

Unscharfe Wortbedeutungen.

Ein quantitatives Verfahren zur lexikalischen Analyse des verwendeten Vokabulars im Rahmen eines Strukturmodells unscharfer (fuzzy) Semantik*

Burghard Rieger

Inhaltsverzeichnis

0	Aufgabenstellung	1
0.1	Kurzfassung der Ergebnisse	2
1	Unschärfeproblematik in der Linguistik	3
1.1	Das Konzept der unscharfen Mengen	4
1.2	Anwendung auf sprachliche Bedeutungen	5
1.3	Empirisch-operationale Rekonstruktion	7
2	Modellbildung einer unscharfen Lexikonstruktur	8
2.1	Der formale Aufbau des Modells	9
2.2	Zugeordnete empirische Verfahren	12
2.3	Quantitative Analyse der Daten	14
3	Darstellung der Ergebnisse	19
3.1	Vokabular-Deskriptoren	19
3.2	Unterscheidende Aspekte	25
3.3	Gemeinsame Aspekte	30
4	Literatur	34

0 Aufgabenstellung

Das Vorhaben der Bonner Forschungsstelle für öffentlichen Sprachgebrauch des Instituts für deutsche Sprache, den öffentlichen Sprachgebrauch beider deutscher Staaten anhand von Presstexten der DDR bzw. BRD zu untersuchen, hat sich zur Aufgabe gemacht, nachweisbare Unterschiede und Abweichungen sprachlicher Entwicklungen aufgrund

*Erschienen in: Hellmann, M.W. (Hrsg.): Ost-West-Wortschatzvergleiche. Maschinell gestützte Untersuchungen zum Vokabular von Zeitungstexten aus der BRD und der DDR. [Forschungsberichte des Instituts für Deutsche Sprache Mannheim 48] Tübingen (Narr) 1984, S. 293–339.

eines aus Texten der Zeitungen ‚Neues Deutschland‘ (ND) und ‚Die Welt‘ (WE) bestehenden Zeitungskorpus aufzudecken und zu beschreiben (SCHAEDER 1978). Im Rahmen des von der Deutschen Forschungsgemeinschaft geförderten Projekts „Ost-West-Wortschatzvergleiche“ taucht dabei unter anderem das Problem auf, jenen Teil der Lexik des Deutschen zu beschreiben, der als BRD bzw. DDR-spezifisch gelten kann (HELLMANN 1976, 1979). Dazu ist ein Vergleich der semantischen Werte notwendig, die sich möglicherweise aufgrund eines unterschiedlichen Gebrauchs gleicher Lexeme im ost- bzw. westdeutschen Sprachraum entwickelt haben.

Für den lexikalisch-semantischen Bereich, in dem sich — anders als in Morphologie und Syntax — sprachliche Unterschiede noch am ehesten niedergeschlagen haben, ergibt sich dabei ein doppeltes Problem, nämlich

1. wie lassen sich mit Hilfe maschineller Verfahren der linguistischen Textanalyse die im Bonner Zeitungskorpus auf der Wortebene vermuteten semantischen BRD- bzw. DDR-Spezifika ermitteln, und
2. wie lassen sich die ermittelten Bedeutungsdifferenzen für ein zu erstellendes Korpuswörterbuch des Ost-West-Wortschatzes am geeignetsten darstellen?

Als zentrale Schwierigkeit für beide Problemaspekte muß dabei die Forderung gelten, daß durch das zu entwickelnde bzw. zu erprobende Verfahren möglichst auch schon geringe Abweichungen und Unterschiede der Lexik sollen ermittelt und abgebildet werden können. dabei ist gleichzeitig das operationale Auflösungsvermögen des Verfahrens nicht so weit zu treiben, daß etwa übergreifende Ähnlichkeiten und Gemeinsamkeiten im Ost-West-Wortschatz nicht mehr erkannt werden können. Das Verfahren zur Analyse wie zur Repräsentation der ermittelten Bedeutungen darf daher weder so grob sein, daß einerseits geringe Bedeutungs-differenzierungen übersehen werden, noch so fein sein, daß andererseits größere Bedeutungszusammenhänge unerkannt bleiben.

Die folgenden Ausführungen sind das Resultat einer Überprüfung und Erprobung der Anwendbarkeit eines vom Verfasser entwickelten maschinellen Verfahrens zur lexikalisch-semantischen Beschreibung von Wortbedeutungen. Für die Zielsetzung des Projekts „Ost-West-Wortschatzvergleiche“ war zu ermitteln, in wieweit das Modell einsetzbar ist im Hinblick auf

3. die Entdeckung unterschiedlichen Sprachgebrauchs im Textmaterial und
4. die Beschreibung unterschiedlichen Sprachgebrauchs im Lexikon/Wörterbuch.

Hierzu wurden zwei Datenbänder zur Verfügung gestellt, die eine Auswahl von Texten (WE und ND) in numerisch aufbereiteter Form zugänglich machten. Die Aufbereitung besteht im wesentlichen in einer formatierenden Übersetzung der Zeitungstexte, d.h. in der Erfassung solcher unter „Lemmata“ zusammengefaßten Wörter und Wortgruppen, deren auf die Stämme morphologisch reduzierten Kerne in jedem der Auswahlkorpora eine Frequenz von mindestens 5 Belegen erreichten. Sie wurden als Vokabular-Einträge erfaßt und durch natürliche Zahlen indiziert. Die Datenbänder verzeichnen daher nurmehr durch Textende-Marken unterteilte Lemma-Nummernfolgen, d.h. die Nummern der Lemmata in der Reihenfolge, in der die zugehörigen Wortbelege in den Texten des bearbeiteten Experimentierkorpus vorkommen. Diese Daten bilden die Grundlage für die vorliegende Pilotuntersuchung.

0.1 Kurzfassung der Ergebnisse

Die aus statistischen Korrelationsanalysen von lemmatisierten Texten der geschriebenen Sprache ermittelten Strukturdaten erlauben eine konsekutive Abbildung der Regularitäten, durch die die Verwendungsweisen von Wörtern, ihre Unterschiede und deren Zusammenhänge als Bedeutungssystem repräsentiert werden können. Diese Beschreibung der Wortbedeutungen in ihrem strukturalen Zusammenhang bedient sich der Notationen der Theorie der unscharfen (fuzzy) Mengen. Zur Kennzeichnung der Bedeutung eines Elements des Vokabulars dient dabei je eine unscharfe Teilmenge des Vokabulars, die sich aus den Regelmäßigkeiten ergibt, mit denen jedes einzelne Wort des Vokabulars durch seine Verwendung in den analysierten Texten zur Konstitution der betreffenden Bedeutung beiträgt.

Neben den Gebrauchsregularitäten $G(x)$ eines Wortes x des Vokabulars werden die Bedeutungen $B(x)$ des Wortes x , sowie dessen Umgebungen $E(x)$ in Form von Wortlisten charakterisiert, die nach abnehmenden Zugehörigkeitsgraden (Korrelations- bzw. Distanzwerte) geordnet sind.

Die für jedes der Zeitungskorpora WE bzw. ND separat erstellten Auswertungen lassen den unterschiedlichen Gebrauch gleicher Wörter in $G(x)$, die unterschiedlichen Bedeutungen in $B(x)$ und die unterschiedlichen Umgebungen in $E(x)$ ablesen. Für die Darstellung im Lexikon sind wohl die in $B(x)$ gegebenen Beschreibungen am geeignetsten. Sie bieten dem Lexikonbenutzer die aus der Gesamtheit der Wortverwendungen kondensierten Daten intersubjektiv überprüfbar an, ohne diese Daten anders als auf der theoretischen Grundlage ihrer methodischen Erhebung und Verarbeitung zu interpretieren. Vor diesem Hintergrund sind die aus ihnen zu ziehenden Schlüsse nicht als für den Benutzer unkontrollierbare Interpretationen des Lexikon-Erstellers vorgegeben, sondern vom Lexikon-Benutzer selbst zu leisten und zu kontrollieren.

1 Unschärfeproblematik in der Linguistik

Für die linguistischen Theorien und Methoden der Beschreibung und Analyse sprachlich-kommunikativer Phänomene ist eine vergleichsweise klassische Sichtweise noch weitgehend bestimmend. Diese als ‚categorical-view‘ (LABOV 1970) bezeichnete Auffassung geht davon aus, daß linguistische Entitäten — auf welcher der möglichen (phonologisch, morphologisch, syntaktisch, semantisch, pragmatisch bestimmten) Ebenen auch immer — diskret, invariant, qualitativ distinkt, aus atomaren Komponenten aufgebaut, durch endliche Konjunktionen definierbar und deswegen mit empirisch-statistischen Verfahren nur unzulänglich erfaßbar seien. Zugehörigkeit solcher Einheiten zu Kategorien bzw. relationale Verknüpfungen zwischen diesen Kategorien werden daher bisher noch — von wenigen Ausnahmen (etwa SUPPES 72, LABOV 73, KLEIN 74) abgesehen — durch binäre Regeln streng deterministischer Art repräsentiert. Diese können folglich kontinuierliche Übergänge wahrscheinlicher oder möglicher Zugehörigkeiten nicht abbilden. Sie erweisen sich denn auch als ungeeignet, Erscheinungen gerade jenes Bereichs adäquat zu erfassen, der unter dem Begriff der sprachlichen *Variabilität* (KLEIN 1976) einerseits, unter dem der *Vagheit* (RIEGER 1974) andererseits im Rahmen der Semantik zunehmend in den Mittelpunkt linguistischen Forschungsinteresses gerückt ist. dabei handelt es sich einmal um die durch individuelle, situative, soziale, historisch oder wie immer beeinflusste Abweichungen von normierend abstrahierten Regeln des bedeutungskonstituierenden Sprachgebrauchs (vgl. RIEGER 1977 b), zum anderen um die in jeder Form natürlichsprachlicher Zeichen- und Symbolverwendung zu kommunikativen

Zwecken enthaltenen semantischen Unschärfen und Ungenauigkeiten.

Beides, Variabilität und Vagheit sind aber Eigenschaften, deren Berücksichtigung als entscheidend für eine Lösung des anstehenden Problems lexikalisch-semantischer Bedeutungs-Analyse und -Repräsentation gelten muß. Diese Eigenschaften, die nach einigen frühen Ansätzen im 19. Jahrhundert Gegenstand einer ersten philosophisch akzentuierten Phase intensiven Studiums während der 20er und 30er Jahre unseres Jahrhunderts waren, wurden erst mit dem Erscheinen der mathematischen Theorie der unscharfen Mengen (fuzzy sets theory) als genuin linguistischer Phänomenbereich Gegenstand eines übergreifenden Neuansatzes. ZADEHs Theorie (1965) und die seither auf ihr aufbauenden Modellbildungen (GAINES 1976), haben inzwischen Notationen bereitgestellt, die es möglich erscheinen lassen, sowohl zu numerischen Analysen als auch zu formalen Darstellungen der semiotischen Prozesse in Zeichenkonstitution und -verwendung zu gelangen.

Für die Bedeutungsanalyse und -repräsentation gerade geringerer lexikalisch-semantischer Differenzierungen in natürlich-sprachlichen Texten ergeben sich daher in der unscharfen Theorie- und Modellbildung vermutlich noch die aussichtsreichsten Ansätze zu formal adäquateren, dabei empirisch fundierten Abbildungen von Wortbedeutungen im Vergleich zu dem, was die bekannten traditionellen Versuche hierzu bisher haben leisten können. Deswegen zunächst eine kurze Einführung der wesentlichen Charakteristika dieses Neuansatzes (1.1), seiner Anwendung innerhalb eines denotativen Bedeutungsmodells (1.2), sowie der Bedingungen seiner empirisch-operationalen Rekonstruktion (1.3).

1.1 Das Konzept der unscharfen Mengen

Der Grundgedanke der Theorie der unscharfen Mengen, die die traditionelle Mengentheorie als Grenzfall enthält, ist denkbar einfach und plausibel. Im Unterschied zur klassischen oder ‚scharfen‘ Mengentheorie, in der ein Individuum alternativ im Hinblick auf eine Menge entweder Element ist oder nicht, kann man in der neuen Theorie die Zugehörigkeit eines Individuums zu einer deswegen ‚unscharf‘ genannten Menge graduell angeben. Danach kann die charakteristische Funktion $\mu_A(x)$ eines Elements x der Menge A nicht nur die Werte 0 (nicht-zugehörig) und 1 (zugehörig) annehmen, sondern auch jeden beliebigen anderen Wert zwischen 0 und 1, wobei $\mu_A(x) = 0,2$ eine geringere Zugehörigkeit des Elements x zur Menge A anzeigt, als $\mu_A(x) = 0,8$.

Allgemein wird eine unscharfe Teilmenge A von X charakterisiert durch die Zugehörigkeitsfunktion

$$\mu_A : X \rightarrow [0, 1] \quad (1)$$

die jedem $x \in X$ einen (und nur einen) Zugehörigkeitswert $\mu_A(x)$ aus dem Intervall $[0, 1]$ zuordnet, der den Grad angibt, mit dem das Individuum x als Element der unscharfen Menge A zu gelten hat. Die unscharfe Menge A besteht also aus der Menge der geordneten Paare

$$A := \{ (x, \mu_A(x)) \} \text{ für alle } x \in X. \quad (2)$$

Gemäß der klassischen Definitionen der Verknüpfungsoperationen wird man auch für beliebige unscharfe Mengen A und B fordern, daß die Zugehörigkeitswerte der neu entstehenden unscharfen Menge C sich bei *Durchschnittsbildung* nicht erhöhen und bei *Vereinigung* nicht vermindern; ZADEH gibt daher die jeweils niedrigst bzw. höchst möglichen

Werte als Definition

$$C = A \cap B \quad :\Leftrightarrow \quad \mu_D(\mathbf{x}) = \min(\mu_A(\mathbf{x}), \mu_B(\mathbf{x})) \quad (3)$$

$$C = A \cup B \quad :\Leftrightarrow \quad \mu_D(\mathbf{x}) = \max(\mu_A(\mathbf{x}), \mu_B(\mathbf{x})) \quad (4)$$

und für die *Komplementbildung* folgende Anweisung

$$B = \neg A \quad :\Leftrightarrow \quad \mu_B(\mathbf{x}) = 1 - \mu_A(\mathbf{x}) \quad (5)$$

wobei sich *Gleichheit* und *Enthaltensein* wie folgt definieren

$$A \equiv B \quad :\Leftrightarrow \quad \mu_A(\mathbf{x}) = \mu_B(\mathbf{x}) \quad (6)$$

$$A \subseteq B \quad :\Leftrightarrow \quad \mu_A(\mathbf{x}) \leq \mu_B(\mathbf{x}) \quad (7)$$

Diese Definitionen reduzieren sich auf die klassischen Mengen, wenn man die Zugehörigkeitsfunktion auf die binär zulässigen Werte 0 und 1 einschränkt. Damit werden nun auch solche Gegenstandsbereiche im Prinzip mengentheoretisch erfaßbar, deren fließende Übergänge, ungenaue Abgrenzbarkeit oder unvollständige Kenntnis als Hauptcharakteristika ihrer Verschwommenheit und Vagheit eine exaktwissenschaftliche Behandlung bisher verhinderten. Das soll im folgenden anhand meines Standardbeispiels (RIEGER 1977a, 1979a, b), eines einfachen Begriffs wie ‚Mittelklassewagen‘, verdeutlicht werden. Dieser Begriff, aus der Automobilwerbung jedem von uns geläufig, ist *unscharf* in bezug auf die Menge derjenigen Fahrzeugtypen, die gemeint sind, wenn von ‚Mittelklassewagen‘ die Rede ist. Einem Fiat 500 beispielsweise würde deshalb in dieser Menge der ‚Mittelklassewagen‘ ein äußerst geringer Zugehörigkeitswert zukommen, weil er als ausgesprochener Kleinwagen gilt. Wenn man seine ‚Mittelklassehaftigkeit‘ anhand des Hubraums bewerten sollte, würde man ihm den Wert 0,0 geben müssen. Ein VW-Golf, beinahe schon ein Mittelklassewagen, zumindest seinem Motorvolumen nach, hätte einen deutlich höheren Wert, etwa 0,5, ein Opel-Rekord, als typischer Mittelklassewagen, den Wert 1,0, ein Mercedes, eher schon ein Luxuswagen, einen wieder deutlich niedrigeren Wert, beispielsweise 0,3, während ein Rolls Royce, als reine Luxuslimousine, etwa einen Zugehörigkeitswert von exakt 0,0 zur unscharfen Menge M der ‚Mittelklassewagen‘ aufwiese.

Trägt man nun zur Veranschaulichung (Fig. 1) den Individuenbereich X der Einfachheit halber als kontinuierliche Hubraumskala von links nach rechts auf der Abszisse und die einzelnen — den verschiedenen Fahrzeugtypen (subjektiv) zugeschriebenen — Zugehörigkeitswerte $\mu_M(x)$ auf der Ordinate auf, so ergibt die unscharfe Menge M eine Kurve, die als formale Darstellung der (subjektiven) referentiellen Bedeutung von ‚Mittelklassewagen‘ über X gelten kann.

1.2 Anwendung auf sprachliche Bedeutungen

Als ZADEH (1971) sein Konzept der unscharfen Mengen erstmals auf linguistische Bedeutungsphänomene anwandte, ging er im Rahmen der Semantik von einem strikt *referenztheoretischen* Modell aus. Vagheit natürlich-sprachlicher Bedeutung ergab sich folglich als eine unscharfe Teilmenge aller möglichen Extensionen, auf die ein sprachlicher Ausdruck referiert.

Thus, a word like ‘green’ is a name for a class in which the transition from membership to non-membership is gradual rather than abrupt. The same is true of phrases such as ‘beautiful women’, ‘tall buildings’, ‘large integers’, etc. In fact, it may be argued that in the case of

	Fiat	VW Golf	Opel Rekord	Mercedes	Rolls Royce
X	f	g	o	m	r
ccm	500	1100	1700	3200	6300
$\mu_M(x)$	0,0	0,5	1,0	0,3	0,0

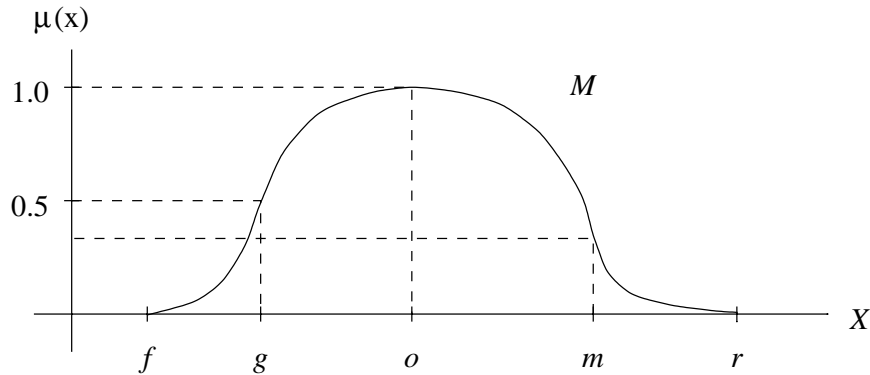


Fig. 1

natural languages, most of the words occurring in a sentence are names of fuzzy rather than non-fuzzy sets, with the sentence as a whole constituting a composite name for a fuzzy subset of the universe of discourse. (S. 160)

Dazu wird zunächst die endliche Menge der sprachlichen Ausdrücke

$$T := \{x_i\}, \quad i = 1, \dots, n \quad (8)$$

und eine Menge von Gegebenheiten (universe of discourse)

$$U := \{y_j\}, \quad j = 1, \dots, m \quad (9)$$

eingeführt, die hier informell als Ansammlung von Objektpunkten charakterisiert werden mag, welche jedoch weit und reich genug definiert ist, daß jeder in ihr mögliche *Begriff* (concept) mit einer unscharfen Teilmenge von U identifiziert werden kann.

Als *Bedeutung* eines Ausdrucks $x \in T$ kann so die unscharfe Teilmenge $M(x)$ in U erklärt werden, die sich über eine auf x restringierte Zugehörigkeitsfunktion $\mu(y_j|x)$ definieren läßt. Von daher erscheint der Ausdruck x als *Name* der unscharfen Menge $M(x)$, die einen *Begriff* in U repräsentiert.

In diesem referenziellen Sinne läßt sich die Bedeutung sprachlicher Ausdrücke nun allgemein als System verstehen, das eine Korrespondenz herstellt zwischen Termen x in T und Mengen von Objektpunkten y in U . Diese Korrespondenz ist eine nicht-eindeutige Beziehung. Sie kann als unscharfe Relation L aufgefaßt werden, die sich über die Zugehörigkeitsfunktion

$$\mu_L : T \times U \rightarrow [0, 1] \quad (10)$$

definieren läßt. Damit wird — entsprechend (1) — jedem geordneten Paar (x, y) ein (und nur ein) Zugehörigkeitswert $\mu_L(x, y)$ in L zugeordnet.

Die unscharfe Relation L induziert also eine Korrespondenz zwischen Elementen $x \in T$ und unscharfen Mengen in U derart, daß einem sprachlichen Ausdruck x die unscharfe Menge $M(x) \subseteq U$ entspricht, die seine *Bedeutung* heie. Deren auf x eingeschränkte Zugehörigkeitsfunktion

$$\mu_{(y_j|x)} = \mu_L(x, y_j) \quad (11)$$

ergibt sich aus der Menge der Zugehörigkeitswerte $\mu_{M(x)}(y_j)$, die den Objektpunkten y_j zur unscharfen Menge $M(x)$ zukommen

$$\mu_{M(x)}(y_j) = \mu_L(x, y_j) \quad (12)$$

Betrachtet man umgekehrt ein bestimmtes Element $y \in U$, dann definiert die auf y eingeschränkte Zugehörigkeitsfunktion der Inversion $\mu_{L^{-1}}(x_i, y)$ eine unscharfe Menge $D(y)$ in T , in der jedes x folgenden Zugehörigkeitsgrad hat

$$\mu_{D(y)}(x_i) = \mu_{L^{-1}}(x_i, y) \quad (13)$$

ZADEH nennt diese unscharfe Menge $D(y)$ *Deskriptormenge* (descriptor set) weil sie dazu geeignet ist, den Grad anzugeben, mit dem jeder Ausdruck in T zur Beschreibung eines bestimmten Objektpunktes in U beiträgt.

In summary, a language L is a fuzzy relation from T to U characterized by a membership function $\mu_L(x_i, y_j)$.

As a relation, L associates with each term x in T its meaning, $M(x)$, which is a fuzzy set in U defined by $\mu_{M(x)}(y_j) = \mu_L(x, y_j)$. Furthermore, L associates with each element y of U a fuzzy descriptor set, $D(y)$, defined by $\mu_{D(y)}(x_i, y)$. (S. 168)

Damit ist eine Abbildung von *Bedeutungen* über unscharfe Mengen von Objektpunkten $M(x)$ bzw. eine sprachliche Repräsentation von Begriffen über unscharfe Mengen von *Deskriptoren* $D(y)$ erklärt. Diese können — auch unabhängig von dem hier zugrundeliegenden referentiellen Semantikmodell — zumindest ‚formal adäquat‘ im Sinne des Phänomens der Vagheit genannt werden, insofern ihr ein ‚unscharfes‘, kein klassisches Mengenkonzept zugrundeliegt. Allerdings bleibt die empirische Seite dieses Angangs undiskutiert, die Frage danach also, wodurch die derart als unscharfe Mengen beschreibbaren Bedeutungen von Ausdrücken festgestellt, wie die Zugehörigkeitsgrade tatsächlich ermittelt und über welchen aller möglichen Deskriptormengen sie beschrieben werden können.

1.3 Empirisch-operationale Rekonstruktion

Wie das zitierte Mittelklassewagen-Beispiel zur Einführung der unscharfen Mengen (Fig. 1) verdeutlicht, wird die oben angesprochene Forderung der Adäquatheit der Beschreibung natürlichsprachlicher Bedeutung einzelner Wörter durch unscharfe Mengen nur zum Teil, nämlich formal erfüllt. Diese formale Darstellung hängt aber entscheidend ab

- a) von dem jeweils zugrunde gelegten Individuenbereich, der als Deskriptormenge fungiert (z.B. Skala der möglichen Motorvolumen), in dem die unscharfe Menge (z.B. Mittelklassewagen) definiert wird und

- b) von dem Verfahren, aufgrund dessen einem Individuum (z.B. Golf = 0,5) ein Zugehörigkeitswert in bezug auf die zu definierende unscharfe Menge (z.B. Mittelklassewagen) zugeschrieben wird.

Beides, die Bestimmung des Individuenbereichs als Deskriptormenge wie das Verfahren zur Ermittlung von Zugehörigkeitswerten, betrifft aber schon Fragen einer Analyse natürlich-sprachlicher Bedeutung von Wörtern, die ohne den semiotischen Zusammenhang von Zeicheninventar, Anwendungskontext und kommunikativem Gebrauch der Sprecher/Hörer bzw. Schreiber/Leser nicht wird beantwortet werden können.

Hier liegt denn auch die methodologische Begründung für eine Abwendung von rein referenztheoretischen Ansätzen im Rahmen lexikalisch-semantischer Analysen und für die Hinwendung zu einer strukturalen Modellbildung. Denn letztere unternimmt es, die in sprachlichen Äußerungen tatsächlicher Sprecher/Hörer in konkreten Kommunikationssituationen geleisteten (durchaus nicht immer einheitlichen) Strukturierungen von Außersprachlichem (universe of discourse) als Resultat und Folge der regelgeleiteten Verwendung des Zeicheninventars zu analysieren und zu beschreiben; erstere aber muß neben dem Zeicheninventar und den Regularitäten seines Gebrauchs das ‚universe of discourse‘ schon als strukturierte Menge nicht-sprachlicher Gegebenheiten (Realität) voraussetzen.

Die Transparenz dieses denotativen, dem Aufbau formaler Sprachen nachgebildeten Analyseansatzes wird dabei freilich durch den Mangel an (nicht-sprachlichen) Daten zunichte gemacht, da eine Strukturierung von außersprachlicher ‚Realität‘ in Objekte und/oder Prozesse anders denn als semiotische Leistung der Bedeutungskonstitution durch natürliche Sprache nicht belegt werden kann. Eben diese Leistung bildet das Fundament des hier zu verfolgenden Ansatzes, allerdings nicht so, als wäre der semiotische Prozeß der Bedeutungskonstituierung selbst simulativ repräsentierbar (RIEGER 1977b), aber doch insofern, als die in ihm wirksamen Prinzipien zur Grundlage der empirischen Analyse und formalen Beschreibung lexikalisch-semantischer Gegebenheiten gemacht werden.

dabei kann bei gegebenen Elementen von der Ausbildung von Regularitäten (des Gebrauchs) über Unterschiedlichkeiten (der Verwendungsweisen) und Redundanzen (des Erwartbaren), zur Konstitution (von Zusammenhangsstrukturen) und Neudefinition (von Einheiten) fortgeschritten werden mit dem Ziel, die ‚Bedeutung‘ eines Lexikoneintrags formal und empirisch als Funktion aller seiner Unterschiede in allen seinen Verwendungsregularitäten zu sämtlichen anderen Einheiten des analysierten Vokabulars innerhalb pragmatisch-homogener Textmengen (s. unten 2.3) zu erklären.

2 Modellbildung einer unscharfen Lexikonstruktur

Die Theorie der unscharfen Mengen kann einer auf ihr aufbauenden lexikalisch-semantischen Bedeutungsnotation einen übergreifenden — von einem jeweils gewählten referentiellen, strukturalen oder auch handlungstheoretischen Semantikmodell unabhängigen — Formalismus bieten. Ihm lassen sich aber nur im Falle strukturaler Bedeutungsmodelle empirisch-quantitative Verfahren der Bedeutungsanalyse zuordnen, die zusammen mit diesem Formalismus das Modell einer unscharfen Lexikonstruktur bilden.

Mit der im folgenden vorgenommenen, in diesem Sinne modifizierten Übertragung des referentiellen Ansatzes von ZADEH auf ein strukturalen Semantikmodell wird deshalb nicht nur eine zunächst denotative Bedeutungsauffassung durch eine konnota-

tive abgelöst, und es findet auch kein bloßer Aspektwechsel statt von einer Sprach-externen Sicht zu einer Sprach-internen Auffassung des Bedeutungsbegriffs mit Wechsel von möglicher extensionaler zu intensionaler Unschärfe. Durch die Hinwendung auf das System semantischer Regularitäten, welche die Wörter eines Vokabulars untereinander ausbilden, wenn sie in Texten als sprachlich-kommunikativen Handlungen benutzt werden, wird vielmehr Sprache selbst in ihrer performativen Leistung der Strukturierung von ‚Realität‘ zur beobachtbaren Gegebenheit erklärt. Ihre bedeutungskonstituierenden Struktureigenschaften — in der Verwendung bei Einzelnen (Idiolekte) oder bei Gruppen und Schichten (Soziolekte), innerhalb eines Zeitabschnitts (synchron) oder über mehrere Zeitabschnitte hinweg (diachron) — werden dabei zu im Prinzip auch empirisch zugänglichen Daten, aus denen trotz der Vagheit und Varietäten gerade im lexiko-semantischen Bereich die lexikalische Struktur eines Vokabulars gleichwohl ermittelt und exakt beschrieben werden kann.

Nun ist Sprache aus semiotischer Sicht keine bloße Menge von Wörtern, die eine Menge nicht-sprachlicher Elemente mittels besonderer Funktionen zugeordnet wäre, sondern Sprache wird begriffen als eine Menge unterscheidbarer sprachlicher Gegebenheiten erst aufgrund und im Rahmen ihres Strukturzusammenhangs, der — auf welcher der betrachteten (phonetischen, morphologischen, syntaktischen, semantischen, pragmatischen) Ebene auch immer — einzelne Elemente bzw. Kategorien von Elementen deswegen zu unterscheiden und zu isolieren erlaubt, weil sie diesen Strukturzusammenhang in jeweils unterschiedlicher Weise abbilden.

Die Rekonstruktion dieses Systems auf lexikalisch-semantischer Ebene muß dabei nicht auf die phonomorphe Konstitution der Wörter rekurrieren, sondern kann von der in natürlich-sprachlichen Äußerungen vorgegebenen Strukturierung in Worteinheiten ausgehen. Deren Verwendung nicht nur in einzelnen Sätzen, sondern in ganzen Texten, welche die Bedingung pragmatischer Homogenität erfüllen sollten, liefert dabei die Basisinformationen des kommunikativen Vokabulargebrauchs zur regelgeleiteten Konstitution von Bedeutungen. Diese können in ihrem systematischen Zusammenhang mit Hilfe unscharfer Relationen zunächst formal entwickelt (2.1), durch Zuordnung numerisch-operationaler Verfahren (2.2) empirisch zugänglich gemacht und sodann aufgrund der mathematisch-statistischen Datenanalyse (2.3) quantitativ beschrieben werden.

2.1 Der formale Aufbau des Modells

Dazu wird hier — anstelle von (8) — aus dem Gesamtwortschatz einer Sprache zunächst ein bestimmtes Vokabular V als Teilmenge aller Wörter der Sprache eingeführt

$$V := \{x_i\}, \quad i = 1, \dots, n \quad (14)$$

und — anstelle von (9) — ein lexikalisches System C (*Corpusraum*)

$$C := \{y_j\}, \quad j = 1, \dots, m \quad \text{wobei} \quad m \geq n \quad (15)$$

und ein semantisches System S (*Bedeutungsraum*) gefordert

$$S := \{y_k\}, \quad k = 1, \dots, l \quad \text{wobei} \quad l \geq m \quad (16)$$

Beide, C und S , werden dabei anhand des Gebrauchs zu definieren sein, den wirkliche Sprecher/Schreiber in tatsächlichen Texten zum Zweck der Kommunikation in bestimmten Anwendungskontexten von V machen.

Während man daher C informell als eine die Verwendungsregularitäten der Wörter aus V repräsentierende strukturierte Ansammlung von *Corpuspunkten* kennzeichnen kann, die es erlaubt, mit jeder in Elementen von V ausdrückbaren Bedeutung eine unscharfe Teilmenge von C zu identifizieren, läßt sich S etwa charakterisieren als ein durch die Unterschiede dieser unscharfen Mengen bestimmtes System von *Bedeutungspunkten*, die je unscharfe Mengen von Corpuspunkten repräsentieren und den Zusammenhang der lexikalisch-semantischen Struktur des Vokabulars in *semantischen Umgebungen* deutlich werden lassen. *Corpuspunkte*, *Bedeutungspunkte* und *semantische Umgebungen* bilden die im folgenden dargelegten Schritte einer konsekutiven Abbildung des Vokabulars, seines Gebrauchs und der sich daraus ergebenden Bedeutungsdifferenzierungen.

Die strukturellen Zusammenhänge zwischen V , C und S sollen zunächst formal über die unscharfen Relationen R , Q und L , die nicht unabhängig voneinander sind, erfaßt und dargestellt werden. Die unscharfe Relation R , die das Vokabular V auf sich selbst abbildet, wird — wie in (10) — über die Zugehörigkeitsfunktion

$$\mu_R : V \times V \rightarrow \mathbb{R} \quad (17)$$

definiert. Sie schreibt jedem geordneten Paar (x, x') einen (und nur einen) Zugehörigkeitswert $\mu_R(x, x')$ als reelle Zahl aus \mathbb{R} zu. R stellt damit eine Korrespondenz her zwischen Elementen aus V und unscharfen Teilmengen von V derart, daß

- 1.) der *Gebrauch* eines Wortes x aus V als unscharfe Teilmenge $G(x)$ in V abgebildet und — wie in (12) — über die Zugehörigkeitsfunktion

$$\mu_{G(x)}(x'_i) = \mu_R(x, x'_i) \quad \text{für alle } x'_i \in V \quad (18)$$

definiert und

- 2.) die *Verwendungsregularität* eines Wortes x' aus V über die unscharfe Teilmenge $D(x')$ in V beschrieben und — wie in (13) — über die Zugehörigkeitswerte der inversen unscharfen Relation R^{-1}

$$\mu_{D(x')}(x_i) = \mu_{R^{-1}}(x_i, x') \quad \text{für alle } x_i \in V \quad (19)$$

bestimmt wird.

Definiert man Gebrauch als Verwendungsregularitäten eines Wortes und damit R als symmetrische Relation, sind $G(x)$ und $D(x)$ identisch. Die unscharfe Relation Q wird — wie in (17) — über die Zugehörigkeitsfunktion

$$\mu_Q : V \times C \rightarrow \mathbb{R}^+ \quad (20)$$

definiert, die jedem geordneten Paar (x, y) einen (und nur einen) Wert $\mu_Q(x, y)$ als reelle, nicht-negative Zahl aus \mathbb{R} zuordnet und damit den Zugehörigkeitsgrad in Q angibt.

Q induziert eine Korrespondenz zwischen V und C derart, daß

- 3.) Elementen von V unscharfe Mengen in C zugeordnet werden, die für jedes $x \in V$ eine unscharfe Teilmenge $B(x)$ in C als die *strukturelle Bedeutung* des Wortes x aus V erklären. Deren Zugehörigkeitsfunktion ergibt sich dabei aus den Zugehörigkeitswerten $\mu_{B(x)}(y_j)$, die jedem Corpuspunkt y_j in der unscharfen Teilmenge $B(x) \subseteq C$ zukommen.

$$\mu_{B(x)}(y_j) = \mu_Q(x, y_j) \quad \text{für alle } y_j \in C \quad (21)$$

Diese Werte geben gleichsam an, in welchem Maße jeder Corpuspunkt im lexikalischen System zur Bedeutung eines bestimmten Wortes des Vokabulars beiträgt; und umgekehrt, daß

- 4.) Elementen von C unscharfe Mengen in V zugeordnet werden, die für jedes $y \in C$ eine unscharfe Teilmenge $D(y)$ in V als *Deskriptormenge* eines Corpuspunktes y erklären. Sie ergibt sich aus den Zugehörigkeitswerten $\mu_{D(y)}(x_i)$, die jedem der Wörter x_i in der unscharfen Teilmenge $D(y) \subseteq V$ über die inverse unscharfe Relation Q^{-1} zukommen

$$\mu_{D(y)}(x_i) = \mu_{Q^{-1}}(x_i, y) \quad \text{für alle } x_i \in V \quad (22)$$

Diese Werte geben also quasi den Grad an, mit dem jedes Wort x_i des Vokabulars zur Unterscheidung eines bestimmten Corpuspunktes im lexikalischen System beiträgt.

Die unscharfe Relation L wird — wie in (17) und (20) — über die Zugehörigkeitsfunktion

$$\mu_L : V \times S \rightarrow \mathbb{R}^+ \quad (23)$$

definiert, die jedem geordneten Paar (x, z) einen (und nur einen) Zugehörigkeitswert $\mu_L(x, z)$ als reelle, nicht-negative Zahl aus \mathbb{R}^+ zuordnet. L induziert damit die Korrespondenz zwischen V und S , wonach

- 5.) Elementen von V unscharfe Mengen in S zugeordnet werden, die für jedes $x \in V$ eine unscharfe Teilmenge $E(x)$ in S als *semantische Umgebung* des Wortes x aus V erklären. Deren Zugehörigkeitsfunktion ergibt sich aus den Zugehörigkeitswerten $\mu_{E(x)}(z_k)$, welche jeder Bedeutungspunkt z_k in der unscharfen Teilmenge $E(x) \subseteq S$ hat.

$$\mu_{E(x)}(z_k) = \mu_L(x, z_k) \quad \text{für alle } z_k \in S \quad (24)$$

Diese Werte geben demnach an, in welchem Maße jeder Bedeutungspunkt im semantischen System zur semantischen Umgebung eines Wortes des Vokabulars gehört; und umgekehrt wonach

- 6.) Elementen von S unscharfe Mengen in V zugeordnet werden, die für jedes $z \in S$ eine unscharfe Teilmenge $D(z)$ in V als *Deskriptormenge* eines Bedeutungspunktes z aus S erklären. Sie ergibt sich aus den Zugehörigkeitswerten $\mu_{D(z)}(x_i)$, welche aufgrund der inversen unscharfen Relation L^{-1} jedem der Wörter x_i in der unscharfen Teilmenge $D(z) \subseteq V$ zukommen

$$\mu_{D(z)}(x_i) = \mu_{L^{-1}}(x_i, z) \quad \text{für alle } x_i \in V \quad (25)$$

Diese Werte geben demnach den Grad an, mit dem jedes Wort x_i des Vokabulars zur Unterscheidung eines bestimmten Bedeutungspunktes $z \in S$ im semantischen System beiträgt.

Der formale Zusammenhang von R , Q und L kann nun von der Menge der Wörter V über die strukturierte Menge der Corpuspunkte C zu der strukturierten Menge der Bedeutungspunkte S als konsekutive Abbildung von Mengensystemen erklärt werden derart, daß die Restriktionen der Relation R auf jedes der x_i in der Menge V die Elemente $y_j \in C$, und daß die Restriktionen der Relation Q auf jedes y in der Menge C die Elemente $z_k \in S$ konstituiert, wenn $i \leq j \leq k$.

So definiert die auf x eingeschränkte unscharfe Relation R , die (18) den *Gebrauch* $G(x)$ eines Wortes x aus V als unscharfe Teilmenge in V abbildet, einen *Corpuspunkt* y im lexikalischen System C

$$\mu_R(x_j|x) = \{\mu_{G(x)}(x_j)\} := y \in C \quad (26)$$

Weiter definiert die auf x eingeschränkte unscharfe Relation Q , die (21) die *Bedeutung* $B(x)$ eines Wortes x aus V als unscharfe Teilmenge in C abbildet, einen *Bedeutungspunkt* z im semantischen System S

$$\mu_Q(y_j|x) = \{\mu_{B(x)}(y_j)\} := z \in S \quad (27)$$

Und letztlich definiert die auf x eingeschränkte unscharfe Relation L (24) die semantische Umgebung $E(x)$

$$\mu_L(z_k|x) = \{\mu_{E(x)}(z_k)\} := E(x) \subseteq S \quad (28)$$

Die *semantische Umgebung* $E(x)$ eines Wortes x als unscharfe Teilmenge von S setzt sich dabei aus der Menge aller Bedeutungen $\{B(x)\}_k$ zusammen

$$\mu_L\left(x, \left\{ \mu_Q\left(x, \left\{ \mu_R(x, x_i) \right\}_j \right\}_k \right) \right) := E(x) \quad (29)$$

die *Bedeutung* $B(x)$ eines Wortes x als unscharfe Teilmenge von C , konstituiert sich aus der Menge aller Gebräuche $\{G(x)\}_j$

$$\mu_Q\left(x, \left\{ \mu_R(x, x_i) \right\}_j \right) := B(x) \quad (30)$$

und der *Gebrauch* $G(x)$ eines Wortes x ist als unscharfe Teilmenge von V bestimmt

$$\mu_R(x, x_i) := G(x) \quad (31)$$

Für die empirische Rekonstruktion dieser formalen Konstrukte wird die Zuordnung von Meßvorschriften wichtig, welche geeignet sind, die unscharfen Relationen R , Q und L operational zu definieren. Erst diese Meßvorschriften erlauben es, bei Anwendung auf konkrete Textdaten die Zugehörigkeitsgrade einzelner Elemente zu bestimmten Teilmengen empirisch zu ermitteln.

2.2 Zugeordnete empirische Verfahren

Anders als innerhalb referenztheoretischer Modelle lassen sich in diesem strukturalen Semantikmodell die unscharfen Relationen R , Q und L über *Meßvorschriften* rekonstruieren. Diese Meßvorschriften operieren Sprach-intern auf Daten konkreter Textcorpora, nicht aber Sprach-extern auf Daten aus Probandentests. Sie liefern die numerischen Werte der jeweiligen Zugehörigkeitsgrade und erlauben damit den zunächst nur über die Zugehörigkeitsfunktion μ_R , μ_Q und μ_L formal dargestellten Zusammenhang nun auch mit empirischen Daten zu verbinden. Die Meßvorschriften der unten detailliert noch einzuführenden empirischen Koeffizienten α , δ und ζ fungieren zunächst wie Allrelationen in V , C und S . Ihre numerischen Werte können aber als quasi ‚verunschärfende‘ Funktionen (fuzzifications) auch scharfer Abbildungen \bar{R} , \bar{Q} , \bar{P} oder deren Kompositionen verstanden werden, welche die Zugehörigkeitsfunktionen der unscharfen Relationen R , Q und L bzw. deren Inversionen in der darzulegenden Weise determinieren.

Ist etwa \bar{R} eine scharfe Abbildung von V auf sich selbst

$$\bar{R}: V \rightarrow V \quad (32)$$

so kann das noch einzuführende Maß α als Fuzzifikation von R dann gedeutet werden, wenn hierdurch jedem geordneten Paar (x, x') als Element von \bar{R} zusätzlich genau ein reeller Zahlenwert als Zugehörigkeitsgrad $\alpha(x, x')$ aus dem Intervall I zugeordnet wird, derart, daß aus der (*scharfen*) Abbildung (32) nun die (*unscharfe*) Relation

$$\alpha: V \times V \rightarrow I, \quad I \subseteq \mathbb{R} \quad (33)$$

hervorgeht. Ist α symmetrisch und (17) empirisch zuordenbar, so ist die Einschränkung des Definitionsbereichs von α auf x

$$\mu_R(x, x'_i) = \mu_{R^{-1}}(x_i, x') := \alpha(x_i|x) \quad (34)$$

Über (34) kann (33) als (*scharfe*) Abbildung

$$\bar{Q}: V \rightarrow C \quad (35)$$

gedeutet werden, deren Inversion Q^{-1} definitionsgemäß (26) unscharf bleibt. Denn vermöge (34) wird zwar jedem Wort x aus V genau ein n -tupel von Zahlenwerten $\langle \alpha(x, x_1), \dots, \alpha(x, x_n) \rangle$ oder kürzer $\alpha_x(x_1, \dots, x_n)$ aus dem Intervall I zugeordnet, das genau einen Corpuspunkt y im Corpusraum C definiert, aber jedem $y \in C$ entspricht damit umgekehrt (22) eine durch dieses n -tupel in V definierte unscharfe *Deskriptormenge* $D(y)$

$$\mu_{D(y)}(x_i) = \mu_{Q^{-1}}(x_i, y) := \alpha(x_i|x) \quad (36)$$

Ein noch anzugebendes Distanzmaß δ kann als symmetrische Fuzzifikation einer Abbildung des so (35) konstituierten C auf sich selbst eingeführt werden

$$\delta: C \times C \rightarrow J, \quad J \subset \mathbb{R}^+ \quad (37)$$

Dieses δ schreibt jedem geordneten Paar (y, y') genau einen reellen, nicht-negativen Zahlenwert als Zugehörigkeitsgrad $\delta(y, y')$ aus dem Intervall J zu.

Als ein auf dem Kreuzprodukt von C definiertes symmetrisches Maß, leistet δ eine zweifache Aufgabe.

Wegen (35) kann es unmittelbar als Fuzzifikation von Q gedeutet werden, was zur *unscharfen* Relation

$$\delta_Q: V \times C \rightarrow J, \quad J \subset \mathbb{R}^+ \quad (38)$$

führt. δ stellt ein μ_Q in (20) zuordenbares Maß dar, dessen auf y resp. x eingeschränkter Definitionsbereich nach (21) die strukturelle *Bedeutung* $B(x)$ bestimmt.

$$\mu_{B(x)}(y_j) = \mu_Q(x, y_j) := \delta(y_j|y) \quad (39)$$

Damit ist zunächst die unscharfe Relation Q (20) über die beiden der Zugehörigkeitsfunktion μ_Q in (39) und ihrer Inversion $\mu_{Q^{-1}}$ in (36) zugeordneten Meßvorschriften als eine reflexive, nichtsymmetrische, unscharfe Abbildung des Vokabulars rekonstruiert.

$$\mu_Q(x, y_j) = \delta(y_j|y) \quad (40)$$

für alle $i = j$

$$\mu_{Q^{-1}}(x_i, y) = \alpha(x_i|x) \quad (41)$$

Weiterhin kann ein auf y eingeschränktes δ (39) aber auch als eine sich aus (37) ergebende (*scharfe*) Abbildung

$$\overline{P}: C \rightarrow S \quad (42)$$

gedeutet werden, die jedem Corpuspunkt y aus C ein n -tupel von Zahlenwerten $\langle \delta(y, y_1), \dots, \delta(y, y_m) \rangle$ bzw. $\delta_y(y_1, \dots, y_m)$ aus dem Intervall J zuordnet, das damit genau einen Bedeutungspunkt z im Bedeutungsraum S definiert.

Aufgrund von (35) und (42) läßt sich eine (*scharfe*) Abbildung von V in S als Komposition von Q und P schreiben

$$\overline{P \circ Q}: V \rightarrow S \quad (43)$$

die demnach aus der Menge der geordneten Paare $\{(x_i, z_k)\}$ besteht, für die es je ein $y \in C$ gibt, für das gilt $(x_i, y) \in \overline{Q}$ und $(y, z_k) \in \overline{P}$, was trivialerweise für $i = k$ erfüllt ist.

Die Inversion $(P \circ Q)^{-1}$ bleibt definitionsgemäß unscharf, denn nach (27) entspricht jedem $z \in S$, eine durch die Restriktion $\mu_Q(x, y_j) = \delta(y_j|y)$ in C definierte unscharfe Menge, die nach (25) mit der Deskriptormenge $D(z) \subseteq V$, identisch ist

$$\mu_{D(z)}(x_i) = \mu_{L^{-1}}(x_i, z) := \delta(y_j|y) \quad (44)$$

Nachdem so durch die Restriktion von δ nicht nur die Zugehörigkeitsfunktion μ_Q (39) der unscharfen Relation Q (20), sondern auch die Zugehörigkeitsfunktion $\mu_{L^{-1}}$ (44) der Inversion der unscharfen Relation L (23) bestimmt wurde, macht die Rekonstruktion der Zugehörigkeitsfunktion μ_L selbst die Einführung eines δ analogen Distanzmaßes ζ erforderlich.

Als ein auf dem Kreuzprodukt von S definiertes symmetrisches Maß leistet nun ζ eine zweifache Aufgabe. So kann es zunächst entsprechend (33) und (37) — als Fuzzifikation einer Abbildung von S auf sich selbst gedeutet werden

$$\zeta: S \times S \rightarrow I, \quad I \subset \mathbb{R}^+ \quad (45)$$

die jedem geordneten Paar (z, z') genau einen reellen, nicht negativen Zahlenwert als Zugehörigkeitsgrad $\zeta(z, z')$ aus dem Intervall I zuschreibt.

Weiterhin kann ζ als Fuzzifikation der (*scharfen*) Komposition $P \circ Q$ dienen, wodurch aus (43) sich dann

$$\zeta_{P \circ Q}: V \times S \rightarrow I, \quad I \subset \mathbb{R}^+ \quad (46)$$

ergibt. $\zeta_{P \circ Q}$ dient so als ein der (*unscharfen*) Relation (23) zuordenbares empirisches Maß, dessen Definitionsbereich auf z eingeschränkt, eine (24) entsprechende Vorschrift ergibt

$$\mu_{E(x)}(z_k) = \mu_L(x, z_k) := \zeta(z_k|z) \quad (47)$$

Damit ist auch die unscharfe Relation L (23) über die beiden Zugehörigkeitsfunktionen μ_L in (47) und ihrer Inversion $\mu_{L^{-1}}$ in (44) zugeordneten Meßvorschriften als eine reflexive, nicht-symmetrische, unscharfe Abbildung des Vokabulars rekonstruiert.

$$\mu_L(x, z_k) = \zeta(z_k|z) \quad (48)$$

für alle $i = k$

$$\mu_{L^{-1}}(x_i, z) = \delta(y_j|y) \quad (49)$$

So ist abschließend die Repräsentation des semantischen Aspektes eines sprachlichen Ausdrucks x über die unscharfen Mengen $G(x)$ seines *Gebrauchs* (18), $B(x)$ seiner *Bedeutung* (21) und $E(x)$ seiner semantischen *Umgebung* (24) als die *formale Strukturbeschreibung* eines Systems konsekutiver Abbildung des Vokabulars geleistet, denen nun über die Anwendung der Koeffizienten α , δ und ζ bei der Analyse des in einem bestimmten Textcorpus angetroffenen Vokabulars *empirische Daten* zugeordnet werden, um so unterschiedliche sprachliche Gegebenheiten als spezifische Ausprägungen derselben abstrakten Repräsentationstruktur darstellen zu können.

2.3 Quantitative Analyse der Daten

Auf den Zusammenhang von *Anwendungskontext*, *kommunikativem Gebrauch* und *Bedeutung* sprachlicher Ausdrücke ist wiederholt hingewiesen worden (RIEGER 1977c). Eine über mathematisch-statistische Verfahren empirisch fundierte und über Notationen der Theorie der unscharfen Mengen formal befriedigende Analyse der lexikalisch-semantischen Bedeutung natürlich-sprachlicher Ausdrücke muß diesen Zusammenhang berücksichtigen. Bedeutung wird daher nicht außerhalb der *Pramatik* analysiert werden können, in der sie sich konstituiert, d.h. im Zusammenhang von *Kommunikation*, und Bedeutung wird anhand von Gegebenheiten analysiert werden müssen, die diesen Zusammenhang empirisch zugänglich repräsentieren, d.h. anhand von ganzen sprachlichen Texten (RIEGER 1979c).

Als *Text* gilt dabei jede Folge von sprachlichen Zeichen, die im Anwendungskontext einer konkreten Situation von tatsächlichen Sprechern/Hörern zum Zweck der Kommunikation geäußert/erkannt werden kann. Mit *Pragmatik* wird dabei der potentiell sowohl durch Ort und Zeit wie Gegenstand, Medium und Beteiligte bestimmbare situative Kommunikationsrahmen bezeichnet, in dem sich Sprachverwendungen ereignen. Unter *Kommunikation* wird der Prozeß zunehmender Einschränkung von Wahlmöglichkeit verstanden, den die daran Beteiligten über Zeichen und Zeichenfolgen (Texte) wechselseitig initiieren und nachvollziehen. Damit braucht als Motivation solcher über Zeichen und Zeichenfolgen initiierten Kommunikation primär nicht mehr ein ‚richtiges‘ oder ‚falsches‘ Verstehen angesetzt zu werden, sondern kommunikative Akte können allgemein alle jene Handlungen heißen, die dem Abbau von Unsicherheit dienen, den die miteinander Kommunizierenden je nach pragmatischen Erfordernissen in ‚größerem‘ oder ‚geringerem‘ Grade anstreben und/oder erzielen. Bedeutung läßt sich aus dieser Sicht folglich nicht als eine *statische* Qualität beschreiben, die Zeichen und Zeichenfolgen nun einmal haben. Vielmehr muß die Analyse und Repräsentation der Bedeutung eines Zeichens, Wortes etc. als eine Art ‚Momentaufnahme‘ verstanden werden, die den im Prinzip andauernden *dynamischen* Prozeß der Bedeutungskonstitution innerhalb des pragmatischen Anwendungskontexts eines ganzen Vokabulars quasi unter dem Blickwinkel dieses betreffenden Zeichens, Wortes etc. über einer Pragmatik abbildet.

Diesen kommunikations- und handlungstheoretisch beschreibbaren Zusammenhang kann eine empirische Analyse der Vagheit natürlichsprachlicher Wortbedeutungen durchaus auch bei Verwendung mathematisch-statistischer Verfahren berücksichtigen. Wie schon in anderem Zusammenhang (RIEGER 1972) näher ausgeführt, läßt sich dabei Gegenstand und Ziel einer Untersuchung mit der weitgehend operationalisierten Begriffsbildung von Stichprobe und Grundgesamtheit in Verbindung bringen, wie dies innerhalb der urteilenden Statistik und ihrer Methodik geschieht.

Jede textstatische Untersuchung kann nun im Prinzip zwar davon ausgehen, daß sich in Texten — im Unterschied zu bloßen Wörter bzw. Zeichenansammlungen — *Ordnungs-*

relationen und *regelmäßige Beziehungen* zwischen den verwendeten Zeichen, Wörtern etc. aufdecken, beschreiben und messen lassen, wie dies etwa in der quantitativen Linguistik schon von HARRIS (1968) und im Rahmen des Document and Information Retrieval etwa von SALTON et al. (1975/1976) und SPARCK JONES/KAY (1973) seit langem praktiziert wird. Wenn aber die so offenbar empirisch ermittelbaren Beziehungen für eine linguistische Theorienbildung in der Semantik, Lexikologie bzw. Lexikographie relevant werden und nicht eine bloße Ansammlung numerischer Fakten und uninterpretierbarer Daten bleiben sollen, muß der jeweilige sprachliche Untersuchungsgegenstand zusätzlichen Forderungen genügen, damit eine statistische Analyse überhaupt sinnvoll genannt werden kann: er muß sich im Sinne statistischer Methodik als zufällige Stichprobe aus einer Grundgesamtheit deuten lassen, über die bestimmte wissenschaftliche Aussagen gemacht werden sollen (RIEGER 1979d). Danach heißt eine *Stichprobe zufällig* dann, wenn die Operation der Auswahl eines *Untersuchungsgegenstandes* (hier etwa der Texte eines Corpus) bei — im Prinzip beliebig häufiger — Wiederholung der Auswahlprozedur auf die *Grundgesamtheit* hin konvergiert. Die *Grundgesamtheit*, die im linguistischen Bereich im allgemeinen immer *fiktiv* sein wird, läßt sich dabei nur vom *Untersuchungsziel* her bestimmen. Denn in statistischer Hinsicht ist ein *Untersuchungsziel* identisch mit dem Vorhaben, bei nur unvollständiger Information intersubjektiv nachprüfbar, optimale Aussagen über eine selbst nicht zugängliche *Grundgesamtheit* zu machen aufgrund der Analyse einzig von daraus entnommenen *zufälligen Stichproben*, die den *Untersuchungsgegenstand* bilden.

Ist — wie hier — das Untersuchungsziel die Analyse und Repräsentation natürlich-sprachlicher Wortbedeutung im Zusammenhang der sie fundierenden Pragmatik (öffentlicher Sprachgebrauch/Presse), dann können nur solche natürlich-sprachlichen Texte den Untersuchungsgegenstand bilden, die von tatsächlichen Sprechern/Schreibern (Journalisten) in einer konkreten (WE bzw. ND), dabei gleichartigen Kommunikationssituation (Zeitungs-Information/-Kommentar) geäußert worden sind. Nur ein solches Textcorpus, das *pragmatisch-homogen* heiße, kann als eine zufällige Stichprobe aller derjenigen sprachlichen Äußerungen gelten, die von diesen Sprechern/Schreibern in dieser Pragmatik tatsächlich gemacht wurden oder doch hätten gemacht werden können, weshalb sie eine fiktive Grundgesamtheit bilden.

Bei der statistischen Analyse kann unter diesen Bedingungen der pragmatischen Homogenität des Textcorpus von einer empirischen Meßvorschrift Gebrauch gemacht werden, die erlaubt, etwa vorhandene Regularitäten der Abhängigkeit zwischen Wörtern aufgrund ihrer Ko-okkurrenzen in den Texten festzustellen und deren unterschiedliche Intensitäten zu messen. So kann die ganze Skala von wechselseitiger Abstoßung über Beziehungslosigkeit bis zur wechselseitigen Anziehung zwischen Lemmata — graduell beliebig genau in numerischen Werten des Intervalls von -1 bis $+1$ ausgedrückt werden. Dies leistet der Korrelationskoeffizient, der — wie (33) vorsieht — die Beziehung jedes Wortes aus dem Vokabular zu jedem anderen verwendeten Wort mißt aufgrund des Gebrauchs, den die Sprecher/Schreiber von ihnen in den analysierten Texten des Corpus machen.

Werde etwa das in (14) schon formal eingeführte Vokabular V der lemmatisierten Worteinträge x_i

$$V := \{x_i\}, \quad i = 1, \dots, n$$

untersucht anhand der Verwendung, die von den unter jedem Lemmaeintrag zusammengefaßten Wörtern in Texten gemacht wird. Bestehe das untersuchte Corpus K aus

Texten t

$$K := \{t\}, \quad t = 1, \dots, T \quad (50)$$

in dem

$$U = \sum_{t=1}^T u_t, \quad 1_t \leq u_t \leq U \quad (51)$$

den Corpusumfang U als Summe aller Textlängen u_t , gemessen in der Anzahl der im t -ten Text verwendeten Wörter (*tokens*) bezeichne, und

$$H_i = \sum_{t=1}^T h_{it}, \quad 1_t \leq h_{it} \leq H_i \quad (52)$$

die Gesamthäufigkeit H_i des i -ten Wortes (type) x als Summe seiner Häufigkeiten h_{it} in allen Texten t angebe, dann lautet der modifizierte *Korrelationskoeffizient* $\alpha(x, x')$ für die Lemmawörter x und x'

$$\alpha(x, x') = \frac{\sum_{t=1}^T (h_{it} - h_{it}^*)(h'_{it} - h'_{it}{}^*)}{\left(\sum_{t=1}^T (h_{it} - h_{it}^*)^2 \sum_{t=1}^T (h'_{it} - h'_{it}{}^*)^2 \right)^{\frac{1}{2}}}; \quad -1 \leq \alpha(x, x') \leq +1 \quad (53)$$

wobei

$$h_{it}^* = \frac{H_i}{U} u_t \quad \text{und} \quad h'_{it}{}^* = \frac{H'_i}{U} u_t.$$

Die Korrelationswerte eines Wortes x zu sämtlichen anderen Wörtern des Vokabulars, die nach (18) und (33) den *Gebrauch* von x als unscharfe Menge $G(x)$ abbilden, liefern so — wie in (35) vorgesehen — die Daten zur Definition eines *Corpuspunktes* y im Corpusraum C . Das hat zur Folge, daß zwei Corpuspunkte y und y' in C umso näher benachbart sind, je weniger unterschiedlich die beiden sie definierenden unscharfen Mengen $G(x)$ und $G(x')$ und damit der Gebrauch der Wörter x und x' sind. Ihre Unterschiedlichkeit läßt sich mit Hilfe verschiedener Maße bestimmen. Hier soll sie als Entfernung zwischen den $G(x)$ und $G(x')$ entsprechenden Corpuspunkten y und y' in C ausgedrückt und etwa — wie in (36) vorgesehen — über das euklidische *Distanzmaß* δ empirisch bestimmt werden

$$\delta(y, y') = \left(\sum_{i=1}^n (\alpha(x, x_i) - \alpha(x', x_i))^2 \right)^{\frac{1}{2}}, \quad 0 \leq \delta(y, y') \leq 2\sqrt{n}. \quad (54)$$

Die Distanzwerte $\delta(y, y_j)$ eines Corpuspunktes y zu sämtlichen anderen Corpuspunkten y_j des Corpusraums, die nach (21) und (38) die *Bedeutung* von x als unscharfe Menge $B(x)$ abbilden, liefern so — wie in (39) vorgesehen — die Daten zur Definition eines *Bedeutungspunktes* z im *Bedeutungsraum* S . Das hat zur Folge, daß zwei Bedeutungspunkte z und z' in S umso näher benachbart sind, je weniger unterschiedlich die beiden sie definierenden unscharfen Mengen $B(x)$ und $B(x')$ und damit die Bedeutungen von x und x' sind. Ihre Unterschiedlichkeit läßt sich daher als der Abstand zwischen den Bedeutungspunkten z und z' in S ausdrücken und etwa — wie in (43) vorgesehen — über das euklidische *Distanzmaß* ζ empirisch bestimmen

$$\zeta(z, z') = \left(\sum_{j=1}^n (\delta(y, y_j) - \delta(y', y_j))^2 \right)^{\frac{1}{2}}, \quad 0 \leq \zeta(z, z') \leq 2n. \quad (55)$$

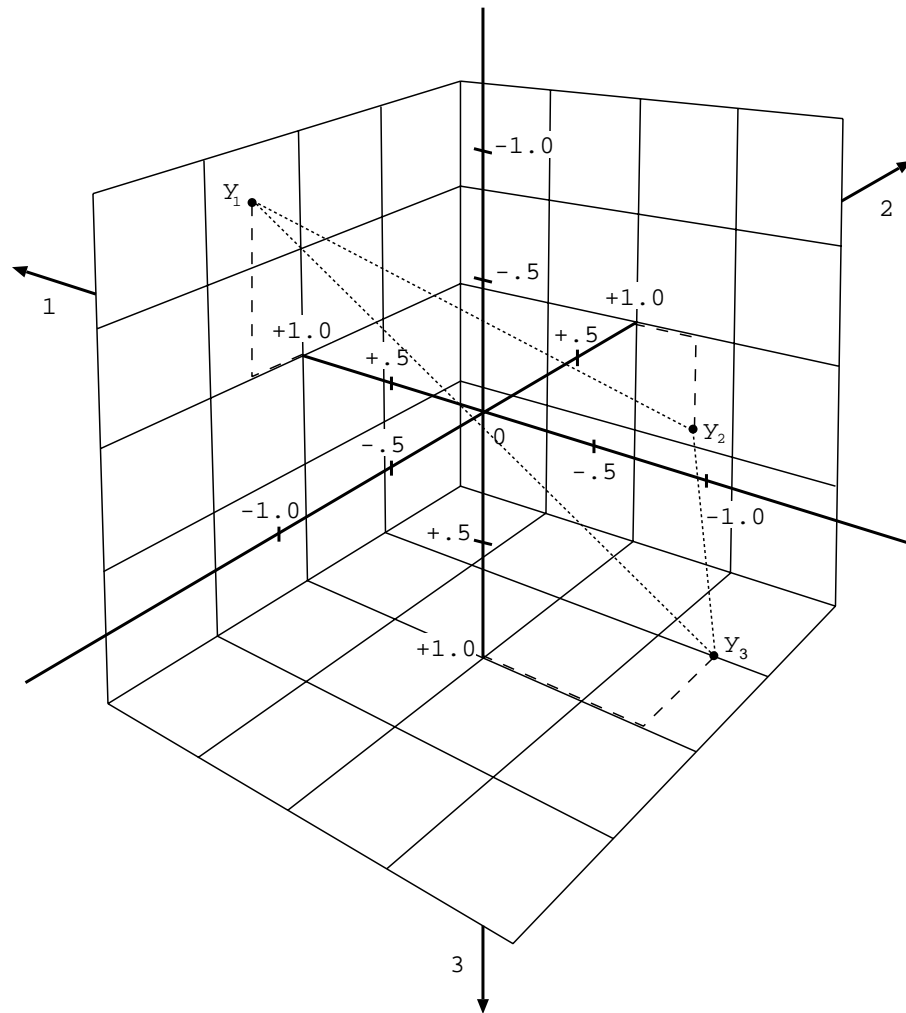


Fig. 2

Die Distanzwerte $\zeta(z, z_i)$ eines Bedeutungspunktes z zu sämtlichen anderen Bedeutungspunkten im Bedeutungsraum S liefern so — wie in (24) und (47) vorgesehen — die numerischen Daten zur Definition der *semantischen Umgebung* eines Wortes x als unscharfer Menge $E(x)$ im Vokabular.

Zur Illustration des Analysegangs und zur Veranschaulichung der konsekutiven Abbildungen des Vokabulars aufgrund seines Gebrauchs in den analysierten Texten des Korpus möge eine vereinfachende, auf drei Dimensionen beschränkte räumliche Darstellung dienen. Dazu wird von einem Textkorpus K ausgegangen, das alle genannten Bedingungen erfüllt, aber aus Texten t einer fiktiven Sprache bestehe, die ein Vokabular V von nur drei Wörtern zur Verfügung hat. Der Gesamtumfang U (51) der in jedem der Texte t des Korpus K (50) verwendeten Wörter (tokens) verteilt sich also (14) auf $n = 3$ Lemmawörter (types), nämlich x_1 , x_2 und x_3 , die das Vokabular der zu analysierenden Lexikonstruktur ausmachen.

Der Korrelationskoeffizient $\alpha(x, x')$ mißt nun paarweise (53) die Beziehung jeden Wortes x_1 , x_2 und x_3 zu allen Wörtern des Vokabulars, und zwar aufgrund ihrer Verwendung in den Texten des Korpus. Das ergibt für jedes Wort folglich $n = 3$ Meßwerte,

für x_1 beispielsweise die Werte α_{11} , α_{12} und α_{13} . Diese Meßwerte werden nun als Koordinaten interpretiert, die — nach (35) — dem Wort $x_1 \in V$ einen Punkt y_1 in einer n -dimensionalen Raumstruktur C zuordnen (Fig. 2), welche durch die den $n = 3$ Worttypen x_1 , x_2 , x_3 entsprechenden Achsen 1, 2 und 3 aufgespannt wird. Nach (34) ist damit sowohl (18) der *Gebrauch* $G(x)$ bzw. (19) die *Verwendungsregularität* $D(x)$ eines Wortes x als unscharfe Teilmenge von V erklärt, die (26) einen Punkt y in C definiert. Nach (15) heißt C *Korpusraum*, jedes seiner Elemente $y_j \in C$ *Korpuspunkt*.

Wie aus Fig. 2 ersichtlich ist, wird die Lage jeden Punktes $y \in C$ in diesem Raum durch das Tripel seiner Korrelationswerte bestimmt. Zwei Korpuspunkte $y, y' \in C$ werden folglich im Korpusraum umso enger benachbart sein, je geringfügiger die Gebräuche bzw. Verwendungsregularitäten der von ihnen repräsentierten Wörter $x, x' \in V$ voneinander abweichen. Als Maß solcher Unterschiedlichkeit in Gebrauch oder Verwendung bietet sich die Entfernung zwischen zwei Korpuspunkten in C an. Diese wird durch die Metrik des Korpusraumes bestimmt und kann beispielsweise als euklidische Distanz gemessen werden (gepunktete Linie in Fig. 2).

Diese δ -Werte stellen nun eine abgeleitete, neue Charakteristik dar, die die paarweise Beziehung mißt (54), die zwischen jedem Korpuspunkt y_1, y_2, y_3 und allen anderen Korpuspunkten besteht aufgrund der Unterschiede zwischen allen durch sie repräsentierten Verwendungsweisen. Das ergibt für jeden Korpuspunkt wiederum $n = 3$ Meßwerte, für y_1 beispielsweise die Werte $\delta_{11}, \delta_{12}, \delta_{13}$. Diese Werte werden wiederum als Koordinaten interpretiert, die — nach (42) — dem Korpuspunkt $y_1 \in C$ einen Punkt z_1 in einer n -dimensionalen Raumstruktur S zuordnen. Nach (38) ist damit durch (21) die *strukturelle Bedeutung* $B(x)$ eines Wortes x als unscharfe Menge von C erklärt, die so (27) den Punkt z in S definiert. Nach (16) heißt S *Bedeutungsraum*, jedes seiner Elemente $z_k \in S$ *Bedeutungspunkt*.

Im Bedeutungsraum wird die Lage jeden Bedeutungspunktes $z \in S$ somit durch die drei Koordinaten bestimmt, die *alle* Unterschiedlichkeiten (δ - oder Distanz-Werte) erfassen, die sich aufgrund *aller* Verwendungsregularitäten des Gebrauchs (α - oder Korrelationswerte) eines Wortes $x \in V$ in den analysierten Texten ergeben.

Um sich eine Vorstellung von der Lage eines Bedeutungspunktes zu machen, genügt es, diejenigen Bedeutungspunkte im Bedeutungsraum festzustellen, die ihm unmittelbar benachbart sind. Sie lassen sich über das Distanzmaß ζ ermitteln, welches die Distanz je zweier Bedeutungspunkte voneinander (55) bestimmt. Die Konfiguration der Bedeutung eines Wortes durch Angabe derjenigen Bedeutungspunkte z_k , die innerhalb eines solchen semantischen Systems S in unmittelbarer Nachbarschaft liegen, hat — wie auch an anderer Stelle (RIEGER 1974, 1975) ausführlich dargelegt — ihre sprachwissenschaftliche Entsprechung in der Konzeption des semantischen Feldes. Es läßt sich innerhalb dieses Modells einer Lexikonstruktur topologisch als Umgebung explizieren. Diese Umgebung beschreiben quasi die Punkthäufungen im Bedeutungsraum, die sich innerhalb einer gewissen Distanz um einen Bedeutungspunkt $z \in S$ scharen. Nach (28) und (29) heißt die topologische Nachbarschaft *semantische Umgebung* $E(x)$ eines Wortes $x \in V$. Sie kann als eine neue, umfassende Repräsentation von Wortbedeutung gelten, deren Systemzusammenhang sie nach (26) bis (31) darstellt aufgrund einer Notation, die Bedeutung von Wörtern als Funktion aller Unterschiede aller Verwendungsregularitäten des Vokabulars in Texten abbildet.

3 Darstellung der Ergebnisse

Wegen der begrenzten Umfänge der aus dem Textmaterial der Zeitungen WE und ND entnommenen Stichproben sind die im folgenden vorgelegten Ergebnisse der Analysen beider Datensätze mit gewissen einschränkenden Vorbehalten zu bewerten. Sie betreffen vornehmlich den — durch den Charakter einer Pilotuntersuchung vorgegebenen — geringen Umfang der aus den jeweiligen Zeitungsnummern des Jahrgangs 1964 entnommenen Textmengen, sowie die Auswahl und den Umfang der hierdurch für beide Datensätze zu analysierenden Vokabulare.

Nach der Ermittlung der als Deskriptormenge der Vokabulare fungierenden Wortlisten (3.1) werden die Analyseergebnisse anhand ausgewählter Beispiele dargestellt und zwar im Hinblick einmal auf mögliche unterscheidende Aspekte (3.2), zum anderen auf mögliche gemeinsame Aspekte (3.3) des jeweiligen Wortgebrauchs im lexikalisch-semanticen Bereich. Die Beispiele sind dem Daten-Supplement entnommen, in dem die vollständigen Daten enthalten sind, die aus den Analysen der Textkorpora gemäß der oben entwickelten Modellbildung einer unscharfen Lexikonstruktur für das WE- und das ND-Material ermittelt wurden.

3.1 Vokabular-Deskriptoren

Aus den Basisdaten des aus 175 Artikel bestehenden WE- und des aus 57 Artikel bestehenden ND-Textmaterials ergeben sich drei Vokabularlisten des verwendeten Wortschatzes, deren Elemente (lemmatisierte Wort-Typen) mit Häufigkeiten ≥ 5 bzw. ≥ 3 in den Textkorpora belegt sind. Diese Wortlisten bilden die für die Bedeutungsbeschreibung verwendeten Deskriptormengen D_1 (analysiertes WE-Vokabular mit 365 Types/7457 Tokens), D_2 (analysiertes ND-Vokabular mit 242 Types/1774 Tokens) und $D_3 = D_1 \cap D_2$ (Schnittmenge der WE und ND gemeinsamen 129 Types), die im Sinne des entwickelten Modells zur Darstellung des lexikalisch-semanticen Zusammenhangs der in den Textkorpora verwendeten Wortbedeutungen, ihrer Unterschiede und Gemeinsamkeiten dienen.

Anmerkung zu Bearbeitung der Materialgrundlage:

Bereitgestellt wurden jeweils eine Auswahl aus den Jahrgangsquerschnitten ND 64 und WELT 64 des Bonner Zeitungskorpus, und zwar 7 Seiten aus ND 64 (jeweils Seite 1) und 15 Seiten aus WELT 64 (jeweils Seite 1 und 2). Die Texte aus diesen beiden Teilmengen wurden lemmatisiert auf folgende Weise: Berücksichtigt wurden nur die Wortformen von Substantiven, Adjektiven und Verben. Sie wurden manuell fortlaufend auf eine Art Stamm-Morphem zurückgeführt, wobei durch die Endungen (nach /) die belegten Wortarten erkennbar blieben. Komposita wurden aufgelöst, wenn ihre Glieder selbständige Wörter waren. Personennamen wurden übergangen. Die Lemmata wurden mit Nummern verschlüsselt, die die Grundlage der statistischen Berechnungen bildeten.

Diese Arbeiten wurden in der F.ö.S. ausgeführt.

D_1 : WE-Vokabular/Deskriptoren

ABEND	ACHTEN/UNG	AKTION
ABFAHRT/EN	AESTHETIK/ISCH	ALLE
ABKOMMEN	AFRIKA/NER/ISCH	ALLGEMEIN

ALPEN
ALT/ER
AMERIKA/ER/ISCH
AMT/IEREN
ANBIETEN/GEBOT
ANTWORT/EN
APPELL/IEREN
ARBEIT/EN
ARTZ/LICH
ATLANTIK/SCH
ATOM/AR
AUFNAHME/EN
AUFTRAG/EN
AUSGABE/GEBEN
AUSLAND/ER/ISCH
AUSSCHLUSS/EN
AUSSCHUSS
AUSSEN/LICH/KT
AUTO
BAD/N
BAHN
BAU/EN
BE/GRUND/UNG
BE/RATEN/UNG
BEACHTEN/UNG
BEGINN/EN
BEHOERDE
BEITRAG/EN
BEKANNT
BEMUEHEN/UNG
BEOBACHTEN/UNG
BEREIT/EN
BERG
BERICHT/EN
BERLIN/ER
BERUF/LICH
BESCHLUSS/EN
BESTIMMEN/UNG
BESUCH/EN
BEWEGEN/UNG
BILD
BILDEN/UNG
BINDEN/BAND/NG
BITTE/N
BLEIBEN
BLUT/EN
BRAUCHEN
BRINGEN
BRITE/ANNIEN

BUCH
BUERO
BUND
BUNDESTAG
CHEF
CHINA/SISCH
COMPUTER
DANK/EN
DAUER/N
DELEGATION/ER
DEMONSTRATION/N
DENKEN/GEDANKE
DEUTSCH/LAND
DIENST/EN
DIPLOM
DIPLOMAT/IE/SCH
DOPPEL
DORF
EHRE/N
EINSATZ/EN
ELEKTRO/NISCH
EMPFANG/EN
ENDE/N/LICH
ENG/E/N
ENTSCHEIDEN/UNG
ENTWICKELN/UNG
ERFAHREN/UNG
ERFOLG/REICH
ERGEBNIS/EN
ERKENNTNIS/EN
ERKLAEREN/UNG
ERNTE/N
ERSCHEINEN/UNG
ERWARTEN/UNG
EUROPA/ISCH
FAEHIG/KEIT
FAHREN/ER
FALL/EN
FDP
FEIER/N
FERNSEHEN
FEST/IGEN/UNG
FILM/EN
FLACH/E
FLIEGEN/FLUG
FLUCHT/EN/LING
FLUGZEUG
FOLGE/N/UNG
FORDERN/UNG

FORM/N/AT/ICH
FRAGE/N
FRAKTION
FRANKREICH/ISCH
FREI/HEIT
FREMDE/E/R
FREUDE/EN
FREUND/SCHAFT
FRIEDE/LICH
FRUEH
FUEHREN/ER/UNG
FUND/EN
GABE/EN
GANZ
GAST
GEBIET
GEFAHR/LICH
GEFUEHL/EN
GEHEN/GANG
GEMEINSCHAFT/CH
GERECHT/IGKEIT
GESCHAEFT/LICH
GESCHEHEN
GETREIDE
GEWINN/EN
GEWOHNHEIT/LICH
GLAUBE/N
GLUECK/LICH
GRENZE/N
GRIECHE/LAND
GROSS/E
GRUSS/EN
GUT
HALB/HAELFTE
HANDEL/N
HAUPT
HAUS
HEGEL
HELFEN/HILFE
HERRSCHAFT/EN
HERSTELLEN/R/NG
HERZ
HOCH/HOEHE
HOEREN
HOFFEN/UNG
HOLLAND/ER/ISCH
HOTEL
INDUSTRIE/IEREN
INFORMATION/REN

INTERESSE/ANT
JAHR/G/LICH
JAZZ
JOURNAL/IST
JUNG/END
KALENDER
KALT/E
KAMPF/EN
KATHOLIK/ISCH
KENNEN/TNIS
KLOSTER/LICH
KNEIPP
KOENIG/IN
KOENNEN
KOMMANDO/EUR/NT
KOMMEN
KOMMUNIST/MUS
KONGO
KONTAKT
KONTROLLE/IEREN
KOSTEN
KRAFT
KRANK/HEIT
KRIEG/ERISCH
KRITIK/ISCH/ER
KUENDIGEN/UNG
KUNST/LERISCH
KUR
KURS
KURZ/E
LAND/SCHAFT
LANG/E
LAUF/EN
LAUT/EN
LEBEN
LEHRE/N/R
LEITEN/R/UNG
LESEN/R/UNG
LETZT
LIEBE/N
LIEGEN/LAGE
LIFT
LINKS
LUFT
MACHEN
MACHT/IG
MANN
MARKT
MASCHINE

MEINEN/UNG
MELDEN/UNG
MENSCH
MILITAER/ISCH
MINISTER/IUM
MITGLIED
MITTEILEN/UNG
MODE/ERN/ISCH
MOEGLICH/KEIT
MORD/EN
MUESSEN
MUSIK/ALISCH/R
NAH/E/RN
NATION/AL
NATUR/LICH
NEHMEN
NENNEN/UNG
NEU
NUTZEN/NUETZEN
OEFFNEN/UNG
OESTERREICH
ORGANISATION
ORT
OSTEN
PARTEI
PASSIERSCHEIN
PERSON/LICH/EIT
PHILOSOPH/IE
PISTE
PLAN/EN
PLATZ
POLITIK/ER/ISCH
POLIZEI/IST
PRAESIDENT
PRAXIS/TISCH
PREIS
PROGRAMM
PROPAGANDA
PROTOKOLL
PUNKT
RAKETE
RAUM/LICH
RECHNEN/UNG
RECHT
REDAKTION/EUR
REDE/N
REFERIEREN/AT
REGEL/N/UNG
REGEN

REGIEREN/UNG
REICH/TUM
REIN/HEIT
REISE/N
REPUBLIK/ANISCH
RETTEN/UNG
REZENSION/NT
RUESTEN/UNG
RUF/EN
RUHE/N
RUNDFUNK
SACHE/LICH
SAGEN
SAMMELN/UNG
SCHLAG/EN
SCHLIESSEN/LUSS
SCHNEE/EN
SCHNELL
SCHOEN
SCHREIBEN
SCHULE/R
SCHWACH/E
SCHWER
SCHWIERIG/KEIT
SEHEN/SICHT
SEKRETAER
SENAT/OR
SENDEN/UNG
SICHER/N/EIT/NG
SITZ/N/UNG
SKI
SOLDAT
SONNTAG
SORGE/N
SOWJET/ISCH/ION
SPIEL/EN/ER
SPORT/LER
SPRACHE/N
STAAT
STADT
STAMM/EN
STARK/E
STEHEN
STEIGEN/UNG
STELLE
STELLEN
STEUER/N
STRAFE/N
STRASSE

STREIK/EN	UEBERZEUGEN/UNG	WAHL/EN
STUECK	UNO	WALD
STUNDE	UNTERNEHMEN/R	WASSER
SUCHE/N	UNTERRICHT/EN	WEG
SYSTEM/ATISCH	URLAUB	WELT
TAG/EN	USA	WERBEN/UNG
TECHNIK/ISCH	VERANTWORTEN/NG	WETTER
TEIL	VERBAND	WIRTSCHAFT
TEILNAHME/N	VERBINDEN/UNG	WISSENSCHAFT
TELEFON	VEREIN	WOCHE
TERMIN	VEREINBAREN/UNG	WOLLEN/WILLE
THEMA	VERHAELTNIS	WUNSCH/EN
THEORIE/TISCH	VERHANDELN/UNG	ZAEHLEN/ZAHL
TOD	VERKEHR/EN	ZAHL/EN
TON/EN	VERSUCH/EN	ZEIT
TOUR	VERTRAG	ZEITUNG
TREFFEN/ER	VERWALTEN/UNG	ZENTRAL/UM
TRITT/EN	VOLK	ZIEL/EN
TUERKEI/ISCH	VORSITZ/ENDER	ZOLL/EN
TUN/TAT/TIGKEIT	WACHSEN/TUM	ZONE
TYP/ISCH	WAFFE	

*D*₂: ND-Vokabular/Deskriptoren

ABEND	BEREIT/EN	DURCHSETZEN
ABGEORDNETER	BERICHT/EN	DYNAMO
ABKOMMEN	BESATZUNG	EHRE/N
ABSICHT	BESCHLUSZ/EN	ELEKTRO/NISCH
AFRIKA/NER/ISCH	BESICHTIGUNG	ENDE/N/LICH
ALLE	BESUCH/EN	ENTSCHEIDEN/UNG
ALT/ER	BETRIEB/EN	ENTWICKELN/UNG
ANTI	BETRIEB/LICH	ERFOLG/REICH
ANTWORT/EN	BEWEGEN/UNG	ERFUELLEN/UNG
ARBEIT/EN	BEZIRK	ERNTE
ARBEITER/IN	BLECH	EUROPA
ATOM/AR	BLEIBEN	EXPORT
ATOMMINE	BRAUCHEN	FAEHIG/KEIT
AUFBAU/EN	BRIGADE	FAHREN/ER/T
AUSSPRACHE/EN	BRINGEN	FAHRT/EN
AUSSTELLUNG	BRUDER/LICH	FARBE/IG
AUSZEICHNUNG/EN	BUND	FASER
BAU/EN	CHEMIE/SCH	FEIER/N
BAUER	DAMM	FERTIGUNG/EN
BAUER/IN	DANK/EN	FEST/IGEN/UNG
BAUSTELLE	DEUTSCH/LAND	FOLGE/N/UNG
BE/RATEN/UNG	DIENST/EN	FORDERN/UNG
BEGINN/EN	DOKTRIN	FRAGE/N
BEISTAND	DONNERSTAG	FRANKREICH/ISCH

FRAU	KOMITEE	POLITIK/ER/ISCH
FREUND/SCHAFT	KOMMEN	PRAESIDENT
FRIEDE/LICH	KONGRESZ	PREIS
FUEHREN/UNG	KONTROLLE/IEREN	PRODUKTION
FUEHRUNG/EN	KONZERT	PROMINENT
FUND/EN	KRAFT	PROTEST
GABE/EN	KREIS	PUNKT
GANZ	KULTUR/EN	QUALITAET
GAST/IEREN	KUNDGEBUNG	RAD
GEBIET	KUNST	RAT
GEHEN	LAND	RECHT
GEHOEREN	LANDWIRT/SCHAFT	REGIEREN/UNG
GEMEINSAM/KEIT	LANG	REVANCHE/IST
GENERAL	LANG/E	SAGEN
GENOSSE/IN	LEBEN	SCHIFF
GERING	LEGEN	SCHWER
GESCHENK/EN	LEISTEN/UNG	SEHEN/SICHT
GESPRAECH/EN	LEITEN/UNG	SEKRETAER
GETREIDE	LETZT	SOLDAT
GEWERKSCHAFT/ER	LIEBE/N	SOWJET/ISCH/ION
GLUECKWUNSCH/EN	LOHN	SOZIALIST/ISCH
GRENZE	LPG	SPALT/EN
GROSZ/E	MACHEN	SPEZIAL/IST
GRUND	MAEHDRESCHER	SPINNDUESE
GRUSZ	MAHD	SPORT/LER
GUERTEL	MANNSCHAFT	SPRACHE/N
GUT	MASCHINENBAU	SPRENGEN
GUT/E	MILITARIST/MUS	STAAT
HALBJAHR	MINISTER	STADT
HALTEN	MITTEL	STARK/E
HAND	MOEGLICH/KEIT	STAUDAMM
HANDEL/N	MOTOR	STEHEN
HAUPT	MUESZEN	STELLE
HAUPTSTADT	NACHT	STELLEN
HAUS	NAH/E/RN	STRASZE
HEBUNG/EN	NATION/AL	STREBEN/UNG
HERR/EN/SCHAFT	NEHMEN	STUNDE
HERZ/LICH	NENNEN/UNG	STURM/ISCH
HOCH/HOEHE	NEU	TAG
HOEREN	NEUJAHR	TECHNIK/ISCH
IMPERIALIST/SCH	NUTZEN/NUETZEN	TEILNAHME/EN
INDUSTRIE/IEREN	OEFFNEN/UNG	TELEGRAMM
JAHR/IG/LICH	OEKONOMISCH	THELMANN
JUNG/END	ORGANISATION	TOD
KAMPF/EN	OSTEN	TREFFEN
KANAL	PARTEI	TREFFEN/R
KLEIN	PIONIER	TRITT/EN
KOENNEN	PLAN	VERANTWORTEN/NG
KOMBINE	PLATZ	VERBINDEN/UNG

VERSUCH/EN
VERTRAG
VERTRETEN/UNG
VOLK
VORSCHLAG
WALZE
WARTEN

WASSER
WATT
WEIHNACHT
WELT
WERK
WERK/ER
WERT/EN

WETTBEWERB
WIRKLICH/KEIT
WIRTSCHAFT
WISSENSCHAFT
WOHNUNG
ZENTRAL

D_3 : Schnittmenge $D_1 \cap D_2$

ABEND
ABKOMMEN
AFRIKA/NER/ISCH
ALLE
ALT/ER
ANTWORT/EN
ARBEIT/EN/ER
ATOM/AR
BAU/EN
BEGINN/EN
BEREIT/EN
BESCHLUSS/EN
BESUCH/EN
BEWEGEN/UNG
BLEIBEN
BRAUCHEN
BRINGEN
BUND
DANK/EN
DEUTSCH/LAND
DIENST/EN
EHRE/N
ELEKTRO/NISCH
ENDE/N/LICH
ENTSCHEIDEN/UNG
ENTWICKELN/UNG
ERFOLG/REICH
ERNTE
EUROPA
FAEHIG/KEIT
FAHREN/ER/T
FEIER/N
FEST/IGEN/UNG
FOLGE/N/UNG
FORDERN/UNG
FRAGE/N
FRANKREICH/ISCH
FREUND/SCHAFT

FRIEDE/LICH
FUEHREN/UNG
FUND/EN
GABE/EN
GANZ
GAST/IEREN
GEBIET
GEHEN
GETREIDE
GRENZE
GROSS/E
GRUSS
GUT/E
HANDEL/N
HAUPT
HAUS
HERR/EN/SCHAFT
HERZ/LICH
HOCH/HOEHE
HOEREN
INDUSTRIE/IEREN
JAHR/IG/LICH
JUNG/END
KAMPF/EN
KOENNEN
KOMMEN
KONTROLLE/IEREN
KRAFT
KUNST
LAND/SCHAFT
LANG
LEBEN
LEITEN/UNG
LETZT
LIEBE/N
MACHEN
MINISTER
MOEGlich/KEIT

MUSSEN
NAH/E/RN
NATION/AL
NEHMEN
NENNEN/UNG
NEU
NUTZEN/NUETZEN
OEFFNEN/UNG
ORGANISATION
OSTEN
PARTEI
PLAN
PLATZ
POLITIK/ER/ISCH
PRAESIDENT
PREIS
PUNKT
RECHT
REGIEREN/UNG
SAGEN
SCHWER
SEHEN/SICHT
SEKRETAER
SOLDAT
SOWJET/ISCH/ION
SPORT/LER
SPRACHE/N
STAAT
STADT
STARK/E
STEHE
STELLE
STELLEN
STRASSE
STUNDE
TAG
TECHNIK/ISCH
TEILNAHME/EN

TOD	VERSUCH/EN	WIRTSCHAFT
TREFFEN/R	VERTRAG	WISSENSCHAFT
TRITT/EN	VOLK	ZENTRAL
VERANTWORTEN/NG	WASSER	ZIEL/EN
VERBINDEN/UNG	WELT	

3.2 Unterscheidende Aspekte

Die erste Analyse-Darstellung hebt auf mögliche Unterschiede der auf der lexikalisch-semanticen Ebene sich konstituierenden Wortbedeutungen ab. Hierzu werden die in den jeweiligen Datensätzen WE und ND enthaltenen Vokabulare lemmatisierter Worttypes als die (unterschiedlichen) Deskriptormengen D_1 und D_2 zur Beschreibung des Gebrauchs $G(x)$, der Bedeutung $B(x)$ und des Umgebungfeldes $E(x)$ jedes Worttypes $x \in D_3$ verwendet. In den folgenden Beispielen, die diese Darstellung veranschaulichen, wurden jedoch nicht alle, sondern nur die jeweils 30 höchstbewerteten Deskriptoren aus entweder D_1 oder D_2 ausgedruckt:

GEBRAUCH $G(x)$		DIE WELT	
$x = \text{EUROPA/ISCH}$			
POLITIK/ER/ISCH	0,557	FRANKREICH/ISCH	0,504
FRAKTION	0,472	DEUTSCH/LAND	0,386
VOLK	0,382	BEWEGEN/UNG	0,343
VORSITZ/ENDER	0,338	APPELL/IEREN	0,336
HOLLAND/ER/ISCH	0,310	FREUND/SCHAFT	0,297
GEMEINSCHAFT/CH	0,292	BLUT/EN	0,291
BE/RATEN/UNG	0,260	BRITE/ANNIEN	0,250
SAGEN	0,250	ERKLAEREN/UNG	0,248
VERTRAG	0,246	KRIEG/ERISCH	0,246
REGIEREN/UNG	0,234	DELEGATION/ER	0,225
TEILNAHME/N	0,222	ENG/E/N	0,219
GEWOHNHEIT/LICH	0,214	STATT	0,213
SEHEN/SICHT	0,212	ENTWICKELN/UNG	0,212
MINISTER/IUM	0,193	GROSZ/E	0,192
SCHWIERIG/KEIT	0,191	ZOLL/EN	0,190

STRUKTURELLE BEDEUTUNG $B(x)$		DIE WELT	
$x = \text{EUROPA/ISCH}$			
ZONE	1,000	POLITIK/ER/ISCH	1,819
DEUTSCH/LAND	1,932	FRAKTION	1,949
APPELL/IEREN	2,114	VERTRAG	2,157
GEMEINSCHAFT/CH	2,177	BLUT/EN	2,193
KRIEG/ERISCH	2,216	VOLK	2,228
TEILNAHME/N	2,239	GEWOHNHEIT/LICH	2,268
HOFFEN/UNG	2,285	MINISTER/IUM	2,309
HOLLAND/ER/ISCH	2,340	ERKLAEREN/UNG	2,370
KATHOLIK/ISCH	2,383	ATLANTIK/SCH	2,403
BUNDESTAG	2,433	NATION/AL	2,455
VORSITZ/ENDER	2,463	STAAT	2,464
KURS	2,476	DIPLOMAT/IE/SCH	2,476
ZOLL/EN	2,483	MILITAER/ISCH	2,488
FDP	2,498	DELEGATION/ER	2,533
AUSSCHUSZ	2,533	STARK/E	2,565

SEMANTISCHE UMGEBUNG $E(x)$		DIE WELT	
$x = \text{EUROPA/ISCH}$			
GEMEINSCHAFT/CH	5,473	DEUTSCH/LAND	5,556
VERTRAG	5,860	KRIEG/ERISCH	5,898
TEILNAHME/N	6,059	FRAKTION	6,260
GEWOHNHEIT/LICH	6,401	DELEGATION/ER	6,760
MINISTER/IUM	6,778	BESUCH/EN	6,820
AMERIKA/ER/ISCH	6,887	FRIEDE/LICH	7,003
ZONE	7,015	HOLLAND/ER/ISCH	7,031
OSTEN	7,089	VERSUCH/EN	7,103
WELT	7,349	VOLK	7,359
MACHT/IG	7,378	KOMMUNIST/MUS	7,405
SOWJET/ISCH/ION	7,439	WACHSEN/TUM	7,480
CHINA/SISCH	7,483	POLITIK/ER/ISCH	7,495
KAMPF/EN	7,503	GROSZ/E	7,556
BUND	7,580	BUNDESTAG	7,589
ENTSCHEIDEN/UNG	7,623	USA	7,628

GEBRAUCH $G(x)$		NEUES DEUTSCHLAND	
$x = \text{EUROPA}$			
IMPERIALIST/SCH	0,962	SPALT/EN	0,946
ANTI	0,898	NATION/AL	0,737
MILITARIST/MUS	0,729	SAGEN	0,722
STREBEN/UNG	0,639	FUEHRUNG/EN	0,629
GANZ	0,565	REVANCHE/IST	0,560
OSTEN	0,539	HOEREN	0,539
HAND	0,539	GEHEN	0,456
KOENNEN	0,452	SCHWER	0,451
BEWEGEN/UNG	0,451	LIEBE/N	0,450
MINISTER	0,450	TRITT/EN	0,449
FRANKREICH/ISCH	0,412	VERSUCH/EN	0,412
HERR/EN/SCHAFT	0,362	ABSICHT	0,360
GABE/EN	0,358	OEKONOMISCH	0,354
HAUPT	0,339	LANG	0,270
NUTZEN/NUETZEN	0,260	FEST/IGEN/UNG	0,254

STRUKTURELLE BEDEUTUNG $B(x)$		NEUES DEUTSCHLAND	
$x = \text{EUROPA}$			
IMPERIALIST/SCH	0,789	SPALT/EN	0,955
ZIEL/EN	1,000	ANTI	1,049
MILITARIST/MUS	1,510	REVANCHE/IST	2,026
FUEHRUNG/EN	2,049	SAGEN	2,148
NATION/AL	2,175	STREBEN/UNG	2,494
GANZ	2,629	SCHWER	2,646
HERR/EN/SCHAFT	2,718	BEWEGEN/UNG	2,753
OSTEN	2,770	FRANKREICH/ISCH	2,818
VERSUCH/EN	2,818	LIEBE/N	2,842
KOENNEN	2,863	MINISTER	2,893
HAND	2,953	GEHEN	3,018
HOEREN	3,077	GABE/EN	3,204
ELEKTRO/NISCH	3,266	ABSICHT	3,358
HAUPT	3,438	TRITT/EN	3,441
LANG	3,464	POLITIK/ER/ISCH	3,530

SEMANTISCHE UMGEBUNG $E(x)$

NEUES DEUTSCHLAND

 $x = \text{EUROPA}$

SPALT/EN	2,672	IMPERIALIST/SCH	2,874
ANTI	3,355	ZIEL/EN	5,555
NATION/AL	6,869	STREBEN/UNG	7,578
MILITARIST/MUS	7,648	GANZ	7,991
SAGEN	8,314	HAND	9,538
FUEHRUNG/EN	9,718	MINISTER	9,825
REVANCHE/IST	10,183	HOEREN	10,256
BEWEGEN/UNG	11,087	KOENNEN	11,615
LIEBE/N	12,155	MACHEN	12,587
ABSICHT	12,614	POLITIK/ER/ISCH	12,635
HERR/EN/SCHAFT	12,797	GABE/EN	12,914
STEHEN	13,115	OSTEN	13,173
FRIEDE/LICH	13,240	SCHWER	13,571
TRITT/EN	13,798	VERSUCH/EN	13,913
FRANKREICH/ISCH	13,913	GEHEN	14,256

GEBRAUCH $G(x)$

DIE WELT

 $x = \text{LEITEN/R/UNG}$

DIPLOM	0,964	ERFAHREN/UNG	0,949
COMPUTER	0,944	KENNEN/TNIS	0,941
FAEHIG/KEIT	0,938	SUCHE/N	0,938
SYSTEM/ATISCH	0,934	TECHNIK/ISCH	0,927
ELEKTRO/NISCH	0,901	INDUSTRIE/IEREN	0,901
BERUF/LICH	0,900	SCHREIBEN	0,879
SCHULE/R	0,876	UNTERRICHT/EN	0,844
BITTE/N	0,837	WUNSCH/EN	0,819
ORGANISATION	0,813	STELLE	0,770
UNTERNEHMEN/R	0,731	VERBAND	0,726
STADT	0,709	GEBIET	0,705
PERSON/LICH/EIT	0,690	AUSGABE/GEBEN	0,689
ALLGEMEIN	0,665	ANBIETEN/GEBOT	0,644
VERANTWORTEN/UNG	0,643	ARBEIT/EN	0,636
EINSATZ/EN	0,588	VERWALTEN/UNG	0,570

STRUKTURELLE BEDEUTUNG $B(x)$

DIE WELT

 $x = \text{LEITEN/R/UNG}$

COMPUTER	0,313	DIPLOM	0,365
ELEKTRO/NISCH	0,408	ERFAHREN/UNG	0,504
SUCHE/N	0,573	FAEHIG/KEIT	0,660
INDUSTRIE/IEREN	0,685	SYSTEM/ATISCH	0,782
SCHREIBEN	0,791	KENNEN/TNIS	0,834
BERUF/LICH	0,875	SCHULE/R	0,902
ZONE	1,000	TECHNIK/ISCH	1,097
WUNSCH/EN	1,304	ORGANISATION	1,323
UNTERRICHTEN	1,461	VERBAND	1,509
BITTE/N	1,933	STADT	1,934
GEBIET	1,952	STELLE	1,967
UNTERNEHMEN/R	2,076	AUSGABE/GEBEN	2,109
VERANTWORTEN/UNG	2,226	ALLGEMEIN	2,276
PERSON/LICH/EIT	2,465	VERWALTEN/UNG	2,483
EINSATZ/EN	2,735	VERKEHR/EN	2,743

SEMANTISCHE UMGEBUNG $E(x)$

DIE WELT

 $x = \text{LEITEN/R/UNG}$

COMPUTER	1,530	SUCHE/N	1,779
FAEHIG/KEIT	2,318	INDUSTRIE/IEREN	2,369
ERFAHREN/UNG	2,394	SYSTEM/ATISCH	2,619
ELEKTRO/NISCH	3,001	DIPLOM	3,498
BERUF/LICH	3,578	KENNEN/TNIS	3,684
SCHULE/R	4,641	TECHNIK/ISCH	4,835
SCHREIBEN	4,989	UNTERRICHT/EN	7,994
ZONE	8,635	ORGANISATION	9,874
BITTE/N	9,908	WUNSCH/EN	10,279
STELLE	13,167	UNTERNEHMEN/R	16,309
STADT	18,243	GEBIET	19,284
VERBAND	19,424	PERSON/LICH/EIT	20,500
AUSGABE/GEBEN	20,981	ANBIETEN/GEBOT	21,857
ALLGEMEIN	23,506	ARBEIT/EN	23,760
VERANTWORTEN/NG	26,253	WERBEN/UNG	26,264

GEBRAUCH $G(x)$

NEUES DEUTSCHLAND

 $x = \text{LEITEN/UNG}$

KONGRESZ	0,832	FRAU	0,660
BRAUCHEN	0,650	STURM/ISCH	0,592
BE/RATEN/UNG	0,566	BESCHLUSZ/EN	0,511
KOMITEE	0,494	GRUND	0,463
LANDWIRT/SCHAFT	0,390	RAT	0,389
BRINGEN	0,374	GETREIDE	0,371
GENOSSE/IN	0,371	AUSSTELLUNG	0,370
STAAT	0,347	SOZIALIST/ISCH	0,321
FOLGE/N/UNG	0,311	BERICHT/EN	0,291
GAST/IEREN	0,289	NUTZEN/NUETZEN	0,284
BESATZUNG	0,284	BETRIEB/LICH	0,266
WISSENSCHAFT	0,252	HERZ/LICH	0,247
OEFFNEN/UNG	0,220	INDUSTRIE/IEREN	0,204
KREIS	0,199	BESUCH/EN	0,194
BAUER	0,188	FRAGE/N	0,186

STRUKTURELLE BEDEUTUNG $B(x)$

NEUES DEUTSCHLAND

 $x = \text{LEITEN/UNG}$

ZIEL/EN	1,000	KONGRESZ	1,479
FRAU	2,045	STURM/ISCH	2,050
BRAUCHEN	2,105	BE/RATEN/UNG	2,570
KOMITEE	2,594	GENOSSE/IN	2,669
BESCHLUSZ/EN	2,724	BERICHT/EN	2,748
RAT	2,861	ARBEITER/IN	2,894
STAAT	2,917	FOLGE/N/UNG	2,933
LANDWIRT/SCHAFT	2,984	GRUND	2,991
AUSSTELLUNG	3,024	LPG	3,062
BESUCH/EN	3,062	KULTUR/EN	3,079
BRINGEN	3,099	GAST/IEREN	3,111
KONZEPT	3,177	GETREIDE	3,185
SOZIALIST/ISCH	3,212	GEHOEREN	3,226
KLEIN	3,234	STELLE	3,273
PREIS	3,301	STUNDE	3,336

SEMANTISCHE UMGEBUNG $E(x)$		NEUES DEUTSCHLAND	
$x = \text{LETTEN/UNG}$			
KONGRESZ	4,334	ZIEL/EN	5,266
BRAUCHEN	6,154	GENOSSE/IN	6,556
FRAU	6,700	STURM/ISCH	6,758
BESUCH/EN	7,741	KOMITEE	7,803
STAAT	8,632	KULTUR/EN	8,686
RAT	8,707	BE/RATEN/UNG	8,722
BESCHLUSZ/EN	8,762	GEHOEREN	8,900
AUSSTELLUNG	8,998	BERICHT/EN	9,109
GRUND	9,181	LANDWIRT/SCHAFT	9,243
GAST/IEREN	9,309	SPEZIAL/IST	9,309
BRINGEN	9,484	BAUER/IN	9,565
STUNDE	9,820	SEKRETAER	9,874
SOZIALIST/ISCH	10,033	FOLGE/N/UNG	10,070
PROMINENT	10,106	NEHMEN	10,191
WARTEN	10,356	BRIGADE	10,356

3.3 Gemeinsame Aspekte

Die zweite Analyse-Darstellung hebt auf mögliche Gemeinsamkeiten der auf lexikalisch-semanticischer Ebene des Sprachgebrauchs sich konstituierenden Wortbedeutungen ab. Hierzu werden die den WE und ND-Texten gemeinsamen Worttypen, die in der Deskriptormenge D_3 enthalten sind, zur Beschreibung des Gebrauches $G(x)$, der Bedeutung $B(x)$ und des Umgebungsfeldes $E(x)$ jedes dieser Worttypes $x \in D_3$ herangezogen. In den folgenden Beispielen, die diese Darstellung veranschaulichen sollen, wurden wiederum nicht alle, sondern jeweils nur die 30 höchstbewerteten Deskriptoren aus D_3 ausgedruckt:

GEBRAUCH $G(x)$		DIE WELT	
$x = \text{MINISTER/IUM}$			
BUND	0,488	POLITIK/ER/ISCH	0,379
DEUTSCH/LAND	0,367	TREFFEN/ER	0,296
VOLK	0,263	VERTRAG	0,256
TEILNAHME/N	0,251	REGIEREN/UNG	0,227
STAAT	0,227	AMERIKA/ER/ISCH	0,225
WELT	0,210	ENDE/N/LICH	0,208
SEHEN/SICHT	0,202	DIENST/EN	0,201
EUROPA/ISCH	0,193	BESUCH/EN	0,187
BESCHLUSZ/EN	0,185	PRAESIDENT	0,179
HAUPT	0,168	GETREIDE	0,149
ERNTE/N	0,139	SPRACHE/N	0,128
HANDEL/N	0,124	VERSUCH/EN	0,122
GROSZ/E	0,121	FRANKREICH/ISCH	0,112
BERICHT/EN	0,111	FRAGE/N	0,102
OSTEN	0,102	FEST/IGEN/UNG	0,095

STRUKTURELLE BEDEUTUNG $B(x)$		DIE WELT	
$x = \text{MINISTER/IUM}$			
BUND	1,639	DEUTSCH/LAND	1,756
POLITIK/ER/ISCH	1,905	VERTRAG	2,144
TEILNAHME/N	2,166	TREFFEN/ER	2,187
REGIEREN/UNG	2,192	PRAESIDENT	2,192
VOLK	2,197	BESCHLUSZ/EN	2,272
EUROPA/ISCH	2,309	BESUCH/EN	2,317
STAAT	2,367	DIENST/EN	2,419
NATION/AL	2,457	AMERIKA/ER/ISCH	2,506
AFRIKA/NER/ISCH	2,526	WELT	2,537
FEIER/N	2,578	GETREIDE	2,581
NENNEN/UNG	2,589	ENDE/N/LICH	2,598
BERICHT/EN	2,606	VERSUCH/EN	2,610
FRIEDE/LICH	2,621	SEHEN/SICHT	2,638
PARTEI	2,638	WASSER	2,639
SOWJET/ISCH/ION	2,643	OSTEN	2,666

SEMANTISCHE UMGEBUNG $E(x)$		DIE WELT	
$x = \text{MINISTER/IUM}$			
BUND	3,821	DEUTSCH/LAND	3,924
PRAESIDENT	4,771	VOLK	5,093
POLITIK/ER/ISCH	5,682	FRIEDE/LICH	5,947
TEILNAHME/N	6,216	SOWJET/ISCH/ION	6,364
AMERIKA/ER/ISCH	6,424	KAMPF/EN	6,661
TREFFEN/ER	6,734	EUROPA/ISCH	6,778
PARTEI	6,940	VERTRAG	7,060
VERSUCH/EN	7,081	REGIEREN/UNG	7,097
BESUCH/EN	7,198	ENTSCHEIDEN/UNG	7,322
DIENST/EN	7,350	FEIER/N	7,434
OSTEN	7,555	ENDE/N/LICH	7,738
STAAT	7,882	NENNEN/UNG	8,040
SCHWER	8,345	BERICHT/EN	8,346
WELT	8,662	ATOM/AR	8,935
BESCHLUSZ/EN	9,439	STARK/E	9,464

GEBRAUCH $G(x)$		NEUES DEUTSCHLAND	
$x = \text{MINISTER}$			
AFRIKA/NER/ISCH	0,600	NATION/AL	0,548
GANZ	0,514	HOEREN	0,494
PRAESIDENT	0,453	EUROPA	0,450
BEWEGEN/UNG	0,435	VERANTWORTEN/NG	0,372
STEHEN	0,348	FRIEDE/LICH	0,288
GABE/EN	0,286	MOEGLICH/KEIT	0,267
SAGEN	0,261	BAU/EN	0,247
BEREIT/EN	0,245	WASSER	0,242
PLATZ	0,236	NUTZEN/NUETZEN	0,225
FUEHRUNG/EN	0,223	POLITIK/ER/ISCH	0,202
FEST/IGEN/UNG	0,199	OSTEN	0,186
KOENNEN	0,172	GROSZ/E	0,170
ENDE/N/LICH	0,163	SCHWER	0,162
FRANKREICH/ISCH	0,161	VERSUCH/EN	0,161
WELT	0,156	LAND	0,152

STRUKTURELLE BEDEUTUNG $B(x)$		NEUES DEUTSCHLAND	
$x = \text{MINISTER}$			
ZIEL/EN	0,143	GANZ	2,468
BEWEGEN/UNG	2,519	NATION/AL	2,521
HOEREN	2,665	AFRIKA/NER/ISCH	2,861
EUROPA	2,893	SAGEN	3,030
STEHEN	3,063	FUEHRUNG/EN	3,121
POLITIK/ER/ISCH	3,124	GABE/EN	3,162
HERR/EN/SCHAFT	3,167	VERANTWORTEN/NG	3,171
ENDE/N/LICH	3,178	SCHWER	3,278
NENNEN/UNG	3,283	PRAESIDENT	3,301
VERSUCH/EN	3,308	FRANKREICH/ISCH	3,308
ELEKTRO/NISCH	3,335	WASSER	3,338
FRIEDE/LICH	3,364	KOENNEN	3,368
MOEGlich/KEIT	3,382	LIEBE/N	3,389
MUESZEN	3,406	ARBEITER/EN	3,414
MACHEN	3,428	LANG	3,446

SEMANTISCHE UMGEBUNG $E(x)$		NEUES DEUTSCHLAND	
$x = \text{MINISTER}$			
ZIEL/EN	0,143	BEWEGEN/UNG	6,923
SAGEN	7,285	GANZ	7,654
NATION/AL	7,861	POLITIK/ER/ISCH	7,861
STEHEN	8,223	FUEHRUNG/EN	8,646
HOEREN	8,925	MACHEN	9,135
KOENNEN	9,258	HERR/EN/SCHAFT	9,336
VERANTWORTEN/NG	9,377	GABE/EN	9,397
EUROPA	9,825	LIEBE/N	9,836
MOEGlich/KEIT	9,996	NENNEN/UNG	10,038
FRIEDE/LICH	10,051	VERSUCH/EN	10,705
FRANKREICH/ISCH	10,705	SCHWER	10,732
LANG	10,745	MUESZEN	11,261
ALLE	11,462	RECHT	11,477
BEREIT/EN	11,490	GUT/E	11,546
VERTRAG	11,820	OSTEN	11,884

GEBRAUCH $G(x)$		DIE WELT	
$x = \text{SOLDAT}$			
FEST/IGEN/UNG	0,261	KOMMEN	0,231
FRIEDE/LICH	0,212	REGIEREN/UNG	0,209
STEHEN	0,207	TRITT/EN	0,204
STELLEN	0,193	GRENZE/N	0,164
JAHR/G/LICH	0,161	LEBEN	0,158
KAMPF/EN	0,147	BERICHT/EN	0,141
BESCHLUSZ/EN	0,139	TREFFEN/ER	0,124
ALT/ER	0,119	ENDE/N/LICH	0,115
SEKRETAER	0,108	PRAESIDENT	0,108
GABE/EN	0,105	ENTSCHEIDEN/UNG	0,101
FEIER/N	0,099	KRAFT	0,092
LIEBE/N	0,091	SCHWER	0,090
DEUTSCH/LAND	0,083	DIENST/EN	0,080
BUND	0,077	LETZT	0,076
MOEGlich/KEIT	0,075	ALLE	0,066

STRUKTURELLE BEDEUTUNG $B(x)$		DIE WELT	
$x = \text{SOLDAT}$			
BESCHLUSZ/EN	2,287	FEST/IGEN/UNG	2,400
STELLEN	2,510	FEIER/N	2,537
BERICHT/EN	2,568	NENNEN/UNG	2,575
TREFFEN/ER	2,595	STEHEN	2,595
WASSER	2,610	BESUCH/EN	2,612
AFRIKA/NER/ISCH	2,627	FRIEDE/LICH	2,652
JAHR/G/LICH	2,660	NATION/AL	2,667
GETREIDE	2,699	OSTEN	2,740
REGIEREN/UNG	2,742	KONTROLLE/IEREN	2,743
KAMPF/EN	2,758	SCHWER	2,777
KRAFT	2,777	ERNTE/N	2,800
OEFFNEN/UNG	2,804	ABKOMMEN	2,813
GRENZE/N	2,822	ABEND	2,826
PRAESIDENT	2,841	STARK/E	2,852
DIENST/EN	2,856	ENDE/N/LICH	2,859

SEMANTISCHE UMGEBUNG $E(x)$		DIE WELT	
$x = \text{SOLDAT}$			
BESUCH/EN	6,053	BESCHLUSZ/EN	6,488
STELLEN	6,519	OSTEN	6,546
FEIER/N	6,560	NENNEN/UNG	7,012
ABKOMMEN	7,074	VERTRAG	7,129
WASSER	7,142	STARK/E	7,155
AFRIKA/NER/ISCH	7,157	NATION/AL	7,211
SCHWER	7,238	KONTROLLE/IEREN	7,273
STEHEN	7,579	TEILNAHME/N	7,592
GETREIDE	7,649	TREFFEN/ER	7,745
OEFFNEN/UNG	7,834	GRENZE/N	7,837
FEST/IGEN/UNG	7,884	ERNTE/N	7,899
VERSUCH/EN	7,973	FRIEDE/LICH	8,189
ERFOLG/REICH	8,343	BERICHT/EN	8,399
KAMPF/EN	8,422	ENTSCHEIDEN/UNG	8,552
BEWEGEN/UNG	8,659	JAHR/G/LICH	8,683

GEBRAUCH $G(x)$		NEUES DEUTSCHLAND	
$x = \text{SOLDAT}$			
GRENZE	0,619	DIENST/EN	0,494
VERANTWORTEN/NG	0,277	HAUPT	0,262
DANK/EN	0,234	ERFOLG/REICH	0,217
BERICHT/EN	0,211	ENTWICKELN/UNG	0,200
STAAT	0,192	JAHR/G/LICH	0,186
SPRACHE/N	0,171	STEHEN	0,167
GUT/E	0,157	BESUCH/EN	0,157
BRINGEN	0,138	STADT	0,128
VOLK	0,105	SAGEN	0,080
FRIEDE/LICH	0,079	SEKRETAER	0,078
ALLE	0,021	NEU	0,011
PREIS	0,007	ARBEITER/EN	0,006
TOD	0,001	AFRIKA/NER/ISCH	-0,001
WASSER	-0,006	ARBEITER/IN	-0,009
ABKOMMEN	-0,009	REGIEREN/UNG	-0,010

STRUKTURELLE BEDEUTUNG $B(x)$		NEUES DEUTSCHLAND	
$x = \text{SOLDAT}$			
SOLDAT	0,000	GRENZE	1,441
DIENST/EN	1,445	DANK/EN	2,456
PREIS	2,509	BERICHT/EN	2,591
ERFOLG/REICH	2,639	ARBEITER/EN	2,732
TOD	2,776	HAUPT	2,816
BLEIBEN	2,828	SEKRETAER	2,890
STELLE	2,904	BESUCH/EN	2,935
ARBEITER/IN	2,974	ENDE/N/LICH	2,980
ABKOMMEN	2,991	JAHR/G/LICH	3,020
BAU/EN	3,030	GUT/E	3,065
VERBINDEN/UNG	3,074	REGIEREN/UNG	3,074
ELEKTRO/NISCH	3,075	AFRIKA/NER/ISCH	3,085
BRINGEN	3,091	NEHMEN	3,094
PRAESIDENT	3,115	WASSER	3,126
PLATZ	3,128	FEIER/N	3,141

SEMANTISCHE UMGEBUNG $E(x)$		NEUES DEUTSCHLAND	
$x = \text{SOLDAT}$			
SOLDAT	0,000	DIENST/EN	3,540
GRENZE	7,006	ARBEITER/EN	7,125
DANK/EN	7,230	TOD	7,292
PREIS	7,366	STELLE	7,553
ABKOMMEN	7,851	BLEIBEN	7,879
BERICHT/EN	8,036	BAU/EN	8,085
AFRIKA/NER/ISCH	8,488	NEHMEN	8,636
ENDE/N/LICH	8,652	REGIEREN/UNG	8,671
ARBEITER/IN	8,758	PRAESIDENT	8,863
HANDEL/N	9,225	WASSER	9,287
PUNKT	9,328	PLATZ	9,338
VERBINDEN/UNG	9,450	FAHRT/EN	9,548
SEKRETAER	9,624	ELEKTRO/NISCH	9,658
KONTROLLE/IEREN	9,666	FEIER/N	9,804
FOLGE/N/UNG	9,871	LANG/E	9,925

Bei der Beurteilung der vorliegenden Bedeutungsrepräsentationen einzelner lemmatisierter Worttypes sollte nicht unerwähnt bleiben, daß die hier vorgelegten Resultate nicht nur durch die zu kleinen, aus dem WE- und ND-Material entnommenen Datensätze beeinträchtigt werden, sondern — im Hinblick auf den angestrebten Wortschatzvergleich — auch durch die unterschiedlichen Umfänge der für diese Pilot-Untersuchung zur Verfügung gestellten und analysierten Datensätze eingeschränkt erscheinen.

4 Literatur

- GAINES, B.R. (1976): „Foundations of fuzzy reasoning“ Int. Journ. of Man-Machine Studies 8, 623–668.
- HARRIS, Z.S. (1968): Mathematical Structures of Language (Interscience Tracts in Pure Mathematics 21). New York/London/Sidney.
- HELLMANN, M.W. (1976): „Möglichkeiten und Probleme der vergleichenden Wortschatzuntersuchung zum öffentlichen Sprachgebrauch in der BRD und DDR“ .

- In: Moser, H. (Hrsg.): Probleme der Lexikologie und Lexikographie (Jahrbuch 1975 des Instituts für deutsche Sprache). Düsseldorf 242–274.
- HELLMANN, M.W. (1979): Zum sprachlichen Ost-West-Problem (Wege der Forschung). Darmstadt (in Vorbereitung).
- KLEIN, W. (1974): Variation in der Sprache (Scripten, Linguistik und Kommunikationswissenschaft). Kronberg/Ts.
- KLEIN, W. (1976): „Sprachliche Variation“. Studium Linguistik 1, 29–46.
- LABOV, W. (1970): „The Study of Language in its Social Context“. Studium Generale 23, 30–87.
- LABOV, W. (1973): „The boundaries of words and their meaning“. In: Bailey/Shuy (Eds.): New Ways of Analyzing Variation in English. Washington 340–373.
- RIEGER, B. (1972): „Warum mengenorientierte Textwissenschaft? Zur Begründung der Statistik als Methode“. Zeitschrift für Literaturwissenschaft und Linguistik (LiLi) 8, 11–28.
- RIEGER, B. (1974): „Eine tolerante Lexikonstruktur. Zur Abbildung natürlich-sprachlicher Bedeutungen auf unscharfe Mengen in Toleranzräumen“. Zeitschrift für Literaturwissenschaft und Linguistik (LiLi) 16, 31–47.
- RIEGER, B. (1975): „On a tolerance topology model of natural language meaning“. Vortrag auf der 2nd Intern. Conf. on Computers in the Humanities (ICCH/2), Los Angeles, USA (vervielf.).
- RIEGER, B. (1977a): „Theorie der unscharfen Mengen und empirische Textanalyse“. In: Klein, W. (Hrsg.): Methoden der Textanalyse. Heidelberg 84–99.
- RIEGER, B. (1977b): „Bedeutungskonstitution. Einige Bemerkungen zur semiotischen Problematik eines linguistischen Problems“. Zeitschrift für Literaturwissenschaft und Linguistik (LiLi) 27/28, 55–68.
- RIEGER, B. (1977c): „Vagheit als Problem der Linguistischen Semantik“. In: Sprengel/Bald/Viethen (Hrsg.): Semantik und Pragmatik. Tübingen 91–101.
- RIEGER, B. (1979a): „Fuzzy Structural Semantics: on a generative model of vague natural language meaning“. Trappl/Hanika/Pichler (Eds.): Progress in Cybernetics and Systems Research, Vol. V. New York/London/Sydney 495–503.
- RIEGER, B. (1979b): „Linguistic Semantics and the Problem of Vagueness: on analysing and representing word meaning“. In: Ager/Knowles/Smith (Eds.): Advances in Computer-Aided Literary and Linguistic Research. Birmingham 271–288.
- RIEGER, B. (1979c): „Revolution, Counterrevolution or a New Empirical Approach to Frame Reconstruction instead?“. In: Petöfi, J.S. (Ed.): Text vs. Sentence. Basic Questions of Textlinguistics, Vol. II. Hamburg 555–571.
- RIEGER, B. (1979d): „Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung“. In: Bergenholtz/Schaeder (Hrsg.): Textcorpora. Materialien für eine empirische Textwissenschaft. Kronberg/Ts. 52–70.
- SALTON, G./YANG, C.S./YU, C.T. (1974): „A Theory of Term Importance in Automatic Indexing“. Intern. Conference on Computational Linguistics (COLING 76). Ottawa (mimeogr.).

- SALTON, G. (1976): „A Comparison of Term Value Measurements of Automatic Indexing“. Intern. Conference on Computational Linguistics (COLING 76).
- SCHAEDER, B. (1978): „Maschinelle Dokumentation und Lexikographie. Ausführungen zum DFG-Projekt: Ost-West-Wortschatzvergleiche“ In: Krallmann, D. (Hrsg.): Kolloquium zur Lage der linguistischen Datenverarbeitung. Essen 110–164.
- SCHAEDER, B. (1979): „Zur Methodik der Auswertung von Textkorpora für die Zwecke der Lexikographie“. In: Bergenholtz/Schaeder (Hrsg.): Empirische Textwissenschaft. Königstein/Ts. 220–267.
- SPARCK-JONES, K./KAY, M. (1973): Linguistics and Information Science. New York.
- SUPPES, P. (1972): „Probabilistic Grammars for Natural Languages“. In: Davidson/Harman (Eds.): Semantics of Natural Language. Dordrecht 741–762.
- ZADEH, L.A. (1965): „Fuzzy Sets“. Information and Control 8, 338–353.
- ZADEH, L.A. (1971): „Quantitative fuzzy semantics“. Information Science 3, 159–176.