in sich im Zusammenhang mit der Erarbeitung der nach ihm benannten Molekülorbitaltheorie dem Benzenmolekül als typischen Vertreter der aromatischen Verbindungen zugewandt [3]. Insbesondere wurde dabei die sogenannte HÜCKEL-Regel aufgefunden, nach welcher eine ungesättigte Kohlenwasserstoffverbindung dann aromatisch ist, wenn ihre Moleküle planare Ringe sind, die  $4n+2$   $\pi$ -Elektronen mit  $n = (0), 1, 2, \dots$  enthalten. 1938 publizierte HÜCKEL eine Zusammenfassung seiner quantenchemischen Arbeiten zu ungesättigten und aromatischen Molekülen [4].

50 Jahre zuvor waren von VICTOR MEYER (8.9.1848/Berlin – 8.8.1897/Heidelberg), dessen 100. Todestages 1997 zu gedenken ist, das vom ihm 1882 entdeckte Thiophen und dessen Derivate der Klasse der aromatischen Verbindungen zugeordnet [5], nachdem bereits 1871 die Verbindungen Pyridin, Chinolin und Isochinolin als aromatisch erkannt worden waren.

Hier läßt sich die Ähnlichkeit zum Benzen bzw. Naphthalen leicht dadurch erklären, daß in deren Molekülen ein formaler Austausch einer Gruppierung

$\text{=CH-}$  durch  $\text{=N-}$  erfolgt ist, z.B.



Dagegen liegen die Verhältnisse beim Thiophen anders, so daß OTTO DIELS (1876–1954) urteilt:

*Höchst auffallend ist es, daß Thiophen, das sich vom Benzol in der Zusammensetzung und in der Zahl der Ringglieder grundlegend unterscheidet, mit Benzol die größte Ähnlichkeit zeigt.* [6].

und WOLFGANG LANGENBECK (1899–1967) schreibt:

*Beim Thiophen ist die Ähnlichkeit mit dem Benzol auch im chemischen Verhalten besonders deutlich.* [7]

Tatsächlich ging auch VICTOR MEYERS erste Nachricht von 1882 über die Entdeckung des Thiophens mit dem bezeichnenden Titel „*Ueber Benzole verschiedenen Ursprungs*“ [8]<sup>1</sup> von dieser Ähnlichkeit aus.

Somit sind die diesbezüglichen Arbeiten und Erkenntnisse VICTOR MEYERS in Hinblick auf die empirische Strukturtheorie chemischer Verbindungen in ihrer Bedeutungen denen KEKULÉS zur Benzen- und Aromatizitätstheorie an die Seite zu stellen. Das kommt in der Äußerung RICHARD MEYERS (1846–1926)

... *das Thiophen und seine Derivate wurde zu einem neuen Kapitel der organischen Chemie, welches sich bis in die Einzelheiten als ein Abbild der Benzolchemie erwies.* [10]

und besonders in der von E. V. MEYER

*Namentlich V. Meyers ausgezeichnete grundlegende Untersuchungen über Thiophen und seine Abkömmlinge haben ... zu einer schärferen Fassung und somit zur Klärung des Begriffs aromatische Verbindungen geführt.* [1, S. 311-312]

treffend zum Ausdruck. Beispielsweise konnten Verbindungen wie Furan und Pyrrol, die wie das Thiophen eine Fünfringstruktur besitzen, in der das S- durch ein O-Atom bzw. durch eine NH-Gruppe ersetzt ist, als aromatisch betrachtet werden.

## 2. ... und gegenwärtiges Anliegen

Das empirische Bild, das auf diese Weise für die chemischen und physikalisch-chemischen Beziehungen zwischen den heterocyclischen aromatischen Verbindungen entstanden ist, zeigt offensichtlich eine Reihe, die sich vom Pyridin ableitet, und eine andere, deren Ausgangsverbindung das Thiophen ist. Diese beiden Verbindungen stehen wiederum dem Benzen sehr nahe.

Die mit diesen Feststellungen angesprochene Klasseneinteilung kann aber wegen der den Beziehungen zugrunde liegenden, in den chemischen und physikalischen Eigenschaften der Heteroaromaten zum Ausdruck kommenden Komplexität sicherlich nur auf Grund einer gewissen Willkür als disjunkt angenommen werden. Auch die Zuordnung der einzelnen heterocyclischen Verbindungen zur Klasse der Aromaten weist erfahrungsgemäß graduelle Unterschiede auf.

Es hat daher nicht an Untersuchungen und Überlegungen gefehlt, den Grad der Aromatizität relativ zu der des Benzens durch spezielle Kriterien zu erfassen. Diese Bemühungen sind insbesondere deswegen von großer Bedeutung für die

<sup>1</sup> In dieser Mitteilung [8] unterschied MEYER noch zwischen „*reines Benzol aus Steinkohlentheer*“ und „*reines Benzol aus Benzoesäure*“. Erst ein Jahr später bestätigte er seine Vermutungen, daß es sich bei dem ersteren um eine vom Benzen verschiedene Verbindung handelt, die er Thiophen nannte [9].

Chemie, da mit dem aromatischen Zustand einer Verbindung ein spezielles Reaktionsverhalten verbunden ist bzw. aus diesem abgeleitet werden kann.

Diese Kriterien verknüpft man mit dem grundlegenden Merkmal des aromatischen Zustandes bzw. des Benzenmoleküls, das theoretisch auf die Delokalisierung der  $\pi$ -Elektronen im Ring zurückgeführt wird. Sie lassen sich daher in drei Klassen einteilen. Die geometrischen Kriterien beziehen sich darauf, daß durch die genannte Delokalisierung die Abstände zwischen den Atomen des Rings von gleichem Betrag sein sollten. Da Einfach- und Doppelbindungen von verschiedener Länge sind, können geometrische Kriterien auch mit dem mehr oder weniger großen Ausgleich der Bindungsordnungen im Ring verbunden sein. Der aromatische Zustand ist durch eine besonders niedrige Energie des molekularen Grundzustandes gekennzeichnet, die sich ebenfalls mit der Delokalisierung in Verbindung bringen läßt. Die energetischen Aromatizitätskriterien benutzen daher die sogenannte Resonanzenergie, die den Unterschied zwischen der tatsächlichen Molekülenergie und der des entsprechenden Moleküls unter der Annahme fixierter Doppelbindungen angibt, als Maß. Die Beobachtung, daß aromatische Verbindungen besondere magnetische Eigenschaften aufweisen, die sich auf der Basis eines Ringstromkonzeptes wiederum durch eine Delokalisierung der Ring- $\pi$ -Elektronen erklären lassen, führt zur Klasse der magnetischen Kriterien.

Trotz ihres gemeinsamen Grundgedankens ordnen aber verschiedene Kriterien nicht in jedem Fall eine gegebene Menge aromatischer Verbindungen weder quantitativ noch qualitativ in gleicher Weise. So waren A. R. KATRITZKY et al. [11] auf Grund von Untersuchungen verschiedener Aromatizitätskriterien mit der Hauptkomponentenanalyse zu der Auffassung zweier „orthogonaler“ Aromatizitätskonzepte, eines „klassischen“, d. h. geometrischen und energetischen, und eines „magnetischen“, geführt worden. Auf ähnliche Weise kamen K. JUG und A. M. KÖSTER [12] zu der Aussage, daß es sich bei der Aromatizität um ein mehrdimensionales Problem handelt. Diese Auffassungen konnten im Prinzip mit Hilfe formalbriffsanalytischer Betrachtungen bestätigt werden [13].

Auch die vorliegende Arbeit ist der Anwendung der Formalen Begriffsanalyse (FBA) auf eine erneute Untersuchung der mit der Aromatizität verbundenen Komplexität gewidmet.<sup>2</sup> Die Erfahrung bei der Verwendung dieser Methode der explorativen Datenanalyse im Zusammenhang mit chemischen Problemen unterschiedlicher Art [13–23] hat ergeben, daß sie gerade für eine derartige Fragestellung besonders geeignet ist.

<sup>2</sup> Die FBA wurde von RUDOLF WILLE an der Technischen Hochschule Darmstadt hundert Jahre nach der Entdeckung des Thiophens entwickelt [24]. An dieser Lehr- und Forschungseinrichtung seiner Geburtsstadt hatte KEKULÉ im Wintersemester 1848/49 Vorlesungen zur Mathematik und zu den Naturwissenschaften gehört. Damals hieß die 1836 gegründete Institution Höhere Gewerbeschule, bevor sie über die Stufen der Technischen Schule (1864) und Polytechnischen Schule (1869) 1877 ihren heutigen Namen und Status erhielt.

Insbesondere sollen die hier dargestellten Untersuchungen ein weiterer Beitrag in einer Reihe von Arbeiten [16–22] sein, die der Objektklassifikation mit Hilfe der FBA gewidmet sind. Von den Vorteilen, die die FBA auch für diese Problematik bietet, seien genannt, daß eine Klassifikation ermöglicht wird, welche die mit den jeweiligen Daten gegebene Hierarchie verlustlos widerspiegelt, da sie eine baumartige Hierarchie nicht erzwingt, und daß bei der Klassifikation von Objekten zugleich die merkmalsbedingten Klasseninhalte ablesbar sind. Weiterhin ist die Klassifikation nicht notwendig disjunkt und kennt keine methoden- oder zufallsabhängige Mehrdeutigkeit, wie man sie etwa bei der Verwendung hierarchischer bzw. partitionierender Verfahren der Cluster-Analyse kennt (vgl. [16, 17, 25–29]).<sup>3</sup>

Der Inhalt und das Ziel der nachfolgenden Darlegungen kann in zwei Punkten zusammengefaßt werden:

- (A) Die Darstellung eines neuen, auf der Formalen Begriffsanalyse basierenden Verfahrens der Gegenstandsklassifikation und
- (B) dessen Demonstration an einem überschaubaren chemischen Beispiel.

Die vorliegende Arbeit versteht sich somit als die erste, möglichst präzise Einführung in ein Klassifikationsverfahren, das später bei der Behandlung umfangreicher Datensätze Anwendung finden soll, wobei es variiert und ausgebaut werden wird. Das Beispiel wurde mit dem Ziel ausgewählt, durch empirische Forschung gewonnene chemische Erfahrung als Ergebnis gewissermaßen wiederzufinden und illustrieren zu können.

Hinsichtlich dieser chemischen Anwendung wird daher die im ersten Abschnitt dargelegte Fragestellung der hierarchischen Ordnung und Klassifikation heterocyclischer aromatischer Verbindungen unter Verwendung<sup>6</sup> einiger ausgewählter Aromatizitätskriterien untersucht. Insbesondere soll dabei die Stellung des Thiophens im Vergleich etwa mit dem Pyridin bzw. Benzen beleuchtet werden. Um die gerade auch dem praktisch arbeitenden Chemiker nützlichen Vorteile, die bei der Benutzung der FBA entstehen, aufzuzeigen, werden den Untersuchungen im Rahmen dieser Methode Ergebnisse vorangestellt, die bei der Anwendung von Cluster-Analyse-Verfahren erhalten wurden.

---

<sup>3</sup> Eine Vergleich der Formalen Begriffsanalyse mit der Hauptkomponentenanalyse und ähnlichen Methoden braucht hier nicht wiederholt zu werden, da er bereits von K. E. WOLFF et al. [41–44] in allgemeiner Form durchgeführt wurde und der Inhalt der vorliegenden Arbeiten einen solchen nicht anstrebt. Das Primat der folgenden Ausführungen liegt vielmehr in der Darstellung des Inhalts und des vorteilhaften Anwendens des neuen Klassifikationsverfahrens auf chemische Fragestellungen. Daß dazu ein Demonstrationsdatensatz Verwendung findet, der zuvor mit der Hauptkomponentenanalyse behandelt wurde ([11, 30] bzw. [12]), ist sekundärer Natur.

Ein Beitrag zu einer umfassenden Theorie der Aromatizität ist also auf keinen Fall das Anliegen dieser Arbeit. Ihre Hauptblickrichtung zielt vielmehr auf praktische Anwendungen im Rahmen der explorativen Datenanalyse.

### 3. Objekte, Merkmale und Werte

Als Untersuchungsgegenstände wurden in Anlehnung an die Arbeiten von A.R. KATRITZKY et al. [11, 30] die in Abb. 1 aufgeführten elf heterocyclischen aromatischen Verbindungen und das Benzen als typischsten Aromaten<sup>4</sup> benutzt.

Diesen sind fünf Aromatizitätskriterien als Merkmale zugeordnet worden, die in Tabelle 1 genannt und erläutert werden (s. auch [11, 30]).

Die bei den vorliegenden Untersuchungen verwendeten Werte für die einzelnen Aromatizitätskriterien wurden der Arbeit [30] entnommen. Sie sind in Tabelle 2 jeweils in der ersten Spalte wiedergegeben. Jeder Wert ist auf ein ordinales Merkmal abgebildet, dessen Ausprägung die jeweilige zweite Spalte angibt.

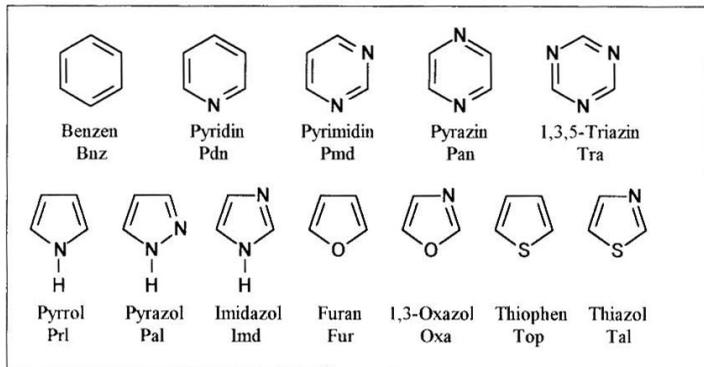


Abb. 1 Die Strukturformeln der betrachteten aromatischen Verbindungen und die verwendeten Abkürzungen

<sup>4</sup> „Es waren die aromatischen Verbindungen, für die Kekulé im Benzol das Urbild sah.“ schrieb der Chemie-Historiker RUDOLF WINDERLICH (1876–1951) [33].

**Tabelle 1** Aromatizitätskriterien

Bezeichnung	Klasse	verwendetes Symbol	Symbol in [11, 30]
Index nach A. F. POŽARSKIJ	geometrisch	<i>P</i>	$\Delta\bar{N}$ <sup>a</sup>
Index nach C. W. BIRD		<i>B</i>	$I_{5(6)}$ <sup>b</sup>
Index nach K. JUG		<i>J</i>	$RC$ <sup>c</sup>
Resonanzenergie nach B. S. A. HESS und L. S. SCHAAD	energetisch	<i>E</i>	$HSRE$ <sup>d</sup>
molare Suszeptibilität	magnetisch	$\chi$	$\chi_M$ <sup>e</sup>

*Erläuterungen:*

- a)  $\Delta\bar{N} = \frac{1}{n} \sum \Delta N$ .  $\sum \Delta N$  Summe der Bindungsordnungsdifferenzen, *n* Anzahl der Bindungen
- b)  $I_t = 100 \cdot (1 - V/V_t^K)$  (*t* = 5, 6),  $V = \frac{100}{nN_0} \sqrt{\sum (N - N_0)^2}$ ,  $V_5^K = 35$ ,  $V_6^K = 33,3$ . *N* Bindungsordnung,  $N_0$  arithmetisches Mittel der Bindungsordnungen, *n* Anzahl der Bindungen,  $V_5^K, V_6^K$  für 5- bzw. 6-Ring-KEKULÉ-Strukturen mit alternierenden Einfach- und Doppelbindungen
- c) *RC*: kleinste Bindungswahl der Bindungen zwischen Atomen in einem monocyclischen Ring
- d)  $HSRE = E_{HMO} - E_{loc}$ .  $E_{HMO}$  HMO-Gesamtenergie der  $\pi$ -Elektronen,  $E_{loc}$  aus empirischen  $\pi$ -Bindungsenergien berechnete Energie
- e)  $\chi_M$ : Absolutbetrag der gemessenen molaren Suszeptibilität, Angaben in  $10^{-6} \frac{cm^3}{mol}$

**Tabelle 2** Ausprägungen der Aromatizitätskriterien

Verbindung	<i>P</i>		<i>B</i>		<i>J</i>		<i>E</i>		$\chi$	
Benzen	0,00	6	100,0	7	1,77	4	0,7	6	54,8	7
Triazin	0,00	6	100,0	7	1,48	3	0,4	3	37,9	1
Pyrazin	0,06	5	90,8	6	1,56	3	0,5	4	37,6	1
Pyridin	0,10	5	84,1	5	1,57	3	0,6	5	49,2	6
Pyrimidin	0,13	5	82,4	5	1,50	3	0,5	4	43,1	3
Pyrazol	0,18	4	74,2	4	1,32	2	0,6	5	42,6	3
Imidazol	0,21	4	68,0	3	1,31	2	0,4	3	44,3	4
Thiophen	0,25	3	67,7	3	1,53	3	0,3	2	57,4	8 (7)
Pyrrol	0,24	3	66,5	3	1,34	2	0,4	3	47,6	5
Thiazol	0,32	2	59,4	2	1,28	2	0,3	2	50,6	6
Oxazol	0,40	1	37,8	1	1,08	1	0,1	1	39,2	2
Furan	0,47	1	35,2	1	1,11	1	0,1	1	43,1	3

Wie die Kopfzeile in Tabelle 2 zeigt, werden für das ursprüngliche Merkmal und dessen Abbildung dasselbe Symbol verwendet. Die genannte Abbildung kann man auch so auffassen, daß jeder Wert eines Kriteriums einem Intervall zugeordnet wird, dessen Nummer durch die Ausprägung des ordinalen Merkmals gegeben ist. Dabei gilt für jedes Kriterium, daß die natürliche Reihenfolge dieser Zahlen ein Anwachsen der Aromatizität bezüglich des entsprechenden Kriteriums verdeutlicht.

#### 4. Klassifikation mit Cluster-Analyse-Algorithmen

Als Daten für die Cluster-Analyse-Verfahren wurden die in Tabelle 2 aufgeführten ursprünglichen Zahlenwerte der Aromatizitätskriterien verwendet.

Die Vorgehensweise zur Ermittlung sogenannter verallgemeinerter Partitionen, d.h. einer Klassenzerlegung der in Abb. 1 gegebenen Menge  $G_A$  von zwölf Verbindungen nach den fünf Eigenschaften  $P$ ,  $B$ ,  $J$ ,  $E$  und  $\chi$ , kann durch folgendes Schema wiedergegeben werden:

- (1) Vorgabe der Anzahl  $c$  von Cluster (disjunkte Klassen), in welche die Menge  $G_A$  mittels des jeweiligen Cluster-Analyse-Algorithmus zerlegt werden soll. Es wurden nacheinander die Zahlen  $c = 2, 3, 4$  gewählt.
- (2) Ermittlung einer Menge  $P(c) = \{P_1^c, P_2^c, \dots, P_2^c\}$  unterschiedlicher Partitionen  $P_i^c$  fester Länge  $c$ . Bei der Benutzung von sieben Verfahren der agglomerativen hierarchischen Cluster-Analyse nach [31, S. 162–189] liefern die einzelnen Verfahren im allgemeinen unterschiedliche Partitionen  $P_i^c$ . Im Falle des partitionierenden Verfahrens KMEANS [31, S. 68–75] ist die berechnete Endpartition  $P_i^c$  von der gewählten Startpartition abhängig, so daß die Menge  $P(c)$  durch eine zufallsgenerierte Anzahl unterschiedlicher Startpartitionen erhalten werden kann. Abstandsberechnungen wurden stets mit dem euklidischen Maß vorgenommen.
- (3) Die gesuchte verallgemeinerte Partition ergibt sich aus  $P(c)$  als Menge aller nicht leeren Mengen, die erhalten werden, wenn der Durchschnitt aller Klassen  $C_{ij}^c \in P_i^c$  einer jeden Partition  $P_i^c \in P(c)$  mit allen anderen Klassen von  $P(c)$  gebildet wird [29, 32, 39].

Die verallgemeinerten Partitionen, die unter Verwendung der sieben agglomerativen Verfahren der hierarchischen Cluster-Analyse ermittelt wurden, zeigt Abb. 2. Bei Benutzung des KMEANS-Algorithmus sind die Ergebnisse erhalten worden, die in Abb. 3 dargestellt sind.

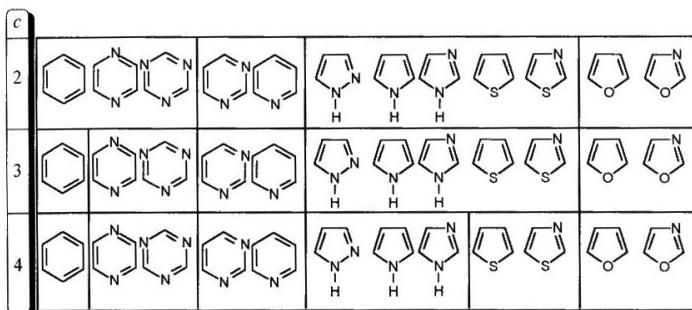


Abb. 2 Mit hierarchischen Cluster-Analyse-Verfahren ermittelte verallgemeinerte Partitionen

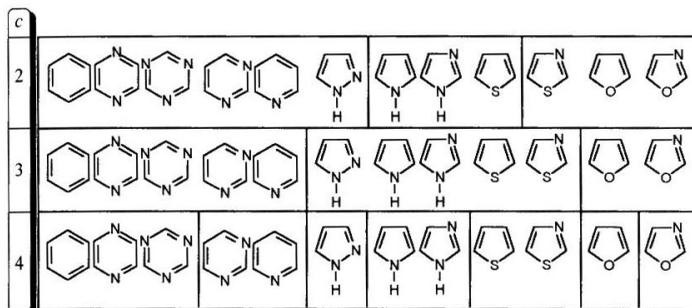


Abb. 3 Mit dem partitionierenden KMEANS-Verfahren ermittelte verallgemeinerte Partitionen

Die Klassifikationen lassen erkennen, daß mit den verwendeten Aromatizitätskriterien offensichtlich bestimmte strukturelle und chemische Eigenschaften der Verbindungen in einem Zusammenhang stehen. Zu diesen gehört die Anzahl der Ringatome, denn nur in einem Fall wurde eine Klasse erzeugt, die 5- und 6-Ringaromaten gleichermaßen erhielt (s. Abb. 3,  $c = 2$ ). Auch die chemische Natur der Heteroatome (N, S, O) hat einen mittelbaren Einfluß auf die Klasseneinteilung.

Es läßt sich aber nicht verkennen, daß die Ergebnisse der Cluster-Analyse im wesentlichen nur die Aussage gestatten, daß die benutzten Kriterien auch in ihrer Gesamtheit anwendbar sind. Es ist aber an den Resultaten nicht ablesbar, welchen Inhalt hinsichtlich der Kriterien den einzelnen Klassen zukommt, d. h.,

welche Merkmalsausprägungen sie kennzeichnen, und wie das Verhältnis der Elementen bezüglich ihrer Klassenzugehörigkeit ist. Aber gerade diese Fragen sind für die chemische Praxis, insbesondere für die Beurteilung des Reaktionsverhalten von außerordentlicher Wichtigkeit.

Auch die empirische Erfahrung der Heterocyclenchemie von der Existenz eines vom Pyridin/Benzen und eines vom Thiophen ableitbaren Zweiges sowie der Komplexität einer hierarchischen Ordnung der Aromaten kommt im Rahmen der Cluster-Analyse nicht oder nur sehr mittelbar zum Ausdruck. Beispielsweise könnte man aus der wechselnden Klassenzugehörigkeit des Pyrazols im Falle der Analyse mit dem KMEANS-Algorithmus (Abb. 3) auf das Vorhandensein komplexer Verhältnisse schließen.

Es soll im weiteren, wie oben angedeutet, gezeigt werden, daß diese Mängel im Rahmen formalbegriffsanalytischer Betrachtungen nicht auftreten.

## 5. Einige Grundlagen der Formalen Begriffsanalyse

Nachdem nunmehr neben dem umfassenden Übersichtsartikel [34] auch die Monographie von B. GANTER und R. WILLE [35] über die FBA sowie eine Reihe von Arbeiten [13–23] zu Anwendungen dieser Methode in der Chemie, die auch deren Grundlagen erläutern (insbesondere s. [13, 16, 23]), vorliegen, genügt es, hier nur eine kurze Wiederholung einiger grundlegender Definitionen der FBA zu geben:

Das Tripel  $\mathcal{C} = (G, M, I)$  ist ein (einwertiger) *Kontext*, wenn  $G$  und  $M$  Mengen sind, die Gegenstands- bzw.  $M$  Merkmalsmenge heißen, und  $I \subseteq G \times M$  gilt. Die Aussage  $(g, m) \in I$  bzw.  $gIm$  wird gelesen: „Der Gegenstand  $g$  hat das Merkmal  $m$ .“

Ist  $H \subseteq G$  und  $N \subseteq M$ , so bezeichnet

$$H' = \{m \in M \mid \forall h \in H: hIm\} \subseteq M$$

die Menge aller Merkmale, welche alle Gegenstände von  $H$  gemeinsam besitzen, und

$$N' = \{g \in G \mid \forall n \in N: gIn\} \subseteq G$$

enthält alle Gegenstände, bei denen alle Merkmale von  $N$  vorkommen. Das Paar  $(X, Y)$  wird ein (formaler) *Begriff* in  $\mathcal{C}$  genannt, falls  $X \subseteq G$ ,  $Y \subseteq M$ ,  $X' = Y$  und  $Y' = X$  gilt. Dabei heißt  $X$  der Umfang und  $Y$  der Inhalt von  $(X, Y)$ . Die Menge aller Begriffe bezüglich  $\mathcal{C}$  sei mit  $\mathcal{B}(\mathcal{C})$  bezeichnet.

Hervorgehoben seien die Gegenstandsbegriffe. Dabei heißt ein Begriff  $\gamma g \in \mathcal{B}(\mathcal{C})$  *Gegenstandsbegriff* zum Gegenstand  $g \in G$ , wenn er der Begriff aus  $\mathcal{B}(\mathcal{C})$  ist, der  $g$  im kleinsten Umfang hat, d. h.  $\gamma g = (\{g\}', \{g\}')$ .

Der Begriff  $(U, V)$  ist in  $\mathcal{C}$  ein *Unterbegriff* des *Oberbegriffs*  $(X, Y)$ , wenn  $U \subseteq X$  bzw. gleichbedeutend  $V \supseteq Y$  gilt, d. h. in symbolischer Schreibweise:

$$(U, V) \leq (X, Y) : \Leftrightarrow U \subseteq X \Leftrightarrow V \supseteq Y.$$

Die Halbordnungsrelation  $\leq$  wird *Unterbegriffsrelation* genannt. Die (halb)geordnete Menge  $(\mathcal{B}(\mathcal{C}), \leq)$  ist ein vollständiger Verband, der *Begriffsverband* zum Kontext  $\mathcal{C}$ , der in Form eines *Liniendiagramms* visualisiert werden kann. (Zum Zeichnen und Lesen von Liniendiagrammen s. z. B. [13, 16, 23, 34–36].)

Einwertige Kontexte, deren zugehörige Begriffsverbände eine einfache und klare begriffliche Bedeutung widerspiegeln, werden als begriffliche *Skalen* bezeichnet. Im Falle der eindimensionalen *Ordinalskala*  $(\{1, 2, \dots, n\}, \{1, 2, \dots, n\}, \leq)$ , wobei  $\{1, 2, \dots, n\} \subset \mathbb{N}$  eine Teilmenge der natürlichen Zahlen  $\mathbb{N}$  und  $\leq$  die gewöhnliche in  $\mathbb{N}$  erklärte Relation sind, ist diese Bedeutung der Ordination. Das Liniendiagramm zum Begriffsverband dieser Skala ist eine Kette.

Das Quadrupel  $(G, A, W, R)$  nennt man einen *mehrwertigen Kontext*, falls  $G$ ,  $A$  und  $W$  Mengen und  $R$  die ternäre Relation  $R \subseteq G \times A \times W$  sind und  $|W| > 1$  gilt.  $G$  heißt die Gegenstands-,  $A$  die Merkmals- und  $W$  die Werte- oder Ausprägungsmenge. Die Aussage  $(g, a, w) \in R$  wird gelesen: „Der Gegenstand  $g$  hat beim Merkmal  $a$  den Wert bzw. die Ausprägung  $w$ .“ Die gleichzeitige Gültigkeit von  $(g, a, w_1) \in R$  und  $(g, a, w_2) \in R$  impliziert  $w_1 = w_2$ .

Es sei  $(G, A, W, R)$  ein mehrwertiger Kontext mit  $|A| = \alpha$  (mehrwertigen) Merkmalen:  $A = \{a_1, a_2, \dots, a_\alpha\}$ . Weiterhin bezeichne  $A_i(W)$  die Menge aller Werte bzw. Ausprägungen aus  $W$ , die das Merkmal  $a_i \in A$  in  $(G, A, W, R)$  aufweist.  $(G, A, W, R)$  werde unter Verwendung der Skalen  $\mathfrak{S}_i = (A_i(W), M_i, I_i)$  für  $i = 1, 2, \dots, \alpha$  *skaliert*, d. h., in einen einwertigen Kontext überführt.

Der aus dem mehrwertigen Kontext  $(G, A, W, R)$  auf diese Weise abgeleitete Kontext sei  $\mathcal{C} = (G, M, I)$ , d. h.,

$$(G, A, W, R) \longrightarrow \mathcal{C} = (G, M, I) = (G, \bigcup_{i=1}^{\alpha} M_i, I).$$

Es gelten somit die Zuordnungen  $(a_i \in A) \rightarrow M_i \subseteq M$  vermöge  $a_i \rightarrow \mathfrak{S}_i$  für alle  $i = 1, 2, \dots, \alpha = |A|$ .

## 6. Gegenstandsordnung und Zerlegung der Gegenstandsmenge

Es soll nun eine neue Vorgehensweise, mit der eine Klassenzerlegung einer Gegenstands- oder Objektmenge  $G$  unter Benutzung der FBA erzeugt werden kann, im Sinne des in Abschnitt 2 genannten Ziels (A) in den folgenden zwei Abschnitten im Hinblick auf spätere Anwendungen möglichst ausführlich beschrieben werden. Besondere Aufmerksamkeit wird dem Erkennen bzw. der Konstruktion einer sogenannten *Sonderklasse* geschenkt werden, einer an der Praxis orientierten Idee, welche in Klassifikationsverfahren sonst keine Berücksichtigung findet. Ähnlich wie in früheren Arbeiten [20, 22] wird hier die Gegenstandsordnung eine zentrale Stellung einnehmen.

Unter der *Gegenstandsordnung* zum Kontext  $\mathfrak{C} = (G, M, I)$  wird dabei die geordnete Menge  $\omega(\mathfrak{C}) = (G, \leq_g)$  verstanden, wobei  $\leq_g$  eine Halbordnungsrelation ist, für die  $g_i \leq_g g_j \Leftrightarrow \gamma g_i \leq \gamma g_j$  gilt. Hierbei sind  $g_i, g_j \in G$  Gegenstände,  $\gamma g_i, \gamma g_j \in \mathcal{B}(\mathfrak{C})$  die entsprechenden Gegenstandsbegriffe und  $\leq$  die Unterbegriffsrelation (s. o.).

Durch die Äquivalenzrelation  $\cong_g$  mit

$$g_i \cong_g g_j \Leftrightarrow \gamma g_i = \gamma g_j \quad (\Leftrightarrow \{g_i\}' = \{g_j\}'),$$

d. h., die Gegenstände  $g_i$  und  $g_j$  aus  $G$  sind äquivalent, wenn sie in  $\mathfrak{C}$  dieselben Merkmale aus  $M$  haben, sei die Menge  $G$  in die Äquivalenzklassen  $G_1, G_2, \dots, G_p$  zerlegt. Dann kann der Gegenstandsordnung  $\omega(\mathfrak{C})$  ein schlichter gerichteter Graph<sup>5</sup>  $\Gamma(\omega(\mathfrak{C})) = [V, B]$  zugeordnet werden. Dabei wird die Knotenmenge  $V$  von den Äquivalenzklassen  $G_1, G_2, \dots, G_p$  gebildet:  $V = \{G_1, G_2, \dots, G_p\}$ . Für die Bogenmenge  $B$  gilt:  $G_i \triangleright G_j \in B \Leftrightarrow \langle G_i \rangle \leq_g \langle G_j \rangle$ . Hier bezeichnet  $\langle G_i \rangle$  einen Repräsentanten der Klasse  $G_i$ . Da keine Schlingen zugelassen sind, muß stets  $i \neq j$  sein.<sup>6</sup>

Bei der graphischen Wiedergabe von  $\Gamma(\omega(\mathfrak{C}))$  wird die Richtung eines Bogens  $G_i \triangleright G_j \in B$  dadurch angezeigt, daß der Startpunkt  $G_i$  ( $\bullet$ ) in der Papierebene

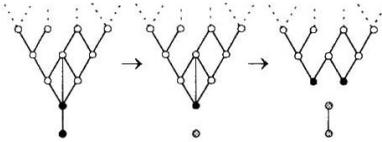
stets unter den Zielpunkt  $G_j$  ( $\circ$ ) gezeichnet wird:   $\equiv$  .

Die Gegenstandsordnung  $\omega(\mathfrak{C})$  bzw. ihre Visualisierung  $\Gamma(\omega(\mathfrak{C}))$  spiegeln eine (im allgemeinen nicht baumartige) Hierarchie der Gegenstände wider und geben zugleich eine Möglichkeit zu deren Klassifikation. Dazu werden solche Gegenstände zu einer Klasse zusammengefaßt, denen Knoten zugeordnet sind, die auf einer elementaren Bahn von  $\Gamma(\omega(\mathfrak{C}))$  liegen. Diese Klassifikation ist im allgemeinen nicht disjunkt.

Es ist leicht einzusehen, daß Gegenstände, deren zugeordnete Knoten die Eingangswalenz null haben, nicht ein und derselben Klasse angehören können. Gibt es in  $\Gamma(\omega(\mathfrak{C}))$  nur einen einzigen Knoten  $G_{s1} \in V$  mit der Eingangswalenz null, so sind die Gegenstände aus  $G_{s1}$  einer Sonderklasse zuzuordnen, da für diese offenbar als besonderes Charakteristikum  $\{\{G_{s1}\}\}' = M$  gilt.

<sup>5</sup> Die verwendeten graphentheoretische Terminologie lehnt sich an [37] an. Sie entspricht auch weitgehend derjenigen in [21] und [23, S. 57–150].

<sup>6</sup> Die Gegenstandsordnung bzw. der ihr zugeordnete Graph zu einem gegebenen Kontext lassen sich mit dem Computerprogramm von P. BURMEISTER [40], das auch im Rahmen der vorliegenden Arbeit verwendet wurde, leicht ermitteln. Dazu sind nacheinander die Optionen (A–V) = (change context data – exchange the meaning of attributes and objects) und (Q) = (calculate attribute order) zu wählen bzw. auszuführen. Mit (Q) wird dann auch die genannte Zerlegung der Gegenstandsmenge durchgeführt und angezeigt.

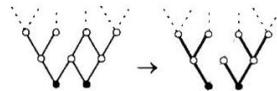


Für weiterführende Untersuchungen sind dann der Knoten  $G_{s_1}$  und alle mit ihm inzidierenden Bögen im Graphen  $\Gamma(\omega(\mathcal{C}))$  zu streichen. Der resultierende Untergraph sei mit  $\Gamma_{-1}(\omega(\mathcal{C}))$  bezeichnet.

Dem isolierten Knoten  $G_{s_1}$  wird der Graph  $\Gamma_{s_1}(\omega(\mathcal{C})) = [V_{s_1}, B_{s_1}]$  mit  $V_{s_1} = G_{s_1}$  und  $B_{s_1} = \emptyset$  zugeordnet. Existiert in  $\Gamma_{-1}(\omega(\mathcal{C}))$  wiederum ein einziger Knoten  $G_{s_2}$  mit der Eingangsvalenz null, so werden auch dieser und die mit ihm inzidierenden Bögen gelöscht, so daß der Untergraph  $\Gamma_{-2}(\omega(\mathcal{C}))$  erhalten wird. Weiterhin wird der Graph  $\Gamma_{s_2}(\omega(\mathcal{C})) = [V_{s_2}, B_{s_2}]$  mit  $V_{s_2} = V_{s_1} \cup G_{s_2}$  und  $B_{s_2} = B_{s_1} \cup G_{s_1} \triangleright G_{s_2}$  gebildet. Dieser Prozeß der Konstruktion von Untergraphen wird solange wiederholt, bis ein Untergraph  $\Gamma_{-z}(\omega(\mathcal{C}))$  entsteht, der mehr als einen Knoten mit der Eingangsvalenz null enthält oder leer ist. Andererseits ist der rekursiv gebildete Graph  $\Gamma_{s_z}(\omega(\mathcal{C})) = [V_{s_z}, B_{s_z}]$  mit  $V_{s_z} = V_{s_{z-1}} \cup G_{s_z}$  und  $B_{s_z} = B_{s_{z-1}} \cup G_{s_{z-1}} \triangleright G_{s_z}$  offensichtlich eine Kette, ein isolierter Knoten oder leer. Durch ihn wird die in  $G$  enthaltene Sonderklasse repräsentiert: Sie umfaßt die Gegenstände, denen Knoten aus  $V_{s_z}$  zugeordnet sind, ihre hierarchische Struktur ist eine ein- oder nulldimensionale Ordination.<sup>7</sup>

Wenn  $\Gamma(\omega(\mathcal{C}))$  als  $\Gamma_0(\omega(\mathcal{C}))$  aufgefaßt und  $\Gamma_{s_0}(\omega(\mathcal{C})) = \emptyset$  (leerer Graph, keine Sonderklasse), d. h.,  $V_{s_0} = \emptyset$  und  $B_{s_0} = \emptyset$ , gesetzt wird, kann allgemein von den Graphen  $\Gamma_{-z}(\omega(\mathcal{C}))$  und  $\Gamma_{s_z}(\omega(\mathcal{C}))$  mit  $z \geq 0$  gesprochen werden.

Da baumartige Hierarchien und disjunkte Klassifikationen sich für die Interpretation von Daten, Zusammenhängen und Gesetzmäßigkeiten in der Forschung und im Alltag bewährt haben, wurde das Ziel gestellt, ein Verfahren zu entwickeln, mit welchem in dem nicht leer vorausgesetzten Graphen  $\Gamma_{-z}(\omega(\mathcal{C}))$  ein Wald  $Y(\omega(\mathcal{C}))$  konstruiert werden kann, dessen Bäume den erwünschten disjunkten Klassen entsprechen.



<sup>7</sup> Da außer im letzten Schritt offensichtlich immer nur genau ein Bogen  $G_{s_{j-1}} \triangleright G_{s_j}$  gelöscht wird, ist in dem Fall, daß  $\Gamma_{-z}(\omega(\mathcal{C}))$  der leere Graph ist,  $\Gamma(\omega(\mathcal{C})) = \Gamma_{s_z}(\omega(\mathcal{C}))$  eine Kette. Die Menge  $G$  läßt sich dann in dem hier dargestellten Sinn nicht weiter in Klassen zerlegen. Es gilt  $G_{s_i} = G_i$  für alle  $i = 1, 2, \dots, p$ . Man kann aber andererseits, wie für jede Sonderklasse  $\Gamma_{s_i}(\omega(\mathcal{C}))$  auch, die  $G_{s_i} \in V_{s_i}$  als disjunkte Klassen auffassen.

Wenn die Bäume dieses Waldes mit  $\Gamma^{(i)}(\omega(\mathcal{U})) = [V_i, B_i]$  für  $i = 1, 2, \dots, q \geq 2$  bezeichnet werden, so kann man schreiben:

$$\bigcup_{j=1}^q V_j = V, \text{ aber } \bigcup_{j=1}^q B_j \subseteq B,$$

$$V_h \cap V_j = \emptyset \text{ und } B_h \cap B_j = \emptyset \text{ für } h, j = 1, 2, \dots, q, h \neq j.$$

Die Anzahl der Knoten mit der Eingangsvalenz null in  $\Gamma_{-z}(\omega(\mathcal{U}))$  entspricht der Anzahl der Bäume  $q$  des erzeugten Waldes  $Y(\omega(\mathcal{U}))$  und zugleich der Anzahl der disjunkten Klassen  $\mathbf{G}^{(i)}$ , in welche die Gegenstandsmenge  $G \setminus V_{sz}$  gemäß

$$G \setminus V_{sz} \rightarrow \mathbf{G}^{(1)}, \mathbf{G}^{(2)}, \dots, \mathbf{G}^{(q)} \text{ mit } g \in \mathbf{G}^{(j)} \subset G : \Leftrightarrow g \in G_h \wedge G_h \in V_j \subset V$$

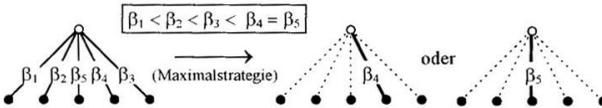
zerlegt wird.

Die prinzipielle Vorgehensweise bei der Umwandlung eines als nicht leer vorausgesetzten Graphen  $\Gamma_{-z}(\omega(\mathcal{U}))$  in den Wald  $Y(\omega(\mathcal{U}))$  ist folgende: (a) Die Bögen  $b_i \in B_z$  von  $\Gamma_{-z}(\omega(\mathcal{U})) = [V \setminus V_{sz}, B_z]$  werden mittels einer Funktion  $\beta$  bewertet, d. h.,  $\beta: \forall b_i \in B_z \rightarrow \beta_i = \beta(b_i) \in \mathbb{R}$ , (b) An jedem Knoten  $G_i \in V \setminus V_{sz}$  werden alle seine nach innen inzidierenden Bögen  $B_i^{(-)} = \{b_j | b_j = G_j \triangleright G_i\} \subseteq B_z$  bis auf einen gestrichen. Dieser übrigbleibende Bogen  $b_e \in B_i^{(-)}$  ist ein solcher mit Extremaleigenschaft, d. h. entweder

$$\beta(b_e) = \max_{b_j \in B_i^{(-)}} \beta(b_j) \text{ oder } \beta(b_e) = \min_{b_j \in B_i^{(-)}} \beta(b_j).$$

Wird die Maximaleigenschaft (Minimaleigenschaft) für die Auswahl des übrigbleibenden Bogens aus  $B_i^{(-)}$  benutzt, so soll von der Verwendung der Maximalstrategie (Minimalstrategie) gesprochen werden.

Gibt es mehrere Bögen aus  $B_i^{(-)}$  mit maximalem bzw. minimalem  $\beta$ , so wird willkürlich einer von ihnen ausgewählt. In diesem Fall gibt es mehrere gleichberechtigte Klassenzerlegungen.



Da  $\Gamma_{-z}(\omega(\mathcal{U}))$  voraussetzungsgemäß mindestens zwei Knoten mit der Eingangsvalenz null besitzt, zerfällt dieser Graph bei dieser Vorgehensweise mindestens in zwei Komponenten  $\Gamma^{(i)}(\omega(\mathcal{U})) = [V_i, B_i]$ . Streicht man in jeder dieser Komponenten einen beliebigen weiteren Bogen, so ist der entstehende Graph nicht mehr zusammenhängend, da jeder Bogen von  $B_i$  nur mit genau einem Kno-

ten aus  $V_i$  nach innen inzident ist. Ein zusammenhängender Graph mit dieser Eigenschaft ist aber ein Baum (vgl. Satz 4.1 in [38, S. 177f.]).

Da in jedem dieser Bäume  $\Gamma^{(i)}(\omega(\mathcal{C})) = [V_i, B_i]$  genau ein Knoten  $G_{0_i} \in V_i$  mit der Einganzvalenz null existiert und alle anderen Knoten die Einganzvalenz eins besitzen, sind die  $\Gamma^{(i)}(\omega(\mathcal{C}))$  Wurzelbäume (vgl. Satz 4.3 in [38, S. 179]). Ein Gegenstand  $g_{0_i} \in G_{0_i}$  der Wurzel kann als Repräsentant der Klasse von Gegenständen angesehen werden, die mit  $V_i$  gegeben ist. Diese Wahl kann deshalb sinnvoll sein, weil  $\{g_{0_i}\}'$  der größte Inhalt von allen mit  $V_i$  verknüpften Gegenstandsbegriffen ist.

## 7. Modelle der Bogenbewertung

Es sei wiederum  $\Gamma_{-z}(\omega(\mathcal{C})) = [V \setminus V_{sc}, B_z]$  der, wie im vorhergehenden Abschnitt erläutert, durch Abspaltung des Untergraphen  $\Gamma_{sc}(\omega(\mathcal{C})) = [V_{sc}, B_{sc}]$  von dem zur Gegenstandsordnung  $\omega(\mathcal{C})$  zum Kontexts  $\mathcal{C} = (G, M, I)$  gehörigen Graphen  $\Gamma(\omega(\mathcal{C})) = [V, B]$  konstruierte Graph, dessen Bögen  $b_{ij} = G_i \triangleright G_j \in B_z$  mit Zahlen  $\beta(b_{ij})$  bewertet werden müssen, um den für die Zerlegung der Menge  $G \setminus V_{sc}$  notwendigen Wald  $Y(\omega(\mathcal{C}))$  zu erhalten.

Im folgenden wird nun ein Ansatz beschrieben, mit dem sich diese Bogenbewertungen  $\beta(b_{ij})$  auf allgemeine Weise ermitteln lassen. Dabei hat die Vorgehensweise zu berücksichtigen, daß das erstrebte Ziel eine Zerlegung von  $G \setminus V_{sc}$  bezüglich bestimmter Merkmale bzw. Merkmalsklassen von  $M$  darstellt, wobei zugleich ein gewisser Grad der Klassenzugehörigkeit der Elemente einer jeden Klasse erkennbar sein soll.

Die Zerlegung von  $G \setminus V_{sc}$  werde hinsichtlich der Merkmalsklassen  $M_i$  der gegebenen Zerlegung  $\mathbf{M} = \{M_1, M_2, \dots, M_r\}$  der Merkmalsmenge  $M$  mit  $r = |\mathbf{M}| \geq 2$  angestrebt.

Existiert der Bogen  $b_{ij} = G_i \triangleright G_j \in B_z$  in  $\Gamma_{-z}(\omega(\mathcal{C}))$ , so besteht in  $\mathcal{C}$  ein (hierarchisch geordneter) Zusammenhang zwischen den Gegenständen aus  $G_i$  und denen aus  $G_j$  bezüglich  $M$  bzw. aller Merkmalsklassen aus  $\mathbf{M}$ . Im weiteren soll nun die (eventuell graduelle) Abhängigkeit bzw. Unabhängigkeit jeder Merkmalsklasse  $M_i \in \mathbf{M}$  untersucht werden. Dabei wird für jedes  $i = 1, 2, \dots, r$  folgendermaßen verfahren:

Es wird der Teilkontext  $\mathcal{C}_i = (G, M \setminus M_i, I \cap (G \times (M \setminus M_i)))$  von  $\mathcal{C}$  gebildet, in welchem gegenüber  $\mathcal{C}$  die Merkmalsklasse  $M_i$  fehlt. Die Gegenstandsordnung zu  $\mathcal{C}_i$  sei  $\omega(\mathcal{C}_i)$  und der ihr zugeordnete Graph  $\Gamma(\omega(\mathcal{C}_i)) = [V^{(i)}, B^{(i)}]$  mit der

Knotenmenge  $V^{(1)} = \{G_1^{(1)}, G_2^{(1)}, \dots, G_{p_i}^{(1)}\}$ . Aus  $\Gamma(\omega(\mathcal{U}_i))$  werden, analog wie oben für  $\Gamma(\omega(\mathcal{U}))$  beschrieben, die beiden Untergraphen  $\Gamma_{s_{z_i}}(\omega(\mathcal{U}_i)) = [V_{s_{z_i}}, B_{s_{z_i}}]$  und  $\Gamma_{-z_i}(\omega(\mathcal{U}_i)) = [V^{(1)} \setminus V_{s_{z_i}}, B_{z_i}]$  ermittelt.

Da die Richtigkeit der Aussage  $g_i \cong_g g_j$  für zwei Gegenstände  $g_i, g_j \in G$  bezüglich  $\mathcal{U}_i$  bestehen bleibt, wenn sie für  $\mathcal{U}$  galt, so muß  $p_i = |V^{(1)}| \leq p = |V|$  sein. Im Fall von  $p_i < p$  existiert also mindestens ein Paar  $\{G_h, G_l\} \subseteq V$  und eine Klasse  $G_l^{(1)} \in V^{(1)}$  mit  $G_h \cup G_l \subseteq G_l^{(1)}$ . Es ist also möglich, daß durch das Streichen der Merkmalsklasse  $M_i$  im Kontext  $\mathcal{U}$  die Anzahl der Gegenstandsbegriffe bezüglich  $\mathcal{U}_i$  verringert wird, wobei dann Umfänge von mindestens zwei Gegenstandsbegriffen von  $\mathcal{U}$  in einem Umfang eines Gegenstandsbegriffs von  $\mathcal{U}_i$  enthalten sind, so daß der Bogen  $G_h \triangleright G_l \in B$  in  $\Gamma(\omega(\mathcal{U})) = [V, B]$  auftritt, nicht aber in dem Graphen  $\Gamma(\omega(\mathcal{U}_i)) = [V^{(1)}, B^{(1)}]$ ,  $G_h \triangleright G_l \notin B^{(1)}$ , wenn  $\{G_h, G_l\} \subseteq V$  die soeben geschilderte Eigenschaft besitzen.

Jedem Bogen  $b_{ij} = G_i \triangleright G_j \in B_{z_i}$  in  $\Gamma_{-z_i}(\omega(\mathcal{U}))$  wird bezüglich aller  $i = 1, \dots, r$  Merkmalsklassen  $M_i$  genau eine Zahl  $\beta_{ij}^{(i)}$  zugeordnet, die Bewertung des Zusammenhangs  $b_{ij}$  bezüglich  $\mathcal{U}$  in  $\mathcal{U}_i$  heißen möge:

$$\forall b_{ij} = G_i \triangleright G_j \in B_{z_i}: b_{ij} \rightarrow \beta_{ij}^{(i)} \quad \text{für } i = 1, 2, \dots, r.$$

$\beta_{ij}^{(i)}$  ist genau einem Element eines Vektors  $\Lambda^{(i)} = (\mathfrak{G}^{(i)}, \lambda_c^{(i)}, \lambda_v^{(i)}, \lambda_m^{(i)}, \lambda_{mv}^{(i)})$  gleich, dessen fünf Elemente den fünf möglichen Fällen entsprechen, die im Anhang (Abschnitt 10) genauer charakterisiert werden.

Die Gesamtbewertung  $\beta(b_{ij})$  des Bogens  $b_{ij} = G_i \triangleright G_j \in B_{z_i}$  in  $\Gamma_{-z_i}(\omega(\mathcal{U}))$  ergibt sich dann als Summe dieser Einzelbewertungen:

$$\beta(b_{ij}) = \sum_{i=1}^r \beta_{ij}^{(i)}.$$

Das Tripel  $\mathbf{E}(\mathcal{U}, r) = (\mathbf{M}, \Lambda, \mathbf{e})$  möge Bewertungsmodell bezüglich des Kontexts  $\mathcal{U} = (G, M, I)$  der Länge  $r = |\mathbf{M}|$  genannt werden. Dabei ist die Menge  $\mathbf{M}$  eine Zerlegung der Merkmalsmenge  $M$ :  $\mathbf{M} = \{M_1, M_2, \dots, M_r\}$ .  $\Lambda$  bezeichnet die  $(5 \times r)$ -Matrix  $\Lambda = \left( \Lambda^{(i)} \right)$  mit  $\Lambda^{(i)} = (\mathfrak{G}^{(i)}, \lambda_c^{(i)}, \lambda_v^{(i)}, \lambda_m^{(i)}, \lambda_{mv}^{(i)})$ .  $\mathbf{e}$  ist ein Element der Menge  $\{max, min\}$ , wobei  $e = max$  die Anwendung der Maximalstrategie und  $e = min$  der Minimalstrategie anzeigt (s. o.).

Bei den nachfolgenden Betrachtungen in dieser Arbeit wird das Bewertungsmodell  $\mathbf{E}_{01}(\mathcal{U}, \alpha) = (\mathbf{M}_\alpha, \Lambda_{01}, max)$  benutzt. Dabei ergeben sich die Bedeu-

tungen von  $\alpha$ ,  $\mathbf{M}_\alpha$  und  $\Lambda_{01}$  aus den folgenden Bemerkungen:

- (1) Es möge der Kontext  $\mathcal{C} = (G, M, I)$  aus dem mehrwertigen Kontext  $(G, A, W, R)$  mit  $|A| = \alpha$  mehrwertigen Merkmalen durch Skalieren, d.h. durch die Zuordnungen  $\forall a_i \in A: a_i \rightarrow \mathfrak{S}_i = (A_i(W), M_i, I_i)$  (Skalen) entstanden sein. Dann kann die Merkmalsmenge  $M$  von  $\mathcal{C}$  als  $M = \bigcup_{i=1}^{\alpha} M_i$  aufgefaßt werden, wobei die Teilmengen  $M_i$  Elemente der Zerlegung  $\mathbf{M}_\alpha = \{M_1, M_2, \dots, M_\alpha\}$  von  $M$  mittels der Zuordnungen  $(a_i \in A) \rightarrow M_i$  für alle  $i = 1, 2, \dots, \alpha$  sind. Jede Merkmalsklasse  $M_i \in \mathbf{M}_\alpha$  ist also genau einem mehrwertigen Merkmal  $a_i \in A$  zugeordnet, so daß mit dem Bewertungsmodell  $\mathbf{E}_{01}(\mathcal{C}, \alpha)$  die Unabhängigkeit bzw. Abhängigkeit der  $a_i$  auf Zusammenhänge zwischen Teilmengen von  $G$  untersucht wird.
- (2) Für alle  $i = 1, 2, \dots, \alpha$  wird  $\mathfrak{g}^{(i)} = 0$  und  $\lambda_c^{(i)} = \lambda_v^{(i)} = \lambda_m^{(i)} = \lambda_{mv}^{(i)} = 1$  gesetzt, d.h., es gilt  $\Lambda_{01}^{(1)} = \Lambda_{01}^{(2)} = \dots = \Lambda_{01}^{(\alpha)} = (0, 1, 1, 1, 1)$ , so daß alle  $\alpha$  Zeilen der Matrix  $\Lambda_{01} = \left( \Lambda_{01}^{(i)} \right)$  einander gleich sind.

In diesem Modell wird somit nur zwischen Abhängigkeit und Unabhängigkeit eines Zusammenhangs  $b_{ij}$  von der Merkmalsklasse  $M_i$  unterschieden, wobei es auch keine graduellen Unterscheidungen hinsichtlich der  $M_i$  gibt. Mit der Zusammenhangsbewertung  $\beta(b_{ij})$  werden folglich alle die Fälle gezählt, in denen  $b_{ij}$  in beliebiger Form von der Existenz einer Merkmalsklasse  $M_i$  (und im Modell  $\mathbf{E}_{01}(\mathcal{C}, \alpha)$  des entsprechenden mehrwertigen Merkmals  $a_i$ ) unabhängig ist.

Die Verwendung der Maximalstrategie ( $e = \max$ ) bewirkt, daß solche Gegenstände, die nicht der Sonderklasse angehören, zu einer Klasse zusammengefaßt werden, deren Zusammenhang in  $\mathcal{C}$  hinsichtlich der Unabhängigkeit von den Merkmalsklassen bzw. den entsprechenden mehrwertigen Merkmalen am stärksten ist. Zugleich geben die Bewertungen  $\beta(b_{ij})$  einen Hinweis auf den Grad der Klassenzugehörigkeit der jeweiligen Gegenstände. Dieser wird um so größer sein, je kleiner die Differenz zwischen  $\beta(b_{ij})$  und dem im Modell  $\mathbf{E}_{01}(\mathcal{C}, \alpha)$  möglichen Maximalwert  $\alpha$  ist, d.h. je unabhängiger der Zusammenhang  $b_{ij}$  von den als gleichwertig angesehenen Merkmalen ist.

Bevor zu dem Anwendungs- und Demonstrationsbeispiel im Sinne des Ziels (B) von Abschnitt 2 übergegangen wird, sei noch einmal darauf hingewiesen, daß diese hier (und im Anhang) gegebenen ausführlichen Darlegungen die möglichst umfassende Grundlage für spätere Anwendungen des neuen Klassifikationsverfahrens geben wollten.

## 8. Der Begriffsverband der heterocyclischen aromatischen Verbindungen

Bei der Anwendung des in den vorhergehenden Abschnitten geschilderten Klassifikationsverfahrens und speziell des Bewertungsmodells  $E_{01}(\mathcal{C}, \alpha)$  auf heterocyclische aromatische Verbindungen wird von der Tabelle 2 als mehrwertiger Kontext  $(G_A, A_A, W_A, R_A)$  ausgegangen, wobei die ordinalen Werte der jeweils zweiten Merkmalspalte verwendet werden sollen. Das bedeutet, daß  $G_A$  zwölf Gegenstände umfaßt, die in Abbildung 1 bzw. Tabelle 2 spezifiziert sind, die Menge  $A_A$  die fünf in Tabelle 1 erläuterten Merkmale, d.h.  $A_A = \{P, B, J, E, \chi\}$ , und  $W_A$  die acht Zahlen  $W_A = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , wobei die Relation  $R_A$  aus Tabelle 2 abgelesen werden kann.

Alle fünf Merkmale werden eindimensional ordinal skaliert (s. Abschnitt 5). Dabei wird, wie oben erläutert, der Kontext  $\mathcal{C}_A$  erhalten:

$$(G_A, A_A, W_A, R_A) \longrightarrow \mathcal{C}_A = (G_A, M_A, I_A) = (G_A, \bigcup_{i=1}^5 M_A^{(i)}, I_A).$$

Für die Elemente der Merkmalsklassen  $M_A^{(i)}$  wurde folgende Notation gewählt, die sich für die Merkmale  $a_A^{(i)} \in \{P, B, J, E, \chi\}$  leicht aus den ihnen zugeordneten Skalen  $\mathcal{S}_A^{(i)}$  in Tabellenform ersehen läßt, wobei  $w_{\max}^{(i)} \in A_A^{(i)}(W_A)$  die Ausprägung von  $a_A^{(i)}$  mit dem größten Zahlenwert darstellt:

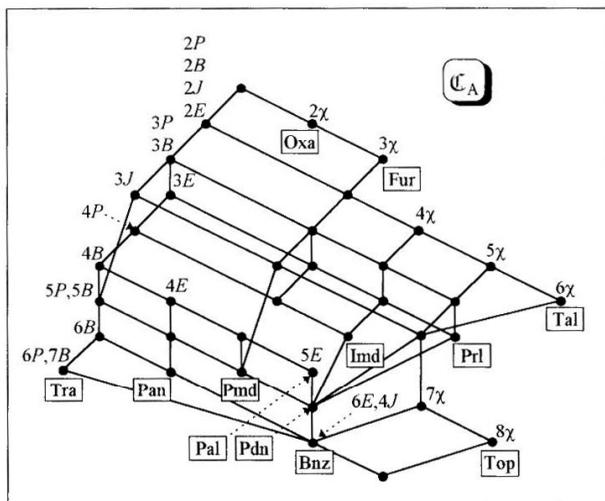
$\mathcal{S}_A^{(i)}$	$1 a_A^{(i)}$	$2 a_A^{(i)}$	$3 a_A^{(i)}$	...	$(w_{\max}^{(i)} - 1) a_A^{(i)}$	$w_{\max}^{(i)} a_A^{(i)}$
1	X			...		
2	X	X		...		
3	X	X	X	...		
⋮	⋮	⋮	⋮	...	⋮	⋮
$w_{\max}^{(i)} - 1$	X	X	X	...	X	
$w_{\max}^{(i)}$	X	X	X	...	X	X

So ist z.B.  $M_A^{(5)} \equiv M_\chi = \{1\chi, 2\chi, \dots, 8\chi\} \subset M_A$ .

Das Liniendiagramm des Begriffsverbands  $(\mathcal{B}(\mathcal{C}_A), \leq)$  zum Kontexts  $\mathcal{C}_A$  zeigt Abbildung 4. Die in [11] und [12] festgestellte sowie in [13]<sup>8</sup> bereits unter

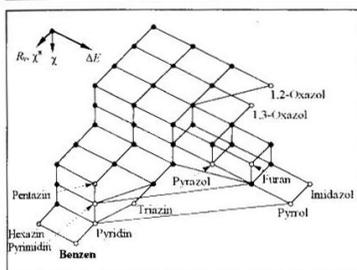
<sup>8</sup> In dieser Arbeit [13] wurde ein Begriffsverband mit folgendem Liniendiagramm errechnet:

Verwendung der FBA diskutierte „Mehrdimensionalität“ der Aromatizität wird auch an der Struktur dieses Begriffsverbandes sofort ablesbar.



**Abb. 4** Das Liniendiagramm des Begriffsverbandes zum Kontext  $C_A$  (bezüglich der Abkürzungen der Gegenstände s. Abbildung 1)

Gleichfalls wird in Abbildung 4 deutlich die chemische Erfahrung wiederspiegelt, nach welcher Benzen und Thiophen sehr ähnliche aromatische Verbindungen sind, von denen sich jeweils eine Klasse weiterer Heterocyclen ableiten lassen. Vom Benzen geht vor allem die Pyridin-Reihe aus, aber auch das Triazin



Zu den hier verwendeten vier Aromatizitätskriterien  $\Delta E$  (energetisch),  $R_v$  (geometrisch, entspricht  $J$ ) und  $\chi^*$  (quantenchemisch berechnete Suszeptibilität) s. [12, 13].  $\chi$  ist die gemessene Suszeptibilität. Die nebenstehend, nicht aber in Abbildung 1 (s.o.) genannten Moleküle sind:



Pentazin



Hexazin



1,2-Oxazol

und das Pyrazin können als Repräsentanten von Heterocyklenklassen angesehen werden. Es ist auch ersichtlich, daß sich die auf das Thiophen folgende Fünfringaromaten, die mindestens ein Heteroatome aus der 6. Hauptgruppe des Periodensystems enthalten, durch kleine *P*-, *B*-, *J*- und *E*-Werte auszeichnen, wobei sie sich durch die Suszeptibilitätswerte unterscheiden. Andererseits besitzen die vom Benzen abgeleiteten Stickstoffaromaten meist kleinere  $\chi$ -Werte, dafür aber höhere Beträge der anderen vier Kriterien, wobei Imidazol und Pyrrol eventuell eine Zwischenstellung einnehmen. Hierin kommt die in [11] angesprochene Unterteilung der Heteroaromaten in „klassische“ und „magnetische“ zum Ausdruck.

Die Klasseneinteilung und Hierarchie der heterocyclischen aromatischen Verbindungen kommt noch besser zum Vorschein, wenn der nicht allzu große Unterschied der  $\chi$ -Werte des Benzens und Thiophens unberücksichtigt bleibt, wie bereits in der letzten Spalte von Tabelle 2 angedeutet. Der auf diese Weise abgeleitete Kontext  $\mathcal{C}_{\bar{A}} = (G_A, M_{\bar{A}}, I_A \cap (G_A \times M_{\bar{A}}))$  unterscheidet sich von  $\mathcal{C}_A$  also nur hinsichtlich der Menge Merkmalsmenge, wobei  $M_{\bar{A}} = M_A \setminus \{8\chi\}$ , d.h. speziell  $M_{\bar{A}}^{(5)} \equiv M_{\bar{\chi}} = \{1\chi, 2\chi, \dots, 7\chi\}$  gilt. Dann ergibt sich ein Begriffsverband, dessen Liniendiagramm Abbildung 5 im Ausschnitt, der die Veränderungen erkennen läßt, zeigt. Hier bildet das Benzen gemäß den Ausführungen in Abschnitt 6 eine Sonderklasse, was seiner besonderen Stellung innerhalb der Klasse der aromatischen Verbindungen durchaus entspricht. Die weiteren Klassen leiten sich wieder vom Thiophen, Pyridin, Triazin und Pyrazin ab.

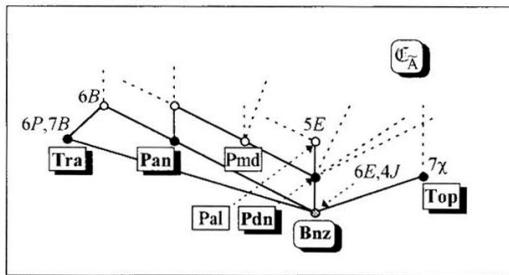


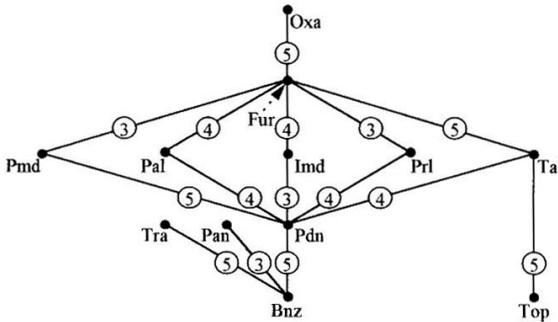
Abb. 5 Liniendiagramm des Begriffsverbandes zum Kontext  $\mathcal{C}_{\bar{A}}$  (Ausschnitt)

## 9. Die formalbegriffsanalytische Klassifikation der Heteroaromaten

Im folgenden werden die Ausführungen der Abschnitte 5 und 6 bezüglich der hier vorgestellten Methode zur disjunkten Klassifikation von Gegenständen am Bei-

spiel der Menge  $G_A$  (Abbildung 1) nach den in Tabelle 2 genannten fünf Aromatizitätskriterien, d.h. basierend auf dem Kontext  $\mathcal{U}_A$  illustriert. Da der sich ergebende Begriffsverband nur 37 Begriffe umfaßt (Abbildung 4), ist seine begriffliche Struktur hinreichend übersichtlich, so daß diese Klassenzerlegung als Interpretationshilfe nicht unbedingt notwendig wäre. Es lassen sich aber gerade an diesem Beispiel gut die Wirksamkeit dieses Verfahrens bzw. dieser Herangehensweise und die mit der FBA verbundenen Möglichkeiten demonstrieren, die bei der Behandlung eines komplexeren Sachverhalts von Nutzen sein werden. Aber auch in diesem Demonstrationsbeispiel wird auf Grund dieser Klassifikation mehr Klarheit erreicht, als sie durch alleiniges Betrachten des Begriffsverbandes erreicht werden kann.

Entsprechend der oben geschilderten Vorgehensweise wird zuerst die Gegenstandsordnung  $\omega(\mathcal{U}_A)$  bezüglich des gesamten Kontexts  $\mathcal{U}_A$  bzw. der zugehörige Graph  $\Gamma(\omega(\mathcal{U}_A))$  ermittelt, der in Abbildung 6 dargestellt ist. Da es in ihm zwei Knoten mit der Eingangswertigkeit null gibt, existiert keine Sonderklasse, und es gilt  $\Gamma(\omega(\mathcal{U}_A)) = \Gamma_{-z}(\omega(\mathcal{U}_A))$  mit  $z = 0$ .



**Abb. 6** Der bogenbewertete Graph  $\Gamma(\omega(\mathcal{U}_A))$  zur Gegenstandsordnung  $\omega(\mathcal{U}_A)$

Entsprechend den beiden Knoten mit der Eingangswertigkeit null wird die Klassenzerlegung zu zwei Klassen führen, die sich vom Benzen bzw. vom Thiophen ableiten. Allerdings ist die Zuordnung der Verbindungen Thiazol, Furan und Oxazol noch nicht klar.

Im Rahmen des Bewertungsmodells  $E_{01}(\mathcal{U}_A, 5)$  müssen die folgenden fünf Teilkontexte von  $\mathcal{U}_A$  gebildet werden:

$$\mathcal{U}_P = (G_A, M_A \setminus M_P, I_A \cap (G_A \times (M_A \setminus M_P))) \quad \text{mit } M_P = \{1P, \dots, 6P\},$$

$$\mathcal{U}_B = (G_A, M_A \setminus M_B, I_A \cap (G_A \times (M_A \setminus M_B))) \quad \text{mit } M_B = \{1B, \dots, 7B\},$$

$$\begin{aligned} \mathcal{C}_J &= (G_A, M_A \setminus M_J, I_A \cap (G_A \times (M_A \setminus M_J))) && \text{mit } M_J = \{1J, \dots, 4J\}, \\ \mathcal{C}_E &= (G_A, M_A \setminus M_E, I_A \cap (G_A \times (M_A \setminus M_E))) && \text{mit } M_E = \{1E, \dots, 7E\}, \\ \mathcal{C}_\chi &= (G_A, M_A \setminus M_\chi, I_A \cap (G_A \times (M_A \setminus M_\chi))) && \text{mit } M_\chi = \{1\chi, \dots, 8\chi\}. \end{aligned}$$

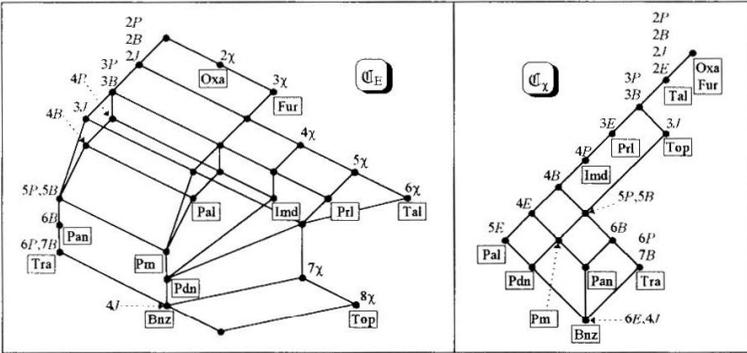
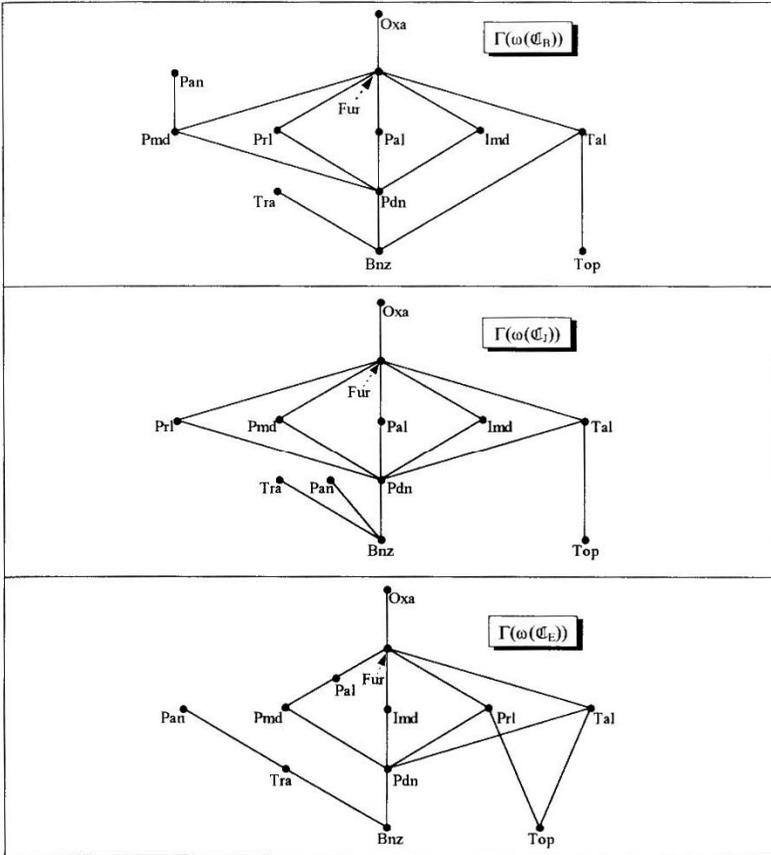


Abb. 7 Liniendiagramme der Begriffsverbände  $\mathcal{B}(\mathcal{C}_E, \leq)$  und  $\mathcal{B}(\mathcal{C}_\chi, \leq)$

In Abbildung 7 sind auswahlweise die Liniendiagramme der Begriffsverbände zu den Kontexten  $\mathcal{C}_E$  und  $\mathcal{C}_\chi$  gezeigt. Der Vergleich mit dem in Abbildung 4 wiedergegebenen Liniendiagramm des Begriffsverbandes zum vollständigen Kontext  $\mathcal{C}_A$  lehrt, daß der Einfluß des magnetischen Kriteriums  $\chi$  auf dessen Struktur wesentlich größer ist als diejenige des energetischen Kriteriums  $E$ . Im ersten Fall ist die Strukturänderung bedeutend umfangreicher als im zweiten. Die Wirkung der geometrischen Kriterien  $P$ ,  $B$  und  $J$  ist derjenigen von  $E$  ähnlich, so daß auch hieran die Unterscheidung von „klassischer“ und „magnetischer“ Aromatizität (s. o.) deutlich wird.

Dieser Tatbestand spiegelt sich selbstverständlich ebenfalls in den Graphen der Gegenstandsordnungen zu den fünf Teilkontexten von  $\mathcal{C}_A$  wider, die in Abbildung 8 gezeigt sind.

Wie in Abschnitt 7 beschrieben, läßt sich mit Hilfe dieser Graphen, die Bewertung der Bögen in  $\Gamma(\omega(\mathcal{C}_A))$  ableiten, die in Abbildung 6 eingetragen wurde. Damit ergeben sich bei Anwendung der Maximalstrategie der aus zwei Bäumen bestehende Wald, den Abbildung 9 zeigt.



**Abb. 8** Graphen der Gegenstandsordnungen zu den Teilkontexten von  $\mathcal{C}_A$ .

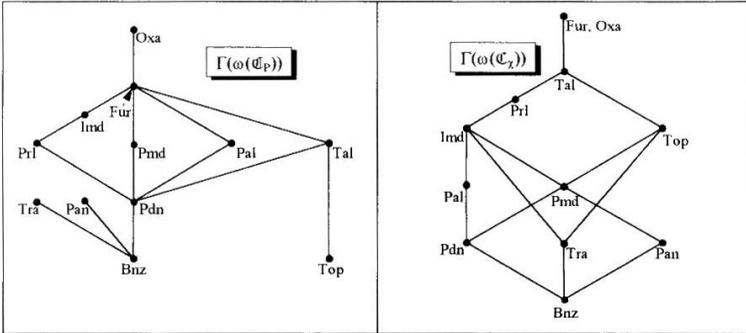


Abb. 8 Graphen der Gegenstandsordnungen zu den Teilkontexten von  $\mathcal{C}_A$  (Fortsetzung)

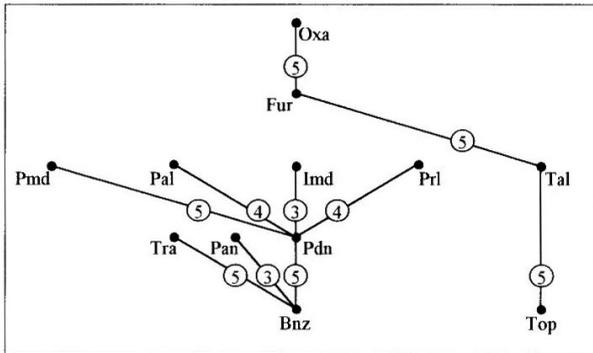
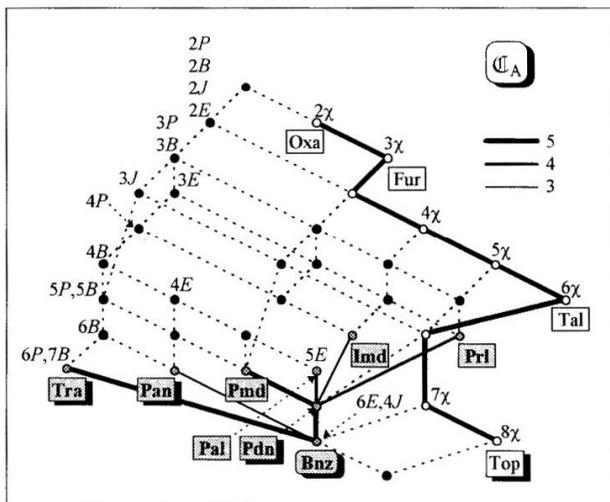


Abb. 9 Der aus  $\Gamma(\omega(\mathcal{C}_A))$  ermittelte zweikomponentige Wald

Es werden somit zwei Klassen erhalten, deren Elemente mit der hinsichtlich der verwendeten Aromatizitätskriterien größten (komplexen) Aromatizität das Benzen und das Thiophen sind, wie es die einleitend erwähnte chemische Erfahrung erwarten läßt. Vom Benzen lassen sich drei Unterklassen ableiten, deren aromatischste Vertreter Sechsringeraromaten mit unterschiedlicher Anzahl an Stickstoffatomen sind, wobei das Triazin und Pyridin stark und das Pyrazin verhältnismäßig schwach mit dem Benzen im Zusammenhang stehen. Auch mit Pyridin sind die weiteren Vertreter dieser Unterklasse in unterschiedlichen Graden zu-

sammenhängend. Nur das Pyrimidin als weiterer Sechsringaromat steht besonders stark mit dem Pyridin in Beziehung. Die Abkömmlinge der Thiophenreihe weisen alle den maximalen Grad der Klassenzugehörigkeit auf.

Den Klasseninhalt hinsichtlich der Aromatizitätskriterien erkennt man, wenn die den Klassen entsprechenden Bäume des Waldes  $\mathcal{Y}(\omega(\mathcal{C}_A))$  in das Liniendiagramm des Begriffsverbandes  $(\mathcal{B}(\mathcal{C}_A), \leq)$  (Abbildung 5) eingetragen werden, wie es in Abbildung 10 geschehen ist.



**Abb. 10** Einordnung der Aromatenklassen in den Begriffsverband  
(Der Klassenzugehörigkeitsgrad wird durch unterschiedliche Linienstärke angedeutet.)

Wie schon oben angedeutet, besitzen die nach der Suszeptibilität hierarchisch geordneten Verbindungen der Thiophenklasse

$$(\text{Thiophen}) = \{\text{Thiophen, Thiazol, Furan, 1,3-Oxazol}\}$$

nur kleine bis mittlere Werte der klassischen Aromatizitätskriterien, so daß diese Klasse in diesem Sinne als die „magnetischen“ Heteroaromaten bezeichnet werden kann. Umgekehrt besitzen alle vom Benzen verschiedenen Vertreter der anderen Klasse nur niedrige bis mittlere Suszeptibilitätswerte. Sie dürfen somit als die klassischen Aromaten angesehen werden. Details, die es hier nicht lohnt

zu diskutieren, aber für die praktische Chemie durchaus nützlich sind, können leicht der Abbildung 10 entnommen werden.

Obwohl das mit dem Kontext  $\mathcal{C}_A$  gewonnene Ergebnis entsprechend der experimentellen Erfahrungen als typischste Vertreter der beiden Klassen das Thiophen und das Benzen aufzeigt, ist es aus mehreren Gründen sinnvoll, das letztere als Sonderklasse zu betrachten. Tatsächlich ist es in der hier betrachteten Menge der aromatischen Verbindungen  $G_A$  die einzige, die kein Heteroatom enthält und zudem, bis auf einen etwas geringeren  $\chi$ -Wert gegenüber dem Thiophen, die größten Werte der Aromatizitätskriterien besitzt. Wie schon oben gezeigt wurde, läßt sich diese Sonderstellung des Benzens bei Verwendung des gegenüber  $\mathcal{C}_A$  nur geringfügig und vertretbar modifizierten Kontexts  $\mathcal{C}_{\bar{A}}$  (s. o.) rechtfertigen bzw. auch sichtbar machen. Das zeigt sich nicht nur an dem in Abbildung 5 angeordneten Begriffsverband  $\mathcal{B}(\mathcal{C}_{\bar{A}}, \leq)$ , sondern auch am Graphen  $\Gamma(\omega(\mathcal{C}_{\bar{A}}))$  zur Gegenstandsordnung  $\omega(\mathcal{C}_{\bar{A}})$  bzw. an den beiden daraus abgeleiteten Untergraphen  $\Gamma_{-1}(\omega(\mathcal{C}_{\bar{A}}))$  und  $\Gamma_{s1}(\omega(\mathcal{C}_{\bar{A}})) = (\{\text{Benzen}\}, \emptyset)$ , die Abbildung 11, zusammen mit der Komponentenerlegung von  $\Gamma_{-1}(\omega(\mathcal{C}_{\bar{A}}))$ , darstellt.

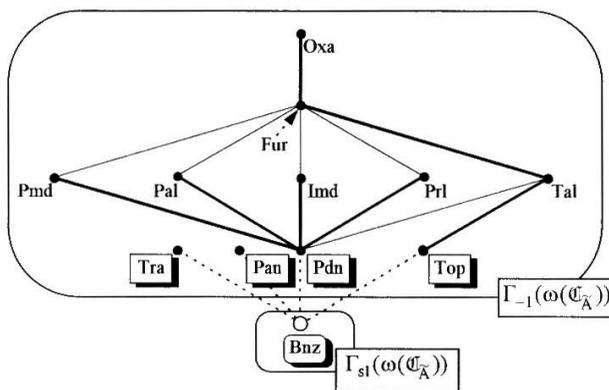
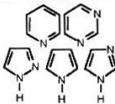
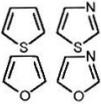


Abb. 11 Die Untergraphen  $\Gamma_{-1}(\omega(\mathcal{C}_{\bar{A}}))$  und  $\Gamma_{s1}(\omega(\mathcal{C}_{\bar{A}}))$  des Graphen  $\Gamma(\omega(\mathcal{C}_{\bar{A}}))$  zur Gegenstandsordnung  $\omega(\mathcal{C}_{\bar{A}})$

Als Resultat werden somit außer der Sonderklasse  $\{\text{Benzen}\}$  vier Klassen aromatischer Verbindungen erzeugt, die in Tabelle 3 einschließlich des aus Ab-

bildung 10 ablesbaren Klasseninhalts und der Klassenstruktur wiedergegeben sind.

**Tabelle 3** Ergebnis der Gegenstandsklassifikation und Klasseninhalte

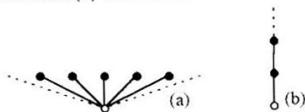
Gegenstände					
klassische Aromatizität	<i>hoch</i>	<i>mittel</i>	<i>hoch</i>	<i>mittel bis hoch</i>	<i>klein bis mittel</i>
magnetische Aromatizität	<i>hoch</i>	<i>mittel</i>	<i>klein</i>	<i>klein</i>	<i>klein bis hoch</i>
Struktur <sup>9</sup>	<i>isoliert</i>	<i>Antikette</i>	<i>isoliert</i>	<i>isoliert</i>	<i>Kette</i>

Berücksichtigt man noch die Möglichkeit, mit Hilfe des Begriffsverbandes Einzelheiten bezüglich der aromatischen Eigenschaften der jeweiligen Vertreter einer Klasse zu erfahren und auf der Basis des Bewertungsmodells Zugehörigkeitsgrade zu einer Klasse abzuschätzen, so wird der Vorteil der formalbegriffsanalytischen Klassifikation gegenüber der Cluster-Analyse im allgemeinen sowie die richtige Widerspiegelung chemischer Zusammenhänge und Erfahrungen und damit der Nutzen für die experimentelle Praxis bei ihrer Benutzung im speziellen offenbar.

## 10. Anhang: Fünf Fälle von Abhängigkeit

Hinsichtlich eines Bogens  $b_{ij} = G_i \triangleright G_j \in B_z$  im Graphen  $\Gamma_{-z}(\omega(\mathcal{U}))$  können beim Übergang von  $\mathcal{U}$  zu  $\mathcal{U}_1$  bzw. von  $\Gamma(\omega(\mathcal{U}))$  zu  $\Gamma(\omega(\mathcal{U}_1))$  fünf Fälle unterschieden werden:

<sup>9</sup> Die Angaben beziehen sich ausdrücklich auf den Kontext  $\mathcal{U}_A$  bzw. auf die untersuchte Gegenstandsmenge  $G_A$ . Unter einer Antikette wird hier die folgende, schematisch wiedergegebene Struktur (a) im Gegensatz zur Kette (b) verstanden:



Fall (A):  $b_{ij}$  ist auch ein Bogen in  $\Gamma_{-z_i}(\omega(\mathcal{C}_i)) = [V^{(1)} \setminus V_{sz_i}, B_{z_i}]$ :  $b_{ij} \in B_{z_i}$ . Der durch  $b_{ij}$  in  $\mathcal{C}$  gegebene Zusammenhang ist (vollkommen) *unabhängig* von der Merkmalsklasse  $M_i$ .  $b_{ij}$  wird die Zahl  $\lambda_c^{(1)}$  zugeordnet, die diese Art der Unabhängigkeit von  $M_i$  bewertet:  $b_{ij} \rightarrow \beta_{ij}^{(1)} = \lambda_c^{(1)} \in \mathbb{R}$ .

Fall (B):  $b_{ij}$  ist ein Bogen in  $\Gamma_{sz_i}(\omega(\mathcal{C}_i)) = [V_{sz_i}, B_{sz_i}]$ ,  $b_{ij} \in B_{sz_i}$ , so daß das Fehlen von  $M_i$  bewirkt, daß die Gegenstände aus  $G_i$  und  $G_j$  nun in einer Sonderklasse vereinigt sind, wobei allerdings die mit  $b_{ij}$  gegebene Ordination erhalten bleibt. Der durch  $b_{ij}$  in  $\mathcal{C}$  gegebene Zusammenhang ist (variiert) *unabhängig* von  $M_i$ .  $b_{ij}$  wird zur Bewertung dieser  $M_i$ -Unabhängigkeit die Zahl  $\lambda_v^{(1)}$  zugeordnet:  $b_{ij} \rightarrow \beta_{ij}^{(1)} = \lambda_v^{(1)} \in \mathbb{R}$ .

Fall (C):  $b_{ij}$  tritt in  $\Gamma(\omega(\mathcal{C}_i))$  nicht auf,  $b_{ij} \notin B^{(1)}$ , es existiert aber ein Knoten  $G_i^{(1)}$  in  $\Gamma_{-z_i}(\omega(\mathcal{C}_i))$ , der durch Kondensation<sup>10</sup> (zumindest) aus den beiden Endknoten  $G_i$  und  $G_j$  von  $b_{ij} = G_i \triangleright G_j$  gebildet gedacht werden kann:  $\exists G_i^{(1)} \in V^{(1)} \setminus V_{sz_i} : G_i \cup G_j \subseteq G_i^{(1)}$ . Der durch  $b_{ij}$  in  $\mathcal{C}$  gegebene Zusammenhang ist (modifiziert) *unabhängig* von  $M_i$ . Die hierarchische Ordnung, welche  $b_{ij}$  ausdrückt, ist verwandelt in eine Äquivalenz bezüglich der Relation  $\cong$ .  $b_{ij}$  wird zur Bewertung dieses Typs der Unabhängigkeit von  $M_i$  die Zahl  $\lambda_m^{(1)}$  zugeordnet:  $b_{ij} \rightarrow \beta_{ij}^{(1)} = \lambda_m^{(1)} \in \mathbb{R}$ .

Fall (D):  $b_{ij}$  tritt in  $\Gamma(\omega(\mathcal{C}_i))$  nicht auf,  $b_{ij} \notin B^{(1)}$ , es existiert aber ein Knoten  $G_{st}^{(1)}$  in  $\Gamma_{sz_i}(\omega(\mathcal{C}_i))$ , von dem man sich vorstellen kann, daß er (zumindest) aus den Endknoten  $G_i$  und  $G_j$  von  $b_{ij} = G_i \triangleright G_j$  durch Kondensation entstanden ist:  $\exists G_{st}^{(1)} \in V_{sz_i} : G_i \cup G_j \subseteq G_{st}^{(1)}$ . Der durch  $b_{ij}$  in  $\mathcal{C}$  gegebene Zusammenhang ist (modifiziert und variiert) *unabhängig* von  $M_i$ .  $b_{ij}$  wird zur Bewertung dieser Art der  $M_i$ -Unabhängigkeit die Zahl  $\lambda_{mv}^{(1)}$  zugeordnet:  $b_{ij} \rightarrow \beta_{ij}^{(1)} = \lambda_{mv}^{(1)} \in \mathbb{R}$ .

<sup>10</sup> Zum graphentheoretischen Begriff der Kondensation vgl. beispielsweise [38, S. 168]. Schematisch läßt sich diese Operation folgendermaßen darstellen:



Fall (E):  $b_{ij}$  tritt in  $\Gamma(\omega(\mathcal{C}_i))$  nicht auf, wohl aber die Endknoten  $G_i$  und  $G_j$  von  $b_{ij}$ :  
 $b_{ij} \notin B^{(1)}$ , aber  $G_i, G_j \in V^{(1)}$ . Der durch  $b_{ij}$  in  $\mathcal{C}$  gegebene Zusammenhang ist *abhängig* von der Existenz der Merkmalsklasse  $M_i$ . Dem Bogen  $b_{ij}$  wird zur Bewertung der Abhängigkeit von  $M_i$  die Zahl  $\mathfrak{G}^{(1)}$  zugeordnet:  $b_{ij} \rightarrow \beta_{ij}^{(1)} = \mathfrak{G}^{(1)} \in \mathbb{R}$ .

Die fünf Fälle der Unabhängigkeit/Abhängigkeit einer Beziehung  $b_{ij}$  werden für die  $i$ -te Komponente der Zerlegung der Merkmalsmenge  $M$  in dem Vektor  $\Lambda^{(1)} = (\mathfrak{G}^{(1)}, \lambda_c^{(1)}, \lambda_v^{(1)}, \lambda_m^{(1)}, \lambda_{mv}^{(1)})$  zusammengefaßt.

## 11. Literatur

- [1] E. v. MEYER, *Geschichte der Chemie von den ältesten Zeiten bis zur Gegenwart, zugleich eine Einführung in das Studium der Chemie*, Verlag von Veit & Comp., Leipzig, 1905<sup>3</sup>
- [2] A. KEKULÉ, *Bull. Soc. Chim.*, 98 (1865), *Liebigs Ann. Chem.* 137, 129 (1866)
- [3] E. HÜCKEL, *Z. Phys.* 70, 204 (1931), 72, 310 (1931), 76, 628 (1932), 83, 632 (1933)
- [4] E. HÜCKEL, *Grundzüge der Theorie ungesättigter und aromatischer Verbindungen*, Verlag Chemie, Berlin, 1938
- [5] V. MEYER, *Die Thiophengruppe*, Verlag F. Vieweg & Sohn, Braunschweig, 1888
- [6] O. DIELS, *Einführung in die organische Chemie*, Verlagsbuchhandlung von J.J. Weber, Leipzig, 1939<sup>9</sup>, S. 268
- [7] W. LANGENBECK, *Lehrbuch der organischen Chemie*, Verlag von Th. Steinkopff, Dresden/Leipzig, 1962<sup>19</sup>, S. 210
- [8] V. MEYER, *Ber. Dtsch. Chem. Ges.* 15, 2893 (1882)
- [9] V. MEYER, *Ber. Dtsch. Chem. Ges.* 16, 1465 (1883)
- [10] R. MEYER, *Victor Meyer – Leben und Wirken eines deutschen Chemikers und Naturforschers 1848–1897* (W. OSTWALD (Hrsg.), *Grosse Männer – Studien zur Biologie des Genies, Vierter Band*), Akademische Verlagsgesellschaft, Leipzig, 1917, S. 379
- [11] A.R. KATRITZKY, P. BARCZYNSKI, G. MUSUMARRA, D. PISANO and M. SZAFRAN, *J. Am. Chem. Soc.* 111, 7 (1989)
- [12] K. JUG und A. M. KÖSTER, *J. Phys. Org. Chem.* 4, 163 (1991)
- [13] H.-G. BARTEL, *match* 30, 9 (1994)
- [14] H.-G. BARTEL, *Concept Analysis and Chemical CAD*, in: *Informatik-Skripten der Humboldt-Universität zu Berlin*, Heft VII, Juni 1989, S. 72-75
- [15] H.-G. BARTEL, *Selbstorganisation* 1, 99 (1990)
- [16] H.-G. BARTEL, *match* 32, 27 (1995)
- [17] H.-G. BARTEL, *Formal Concept Analysis Studies: Roman Bricks and Wall Slabs*, in: Ş. DEMIRCI, A. M. ÖZER und G. D. SUMMERS (eds.), *Archaeometry 94. The Proceedings*

- of the 29<sup>th</sup> International Symposium on Archaeometry, Ankara 9-14 May 1994, Tübitak, Ankara 1996, S. 309-316
- [18] H.-G. BARTEL, *Archäometrie und Denkmalpflege* 1995, 159
- [19] H.-G. BARTEL und M. NOFZ, *Chemometrics and Intelligent Laboratory Systems*, im Druck
- [20] H.-G. BARTEL, *Explorative Data Analysis Studies: Chemical Composition of Ancient Egyptian Bronze Artifacts*, in: Proceedings of the 1<sup>st</sup> International Conference on Ancient Egyptian Mining, Metallurgy and Conservation of Metallic Artifacts, 10-12 April 1995, Cairo, im Druck
- [21] H.-G. BARTEL, *match* 34, (1996), im Druck
- [22] H.-G. BARTEL, *Über Möglichkeiten der Formalen Begriffsanalyse in der Mathematische Archäochemie*, in: R. WILLE (Hrsg.), *Begriffliche Wissensverarbeitung - Methoden und Anwendungen*, Springer-Verlag, Heidelberg, im Druck
- [23] H.-G. BARTEL, *Mathematische Methoden in der Chemie*, Spektrum Akademischer Verlag, Heidelberg/Berlin/Oxford, 1996, S. 243-301
- [24] R. WILLE, *Restructuring lattice theory: An approach based on hierarchies of concepts*, in: I. RIVAL (ed.), *Ordered Sets*, Reidel, Dordrecht-Boston, 1982, S. 445-470
- [25] H.-G. BARTEL, M. NOFZ, R. STÖBER und M. KUPFER, *Z. phys. Chem. (Leipzig)* 271, 159 (1990)
- [26] H.-G. BARTEL, *Monatsh. Chem.* 121, 237 (1990)
- [27] H.-G. BARTEL, M. NOFZ und F.G. WIHSMANN, *Chemometrics and Intelligent Laboratory Systems* 8, 73 (1990)
- [28] H.-G. BARTEL, *Syst. Anal. Model. Simul.* 7, 557 (1990)
- [29] H.-G. BARTEL, *Syst. Anal. Model. Simul.* 8, 139 (1991)
- [30] A.R. KATRITZKY, M. KARELSON und N. MALHOTRA, *Heterocycles* 32, 127 (1991)
- [31] H. SPÄTH, *Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion*, R. Oldenbourg Verlag, München/Wien, 1975
- [32] F. EHRENTREICH, M. NOFZ und H.-G. BARTEL, *Chemometrics and Intelligent Laboratory Systems* 15, 61 (1992)
- [33] R. WINDERLICH, *Kekulé*, in: G. BUGGE (Hrsg.), *Das Buch der großen Chemiker*, 2. Band, Verlag Chemie, Weinheim/Bergstraße, 1955<sup>2</sup>, S. 206
- [34] R. WILLE, *Bedeutungen von Begriffsverbänden*, in: B. GANTER, R. WILLE und K.E. WOLFF (Hrsg.), *Beiträge zur Begriffsanalyse*, BI-Wissenschaftsverlag, Mannheim, 1987, S. 161-211
- [35] B. GANTER und R. WILLE, *Formale Begriffsanalyse - Mathematische Grundlagen*, Springer-Verlag, Berlin/Heidelberg/New York, 1996
- [36] K.E. WOLFF, *A first course in formal concept analysis - How to understand line diagrams*, in: F. FAULBAUM (ed.), *SoftStat '93, Advances in Statistical Software* 4, 429 (1994)
- [37] H. SACHS, *Einführung in die Theorie der endlichen Graphen, Teil 1*, B.G. Teubner Verlagsgesellschaft, Leipzig, 1970
- [38] C. POSTHOFF, D. BOCHMANN und K. HAUBOLD, *Diskrete Mathematik* (Mathematisch-Naturwissenschaftliche Bibliothek, Bd. 70), B.G. Teubner Verlagsgesellschaft, Leipzig, 1986

- [39] H.-G. BARTEL, A. PAPE, G. SCHNEIDER und C. WINKLER, *Eine mathematisch explorative Analyse archäometrischer Daten von Keramikfunden aus Tell el-Dab'a*, in: F. TIRADRITTI (ed.), *Informatica ed egittologia all'inizio degli anni '90*, Bulzoni Editore, Rom, 1996, S. 19-32
- [40] P. BURMEISTER, *Conlmp – Programm zur Formalen Begriffsanalyse einwertiger Kontexte*, TH Darmstadt, FB Mathematik, 1991
- [41] N. SPANGENBERG und K. E. WOLFF, *Comparison between Principal Component Analysis and Formal Concept Analysis of Repertory Grids*, in: W. LEX (Hrsg.), *Arbeitstagung Begriffsanalyse und Künstliche Intelligenz*, Informatik-Berichte 89/3. TU Clausthal, 1991, S. 127-134
- [42] N. SPANGENBERG und K. E. WOLFF, *Comparison between Biplot Analysis and Formal Concept Analysis of Repertory Grids*, in: H. H. BOCK und P. IHM (Hrsg.), *Classification, Data Analysis, and Knowledge Organization*, Springer-Verlag, Berlin/Heidelberg/New York, 1991b, S. 104-112
- [43] K. E. WOLFF, *Comparison of Graphical Data Analysis Methods*, in: F. FAULBAUM und W. BANDILLA (Hrsg.), *SoftStat '95, Advances in Statistical Software* 6, 139 (1996)
- [44] K. E. WOLFF und S. GABLER, *Comparison of Visualizations in Formal Concept Analysis und Correspondence Analysis*, in: M. GREENACRE und J. BLASIUS (eds.), *Visualization of Categorical Data*, Academic Press, Boston, im Druck