

Warum Fuzzy Linguistik? Überlegungen und Ansätze einer computerlinguistischen Neuorientierung*

Burghard Rieger**

Lehrstuhl für Computerlinguistik

Fachbereich II: Linguistische Datenverarbeitung – Universität Trier

1 Zur Situation der Computerlinguistik

In der computerlinguistischen Forschung zeichnen sich derzeit weltweit Veränderungen ab, die übergreifend durch Verschiebungen ihres *Erkenntnisinteresses*, durch Kritik ihrer *Untersuchungsmethoden* sowie durch Neubestimmungen ihrer *Forschungsgegenstände* charakterisierbar sind. Als Indikatoren dieser Veränderungen können zum einen die Erfolge gelten, die dem *neue Konnektionismus* mit der Modellierung von Phänomenen gelangen, welche die natürliche Sprache und ihre Verwendung ganz offensichtlich bestimmen (Lernbarkeit, Veränderbarkeit, Fehlertoleranz, Unschärfe, etc.), in der Computerlinguistik aber gleichwohl (zunächst) unberücksichtigt blieben. Zum anderen bilden natürlich-sprachliche Textkorpora in bisdahin nicht vorstellbaren, geschweige denn verfügbaren Umfängen von mehreren 10- oder gar 100-Millionen Wörtern einen neuen sprachwissenschaftlichen Gegenstandsbereich, dessen Datenmassen empirisch-quantitative Analysen nahelegten und zur Entwicklung auch neuer Untersuchungsmethoden Anlaß gaben und noch geben. In der (Muster-basierten) maschinellen Übersetzung führten diese etwa zu Ergebnissen, die denen (Regel-basierter) Ansätze mehr als nahekommen.

1.1 Die Modellbildung

In beiden Bereichen kam es so zu Modellbildungen, die traditionelle Ansätze der Computerlinguistik wie der sprachverarbeitenden künstlichen Intelligenz in sehr grundsätzlicher Weise in Frage stellen. Maschinelle Sprachanalyse- und -Synthese-Systeme basieren bekanntlich auf korrekten Beschreibungen syntaktischer Strukturen und den ihnen zuzurechnenden semantischen Interpretationen. Hierzu wurden Repräsentationen des (syntaktischen und lexikalischen) *Sprachwissens* in Form von Ersetzungsregeln und deren kontrollierter Abarbeitung (Grammatikformalismen) einerseits und Repräsentationen des

*Der Aufsatz nimmt Ideen, Gedanken und Skizzen des Autors auf, die—im Laufe der Jahre entstanden und aus Anlaß verschiedener Vortragsverpflichtungen entwickelt—in einer früheren Fassung schon auf der Jahrestagung 1995 der *Gesellschaft für Linguistische Datenverarbeitung (GLDV)* in Regensburg mitgeteilt aber nicht veröffentlicht wurden. Die vorliegende Überarbeitung des auf dem Essener *Gerold-Ungeheuer-Symposium* gehaltenen Vortrags stellt überdies Fragen der adäquaten *linguistischen* Kategorienbildung [Rie95a] in den Zusammenhang *semiotisch* angemessener Modellierungen, wobei Planungsergebnisse eines neuen Forschungsprojekts zum Thema sowie die Einrichtung eines Arbeitskreises *Fuzzy Linguistik* in der *GLDV* beträchtlichen Einfluß hatten.

Erschienen in: Krallmann, D./Schmitz, H.W. (Hrsg.): Vorträge des Internationalen Gerold-Ungeheuer-Symposium, Essen 1995, Münster (Nodus Wissenschaftsverlag) 1996.

**E-mail: rieger@ldv01.Uni-Trier.de

(referentiellen und situativen) *Weltwissens* in Form von prädikatenlogisch motivierten Strukturen und deren inferenzieller Verarbeitung (Wissensrepräsentationsformalismen) andererseits benötigt und bereitgestellt. Derartige Repräsentationsformalismen sind Bestandteil der eben deswegen *wissensbasiert* genannten, kognitiven Modellbildung, die auf der Grundlage und mit Hilfe symbolanalytischer Techniken (monotone Logiken, symbolische Repräsentationen, regelbasierte Operationen, serielle Verarbeitung, etc.) zunehmend komplexere Systeme entwickeln und erproben ließ. Diese sehen sich—wenngleich aus unterschiedlichen Gründen—von den konnektionistischen ebenso wie den empirischen Ansätzen und deren Resultaten herausgefordert.

Während konnektionistische Modelle durch den Einsatz prozeßsimulierender Techniken (Theorie der dynamischen Systeme, sub-symbolische oder verteilte Repräsentationen, kontinuierlich-numerische Operationen, parallele Verarbeitung) vor allem dynamische Eigenschaften sprachverarbeitender kognitiver Systeme nachzubilden vermögen [NCCH89], [PSFS89], [CG91], haben die empirisch-quantitativen Analysen sehr großer Sprachkorpora (wahrscheinlichkeits- und possibilitätstheoretische Basis, stochastische bzw. fuzzy Modelle, numerisch-mathematische und fuzzy-logische Operationen, strenge Verfahren der Hypothesenbewertung) vor allem den Reichtum an funktionalen Zusammenhängen sprachlicher Regularitäten und Strukturen erkennen lassen [Arm93] [KR93], der große Textmengen tatsächlicher Sprachverwendung beobachtbar auszeichnet.

1.2 Die Diskussion

Die Unterschiede konnektionistischer vs. regelbasierter Ansätze und Modellbildungen [Smo88], [FP88], [Smo89a], [Sch90] werden dabei von Positionen aus diskutiert, für die ein kognitiv-linguistisches Erkenntnisinteresse verbindlich zu sein scheint. Für Vertreter beider Positionen steht offenbar die Erforschung des Sprachvermögens, seiner Komponenten und deren Organisation im Vordergrund. Die Diskussion konzentriert sich daher vor allem auf die mit den unterschiedlichen, auch *hybriden* Modellbildungen verbundenen Erklärungsansprüche [PSFS89], [Bun92], [Hin92]. Die Auseinandersetzung mit den empirisch-linguistischen Ansätzen, die nicht nur wegen der für Linguisten neuen und lernaufwendigen Untersuchungsmethodik erschwert wird, hat kaum begonnen. Sie wird durch die mit der Analyse von Massendaten veränderte linguistische Gegenstandsbestimmung, der bisher ein noch uneinheitliches Erkenntnisinteresse [CM93] entspricht, zudem weiter verzögert.

Immerhin beginnt sich unter dem Eindruck des formalen Aufwands wie der praktischen Beschränktheit wissensbasierter Modelle einerseits, der überraschenden Leistungsfähigkeit stochastischer Parser [Chu88], [BC93] und statistischer Übersetzungssysteme [Bro88], [BDDM93] andererseits, eine mögliche Revision der für die kognitive Linguistik verbindlichen Grundhypothese abzuzeichnen. Diese besagt nach [Cho88], daß die natürliche Sprache vor allem als Problem der Struktur und des Erwerbs von Sprachvermögen (*Kompetenz*) zu erforschen und daß dies o h n e die Kenntnis empirisch überprüfbarer Parameterwerte aus der Anwendung und Realisation solchen Vermögens in Situationen kommunikativer Sprachverwendung (*Performanz*) auch (theoretisch) zu analysieren möglich und (formal) zu charakterisieren nötig sei.

1.3 Das Resultat

Ein Ergebnis dieser Auseinandersetzungen ist die wachsende Einsicht [Sch94b], daß die Basis bisheriger kognitiv-linguistischer Forschung sich als eine zu weitgehende Abstraktion der sprachlich-kommunikativen Realität erweisen könnte. Angesichts von nur empirisch nachweisbaren, durch etablierte linguistische Kategorien aber nicht erfäßbare sprachliche Strukturen,¹ und vor dem Hintergrund komplexer regelbasierter Modellbildungen in der kognitiven Linguistik (KL), Computerlinguistik (CL) und der sprachverarbeitenden künstlichen Intelligenzforschung (KI) [Rie88], bestehen die Möglichkeiten einer empirisch-performativen Linguistik gerade in ihren komplementären Forschungsgegenständen und -methoden. Sie erlauben quantitativ-statistische wie auch fuzzy-theoretische Modellbildungen und könnten damit eine *semiotische* Neu- oder Umorientierung des kognitiv-linguistischen Erkenntnisinteresses einleiten.

Durch die Verfügbarkeit sehr großer Textkorpora², die natürlichsprachliche Eigenschaften und Phänomene auf der Grundlage von Datenmengen studieren lassen, deren Umfang allein schon eine Neubestimmung dieser Erscheinungen und ihrer Strukturen fordert, erweisen sich bisherige Modellierungsformen zunehmend als inadäquat. Sie führen vermehrt zu Randunschärfen, großen Variationsbreiten und vielfältigen Ambiguitäten bei der Beobachtung sprachlicher Strukturen in performativen Massendaten. Klassisch-kategoriale Konzepte der Linguistik sind daher in einem ersten Schritt zu überprüfen und möglicherweise durch Konzepte *weicher Kategorien* zu erweitern mit Aussicht auf eine bessere Chance, in einem zweiten Schritt Wissen (über Sprache und die Welt) als eine prozedurale Form der kognitiven Strukturierungsleistung *semiotischer Prozesse* verstehen und erklären zu können.

1.4 Die Aufgabe

Die Modellierung derartiger *performativer* Prozesse unterscheidet sich von der Wertzuweisenden Auffüllung schon vorgegebener, weil *kompetenztheoretisch* vorausgesetzter Repräsentationsstrukturen dadurch, daß (diesen vergleichbare, neuartige) Repräsentationen als Strukturierungsleistung von den prozeßmodellierenden Prozeduren selbst erbracht werden. Ihre Resultate werden zunächst in Form von Verteilungen oder Vektoren notiert, die in der Regel hochdimensionierte (metrische) Raumstrukturen (*semiotische Räume*) bilden und deren Elemente als *unscharfe (fuzzy)* Mengen gedeutet werden

¹Phänomene etwa der linearen Nah-Ordnung sprachlich-performativer Einheiten (z.B. Kokkurenzen), deren Regularitäten sich regelbasierten Notationen entziehen aber in numerischen Ausdrücken von Korrelationen beliebig genau notiert und verarbeitet werden können, erweisen sich als Resultate von Prinzipien, die kompetenztheoretisch bisher nur deswegen übersehen wurden, weil sie keine der linearen Fern-Ordnungen linguistischer Kategorien betreffen, wie sie gängige Grammatikformalismen erfassen [Fux54], [CH90], [HR93].

²Das Trierer *dpa*-Korpus besteht beispielsweise aus Textmaterial des sog. Basisdienstes der Jahrgänge 1990–93 (720.000 Dokumente), die dem Autor von der Deutschen Presseagentur (*dpa*), Hamburg, dankenswerterweise zu Forschungszwecken zur Verfügung gestellt wurde. Nach Säuberung (von Formatier- und Steuerzeichen) umfaßt das Korpus rund 180 Mio. laufende Wörter (*Tokens*), für die derzeit ein maschinelles Lemmatisierungs- und Tagging-Verfahren entwickelt wird. Dieses Textkorpus bildet das schriftsprachliche Datenmaterial schon laufender und erst noch geplanter *fuzzy-linguistischer* Untersuchungen.

können. So werden sich anhand von (*morphischen*) Vektoren, die auf n-Grammen von Buchstaben und anhand von (*semischen*) Vektoren, die aus Korrelationen von Wörtern berechnet wurden, vermutlich jene Zusammenhänge studieren lassen, die erst durch vektorieLL repräsentierte *fuzzy* Elemente und Relationen des *morphologischen/ bzw. semantischen* Raumes erkennbar werden. Inwieweit dabei Modelleigenschaften denen der Modellobjekte entsprechen, wird zu prüfen und in jedem Fall schwierig zu entscheiden sein. Sie ist aber eine Voraussetzung der Untersuchung von *semiotischen* Funktionen solcher Entitäten, die unterschiedliche Ordnungen, Strukturen und Bedeutungen konstituieren, vielfältiger und in einem (vermutlich) größerem Umfang, als linguistische Beschreibungen bisher zu erfassen, zu analysieren und zu repräsentieren vermochten.

2 Zur semiotischen Dimension

Alle Wissenschaften arbeiten mit Begriffen, die im Hinblick auf einige der Entitäten (Objekte, Eigenschaften, Prozesse, Sachverhalte, etc.), mit denen sie umgehen, als *Idealisierungen* gelten müssen. Objektwissenschaften—wie Metawissenschaften—stehen daher vor der Aufgabe, die von ihnen verwendeten Begriffe einerseits in Form von bestimmten Zeichen (Designatoren) einzuführen und andererseits zu bestimmen, was diese Zeichen jeweils bezeichnen (Designation). Das geschieht im allgemeinen durch Angabe sog. fachsprachlicher Begriffs- oder Bedeutungsdefinitionen, die in Form von Regeln, Verfahren und/oder Operationen explizieren, wie welche Zeichen (und Zeichenaggregate) sich auf welche Designationen beziehen (lassen). Dabei ist es zweckmäßig, *Zuordnungsbeziehungen*, welche die Möglichkeiten festlegen, korrekte Zeichen-Zeichen-Verbindungen zu aggregieren (*Syntaktik*)³, von den *Korrespondenzrelationen* zu unterscheiden, welche die Beziehungen zwischen Zeichen (und Zeichenaggregaten) einerseits und dem von diesen Bezeichneten andererseits regeln (*Semantik*). Innerhalb jeder Disziplin ergibt sich mit der Ausbildung ihrer jeweiligen Fachsprache eine quasi metawissenschaftliche Aufgabenstellung, der sich die Sprachwissenschaft im Hinblick auf die natürliche Sprache auch objektwissenschaftlich gegenüber sieht. In dieser doppelten Aufgabenstellung linguistischer Forschung besteht die weit über die Linguistik hinausweisende Bedeutung ihrer Begriffs- und Kategorienbildung.⁴

2.1 Überprüfbarkeit

Insbesondere für die empirischen Objektwissenschaften erweist sich, daß sie dem Anspruch der (durch intersubjektiven Zustimmungszwang qualifizierten) Überprüfbarkeit ihrer Aussagen nur in dem Maße zu genügen vermögen, wie sie die Klärung der *Zuordnungs- und Korrespondenzregeln* für ihre fachwissenschaftlichen Notationssysteme haben vorantreiben können. Mit der möglichst eindeutigen Bestimmung und definitonischen Festlegung dieser Beziehungen ergeben sich zahlreiche Aufgaben- und Fragestellungen,

³Im Unterschied dazu sprechen wir von *Syntax*, wenn diese Zuordnungsbeziehungen in Form von (Ersetzungs-)Regeln, Produktionen, etc., d.h. selber wiederum in (formal-)sprachlichen Notationen repräsentiert werden.

⁴Vgl. [Rie95a]

denen sich u.a. Sprachphilosophie, Psychologie und Linguistik widmen. Was dabei die Schwierigkeit der Bestimmung dieser *Zuordnungs-* und *Korrespondenz-*Beziehungen, die in allen Disziplinen mit der erfolgreichen Theorienkonstruktion und Modellüberprüfung verbunden ist, zu einem Disziplinen übergreifenden und wissenschaftstheoretisch relevanten Problem macht, ist seine *semiotische* Dimension, die ontologisch begründet ist. Denn jede (wenn auch unbefriedigende) Lösung dieses Problems kann als eine Bedingungen der Möglichkeit gelten, Erfahrungen, die im Umgang mit der Welt gemacht werden, überhaupt repräsentieren bzw. sprachlich darstellen, über diese Darstellungen vermitteln und aufgrund der diese Vermittlung regelnden Verstehensprozesse auch überprüfen zu können.

Im Hinblick auf diejenigen Prozesse, die das Verstehen natürlichsprachlicher Repräsentationen ausmachen, hat sich die Sprachwissenschaft und Linguistik engagiert und zum Objekt ihrer wissenschaftlichen Untersuchungen gemacht, ohne deshalb eine metawissenschaftliche Disziplin zu begründen. Über die beschreibende Analyse und (re-)konstruktive Modellierung der sog. *Form-Funktion-* bzw. *Ausdruck-Inhalt-*Beziehungen haben insbesondere die computerlinguistischen Forschungen etwa zur *Syntax* der natürlichen Sprachen mit der Theorie der formalen Sprachen als automatentheoretischem Modell, in der *Semantik* der natürlichen Sprache mit referenztheoretisch oder bedeutungsstrukturell modellierenden Repräsentationsformalismen entscheidende Beiträge geliefert. Dabei wird bis heute mit Begriffen und Kategorien gearbeitet, die als Idealisierungen nicht etwa eine Konsequenz strikter Theorien- und Modellbildungen sind, sondern oft selber den Grund bilden für eine erschwerte bis unmögliche Hypothesenbildung mit deswegen kaum forderbarer bis leistbarer strikten Überprüfbarkeit.

2.2 Notation

Solange eine umfassende Sprachtheorie fehlt, welche die formal erklärenden Terme (ihres *theoretischen* Teils) über wohldefinierte Meßoperationen (eines *Zuordnungsteils*) mit quantitativ numerischen Termen (ihres *empirischen* Teils) verbindet, ist es (noch) nicht möglich, linguistische Gesetzeshypothesen so zu formulieren, daß daraus beobachtbare Zusammenhänge deduktiv prognostiziert und in Form experimenteller Versuche auch getestet werden könnten. Zusammenhänge, die daher nur aufgrund bestimmter Beobachtungen und eher rudimentärer theoretischer Vorstellungen bestenfalls vermutet werden können, werden—in Ermangelung theoretischer Gesetzeshypothesen, aus denen sie ableitbar wären, und in Ermangelung auch einer damit bereitstehenden formalen Notationsform, welche sie überprüfbar zu repräsentieren erlaubte—deshalb nur als Zusammenhang von Beobachtungen mitteilbar sein oder als Resultat jener Prozesse beschrieben werden können, die für das Zustandekommen dieser Beobachtungen vermutet werden. Dabei zeigt sich leider, daß das Vermögen, mittels natürlichsprachlicher Zeichen und deren Aggregation *Bedeutungen* konstituieren zu können⁵, wissenschaftliche Aussagen ins-

⁵Obwohl der Prozeß der *Bedeutungskonstitution* in [Rie94a] als herausragende Leistung und *semiotische* Dimension der natürlichen Sprache gekennzeichnet wird, ist es gerade diese Eigenschaft, welche sich einer Modellierung (bisher) entzog und Inhalte neuer *Bedeutungen* natürlichsprachlicher Ausdrücke auch weiterhin ungeeignet sein läßt, etwa durch ein *formales* Verfahren wie der Wahrheitswert-Zuweisung für prädikaten- und aussagen-logische Terme, bewertet, d.h. anders als durch interpretatives *Verstehen*

besondere solcher Disziplinen zu überprüfen erschwert, deren *Forschungsgegenstand* aus natürlichsprachlichen Zeichenketten besteht, deren *Untersuchungsmethoden* auf ein analysierendes Segmentieren und Kategorisieren sprachlicher Strukturen hinausläuft, und deren *Erkenntnisinteresse* dem Erkennen und Verstehen dieser Strukturen und ihrer Funktionen gilt.

Für die Formulierung diesbezüglicher Hypothesen, die als Behauptungen *vermuteter Zusammenhänge* ein überprüfbares Modell sollen bilden können, ist deswegen eine besondere Form ihrer Repräsentation Voraussetzung. Erst sie kann gewährleisten, daß zwischen denjenigen Bedingungen, die eher beim Zustandekommen der *Vermutung* beteiligt sind, und solchen Bedingungen, die eher für die Entstehung der *Zusammenhänge* maßgeblich sind, unterschieden wird. Natürlichsprachlich formulierte Hypothesen können diese Unterscheidung immer nur natürlichsprachlich vermitteln, nicht aber formal explizieren. Sie müssen die Plausibilität daher dem individuellen Akt einer verstehenden Interpretation solcher Hypothesen überlassen anstatt ihre Überprüfung operational kontrolliert an Kalküle oder Testprozeduren zu delegieren. Dazu wäre ihre formalsprachliche Präzisierung notwendig, die diese Unterscheidung wie den Grad ihrer Plausibilität explizieren und notieren können müßte. Formalen Notationssysteme mit zugeordneten Meßoperationen sind daher eine wesentliche Voraussetzung umfassender wissenschaftlicher Theoriebildung und experimenteller Überprüfbarkeit ihrer Modelle.

2.3 Optimierung

Die Entwicklung und Differenzierung von Wissenschaft in eine Vielzahl (mehr oder weniger) eigenständiger Disziplinen stellt sich als kontinuierlicher Prozeß primären Gewinnens, vermittelnden Darstellens, lernenden Erwerbs und überprüfenden Veränderns von neuen Erkenntnissen der an diesem Prozeß Beteiligten dar. Dieser durch kritisches Verstehen gekennzeichnete historische Prozeß kann seine optimierende Wirkung offenbar deswegen entfalten, weil die zeichenhaft-symbolische Repräsentation individueller Erkenntnisse sie aus der Gebundenheit an ihre Produzenten und Rezipienten (weitgehend) befreit und (durch zunehmend höhere Auflösungsvermögen der verwendeten Notationssysteme) in eine besondere Form (*Wissen*) überführt, welche die Intersubjektivität des Zugangs (*Verstehen*) wie der Überprüfbarkeit (*Kritik*) sichert. Dies leistet im wesentlichen eine von den Erkenntnisträgern abstrahierende Transformation, durch die die situative (Orts- und Zeit-) Gebundenheit (*Kontexte*) aller Erfahrung in vereinheitlichende Ordnungen (*Strukturen*) überführt und übergreifend in abstrakten Repräsentationen (*Systeme* oder *Theorien*) aufgehoben wird⁶, die zwischen den Ausgangsbedingungen (*Erkenntnisinteresse*), den Objekten (*Forschungsgegenstand*) und den Verfahren (*Untersuchungsmethoden*) zu unterscheiden und sie zu spezifizieren erlaubt. Erst eine (auch notationelle) Differenzierung dieser Bestimmungsstücke von Erkenntnis läßt die (metawissenschaftliche) Präzisierung ihrer wechselseitigen Abhängigkeiten zu und erlaubt die Festlegung von Kriterien der Überprüfbarkeit von Aussagen, die innerhalb so bestimmter

realisiert und eingeordnet zu werden.

⁶Vgl. hierzu schon [Rie72] und die in [Rie89] vorgelegte semantische Vagheitstheorie mit ausgearbeiteten formal-theoretischen und quantitativ-empirischen Teilen sowie den beide verbindenden und spezifizierenden Zuordnungs- und Korrespondenzregeln.

Disziplinen als wissenschaftliche Hypothesen gelten können. Was so durch die einander ablösenden wissenschaftlichen Theorien in der Zeit [Kuh62] als Optimierung erscheint, hat seine Entsprechung im Prozeß der Bedeutungskonstitution und des Verstehens, der den situationsangepaßten und deswegen kommunikativen Gebrauch natürlicher Sprache zur (kognitiven) Orientierung in der Welt kennzeichnet.

Übergreifend kann man daher die zureichende Bestimmtheit *syntaktischer Zuordnungen* und *semantischer Korrespondenzen* als Bedingung insbesondere solcher Prozesse verstehen, die durch standardisierte Zeichenverwendung eben diese Standards zu verändern vermögen [RT89] [Rie91a]. Das ist freilich kein Selbstzweck, weder der Wissenschaft noch der Sprache, sondern Ergebnis und Voraussetzung kommunikativer Interaktionen zeichenverwendender, informationsverarbeitender Systeme mit dem Ziel, durch Veränderungen solcher Standards den variablen Umgebungsansprüchen effizienter genügen und unterschiedlichen Systembedürfnissen effektiver entsprechen zu können [Rie94a] [Rie95c]. Zur Charakterisierung der Dynamik derart wechselseitig sich anpassender Veränderung, die nicht auf einem bloß kybernetischen Ausgleich vorgegebener *Soll-Ist*-Größen von im übrigen stabilen System- und Umgebungsstrukturen beruht, sondern selbst die Unterscheidung zwischen systemtheoretisch so grundlegenden Komponenten wie *System* und seiner *Umgebung* als Resultat *emergenter* Strukturierungen deutet, die erst durch *selbstorganisierende* Prozesse der Informationsverarbeitung bewirkt wird, bietet sich das *ökologische* Paradigma der Beschreibung, Analyse und Modellbildung [Rie95b] [Rie96] an.

3 Zum Defizit der Linguistik

Anders als in den mit *nicht-semiotischen* Forschungsgegenständen befaßten Objektwissenschaften (wie etwa der Experimentalphysik, Chemie, Biologie, etc.) aber auch anders als in den von *semiotisch* vermittelten Forschungsgegenständen bestimmten Beobachtungswissenschaften (wie etwa Psychologie, Soziologie, Ethnologie, etc.) hat die Linguistik—wie durch ihren Forschungsgegenstand *Sprache* verführt—sich von bestimmenden Entwicklungen der modernen Wissenschaftstheorie seit Ende der 70er Jahre⁷ nahezu unberührt gezeigt. Dabei scheint die traditionelle *langue-parole-language*-Differenzierung DESAUSSURES ebenso wie die nicht minder einflußreiche *competence-performance*-Unterscheidung CHOMSKYS eine ganze Klasse von Problemen nachhaltig überdeckt zu haben. Hierzu zählen insbesondere jene, die mit der Frage entstehen, wie denn eine überprüfbare, durch kontrollierte Verfahren wirksame, wechselseitige Vermittlung von formaltheoretischen Konstrukten einerseits und empirisch-deskriptiven Begriffen andererseits so gewährleistet werden könne, daß deren *Theorie*-geleitete, *methodische* Anwendung auf *empirische* Gegenstandsbereiche—durch experimentelle Erprobung in hierzu abgeleiteten Modellen—zur kontinuierlichen Verbesserung dieser Vermittlungsleistung beitrage.

⁷Vgl. etwa die bisdahin rege Diskussion zumindest in Deutschland um wissenschaftstheoretische Positionen der Sprachwissenschaft und Linguistik [PR70], [Rie72] [Wun76], [Rie77], [Fin79], [BS79] etc.

3.1 Zuordnung und Korrespondenz

Als ein möglicher Grund für die Leichtigkeit, mit der die auch in der Sprachwissenschaft anstehenden Probleme übersehen wurden, mag gelten, daß linguistische Konzepte und Abstraktionen den Mangel sprachwissenschaftlicher Begriffs- und Kategorienbildungen offenbar nachhaltig verdeckten. Da sie ja (meist) jeweils einem der beiden für die linguistische Beschreibung unterschiedenen Untersuchungsbereiche des *Sprachvermögens* (*Kompetenz*) oder des *Sprachgebrauchs* (*Performanz*) zugeordnet werden können⁸, blieb auch das für die Sprachwissenschaft insgesamt charakteristische unvermittelte Nebeneinander von (spekulativer) *Theorie* und (weicher) *Empirie* scheinbar unproblematisch, obwohl den formalen Termen theoretischer Aussagen keine Meßoperationen oder Beobachtungsterme empirischer Aussagen in überprüfbarer Weise systematisch zugeordnet werden können. Weitgehend übersehen wurde und wird dabei, daß auf den unteren *semiotischen* Beschreibungsebenen (Phonologie, Morphologie, IC-Syntax)—durch die klassisch strukturalen Verfahren der *Segmentierung* und der *Klassifizierung* vermittelt—noch eine (wenn auch schwach) operationale Klärung der fachterminologischen *Korrespondenzen* zwischen theoretischem Konstruktbezug und empirischem Beobachtungsbegriff methodisch etabliert wird, während diese auf den höheren Beschreibungsebenen (PS-Syntax, Semantik, Pragmatik) fast völlig fehlt. Für diese sind allerdings die *Zuordnungen* zum Teil eingehender spezifiziert, welche die Bildung korrekter formaler Konstrukte und fachsprachlich Ausdrücke für den jeweiligen Untersuchungsbereich regeln.

Jedenfalls erweisen sich die weitgehend kompetenztheoretischen Modellierungen linguistischer Begriffe zunehmend als inadäquat und revisionsbedürftig angesichts der Probleme, welche die mit ihnen etablierten Kategorisierungen zur Charakterisierung performativer Sprachdaten aufwerfen. Regelbasierte Grammatikformalismen, symbolische Repräsentationen, monotone Logiken und deterministische Strukturmodelle abstrakter Einheiten führen—trotz ihrer in formal-theoretisch ausgearbeiteten Notationssystemen geklärten *Zuordnungen*—vermehrt zu Randunschärfen, großen Variationsbreiten und vielfältigen Ambiguitäten bei dem Versuch, derartig konzipierten linguistischen Kategorien die beobachtbaren Erscheinungen sprachlicher Phänomene—ohne eine ausgearbeitete Systematik methodisch konstituierter *Korrespondenzen*—zu subsumieren.

3.2 Theorie und Empirie

Dieser Umstand erschien kaum beunruhigend, wenn die Linguistik mit Hilfe ihrer bisherigen Kategorien zur Entwicklung anwendungsstarker Verfahren in der Lage gewesen wäre, was aber nicht der Fall ist. Gerade im Hinblick auf die heute verfügbaren, sehr großen natürlichsprachlichen Textkorpora erweisen sich selbst diejenigen Modellbildungen als bestenfalls theoretisch motiviert, über die wir dank langjähriger computerlinguistischer Forschung verfügen. Empirisch testbar—und damit im Sinne wissenschaftlicher Hypothesenbildung überprüfbar—sind die Implementationen der bisher entwickel-

⁸So etwa kategoriale Begriffe theoretischer Konstrukte (wie *Phonem*, *Morphem*, *Wort*, *Satz* etc.) zur Bezeichnung abstrakter Konzepte im System der *langue/competence* und deskriptive Begriffe empirischer Größen (wie etwa *Laut*, *Phon*, *Morph*, *Äußerung* etc.) zur Bezeichnung beobachtbarer Erscheinungen (als vermeintlichen Realisationen der Konzepte) in der *parole/performance*.

ten Grammatikformalismen allerdings nicht, solange ihr Versagen bei Anwendung auf jeden nicht fragmentierend vor-ingeschränkten Ausschnitt der Erscheinungsvielfalt performativer Sprachdaten als ein (aus den unterschiedlichsten Gründen) nicht angemessenes Verfahren erklärt wird.

So legt die Divergenz zwischen kategorialer Begriffswelt linguistischen Wissens und der Welt der beobachteten Spracherscheinungen die Vermutung nahe, daß kompetenztheoretisch motivierte, mit kategorialen linguistischen Begriffen operierende Theorien- und Modellbildungen den Gegenstandsbereich *Sprachkenntnis* zwar (grob) strukturieren. Den erfahrbaren Phänomenbereich performativer *Sprachrealität* vermögen diese Kategorien aber nur sehr unvollkommen zu erfassen, weil hierzu eine sie korrigierende empirische Komponente nötig wäre, die bisher (weitgehend) fehlt. Daß zur Überbrückung dieser Divergenz probabilistische Ansätze nur sehr bedingt einsetzbar sind, liegt in deren wahrscheinlichkeitstheoretischer Fundierung begründet. (Daten-)Schätz- und (Hypothesen-)Test-Verfahren setzen gleichermaßen die Kenntnis jener (theoretischen) Wahrscheinlichkeitsverteilungen von Zufallsvariablen voraus, als deren Realisierungen man die (empirischen) Häufigkeiten von Elementen deuten müßte, welche die traditionellen linguistischen Begriffs- und Kategorienbildungen in Sprachkorpora zu identifizieren erlauben [Rie79a]. Beide Konzepte, Zufallsvariable und linguistische Einheit, erweisen sich als—zumindest zeichentheoretisch [Rie95a]—unangemessene Abstraktionen.

3.3 Zeichen und Funktion

Warum sollte eine zeichentheoretische Erweiterung der kognitiven Perspektive gerade für die sprachwissenschaftlichen Disziplinen eine Chance eröffnen, das Wissen (über Sprache und die Welt) zu vertiefen und zu erweitern? Es geschieht sicher nicht dadurch, daß die phänomenal begegnende *Mannigfaltigkeit* der beobachtbaren Objektwelt wieder nur—quasi ARISTOTELISCH—unter dem Aspekt der diese Vielheit vereinheitlichenden *Ordnung* beschrieben und—quasi AUGUSTINISCH—unter (immer wieder neuen Begriffsbildungen von) Typen, Klassen und Kategorien subsummiert wird. Ebenso wenig genügt es, Vielheiten aufgrund von Prinzipien ihrer Geordnetheit (von Einheiten über die Eigenschaften und Merkmale zu deren Beziehungen und Relationen) zu analysieren und in Form von *Strukturen* zu repräsentieren, für die eine in ihren *Zuordnungen* und *Korrespondenzen* wohldefinierte formale Zeichennotation gefunden werden kann. Mit *Semiotisierung*, wie sie in [Rie89] entwickelt wurde, ist vielmehr eine Funktion gemeint, die über einen durch zeichenhafte Repräsentationen vermittelten Zusammenhang läuft, der—nach den Reduktionsformen *Ordnung* und *Struktur*—eine übergreifende Kennzeichnung von Vielheiten durch Bestimmung und Angabe von *Prozessen* erlaubt, und dessen Besonderheit darin zu liegen scheint, daß es ihn überhaupt gibt. Ein Erklärungsanspruch wird (vorerst) nur darin bestehen können, daß dieser Zusammenhang Prozesse betrifft, die in Form von *Prozeduren* modellierbar sind.

Prozeduren stellen eine besondere Notationsform dar [Eik85], [Rie85a], welche von der Zeitlichkeit von Prozessen abstrahieren und diese formal manipulieren läßt, die aber über Implementationen—geeignete Algorithmen, Programmiersprachen und Automaten vorausgesetzt—wieder in die zeitlichen Abläufe von Prozessen überführt werden können. Prozedurale Modellierung kognitiver Prozesse erlaubt so deren Simulation unter kon-

trolliert veränderten Bedingungen mit der quasi empirischen Möglichkeit, beobachtete (oder vermutete) Ergebnisse dieser Prozesse anhand von Resultaten experimenteller Modellläufe zu überprüfen und wechselseitig zu bewerten.

4 Zur kognitiven Perspektive

Im Rahmen der *Kognitionswissenschaften*, deren Erkenntnisinteresse sich auf die Erweiterung der Kenntnisse über das *Wissen* (seine Formen, seine Strukturen, seines Erwerbs, seiner Anwendungen und seiner Realität im Denken) richtet [JW77], wird unter durchaus uneinheitlicher Wissenschaftsauffassungen anhand unterschiedlichster Forschungsgegenstände und mit den verschiedensten Untersuchungsmethoden an der theoretischen wie praktischen Präzisierung dessen gearbeitet, was *Kognition* ausmacht.

4.1 Kognitive Prozesse

Verbindlich scheint ein prozessuales Verständnis von Kognition zu sein. *Kognitive Prozesse* können danach—im weiten Sinn: systemtheoretisch—als adaptive Leistung von informationsverarbeitenden Systemen (*Agenten*) verstanden werden, die ihre Umgebungen in jeweils dem Maße in strukturierte Repräsentationen zu überführen vermögen, wie dies ihre jeweils eigene Strukturiertheit erlaubt [Rie95b] [Rie95c]. In den Kognitionswissenschaften heißen diese Prozesse *mentale Prozesse*. Sie gelten als Bedingung dafür, daß Regularitäten, Ordnungen und Strukturen nicht nur entstehen, sondern auch repräsentiert werden. Sofern es sich dabei um (nicht notwendigerweise auch zeichenhaft-symbolische) *Repräsentationen* von für das Zustandekommen von Erkenntnis notwendigen (Zwischen-)Resultaten handelt, wird angenommen, daß auch ihnen (andere) *Prozesse* zugeordnet sind, die einerseits das Entstehen solcher Repräsentationen, andererseits ihre Verarbeitung kontrollieren [Har82]. Die *Semiotisierung* der kognitionswissenschaftlichen Sicht besteht—verkürzt gesagt—in der Suche nach solchen Prozessen (und ihren Modellen), die durch die Verarbeitung (von schon repräsentierten) Einheiten der einen Stufe die Repräsentationen (von Einheiten) einer anderen Stufe erst schaffen.

4.2 Kognitive Linguistik

In ihren Modellbildungen geht die kognitive Linguistik von formal konstruierbaren, *mentalen Repräsentationen* aus, welche eine Bedingung der Möglichkeit der auf ihnen operierenden *mentalen Prozesse* bilden. Einer durch das Erkenntnisinteresse der theoretischen Linguistik motivierten Perspektive [Cho88], wonach sich die Fragen nach Struktur und Erwerb von Sprachkenntnis (*Kompetenz* als *Sprachwissen*) weitgehend *unabhängig von* und jedenfalls *vor* denen nach ihrer Anwendung und möglichen (materialen) Realisierung (*Performanz* als *Sprachfertigkeit*) stellen, entsprechen auch computerlinguistische Modellbildungen [Hab88] [Usz88], welche vor allem die theoretischen Möglichkeiten und formalen Bedingungen des Sprachwissens zu klären suchen.

Ein *semiotisches* Defizit dieser kognitiv-linguistischen Modellierungen des Sprachwissens kann anhand der Forderungen sichtbar gemacht werden, die als *Berechenbarkeit*,

Berechenbarkeitsebene:	<i>LE</i>	<i>G</i>
Algorithmisierungsebene:	<i>ML</i>	<i>P</i>

Abbildung 1: Schema der *Inkongruenz* von Berechenbarkeit und Algorithmisierung sprachlicher Strukturbeschreibungen (nach [Bie88]).

Algorithmisierbarkeit und *Implementierbarkeit* [Mar82] an prozedurale Modelle kognitiver Prozesse allgemein zu stellen sind. Sie lassen eine eigentümliche Inkompatibilität hervortreten, die in [Bie88] zunächst nur als formales Problem konstatiert wurde. Es stellt sich (Abb. 1) als Inkongruenz der Unterscheidung von sprachlichen Einheiten im *Lexikon (LE)* und den Verkettungsregeln der *Grammatik (G)* auf der Modellebene *berechenbarer Repräsentationen* einerseits und den Einheiten einer Gedächtnisstruktur als *mentalem Lexikon (ML)* und der Abarbeitung von Regeln in *Parsern (P)* auf der Ebene der *algorithmisierbaren Prozeduren* andererseits dar. Diese Inkongruenz benennt eine Algorithmisierungs- bzw. Berechenbarkeitslücke im Modellierungsansatz der kognitiven Linguistik, die deren *semiotisches Defizit* aufdeckt⁹. Daß diese Lücke geschlossen werden könnte bzw. gar nicht erst zu entstehen brauchte, wenn geeignetere (als symbolische und regelbasierte) Repräsentationen sprachlicher Einheiten und ihrer Strukturen gewählt würden, ist eine der Arbeitshypothesen für eine semiotisch-funktionale Modellbildung in der (Computer-)Linguistik.

Das *empirische Defizit* dieser Modellbildungen wird darin faßbar, daß anstelle empirisch-operationaler Kriterien der Überprüfbarkeit von Hypothesen und Theorien für die kognitive Linguistik—trotz ihrer dem realistischen Paradigma verpflichteten Wissenschaftsauffassung—im wesentlichen nur zwei (voneinander nicht unabhängige, zudem mittelbare) Adäquatheitskriterien wie *Lernbarkeit* und *Verarbeitbarkeit* gelten sollen [FHR94], die im Hinblick auf prozedurale Modelle sprachverarbeitender Prozesse nicht nur nicht testbar, geschweige denn entscheidbar sind¹⁰. Sie reduzieren vielmehr jede Überprüfung der solchen Modellen zugrundeliegenden Hypothesen auf die bloße Anerkennung oder Ablehnung ihrer theoretischen Prämissen (hier: symbolische Repräsentation, regelbasierte Verarbeitung) und machen damit eine operationale Überprüfung des empirischen Gehalts der Hypothesen mit dem Ziel ihrer intersubjektiv nachvollziehbaren Bewertung

⁹Diese formale Inkongruenz kann als *semiotisch* begründet deswegen gelten, weil sie besagt, "daß die Regeln und Prinzipien aus *G* auch den *Inhalt* von *ML* kontrollieren, daß der *Parser* aber nicht allen Regeln und Prinzipien von *G* korrespondiert" ([Bie88], S.203).

¹⁰*Lernbarkeit*: "...eine Aussage über die mentale Repräsentation sprachlichen Wissens ist nur dann adäquat, wenn die entsprechende Repräsentation prinzipiell lernbar ist" und *Verarbeitbarkeit*: "...eine vorgeschlagene Repräsentation muß insbesondere durch Prozessoren, etwa einen plausiblen Parser oder Formulator, verarbeitet werden können" ([FHR94], Einleitung). Damit wird *Adäquatheit* der Schnittmenge $R_L \cap R_A$ aller prinzipiell lernbaren Repräsentationen R_L (unabhängig von ihrem Notationsformat) mit allen algorithmisch verarbeitbaren Repräsentationen R_A (sofern es nur plausible Implementationen solcher Algorithmen gibt) zugesprochen. Die Entscheidung, wonach bestimmten Repräsentationen, Modellen, Systemen der Vorzug vor anderen zu geben wäre, wird nicht an ihre (oder der ihnen zugrundeliegenden Hypothesen) Überprüfbarkeit gebunden, sondern als von der "Plausibilität" von Parsern und Formulatoren abhängig erklärt.

(hier: Bestätigen oder Verwerfen) unmöglich.

4.3 Kognitive Sprachverarbeitung

Das übergreifende Paradigma von Kognition als (zu modellierendes) System wissenschaftlicher Prozesse wird auch von der kognitiven Sprachverarbeitung akzeptiert. Sie konfrontiert jedoch kompetenztheoretisch relevante Ergebnisse und Resultate der kognitiven Linguistik mit der performativen Praxis der kommunikativen Produktion und Rezeption natürlichsprachlicher Texte, zu deren Erklärung sie eine *ökologisch* orientierte Kognitionstheorie fordert [RS93]. Diese ökologische Einbettung, die erstmals in [Rie89] und thematisch in [Rie94a] und [Rie96] als *semiotischer* Ansatz informationsverarbeitender Systeme entwickelt wurde, bezieht konsequenterweise die Bedingungen wissenschaftlicher Kommunikation und Modellpluralität als besondere, weil explizierte Form situierter Produktion und Rezeption von Zeichen- und Symbolaggregationen ein. Dem ökologischen Paradigma [Bat79], [MV80] dieser Wissenschaftsauffassung entspricht dabei eine dynamische Konzeption von Modellierung, deren Überprüfbarkeit weitgehend in der methodischen Realisierung kontextuell situierter *Prozeßsimulationen* begründet ist. Diese können als *semiotische* Erklärung für das Entstehen von Strukturen, Ordnungen und Einheiten aus Vielheiten dann gelten, wenn sie—unabhängig von allen übrigen Erklärungsparadigmen—einen durch *Berechenbarkeit* kontrollierbaren, durch *Algorithmisierbarkeit* modellierbaren und durch seine *Prozeduralität* vermittelten Zusammenhang herstellen zwischen Repräsentationen unterschiedlicher Ebenen.

Während das Prinzip der Berechenbarkeit (*computational level*) mögliche Formate, Einheiten und Operationen über die Repräsentationen aller Beschreibungsebenen sprachlicher Phänomene festzulegen erlaubt, sind aus der Menge ihrer möglichen Algorithmisierungen (*algorithmic level*) nur diejenigen *semiotisch* interessant, die durch Verarbeitung der Einheiten einer Repräsentationsebene Einheiten einer anderen Ebene erst konstituieren. Sie bilden offenbar eine Teilklasse der Algorithmen, die in Modellen symbolischer Repräsentationen von Einheiten und der regelbasierten (syntaktischen) Festlegung ihrer Konkatenationen gar nicht vorkommen, sondern überhaupt nur in Modellen mit verteilt repräsentierten Einheiten und nicht-syntaktischen Agglomerationen benötigt werden. Denn während die regelverarbeitenden Algorithmen in den symbolischen Modellen den Bereich der *Zuordnungen* abdecken, vermögen Algorithmen, die auf sub-symbolischen oder verteilten Repräsentationen operieren, offenbar den Bereich der *Korrespondenzen* zu modellieren. Diese *semiotischen* Algorithmen setzen im wesentlichen Einheiten unterschiedlicher Repräsentationsebenen derart zueinander in Beziehung, daß sie (mindestens) eine dieser Ebenen mit ihren Einheiten erst schaffen. Die Rede ist von *emergenten* Strukturen, welche bisher nicht-unterscheidbare Einheiten dadurch unterscheiden lassen, daß sie als Resultate von Prozessen erscheinen, welche die Daten, die sie verarbeiten, systematisch verändern.

5 Zur unscharfen Modellierung

Weitgehend ungeklärt ist bisher, ob—und gegebenenfalls wie—*semiotische* Modellbildungen ein stufenweises Entstehen von Strukturen aus Ordnungen und dieser Ordnungen aus Regularitäten von Vielheiten beobachtbarer Entitäten erklären können. Es kann aber aufgrund vorliegender Untersuchungen vermutet werden, daß diese Prozesse der Identifikation von Regelmäßigkeiten und deren Zusammenfassung in solchen (Zwischen-)Repräsentationen, denen wiederum Eigenschaften von beobachtbaren Entitäten zukommen, für das Entstehen und die Verwendung zeichenhaft-funktionaler Strukturen in natürlich-sprachlichen Systemen verantwortlich, wenn nicht mit ihnen gar identisch sind. Eine fundierte Hypothese dazu ist, daß sie—unabhängig von allen übrigen Erklärungsparadigmen—einen durch *Berechenbarkeit* formal kontrollierbaren, durch *Algorithmisierbarkeit* prozedural modellierbaren und durch ihre *Implementationen* praktisch vermittelten Zusammenhang überprüfbar herstellen zwischen Repräsentationen verschiedener Ebenen, die sie selbst erzeugen.

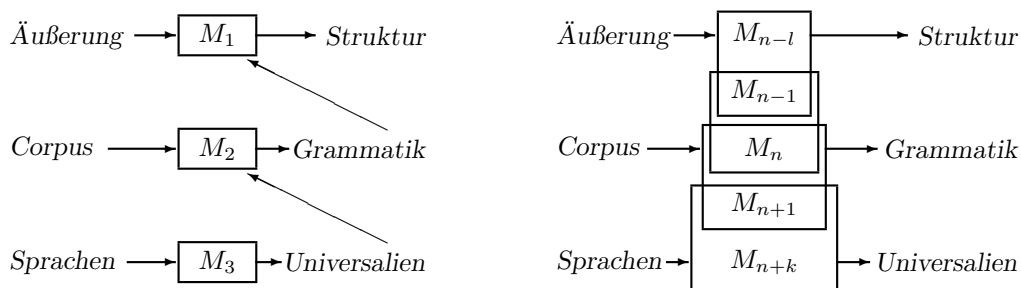


Abbildung 2: Gegenüberstellung der Stufen modular separierter Mechanismen M_1 bis M_3 symbolisch-regelbasierter Modellbildung mit der vielfachen Überdeckung numerisch-prozedural realisierter Vektorräume M_{n-l} bis M_{n+k} in semiotisch-funktionalen Modellierungen einer *fuzzy* Linguistik.

Für kognitive Prozesse des Sprachverstehens bietet sich eine Modellierung in Form von mehrstufigen Verarbeitungs- und Repräsentationsebenen¹¹ an, weil sie auf Gegebenheiten aufsetzen können, die als sprachliche Manifestationen selber schon (Zwischen-)Repräsentationen sind. Es ist zu untersuchen, ob die klassische Dreistufigkeit *linguistischer* Analysemechanismen M_1 bis M_3 nach [Bie81] (Abb. 2 links) nicht besser durch eine ebenfalls mehrstufige, aber vielfach interdependente Überdeckung *semiotischer* Prozeduren M_{n-l} bis M_{n+k} (Abb. 2 rechts) modelliert werden könnte. Anders als kompetenztheoretische Idealisierungen von Einheiten (*Universalien* → *Grammatik* → *Satz*), welche sprachliche Erscheinungsformen auf das Prinzip *symbolisch* repräsentierbarer und *syntaktisch* überprüfbarer Korrektheit hin abstrahieren, wird deshalb versucht werden, in einem performanzorientierten Ansatz zunächst die Zeichen- und Bedeutungs-konstituierenden Leistungen kognitiver Prozesse in Form von Funktionen und Funktionensystemen zu modellieren.

¹¹[Bie81] spricht im Anschluß an [Cho65], [Cho75] von "Kaskaden der Vermittlungsstufen" (S.178)

5.1 Quantitative Verfahren

Die quantitative Beschreibung und numerische Analyse sprachlicher Elemente, Einheiten und Strukturen bietet sich an, wenn es darum geht, Eigenschaften ihrer Verwendung, ihres Gebrauchs und der damit verbundenen Zusammenhänge zu ermitteln, die als (nicht unmittelbar beobachtbare) abgeleitete Funktionen ihres (beobachtbaren) Vorkommens beschrieben werden können. In Verbindung mit den *fuzzy-theoretischen* Möglichkeiten der Modellierung [KGK93] [Nov89] erlauben diese Verfahren die Definition von elastischen Einheiten [Zad75]—den *soft constraints* [Smo89b] in sub-symbolischen Modellen entsprechend—durch numerische Spezifizierungen und erhöhtes Auflösungsvermögen von Zugehörigkeitsgraden verbunden mit größeren Toleranzen der Kategorisierung und der Verarbeitung [Zad94]. *Weiche Kategorien* lassen sich so in ihren Reichweiten (Umfang) über die gleichen Formen numerischer Bestimmung kennzeichnen, wie die Arten der ihnen subsumierbaren Elemente (Inhalt). *Unschärfe (fuzzy) Kategorien* heißen dabei solche abstrakten Zuordnungen, deren (leeren) Strukturen ebenso wie deren mögliche Füllungen als Resultate von Prozessen erscheinen, die in Form von *Prozeduren* dargestellt werden können. Die *prozedurale* Form erlaubt dabei,

- ▷ veränderliche strukturelle Zusammenhänge als *dynamisch* sich verändernde Resultate von Prozessen zu erklären, deren (wiederholter) Ablauf Veränderungen der (metrischen, topologischen) Struktur der Daten, auf denen sie operieren, zur Folge hat;
- ▷ diese Prozesse als Modell *kognitiver* Leistungen zu deuten, wodurch die Elemente (und Elementverbindungen) einer Ebene Strukturen anderer Ebenen zugeordnet werden, die durch diese Zuordnung erst entstehen;
- ▷ diese Zuordnung als prozedurale Explikation einer *Bedeutungskonstitution* zu verstehen, insofern bisher verborgenen (*hidden*) und uninterpretierten Einheiten und Strukturen durch die sie konstituierenden Prozeduren erkennbare Funktionen erst zukommen.

Kategoriale Konstrukte und Begriffsbildungen der strukturellen Linguistik, die immerhin aufgrund operationaler (wenn auch nicht streng algorithmisierbarer) Ansätze gefunden wurden, können hierbei als Leitlinie der Suche dienen, wenn es darum geht, operational definierte, prozedural (re)konstruierende Konzepte zu entwickeln, denen überprüfbare, gegebenenfalls *weiche Kategorien* entsprechen und für eine semiotische Theorie linguistischer Performanz relevant werden könnten.

Die Verbindung eines *semiotisch* fundierten Modells vektorieller Funktionsrepräsentationen linguistischer Einheiten mit einer dem semiotischen Gegenstandsbereich angemessenen mathematischen Analyse dieses Repräsentationsformats eröffnet begründete Aussicht darauf, neuartige Zusammenhänge zwischen den Erscheinungen *performativer* Sprachverwendung und den Prinzipien *kompetenten* Sprachvermögens aufzuweisen, deren quasi empirische Überprüfbarkeit durch die prozedurale Modellierung und experimentelle Simulation Zeichen-konstitutiver Prozesse gewährleistet werden kann.

5.2 Rekonstruktive Modellierung

Für die Modellbildung kann auf ein Grundprinzip sprachlicher Strukturbildung zurückgegriffen werden, dessen Universalität¹² in der spezifischen Form der Einschränkung liegt, welche die in beobachtbaren sprachlichen Regularitäten tatsächlichen realisierten Kombinationen von den theoretisch möglichen Kombinationen dieser sprachlichen Einheiten unterscheidet. Diese als lineare Verkettungsrelationen (*Syntagmatik*) und als selektive Ersetzungsrelationen (*Paradigmatik*) von (eben hierdurch unterschiedenen) linguistischen Einheiten wurde schon von den Begründern der strukturellen Linguistik in ihrer Systematik erkannt und zur Konstitution verschiedener Beschreibungsebenen sprachlicher Erscheinungen und ihrer Kategorisierung (Segment und Klasse) genutzt. Diese Unterscheidung kann durch den fuzzy-theoretischen Ansatz für die semiotische Modellbildung verschärft und numerisch präzisiert werden¹³. Die hierbei zu verarbeitenden primären Eingabedaten sind (eindimensionale) Verteilungen von Elementen, die aufgrund *syntagmatischer* und *paradigmatischer* Restriktionen erst als Einheiten unterschieden werden. Verarbeitungsergebnisse geeigneter Prozeduren werden wiederum als (diesmal: zweidimensionale) Verteilungen (Matrizen) ausgegeben, welche als Relationen, Vektoren oder hochdimensionierte Raumstrukturen interpretiert werden können. Diese *semiotische Räume* werfen in ihrer Topologie wie in ihrer Metrik noch beträchtliche Probleme auf, die bei der Deutung und Interpretation (durch veranschaulichende Transformationen) selbst kleiner Ausschnitte—möglicherweise aber auch wegen dieser Fragmentierung—auftreten.

Einige auf *semantischer* und *morphologischer* Ebene durchgeführte Untersuchungen, die auf der Grundlage schriftsprachlicher Textdaten des Deutschen basieren, haben die (graphischen) Analoga so unterschiedlicher linguistischer Konzepte wie *Wortbedeutung* und *Silbe* als vektorbasierte Strukturierungsleistung von Prozeduren modellieren können, indem sie auf linearen Folgen von Elementen (*Wörter* und *Graphen* bzw. Buchstaben) in natürlichsprachlichen Texten großer Korpora operieren. Die vorliegenden Implementierungen solcher prozeduralen Modellierungen von *weichen* (*fuzzy*) linguistischen Kategorien beruhen dabei auf verteilten Repräsentationen, welche die Zustände eines dynamischen Systems möglicher Graphen-Agglomerationen bzw. Wort-Zusammenhängen partiell festlegen.

Auf zwei sehr unterschiedlichen Analyseebenen—dem *morphologischen Raum* mit (*morphischen*) Vektoren aus Prozeduren auf n-Grammen von Buchstaben [Rie95a] und dem *semantischen Raum* mit (*semischen*) Vektoren aus Prozeduren auf Korrelationen von Wörtern [Rie91a], [Rie91b]—hat sich gezeigt, daß die Fülle der in diesem Repräsentationsformat enthaltenen Informationen nur in dem Maße zugänglich und nutzbar ist, wie die Verfügbarkeit von Prozeduren zu ihrer Überführung in Strukturen, die (potentiell) mathematisch und/oder linguistisch¹⁴ interpretierbar sind.

¹²Es ist keine Sprache dieser Welt bekannt, die nicht den Strukturierungsprinzipien der *syntagmatischen* und *paradigmatischen* Restriktionen unterläge, wengleich wesentliche sprachtypologische Unterschiede auf dem unterschiedlichen Gebrauch beruhen, der in den verschiedenen Sprachen von diesen Prinzipien gemacht wird.

¹³Daß damit kein Widerspruch zur *fuzzy* Modellierungstechnik entsteht, ist in der Konzeption *unscharfer Mengen* [Zad65] begründet, für die die Verbindung von Ausdrücken diskret algebraischer Entitäten mit numerischen Termen mathematischer Kontinua kennzeichnend ist.

¹⁴Als Struktur-analyisierende und -generierende Transformationen tatsächlicher Sprachdaten, welche

5.2.1 Der morphologische Raum

Auf dieser Modellierungsebene wurde bisher Teilaspekte (u.a. [Oer94], [Sch94a], [Zub94], [Wag95]) dessen untersucht, wie—zunächst von Kategorien und Konzepte der linguistischen Theorienbildung ausgehend—sich Maße, Strukturen und Repräsentationen für Bindungskräfte auf der Graphen-(Buchstaben-)ebene entwickeln ließen, welche die syntagmatischen und paradigmatischen Restriktionen von Einheiten zusammenfassen, aus denen diese (und andere neue) im Deutschen gebildet werden (können). Ausgehend von der graphemischen Umsetzung der traditionell lautlichen Einheit *Silbe* und unter Nutzung bisheriger phonetischer, phonemischer und morphophonematischer Begriffsbestimmungen und Theoriebildungen als heuristischer Hilfe bei der Suche nach Regularitäten, die in schriftsprachlichen Daten wie dem Trierer *dpa*-Korpus vorliegen, wurden eine *dynamische* Kontrollstruktur der positionsabhängigen Bindungsneigungen von Graphemen entwickelt. Sie basiert auf der iterativen Agglomeration von Elementen (Buchstaben), welche die syntagmatischen und paradigmatischen Restriktionen kombinatorisch möglicher (aber nicht realisierter) Verbindungen dieser Einheiten zusammenfaßt und in ihrer Gesamtstruktur wiedergibt.

Die Übersicht einiger notationeller und definitorischer Vereinbarungen des Modells betrifft

n-Gramme als *n*-gliedrige Ketten benachbarter Entitäten. Für $n \geq 2$ lassen sie sich als geordnete Paare benachbarter Einheiten (Lettern, Graphen, Zeichenketten, Wortketten, etc.) erfassen, welche die Basis bilden, auf der

Abstraktionen über diesen Einheiten als *weiche* kategoriale Typisierungen (vergleichbar Buchstaben, Graphemen, Morphemen, Silben, Wörtern, etc.) prozedural modellierbar sind. Sie können formal erklärt werden als *dispositionelle Dependenz-Strukturen (DDS)* [Rie85b] [Rie89] von

fuzzy (Teil-)Mengen des jeweiligen Zeichen-Repertoires Z^n verschiedener Dimensionen $n \geq 1$

$$\tilde{X}_n := \{(x, \mu_n(x)) : x \in Z^n\} \subseteq Z^n \times [0, 1], \quad (1)$$

wobei

$$\mu_n : Z^n \rightarrow [0, 1] \quad (2)$$

den Grad der Zugehörigkeit angibt. Die

Zugehörigkeitswerte $\mu_n(x)$ lassen sich für einen gegebenen Korpus aus den Übergangneigungen aller linear verketteten Einheiten wie folgt induktiv berechnen. Für eine *n*-elementige Zeichenkette $x \in Z^n$ gebe $H_n(x)$ die Häufigkeit des Auftretens von x in dem zugrundeliegenden Korpus an. Man setze nun für ein

Konzepte und Techniken der *unscharfen* Modellierung nutzen, gehören diese Untersuchungen in den Bereich der empirisch, (re-)konstruktiv arbeitenden Computerlinguistik, die als *Fuzzy Linguistics* sich derzeit erst abzuzeichnen beginnt.

Bi-Gramm $x = (y, z) \in Z^n$, $y \in Z^{n-1}$, $z \in Z$,

$$\mu_n(x) = \frac{H_n(x)}{H_{n-1}(y)}. \quad (3)$$

Mit $Z = \{z_1, \dots, z_m\}$ erhält man so für jedes $y \in Z^{n-1}$ einen Vektor

$$(\mu_n(y, z_1), \dots, \mu_n(y, z_m))^T \in \mathbb{R}^m. \quad (4)$$

In ihrer Gesamtheit spiegeln diese Vektoren die morphologische Struktur des gegebenen Korpus wider. Auf diese Weise werden

weiche Kategorien konstituiert, die als Systeme von *Fuzzy-Mengen* der beobachtbaren Verkettungs-Regularitäten definiert sind und aus der Bindungsneigung von Einheiten auf der jeweiligen Ebene *elastische Einschränkungen* repräsentieren, welche die Systemstruktur bestimmen.

Die Darstellung der Entwicklung *weicher Kategorien* als *elastische Constraints* (verschiedener Ebenen) wird über ihre formale Darstellung als (mehrstellige) *fuzzy-Relationen* und den diesen entsprechenden numerischen Formaten in Übergangsmatrizen (auch höherer Ordnungen) erleichtert. Hierzu wird der Raum aller theoretisch möglichen Kombinationen nicht nur auf den Raum der faktisch möglichen, sondern weiter auf den der real auch belegten Kombinationen eingeschränkt, den die beobachteten Einheiten in den analysierten Texten des Korpus tatsächlich aufspannen. Die Umfänge der theoretisch (T_n) und faktisch (F_n) möglichen sowie real (in einem exemplarisch zugrundeliegenden Teil des Trierer *dpa*-Korpus) belegten (R_n) Mengen unterschiedlicher Zeichenketten gleicher Längen (*n*-Gramm-*types*) sind in Tab. 1 zusammengestellt.

<i>n</i> - Gramme	$ R_n $ (real belegt)	$ T_n = Z^n $ (theor. möglich)	$100 \cdot \frac{R_n}{T_n}$ Prozent	$F_n = m \cdot R_{n-1} $ (fakt. möglich)	$100 \cdot \frac{R_n}{F_n}$ Prozent
1	31	31	100,000	31	100,000
2	817	961	85,015	961	85,015
3	10.175	29.791	34,154	25.327	40,174
4	54.470	923.521	5,898	315.425	17,268
5	164.045	28.629.151	0,572	1.688.570	9,715
6	357.632	887.503.681	0,040	5.085.395	7,032
7	634.767	27.512.614.111	0,002	11.086.592	5,725
Test-Korpus (Umfang):		3.648.326	(Zeichen)	502.587	(Wörter)

Tabelle 1: Graphen-(Buchstaben-)Kombinatorik und ihre (theoretisch und faktisch) möglichen, sowie die davon real belegten *types* von *n*-Grammen in einer Teilmenge der Daten des Trierer *dpa*-Korpus

Für schriftsprachliche Texte des Deutschen ergäben sich auf Buchstabenebene bei m unterschiedenen Zeichen (Graphen) und einer maximalen Länge n der betrachteten Zeichenfolgen für $m = 31$ und $n = 7$ zwar *theoretisch* die möglichen Kombinationen (Tab. 1, Spalte T_n) oder scharfen Relationen $T_n = Z^n$, nämlich

$$\begin{aligned}
Z = T_1 &= \{x_1 && : x_1 \in Z\} \\
T_2 &= \{(x_1, x_2) && : x_1, x_2 \in Z\} \\
T_3 &= \{(x_1, x_2, x_3) && : x_1, x_2, x_3 \in Z\} \\
&\vdots && \vdots \\
T_{n-1} &= \{(x_1, \dots, x_{n-1}) && : x_1, \dots, x_{n-1} \in Z\} \\
T_n &= \{(x_1, \dots, x_n) && : x_1, \dots, x_n \in Z\}.
\end{aligned}$$

Faktisch erfaßt werden müßten daraus aber nur die auch *praktisch* auf jeder n -ten Ebene noch zu realisieren möglichen Zeichen-Kombinationen $F_n \subseteq F_{n-1} \times Z$ (Tab. 1, Spalte F_n), nämlich

$$\begin{aligned}
Z = F_1 &\subseteq \{x_1 && : x_1 \in Z\} \\
F_2 &\subseteq \{(x_1, x_2) && : x_1 \in F_0, x_2 \in Z\} \\
F_3 &\subseteq \{(x_1, x_2, x_3) && : (x_1, x_2) \in F_1, x_3 \in Z\} \\
&\vdots && \vdots \\
F_{n-1} &\subseteq \{(x_1, \dots, x_{n-1}) && : (x_1, \dots, x_{n-2}) \in F_{n-2}, x_{n-1} \in Z\} \\
F_n &\subseteq \{(x_1, \dots, x_n) && : (x_1, \dots, x_{n-1}) \in F_{n-1}, x_n \in Z\}.
\end{aligned}$$

Die *fuzzy* Modellierung gemäß (3) wird über die relativen Übergangs-Häufigkeiten erreicht, welche die *unscharfen* Relationen für jede der $n-1$ langen Ketten bestimmt. Dabei brauchen die Zugehörigkeitswerte gemäß (2) der betreffenden $(n-1)$ -stelligen Relationen nurmehr für *Bi-Gramme* erhoben und berechnet zu werden. Denn es läßt sich der Umstand ausnutzen, daß in den (jeweils höher-stelligen) Relationen das jeweils erste Glied aller geordneten Paare, die Elemente dieser Relationen sind, aus einer (schon berechneten) *Agglomeration* von Elementen aus jeweils niedriger-stelligen Relationen besteht. Dieses nicht nur formale Prinzip der prozeduralen *Selbstähnlichkeit* von n Agglomerations-schritten erlaubt darüber hinaus—aufgrund der besonderen, nach (2) numerisch präzisierten Bindungsneigungen so konstituierter Einheiten—die systematischen Strukturen unterscheidbarer Elemente auf der niedrigeren, $(n-1)$ -ten Ebene mit den neuen Einheiten auf der nächst höheren, n -ten Ebene in Beziehung zu setzen und als *elastische Einschränkungen* bzw. *weiche* Typisierungen nicht nur zu modellieren, sondern auch so zu verarbeiten (Tab. 1, Spalte R_n) und als Baumgraph zu repräsentieren (Abb. 3).

Auf dieser Modellierungsebene werden durchaus differenziertere Maße (als die hier verwendeten relativen Häufigkeiten) zur Erfassung von Bindungskräften und Übergangsneigungen entwickelt werden können, die der Erfassung und Verarbeitung größerer Agglomerationen als (diskreter) Einheiten in kognitiven Prozessen besser Rechnung tragen. Grundüberlegung dabei ist, daß alle *sprachperformativen* Entitäten ganzheitlich wahrgenommen und verstanden werden und als aus anderen Einheiten aufgebaute Zusammensetzungen in der Regel nur dann analysiert werden, wenn eine Störung (*defizienter Modus*) der Wahrnehmung bzw. des Verständnisses dies erzwingt.

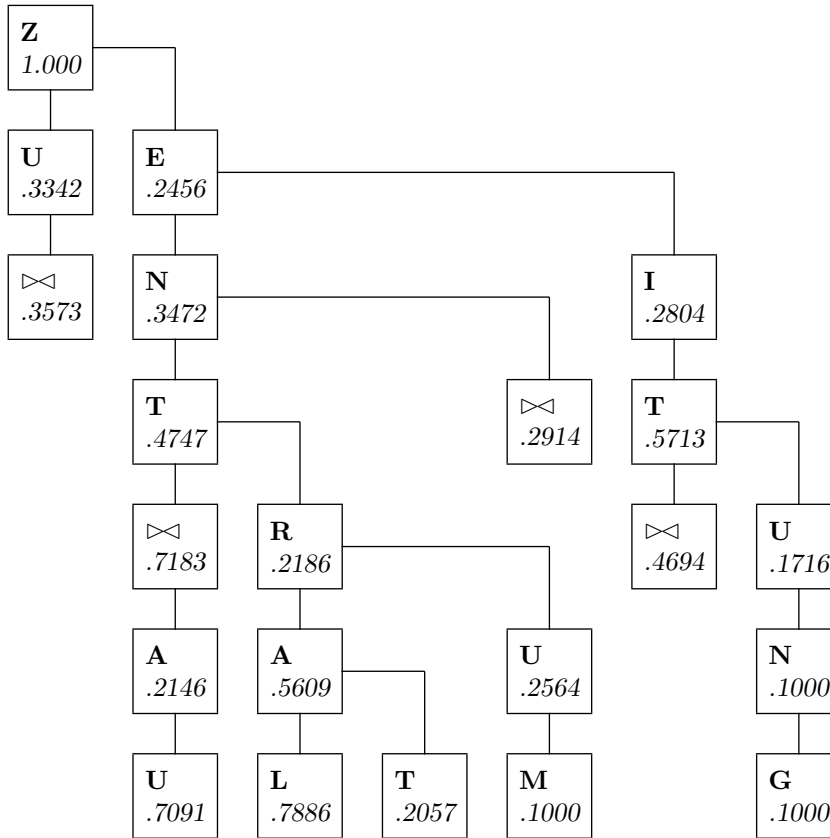


Abbildung 3: Weiche Kategorie $\cong Z$ als Hierarchie abnehmender Übergangneigungen des Buchstaben Z (für 7-Gramme) in Form eines allgemeinen Baumgraphen ermittelt anhand eines Ausschnitts des Trierer *dpa*-Korpus.

5.2.2 Der semantische Raum.

Anknüpfend an schon früher vorgelegte Operationalisierungen dieses Prinzips auf der Ebene lexikalisch-semantischer Einheiten bietet die Theorie der unscharfen (*fuzzy*) Mengen in Verbindung mit statistischen Verfahren der quantitativen Analyse großer Textmengen die Möglichkeit, natürlich-sprachliche *Wortbedeutungen* in Form von Vektoren zu repräsentieren, deren Komponenten als Funktionswerte *syntagmatischer* und *paradigmatischer* Restriktionen von Wortverwendungsweisen berechnet wurden [Rie81a], [Rie82a], [Rie82b]. Die besonderen (formalen und inhaltlichen) Eigenschaften derart modellierter lexikalisch-semantischer Beziehungen haben eine Reihe sehr interessanter Eigenschaften, deren distanzbestimmter Zusammenhang etwa dem grundlegenden Charakteristikum von *Stereotypen* [Rie85d], [Rie85b], [Rie85e] entspricht, wie ebenso deren semantische Verwandtschaften sich als strukturelle Eigenschaft des vieldimensionalen Raumes ergeben und zur Modellierung von *semantischen* (im Unterschied zu *syntaktischen*) Inferenzprozessen [Rie84], [Rie83], [Rie85c] genutzt werden konnten.

Entscheidender als diese auf dem Repräsentationsformat aufsetzenden Operationalisierungen sind die Prozesse, die—als Prozeduren modelliert—diese Repräsentationen erst

liefern und zwar als variable Resultate der textanalytischen Verarbeitung sprachlicher Eingabeketten. Dies geschieht durch die algorithmische Rekonstruktion der *syntagmatischen* und *paradigmatischen* Beschränkungen als formalen Abstraktionen auf (Zwischen-)Repräsentationen, die einen zweistufigen Prozeß der *Emergenz* von Zusammenhängen prozedural modellieren. Diese können als Bedeutungen [Rie91a] deswegen gelten, weil sie als vektorieell repräsentierte und unterschiedene Strukturen bestimmten Zeichen (und Zeichenketten) zugeordnet sind, als deren (Verwendungs-)Regularitäten sie sich definieren [Rie91b], [Thi92], [RBR93], [Rie94b].

Die textanalytischen Verfahren, die hierzu auf der Wortebene angewandt wurden, sind deskriptiv-statistischer Natur und beruhen im wesentlichen auf einer Korrelationsmessung α von (Wort-)Token in einem Korpus *pragmatisch-homogener* Texte sowie auf einer *verteilten* Repräsentation der Verwendungsregularitäten ihrer (Wort-)Typen, deren Unterschiede—über ein Distanzmaß δ numerisch präzisiert—den *semantischen Raum* $\langle S, \zeta \rangle$ mit abstrakten Bedeutungsrepräsentationen (*Bedeutungspunkten*) konstituieren.

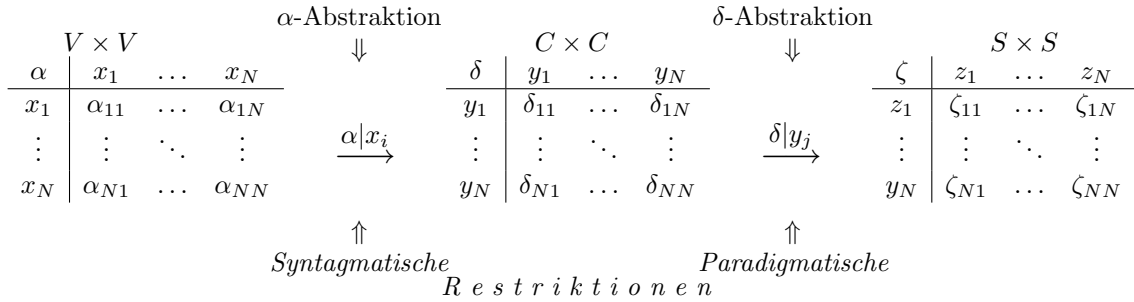


Abbildung 4: Formalisierung (*syntagmatischer/paradigmatischer*) Restriktionen (*constraints*) durch die zweistufige (α - und δ -) Abstraktion von Verwendungsregularitäten x_i und deren Ähnlichkeiten/Unterschieden y_j .

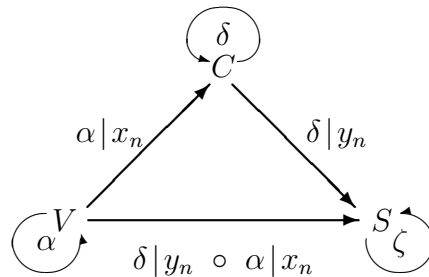


Abbildung 5: Abbildungsrelationen α and δ zwischen den strukturierten Mengen des Vokabulars $x_n \in V$, seiner Verwendungs-Regularitäten $y_n \in C$ und seiner Bedeutungspunkte $z_n \in S$.

Als Resultat des daraus abgeleiteten zweifachen Abstraktionsschritts (*Abb. 4*) läßt sich die Lage und Position jedes Bedeutungspunkts $z \in S$ in der vieldimensionalen me-

trischen Struktur des *semantischen Raum* $\langle S, \zeta \rangle$, auch als Funktion aller Unterschiede (δ - oder Distanzwerte) aller Verwendungsregularitäten (α - oder Korrelationswerte) der in den untersuchten Texten verwendeten Wörter des Vokabulars $x \in V$ deuten (Abb. 5). Als Zuordnung von *Wörtern* zu ihren *Bedeutungspunkten* stellt diese Funktion eine (mögliche) formale Repräsentation und empirisch überprüfbare Operationalisierung der *syntagmatischen* und *paradigmatischen* Restriktionen dar, welche Folgen von Wörtern in Texten pragmatisch homogener Korpora erkennen lassen. Die Visualisierung *unscharfer* Konzepte mithilfe eines Bäume generierenden Algorithmus, der die Bedeutungspunkte nach ihrer Relevanz (*Kriterialität*) für die Bedeutung eines Konzepts (Wurzelknoten) reorganisiert stellt dabei bisher die einzige Möglichkeit dar, um eine anschauliche Vorstellung *relevanter* Ausschnitte des semantischen Raumes zu vermitteln (Abb. 6).

Der systemtheoretische Zusammenhang der α , δ und ζ zugeordneten Analyse- und Repräsentationsalgorithmen kann so als ein prozedurales Modell der *Verstehensfähigkeit* gedeutet werden, wonach ein System mit α -, δ - und ζ -Eigenschaften über Analyse-, Repräsentations- und Kontroll-Mechanismen verfügt, die es ihm—ausschließlich durch Verarbeitung der natürlichsprachlichen Texteingaben—erlauben, strukturelle Information seiner (textuellen) Umgebung zu erkennen, zu verarbeiten und als sein (semantisches) Wissen zu repräsentieren. Daß das System dabei in der Lage ist, aufgrund veränderter (textueller) Umgebungen auch sein eigenes (semantisches) Wissen zu modifizieren, macht die (adaptive) Dynamik dieses Modells aus.

Der Prozeß des *Verstehens* kann—im weiten Sinn: *systemtheoretisch*—als adaptive Leistung von informationsverarbeitenden, kognitiven Systemen gedeutet werden, durch welche sie ihre Umgebungen in strukturierte Repräsentationen überführen, und zwar in jeweils dem Maße, wie ihre eigene Strukturiertheit dies erlaubt. Aufgrund sehr früher Untersuchungen zur Bedeutungskonstitution [Rie78], [Rie79b], [Rie80] und ihrer späteren Wiederaufnahme und Weiterentwicklung [Rie89], [RT89], [Rie91a], [Rie91b] erscheint es plausibel anzunehmen, daß *Bedeutungen* sprachlicher Einheiten als Resultate einer Klasse kognitiver Prozesse verstanden werden können, deren prozedurale Modellierung darin besteht, durch die Verarbeitung von schon repräsentierten Einheiten (der einen Stufe) die Repräsentationen von Einheiten (einer anderen Stufe) erst zu schaffen [RT93], [Rie94b].¹⁵

Anders als in den auf die Dynamik *konnektionistischer* Systeme konzentrierten Charakterisierungen der Entwicklung und Veränderbarkeit von Zustandsräumen, die auch durch Differenzen- und Differentialgleichungen beschrieben werden, richtet sich die *semiotische* Modellierung auf Prozesse, die quasi ontische Differenzierungen unterschiedlicher Seinsweisen von Entitäten (Zeichen und Zeichenagglomerationen, sowie deren Funktionen) betreffen.

Da für die prozedurale Modellierung der *syntagmatischen* und *paradigmatischen* Restriktionen in natürlichsprachlichen Texten, deren mehrstufige Verbindung das *semioti-*

¹⁵Die in dieser Sprechweise suggerierte Abfolge primärer **vor** sekundären Repräsentationen ist der Prozeduralität dieser Phänomens eigentlich unangemessen; die Unterscheidung sowohl der lokalen Repräsentationsebenen als auch ihrer zeitlichen Relate sind Kennzeichen eher der re-konstruktiven *Modellierung* als des zu modellierenden kognitiven *Prozesses*, dessen formale Spezifikation als *Prozedur* von beidem zu abstrahieren erlaubt.

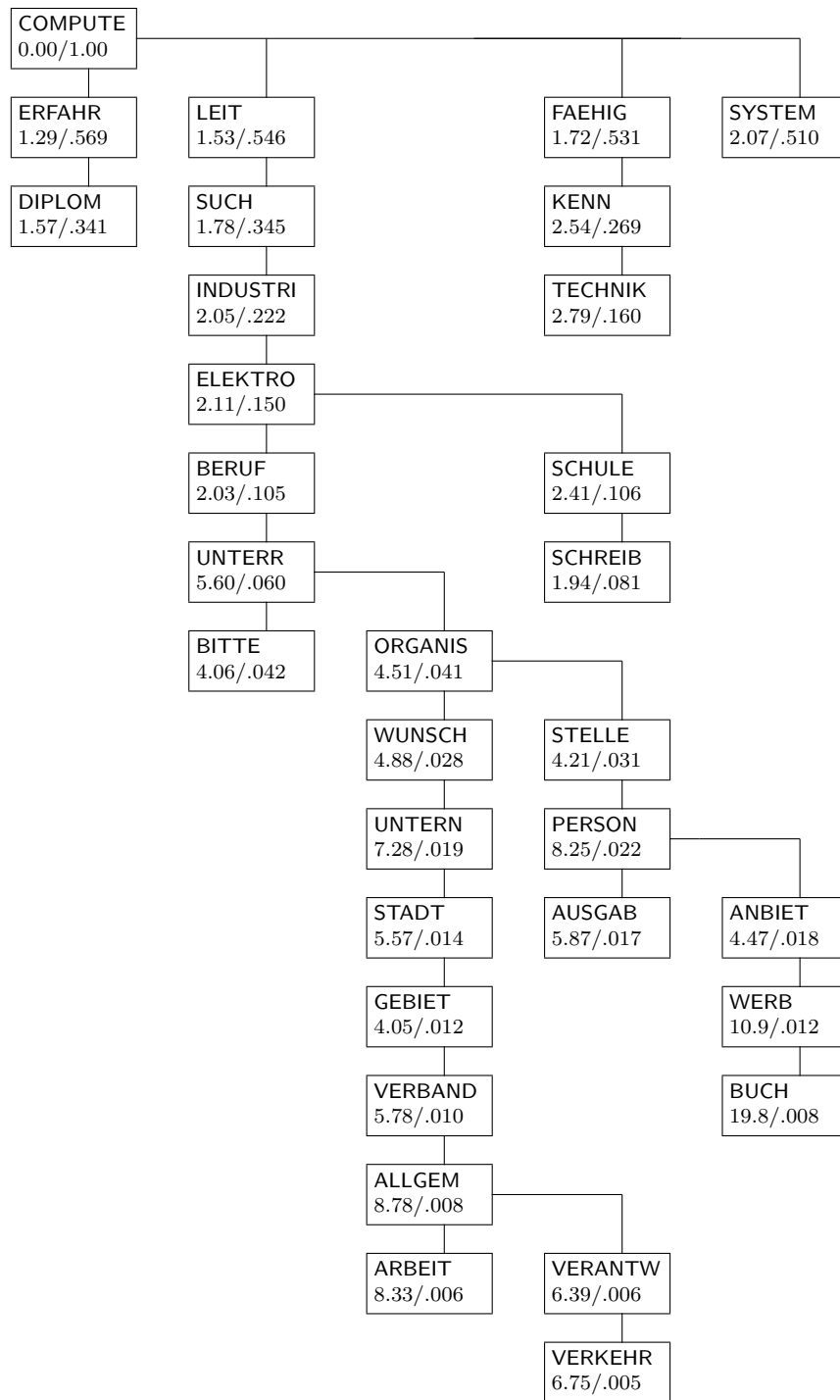


Abbildung 6: DDS-Baumgraph semantischer Dispositionen des Bedeutungspunktes COMPUTER in $\langle S, \zeta \rangle$ mit Punktdistanzen (erster Wert) und Bedeutungskriterialitäten (zweiter Wert) ermittelt anhand eines Teilkorpus deutscher Zeitungstexte (DIE WELT, 1964).

Neuere Linguistik:	(Basis: Grammatikalitätsurteile)
E-Interesse \implies	Theorie der Kompetenz
F-Gegenstand \implies	Sprachvermögen
U-Methoden \implies	Beschreibung, Analyse, Kategorisierung, Formalisierung, etc.
Computerlinguistik:	(Basis: Berechenbarkeit)
E-Interesse \implies	Prinzipien sprachlicher Kognition
F-Gegenstand \implies	Repräsentationen und Prozesse
U-Methoden \implies	Operationalisierung, Prozeduralisierung, Algorithmisierung, Implementierung, etc.
Quantitative Linguistik:	(Basis: Zählbarkeit)
E-Interesse \implies	Gesetze der Sprache(erscheinungen)
F-Gegenstand \implies	Sprachmaterial
U-Methoden \implies	diskrete Mathematik, Statistik, Analysis, Systemtheorie, etc.
Fuzzy Linguistik:	(Basis: Emergenz)
E-Interesse \implies	Theorie der Performanz
F-Gegenstand \implies	situierte (sprachliche) Kommunikation
U-Methoden \implies	Fuzzy Modellierung (auch von Methoden der NL, CL und QL)

Tabelle 2: Tabellarische Übersicht unterschiedlicher *Erkenntnisinteressen*, *Forschungsgegenstände* und *Untersuchungsmethoden* von traditioneller Linguistik, Computerlinguistik, Quantitativer Linguistik und Fuzzy Linguistik.

sche Modell der Zustandsvektoren im *semantischen Raum* liefert, das Auflösungsvermögen entscheidend ist, mit dem die Erfassung primärer sprachlicher Daten erfolgt, wird nach den Erfahrungen mit den *internen*, auf der textuellen Strukturiertheit von Korpora beruhenden Maßen des modifizierten Korrelationskoeffizienten α [Rie81b], [Rie89] und der Kollokationsgewichtung *conf* [RT89], [Thi92] vor allem zu untersuchen sein, inwieweit *externe*, auf ganze Korpora aufsetzende, aber die Daten über variable Fenstergrößen und Schrittweiten erfassende Maße etwa der Verbundinformation (*mutual information*) [Fan61], [CH90], [Sma91] bessere Resultate erzielt werden. Die Hypothese ist, daß die internen Maße, die wegen der bisher nur verfügbaren kleineren Textkorpora ($\leq 10^4$ Worttokens) die notwendige Datenkonzentration durch Lemmatisierung erreichen, den externen Maßen weitgehend gleichwertig sind, obwohl diese ohne Lemmatisierung auskommen, dies aber durch extrem große Textkorpora ($\geq 10^7$ Worttokens) ausgleichen. In jedem Fall wird zu studieren sein, ob (und wenn ja, wie) ein externes Maß durch Kombination mit Parametern der internen Korpusstrukturierung (Textlängen, n -Gramme) zu optimieren ist mit dem Ziel, solche syntagmatischen und paradigmatischen Strukturmerkmale zu erfassen, die möglichst unabhängig von den durch ihre prozedurale Modellierung konsti-

tuierten Repräsentationsebenen sind.

6 Zusammenfassung

Zieht man zur Charakterisierung der beteiligten linguistischen Wissenschaftszweige wiederum die Bestimmungsstücke ihres jeweils dominierenden *Erkenntnisinteresses*, des dabei isolierten *Forschungsgegenstands* und der hierzu verwendeten *Untersuchungsmethoden* heran, so läßt sich—mit allem Vorbehalt, den schematische Verkürzungen dieser Art (Tab. 2) nahelegen sollten—folgende Übersicht der sie unterscheidenden Positionen geben. Die hierzu versuchsweise auch genannten erkenntnistheoretisch wie forschungspraktisch wirksamen *Basen* der verschiedenen Linguistiken lassen deren Theorie- und Modellbildung eher einordnen und die damit verbundenen Erklärungsansprüche besser abschätzen:

Neuere Linguistik: ihr *Erkenntnisinteresse* richtet sich auf eine Theorie der *Kompetenz* (*Universalgrammatik*). Ihren *Forschungsgegenstand* bildet daher das Sprachvermögen, das mithilfe der *Untersuchungsmethoden* von Introspektion, analysierender Beschreibung und Kategorisierung, formaler (regelbasierter) Repräsentation, etc. studiert wird. Als *Basis* dieses Ansatzes kann das Postulat natürlichsprachlicher Korrektheit gelten, die über Grammatikalitätsurteile scheint zugänglich gemacht werden zu können.

Computerlinguistik: ihr *Erkenntnisinteresse* gilt den Prinzipien der Kognition sofern sie das Sprachverstehen betreffen. Ihren *Forschungsgegenstand* bilden mentale Repräsentationen und Prozesse, die formal zu modellieren sind. Als *Untersuchungsmethoden* stehen Operationalisierung, Prozeduralisierung, und Algorithmisierung im Vordergrund. Die *Basis* dieses Ansatzes bildet das automatentheoretisch begründete Berechenbarkeitspostulat.

Quantitative Linguistik: ihr *Erkenntnisinteresse* richtet sich auf die Sprachgesetze, d.h. auf (universell gültige) Prinzipien von Spracherscheinung. Den *Forschungsgegenstand* bilden daher Sprachmaterialien nahezu jeder Art, die mit *Untersuchungsmethoden* der Statistik, der mathematischen Analysis, der Systemtheorie, etc. analysiert und charakterisiert werden. *Basis* dieses Angangs ist die diskretisierende Kategorisierung sprachlicher Entitäten und damit ihre Abzählbarkeit.

Fuzzy Linguistik: ihr *Erkenntnisinteresse* ist ein integratives, das sich auf eine Theorie der *Performanz* richtet. Dieser liegt aber nicht primär ein Vermögen zur Produktion *korrekter* Sätze zugrunde, sondern die kommunikative Kompetenz des sinnvollen Gebrauchs pragmatisch-funktionaler, d.h. *bedeutsamer* sprachlicher Äußerungen. Den *Untersuchungsgegenstand* bilden (verbale/schriftliche) Zeugnisse situierter sprachlicher Kommunikation, wobei die *Untersuchungsmethoden* alle Techniken der unscharfen (*fuzzy*) Modellierung umfassen, also auch solcher Verfahren, die von der *Neueren Linguistik*, *Computerlinguistik* und *Quantitative Linguistik* eingesetzt werden. *Basis* dieses Ansatzes bildet eine Hypothese, wonach semiotische Prozesse die Repräsentationen, auf denen sie operieren, selbst liefern. Modularität,

Rekursivität, und Emergenz sind daher Grundlagen der Modellierung *semiotischer, kognitiver, informationsverarbeitender Systeme (SCIPS)*.

Literatur

- [Arm93] S. Armstrong-Warwick (Hrsg.) (1993): *Using Large Corpora I & II*. [Special Issues: *Computational Linguistics*, 19:1 & 2], Boston, MA, (MIT Press for ACL).
- [Bat79] G. Bateson (1979): *Mind and Nature: A Necessary Unit*. NewYork, (Dutton).
- [BC93] T. Briscoe, J. Carroll (1993): Generalized Probabilistic LR Parsing of Natural Language (Corpora) with Unification-Based Grammars. *Computational Linguistics*, 1993(1), S. 25–59.
- [BDDM93] P. Brown, S. DellaPietra, V. DellaPietra, R. Mercer (1993): The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 1993(2), S. 263–311.
- [BG93] H. Bandemer, S. Gottwald (1993): *Einführung in die Fuzzy-Methoden*. 4.überarb.erweit. Aufl., Berlin, (Akademie Verlag).
- [Bie81] M. Bierwisch (1981): Logische und psychologische Determinanten der Struktur natürlicher Sprachen. In: J. Scharf (Hrsg.), *Naturwissenschaftliche Linguistik*. Halle, (J.A. Barth), S. 176–187; 841–851.
- [Bie88] M. Bierwisch (1988): Auspizien einer kognitiven Orientierung der Computerlinguistik. In: I. Batori, U. Hahn, M. Pinkal, W. Wahlster (Hrsg.), *Computerlinguistik und ihre theoretischen Grundlagen*, DGfS, GI, GLDV, Berlin / Heidelberg / NewYork, (Springer), S. 198–203.
- [Bro88] P. E. Brown (1988): A statistical approach to language translation. In: D. Vargha (Hrsg.), *COLING-88 Budapest*, J.v.Neumann Society of Computing Science, Budapest, (ICCL), S. 71–76.
- [BS79] H. Bergenholtz, B. H. Schaefer (Hrsg.) (1979): *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*. [Monographien Linguistik und Kommunikationswissenschaft 39], Königstein, Ts., (Scriptor).
- [Bun92] H. Bunke (Hrsg.) (1992): *Advances in Structural and Syntactic Pattern Recognition*. Singapore, (World Scientific).
- [CG91] G. A. Carpenter, S. Grossberg (Hrsg.) (1991): *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA, (MIT Press).
- [CH90] K. W. Church, P. Hanks (1990): Word association norms, mutual information, and lexicography. *Computational Linguistics*, 1990(1), S. 22–29.
- [Cho65] N. Chomsky (1965): *Aspects of the Theory of Syntax*. Cambridge, MA, (MIT).
- [Cho75] N. Chomsky (1975): *The Logical Structure of Linguistic Theory*. NewYork, (Plenum).

- [Cho88] N. Chomsky (1988): *Language and Problems of Knowledge. The Managua Lectures*. [Current Studies in Linguistics 16], 2nd. Aufl., Cambridge, MA / London, (MIT Press).
- [Chu88] K. Church (1988): A stochastic parts program and noun phrase parser for unrestricted text. In: *Second Conference on Applied Natural Language Processing*, Austin, TX, (ACL), S. 136–143.
- [CM93] K. W. Church, R. L. Mercer (1993): Introduction to the Special Issue on Computational Linguistics Using Large Corpora. *Computational Linguistics*, 1993(1), S. 1–24.
- [Eik85] H. Eikmeyer (1985): Prozedurale Semantik. In: B. Rieger (Hrsg.), *Dynamik der Bedeutungskonstitution*. Hamburg, (Buske), S. 31–46.
- [Fan61] R. Fano (1961): *Transmission of Information: A Statistical Theory of Communication*. Cambridge, MA, (MIT Press).
- [FHR94] S. W. Felix, C. Habel, G. Rickheit (Hrsg.) (1994): *Kognitive Linguistik. Repräsentation und Prozesse*. [Psycholinguistische Studien], Opladen, (Westdeutscher Verlag).
- [Fin79] P. Finke (1979): *Grundlagen einer linguistischen Theorie. Empirie und Begründung in der Sprachwissenschaft*. [Wissenschaftstheorie und Philosophie 15], Braunschweig/ Wiesbaden, (Vieweg).
- [FP88] J. Fodor, Z. Pylyshyn (1988): Connectionism and cognitive architecture: A critical analysis. *Cognition*, 1988(28), S. 3–71.
- [Fux54] W. Fucks (1954): On Nahordnung and Fernordnung in samples of literary texts. *Biometrika*, 1954(2), S. 116–132.
- [Hab88] C. Habel (1988): Kognitionswissenschaft als Grundlage der Computerlinguistik. In: I. Batori, U. Hahn, M. Pinkal, W. Wahlster (Hrsg.), *Computerlinguistik und ihre theoretischen Grundlagen*, DGfS, GI, GLDV, Berlin / Heidelberg / NewYork, (Springer), S. 204–209.
- [Har82] S. Harnad (1982): A Unifying Theme for the Cognitive Science. In: T. Simon, R. Scholes (Hrsg.), *Language, Mind, and Brain*. Hillsdale, NJ, (Erlbaum), S. 1–12.
- [Hin92] G. Hinton (Hrsg.) (1992): *Connectionist Symbol Processing*. Cambridge, MA, (MIT Press).
- [HR93] D. Hindle, M. Rooth (1993): Structural Ambiguity and Lexical Relations. *Computational Linguistics*, 1993(1), S. 103–120.
- [JW77] P. Johnson-Laird, P. Wason (Hrsg.) (1977): *Thinking. Readings in Cognitive Science*. [Open University Set Book], Cambridge / London / NewYork, (Cambridge UP).
- [KGK93] R. Kruse, J. Gebhardt, F. Klawonn (1993): *Fuzzy Systeme*. [Leitfäden und Monographien der Informatik], 1.. Aufl., Stuttgart, (B.G.Teubner).

- [KR93] R. Köhler, B. Rieger (Hrsg.) (1993): *Contributions to Quantitative Linguistics. Proceedings of the 1st Quantitative Linguistics Conference – QUALICO-91*. Dordrecht, (Kluwer Academic Publishers).
- [Kuh62] T. S. Kuhn (1962): *The Structure of Scientific Revolution*. [Intern. Encyclopedia of Unified Science 2/2], Chicago/ London, (Chicago UP).
- [Mar82] D. Marr (1982): *Vision*. . Aufl., SanFrancisco, (Freeman).
- [MV80] H. Maturana, F. Varela (1980): *Autopoiesis and Cognition*. [Boston Studies in the Philosophy of Science 42], Dordrecht/ Boston/ London, (Reidel).
- [NCCH89] L. Nadel, L. Cooper, P. Culicover, R. Harnish (Hrsg.) (1989): *Neural Connections, Mental Computation*. Cambridge, MA, (MIT Press).
- [Nov89] V. Novák (1989): *Fuzzy Sets and their Applications*. Bristol/ Philadelphia, (A. Hilger).
- [Oer94] B. Oerder (1994): *Zum Problem linguistischer Kategorien: Untersuchungen der Silbenkonstitution im Deutschen und ihre unscharfe Modellierung*. Magister-Arbeit, Linguistische Datenverarbeitung/ Computerlinguistik, FB II: Universität Trier.
- [PR70] H. Pilch, H. H. Richer (Hrsg.) (1970): *Theorie und Empirie in der Sprachwissenschaft*. Basel / München / Paris, (Karger).
- [PSFS89] R. Pfeiffer, Z. Schreter, F. Fogelman-Soulié, L. Steels (Hrsg.) (1989): *Connectionism in Perspective..* Amsterdam, (North Holland).
- [RBR93] B. Rieger, P. Badry, M. Reichert (1993): Bedeutungserwerb mit referentieller Kontrolle. Ein System experimentell überprüfbarer Bedeutungskonstitution. In: H. P. Pütz, J. H. Haller (Hrsg.), *Sprachtechnologie: Methoden, Werkzeuge, Perspektiven*. Hildesheim, (Georg Olms), S. 185–245.
- [Rie72] B. Rieger (1972): Warum mengenorientierte Textwissenschaft? Zur Begründung der Statistik als Methode. *Zeitschrift für Literaturwissenschaft und Linguistik*, 1972(8), S. 11–28.
- [Rie77] B. Rieger (1977): Bedeutungskonstitution. Einige Bemerkungen zur semiotischen Problematik eines linguistischen Problems. *Zeitschrift für Literaturwissenschaft und Linguistik*, 1977(27/28), S. 55–68.
- [Rie78] B. B. Rieger (1978): Feasible Fuzzy Semantics. In: K. Heggstad (Hrsg.), *COLING-78 7th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, Bergen, (ICCL), S. 41–43.
- [Rie79a] B. Rieger (1979): Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung. In: H. Bergenholtz, B. Schaefer (Hrsg.), *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*. Königstein, (Scriptor), S. 52–70.
- [Rie79b] B. B. Rieger (1979): Fuzzy Structural Semantics. On a generative model of vague natural language meaning. In: R. Trappl, P. Hanika, F. Pichler (Hrsg.), *Progress in Cybernetics and Systems Research*. Washington/ NewYork/ London, (Wiley & Sons), S. 495–503.

- [Rie80] B. B. Rieger (1980): Fuzzy Word Meaning Analysis and Representation in Linguistic Semantics. An empirical approach to the reconstruction of lexical meanings in East- and West-German newspaper texts. In: M. Nagao, K. Fuchi (Hrsg.), *COLING-80 8th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, Tokyo, (ICCL), S. 76–84.
- [Rie81a] B. Rieger (1981): Unscharfe Semantik natürlicher Sprache. Zum Problem der Repräsentation und Analyse vager Wortbedeutungen. In: J. Scharff (Hrsg.), *Naturwissenschaftliche Linguistik. Leopoldina Symposium 1976*. Halle, (J. Ambrosius Barth), S. 251–276.
- [Rie81b] B. B. Rieger (1981): Feasible Fuzzy Semantics. On some problems of how to handle word meaning empirically. In: H. Eikmeyer, H. Rieser (Hrsg.), *Words, Worlds, and Contexts. New Approaches in Word Semantics*. Berlin/ NewYork, (de Gruyter), S. 193–209.
- [Rie82a] B. B. Rieger (1982): Fuzzy Representation Systems in Linguistic Semantics. In: R. Trappl, N. Findler, W. Horn (Hrsg.), *Progress in Cybernetics and Systems Research*. Washington/ NewYork/ London, (McGraw-Hill Intern.), S. 249–256.
- [Rie82b] B. B. Rieger (1982): Procedural Meaning Representation. An empirical approach to word semantics and analogical inferencing. In: J. Horecky (Hrsg.), *COLING-82 Proceedings 9th International Conference on Computational Linguistics*. Amsterdam/ NewYork, (North Holland), S. 319–324.
- [Rie83] B. B. Rieger (1983): Generating Dependency Structures of Fuzzy Word Meanings in Semantic Space. In: S. Hattori, K. Iounu (Hrsg.), *Proceedings of the XIIIth International Congress of Linguists*, Comité International Permanent des Linguistes, Tokyo, (CIPL), S. 543–548.
- [Rie84] B. B. Rieger (1984): Semantic Relevance and Aspect Dependency in a Given Subject Domain. In: D. Walker (Hrsg.), *COLING-84 10th International Conference on Computational Linguistics*, International Committee of Computational Linguistics, Stanford, (ACL-ICCL), S. 298–301.
- [Rie85a] B. Rieger (1985): Einleitung. In: B. Rieger (Hrsg.), *Dynamik in der Bedeutungskonstitution*. Hamburg, (Buske), S. 1–17.
- [Rie85b] B. Rieger (1985): Semantische Dispositionen. Prozedurale Wissenstrukturen mit stereotypisch repräsentierten Wortbedeutungen. In: B. Rieger (Hrsg.), *Dynamik in der Bedeutungskonstitution*. Hamburg, (Buske), S. 163–228.
- [Rie85c] B. B. Rieger (1985): Inducing a Relevance Relation in a Distance-like Data Structure of Fuzzy Word Meaning Representations. In: R. Allen (Hrsg.), *Data Bases in the Humanities and Social Sciences. 4th International Conference on Data Bases in the Humanities and Social Sciences*, ICDBHSS/83, Osprey, FL, (Paradigm Press), S. 374–386.
- [Rie85d] B. B. Rieger (1985): Lexical Relevance and Semantic Disposition. On stereotype word meaning representation in procedural semantics. In: G. Hoppenbrouwes, P. Seuren, T. Weijters (Hrsg.), *Meaning and the Lexicon*. Dordrecht, (Foris Publications), S. 387–400.

- [Rie85e] B. B. Rieger (1985): Stereotype representation and dynamic structuring of fuzzy word meanings for contents-driven semantic processing. In: J. Agrawal, P. Zunde (Hrsg.), *Empirical Foundations of Information and Software Science*. NewYork/London, (Plenum Press), S. 273–291.
- [Rie88] B. Rieger (1988): Theoretische Grundlagen der Computerlinguistik: eine Paneldiskussion mit M. Bierwisch, C. Habel, B. Rieger, H. Uszkoreit und W. Wahlster. In: I. Bátori, U. Hahn, M. Pinkal, W. Wahlster (Hrsg.), *Computerlinguistik und ihre theoretischen Grundlagen*. Berlin/ Heidelberg/ NewYork, (Springer), S. 192–218.
- [Rie89] B. Rieger (1989): *Unscharfe Semantik. Die empirische Analyse, quantitative Beschreibung, formale Repräsentation und prozedurale Modellierung vager Wortbedeutungen in Texten*. Frankfurt a.Main/ Bern/ Paris, (Peter Lang).
- [Rie91a] B. B. Rieger (1991): Distributed Semantic Representation of Word Meanings. In: J. D. Becker, I. Eisele, F. W. Mündemann (Hrsg.), *Parallelism, Learning, Evolution. Evolutionary Models and Strategies, WOPPLOT-89*. Berlin/ Heidelberg/ NewYork, (Springer), S. 243–273.
- [Rie91b] B. B. Rieger (1991): *On Distributed Representation in Word Semantics*. Berkeley, CA, International Computer Science Institute (ICSI), Forschungsbericht TR-91-012.
- [Rie94a] B. Rieger (1994): *Situationssemantik und Computerlinguistik unterwegs zu einer prozeduralen Semiotik? Eine ökologische Perspektive aus Sicht der Theorie informationsverarbeitender Systeme*. Universität Trier, Linguistische Datenverarbeitung & Computerlinguistik, Fachbereich II, Forschungsbericht TR-04-1994.
- [Rie94b] B. B. Rieger (1994): Fuzzy Computational Semantics. In: H. Zimmermann (Hrsg.), *Fuzzy Systems. Proceedings of the Japanese-German-Center Symposium*. Berlin, (JGCB), S. 197–217.
- [Rie95a] B. Rieger (1995): Fuzzy Modellierung linguistischer Kategorien. In: W. Feldmann, H. Hinrichs (Hrsg.), *Text und Lexikon*. Tübingen, (Niemeyer), S. [im Druck].
- [Rie95b] B. B. Rieger (1995): Meaning Acquisition by SCIPS. In: B. M. Ayyub (Hrsg.), *ISUMA-NAFIPS-95*, IEEE-Transactions: Joint Intern. Conf. on Uncertainty Modeling and Analysis, North American Fuzzy Information Processing Society, Los Alamitos, CA, (IEEE Computer Society Press), S. 390–395.
- [Rie95c] B. B. Rieger (1995): Situations, Language Games, and Semiotic Agents. In: J. Albus, A. Meystel, D. Pospelov (Hrsg.), *Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems*, 10th IEEE ISIC Workshop, Bala Cynwyd, PA, (AdRem Inc.), S. 130–138.
- [Rie96] B. B. Rieger (1996): Situation Semantics and Computational Linguistics: towards Informational Ecology. A semiotic perspective for cognitive information processing systems. In: K. Kornwachs, K. Jacoby (Hrsg.), *Information. New Questions to a Multidisciplinary Concept*, E.W.-Heraeus-Stiftung, Deutsche Physikalische Gesellschaft, Deutsche Gesellschaft für Systemforschung, Berlin, (Akademie), S. 285–315.

- [RS93] G. Rickheit, H. Strohner (1993): *Grundlagen der kognitiven Sprachverarbeitung*. [UTB 1735], Tübingen / Basel, (Francke).
- [RT89] B. B. Rieger, C. Thiopoulos (1989): Situations, Topoi, and Dispositions. On the phenomenological modelling of meaning. In: J. Retti, K. Leidlmaier (Hrsg.), *5.Österreichische Artificial-Intelligence-Tagung, Innsbruck-Igls*. Berlin/ Heidelberg/ NewYork, (Springer), S. 365–375.
- [RT93] B. B. Rieger, C. Thiopoulos (1993): Semiotic Dynamics: a self-organizing lexical system in hypertext. In: R. Köhler, B. Rieger (Hrsg.), *Contributions to Quantitative Linguistics. Proceedings of the 1st Quantitative Linguistics Conference – QUALICO-91*. Dordrecht, (Kluwer Academic Publishers), S. 67–78.
- [Sch90] H. Schnelle (1990): Die Herausforderung des Konnektionismus in der kognitiven Linguistik. In: H. Friemel, G. Müller-Schönberger, A. Schütt (Hrsg.), *Forum-90: Wissenschaft und Technik*. Berlin/ Heidelberg/ NewYork, (Springer), S. 42–58.
- [Sch84] K. Schmucker (1984): *Fuzzy Sets, Natural Language Computations and Risk Analysis*. Rockville, AL, (Computer Science Press).
- [Sch94a] R. Schneider (1994): *Zur empirischen Rekonstruktion der linguistischen Einheit 'Wort' als unscharfer linguistischer Kategorie*. Magister-Arbeit, Linguistische Datenverarbeitung/ Computerlinguistik, FB II: Universität Trier.
- [Sch94b] H. Schnelle (1994): Welcher Sprachwissenschaft auf der Spur? Plädoyer für größere Klarheit und Toleranz. *Zeitschr.f.Sprachwissenschaft*, 1994(1), S. 110–120.
- [Sma91] F. Smadja (1991): *Retrieving collocational knowledge from textual corpora. An application: Language generation*. PhD. Diss. CS Dept. Columbia, NewYork.
- [Smo88] P. Smolensky (1988): On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 1988(11), S. 1–74.
- [Smo89a] P. Smolensky (1989): Connectionism and Constituent Structure. In: R. E. Pfeiffer (Hrsg.), *Connectionism in Perspective*. Amsterdam/ NewYork, (North Holland & Elsevier), S. 3–24.
- [Smo89b] P. Smolensky (1989): Connectionist Modelling: Neural Computation / Mental Connections. In: L. Nadel, L. Cooper, P. Culicover, R. Harnish (Hrsg.), *Neural Connections, Mental Computation*. Cambridge, MA / London, (MIT Press), S. 49–68.
- [Thi92] C. Thiopoulos (1992): *Semiosis und Topoi*. (Diss. Universität Trier), Pfaffenweiler.
- [Usz88] H. Uszkoreit (1988): Computerlinguistik, Linguistik und künstliche Intelligenz. In: I. Batori, U. Hahn, M. Pinkal, W. Wahlster (Hrsg.), *Computerlinguistik und ihre theoretischen Grundlagen*, DGfS, GI, GLDV, Berlin / Heidelberg / NewYork, (Springer), S. 210–214.
- [Wag95] R. Wagner (1995): *Harris'sche Konstitutions- und Segmentierungsalgorithmen. Eine Überprüfung und Umsetzung für das Deutsche*. Magister-Arbeit, Linguistische Datenverarbeitung/ Computerlinguistik, FB II: Universität Trier.

- [Wun76] D. H. Wunderlich (Hrsg.) (1976): *Wissenschaftstheorie der Linguistik*. Kronberg, Ts., (Athenäum).
- [Zad65] L. A. Zadeh (1965): Fuzzy Sets. *Information and Control*, (8), S. 338–353.
- [Zad75] L. Zadeh (Hrsg.) (1975): *Fuzzy Sets and their Application to Cognitive and Decision Processes*. New York/ San Francisco, (Academic Press).
- [Zad94] L. A. Zadeh (1994): Fuzzy Logic, Neural Networks, and Soft Computing. *Comm. of the ACM*, 37(3), S. 77–84.
- [Zub94] D. Zubak (1994): *Zwischen Graphemen und Silben: vom distributionellen Muster zur fuzzy Modellierung*. Magister-Arbeit, Linguistische Datenverarbeitung/ Computerlinguistik, FB II: Universität Trier.