

Fuzzy Modellierung linguistischer Kategorien*

Burghard Rieger

Linguistische Datenverarbeitung / Computerlinguistik
FB II: Sprach- und Literaturwissenschaften der Universität Trier

rieger@ldv01.Uni-Trier.de

1 Die semiotische Dimension des Ansatzes

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 — ebenso 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*¹), von den *Korrespondenzrelationen* zu unterscheiden, welche die Beziehungen zwischen Zeichen (und Zeichen-Aggregaten) einerseits und dem von diesen Bezeichneten andererseits regeln (*Semantik*). Innerhalb jeder Disziplin ergibt sich im Hinblick auf die Ausbildung ihrer jeweiligen Fachsprache eine quasi metawissenschaftliche Aufgabenstellung, der sich die Sprachwissenschaft im Hinblick auf die natürliche Sprache objektwissenschaftlich gegenüber sieht. Aus dieser doppelten Aufgabenstellung linguistischer Forschung ergibt sich die weit über die Linguistik hinausweisende Bedeutung ihrer Begriffs- und Kategorienbildung.

1.1 Überprüfbarkeit

Insbesondere für die empirischen Objektwissenschaften erweist sich, daß sie dem Anspruch der Überprüfbarkeit ihrer Aussagen nur in dem Maße zu genügen vermögen,

*Erschienen in: Hinrichs, E./Feldweg, H. (Hrsg.): Lexikon & Text: wiederverwendbare Methoden und Ressourcen zur linguistischen Erschließung des Deutschen. [Lexicographica Series Maior] Tübingen (Niemeyer), S. 155-169, 1996.

¹Im Unterschied dazu sprechen wir von *Syntax*, wenn diese Zuordnungsbeziehungen in Form von Regeln, Produktionen, etc. selber wiederum in (formal-)sprachlichen Notationen repräsentiert werden.

wie sie die Klärung der *Zuordnungs-* und *Korrespondenzbeziehungen* für ihre fachwissenschaftlichen Notationssysteme haben vorantreiben können. Mit der möglichst eindeutigen Bestimmung und definatorischen Festlegung dieser Beziehungen ergibt sich eine Aufgabenstellung, die von der Linguistik angegangen wird. Was sich in allen wissenschaftlichen Disziplinen als eine wichtige Voraussetzung der Theorienkonstruktion und Modellüberprüfung erwiesen hat, kann daher als ein sehr grundlegendes Problem insofern gelten, als seine Lösung eine Bedingung dafür ist, daß Erfahrungen, die im Umgang mit der Welt gemacht werden, symbolisch dargestellt, über diese Darstellung vermittelt und aufgrund der diese Vermittlung regelnden Verstehensprozesse durch Nachvollzug auch überprüft werden können.

Im Hinblick auf diejenigen Prozesse, die das Verstehen natürlichsprachlicher Repräsentationen vermitteln, 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 (referenztheoretischen und bedeutungsstrukturalen) Repräsentationsformalismen oder auch in einem diese beiden einschließenden, empirisch wie theoretisch begründeten Gegenstandsbereich der *experimentellen Semiotik*² mit prozeduralen Modellen der Prozeßsimulation entscheidende Beiträge geliefert. Dabei wird bis heute mit Begriffen und Kategorien gearbeitet, deren Idealisierungen nicht die Konsequenz strikter Theorien- und Modellbildungen sind, sondern oft selber den Grund bilden für eine erschwerte bis unmögliche Hypothesenbildung, die strikte Überprüfung deswegen kaum zuläßt.

1.2 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 Gesetzhypothesen generell 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 Gesetzhypothesen, aus denen sie ableitbar wären, und in Ermangelung auch einer damit bereitstehenden formalen Notation, welche sie zu repräsentieren erlaubte—deshalb nur als Zusammenhang von Beobachtungen mitteilbar sein oder als Resultat von 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 können,³ wissenschaftli-

²(Rie95)

³Obwohl an anderer Stelle (Rieg:93), S.185ff als besondere Leistung und *semiotische* Dimension der natürlichen Sprache hervorgehoben, ist es gerade diese Eigenschaft, welche die Bedeutungen

che Aussagen insbesondere solcher Disziplinen zu überprüfen erschwert, deren Forschungsgegenstand "natürlichsprachliche Zeichenketten" bilden, 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 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 explizieren, müssen sie daher dem individuellen Akt einer verstehenden Interpretation solcher Hypothesen überlassen, anstatt sie an Kalküle zu delegieren. Denn dazu wäre ihre Reformulierung in einer formalen Sprache notwendig, die diese Unterscheidung explizit macht und zu notieren erlaubt. Notationssysteme solcher Art sind daher eine wesentliche Voraussetzung umfassender wissenschaftlicher Theorienbildung.

1.3 Optimierung

Die geschichtliche Entwicklung der Wissenschaft in eine Vielzahl (mehr oder weniger) eigenständiger Disziplinen stellt sich als kontinuierlicher Prozeß primären Gewinns, vermittelnden Darstellens, lernenden Erwerbs und überprüfenden Veränderens 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 aller Erkenntnis (die *Kontexte*) vereinheitlichend in Repräsentationen (die *Theorien*) überführt wird,⁴ die zwischen den Ausgangsbedingungen (*Erkenntnisinteresse*), den Objekten (*Forschungsgegenstand*) und den Verfahren (*Untersuchungsmethoden*) zu unterscheiden und sie zu spezifizieren erlauben. Erst eine (auch notationelle) Differenzierung dieser Bestimmungsstücke von Erkenntnis läßt die (metawissenschaftliche) Präzisierung⁵ ihrer wechselseitigen Abhängigkeiten zu und erlaubt die Festlegung von Kriterien der Überprüfbarkeit von Aussagen, die innerhalb so bestimmter Disziplinen als wissenschaftliche Hypothesen gelten können.

natürlichsprachlich formulierte Aussagen (bisher) ungeeignet sein läßt, etwa durch ein *formales* Verfahren, d.h. anders als durch ihr *inhaltliches* Verstehen beurteilt zu werden.

⁴vgl. hierzu etwa die in (Rieg:89) vorgelegte semantische Vagheitstheorie mit ausgearbeitetem formalen und empirischem Teil sowie den beide spezifizierenden Zuordnungs- und Korrespondenz-Teilen

⁵vgl. hierzu schon (Rie72), (Rie79)

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.⁶ Das ist freilich kein Selbstzweck, 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.⁷ 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 Differenzierung systemtheoretisch so grundlegender Unterscheidungen wie zwischen *System* und seiner *Umgebung* als Resultat *emergenter* Strukturierungen durch *selbstorganisierende* Prozesse der Informationsverarbeitung deutet, bietet sich nach (Rie95) das *ökologische* Paradigma der Beschreibung und Analyse an.

2 Das 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, Etnologie, etc.) hat die Linguistik—wie durch ihren *semiotischen* Forschungsgegenstand 'Sprache' verführt—sich von bestimmenden Entwicklungen der modernen Wissenschaftstheorie seit Ende der 70er Jahre⁸ 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, die mit der Frage entstehen, wie denn eine überprüfbare, durch kontrollierte Verfahren wirksame, wechselseitige Vermittlung von formal-theoretischen Konstrukten einerseits und empirisch-deskriptiven Begriffen andererseits so gewährleistet werden kann, daß deren *Theorie*-geleitete, *methodische* Anwendung auf *empirische* Gegenstandsbereiche —durch experimentelle Erprobung hierzu abgeleiteter Modelle gesichert—zur kontinuierlichen Verbesserung dieser Vermittlungsleistung beizutragen vermag.

2.1 Zuordnung und Korrespondenz

Als ein möglicher Grund für die Leichtigkeit, mit der die auch in der Sprachwissenschaft anstehenden Probleme übersehen wurden, mag der folgende Umstand gelten: linguistische Kategorien und Konzepte—obwohl schon in der Vergangenheit als defizitär erkannt und auch bis heute aufgewiesen⁹ —verdeckten den Mangel sprachwissenschaftlicher Begriffs- und Theorienbildung offenbar dadurch, daß sie ja meist

⁶(RT89)

⁷(Rie91)

⁸vgl. etwa die bis dahin rege Diskussion zumindest in Deutschland um wissenschaftstheoretische Positionen der Sprachwissenschaft Linguistik (PR70), (Wun76), (BS79) etc.

⁹etwa (Sch70), (Lab73), (Wun76) aber etwa auch aus KI-Sicht (Geh89), (Tro89)

jeweils einem der beiden für die linguistische Beschreibung des Phänomenbereichs *Sprache* unterschiedenen Untersuchungsgegenstände des Sprachvermögens (*Kompetenz*) oder des Sprachgebrauchs (*Performanz*) zugeordnet werden können.¹⁰ 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 *Kategorisierung* vermittelt—noch eine (wenn auch schwach) operationale Klärung der fachterminologischen *Korrespondenzen* zwischen theoretischem Konstruktbegriff und empirische Beobachtungsbegriff methodisch etabliert wird, während dies auf den höheren Beschreibungsebenen (PS-Syntax, Semantik, Pragmatik) fast völlig fehlt. Für diese sind allerdings die *Zuordnungen* meist eingehender spezifiziert, welche die Bildung fachsprachlich korrekter 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 diskreter 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—auch ohne eine ausgearbeitete Systematik methodisch konstituierter *Korrespondenzen*— zu subsumieren.

2.2 Theorie und Empirie

Dieser Umstand erschiene 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 (mit Sprachdaten in Umfängen von $\geq 10^7$ Wörtern) erweisen sich daher 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 entwickelten Grammatikformalismen allerdings nicht, solange ihr Versagen bei Anwendung auf jeden nicht fragmentierend vor-ingeschränkten Ausschnitt performativer Sprachdaten als natürliche Konsequenz dieser (aus den unterschiedlichsten Gründen) für unangemessen gehaltenen Überprüfung erklärt wird.

So legt die Divergenz von der kategorialen Begriffswelt linguistischen Wissens mit 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* aber

¹⁰So etwa theoretische Begriffe (wie *Phonem*, *Morphem*, *Wort*, *Substantiv* und *Verb* etc.) zur Bezeichnung abstrakter Konzepte der *langue/competence* und eher empirische Begriffe (wie etwa *Laut*, *Intonation*, *Phon*, *Morph*, *Äußerung* etc.) zur Bezeichnung beobachtbarer Realisationen der *parole / performance*.

nur sehr unvollkommen zu erfassen vermögen. Hierzu wäre eine empirische Komponente in der Theorien- und Modellbildung nötig, die bisher (weitgehend) fehlt.

Daß zur Überbrückung dieser Divergenz statistische 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 möchte, welche die traditionellen linguistischen Begriffs- und Kategorienbildung in Sprachkorpora zu identifizieren erlaubt. Beide Konzepte, Zufallsvariable und linguistische Einheit, sind aber möglicherweise—wie unten deutlich werden wird—zumindet zeichentheoretisch unangemessene Abstraktionen.

3 Die zeichentheoretische Revision

Warum sollte eine semiotische Perspektive gerade für die sprachwissenschaftlichen Disziplinen eine Chance eröffnen, das Wissen (über Sprache und die Welt) zu vertiefen und zu erweitern? Sicherlich nicht allein dadurch, daß sie phänomenal begegnende *Mannigfaltigkeit* der beobachtbaren Objektwelt wieder nur—quasi ARISTOTELISCH—unter dem Aspekt der diese Vielheit vereinheitlichenden *Ordnung* beschreibt und—quasi AUGUSTINISCH—unter (immer wieder neuen Begriffsbildungen von) Typen, Klassen und Kategorien subsummiert. 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 ein Umorientierung gemeint, die über einen durch zeichenhafte Repräsentationen vermittelten Zusammenhang läuft, der—nach den Reduktionsformen *Ordnung* und *Struktur*—die übergreifende Kennzeichnung von Vielheiten durch Bestimmung und Angabe von *Prozessen* sieht, deren Besonderheit darin zu liegen scheint, daß sie in Form von *Prozeduren* repräsentiert werden können. Sie stellen eine neue Notationsform bereit, welche von der Zeitlichkeit von Prozessen zu abstrahieren erlaubt, und deren Simulation in experimentell kontrollierbaren Modellläufen eine neue, quasi-empirische Überprüfung ihres (möglichen) Erklärungswerts darstellt.

3.1 Zur Kritik der 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,¹¹ wird unter sehr unterschiedlichen Wissenschaftsauffassung und aufgrund der durch sie bestimmten Forschungsgegenstände und Untersuchungsmethoden an der theoretischen wie praktischen Entfaltung dessen gearbeitet, was unter *Kognition* zu verstehen ist.

Verbindlich scheint dabei ein prozessuales Verständnis von Kognition zu sein. *Kognitive Prozesse* können daher—im weiten Sinn: *systemtheoretisch*—als adapti-

¹¹vgl. (JLW77)

ve Leistung von informationsverarbeitenden Systemen verstanden werden, die ihre Umgebungen in jeweils dem Maße in strukturierte Repräsentationen zu überführen vermögen, wie dies ihre jeweils eigene Strukturiertheit erlaubt.

In den Kognitionswissenschaften heißen diese Prozesse *mentale*. Sie gelten als Bedingung dafür, daß *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. Die *Semiotisierung* der kognitionswissenschaftlichen Sicht besteht nun—verkürzt gesagt—in der Suche nach solchen Prozessen (bzw. ihren prozeduralen Modellierungen), die durch die Verarbeitung von schon repräsentierten Einheiten (der einen Stufe) die Repräsentationen von Einheiten (einer anderen Stufe) erst schaffen.¹²

3.2 Zur kognitiven Linguistik

Die *kognitive Linguistik* geht in ihrer Modellbildung von formal konstruierbaren, *mentalen Repräsentationen* aus, welche die Bedingungen und die Möglichkeit *mentaler Prozesse* bilden. Einer aus der Sicht der theoretischen Linguistik motivierten Perspektive,¹³ der sich die Fragen nach der Struktur und dem Erwerb von Sprachkenntnis (als *Wissen*) vor denen nach ihrer Anwendung und möglichen (materialen) Realisierung (als *Fertigkeit*) stellen, entsprechen die kognitiven Modellbildungen, welche vor allem die theoretischen Möglichkeiten und formalen Bedingungen des Sprachwissens zu klären suchen. Aufgrund der Prinzipien der *Berechenbarkeit*, *Algorithmisierbarkeit* und *Implementierbarkeit*¹⁴ von Modellierungen dieser Prozesse wird dabei ein *semiotisches* Problem in der kognitiv-linguistischen Modellierung des Sprachwissens sichtbar, auf das Bierwisch (1988) erstmals hinweist. Es stellt sich aufgrund der postulierten Berechenbarkeit und Algorithmisierbarkeit von Sprachwissen als eine *strukturelle Inkongruenz* dar, die auf der Ebene formaltheoretischer Repräsentationen die Unterscheidung von berechenbaren sprachlichen Einheiten (Lexikon) und deren Verkettungsregeln (Grammatik) einerseits betrifft, welche andererseits der Unterscheidung von Gedächtnisstruktur (mentalem Lexikon) und Regelverarbeitung (den Parsern) auf der Ebene der algorithmisierbaren Prozeduren nicht genau entspricht.¹⁵

¹²Die in dieser Sprechweise suggerierte Abfolge primärer vor sekundären Repräsentationen ist der Prozeduralität dieses Phänomens eigentlich unangemessen; die Unterscheidung sowohl der Repräsentationsebenen als auch ihrer zeitlichen Relation sind Kennzeichen der re-konstruktiven Modellierungen der Prozesse, nicht der diesen zugrundeliegenden Prozedur, die von beidem zu abstrahieren erlaubt.

¹³vgl. etwa (Cho88), Lecture 1 (S.1–34), wo diese entfaltet wird.

¹⁴nach (Mar82)

¹⁵Diese Inkongruenz kann als *semiotisch* begründet deswegen gelten, weil sie besagt, "daß die Regeln und Prinzipien aus *G* [der *Grammatik*] auch den *Inhalt* von *ML* [dem *mentalen Lexikon*] kontrollieren, daß der *Parser* aber [gerade] nicht allen Regeln und Prinzipien von *G* korrespondiert" ((Bie88), S.203 [meine Hervorh.]), also eine Algorithmisierungs- bzw. Berechenbarkeits-Lücke der Modellierungen benennt, welche Prozeduren betreffen, die sprachliche Inhalte und Funktionen möglicherweise nicht nur kontrollieren sondern konstituieren.

Gleichzeitig werden anstelle empirischer Kriterien der Überprüfbarkeit von Hypothesen und Theorien der kognitiven Linguistik—trotz ihrer dem realistischen Paradigma verpflichteten Wissenschaftsauffassung—im wesentlichen nur zwei (voneinander nicht unabhängige, zudem indirekte) Adäquatheitskriterien für die Beurteilung von Aussagen formuliert,¹⁶ die eine intersubjektive Übereinstimmung in der Bewertung der ihnen zugrundeliegenden Hypothesen (*aufrechterhalten / verwerfen*) nur durch die Anerkennung ihrer theoretischen Prämissen, nicht durch ihren empirischen Gehalt anbieten können.

3.3 Zur kognitiven Sprachverarbeitung

Die *kognitive Sprachverarbeitung* konfrontiert die eher kompetenztheoretischen Modelle der kognitiven Linguistik mit der performativen Praxis der kommunikativen Produktion und Rezeption natürlichsprachlicher Texte,¹⁷ zu deren Erklärung sie eine *ökologisch* orientierte Kognitionstheorie fordert. Der kognitiv-sprachverarbeitende Ansatz, der konsequenterweise die Bedingungen wissenschaftlicher Kommunikation und deren Modellpluralität einbezieht und als eine besondere, weil explizierte Form situierter Produktion und Rezeption von Zeichen- und Symbolaggregation zu analysieren und zu beschreiben unternimmt, vertritt—im ökologischen Paradigma seiner Wissenschaftsauffassung—eine dynamische Konzeption von Modellierung, deren Überprüfbarkeit weitgehend in der methodischen Realisierung kontextuell situierter *Prozeßsimulationen* begründet ist. Diese können dabei als *semiotische* Erklärung für das Entstehen von Strukturen, Ordnungen und Vielheiten deswegen gelten, weil 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 auf den Repräsentationen verschiedener Ebenen festzulegen erlaubt, sind aus der Menge ihrer möglichen Algorithmisierungen (*algorithmic level*) 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 Regel-verarbeitenden Algorithmen in den symbolischen Modellen den Bereich der *Zuordnungen* abdecken, vermögen in den sub-

¹⁶*Lernbarkeit*: "eine Aussage über die mentale Repräsentation sprachlichen Wissens ist nur dann adäquat, wenn die entsprechende Repräsentation prinzipiell lernbar ist"; *Verarbeitbarkeit*: "eine vorgeschlagene Repräsentation muß insbesondere durch Prozessoren, etwa einen plausible Parser oder Formulator, verarbeitet werden können" ((FHR94), Einleitung). Damit wird aber *Adäquatheit* der Schnittmenge aller prinzipiell lernbaren Repräsentationen (unabhängig von ihrem Notationsformat) mit allen algorithmisch verarbeitbaren Repräsentationen (sofern es nur plausible Implementationen solcher Algorithmen gibt) zugesprochen, was die Frage der Auswahl, wonach bestimmten Modellen der Vorzug vor anderen zu geben wäre, nicht beantwortet, sondern nur auf das Kriterium der "Plausibilität" von "Parsern und Formulatoren" verschiebt.

¹⁷vgl. (RS93)

symbolischen Modellen diese Algorithmen offenbar den Bereich der *Korrespondenzen* in dem Maße zu konstituieren, in dem sie Einheiten unterschiedlicher Repräsentationsebenen dadurch zueinander in Beziehung setzen, daß sie (mindestens) eine dieser Ebenen erst schaffen. Die Rede ist von *emergenten* Strukturen, welche bisher nicht-unterscheidbare Einheiten dadurch unterscheiden lassen, daß sie als Resultate von Prozeduren erscheinen, welche die Daten, auf denen sie operieren, verändern.

4 Unscharfe (*fuzzy*) Kategorien

Der Einsatz *quantitativer* Verfahren der Analyse und Beschreibung sprachlicher Elemente und deren beobachtbaren Vorkommens in sehr großen Textkorpora in Verbindung mit den *fuzzy-theoretischen* Möglichkeiten der Modellierung von elastischen Einheiten können dagegen ein durch numerische Spezifizierungen erhöhtes Auflösungsvermögen mit den größeren Toleranzen kombinieren, welche *weiche* Begriffe dadurch erlauben, daß sie Form (Umfang) und Funktion (Inhalt) über die gleiche Bestimmung kennzeichnen lassen.

4.1 Ansatz und Modellierung

Unscharfe (fuzzy) Kategorien heißen dabei solche abstrakten Zuordnungen, deren (leere) Strukturen ebenso wie deren mögliche Füllungen als Resultate von Prozessen erscheinen, die in Form von *Prozeduren* dargestellt werden können. Diese *prozedurale* Form erlaubt

- veränderliche strukturelle Zusammenhänge als *dynamisch* sich verändernde Resultate von Prozessen zu erklären, deren (repetitiver) Ablauf die (topologische) Struktur der Daten, auf denen sie operieren, verändert;
- diese Prozesse als Modell *kognitiver* Leistungen zu deuten, wodurch Elemente (und Elementverbindungen) einer Ebene Elementen einer anderen zugeordnet werden, die beide (Elemente wie Ebene) durch diese Zuordnung erst entstehen;
- diese Zuordnung als prozedurale Explikation einer *Bedeutungskonstitution* zu verstehen, insofern bisher verborgenen (*hidden*) und uninterpretierten Einheiten durch die Prozedur Funktionen zukommen.

Es scheint daher an der Zeit, traditionell-kategoriale Konstrukte und Begriffsbildungen der Linguistik zu überprüfen und gegebenenfalls als *weiche Kategorien* neu zu modellieren.

In diesem Zusammenhang werden in Trier eine Reihe verschiedener Untersuchungen durchgeführt, die in einem ersten Angang—ausgehend von den durchweg traditionellen Begriffsbildungen der strukturellen Linguistik—anhand extrem umfänglicher Datensätze sehr großer Textcorpora¹⁸ und mittels strikter Operationalisierung

¹⁸Das Trierer *dpa*-Korpus umfaßt derzeit das Sprachmaterial von 357.275 Meldungen mit insgesamt rund 68 Mio Wörtern des sog. *Basisdienstes 1990/91*, welches der LDV/CL der Universität Trier dankenswerterweise zu Forschungszwecken von der Deutschen Presseagentur (*dpa*) Hamburg zur Verfügung gestellt wurde (Rie92), (Gal93), (Jak94).

von Ideen, Vorstellungen und Ansätzen des Distributionalismus nach Strukturzusammenhängen fanden, welche von nicht-quantitativen Ansätzen übersehen werden müssen. Anders als durch den Einsatz bloß numerischer Verfahren des Zählens und Messens brauchen prozedurale Modellierungen unscharfer (fuzzy) Einheiten dabei keine Kontrollstrukturen vorauszusetzen, sondern können Strukturierungsleistung selbst erbringen, wobei deren Resultate —als Zwischenrepräsentationen wie Filter abrufbar—zur optimierenden Kontrolle des modellierten Prozesses (rückkoppelnd) eingesetzt werden. Als Strukturen generierende Analysen tatsächlicher Sprachdaten, welche 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.

Ausgehend von der traditionellen Kategorie "Silbe" und unter Nutzung bisheriger phonetischer, phonemischer und morphophonematischer Begriffsbestimmungen und Theoriebildungen¹⁹ als heuristische Hilfen bei der Suche nach solchen Regularitäten, die in schriftsprachlichen Daten wie dem Trierer *dpa*-Korpus vorliegen, konnte eine *dynamische* Kontrollstruktur entwickelt werden. Sie basiert auf der iterativen Agglomeration von Elementen (Buchstaben plus Leerzeichen), welche die syntagmatischen und paradigmatischen Restriktionen kombinatorisch möglicher (aber nicht realisierter) Verbindungen jener Einheiten zusammenzufassen erlaubt, aus denen Silben und Wörter im Deutschen gebildet werden (können).

4.2 Notationen und Definitionen

Zunächst einige notationelle und definatorische Vereinbarungen, die zum Verständnis der im folgenden entwickelten und beschriebenen Schritte zur Entfaltung des Konzepts *weicher Kategorisierung* und seines formalen Aufbaus notwendig sind:

Als *n*-Gramme werden *n*-gliedrige Ketten benachbarter Entitäten bezeichnet. Bei $n \geq 2$ lassen sich

Bi-Gramme 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 (von Buchstaben, Graphemen, Morphemen, Silben, Wörtern, etc.) prozedural modellierbar sind. Sie können formal erklärt werden als *dispositionelle Dependenz-Strukturen (DDS)*²⁰ von

fuzzy (Teil-)Mengen des jeweiligen Zeichen-Repertoires $\forall x_i \in G^n ; i, j, k = 1, \dots, m$ verschiedener Dimensionen $n \geq 1$

$$\tilde{X} := \{(x_i, \mu_X(x_i))\} \subseteq G^n \quad (1)$$

die aufgrund $(n - 1)$ -stelliger

¹⁹In zwei Magisterarbeiten der LDV/CL in Trier (Oer94) und (Zub94) wurden unterschiedliche Fragestellungen dieses Ansatzes bisher untersucht und anhand des Trierer *dpa*-Korpus überprüft, dem auch die unten vorgelegten Beispiele entstammen.

²⁰vgl. (Rie89), Kap.9, S.237–289

fuzzy Relationen definiert sind

$$\tilde{X}_{(n-1)} \stackrel{\text{def}}{=} \tilde{R}_{(n-1)} \subseteq \underbrace{G \times \dots \times G}_{n\text{-mal}} \quad (2)$$

Deren Elemente bestimmen sich über

Zugehörigkeitswerte , die ein numerische Ausdruck

$$\mu_{\tilde{R}_{(n-1)}}(x_i) : \underbrace{G \times \dots \times G}_{n\text{-mal}} \longrightarrow [0, 1] \quad (3)$$

der Übergangsneigungen aller linear verketteten Einheiten $x_{(n-1)i}, x_{ni} \in G^n$ sind, gemäß

$$\mu_{\tilde{R}_{(n-1)}}(x_i) = \frac{H_n(x_i)}{H_{(n-1)}(x_i)} \quad (4)$$

wobei $H_n(x_i) = \sum x_i$ mit $x_i \in G^n$ und $H_{(n-1)}(x_i) = 1$ für $n < 2$

Zeilen- und Spalten-Vektoren der so erklärten Übergangsmatrizen

$$\mathcal{R}_{(n-1)} = \begin{bmatrix} \mu_{\tilde{R}_{n-1}} 11 & \cdots & \mu_{\tilde{R}_{n-1}} 1m \\ \vdots & \ddots & \vdots \\ \mu_{\tilde{R}_{n-1}} m1 & \cdots & \mu_{\tilde{R}_{n-1}} mm \end{bmatrix} \quad (5)$$

lassen weiche Typisierungen $\tilde{R}_{(n-1)}^i$ dadurch definieren, daß

Restriktionen erklärt werden

$$\tilde{R}_{(n-1)}^i = \tilde{R}_{(n-1)} \parallel x_i \stackrel{\text{def}}{=} \{ \mu_{\tilde{R}_{(n-1)}}(x_i, x_j) \mid (x_i, x_j) \in \tilde{R}_{(n-1)} \wedge x_i \in G^{(n-2)} \} \quad (6)$$

welche die unscharfen Relationen $\tilde{R}_{(n-1)} \subseteq G^n$ auf jeweils ein Element x_i ihres Vorbereichs $G^{(n-1)}$ einschränken und so als Mengensystem unscharfer (Teil-)Mengen $[\tilde{R}_{(n-1)}^i] \in G^{(n-1)}$ etablieren.

α -Schnitte stellen sicher, daß ausschließlich Bindungswerte erfaßt werden,

$${}^\alpha \tilde{R}_{(n-1)}^i \stackrel{\text{def}}{=} \{ \tilde{R}_{(n-1)}^i \mid \mu_{\tilde{R}_{(n-1)}}(x_i) \geq \alpha \} \quad (7)$$

die über der durch α gesetzten (unteren) Schwelle liegen.

Normalisierung der unscharfen Mengen ${}^\alpha \tilde{R}_{(n-1)}^i$ sorgt schließlich dafür, daß

$$\mu_{\tilde{R}_{(n-1)}^i} = \frac{\mu_{\tilde{R}_{(n-1)}}(x_i)}{\sup_{x \in G^{(n-1)}} \mu_{\tilde{R}_{(n-1)}}(x_i)} ; \forall x_i \in {}^\alpha \tilde{R}_{(n-1)}^i \subseteq G^n \quad (8)$$

diese weichen Typisierungen mindestens ein Element enthalten, das den (reflexiven) Bindungswert 1.0 annimmt. Auf diese Weise werden

weiche (fuzzy) 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 Constraints* ableiten, welche ihre Systemstruktur bestimmen.

n - Gramme	$R = H(G^n)$ (real belegt)	$T = K(G^n)$ (theor. möglich)	$100 \cdot \frac{R}{T}$ Prozent	$F = G \cdot H(G^{n-1})$ (fakt. möglich)	$100 \cdot \frac{R}{F}$ 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
<i>Teilkorpus-Umfang</i> : 3.648.326 Zeichen 502.587 Wörter					

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

4.3 Darstellung und Repräsentation

Die Darstellung der Entwicklung *weicher* Kategorien als *elastische Constraints* (verschiedener Ebenen) wird über deren formaler Darstellung als (mehrstellige) *fuzzy*-Relationen \tilde{R}_{n-1} und den diesen entsprechenden numerischen Formaten in Übergangsmatrizen $\mathcal{R}_{(n-1)}$ erleichtert. Hierzu wird der Raum aller theoretisch möglichen Kombinationen mit $\tilde{R}_{(n-1)} \subseteq G^n$ nicht nur auf den Raum der real möglichen sondern der faktisch auch realisierten Kombinationen ${}^\alpha \tilde{R}_{(n-1)} \subseteq {}^\alpha \tilde{R}_{(n-2)} \times G^n$ eingeschränkt, den die beobachteten Einheiten in den analysierten Texten des Korpus tatsächlich aufspannen. Die Umfänge der theoretisch und faktisch möglichen sowie real belegten Mengen unterschiedlicher Zeichenketten gleicher Längen (n -Gramm-*types*) im Trierer *dpa*-Korpus sind in Tab. 1 zusammengestellt.

So ergäben sich für schriftsprachliche Texte des Deutschen auf Buchstabenebene bei m unterschiedenen Zeichen (Graphen: Buchstaben plus Leerzeichen) und einer maximalen Länge n der betrachteten Zeichenverkettungen $\forall x_i \in G^n ; i, j, k = 1, \dots, m = 31$ und $n = 1, \dots, 7$ zwar *theoretisch* die scharfen Relationen.²¹

$$\begin{aligned}
Id = R_0 &:= \{x_i\}; & x_i &\in G \\
R_1 &:= \{(x_i, x_j)\}; & (x_i, x_j) &\in G \times G \\
R_2 &:= \{(x_i, x_j, x_k)\}; & (x_i, x_j, x_k) &\in G \times G \times G \\
&\vdots & &\vdots \\
R_{n-2} &:= \{(x_1, \dots, x_{m-1})\}; & (x_1, \dots, x_{m-1}) &\in G^{(n-1)} \\
R_{n-1} &:= \{(x_1, \dots, x_m)\}; & (x_1, \dots, x_m) &\in G^n
\end{aligned}$$

Tatsächlich erfaßt werden müßten daraus aber nur nur die auch *praktisch* auf jeder der n Ebenen noch zu realisieren möglichen Zeichen-Kombinationen.²²

²¹vgl. die *theoretisch* möglichen Häufigkeiten unterschiedlicher n -Gramme in Tab.1, Spalte T

²²vgl. die *faktisch* möglichen Häufigkeiten unterschiedlicher n -Gramme in Tab.1, Spalte F

$$\begin{aligned}
Id = R_0 &:= \{x_i\}; & x_i &\in G \\
R_1 &:= \{(x_i, x_j)\}; & (x_i, x_j) &\in R_0 \times G \\
R_2 &:= \{(x_i, x_j, x_k)\}; & (x_i, x_j, x_k) &\in R_1 \times G \\
&\vdots & & \vdots \\
R_{n-2} &:= \{(x_1, \dots, x_{m-1})\}; & (x_1, \dots, x_{m-1}) &\in R_{(n-3)} \times G \\
R_{n-1} &:= \{(x_1, \dots, x_m)\}; & (x_1, \dots, x_m) &\in R_{(n-2)} \times G
\end{aligned}$$

Die *unscharfe (fuzzy)* Modellierung gemäß (4) wird über die relativen Übergangshäufigkeiten erreicht, welche die *unscharfen* Relationen für jede der $n - 1$ langen Ketten bestimmt. Dabei brauchen die Zugehörigkeitswerte demäß (3) der betreffenden $n - 1$ -stelligen *fuzzy* Relationen $\tilde{R}_{(n-1)}$ durch α -Schnittbildung gemäß (7) und Normalisierung gemäß (8) eingeschränkt—nurmehr als *Bi-Gramme* erhoben und berechnet zu werden.

Durch die jeweiligen Restriktionen gemäß (6) kann darüber hinaus ausgenutzt werden, daß in den (jeweils höher-stelligen) *fuzzy* Relationen $\tilde{R}_{(n-1)} \subseteq G^n$ das jeweils erste Glied aller geordneten Paare, die Elemente dieser Relationen sind, aus einer (schon berechneten) *Agglomeration* von Elementen aus jeweils niedriger-stelligen Relationen $\tilde{R}_{(n-2)} \subseteq G^{(n-1)}$ besteht. Dieses nicht nur formale Prinzip der prozeduralen *Selbstähnlichkeit* von n Agglomerationsschritten erlaubt darüber hinaus, aufgrund der besonderen, nach (3) 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.²³

$$\begin{aligned}
Id = \alpha\tilde{R}_0^i &:= \{\mu(x_i)\}; & x_i &\in G \\
\alpha\tilde{R}_1^i &:= \{\mu(x_i, x_j)\}; & \mu(x_i) &\in \alpha\tilde{R}_0^i \times G \\
\alpha\tilde{R}_2^i &:= \{\mu(x_i, x_j, x_k)\}; & \mu(x_i, x_j) &\in \alpha\tilde{R}_1^i \times G \\
\alpha\tilde{R}_3^i &:= \{\mu(x_i, x_j, x_k)\}; & \mu(x_i, x_j, x_k) &\in \alpha\tilde{R}_2^i \times G \\
&\vdots & & \vdots \\
\alpha\tilde{R}_{n-2}^i &:= \{\mu(x_1, \dots, x_{m-1})\}; & \mu(x_1, \dots, x_{m-1}) &\in \alpha\tilde{R}_{n-3}^i \times G \\
\alpha\tilde{R}_{n-1}^i &:= \{\mu(x_1, \dots, x_m)\}; & \mu(x_1, \dots, x_m) &\in \alpha\tilde{R}_{(n-2)}^i \times G
\end{aligned}$$

Das führt zum Aufbau einer mengentheoretischen *Komposition unscharfer* Relationen oder dem *Produkt*

$$\prod_{n=(n-1)}^n \alpha\tilde{R}_{(n-1)}^i = \{\alpha\tilde{R}_{(n-1)}^i \circ \{\alpha\tilde{R}_{(n-2)}^i \circ \dots \circ \{\alpha\tilde{R}_2^i \circ \{\alpha\tilde{R}_1^i \circ \{\alpha\tilde{R}_0^i\}\}\}\}\} \quad (9)$$

das durch seine (über die Inklusionsbeziehung) agglomerierende Struktur (relativ elegant) algorithmisiert und (relativ effizient) implementiert werden kann.²⁴ Repräsentationen der *fuzzy Kategorien*, die die Struktur des *Produkts* (9) in Form von Baumgraphen darstellen, sind in den beiden Abbildungen 1 und 2 für die beiden Buchstaben D und Z vorgelegtr. Die Tiefenpfade der Bäume lassen (bei unterschiedlichen

²³vgl. die real belegten Häufigkeiten unterschiedlicher n -Gramme in Tab.1, Spalte R

²⁴Der in (Oer94) entwickelte und auf einer PCS-Cadmus Risc3000 unter UNIX in C implementierte Algorithmus setzt dabei das schon für die mengentheoretische Rekonstruktion bestimmende Prinzip der formalen *Selbstähnlichkeit* auch für die prozeduralen Verarbeitung vorteilhaft um und nutzt es gleichzeitig zur Generierung der Baumgraphen.

α -Schwellen) die differierenden Silben- bzw. Wortbildungs-Neigungen erkennen, die sich—zunächst nur für die hier dargestellten links-nach-rechts Verkettung—ergeben.

Literatur

- [1] H. Bergenholtz and B. H. Schaefer, editors. *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora. Monographien Linguistik und Kommunikationswissenschaft 39*, (Scriptor), Königstein, Ts., 1979.
- [2] S. W. Felix, C. Habel, and G. Rickheit, editors. *Kognitive Linguistik. Repräsentation und Prozesse. Psycholinguistische Studien*, (Westdeutscher Verlag), Opladen, 1994.
- [3] R. Kruse, J. Gebhardt, and F. Klawonn. *Fuzzy Systeme. Leitfäden und Monographien der Informatik*, (B.G.Teubner), Stuttgart, 1. edition, 1993.
- [4] W. Labov. “the boundaries of words and their meaning”. In C. Bailey and R. Shuy, editors, *New Ways of Analyzing Variation in English*, pages 340–373, (Georgetown U.P.), Washington, 1973.
- [5] B. Oerder. *Zum Problem linguistischer Kategorien: Untersuchungen der Silbenkonstitution im Deutschen und ihre unscharfe Modellierung*. Master’s thesis, (Computerlinguistik Mag.), FB II: Universität Trier, 1994.
- [6] H. Pilch and H. H. Richer, editors. *Theorie und Empirie in der Sprachwissenschaft*. (Karger), Basel / München / Paris, 1970.
- [7] G. Rickheit and H. Strohner. *Grundlagen der kognitiven Sprachverarbeitung. UTB 1735*, (Francke), Tübingen / Basel, 1993.
- [8] B. Rieger. “editorial”. *LDV-Forum der Gesellschaft für Linguistische Datenverarbeitung*, 9(2):1–2, 1992.
- [9] B. Rieger. “repräsentativität: von der unangemessenheit eines begriffs zur kennzeichnung eines problems linguistischer korpusbildung”. In H. Bergenholtz and B. Schaefer, editors, *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*, pages 52–70, (Scriptor), Königstein, Ts., 1979.
- [10] B. Rieger. *Unschärfe Semantik. Die empirische Analyse, quantitative Beschreibung, formale Repräsentation und prozedurale Modellierung vager Wortbedeutungen in Texten*. (Peter Lang), Frankfurt a.Main / Bern / Paris, 1. edition, 1989.
- [11] B. Rieger, P. Badry, and M. Reichert. “bedeutungserwerb mit referentieller kontrolle. ein system experimentell überprüfbarer bedeutungskonstitution.”. In H. P. Pütz and J. H. Haller, editors, *Sprachtechnologie: Methoden, Werkzeuge, Perspektiven*, pages 185–245, (Georg Olms), Hildesheim, 1993.
- [12] H. Schnelle. “theorie und empirie in der sprachwissenschaft”. In H. Pilch and H. Richter, editors, *Theorie und Empirie in der Sprachforschung*, pages 51–65, (Karger), Basel / München / Paris, 1970.

- [13] D. H. Wunderlich, editor. *Wissenschaftstheorie der Linguistik*. (Athenäum), Kronberg, Ts., 1976.

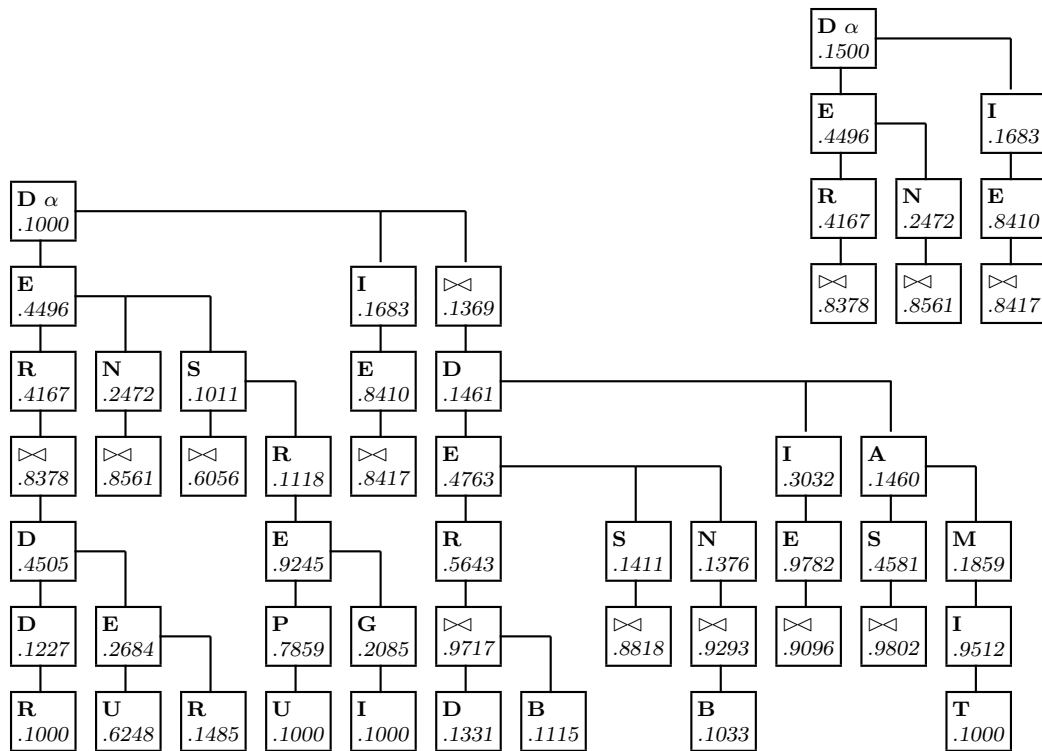


Abbildung 1: Hierarchien höchster Übergangsnegungen des Buchstaben D in Form allgemeiner Baumgraphen (*links*: mit α -Schnitt bei 0.1, *rechts*: mit α -Schnitt bei 0.15); die Ziffern geben die Zugehörigkeitswerte der betreffenden Buchstaben zu den jeweiligen weichen Typisierungen einer Agglomerationsstufe an.

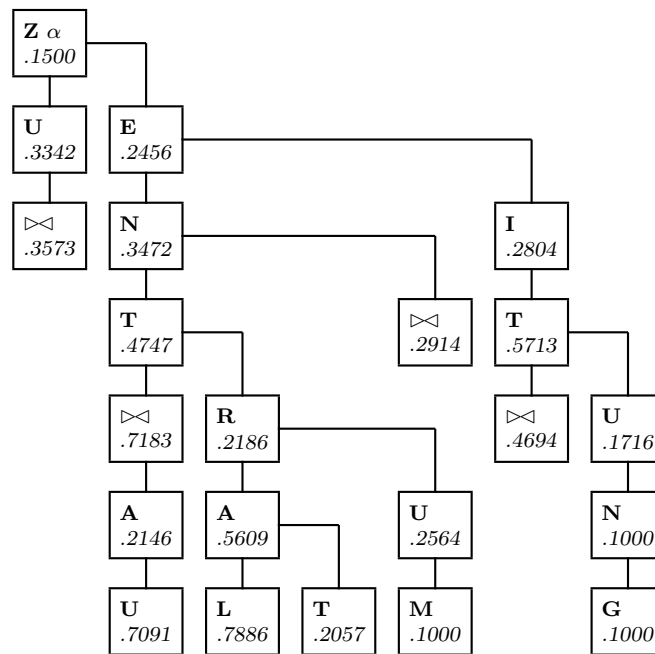


Abbildung 2: Hierarchien höchster Übergangsnennungen des Buchstaben Z in Form allgemeiner Baumgraphen (mit α -Schnitt bei 0.15); die Ziffern geben die Zugehörigkeitswerte der betreffenden Buchstaben zu den jeweiligen *weichen* Typisierungen einer Agglomerationsstufe an.