

Fakultät für Linguistik und Literaturwissenschaften, Universität Bielefeld

Magisterarbeit

**Modellierung der Sprachproduktion:  
Perspektiven neuerer kognitionswissenschaftlicher  
Ansätze**

Marc Kupietz

Oktober 1996

Gutachter: Dr. H.-J. Eikmeyer  
Dr. U. Schade

Hiermit versichere ich an Eides statt, daß ich die nachfolgende Arbeit selbstständig verfaßt habe und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Bielefeld, 3. November 1996,

# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
<b>I. Paradigmen in den Kognitionswissenschaften</b>	<b>2</b>
<b>1. Der klassische Ansatz</b>	<b>3</b>
1.1. Der Ausgangspunkt: Design vs. Evolution . . . . .	3
1.2. Modularität . . . . .	7
1.2.1. Doppelte Dissoziationen . . . . .	8
1.2.2. Die Modularitätsannahme als Instrument . . . . .	11
1.3. Ontogenese . . . . .	11
1.4. Erklärungsebenen . . . . .	12
1.5. Die „Language of Thought“-Hypothese . . . . .	15
1.6. Die Computer-Metapher . . . . .	19
1.6.1. Die Church/Turing-These . . . . .	19
1.6.2. Die „Physical-Symbol-System“-Hypothese . . . . .	20
1.6.3. Das „Chinese-Room“-Gedankenexperiment . . . . .	21
1.6.4. Funktionalismus . . . . .	22
1.6.5. „Symbol-Grounding“ . . . . .	23
1.7. Zusammenfassung . . . . .	25
<b>2. Neuere Ansätze</b>	<b>26</b>
2.1. Konnektionismus . . . . .	26
2.1.1. Charakteristika . . . . .	27
2.1.2. Die Attraktivität des Ansatzes . . . . .	27
2.1.3. Zur Existenz von Regeln und Symbolen . . . . .	28
2.1.4. Kritikpunkte . . . . .	29
2.2. „Artificial Life“ und Situierete Agenten . . . . .	30
2.2.1. Biologische Plausibilität . . . . .	31
2.2.2. Situiertheit . . . . .	32
2.2.3. „Evolutionary“ Robotics . . . . .	33
2.2.4. Implikationen für die Untersuchung höherer kognitiver Fähigkeiten . . . . .	34

2.3.	Allgemeine Dynamische Systeme . . . . .	34
2.3.1.	Dynamische Systeme und Konnektionismus . . . . .	35
2.3.2.	Die Dynamik situierter Agenten . . . . .	36
2.4.	Ein Versuch, das Puzzle zusammenzusetzen . . . . .	37
<b>II.</b>	<b>Modellierung der Sprachproduktion</b>	<b>38</b>
<b>3.</b>	<b>Die Entwicklung</b>	<b>39</b>
<b>4.</b>	<b>Ein PDP–Ansatz zur Erklärung von „Frame–Constraints“</b>	<b>41</b>
4.1.	Gegenstand der Analyse . . . . .	41
4.2.	„Frame Constraints“ . . . . .	41
4.2.1.	Versprecherdaten . . . . .	42
4.3.	Das Modell . . . . .	44
4.4.	Simulationsstudien . . . . .	45
4.4.1.	Studie 1: Variation der Simulationsfaktoren . . . . .	45
4.4.2.	Studie 2: Fehler–Generierung durch Rauschen . . . . .	47
4.4.3.	Studie 3: Größeres Vokabular . . . . .	47
4.5.	Diskussion . . . . .	48
4.5.1.	Linguistik vs. Psycholinguistik . . . . .	48
4.5.2.	Ein Vergleich mit alternativen Ansätzen . . . . .	49
<b>5.</b>	<b>PDP–Modellierung der Sprachproduktion</b>	<b>52</b>
5.1.	Segmentbewegungen . . . . .	52
5.2.	Simulationsstudien . . . . .	53
5.2.1.	Studie 1: Reproduktion . . . . .	53
5.2.2.	Studie 2: Bewegungsfehler durch Kontamination der Input–Schicht . . . . .	55
5.2.3.	Studie 3: Bewegungsfehler durch Kontamination der Hidden–Schicht . . . . .	56
5.2.4.	Studie 4: Größeres Vokabular, unterschiedliche Segmentanzahl . . . . .	56
5.3.	Diskussion . . . . .	58
5.3.1.	Modellbewertung . . . . .	58
5.3.2.	Möglichkeiten des psychologischen Erkenntnisgewinns . . . . .	61
5.3.3.	Zusammenfassung . . . . .	66
5.3.4.	Schlußfolgerungen und Perspektiven . . . . .	68
<b>6.</b>	<b>Zusammenfassung und Schlußfolgerungen</b>	<b>69</b>
	<b>Literaturverzeichnis</b>	<b>71</b>

# Einleitung

„Cognitive Science is the goldrush of the mind. Everybody's searching for it. Worse still, everybody (symbolic AI workers, subsymbolic AI workers, neuroscientists, naturalistic philosophers, and so forth) claims to be finding it. Or at least, they claim to know where to look.“ (Clark 1989)

Clarks Analogie illustriert treffend die derzeitige Situation in der Kognitionswissenschaft. Zehn Jahre nach der Konnektionismus–Symbolismus–Debatte, gibt es heute eine fast unüberschaubare Menge verschiedener Ansätze unterschiedlichster Herkunft.

Das wesentliche Ziel dieser Arbeit ist, eine Strukturierung dieser Paradigmenvielfalt zu leisten und darüber hinaus *die* Prinzipien der verschiedenen Ansätze zu extrahieren, die für eine zukünftige Kognitionswissenschaft im allgemeinen und für die Sprachproduktionsforschung im besonderen erfolgversprechend sind.

Der erste Teil der Arbeit soll zunächst einen detaillierten Überblick über die verschiedenen Paradigmen in der Kognitionswissenschaft verschaffen. Dabei wird nicht die Argumentation für einen bestimmten Ansatz im Vordergrund stehen. Stattdessen werden die Grundsätze der verschiedenen Paradigmen im einzelnen analysiert, bewertet und zueinander in Beziehung gesetzt. Diese Vorgehensweise ermöglicht einerseits eine undogmatische Beurteilung der Paradigmen und ermöglicht andererseits eine paradigmunenabhängige Auswahl von Prinzipien und Methoden für eine vielversprechende Erforschung von Kognition.

Im zweiten Teil der Arbeit untersuche ich dann die Anwendbarkeit dieser allgemeinen Prinzipien und Methoden auf das Gebiet der Sprachproduktion. Dieses Gebiet bietet sich unter anderem deshalb für eine solche Untersuchung an, weil es bislang weitgehend von einer theoretischen Diskussion — gerade was die Implikationen neuerer Ansätze betrifft — ausgenommen wurde. Die Analyse von Simulationsstudien zu einem konkreten Ansatz der Modellierung von Sprachproduktionsprozessen (Dell, Juliano & Govindjee 1993) und deren Diskussion führt zu Schlüssen, welche einerseits eine lokal konnektionistische Modellierung nahelegen und andererseits die Forderung unterstützen, Kognition nicht als isoliertes mentales Phänomen zu betrachten, sondern als Interaktion mit der realen Welt.

**Teil I.**

**Paradigmen in den  
Kognitionswissenschaften**

# 1. Der klassische Ansatz

In den folgenden Abschnitten stelle ich die wesentlichen Charakteristika des klassischen Ansatzes zur Kognitionswissenschaft dar, der je nach Perspektive und Ziel mit „*symbolic paradigm*“ (Smolensky 1988, S. 3), „*computational model of the mind*“ (Pylyshyn 1986), „*GOFAI (good old fashioned AI)*“ (Haugeland 1985), „*High Church Computationalism*“ (Dennett 1986), „*semantically-transparent-systems-approach*“, „*classical cognitivism*“ (Clark 1989) etc. bezeichnet wird. Im Unterschied zur vorhandenen Literatur sollen dabei nicht lediglich die Unterschiede zwischen symbolischen und subsymbolischen, parallel-distribuiert konnektionistischen Ansätzen herausgearbeitet werden. Vielmehr soll ein Gesamteindruck des klassischen Ansatzes entstehen, der es ermöglicht, diesen zum einen unabhängig von Alternativen einzuschätzen und zum anderen ihn perspektivisch entlang verschiedener Dimensionen mit möglichen Alternativen zu vergleichen. Dies ist notwendig, da erstens Symbolismus und Konnektionismus nicht mehr „*the only games in town*“ sind (ich denke z.B. an „*A-Life*“ und systemtheoretische Ansätze) und weil zweitens konkrete konnektionistische Modelle in der Regel nicht alle die Vorteile haben, die sie von ihren klassischen Konkurrenten abheben sollten.

## 1.1. Der Ausgangspunkt: Design vs. Evolution

Eine Eigenschaft klassischer Modelle ist, daß ihre Erklärungen hauptsächlich auf einer systematischen funktionalen Dekomposition einer Aufgabenanalyse beruhen. In Levelts Sprachproduktionsmodell (Levelt 1989) wird zum Beispiel die Aufgabe, aus einer Absicht eine konkrete sprachliche Artikulation zu generieren, in die Unteraufgaben „Konzeptualisierung“, „Formulierung“ und „Artikulation“ zerlegt. Das Modell postuliert dann die Existenz von Systemkomponenten, die jeweils für die Lösung einer dieser Subaufgaben zuständig sind — nämlich den *Konzeptualisator*, den *Formulator* und den *Artikulator* — und die Art der Kommunikation zwischen diesen Komponenten. Die Wahl eben dieser Dekomposition erscheint zunächst willkürlich, oder wie Levelt selbst es formuliert:

„The architecture in figure 1.1 may, on first view, appear to be rather arbitrary and at this stage it is.“ (S. 14)

Andererseits gibt es natürlich Gründe, eine bestimmte Ausgangsarchitektur anderen vorzuziehen. Als Grundlage aller weiteren Überlegungen dient in den meisten Fällen das,

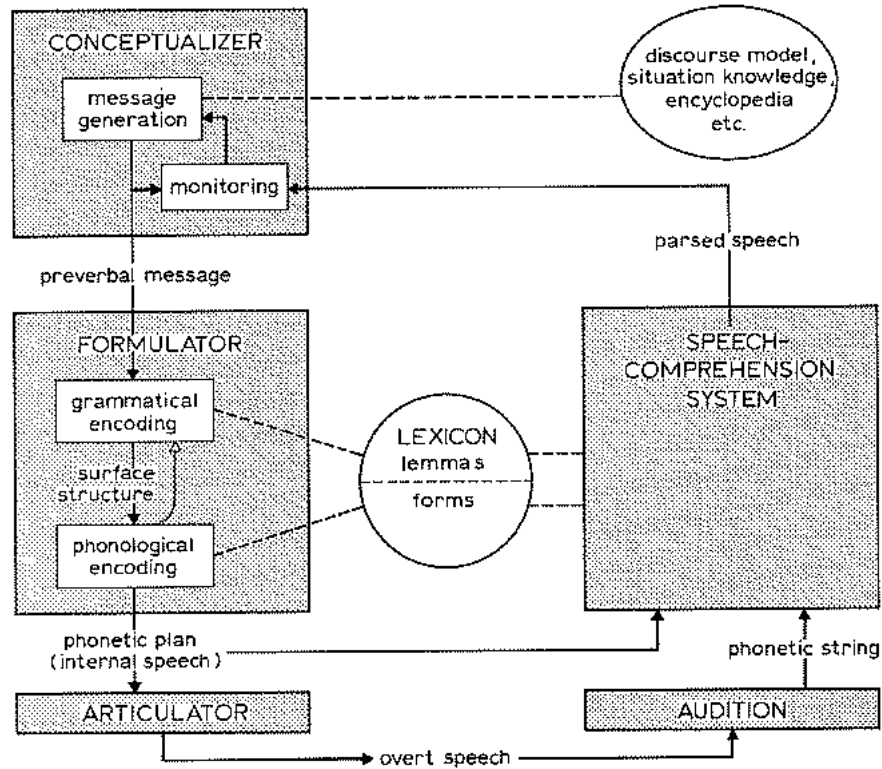


Abbildung 1.1.: Levelts Entwurf eines Sprechers. Kästchen repräsentieren Verarbeitungskomponenten; Kreise und Ellipsen repräsentieren Wissenspeicher. (Levelt 1989, S. 9)



was Dennett (1978) den Designstandpunkt („*design stance*“) nennt. Von diesem Standpunkt aus stellt man sich die Frage, wie das zu verstehende System beschaffen sein sollte, also z.B., aus welchen Komponenten es bestehen sollte, um seine Aufgabe optimal zu bewältigen.

Da die Wahl eines Ausgangspunktes in der Modellbildungspraxis äußerst kritisch ist, um gewissermaßen nicht mit dem falschen Fuß aufzustehen, soll auf die oben skizzierte, typische Wahl im folgenden ausführlich eingegangen werden.

Es stellt sich also die Frage, inwieweit die Einnahme eines „Designer-“ bzw. eines „Ingenieursstandpunkts“ die Menge möglicher Architekturen und damit die Menge der erklärbaren Phänomene einschränkt und inwieweit diese Einschränkung gerechtfertigt ist.

Bei der Wahl einer Designer-Perspektive geht man davon aus, daß es erstens einen kognitiven Mechanismus gibt, der speziell für die zu erklärende Aufgabenlösung zuständig ist — bzw. „designed“ wurde —, und daß zweitens dieser Mechanismus ein mehr oder weniger optimales Design hat. Die Diskussion der ersteren Annahme soll bis zum Abschnitt über Modularität (§1.2) aufgeschoben werden. Daß letzteres nicht unbedingt der Fall sein muß, soll im folgenden durch zwei von Dawkins (1986, S. 92-95) beschriebene Beispiele illustriert werden, die vom menschlichen Auge bzw. von Plattfischen handeln.

Im menschlichen Auge sind die lichtempfindlichen Zellen vom Licht fort gerichtet, wohingegen die Verbindungen zum optischen Nerv zum Licht hin gerichtet sind und über die Oberfläche durch ein Loch in der Retina (dem blinden Fleck) in Richtung Gehirn führen. Diese Konstruktion, die vermutlich aus der Notwendigkeit heraus evolviert ist, mit wenig Veränderungen in kurzer Zeit ein funktionierendes System zu schaffen, erscheint vom Ingenieursstandpunkt umständlich und abwegig.

Dawkins zweites Beispiel betrifft das Auge von flachen Knochenfischen. Für Fische, die dicht am Meeresboden leben, ist es von großem Vorteil, flach zu sein, um sich den Konturen des Untergrundes anpassen zu können. Eine mit den Haien verwandte Art von flachen Fischen, nämlich Rochen und andere Knorpelfische, hat ihre Flachheit naheliegenderweise entwickelt, indem sie sich seitlich ausgebreitet hat, so als ob ihr eine Dampfwalze über den Rücken gefahren wäre. Nicht jedoch die flachen Knochenfische, wie z.B. Scholle oder Seezunge. Sie sind in horizontale Richtung abgeflacht und halten sich gewissermaßen auf der Seite liegend am Meeresgrund. Für die Vorfahren von Scholle und Seezunge war dies zunächst auch eine naheliegende „Lösung“ für das Problem, sich am Boden verstecken zu müssen, da sie im Gegensatz zu den Haien höher als breit waren. Sie brachte für die Fische jedoch die Schwierigkeit mit sich, daß ein Auge immer in den Sand gerichtet und somit von geringem Nutzen war. Die offensichtliche Methode, dieses neue Problem zu beseitigen, die Andy Clark im Gegensatz zur Ingenieurs- die Flickschusterlösung („*tinkerer's solution*“, Clark 1989, S. 71) nennt, ist, einfach das untere Auge langsam auf die andere Seite nach oben wandern zu lassen. Auch dies ist eine ziemlich unelegante Lösung, die zeigt, daß die Evolution einfache, billige und kurzfristige Lösungen vorzieht, selbst wenn diese neue Probleme mit sich bringen, die dann wiederum durch weiteres „Herumbasteln“ gelöst werden müssen. Diese seltsame Entwicklungsgeschichte wird von jedem flachen Knochenfisch rekapituliert: er beginnt sein Leben als symme-

trischer an der Wasseroberfläche schwimmender Fisch und läßt sich dann, nachdem sich langsam sein Skelett verändert hat und sein Auge über den Kopf auf die andere Seite gewandert ist, auf dem Meeresboden nieder. Dawkins zieht daraus folgendes Fazit:

„The whole skull of a bony flatfish retains the twisted and distorted evidence of its origins. Its very imperfection is powerful testimony of its ancient history, a history of step by step change rather than deliberate design. No sensible designer would have conceived such a monstrosity if given a free hand to create a flatfish on a clean drawing board.“ (S. 92)

Die Frage ist nun: Warum sollten sich kognitive Funktionen nicht auf ähnliche Weise entwickelt haben? Ich denke, es ist eine plausible Vorstellung, daß das menschliche kognitive System, welches ursprünglich nur grundlegende Probleme der Lebens- bzw. Arterhaltung lösen mußte, im Laufe der Zeit sich immer wieder in kleinen Schritten auf höchst unvollkommene Weise ohne größere Umstrukturierungen erweitert und verändert hat. Damit ist es von der Designerperspektive aus betrachtet heute vermutlich so wenig optimal wie das Auge, wie flache Knochenfische oder wie ein uraltes Computerprogramm, das ursprünglich zum Schreiben von Rechnungen in einem kleinen Weinladen entwickelt wurde, und nun, ohne daß es je reimplementiert wurde, sämtliche Geschäfte einer Getränkehandelskette verwaltet.

Die Frage, die sich damit für Kognitionswissenschaftler stellt, ist, ob sich kognitive Funktionen überhaupt ahistorisch, das heißt ohne Berücksichtigung von Phylogenese und Ontogenese, erklären lassen. Eine radikale Antwort, die unter anderem von Brooks (1991a, 1991b, 1992) vertreten wird, ist, daß man höhere kognitive Fähigkeiten erst dann erklären kann, wenn man die Funktionsweise der Fähigkeiten, auf denen diese aufbauen, verstanden hat, da die Funktionsweise grundlegender Fähigkeiten im wesentlichen die Struktur später daraus entwickelter Lösungen bestimmt.

Die Erkenntnis, daß die Evolution kein Designer ist, muß allerdings noch nicht zu der Schlußfolgerung führen, daß klassische kognitive Modellierung zum Scheitern verurteilt ist. Es ist z.B. denkbar (wenn auch nur theoretisch), Wissen über die funktionale Phylogenese des Geistes in klassische Erklärungsansätze einzubeziehen. Tatsächlich wird aber — meist implizit — davon ausgegangen, daß sich Kognition ohne Berücksichtigung ihrer Entwicklung erklären läßt. Diese Annahme bzw. dieses Verständnis von „Kognition“, das vor allem von Pylyshyn (1986) geprägt wurde, schränkt die Informationsgrundlage einer Modellbildung soweit ein, daß es notwendig wird, die Modellarchitektur nach Design-Gesichtspunkten auszuwählen. Eine so gewonnene, „uninformierte“ Ausgangsarchitektur, ist zunächst allerdings nur ein schlechter Startpunkt.

Angenommen, ein Software-Wissenschaftler würde versuchen, die genaue Funktionsweise des Kommissionierungsmoduls (er nimmt an, daß es ein solches gibt) des oben erwähnten Getränkehandelsketten-Programms zu verstehen, ohne jedoch die Maschinsprache zu kennen, in der das Programm geschrieben ist (immerhin ist er in der glücklichen Lage, eine ungefähre Ahnung davon zu haben, wie Computer funktionieren!). Er analysiert also zunächst die Aufgaben, die das „Modul“ bewältigt. Dann erstellt er anhand

der Analyse und anhand seiner Vorstellung davon, wie so etwas programmiert sein sollte, ein (vorläufiges) Datenflußdiagramm, in dem er die Aufgaben auf viele Unterprogramme verteilt und die Kommunikation unter diesen festlegt. Um nun herauszufinden, wie gut sein Modell ist, erstellt er eine Computersimulation und vergleicht das Verhalten seiner Simulation, also gewissermaßen die Vorhersagen seines Modells mit dem Verhalten des echten Programms unter allen Bedingungen und mit allen Meßtechniken, die ihm einfallen. Wenn er auf Unterschiede stößt, verändert er sein Modell und beginnt von neuem.

Im Prinzip sollte der Software-Wissenschaftler so in der Lage sein, seine Vorstellung von dem Programm, das er verstehen möchte, bis hin zu einem (un-)gewissen Niveau zu verbessern. Vielleicht würde er sogar herausfinden, daß es sich bei dem Kommissionsverwaltungs-„Modul“ eigentlich um eine auf „schmutzige“ Weise erweiterte und nicht dekomponierbare (tausende goto's) Version des ursprünglichen Programms zum Schreiben von Rechnungen handelt.

Der Unterschied zwischen einem solchen Computerprogramms und Kognition besteht allerdings darin, daß sich das Computerprogramm vermutlich ohne Nutzung weiterer Informationsquellen allein anhand seines Ein- und Ausgabeverhaltens komplett nachbilden läßt. Daß dies auch für Kognition gilt, ist angesichts des Scheiterns der klassischen KI auf dem Gebiet der Robotik (vgl. Brooks 1991b; Dennett 1984) unwahrscheinlich. Es gibt jedenfalls keinen Grund, bei der Erklärung von Kognition Erkenntnisse über ihre Phylogenese zu ignorieren.

## 1.2. Modularität

Eine der theoretischen Grundlagen und Inspiration solcher, wie in Abschnitt 1.1 skizzierter, auf Dekomposition basierender psychologischer Modelle bildet in der Regel die Idee der Modularität des Geistes. Die gemeinsame Annahme der z.T. recht unterschiedlichen Modularitätstheorien (Chomsky 1980; Coltheart 1985; Fodor 1983; Marr 1982; Minsky 1985; Morton 1981) ist, daß das kognitive System aus einer Anzahl relativ *unabhängiger* Komponenten oder Moduln zusammengesetzt ist. Im folgenden soll vor allem auf Jerry Fodors Auffassung der „Starken Modularität“ eingegangen werden, da sie am ausgeprägtesten ist und vermutlich die weitreichendsten Konsequenzen für die Kognitions- und Neurowissenschaften impliziert.

In seinem Buch „*The Modularity of Mind*“ (1983) argumentiert Fodor für eine dreigeteilte funktionale Architektur kognitiver Prozesse. Fodor postuliert eine Trennung zwischen sogenannten Input- bzw. Output-Moduln/ Systemen als Vermittler zwischen Sensorik bzw. Motorik, auf der einen Seite, und dem zentralen kognitiven System, auf der anderen Seite. In die eine Richtung transformieren diese Systeme also sensorische Signale in Hypothesen über die externe Welt, formuliert in einem propositionalen Code, der „*Language-of-Thought*“ (siehe §1.5), auf dem das zentrale kognitive System operieren kann. In die andere Richtung transformieren sie in der *lingua mentis* formulierte Anweisungen des zentralen Systems in Signale zur Steuerung der Motorik. Fodors Hauptgrund,

eine solche dreigeteilte Architektur des Geistes vorzuschlagen, ist sein epistemologisches Anliegen, daß ohne sie die, seiner Meinung nach unerläßliche, Unterscheidung zwischen Perzeption und Kognition verwischt wird; ein Fehler, der unter dem Namen „New Look“ und „Interactionism“ die Kognitionswissenschaften Jahrzehnte lang dominiert hat, so Fodor (1985, S. 1). Die Unterscheidung ist für ihn deshalb so wichtig, weil, seiner Auffassung nach, die Komponenten des Input-/Outputsystems, sprich die Moduln, im Gegensatz zum zentralen System folgende Eigenschaften aufweisen: Sie sind . . .

- *domänenspezifisch*: z.B. hat das „Sprach-System“ wenig mit dem „Seh-System“ gemein
- *angeboren*: ihre Genese ist endogen determiniert und folgt einem bestimmten Tempo und einer bestimmten Reihenfolge (z.B. Spracherwerb),
- *„hart-verdrahtet“*: sie sind mit spezifischen neuroanatomischen Mechanismen, bzw. mit bestimmten neuroanatomischen Loci („*strong Modularity*“) assoziiert,
- *autonom*: der Ausfall eines Moduls ruft ein spezifisches Fehlermuster hervor (z.B. Aphasie vs. Agnosie) und beeinflusst die Funktion des restlichen Systems nur eingeschränkt
- *nicht zusammengesetzt*
- *informationell abgeschlossen*: ihre Berechnungen sind unbeeinflusst vom globalen Informationsstand des Systems (man kann nicht umhin, trotz besseren Wissens, Sinnestäuschungen zu erliegen).

Ferner konstituiert die Untersuchung dieser Moduln die wahre Domäne einer wissenschaftlich durchführbaren Psychologie, da laut Fodors „*First Law of the Nonexistence of Cognitive Science*“ (1983, S. 107) die Untersuchung des zentralen Systems keine respektable Wissenschaft hervorbringen wird, während die Untersuchung der sensomotorischen Transduktoren keine echte Psychologie hervorbringen wird.

Die Argumente, die Fodor nun zur Unterstützung seiner Theorie beibringt, beschränken sich allerdings auf die Berufung auf die eher zweifelhafte Tradition der Talent-Psychologie (speziell auf den Phrenologen Franz Josef Gall) und einige illustrative Beispiele für „informationelle Abgeschlossenheit“ bei Sinnestäuschungen. Meyering (1994, S. 43) konstatiert, daß Fodors Postulat der starken Modularität, das die Isomorphie zwischen funktionaler und neuroanatomischer Organisation einbezieht (siehe obige Auflistung), vermutlich als einziges Argument zugunsten der für sein Modularitätskonzept kritischen Eigenschaft der informationellen Abgeschlossenheit übrig bleibt.

### 1.2.1. Doppelte Dissoziationen

Das Argument zugunsten einer isomorphen Modularität neuroanatomischer und kognitiver Strukturen beruht auf der Beobachtbarkeit von sogenannten *doppelten Dissoziationen*

(Teuber 1955). Von einer einfachen Dissoziation spricht man, wenn, als Resultat eines Gehirnschadens, die Performanz eines Patienten bei einer kognitiven Aufgabe signifikant stärker beeinträchtigt ist als bei einer anderen. Z.B., konnte Patient PW (Patterson & Marcel 1977) nach einer schweren Schädigung der linken Hemisphäre 67% der ihm vorgelegten Wörter mit konkreter Bedeutung (z.B. „TABLE“) korrekt laut lesen, aber nur 13% der Wörter mit abstrakter Bedeutung (z.B. „TRUTH“). Der Befund, daß das Lesen von konkreten und abstrakten Wörtern unterschiedlich empfindlich für Gehirnschädigungen ist, legt folgende Interpretationen nahe: 1. abstrakte und konkrete Wörter sind im Gehirn separat repräsentiert; oder 2. das Lesen beider Arten von Wörtern ist gleichermaßen beeinträchtigt, aber es ist schwieriger, abstrakte Wörter auszusprechen. Nimmt man letztere Interpretation an, sollte man nicht erwarten, eine umgekehrte Dissoziation zu beobachten, also bessere Leseleistungen bei abstrakten als bei konkreten Wörtern nach einer Gehirnschädigung. Dies jedoch ist genau das, was bei Patient CAV (Warrington 1981) beobachtet werden konnte: er las 36% der konkreten und 55% der abstrakten Wörter korrekt. Zusammengenommen konstituieren PW und CAV eine doppelte Dissoziation beim Lesen konkreter vs. abstrakter Wörter. Ähnlich dieser wurden doppelte Dissoziationen bei einer Reihe von verschiedenen Aufgabenpaaren festgestellt (Plaut 1995, S. 292, für Beispiele).

Es ist sicherlich naheliegend, solch spezifische Störungen zunächst mit der Existenz unabhängiger Subsysteme zu erklären. Andererseits, je spezifischer die Ausfälle werden, desto weniger plausibel wird es, sie durch den Ausfall eines „Moduls“ zu erklären. So ist es z.B. höchst unplausibel, aus der Tatsache, daß Patient DL die Fähigkeit verloren hat, Gemüse und Früchte zu benennen (Hart, Berndt & Caramazza 1985), die Existenz eines separaten lexikalischen Früchte/Gemüse-Moduls zu folgern, zumal man, Fodors Theorie nach, annehmen müßte, daß dieses dann angeboren ist.

Zumindest in solchen Fällen sind gerade nicht-modulare Erklärungsansätze offensichtlich fruchtbarer. David Plaut (1995) z.B. simuliert in einem distribuiert konnektionistischem Modell die doppelte Dissoziation vom Lesen abstrakter vs. konkreter Wörter und zeigt außerdem, daß innerhalb seines Modells die Beschädigung von Leitungen an ein und derselben Lokalität in einigen Fällen zu einer besseren Performanz bei abstrakten und in anderen Fällen zu einer besseren Performanz bei konkreten Wörtern führen kann. Dies läßt erahnen, daß die Relation zwischen neuroanatomischer und kognitiv funktionaler Struktur nicht so einfach sein muß, wie die, die die starke Version von Fodors Theorie des Geistes postuliert.

Wie Plauts Simulation zeigen die in Abbildung 1.2 skizzierten Systeme, die, obwohl sie nicht modular sind, doppelte Dissoziationen hervorrufen können, daß es nicht zulässig ist, von der Existenz doppelter Dissoziationen auf die Existenz von Modulen zu schließen. Damit liefern doppelte Dissoziationen keinerlei Evidenz für eine modulare Organisation des kognitiven Systems und es erscheint angebracht, an dieser Stelle T.C. Meyering zu zitieren:

*„The Modularity of Mind, then, provides no real argument. Instead, what*

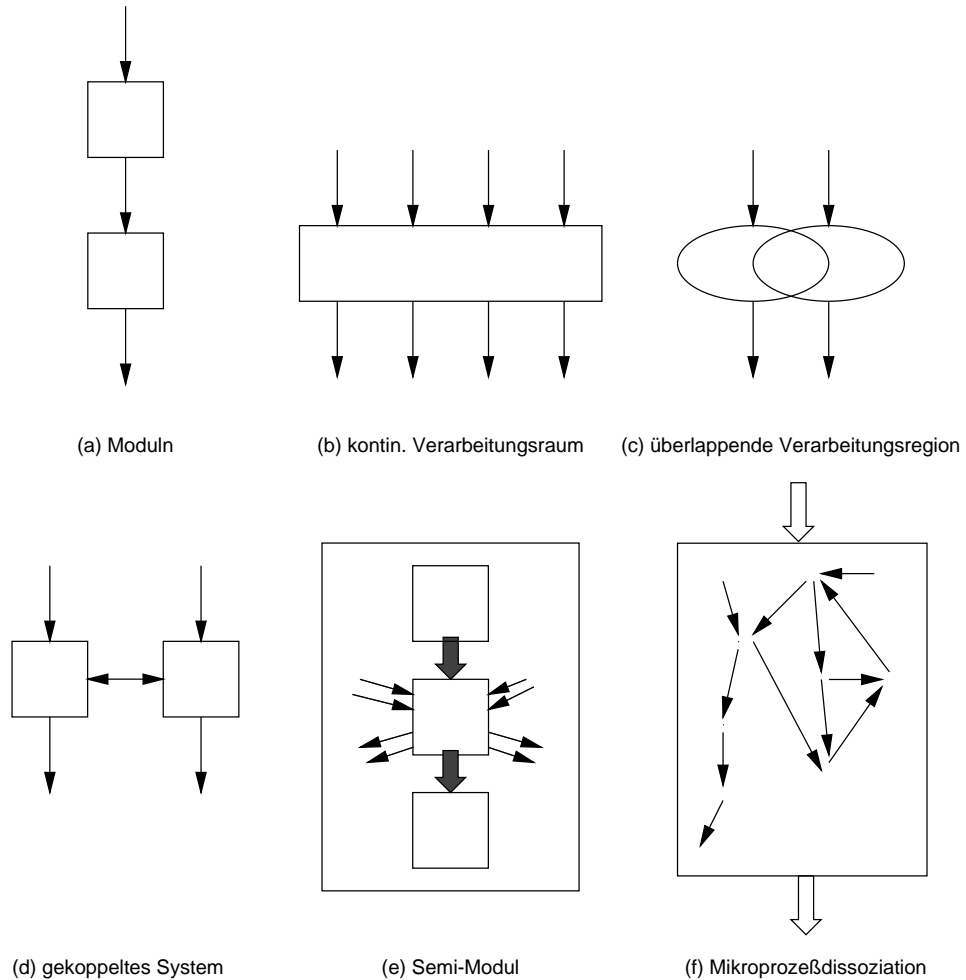


Abbildung 1.2.: Sechs verschiedene Systeme, die doppelte Dissoziationen hervorrufen können (Shallice 1988, S. 250) (a) Traditionelle Moduln. (b) Kontinuierlicher Verarbeitungsraum. Z.B. würde eine Verletzung der Retina bei  $9^\circ$  vs.  $15^\circ$  die visuelle Verarbeitung selektiv an diesen Stellen stören, ohne daß jeweils ein separates Modul gäbe. (c) Überlappende Verarbeitungsregion. Wenn die Verarbeitungsregionen zuständig für zwei Aufgaben sich partiell überlappen, würde eine selektive Schädigung der nicht überlappenden Bereiche die Aufgaben doppelt dissoziieren, obwohl keine Region ein Modul konstituiert. (d) Gekoppelte Systeme. Wenn zwei Subsysteme (z.B. visuelles und auditorisches Lexikon) eng verkoppelt sind, so daß sie sich widersprechende Information nicht verarbeiten können, wohl aber in Isolation korrekt arbeiten können, so wären sie isolierbar (Shallice 1979), ohne jedoch Moduln zu sein. (e) Semimoduln. Subsysteme können zu einem gewissen Grad Moduln sein, abhängig davon, wie weit ihr korrektes Arbeiten von internen vs. externen Einflüssen abhängt. (f) Multilevel System. Die Verarbeitung zweier Aufgaben hängt von selektiv störbaren Eigenschaften ein und desselben Systems auf verschiedenen Beschreibungsebenen ab.

it offers is at best characterized as an intriguing *just so* story, emphatically endorsed with a grandmotherly 'what I tell you three times is true'. Such grannies may well win our hearts, but not necessarily our minds.“

### 1.2.2. Die Modularitätsannahme als Instrument

Trotz ihrer fragwürdigen Qualität als Grundlage einer ontologisch realistischen Theorie des Geistes, dient eine Form der Modularitätsannahme in vielen psychologischen Erklärungen als Startpunkt bzw. als Instrument für die Generierung (falsifizierbarer) Hypothesen. Ein Aspekt eines solchen Ausgangspunktes, nämlich der *design stance*, wurde bereits in Abschnitt 1.1 behandelt. Im verbleibenden Teil dieses Abschnitt soll analysiert werden, welche Implikationen ein mehr oder weniger modularer Ausgangspunkt für die daraus resultierenden Modelle haben kann.

Die „Modul-Eigenschaft“, die am häufigsten explizite Verwendung in der Modellbildung findet, ist „*informational encapsulation*“. Levelt z.B. bezieht sich bei seiner Dekomposition des Sprachproduktionsprozesses (siehe Abbildung 1.1) *explizit* auf allein diesen Aspekt der Modularität (Levelt 1989, S. 22). Sein Modell ist also so zu interpretieren, daß es sich bei den Komponenten um informationell abgeschlossene Einheiten handelt. Das heißt, daß erstens die Art des charakteristischen Inputs einer Komponente maximal restringiert ist, und daß zweitens der Einfluß anderer Komponenten auf die Operationen einer Komponente minimal ist. Levelts Modell enthält demnach die folgenden Hypothesen: Jede Komponente verarbeitet einen charakteristischen Input: der Grammatische Enkodierer die präverbale Botschaften, der Phonologische Enkodierer die Oberflächenstrukturen, der Artikulator die phonetischen Pläne, usw. Die Operationen dieser Komponenten sind nur minimal oder gar nicht von anderen Komponenten beeinflussbar. Es gibt also z.B. kein Feedback zwischen den drei Hauptkomponenten.

Obwohl Levelts Modell aus der oben verworfenen Designerperspektive konstruiert wurde,<sup>1</sup> muß zunächst anerkannt werden, daß es einige harte, möglicherweise recht theorieunabhängige, empirisch überprüfbare Hypothesen liefert: z.B., daß der „Konzeptualisator“ Versprecher erst über das Sprach-Verstehens-System gemeldet bekommt.<sup>2</sup>

## 1.3. Ontogenese

Parallel zur Nichtzurkenntnisnahme der Phylogenese in klassisch kognitivistischen Ansätzen wird dort in der Regel auch die Ontogenese ignoriert — wie übrigens auch in Levelts

---

<sup>1</sup>Dennett schreibt einige Seiten nach einer Karikierung Levelts Modells — allerdings ohne daß ein Zusammenhang bestehen muß: „Fodor's modules are a bureaucrat's dream“. (Dennett 1991a, S. 260)

<sup>2</sup>Außerdem soll natürlich nicht der Eindruck entstehen, daß mit der Skizze in Abbildung 1.1 die Erklärungsfähigkeit des Ansatzes bereits ausgeschöpft ist. Es werden zusätzlich Aussagen gemacht über die Form von Input und Output der Komponenten, und darüber, wie die Komponenten aus ihrem Input ihren Output generieren.

„Speaking“. Die eine Seite der Ursache dafür ist, daß die Entwicklungspsychologie noch heute als eigener Zweig oder als eigene Disziplin innerhalb der Psychologie angesehen wird. Die andere Seite ist, daß sich angesichts der Abwesenheit entwicklungspsychologischer Konzepte und Erkenntnisse sich „agenetische“ oder sogar „antigenetische“ Theorien in den restlichen Gebieten der Psychologie durchsetzen konnten. Die Argumente für diese mehr (Fodor 1975; Chomsky 1986) oder weniger nativistischen (Carey 1990) Ansätze, die dem klassischen Kognitivismus zugrunde liegen, sind in der Regel ein „*ex nihilo, nihil*“-Argument und ein „*poverty of the stimulus*“-Argument. Die Verbindung dieser beiden führt zu einer Charakterisierung eines durch Geburt festgelegten Initialzustands, der via Abstraktion und Generalisierung mit seiner Umwelt interagiert (Plunkett & Sinha 1992). Der Prozeß der Wissensaquisition selbst hat innerhalb solcher Theorien wenig Einfluß darauf, welches Wissen angeeignet wird. Der Aquisitionsprozeß wird vielmehr als durch den Finalzustand der ausgereiften Kompetenz determiniert betrachtet.

Die Tatsache, daß der Gegenstand klassisch geprägter psychologischer Erklärungsversuche, speziell in der (Psycho-)Linguistik die ausgereifte Kompetenz im Gegensatz zur Performanz in Relation zu verschiedenen Parametern, wie z.B. der ontogenetischen Entwicklungsstufe ist, ist — meiner Ansicht nach — allerdings nicht darauf zurückzuführen, daß unter den kognitiven Psychologen eine explizite Nativismus-Akzeptanz oder ein explizites Genese-Desinteresse herrscht. Der Grund ist vielmehr, daß die übliche Methodologie allein Beschreibungen auf einer *Kompetenzebene* zuläßt und sich diese nicht weiter reduzieren lassen. Die Richtigkeit dieser Behauptung soll in den folgenden zwei Abschnitten belegt werden.

## 1.4. Erklärungsebenen

Erklärungen sind offenbar eine vielschichtige Angelegenheit. So ist es möglich, ein einzelnes Phänomen unter einer Reihe von in ihrer Generalität ansteigenden Erklärungsschemata zu subsumieren. Je niedriger man die Ebene wählt, desto detaillierter und je höher, desto genereller applizierbar sind die möglichen Beschreibungen. Jede einzelne Ebene hat ihre Vor- und Nachteile; einige Erklärungen sind nur auf einer bestimmten Ebene gewinnbar, während die Varietäten der sich darunter subsumierten individuellen Phänomene nur durch einen Abstieg auf der Leiter der explanatorischen Generalität erklären lassen.

Darwins Theorie der natürlichen Auslese z.B. ist auf einer sehr hohen Erklärungsebene angesiedelt. Sie beschreibt sehr allgemein, wie durch angepaßtheitsabhängige Reproduktion und einen Mechanismus der Übertragung von Eigenschaften auf die Nachkommenschaft „blinde“ Auslese scheinbar teleologische (oder absichtsvolle) evolutionäre Änderungen hervorbringt. Diese äußerst allgemeine, umgreifende Idee hat den Vorteil, daß sie sich auf eine Vielzahl einzelner Fälle anwenden läßt, die z.B. in ihren Übertragungsmechanismen und auch sonst sehr stark differieren können, aber dadurch vereint sind, daß sie die Darwinschen Kriterien erfüllen.

Die Frage, die Darwins Theorie allerdings offen läßt, ist, *wie* die Darwinschen Krite-



rien erfüllt werden. Das heißt, sie macht keine Aussagen darüber, wie z.B. Eigenschaften übertragen bzw. vererbt werden. Darüber hinaus gibt es noch Fakten bei einigen spezifischen Klassen von Fällen, die von der Theorie nicht vorhergesagt werden, wie z.B. die rezessiven Eigenschaften bei Mendels Erbsen. Die Nachteile machen detailliertere Ansätze, wie die Mendelsche Genetik notwendig, die z.B. anhand von Paaren von Genen (Genotypen), in denen eins dominant sein kann, erklärt, wie es zu rezessiven Eigenschaften kommt. Nun ist natürlich auch Mendels Theorie nicht die unterste Sprosse der Erklärungsleiter, sondern wurde interessanterweise ursprünglich als Spezifikation der auf einer niedrigeren Ebene stattfindenden DNA-basierten Vererbung — also sozusagen als Spezifikation der Eigenschaften ihrer Hardware-Realisation — betrachtet. Diese Relation, in der Gene als, wie Dennett es auf den Punkt bringt, „the language of inheritance, straightforwardly realized in hunks of DNA“ (Dennett 1988, S. 385) betrachtet werden, entspricht nach Andy Clark dem, was er die *classicist vision* (Clark 1993, S. 43) der Beziehung zwischen abstrakteren kognitionswissenschaftlichen Theorien (Kompetenz-Theorien) und tatsächlichen Verarbeitungsstrategien nennt. Dennett schreibt weiter:

„There are theoretically important mismatches between the language of bean-bag-genetics' and the molecular and developmental details — mismatches serious enough to suggest that, all things considered, there don't turn out to be genes (classically understood) at all.“ (Dennett 1988, S. 385)

In den klassischen Kognitionswissenschaften werden nach Marrs (1982) mittlerweile zum Standard gewordenem Ansatz drei Ebenen des Verstehens unterschieden:

1. „*Computational Theory*“: Auf dieser Ebene soll beschrieben werden *was* berechnet wird und *warum*.
2. *Repräsentation und Algorithmus*: Hier wird beschrieben welche Berechnungsschritte erforderlich sind, um die Aufgabe zu lösen. Außerdem gibt es hier Details darüber, wie Input und Output repräsentiert sein müssen, so daß der Algorithmus die Transformation leisten kann.
3. *Implementation*: Diese Ebene zeigt, wie die Berechnungen durch die konkrete „Hardware“ einer realen Maschine durchgeführt werden können.

Der ideale Erklärungsansatz sollte nun eine Kaskade über diese drei Ebenen ermöglichen. In der Realität liegen allerdings Ausgangspunkt, Zentrum des Theoretisierens und in vielen Fällen auch der Endpunkt zwischen Ebene 1 und 2. Die Theorien auf dieser „Ebene 1.5“ (Peacocke 1986) sind vergleichbar mit Chomskys *Kompetenz-Theorien* (Chomsky 1986). Sie gehen über die Beantwortung der Frage, was berechnet wird, hinaus, indem sie nicht allein die Funktion zwischen Ein- und Ausgabe in ihrer Extension beschreiben, sondern zusätzlich Aussagen darüber machen, wie die Informationen, auf denen die Funktion operiert, strukturiert sind. Eine Erklärung auf der Ebene 1.5 für die Bildung von Vergangenheitsformen englischer Verben im Infinitiv würde also typischerweise auf

Konzepte wie *Stamm* und *Endung* zurückgreifen, ohne allerdings explizite Aussagen über den Algorithmus zu machen, der Stamm und Endung miteinander verbindet. Der Schritt zur Ebene 1.5 erscheint an dieser Stelle klein und völlig unverfänglich, da die Strukturierung der Information lediglich die umständliche Auflistung aller Paare von Infinitiven und Vergangenheitsformen erspart. Problematisch wird der Schritt erst dann, wenn man die eingeführten Konzepte nicht allein als instrumentelle Konstrukte gebraucht, die einem die Beschreibung der Daten erleichtern, sondern die damit eingeführte Strukturierung als psychologisch real betrachtet und damit den Schritt von einer Beschreibung externer, d.h. nicht-kognitiver Phänomene zu einer Theorie über Kognition macht.<sup>3</sup> Der Abstieg in der Erklärungshierarchie zur Ebene 2 ist dann ebenso leicht gemacht. Ein Algorithmus zur Berechnung der Vergangenheitsformen flektierter englischer Verben sähe z.B. folgendermaßen aus:

$$\text{IF (Stamm + Endung) THEN (Stamm + 'ed')} \quad (1.1)$$

Der Unterschied zwischen einer Level-1.5- und einer Level-2-Theorie ist dabei lediglich der, daß eine Level-1.5-Theorie ausschließlich explizite Aussagen über die strukturierten Repräsentationen macht, auf denen das zu beschreibende System operiert (Stamm, Endung, (Stamm + Endung), (Stamm + 'ed')), während die entsprechende Level-2-Theorie zusätzlich aussagt, daß das zu beschreibende System explizit die IF-THEN-Regel repräsentiert. Ob es sich bei der oben dargestellten Regel um eine Hypothese über Algorithmus und Repräsentation oder nur um eine Hypothese über Repräsentation handelt, ist also einstellungsabhängig.

Wichtig ist, festzustellen, daß die zu einer Level-2 Interpretation korrespondierende, stärkere Position („*propositional psychological Realismus*“), die eine explizite Existenz von Regeln postuliert, kaum (noch) vertreten wird.<sup>4</sup> Die hauptsächlich vertretene Version ist der zu einer Level-1.5-Erklärung korrespondierende *strukturell psychologische Realismus* (Clark 1989, S. 162), der lediglich die Existenz der strukturellen Beschreibungen (also z.B. Stamm, Endung, usw.) postuliert (Katz 1964; Pinker & Prince 1988; Fodor & Pylyshyn 1988). Besonders explizit sind hier Fodor und Pylyshyn:

„Classical machines can be *rule implicit* with respect to their programs. [...] What *does* need to be explicit in a classical machine is not its program but the symbols that it writes on its tapes (or stores in its registers). These, however, correspond not to the machine's rules of state transition but to its data structures.“ (S. 61)

---

<sup>3</sup>Zum Problem wird dies, meiner Beobachtung nach, auch dadurch, daß der Schritt, der auf der Beschreibungsebene/Erklärungsebene unbemerkt geschieht, vor allem unter Linguisten den Schritt auf der Objektebene oft genauso unbemerkt und unreflektiert nach sich zieht. Andy Clark bemerkt dazu, daß es sich bei dem, was „offiziell“ eine „Top-Level“-Theorie, also nach Marr eine Ebene-1-Theorie, sein sollte, es sich in der Praxis um Ebene 1.5-Theorien handelt (Clark 1993, S. 44).

<sup>4</sup>Möglicherweise hat Fodor selbst diese Position zunächst vertreten: „what distinguishes what organisms do [...] is that the representations of the rules they follow constitutes one of the causal determinants of their behavior.“ (Fodor 1975, S. 74)

Demnach ist die Explizitheit von Regeln kein Charakteristikum klassischer Kognitionswissenschaft, und es ist folglich ein Fehler, klassische Ansätze von z.B. konnektionistischen anhand der Explizitheit vs. Nicht-Explizitheit von Regeln zu unterscheiden.

## 1.5. Die „Language of Thought“-Hypothese

„Quatsch. Kannst du meine Gedanken lesen, Finne?“

...

„Gedanken lassen sich nicht lesen. Guck, du hast immer noch die Muster des gedruckten Worts in dir, obwohl du kaum sehr belesen bist.“

(William Gibson, *Neuromancer*)

Ein Charakteristikum klassischer kognitionswissenschaftlicher Theorien, oder vielmehr ihre konstituierende Eigenschaft ist dagegen die Isomorphie von Aufgabenanalyse, Kompetenztheorie und ihrer algorithmischen Implementation (siehe Beispiel 1.1). Diese Eigenschaft, Clark (1989, S. 18) nennt sie *semantische Transparenz*, ist darauf zurückzuführen, daß sowohl Aufgabenanalyse als auch Level-1.5- und Level-2-Theorie auf einer Art von Sprache beruhen. Als Apparat zur Beschreibung von Aufgaben auf der, wie Smolensky sie nennt, *konzeptuellen Ebene* (Smolensky 1988) dient gewöhnlich die natürliche Alltagssprache („Man hängt einfach `ed' an den Stamm an“). Für das tatsächliche „*in-the-head-processing*“, auf Ebenen 1.5 bis 2 wird dann die eine sogenannte *language of thought* („LOT“)<sup>5</sup>, die auf eben den strukturellen Einheiten operiert, die bei der Analyse auf der konzeptuellen Ebene festgelegt wurden. Das wichtigste Merkmal einer solchen linguistischen Theorie mentaler Repräsentationen ist, daß sie ähnlich wie natürliche oder formale Sprachen über eine kombinatorische Syntax und Semantik verfügt, so daß molekulare Repräsentationen aus Konstituenten aufgebaut werden, die wiederum zusammengesetzt molekular oder atomar sein können. Die Regeln, die auf diesen Konstituenten operieren, sind dabei rein syntaktisch, das heißt, sie beziehen sich ausschließlich auf die Form und nicht auf die Semantik bzw. den Inhalt der Ausdrücke. Außerdem sind diese Konstituenten „transportabel“, das heißt, sie können in verschiedenen mentalen Ausdrücken wiederverwandelt werden. Oder wie Fodor selbst es ausdrückt:

„[...] they can *share* the constituents they contain, since, presumably, the subexpression that denotes `foot' in `I raise my left foot' is a token of the same type as the subexpression that denotes `foot' in `I raise my right foot.'“

---

<sup>5</sup>Eine „Language of Thought“ ist keinesfalls mit privater Sprache zu verwechseln. Sie liegt auf einer viel tieferen Ebene und ist der Introspektion unzugänglich. Fodor selbst hat einmal zugegeben, daß, als er einmal so stark nachdachte wie er konnte, das einzige irgendwie linguistische was ihm bewußt war so etwas wie „C' mon, Jerry, you can do it!“ war. Auch wenn das in dem Augenblick seine „Gedanken“ waren, und sie ihm vermutlich bei der Lösung seines Problems irgendwie behilflich waren, war das sicherlich nicht die Art von Gedanken, die mittels Hypothesentests und Inferenzen und anderen postulierten Features einer LOT letztendlich sein Problem gelöst haben, oder auch nicht

(Similarly, *mutatis mutandis*, the ' $P$ ' that expresses the proposition  $P$  in the formula ' $P$ ' is a token of the same type as the ' $P$ ' that expresses the proposition  $P$  in the formula ' $P \& Q$ '.)“ (Fodor 1987, S. 137)

Die Vorteile, die man sich mit der Annahme einer LOT einkauft, liegen auf der Hand. In Alltagssprache formulierte, naiv psychologische Beschreibungen, wie z.B. „Peter geht zum Kühlschrank, weil er Durst hat und sich ein Bier holen will“ können quasi 1:1 in wissenschaftlich psychologische Theorien übersetzt werden. Es wird folgerichtig angenommen, daß die Zeichenketten innerer Symbole in einer so engen Beziehung zu den Inhalten unserer mentalistischen Alltagssprache stehen, daß erstens der innere Code „morphosyntaktisch ununterscheidbar von englisch“ („morphosyntactically indistinguishable from English“ (Fodor 1991, S. 316)) ist und zweitens, daß die Tatsache die Proposition „ich hebe linken meinen Fuß“ codiert in „mentalesisch“ meinem Absichtskästchen („*intention box*“, Fodor 1987, S. 136f.) zu haben, die *kausale* Kraft hat, zu verursachen, daß ich meinen Fuß hebe.

Was die Annahme einer linguistischen Theorie des Geistes außerdem noch attraktiv macht, ist, daß sich (putative) Eigenschaften des menschlichen Denkens wie *Produktivität* und *Systematizität* sehr leicht beschreiben lassen. Genauso wie natürliche Sprache produktiv ist (Chomsky 1957), das heißt, es gibt trotz eines endlichen Vokabulars unendlich viele wohlgeformte Sätze, und wir damit im Prinzip in der Lage sind, unendlich viele Sätze zu produzieren/ verstehen, können wir in unserer LOT qua ihrer rekursiv kombinatorischen Struktur *im Prinzip*<sup>6</sup> unendliche viele verschiedene Gedanken hegen. Ebenso, so die Theorie, ist analog zu unserer Fähigkeit, Sätze zu produzieren/ zu verstehen, auch unsere Fähigkeit Gedanken zu hegen systematisch. Das heißt, daß unsere Fähigkeit, einige Gedanken zu haben, *intrinsisch* mit der Fähigkeit verknüpft ist, viele andere Gedanken zu haben. Konkret: wer denken kann „Tarzan liebt Jane“ *muß* auch „Jane liebt Tarzan“ denken können. Außerdem bietet eine Sprache, dadurch, daß ihre Sätze aus kontextinvarianten Symbolen aufgebaut ist, die Möglichkeit, wahrheitserhaltende Inferenzregeln anzuwenden.

In diesem Zusammenhang ist es wichtig, festzustellen, daß die Argumentation für eine LOT nicht so verläuft, daß die Theorie deshalb wahr ist, weil man mit ihr die Beobachtungen am *leichtesten* beschreiben kann, sondern, um Fodor selbst zu Wort kommen zu lassen:

„Here's how I propose to argue:

- a There's a certain property that linguistic capacities have in virtue of the fact that natural languages have a combinatorial semantics.
- b Thought has this property too.
- c So thought must have a combinatorial semantics.“

---

<sup>6</sup>Einige Selbstzweifel bezüglich Sterblichkeit, usw. finden sich im mit „*Why There Still Has to Be a Language of Thought*“ betitelten Anhang Fodors „*Psychosemantics*“ (Fodor 1987).

Das grundlegende Problem an Fodors Argumentation ist, daß sie stark in die Nähe eines infiniten Regresses kommt, der auf der Verwechslung von Denken und der Zuschreibung von Denken („*thought ascription*“) beruht. Der Ursprung aller angesprochenen linguistischen Eigenschaften ist nicht das Denken, daß sich in unserem Handeln ausdrückt, sondern unsere (inner-)sprachlichen Zuschreibungen von Gedanken, die auf Interpretationen unseres Handelns beruhen, die aus diesem „Sinn“ machen sollen. Das heißt, daß der gewöhnliche Zugang zu unserem Denken, nämlich naiv-psychologische Interpretationen unseres Handelns — die Basis der „Language of Thought“-Hypothese (!) — keine anderen Schlußfolgerungen zuläßt als, daß unser Denken z.B. systematisch ist. Wenn nämlich jemand (z.B. Cheetah) denken kann, daß Tarzan Jane liebt, aber nicht denken kann, daß Jane Tarzan liebt (ob dies auch bei Menschen möglich ist, oder nicht, spielt keine Rolle), dann würden wir nicht daraus folgern, daß Cheetah nicht systematisch denkt, sondern ihr auch die Fähigkeit „Tarzan liebt Jane“ zu denken, absprechen, da Cheetah offensichtlich eins der dazu notwendigen Symbole oder Konzepte: „Tarzan“, „Jane“ oder „lieben“ fehlt. Das, was Fodor und Pylyshyn als *empirische* Beobachtungen darstellen, nämlich „nobody has a *punctuate* intellectual competence“ und „you don't find cognitive capacities that consist of the ability to think seventy-four unrelated thoughts“ (Fodor & Pylyshyn 1988, S. 42), sind also keine *empirischen* Fakten, sondern aus der Natur des Gedanken-Zuschreibens folgende, sogenannte *konzeptuelle* Fakten (Clark 1989, S. 147).

Daraus folgt, daß Mentalesisch und unsere naiv-psychologische Beschreibungssprache sich in ihrer Relation zu dem, was tatsächlich in unserem Kopf passiert, nicht unterscheiden, und — hier kommt der infinite Regreß ins Spiel — daß der Erklärungsgegenstand einer LOT nicht die Inhalte unseres Denkens sein können, sondern lediglich die zur Theorie gehörenden Equivalente zu den Inhalten naiv-psychologischer Alltagssprache. Der „Language of Thought“ fehlt demnach die innere „Verankerung“. Sie kann keine Naturalisierung der „*folk psychology*“ leisten.

Selbst wenn Systematizität nicht *notwendigerweise* lediglich ein Feature gedanken-zuschreibender Sätze ist, sollte klar geworden sein, daß das, was eigentlich erklärt werden sollte, das *Verhalten* ist, das diese Sätze gewährleistet. Eine „*Language of Thought*“ erklärt nicht die *Systematizität des Denkens*, sondern setzt diese voraus. An dieser Stelle laufen einige Stränge der Kritik am klassischen Paradigma zusammen. Erstens wird deutlich, daß der Schritt von einer „externalistischen“ Level-1.5-Beschreibung zu einer kognitiven Level-1.5 Theorie in Wirklichkeit sehr gewagt ist, da die Annahmen, die diesen Schritt unterstützen, nicht unzweifelhaft sind, und zweitens, daß die Kaskade klassischer Erklärungen möglicherweise schon auf Ebene 1.5 ein jähes Ende findet. Dies nicht nur weil jeder weitere Schritt biologisch unplausibel erscheint (siehe §2.2), sondern weil jeder weitere Schritt lediglich die Granularität des verwendeten Beschreibungsvokabulars verändern kann. Was aber immer zurückbleibt und damit den „Kellerboden“ bildet, ist eine postulierte angeborene Sprache. Zudem ist die Semantik dieser Sprache auf der tiefst möglichen Beschreibungsebene, wenn man *nicht* den *propositional psychologischen Realismus* vertritt, *dieselbe* wie der einer Erklärung auf Ebene 1.5. Um diesen Punkt voll genießen zu können, ist es unerlässlich, Smolensky zu zitieren:

„A symbolic model is a *system* of interacting processes, all with the same conceptual-level semantics as the task behavior being explained. Adopting the terminology of Haugeland (1978), this *systematic explanation* relies on a *systematic reduction* of the behavior that involves no shift of semantic domain or *dimension*. Thus a game-playing program is composed of subprograms that generate possible moves, evaluate them, and so on. In the symbolic paradigm these systematic reductions play the major role in explanation. The lowest-level processes in the systematic reduction, still with the original semantics of the task domain, are then themselves reduced by *intentional instantiation*: they are implemented exactly by other processes with different semantics but the same form. Thus a move generation subprogram with game semantics is instantiated in a system of programs with list-manipulating semantics. This intentional instantiation typically plays a minor role in the overall explanation, if indeed it is regarded as a cognitively relevant part at all. Thus cognitive explanations in the symbolic paradigm rely primarily on reductions involving no dimensional shift.“ (Smolensky 1988, S. 11)

Die Essenz dieses Punktes betrifft die *prinzipielle Erklärungsfähigkeit* klassischer Theorien — unabhängig von ihrer Wahrheit. Ich meine damit, daß erstens der Erklärungsgegenstand der falsche ist („thought ascriptions“ bzw. konzeptuelle Aufgabenanalyse anstatt Verhalten) und daß es sich zweitens nur in einem sehr schwachen Sinne um wirkliche *Erklärungen* handelt, da sich die „Erklärungen“ und ihr Gegenstand nicht wirklich voneinander unterscheiden (kein „*dimensional shift*“). Die Bedeutung des Punktes für die Praxis psychologischer Modellbildung werde ich nun versuchen, an einem Beispiel klarzumachen.

Die Zirkularität einer auf der „Language of Thought“ basierenden Erklärung ist besonders frappierend in der Psycholinguistik, wo letztendlich eine Sprache mit einer anderen Sprache „erklärt“ werden soll. In seinem Buch *Consciousness Explained* zeichnet Dennett eine Karrikatur Levelts Sprachproduktionsmodells (Dennett 1991 a, S. 234f.). Dennetts Ausgangspunkt ist eine Situation, in der sich der Konzeptualisator (siehe Abb. 1.1) dazu entscheidet, einen Sprechakt auszuführen, der den Gesprächspartner beleidigen soll, indem er einen Kommentar zur Größe seiner Füße abgibt. Der Konzeptualisator schickt also der nächst-niedrigeren Stufe in der Hierarchie, dem Formulator, das Kommando: „Tell this Bozo his feet are too big“. Der Formulator übernimmt den Job und versucht, die geeigneten Worte zu finden: ein Possesivpronomen der 2. Person Singular „*your*“; ein hübsches Wort für Füße, so wie „*feet*“; die richtige Pluralform des Verbs „to be“, nämlich „*are*“; und das geeignete Adverb und Adjektiv: „*too big*“. Diese Worte kombiniert der Formulator dann geschickt mit einem beleidigenden Unterton und führt aus:

„Your feet are too big!“

Doch war das nicht ein bißchen zu einfach? Wenn der Konzeptualisator seinen Befehl (Levelts *präverbale Botschaft* !) auf Englisch abgibt, bleibt dem Formulator außer ein

paar trivialer Justierungen ja kaum noch Arbeit! Welches Repräsentationsmedium der Konzeptualisator auch wählt, es muß geeignet sein, die notwendigen Inhalte zu transportieren und der Formulator muß es verstehen können. Levelt kommt zu dem Schluß, daß es eine Art von „Language of Thought“ sein muß (Levelt 1989, S. 73–74), vielleicht eine, die nur zur Anforderung von Sprechakten verwandt wird (S.74). Der Formulator erhält also die „präverbale Botschaft“, also eine detaillierte mentalesische Anweisung, eine englische Äußerung zu machen. Das läßt ihm schon ein wenig mehr Arbeit übrig, aber letztendlich verdeckt dies nur den drohenden Regreß. Dennett resümiert:

„How does the Conceptualizer figure out which words of Mentalese to use to give the order? There had better not be a smaller duplicate of Levelt's whole blueprint hidden in the Conceptualizer's *message generation* box (and so on, ad infinitum).“ (Dennett 1991a, S. 234)

Man muß Levelt an dieser Stelle zugutehalten, daß er den Leser frühzeitig warnt, daß der Konzeptualisator eine „reification in need of further explanation“ (Levelt 1989, S. 9) ist. Auch der Verdacht, daß in der „message generation box“ ein Geist wohnt, ist kein wirkliches Problem für sein Modell, da es sich mehr oder weniger explizit tiefere Ebenen der Sprachproduktion focussiert. Wahrscheinlich sollte Levelt sich hier trotzdem strikter an Fodors 1. Gesetz der Nicht-Existenz der Kognitionswissenschaft (siehe S. 8 oben) halten, nach dem die Untersuchung zentraler Prozesse keine respektable Wissenschaft hervorbringen kann. Fodor tut dies allerdings auch nicht immer und führt dadurch seine LOT-Theorie fast ad absurdum:

„While some interesting things have been learned about the psychology of input analysis — primarily about language and vision — the psychology of thought has proved quite intractable.“ (Fodor 1983, S. 126)

„The cost of not having a Language of Thought is not having a theory of thinking.“ (Fodor 1987, S. 147)

## 1.6. Die Computer-Metapher

Noch grundlegender für die Kognitionswissenschaft im klassischen Sinne als die Annahme einer „Language of Thought“ war die Übertragung der sogenannten Church/Turing-These (CTT) (Church 1956) auf Kognition und die dadurch eingeleitete Geburt der sogenannten *Computer-Metapher*.

### 1.6.1. Die Church/Turing-These

Church und Turing entwickelten Mitte der dreißiger Jahre sehr unterschiedliche Systeme (*Lambda-Kalkül* und *Turing-Maschine*) von denen sie jeweils — in etwa — behaupteten,

daß jede menschlich oder mechanisch mögliche Berechnung (eigentlich „*Computation*“) auch innerhalb ihrer Systeme berechnet werden kann. Diese These ist zwar nicht beweisbar, da man dazu alle „menschlich oder mechanisch“ möglichen Berechnungen aufzählen müßte. Dennoch spricht vieles für ihre Richtigkeit und damit dafür, daß alle möglichen „Berechnungen“, die Mathematiker, Logiker, Philosophen usw. bisher auf unterschiedlichste Weise charakterisiert und formalisiert hatten, von einer Turing-Maschine (Turing 1937) berechenbar und damit in dieser Hinsicht äquivalent sind (siehe auch Galton 1990). Außerdem folgt aus der These, die hier nicht angezweifelt werden soll, daß „*computation*“ die Manipulation von Symbol-Tokens basierend auf ihrer Form ist. Denn eine Turing-Maschine hat ein Band, auf das Symbole geschrieben sind; sie kann das Band vorwärts oder rückwärts bewegen oder anhalten; und sie kann die Symbole auf dem Band lesen oder überschreiben. Die Maschine ist dabei so konstruiert, daß das, was sie tut (lesen, schreiben, vorwärts, rückwärts, halt), dadurch determiniert ist, in welchem Zustand sie sich gerade befindet, in welche anderen Zustände sie übergehen kann und welches *Symbol* sie gerade gelesen hat.

### 1.6.2. Die „Physical-Symbol-System“-Hypothese

Trotz ihrer suggestiven Kraft hielt die CTT selbst nach der Entwicklung digitaler Computer und höherer Programmiersprachen nicht direkt expliziten Einzug in Psychologie.<sup>7</sup> Erst Fodor (1975), vor allem aber Newell und Simon (1976) und später Newell (1980) schafften ein *psychologisches*, theoretisch fundiertes Selbstbewußtsein der sich schon seit etwa 1960 entwickelnden „*Künstlichen* Intelligenz“. Newell und Simon übertrugen die Church/Turing-These konsequent auf Phänomene des Geistes (Glauben, Wissen, Hoffen, Fürchten, Planen, Erkennen, Sehen usw.), indem sie behaupteten, daß die wissenschaftliche Essenz menschlichen Denkens und menschlicher Intelligenz dieselbe ist wie die Essenz der Funktion einer Turing-Maschine oder eines gewöhnlichen Computers, nämlich: die Fähigkeit Symbole zu manipulieren — genauso wie die wissenschaftliche Essenz von Wasser H<sub>2</sub>O ist. Newells und Simons „*Physical-Symbol-System*“-Hypothese lautet folgendermaßen:

„[...] the necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system [...]“ (Newell & Simon 1976, S. 42)

Mit „general intelligent action“ soll dabei „the same scope of intelligence seen in human action“ impliziert sein. Außerdem soll die Hypothese nicht allein aussagen, daß Intelligenz durch eine universelle Maschine, also auch durch ein „Physical Symbol System“ (PSS) realisierbar ist, sondern eine spezifisch ontologische Aussage über die Natur intelligenter Systeme machen (siehe S. 42). Clark (1989, S. 12) expliziert letzteres in einer

---

<sup>7</sup>In der Psychologie herrschten noch Behaviourismus (Skinner 1933) und frühe Informationsverarbeitungstheorien vor; Ansätze die ohne interne Repräsentationen auskamen.



Version der Hypothese, die er „*The strong-physical-symbol-system (SPSS) hypothesis*“ nennt wie folgt:

“A virtual machine engaging in the von Neuman-style manipulation of standard symbolic atoms has the direct and necessary and sufficient means for general intelligent action.“<sup>8</sup>

Das bedeutet, daß es, um intelligentes Verhalten zu erklären bzw. zu *erzeugen*, nur des richtigen Computerprogramms bedarf. Oder anders formuliert: Ein geeignet programmierter Computer hat buchstäblich kognitive Zustände. Damit erklären solche geeigneten Programme menschliche Kognition (Hypothese der starken KI).

### 1.6.3. Das „Chinese-Room“-Gedankenexperiment

*„I imagine a future debate in which Professor Searle, staunch to the end, succumbs to the `mere imitation' of strangulation at the hands of an insulted and enraged robot controlled by the `mere imitation' of thought and emotion.“*  
(Hans Moravec)

In einer Reihe von Publikationen (1980, 1983, 1984, 1990, 1993) hat sich vor allem John Searle als hartnäckigster Gegner aller Formen von „Computation = Cognition“-Hypothesen etabliert. Grundlage bildet dabei immer sein inzwischen berühmtes „*Chinese-Room*“-Gedankenexperiment.<sup>9</sup>

Searles Gedankenexperiment lädt uns ein, uns vorzustellen, wie er, ein englischer Muttersprachler, der überhaupt keine Ahnung von Chinesisch hat, zusammen mit Körben voll chinesischer Symbole und einem in Englisch geschriebenen Regelwerk, das chinesische Symbole mit anderen chinesischen Symbolen korreliert, eingeschlossen in einem Zimmer sitzt. Die Personen außerhalb des Zimmers, allesamt Chinesen, stellen Searle Fragen, indem sie ihm Zettel mit chinesischen Symbolen ins Zimmer reichen. Searle manipuliert diese Symbole, die er nur anhand ihrer Form erkennen kann, daraufhin entsprechend seines Regelbuchs und reicht andere Zettel mit chinesischen Symbolen als Antworten heraus, so daß jeder außerhalb des Zimmers überzeugt ist, daß Searle Chinesisch versteht. Aber Searle versteht tatsächlich immer noch kein Wort Chinesisch und hat auch keine Möglichkeit, je einem der Symbole, die er manipuliert, Bedeutung zuzuweisen. Die Moral von der Geschichte ist, daß genauso wenig wie „Searle-im-Zimmer“ Chinesisch verstehen kann, ein Computer, auf dem z.B. Schanks Programm zum Verstehen von Geschichten (Schank & Abelson 1977) läuft, etwas über Restaurants verstehen kann.

---

<sup>8</sup>Ob es gerechtfertigt ist, Newell und Simon diese stärkere Version der Hypothese zuzuschreiben, diskutiert Clark auf S. 12–13. Unabhängig davon — zu diesem Schluß kommt auch Clark — beschreibt die SPSS-Hypothese sehr gut die Haltung vieler Philosophen, Psychologen und KI-Forscher, die sich in einem großen Teil klassischer Unternehmungen auf diesen Gebieten widerspiegelt.

<sup>9</sup>Wer das Gedankenexperiment kennt, sollte den nächsten Absatz überlesen..

Searle macht damit nicht lediglich einen Punkt gegen einen ultraorthodoxen, möglicherweise fiktiven, KI-gläubigen Strohmann, der die These vertritt, daß ein Computer, wenn er nur richtig programmiert ist, tatsächlich mentale Zustände hat. Searles Gedankenexperiment stellt außerdem, und das ist wichtig zu erkennen, auf Computer-Metaphern basierende *psychologische Erklärungsansätze* in Frage.<sup>10</sup> Denn wenn, „Computation  $\neq$  Cognition“ und damit „Computation“ kein hinreichendes Kriterium für Kognition ist, muß ein auf „Computation“ basierendes psychologisches Modell, wie adäquat es auch sein mag, das heißt selbst wenn es wie „Searle-im-Zimmer“ den Turing-Test (Turing 1964) besteht, nicht notwendigerweise *Kognition* erklären. Was ein solches Modell zunächst nur leistet, ist, Input/-Output-Verhalten zu *simulieren*. An darüberhinausgehenden Zusammenhängen zwischen Modell und psychologischer Realität läßt das „chinesische Zimmer“ zumindest intuitive Zweifel aufkommen.

#### 1.6.4. Funktionalismus

Einer der Gründe, warum laut Searle auf „Computation“ basierende Modelle nicht Kognition erklären können, ist, daß „Computation“ als unabhängig von einer spezifischen Hardware-Implementation definiert ist, wohingegen Geist, so Searle, entscheidend auf seiner physischen Instantiierung basiert. Das Argument richtet sich zum einen gegen einen Aspekt des „*Computationalism*“, der vor allem in der ihm verwandten, von Putnam (1960, 1967) geprägten, „(Turing-)Maschinen-Funktionalismus“-Theorie (siehe aber auch Fodor 1968a) explizit gemacht wird.<sup>11</sup> Entsprechend dem Maschinen-Funktionalismus, läßt sich Denken essentiell auf einer *funktionalen* (oder „*computationellen*“) Ebene völlig unabhängig von der physischen Materie beschreiben, in der diese Funktionen realisiert sind („we could be made of cheese and it wouldn't matter“, Putnam 1975, S. 291). Searle dagegen argumentiert, daß Kognition *kausal* genauso von den biochemischen Eigenschaften des Gehirngewebes abhängig ist, wie Photosynthese von biochemischen Eigenschaften gewisser anderer Substanzen. Eine formaler Erklärungsansatz für Kognition ist folglich genau so unsinnig wie eine entsprechende formale Photosynthese-Theorie:

„If subjected to sunlight, then produce energy.“

---

<sup>10</sup>Searle (1980) greift zwar nur die These der „Strong AI“ an, nach der der Geist ein Computerprogramm *ist*, und nicht die schwächere Position, nach der Computer ein nützliches Werkzeug in der Untersuchung des Geistes ist (Searle 1990, S. 20), aber es bleibt angesichts seiner Zweifel unklar, *zu was* dieser nützlich sein soll (siehe auch Harnad 1989, §2.5).

<sup>11</sup>Maschinen-Funktionalismus und — in geringerem Maße — einige andere Funktionalismus-Versionen, z.B. „*Homuncular Functionalism*“ (Fodor 1968b; Cummins 1983), haben viele direkte Bezüge zu „*Computationalism*“, „*Language of Thought*“-Theorien und Erklärungssebenen-Theorien à la Marr. Sie haben sich jedoch eine etwas andere Tradition und andere Schwerpunkte, wie z.B. das Geist-Körper-Problem, Qualia und Intentionalität. „The functionalist is in many ways the natural bedfellow of the proponent of the physical-symbol-system hypothesis.“ (Clark 1989, S. 21)

Diese Analogie scheint zwar überzeugend, die Ursache ihrer Überzeugungskraft ist meiner Ansicht nach jedoch nicht notwendigerweise, daß, wie Searle behauptet, Kognition wie Photosynthese sich bestenfalls auf der Ebene biochemischer Prozesse erklären läßt. Die Analogie deutet vielmehr an, daß die Granularität des klassischen Erklärungsvokabulars zu grob ist, so daß weder die entscheidenden Phänomene erklärt werden noch überhaupt Erklärungen im eigentlichen Sinne möglich sind („no dimensional shift“, siehe §1.5; Smolensky 1988). Searle macht also einen Punkt gegen den klassischen Kognitivismus/ Funktionalismus indem er seine Wahl der Erklärungsebene trifft. Ob es jedoch nicht möglich ist, eine formale („mikrofunktionalistische“) Beschreibung von Kognition zu finden, die die relevanten Kausalitäten erfaßt, bleibt als empirische Frage offen (siehe §2.3.1 oder auch Clark 1989, S. 33). Denn genauso wie für Photosynthese nicht alle Eigenschaften einer Pflanze kausal relevant sind, müssen nicht alle Eigenschaften des Gehirns notwendig für Kognition sein.

Aber das Photosynthese-Beispiel veranschaulicht den zweiten Teil von Searles Argumentation gegen die Eigenschaft der Implementationsunabhängigkeit symbolischer Systeme, bzw. ihres Mangels an *relevanten kausalen Eigenschaften*: *simulierte Photosynthese erzeugt keine Energie*, genauso wenig, wie die Simulation eines Waldbrands Hitze erzeugt. Es besteht also ein großer Unterschied zwischen Simulation und Realität bzw. Simulation und Duplikat. Dieses Argument reduziert sich allerdings auf die Frage, welches die *relevanten Interaktionen* des zu Simulierenden mit seiner Umgebung sind und auf welcher *Ebene* diese Interaktionen am besten beschreibbar sind. Die sich ergebende Frage lautet also konkret: Wenn die für Photosynthese relevanten Interaktionen einer Pflanze mit ihrer Umwelt sich am adäquatesten auf einer biochemischen Ebene erfassen lassen und die relevanten Interaktionsgegenstände Kohlenstoff, Kohlendioxid, Sauerstoff, Licht, Energie usw. sind,<sup>12</sup> was sind dann die für Kognition relevanten Interaktionsebenen und Interaktionsgegenstände eines Organismus mit seiner Umwelt? Daß zu einer adäquaten Erfassung der relevanten Eigenschaften von Kognition weder symbolisch verbales Verhalten als Interaktionsgegenstand noch eine symbolische Beschreibungsebene der Interaktion im allgemeinen ausreichen, suggeriert eindrucksvoll Searles „Chinese-Room“-Gedankenexperiment. Um die gestellte Frage positiv beantworten zu können, muß aber zunächst noch Searles drittes Argument gegen die „Cognition = Computation“-Hypothese berücksichtigt werden.

### 1.6.5. „Symbol-Grounding“

Searles wichtigster Einwand gegen die These der starken KI ist, daß Computerprogramme und die Repräsentationen, auf denen sie operieren, *symbolisch* sind, das heißt arbiträr in ihrer Relation zum Bezeichneten. Computer wissen nicht, auf welche Entitäten in der Welt die Symbole, auf denen sie operieren, referieren. Analog weiß „Searle-im-Zimmer“

---

<sup>12</sup>Die Menge der relevanten Interaktionsgegenstände ist also eine echte Obermenge der Menge der relevanten Ein- und Ausgaben.

nicht, auf was die Striche und Schnörkel, mit denen er arbeitet, referieren. Die Operationen auf den Symbolen sind jeweils rein syntaktisch (siehe §1.5) und nicht semantisch. Die Bedeutung der Symbole ist lediglich in weiteren „bedeutungslosen“ Symbolen verankert.<sup>13</sup> Der einzige Ort an dem diese Symbole, genauso wie Symbole in einem Buch, wirklich Bedeutung erlangen, ist der „Kopf“ der Person, die sie interpretiert. Steven Harnad hat das zugrundeliegende Problem in einer Reihe von Publikationen (1989, 1990, 1992, 1994, 1995) als das „*symbol grounding problem*“ bekannt gemacht:

„[...] the symbol grounding problem: The interpretation of a symbol system is not intrinsic to the system; it is projected onto it by the interpreter. This is not true for our thoughts. We must accordingly be more than just computers.“  
(Harnad 1994, S. 379)

„Now here is the critical divergence point between computation and cognition: I have no idea what my thoughts are, but there is one thing I can say for sure about them: They are thoughts about something, they are meaningful, and they are not about what they are about merely because they are systematically interpretable by you as being about what they are about. They are about them autonomously and directly, without any mediation. The symbol grounding problem is accordingly that of connecting symbols to what they are about without the mediation of an external interpretation.“ (Harnad 1992)

Harnads Vorschlag zur Lösung des Problems ist, die Hürde für einen „Cognizer“ von einem gewöhnlichen linguistischen Turing-Test ( $T^2$ ) auf einen „totalen“, „robotischen“ Turing-Test ( $T^3$ ) anzuheben, welcher die Turing-Ununterscheidbarkeit symbolischer *und sensomotorischer* Fähigkeiten erfordert. Für einen Computer bedeutet das, daß es nicht ausreicht, die richtigen Antworten auf seinem Display auszugeben, um Harnad von der Existenz seines Geistes zu überzeugen. Zusätzlich müßte er mit Sensoren und Aktuatoren ausgerüstet sein und durch sein *Verhalten* überzeugen.<sup>14</sup>

In Bezug auf die These der starken KI stimmt Harnad also Searle zu: Kognition ist nicht gleich „Computation“, bzw. Kognition ist mehr als „Computation“. Die Frage die sich für die Kognitionswissenschaften stellt, ist, welche Aspekte von Kognition man überhaupt erklären kann, wenn man allein symbolisches Verhalten betrachtet oder zumindest den Begriff Kognition gemäß dem 1. Gesetz der Nichtexistenz der KWn auf sehr interne Prozesse beschränkt. Dies ist nicht nur eine Frage der Begriffs- bzw. Interessendefinition („Hat Kognition mehr mit Schachspielen oder adaptiven Handeln in der Welt zu tun?“) sondern auch eine erkenntnistheoretische Frage. Wie das „Chinese-Room“-Gedankenexperiment zeigt, ist Kognition („strong AI“), bzw. eine Kognitionstheorie („weak AI“) durch symbolische Verhaltensdaten stark unterdeterminiert. Daß

---

<sup>13</sup>Dies ist ein weiter Aspekt des in Abschnitt 1.5 beschriebenen Regresses.

<sup>14</sup>Harnads Vorschlag ist gegenüber Searles Argumentation immun, da „Searle-im-Zimmer“, um den  $T^3$  zu bestehen, entweder sensorische Transduktoren und motorische Effektoren umfassen muß und somit nicht umhin kann zu verstehen oder nur einen Teil des Systems bildet und somit aufgrund des „Other Minds“-Problems (siehe Churchland 1984, S. 69) nicht wissen kann, ob das gesamte System versteht.

sich trotz der schlechten Datenlage, also gewissermaßen zufällig, innerhalb des klassischen Paradigmas „gute“ Kognitionstheorien entwickeln werden, halte ich angesichts der in diesem Kapitel vorgestellten Argumente für sehr unwahrscheinlich.

## 1.7. Zusammenfassung

In diesem Kapitel wurde dafür argumentiert, daß der traditionelle symbolische Ansatz psychologischer Modell- und Theoriebildung wesentliche Unzulänglichkeiten aufweist. Hauptsächlich wurde kritisiert, daß von sehr wahrscheinlich falschen Prämissen, nämlich *optimalem Design*, *starker Modularität*, *starkem Nativismus*, „*Implementationsunabhängigkeit*“ und einer *Language-Of-Thought* ausgegangen wird. Weiterhin wurde demonstriert, daß die *Voraussetzung* dieser Prämissen — unabhängig von ihrer psychologischen Realität — nur sehr oberflächliche Erklärungen zuläßt, die die Kognitionswissenschaft von relevanten Erkenntnissen aus benachbarten Wissenschaftsgebieten wie der Biologie isoliert und die Domäne von Vergleichsdaten ungerechtfertigt einschränkt.

## 2. Neuere Ansätze

### Einleitung

In diesem Kapitel sollen einige Alternativen zum klassischen Zugang zur Kognitionswissenschaft vorgestellt werden. Bei der Darstellung werden vor allem die Punkte berücksichtigt, die am symbolischen Ansatz kritisiert wurden. Das Ziel soll dabei weniger eine detaillgetreue Erörterung der Vor- und Nachteile der einzelnen Paradigmen in ihrem momentanem Stadium sein. Es soll vielmehr eine *Perspektive* aufgezeigt werden, wie in Zukunft ein Zusammenspiel dieser noch jungen Ansätze einen Fortschritt in der Kognitionsforschung bewirken kann. Aus diesem Grunde ist es notwendig, auch Forschungsansätze zu berücksichtigen, die sich heute nur entfernt mit „Kognition“ im herkömmlichen Sinne befassen.

Im Abschnitt zu „*Artificial Life*“ soll z.B. eine alternative Sichtweise von „Kognition“ vermittelt werden und untersucht werden, wie ausgehend von dieser Ergebnisse aus anderen Wissenschaftsbereichen genutzt werden können. Im Abschnitt zu „*Dynamischen Systemen*“ steht im Vordergrund zu zeigen, daß in bezug auf neurobiologische Plausibilität und repräsentationelle Vorurteilsfreiheit mit dem Konnektionismus das Ende des Weges noch nicht erreicht ist. Außerdem wird dort eine alternative Konzeptualisierung psychologischer Prozesse vorgestellt, die andere Erklärungen liefert als klassische Level-1.5-Theorien.

Im nun folgenden Abschnitt soll zunächst aber ein Überblick über den inzwischen etablierten Ansatz „*Konnektionismus*“ gegeben werden. Der Überblick wird sehr kurz sein, da einige wesentliche konnektionistische Ideen noch im Zusammenhang mit „*Artificial Life*“ und den „*Dynamischen Systemen*“ diskutiert werden. Eine detaillierte Analyse des heutigen Forschungsstands findet sich außerdem — am Beispiel von Sprachproduktionsmodellen — im zweiten Teil dieser Arbeit.

### 2.1. Konnektionismus

„In short, what are the relations between GOFAI and connectionism? Are they, as sometimes claimed, two different paradigms? Are they respectively 'False Starts and Real Foundations,' as the subtitle of a recent collection of essays on AI (Graubard 1988) would have it? Or are they complementary

partners in a common intellectual enterprise? And if cooperation is in question, are they strange bedfellows who will never find real ease with each other? Or are they siblings sprung from the same intellectual roots, sadly separated in childhood but destined to be united? (Hankies out for the happy ending.)“ (Boden 1991)

Die Frage nach der Beziehung zwischen GOFAI und Konnektionismus ist ontologisch nicht befriedigend zu beantworten, da der Begriff „Konnektionismus“ inzwischen für eine sehr heterogene Klasse von Modellen benutzt wird. Aus diesem Grund soll hier nur kurz auf die gemeinsame Motivation und Entwicklung der verschiedenen Spielarten eingegangen werden, um dann in den folgenden Abschnitten einzelne, wichtige ideosynkratische Aspekte genauer zu untersuchen und zu vergleichen.

### 2.1.1. Charakteristika

Der Terminus „konnektionistisches Modell“ wurde von Feldman (Feldman 1981; Feldman & Ballard 1982) geprägt. Er bezeichnete zunächst solche Modelle die durch die Struktur des Gehirns zumindest inspiriert (Norman 1986) sind. Es gibt also keinen atomaren Zentralprozessor, sondern viele, simple, parallel arbeitende, idealisierte neuronartige Einheiten: „Knoten“, von denen jeder einen Aktivierungswert besitzt und andere Knoten abhängig von diesem und den Leitungsstärken zu ihnen beeinflusst. Man unterscheidet vor allem zwischen sogenannten *lokalen* und sogenannten *distribuierten* oder „parallel-distributed-processing“ („PDP“) Netzwerken. In lokalen Modellen ist typischerweise der Aktivierungswert eines einzelnen Knoten interpretierbar, während in PDP-Modellen selbst „atomare“ Informationen über mehrere Knoten verteilt sind (s. §5.3.2.1).

### 2.1.2. Die Attraktivität des Ansatzes

Verantwortlich für die große Attraktivität einer konnektionistischen Theorie des Geistes im Vergleich zur traditionellen Sichtweise waren folgende Eigenschaften, bzw. Möglichkeiten konnektionistischer Modelle:

- Die *größere biologische Plausibilität* (siehe §1.1; Clark 1989; Dennett 1991b),
- die *Modellierbarkeit gradueller Aspekte menschlichen Denkens*,
- die *Modellierbarkeit non-verbaler Fähigkeiten* (McClelland, Rumelhart & Hinton 1986),
- die *Modellierbarkeit grundlegender kognitiver Fähigkeiten*, wie z.B. Mustererkennung (siehe §1.1; Dennett 1991b),

- die *Robustheit* konnektionistischer Netze gegenüber Beschädigung und unvollständigen oder verrauschten Daten,
- die Ermöglichung des *Verzichts auf „Humunculus“-Erklärungen* (siehe §1.5; Smolensky 1988; Sterelny 1990)
- die weitgehende *Vermeidung repräsentationeller Vorurteile* — also der Verzicht auf die „Language-of-Thought“-Prämisse und Symbole im allgemeinen (siehe §1.5; Rumelhart & McClelland 1986; Smolensky 1988),
- der *Verzicht auf Design-Vorurteile* bezüglich Optimalität (s. §1.1; Dennett 1991b) und Modularität (siehe §1.2; Plaut 1995) und
- die Möglichkeit der *Eliminierung von „Folk-Psychology“* in „wissenschaftlich“ psychologischen Erklärungen (siehe §1.5; Churchland 1981; Stich 1983; Clark 1987; Ramsey, Stich & Garon 1991).

### 2.1.3. Zur Existenz von Regeln und Symbolen

Trotz dieser grundlegenden Unterschiede zum klassischen Kognitivismus wurde der Konnektionismus zunächst weniger als konkurrierendes Paradigma verstanden, da dieser zunächst zur Erklärung „niedrigerer“ Aspekte von Kognition, wie „Pattern-Matching“ verwendet wurde (siehe z.B. McClelland & Rumelhart 1981). Dies änderte sich mit Rumelhart und McClellands (1986) Modell des Lernens von Vergangenheitsformen englischer Verben. Radikal daran war, daß die Existenz von Symbolen und Regeln auf „höchster Kognitionsebene“, also in einer Domäne angezweifelt wurde, die bis dahin souverän von Regeln und Symbolen regiert wurde:

„[...] lawful behavior and judgements may be produced by a mechanism in which there is no explicit representation of a rule“ (S. 217)

„We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a `rule' as anything more than a *description* of the language.“ (S. 267)

Dies bedeutete eine Umdefinition des Verhältnisses zwischen Symbolismus und Konnektionismus von „complementary partners“ zu „False Starts and Real Foundations“. Rumelhart und McClelland (1986) behaupteten, daß allein konnektionistische Erklärungen *genuin psychologisch* sind, während symbolische allenfalls *externe, nicht-kognitive* Phänomene beschreiben. Die Unterscheidung zwischen Regel-folgendem und durch Regeln beschreibbares Verhalten wird häufig illustriert, indem darauf hingewiesen wird, daß Planeten ihre Orbits *nicht* berechnen, selbst wenn *wir* in der Lage sind, dies zu tun (siehe z.B. Dennett 1986, S. 73–74).<sup>1</sup>

---

<sup>1</sup>Dennett selbst favorisiert allerdings eine stärkere Differenzierung.



Bezüglich der Rolle von Regeln und symbolischen Repräsentationen<sup>2</sup> in psychologischen Erklärungen und dem Verhältnis zwischen Konnektionismus und Symbolismus im allgemeinen gibt es unter den Vertretern des Konnektionismus allerdings unterschiedliche Standpunkte. Bechtel und Abrahamsen (1991, S. 226) unterteilen diese in drei Klassen:

1. Der *externalistische* Standpunkt: konnektionistische Netze interpretieren und produzieren *externe* Symbole, z.B. innere oder egozentrische Sprache. Auf diesen Symbolen operiert ein „*conscious rule interpreter*“ (Clark 1989, S. 137–139; Smolensky 1988). Dieser Standpunkt ist weitgehend kompatibel mit den anderen.
2. Der *kompatibilistische* Standpunkt: sowohl symbolische als auch subsymbolische Mechanismen sind notwendig zur Erklärung menschlicher Kognition. Typischerweise benutzen entsprechende Modellierungsansätze konnektionistische Netze, als inhaltsadressierbare Speicher, auf deren Basis Symbole von einem Regel-folgendem Mechanismus extrahiert und manipuliert werden (siehe z.B. Touretzky 1990).
3. Der *approximationistische* Standpunkt: konnektionistische Modelle sind der akurateste Ansatz zur Erklärung von Kognition. Symbolische Ansätze können bestenfalls approximative Erklärungen liefern.

#### 2.1.4. Kritikpunkte

Die Form der Kritik des symbolischen Lagers am Konnektionismus (siehe z.B. Fodor & Pylyshyn 1988; Lachter & Bever 1988; Pinker & Prince 1988; Levelt 1991), der sich hauptsächlich gegen den *approximationistischen* Standpunkt richtet, läßt sich auf folgendes Dilemma reduzieren:

*Entweder* konstituiert Konnektionismus ein neues aber inadäquates Modell der Kognition,  
*oder* er konstituiert ein adäquates Modell, indem er ein klassisches implementiert, aber ist deshalb nichts Neues.

Die Kritik an der psychologischen *Adäquatheit* konnektionistischer Modelle im allgemeinen bezieht sich vor allem darauf, daß in diesen — im Gegensatz zu symbolischen Modellen und zum menschlichen Denken, so Fodor und Pylyshyn (1988) — die Eigenschaften *Produktivität*, *Systematizität* und *Kompositionalität* nicht inhärent vorhanden sind. Weitergehend behaupten Fodor und Pylyshyn (1988), daß konnektionistische Systeme, die diese Eigenschaften aufweisen, nichts anderes sind als *Implementationen* klassischer symbolischer Systeme. Wobei die Ebene der Implementation für *psychologische* Erklärungen irrelevant ist.

---

<sup>2</sup>Daß auch Vertreter des Symbolismus keine explizit repräsentierten Regeln postulieren, wurde bereits in Abschn. 1.4 dargelegt.

Gegen diese Einordnung in die Implementationsebene innerhalb der Marr-Hierarchie (siehe §1.4) richtete sich vor allem Smolenskys (1988) Erwiderung. Er schlägt stattdessen eine Unterscheidung zwischen Erklärungen auf *konzeptueller* und *subkonzeptueller* Ebene vor und postuliert, daß eine konzeptuelle Analyse eines subkonzeptuellen Netzwerks sich stark von einer klassischen „conceptual-level“ Erklärung unterscheiden kann. Ob ein PDP-Netzwerk lediglich die Implementation eines symbolischen Modells ist, ist demnach eine empirische Frage.

Die Argumente gegen eine Voraussetzung von Produktivität und Systematizität als „a-priori Eigenschaft“ von Kognition wurden bereits in Abschnitt 1.5 ausführlich dargestellt. Wie konnektionistische Netzwerke Kompositionalität modellieren können, ohne dabei auf ein symbolisches Modell reduzierbar zu sein zeigte van Gelder (1990). Die übrige Kritik aus dem symbolischen Lager bezieht sich, wenn man polemische Angriffe wie z.B. Levelt (1991) vernachlässigt, auf spezielle Eigenschaften spezieller Modelle, die im einzelnen diskutiert werden müssen. Diese Kritikpunkte sollen in Abschnitt 5.3 innerhalb der Analyse eines konnektionistischen Sprachproduktionsmodells besondere Berücksichtigung finden.

In den folgenden zwei Abschnitten werden zwei Forschungsparadigmen vorgestellt, die sich typischerweise konnektionistischer Modellierungsmethoden bedienen, aber versuchen, die „konnektionistische Idee“ noch konsequenter in die Tat umzusetzen, als dies im „klassischen“ konnektionistischen Modellen getan wird.

## 2.2. „Artificial Life“ und Situierete Agenten

Als Antwort auf eine von biologischen und phänomenologischen Grundlagen abstrahierete, gewissermaßen in einem biologischen Vakuum agierende KI entwickelte sich Ende der achtziger Jahre unter der Schirm-Bezeichnung „*Artificial Life*“ eine Gegenbewegung zur traditionellen „Artificial Intelligence“. Die gemeinsame Idee der ansonsten höchst unterschiedliche Interessen umfassenden Bewegung — von DNA-Replikationsmodellen und sensomotorischer Aktivität bishin zu kollektiver Intelligenz und Populationsdynamik — ist, daß man intelligente Systeme nicht anhand ihrer subjektinternen, entkörperlichten, abstrakten Eigenschaften („*DAFT: disembodied abstract features of thought*“, Wheeler 1994b) verstehen kann, sondern daß man, um *Intelligenz* zu verstehen, die spezifisch physische Verkörperung dieser Systeme, ihre sensomotorische Verkopplung mit der Umwelt und die organisatorischen Möglichkeiten, die sich aus diesen Eigenschaften ergeben, untersuchen muß. Gemäß dem Motto „*Ask not what is inside your head but rather what your head is inside of*“ (Mace 1977, zitiert in Clark 1989, S. 66) sind also nicht mehr interne Repräsentationen sondern die *Interaktionen* mit der Umwelt Schwerpunkt des Interesses. In A-Life ist ein *autonomer Agent* ein vollständig integriertes, selbst-kontrolliertes, adaptives System, das sich während einer andauernden Interaktion mit seiner Umwelt aktiv so verhält, daß es bestimmte Ziele erreicht und die Chancen, in seiner dynamischen, unsicheren Umgebung zu überleben (um sich zu reproduzieren),

erhöht. Die A–Life–Methodologie besteht darin, komplette Kontrollsysteme für künstliche autonome Agenten („Animats“) zu entwickeln, wobei es sich um Roboter oder auch Simulationen handeln kann.

### 2.2.1. Biologische Plausibilität

Aus bereits in Abschnitt 1.1 dargelegten Gründen scheint es vernünftig anzunehmen, daß deliberatives Denken und linguistische Kompetenz beim Menschen auf früheren in bezug auf das „Überleben“ fundamentalen Fähigkeiten aufbauen. Unter dem Einfluß dieser Art von Gedanken beginnt die A–Life–Forschung nicht mit der Untersuchung allein dem Menschen vorbehaltener Aspekte von Intelligenz, sondern dem adaptiven Verhalten simpler situierter Agenten, die wahrnehmen und handeln. Diese Fokusverschiebung bedeutet zugleich eine Verlagerung auf biologisch fundierteres Terrain, die auch ausgenutzt wird, da die biologische „Realisierung“ der untersuchten Fähigkeiten nicht wie in der durch „Computationalism“ geprägten orthodoxen Kognitionswissenschaft als „Implementationsdetail“ betrachtet wird. Im Gegenteil, biologische Faktoren werden als unverzichtbar für das Verständnis der zur Frage stehenden Phänomene adaptiven Verhaltens betrachtet.

Die große Hoffnung, die man mit der A–Life Bewegung verbinden kann, ist nicht nur das Entstehen biologisch plausibler Theorien über Kognition zu ermöglichen, sondern auch eine Verankerung der Kognitionswissenschaft in der Biologie und damit eine gewissermaßen symbiotische Beziehung speziell zu den Evolutionswissenschaften, zur Ethologie, zur Neurobiologie und zur Verhaltensökologie. In bezug auf einige Teile der Kognitionswissenschaften hat sich diese Hoffnung bereits als berechtigt erwiesen. Insbesondere die Untersuchung von visuellen, motorischen Fähigkeiten und visuomotorischer Koordination im Rahmen der *Computational Neuroethology* (Cliff 1990, 1995) hat sowohl für die Biologie als auch für die Kognitionswissenschaften interessante Erkenntnisse hervorgebracht. Im Bereich „Vision“ z.B. haben die Arbeiten von Arbib (1987) über Frösche und Kröten, von Cliff (1992) über Schwebefliegen, von Franceschini, Pichon und Blanes (1992) über Fliegen und von Brooks (1989) über seine eigenen „Kreaturen“ gezeigt, daß visuell vermitteltes komplexes Verhalten weder hochauflösende Kameras noch interne Rekonstruktionen der Welt benötigt und daß auf Marrs (1982) Vorstellung basierende *Computer Vision* Systeme nicht nur mit Sicherheit biologisch und psychologisch inadäquat, sondern auch äußerst wenig effektiv sind. Die oben genannten Arbeiten zeigen aber vor allem auch eine Alternative. Sie zeigen, wie durch eine kleine Fokusverschiebung so viele gesicherte Daten, z.B. aus der Biologie, nutzbar gemacht werden können, daß der Suchraum nach einem erfolgreichen „Seh–System“ soweit eingeschränkt werden kann, daß man sich weder auf ungesicherte „a priori“-Annahmen noch auf sein Glück verlassen muß. Während Marrs Vorstellung allein durch seine eigene Ratio eingeschränkt ist (unter anderem auch, da sie sich durch die Konzentration auf das Artefakt „Szenenanalyse“ vom Geschehen in der Welt abkoppelt), sind die oben genannten Systeme und auch die Schlußfolgerungen, die aus diesen Systemen gezogen werden können, abgesehen von vielen Detail–Fakten z.B. durch folgende grundlegenden Erkenntnisse fundiert:

die Schwebefliege benutzt ähnliche visuelle Informationstechniken wie der Mensch (z.B. „Augenbewegungen“) aber sie hat nur ca. 100000 Neuronen; bei Fröschen und Kröten gibt es unterschiedliche perzeptuelle Mechanismen für verschiedene visuelle Stimuli (z.B. der Abstand von Beute und Hindernissen wird über unterschiedliche Verarbeitungskanäle extrahiert); Spezies, deren Entwicklungszweig sich vor über 600 Millionen Jahren vom menschlichen Entwicklungszweig getrennt haben, haben das selbe „Augen-Design“ wie wir, usw. Zusätzlich zu dem Vorteil einfacherer Validierbarkeit gewonnener Hypothesen wegen der geringen Komplexität der zugrundeliegenden Systeme und des größeren Arsenal an Untersuchungsmethoden („Tierversuche“) ergibt sich der für AL konstitutive Hauptvorteil aber aus der situativen Einbettung der AL-Systeme in eine Welt.

### 2.2.2. Situiertheit

Klassische Modelle und der größte Teil konnektionistischer Modelle können essentiell als Transformationen von Repräsentationen beschrieben werden: Eingabe-Daten werden dem Modell in einem bestimmten Format präsentiert, und das Modell wird als erfolgreich betrachtet, wenn seine Ausgaben als korrekte Repräsentationen des Resultats der Ausführung der gewünschten Transformation interpretiert werden können. In den meisten Fällen sind dabei Ein- und Ausgaberepräsentation vom Experimentator selbst bestimmt.<sup>3</sup> Wenn solche Modelle zum Verständnis biologischer Agenten oder zum Verständnis intelligenten Verhaltens im allgemeinen dienen sollen, muß angenommen werden, daß es im Prinzip möglich ist, eine Pipeline von solchen Transformatoren, von Sensoren und Aktuatoren zu konstruieren, daß das gesamte System adaptives bzw. intelligentes Verhalten produziert. Abgesehen von der Mißachtung der Wichtigkeit sensomotorischer Prozesse in der klassischen KI ist der wichtigste Aspekt dieser Annahme die Abhängigkeit von vermittelnden Repräsentationen, die möglicherweise nicht zu rechtfertigen sind; neuronale adaptive Verhalten generierende Pfade senso-motorische Pfade sind möglicherweise nicht in entsprechende repräsentationstransformierende Moduln partitionierbar:

„[. . .] when we `open up the black box' we may not find any patterns of activity identifiable as a representation in the conventional sense, and even if we do, there is no guarantee that they will be in strong accordance with representations chosen *a priori* by connectionist modellers.“ (Cliff 1995, §2)

Das soll nicht bedeuten, daß Repräsentationen unnötig sind, sondern lediglich, daß selbst bei konnektionistischen Modellen die Gefahr besteht, durch *a priori* Repräsentationsannahmen irregeführt zu werden. Ein sicherer Weg, dieses Risiko zu vermeiden, besteht darin, den gesamten Pfad vom sensorischen Input zum motorischen Output nachzubilden. Ein solches System ist dann erfolgreich, wenn es sich erfolgreich in seiner Umgebung verhält; seine Ausgaben sind unabhängig von „*a priori*“-Repräsentationsannahmen und „*a posteriori*“-Interpretationen (siehe §1.6.5).

---

<sup>3</sup>Dies trifft nur eingeschränkt für selbstorganisierende neuronale Netzwerke und in den wenigen Fällen biologisch inspirierter Ein- und Ausgaberepräsentationen zu.

Ein weiterer Aspekt des Situertheitsgedankens ist, daß die Ursache für komplexes, scheinbar intelligentes Verhalten nicht mehr in komplexen internen Prozessen gesucht wird, sondern in der Interaktion einfacher Prozesse mit einer komplexen Umwelt. Zum einen beruht diese Idee auf Erkenntnissen, wie z.B. daß der komplex erscheinende Weg einer Ameise über den Strand nicht Ausdruck eines komplexen Gehirns, sondern Ausdruck einer komplexen Umgebung ist und somit der Ausdruck biologischer Erkenntnisse allgemeinerer Natur, wie z.B. Clarks (1989, S. 64) *007-Prinzip*:

„*The 007 principle*. In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done.“

Angesichts einer komplexen, unvorhersagbaren, sich ständig verändernden Welt scheint es auch nicht sinnvoll bzw. sogar unmöglich, Handlungen weit voranzuplanen. Viel einfacher ist, es weitestgehendst auf Planungen und Repräsentationen zu verzichten und stattdessen auf die Welt direkt zu reagieren, denn: „*The world is its own best model*“ (Brooks 1991a, S. 583).

### 2.2.3. „Evolutionary“ Robotics

Wie bereits deutlich geworden sein sollte, ist eines der Haupt-Prinzipien des A-Life Ansatzes, die Modellierung von Kognition bzw. adaptiven Verhaltens von möglichst wenigen „*a priori*“-Annahmen abhängig zu machen. Der A-Life-Zweig, der diesem Grundsatz am meisten Rechnung zu tragen versucht ist „*Evolutionary Robotics*“ (Cliff, Harvey & Husband 1993; Nolfi & Parisi 1991). Die Methodik dieses Zweiges besteht darin, statt Systeme sozusagen mit der Hand zu designen, sei es auch mit einer biologischen Inspiration im Hinterkopf (wie z.B. Brooks (1986, 1990, 1991a, 1991b) mit seinen auf der „terminal addition“-Idee beruhenden hierarchischen Kontrollschichten), unter Ausnutzung von durch Ideen über Evolution inspirierten Prozessen automatisch zu generieren. Dies funktioniert ungefähr folgendermaßen: Zunächst muß ein Verfahren gefunden werden, mit dem man neuronale Netzwerke als Genotypen encodieren kann, um dann ausgehend von einer zufällig generierten Population von Kontroll-Netzwerken<sup>4</sup> und einer Evaluationsaufgabe einen Selektionszyklus so zu implementieren, daß die bei der Aufgabe erfolgreicheren eine proportional größere Chance haben, zum genetischen Material der folgenden Generationen beizutragen. Im Laufe der Generationen werden so immer erfolgreichere Kontroll-Netzwerke entwickelt. Der Clou des Verfahrens ist, daß die so evolvierten *Animats* nicht auf eine von Menschen bevorzugte Weise („Design“ oder „Behaviors“ — siehe §1.1 bzw. Brooks 1986-94) modular dekomponierbar sein müssen. Dies soll nicht heißen, daß sie notwendigerweise keine Form modularer Struktur aufweisen, sondern nur, daß diese nicht von vornherein festgelegt wurde.

<sup>4</sup>In ER-Ansätzen sind Sensoren und Effektoren natürlich Teil dieser Kontroll-Netzwerke.

#### 2.2.4. Implikationen für die Untersuchung höherer kognitiver Fähigkeiten

Natürlicherweise ist der Stand der Forschung in *A-Life* bzw. *Computational Neuroethology* noch weit davon entfernt, alle seine Prinzipien auf höhere kognitive Fähigkeiten anwenden zu können. Z.B. ist es nicht möglich, basierend auf gesicherten biologischen Daten komplette „sensory–motor pathways“ von Menschen oder anderen großen Säugtieren nachzubilden – die Forschung konzentriert sich auf Arthropoden und Amphibien. Dennoch sind *A-Life* und CNE äußerst relevant auch für das Verständnis speziell menschlicher Aspekte von Intelligenz. Zunächst einmal sind die in der CNE gewonnenen Erkenntnisse über Tier-Kognition, wie z.B. die oben zitierten über *Vision* alle mehr oder weniger direkt in bezug auf menschliche Kognition interpretierbar; vor allem in der Hinsicht, inwiefern *scheinbar* intelligentes Verhalten überhaupt auf sogenannte höhere kognitive Fähigkeiten zurückzuführen ist, bzw. zu welchem Teil komplexes menschliches Verhalten nicht aus reaktivem Verhalten gegenüber einer komplexen Umwelt gegenüber oder aus der Interaktion einfacher, entwicklungsgeschichtlich früherer Fähigkeiten emergiert. Selbst wenn sich herausstellt, daß große Teile menschlicher Intelligenz so nicht zu erklären sind, ermöglicht erst der Vergleich z.B. mit Tier-Intelligenz, vorurteilsfrei zu definieren, was speziell menschliche Intelligenz ist und was somit andere Erklärungen erfordert. Wie zu diesen Erklärungen durchaus auch evolutionspsychologische und ethologische Forschung beitragen kann, zeigen z.B. Noble und Cliff (1996) in Bezug auf die Evolution von Kommunikation, Arbeiten speziell von Miller (1993, 1996) über den Zusammenhang zwischen der Entwicklung kreativer menschlicher Intelligenz und einer sogenannten „*runaway sexual selection*“ oder indirekt auch die Forschung von Cheney und Seyfarth (1990) über die Kommunikation von Vervet-Affen in ihrer natürlichen Umgebung.

Ein Beispiel für die konkrete Anwendung vieler CNE-Prinzipien auf die Untersuchung höherer kognitiver Fähigkeiten ist das Cog-Projekt am MIT, daß sich zum Ziel gesetzt hat, einen humanoiden Roboter inspiriert von Theorien über Phylo- und Ontogenese zu konstruieren (siehe Brooks 1994; Brooks & Stein 1993; Rutkowska 1995).

### 2.3. Allgemeine Dynamische Systeme

„Most naturally occurring phenomena are most successfully described as dynamical systems, so why shouldn't be cognition?“

„Dynamics is a field emerging somewhere between mathematics and the sciences. In our view, it is the most exciting event on the concept horizon for many years. The new concepts appearing in dynamics extended the conceptual power of our civilization and provide new understanding in many fields.“  
(Abraham & Shaw 1992)

Eine Gruppe mathematischer Konzepte, Methoden und Resultate bekannt unter dem Namen „Theorie dynamischer Systeme“, die bisher vor allem in der Physik Anwendung fand<sup>5</sup>, beginnt nun eine immer wichtigere Rolle in den Kognitions- und Neurowissenschaften zu spielen. Ein dynamisches System ist nach Wheeler (1994a, S. 7) essentiell ein System, für das es wenigstens potentiell die Möglichkeit einer rigorosen Analyse seiner Entwicklung über die Zeit gibt. Formal ist dies jedes System, für das man (1) eine endliche Anzahl von Zustandsvariablen, die — bezüglich des Interesses des Beobachters — den Zustand eines Systems zu einer gegebenen Zeit adäquat erfassen, und (2) eine Menge von Zustands-Evolutions-Gleichungen finden kann, die beschreiben wie sich die Werte der Zustandsvariablen über die Zeit ändern.<sup>6</sup> (1) ermöglicht die Konstruktion des sogenannten *Zustandsraums* des Systems, eines geometrischen Modells der Menge aller möglichen Zustände des Systems. (2) und zusätzlich ein Initialzustand, gegeben als Punkt im Zustandsraum, ermöglichen die Darstellung der Veränderung des Systems über die Zeit als Kurve im Zustandsraum. Eine solche Kurve wird *Trajektorie* des Systems genannt. Die Menge aller solcher Trajektorien heißt *Phasen-Portrait* des Systems. Ein *Attraktor* ist ein Zustand des Systems zu dem hin nah entlang verlaufende Trajektorien tendieren zu konvergieren. Nehmen wir nun einen beliebigen Attraktor P. Die Menge der Zustände, so daß, wenn sich das System in einem dieser Zustände befindet, das System zum Attraktor P konvergiert, heißt *Attraktionsbecken* von P. Die Trajektorien, die durch Punkte des Attraktionsbeckens auf ihrem Weg zum Attraktor verlaufen, aber nicht auf dem Attraktor selbst liegen, heißen *Transienten* des Systems. Innerhalb eines Zustands-Raums kann es mehrere Attraktoren geben, die wiederum verschiedenen Typs sein können. *Punkt-Attraktoren* sind z.B. einzelne Punkte im Zustands-Raum, die konstante Lösungen des Systems repräsentieren, während *periodische Attraktoren* oszillatorische Lösungen repräsentieren. *Chaotische Attraktoren* repräsentieren das komplexe Verhalten eines Systems, wenn dieses *sensitiv abhängig von Anfangsbedingungen* ist. Innerhalb eines solchen Systems nehmen zwei von beliebig nah aneinander liegenden Initialzuständen ausgehende Trajektorien im Mittel exponentiell divergierende Wege durch den Zustandsraum des Systems und bleiben dabei ohne sich zu kreuzen an den Attraktor gebunden. Sie falten sich also auf sich selbst zurück und schaffen damit einen unendlich geschichteten chaotischen Attraktor. Solche Attraktoren sind häufig in hochdimensionalen nicht-linearen Systemen.<sup>7</sup>

### 2.3.1. Dynamische Systeme und Konnektionismus

Auch bei komputationellen Systemen (im Sinne von Turing Maschinen — siehe §1.6) handelt es sich um dynamische Systeme. Allerdings füllt die Menge solcher Systeme nur eine extrem kleine Ecke im Raum aller möglichen dynamischen Systeme (siehe van

---

<sup>5</sup>Es gibt aber auch eine enge Verwandtschaft zur Systemtheorie im allgemeinen und zum *Radikalen Konstruktivismus* (siehe Maturana und Varela (1988) und Roth (1986, 1987)).

<sup>6</sup>Eine detaillierte Einführung in die Konzepte der Theorie Dynamischer Systeme findet sich in Norton (1995).

<sup>7</sup>siehe z.B. Devaney (1986) oder, für eine „freundliche“ Erklärung, Abraham und Shaw (1992).

Gelder 1992). Die Beschreibungssprache allgemeiner dynamischer Systeme stellt also ein wesentlich reicheres konzeptuelles Gerüst zur Verfügung als dies orthodoxe Ansätze tun. Die Frage, um die es zunächst gehen soll ist, ob konnektionistische Netzwerke außerhalb der vom klassischen Ansatz besetzten Ecke liegen.

Bei typischen distribuierten neuronalen Netzwerken legt ein Mensch Input-Daten an das System an. Dies setzt den Zustand des Netzwerks auf seinen Ausgangspunkt im Aktivationsraum — einem Zustands-Raum der so viele Dimensionen hat wie das Netz Knoten. Wenn nun das Netzwerk erfolgreich trainiert wurde, liegt dieser Ausgangspunkt im Attraktionsbecken eines Punkt-Attraktors, der bezüglich einer geeigneten (externen) semantischen Interpretation die korrekte Lösung repräsentiert. Die sukzessiv aufeinanderfolgenden Zustände des Systems auf seinem Weg zum Punkt-Attraktor, an dem das System zum Stillstand kommt, lassen sich als Trajektorie interpretieren. Typische PDP-Netzwerke lassen sich offenbar ohne Probleme sowohl mit Hilfe orthodoxer Begriffe wie auch dynamischer Begriffe beschreiben. Die Frage ist nur, warum die Aktivationsraum-Dynamik künstlicher neuronaler Netze auf unpertubierte Trajektorien und Punktattraktoren beschränkt sein sollte? Sicher ist, daß im Zuge der Konnektionismuseuphorie des letzten Jahrzehnts die biologische Adäquatheit konnektionistischer Netze vielfach überbewertet wurde. Zwar ähneln Standardnetzwerke der abstrakten Struktur des Gehirns, andererseits legen z.B. Beer (1992), Cliff, Husbands und Harvey (1993), Husbands, Harvey und Cliff (1995) sowie Skarda und Freemans (1987) „*How Brains Make Chaos to Make Sense of the World*“<sup>8</sup> die Idee nah, daß gerade die meist nicht berücksichtigten Eigenschaften realer neuronaler Netze (z.B. ständiges Hintergrundrauschen, kontinuierliche Evolution in der Zeit, zeitverzögerte Aktivierung abhängig vom Abstand der Einheiten, nicht-uniforme Aktivierungsfunktionen usw.) entscheidend für die Erklärung kognitiver Fähigkeiten sein können. Wie bereits in §1.6.4 deutlich gemacht wurde, soll dies nicht bedeuten, daß jegliche Abstraktion unzulässig ist, es geht vielmehr darum herauszufinden, von welchen Eigenschaften man abstrahieren kann und welches die Schlüsseigenschaften in der Dynamik biologischer neuronaler Netze sind. Welches diese Schlüsselaspekte sind, sollte allerdings durch empirische Forschung entschieden werden und nicht durch philosophische Vorurteile. Das konnektionistische Lager muß sich in dieser Hinsicht denselben Vorwurf gefallen lassen, den sie dem orthodoxen Lager macht, nämlich, daß es von empirisch nicht gerechtfertigten Annahmen ausgeht.

### 2.3.2. Die Dynamik situierter Agenten

Die Beschreibungsvokabular dynamischer Systeme wird häufig zur Analyse von Verhalten evolvierter Agenten im Bereich der Evolutionären Robotik angewandt, da dort in der Regel eine Oberklasse von klassischen neuronalen Netzwerken als Ausgangs-Kontrollstruktur benutzt wird. Ein weiterer Vorteil, die Theorie dynamischer Systeme für die Modellierung

---

<sup>8</sup>Skarda und Freemans Modell der Geruchserkennung und -Unterscheidung bei Hasen impliziert, daß die chaotische Dynamik des olfaktorischen Systems es ermöglicht, neue Gerüche zu den bereits gelernten hinzuzufügen, ohne daß die neuen die alten überlagern oder zu diesen konvergieren.



und Analyse situierter Agenten zu verwenden ist, daß sie gewissermaßen umsonst eine Konzeptualisierung und eine mathematische Beschreibungstechnik der in A-Life fokussierten Interaktion zwischen Agent und Umwelt liefert: Agent und Umwelt werden als *gekoppelte dynamische Systeme* betrachtet. Ein „evolutionary–robotics“-Szenario sähe dann so aus, daß durch die sensomotorische Aktivität des *Animats* die Dynamik seines Nervensystems ständigen, sogenannten *Perturbationen* unterworfen ist, deren Art durch den simulierten Evolutionsprozeß bestimmten Agenten–Umgebungs–Kopplungs–Eigenschaften festgelegt ist.<sup>9</sup>

## 2.4. Ein Versuch, das Puzzle zusammenzusetzen

Ziel dieses Kapitels war, einige Perspektiven aufzuzeigen, auf welche Weise der Kognitionswissenschaft neue Potentiale des Erkenntnisgewinns erschlossen werden können. Die drei dargestellten Ansätze sollten dabei nicht als Alternativen verstanden werden, sondern zusammen ein Bild vielversprechender Methoden und Prinzipien ergeben. Innerhalb dieses Bildes stellen *konnektionistische Netzwerke* und *allgemeine dynamische Systeme* die Methode der Modellierung dar, die im Gegensatz zu klassischen, symbolischen Modellen an der Struktur des Gehirns orientiert sind, Erklärungen im eigentlichen Sinne zulassen und auf die Annahme einer „Language of Thought“ verzichten können. Der Abschnitt zu „A-Life“ sollte vor allem gezeigt haben, daß Kognition kein rein mentales Phänomen ist, sondern wesentlich durch die Interaktion eines Organismus mit seiner Umwelt bestimmt ist. Im Abschnitt zu *dynamischen Systemen* wurde dann skizziert, wie sich Prozesse in konnektionistischen Modellen auf einer konzeptuellen Ebene mathematisch beschreiben lassen und wie sich die Interaktion von simulierten Agenten mit ihrer Umwelt formalisieren läßt.

---

<sup>9</sup>Bei Wheeler (1994a) findet sich eine relativ informelle Beschreibung einer solchen Kopplung.

## **Teil II.**

# **Modellierung der Sprachproduktion**

### 3. Die Entwicklung

Auch in der Modellierung der Sprachproduktion vollzog sich in den letzten Jahren eine Entwicklung von der Vorherrschaft rein symbolischer Ansätze (Garrett 1976, 1980, 1988; Levelt 1989) über lokal konnektionistische „*spreading activation*“ Modelle (Dell & Reich 1980; Stemberger 1982, 1985; Harley 1984; Dell 1986, 1988; MacKay 1987; Schade 1992) bis hin zu ersten parallel distribuierten Ansätzen (Dell et al. 1993). Dennoch unterscheidet sich diese Entwicklung von der in anderen Zweigen peripherer Psychologie durch die relativ früh erlangte und andauernde Wichtigkeit gerade lokaler konnektionistischer Ansätze. Eine der Ursachen dieses Unterschieds ist, daß sich im Gegensatz z.B. zur Sprachrezeption die normale Produktion von Sprache auf der zumeist avisierten lexikalischen Ebene sehr leicht „erfolgreich“ simulieren läßt. Wie bereits in Abschnitt 1.5 beschrieben, ist es trivial ausgehend von der Computer–Metapher einen Rechner so zu programmieren, daß er aus einer internen symbolischen Repräsentation von „Your feet are too big.“ die Symbolfolge „Your feet are too big.“ generiert. Da also „Ausgabeadäquatheit“ als Qualitätskriterium nicht ausreicht, müssen andere Domänen von Vergleichsdaten genutzt werden, um die Freiheitsgrade möglicher, alternativer (equiparametrischer) Theorien zu reduzieren. Als übliche Methode bietet sich an, weitere Daten mittels psychologischer Experimente zu gewinnen. Levelt und Mitarbeiter (Levelt et al. 1991; Schriefers, Meyer & Levelt 1990) schlagen z.B. diesen Weg ein, um ihre Theorie des Lexikonzugriffs zu bestätigen und interaktive konnektionistische Modelle zu widerlegen. Eine sehr viel einfachere Möglichkeit ist jedoch (wie z.B. Dell & Reich 1980; Schade 1992), das bereits vorhandene Wissen über die Systematik von Versprechern (Meringer & Mayer 1895; Meringer 1908; Garrett 1980) auszunutzen und Modelle zu konstruieren, die sowohl normale als auch in diesem Sinne gestörte Produktion erklären und somit durch die Versprecherdaten zusätzlich restringiert sind. Für klassische, regelbasierte Ansätze scheidet diese Möglichkeit prinzipiell aus, da man, um Fehler zu ermöglichen, Regeln einführen müßte, die *gelegentlich* Fehler erzeugen (siehe auch §1.5, §1.4). Weitere Daten, die für klassische Modelle im Gegensatz zu konnektionistischen nicht oder nur wenig nutzbar sind, liefert die Untersuchung der Sprachentwicklung (siehe z.B. Stemberger 1989, 1992) und in zunehmendem Maße die Untersuchung aphasischer Sprache (Stemberger 1984; Harley 1990, 1993; Schade 1995). Aufgrund dieses großen, leicht nutzbaren Datenpotentials konnten sich sehr schnell sehr detaillierte konnektionistische Sprachproduktionsmodelle entwickeln, deren, zumeist durch Computersimulationen gewonnene, Vorhersagen entsprechend der sogenannten „experimentell–simulativen Me-

### 3. Die Entwicklung

---

thode“ (vgl. Rickheit & Strohner 1993, §15.2; Eikmeyer & Schade 1993; Schade 1995, §1.2.2) weiterführende empirische Fragestellungen lieferten. Die Rolle symbolischer Ansätze ist heute unterhalb der Phrasensyntaxebene auf die Untersuchung modularer Dekompositionen des Sprachproduktionsapparats (siehe §1.1, §1.2.2) und die Untersuchung der Struktur von Repräsentationen beschränkt.

Das relativ späte Aufkommen von PDP-Modellierungsansätzen ist vermutlich auf die frühe Präsenz lokal konnektionistischer Modelle und der damit verbundenen geringen wissenschaftspolitischen Dringlichkeit zurückzuführen. Mitverantwortlich ist dabei auch, daß das für die Sprachproduktion entscheidende Problem der *Sequentialisierung* (Lashley 1951; Eikmeyer & Schade 1991; Bruce 1994; Dell, Burger & Svec 1995) auf „genuin“ konnektionistische Weise zuerst bei Elman (1990)<sup>1</sup> zufriedenstellend gelöst wurde,<sup>2</sup> während es innerhalb lokaler Ansätze zunächst vernachlässigt werden konnte (siehe z.B. Stemberger 1982; ein Überblick findet sich bei Schade 1992, §3.4.2). Der erste und bisher einzige PDP-Ansatz zur Sprachproduktionsmodellierung soll in den folgenden Abschnitten genauer betrachtet werden.

---

<sup>1</sup>Siehe aber auch Jordan (1986), Cleermans, Servan-Schreiber und McClelland (1989), Pearlmuter (1989).

<sup>2</sup>McClelland und Rumelhart (1981) z.B. umgehen dieses Problem noch auf sehr unbefriedigende Weise, indem sie für Laute an verschiedenen Positionen verschiedene Laut-Knoten einführen ( $/l_1/, /l_2/, \dots /l_n/$ ), bzw. indem sie auf der Input-Seite Zeit räumlich enkodieren (siehe auch Elman 1995, §8.3; Elman 1990; Schade 1992, S. 93).

## 4. Ein PDP–Ansatz zur Erklärung von „Frame–Constraints“

### 4.1. Gegenstand der Analyse

Im Hinblick auf die im ersten Teil dieser Arbeit skizzierten Entwicklung in den Kognitionswissenschaften von LOT–basierten Theorien zu einem immer stärkeren Verzicht auf „a priori“ Annahmen liefert Dell et al. (1993) sozusagen den „modernsten“ Beitrag zur Sprachproduktionsforschung. Dells Modell ist allerdings nicht als vollständiges Sprachproduktionsmodell zu werten, da es auf die Produktion von 3-Segment–Wörtern beschränkt ist und explizit zunächst nur eine alternative Theorie zu sogenannten „frame constraints“ (s.u.) unterstützen soll. Ich werde daher nach einer Darstellung des Originals meine eigene Erweiterung vorstellen. Das erweiterte Modell werde ich dann auf seine Beschreibungs- und Erklärungsadäquatheit bezüglich Versprecherdaten untersuchen, um zuletzt die Ausbaufähigkeit des Ansatzes in Hinblick gängige Modellbewertungskriterien (vgl. etwa Thagard 1988; Jacobs & Grainger 1994) und die im ersten Teil dieser Arbeit dargestellten Prinzipien zu diskutieren.

### 4.2. „Frame Constraints“

Eine Standardannahme der verschiedenen Theorien der Sprachproduktion ist, daß die Generierung einer Äußerung einen Prozeß erfordert, der getrennt repräsentierte linguistische Strukturen und linguistische Inhalte zusammenfügt. Auf der syntaktischen Ebene wird z.B. angenommen, daß zunächst eine Phrasenstruktur des Satzes anhand semantischer Repräsentationen aufgebaut wird. Dieses Gerüst („Frame“) enthält freie Positionen — sogenannte „Slots“ — für Lexeme bestimmter grammatischer Kategorien. Die geeigneten Lexeme werden dann aus dem mentalen Lexikon abgerufen und in die Slots gefüllt. Es gibt also eine Trennung zwischen Struktur (dem Frame) und Inhalt (den Lexemen).

Ein analoger Frame–Slot–Mechanismus wird sowohl in konnektionistischen (Berg 1988; Dell 1988; MacKay 1987; Schade 1992; Stemberger 1985) wie auch in eher symbolischen Ansätzen (z.B. Meyer 1990; Shattuck-Hufnagel 1979) für die phonologische Realisierung von Lexemen angenommen: Ein phonologischer Frame (wie in 4.1 abgebildet) wird zunächst entweder nach den phonotaktischen Regeln der jeweiligen Sprache

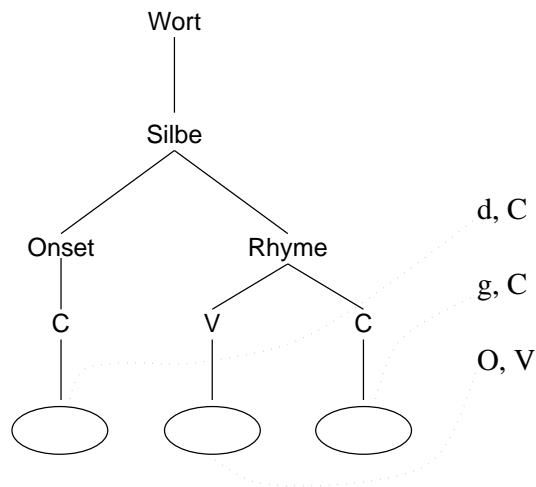


Abbildung 4.1.: Lexemrepräsentation mit Struktur-Inhalt-Trennung

zusammengesetzt oder komplett aus einem Inventar abgerufen. Danach werden phonologische Segmente in den Frame eingefügt. Diese Theorien postulieren also eine Trennung zwischen der phonologischen Struktur und dem phonologischen Inhalt eines Wortes.

Dell et al. (1993) versuchen zu zeigen, daß viele der psychologischen Daten, von denen angenommen wurde, daß sie eine phonologische Struktur-Inhalt-Trennung unterstützen, durch einen Mechanismus erklärt werden können, der ohne eine solche explizite Trennung auskommt. Insbesondere soll demonstriert werden, daß ein PDP-Mechanismus, der von lexikalischen Repräsentationen auf Sequenzen phonologischer Features abbildet und keine Trennung zwischen phonologischen Frames und phonologischen Segmenten aufweist, eben das Versprecher-Profil erzeugt, das bisher als starke Unterstützung einer solchen Trennung gewertet wurde.

#### 4.2.1. Versprecherdaten

Die Postulation von Frame-Slot-Mechanismen stützt sich wie gesagt im wesentlichen auf Versprecherdaten, genauer gesagt auf sogenannte phonologische Frame-Constraints, die eine generalisierende Interpretation von Versprecher-Phänomenen darstellen (vgl. Level-1-Theorie, 1.4!). Dell et al. (1993) versuchen, mittels ihres Modells zunächst die Frame-Constraints von Versprechern zu erklären, die keine Segment-Bewegungen involvieren. Die Versprecherklassen Antizipation, Perseveration und Vertauschung sind also nicht Gegenstand der Modellierung. Die für die sogenannten „nonmovement errors“ relevanten Constraints sind in Tabelle 4.1 dargestellt.

4. Ein PDP-Ansatz zur Erklärung von „Frame-Constraints“

---

<b>Effekt</b>	<b>Erläuterung</b>	<b>Standardinterpretation</b>
<i>Phonotaktische Regularität</i>	Phonologische Versprecher erzeugen fast immer in der jeweiligen Sprache vorkommende Lautsequenzen (siehe z.B. Wells 1951).	Phonotaktisch irreguläre Sequenzen werden durch die Anwendung von Frames und Regeln ausgeschlossen.
<i>Kategorie-Erhaltung</i>	Fehler- und Ziel-Segmente gehören praktisch immer derselben Konsonant-Vokal-Kategorie an.	V-Slots und C-Slots akzeptieren nur entsprechende V- bzw. C-Segmente
<i>Silben-Konstituenten-Effekt</i>	Wenn aneinander angrenzende Vokale und Konsonanten in einem Versprecher gemeinsam ersetzt werden, handelt es sich im Englischen wahrscheinlicher um eine VC- als um eine CV-Kombination (vgl. Nootboom 1969; Shattuck-Hufnagel 1983; Stemberger 1983).	CVC-Silbenframes weisen eine strukturelle Trennung zwischen Onset und Rhy-me auf (siehe Abb. 4.1).
<i>Initialeffekt</i>	Konsonanten am Wort- und Silbenanfang fallen häufiger Versprechern zum Opfer als Konsonanten an anderen Positionen.	Aufgrund der Onset-Rhyme-Struktur von CVC-Silben-Frames, sind Onset Konsonanten leichter abtrennbar (siehe Abb. 4.1).

Tabelle 4.1.: Frame-Constraints bei Fehlern ohne Segmentbewegungen

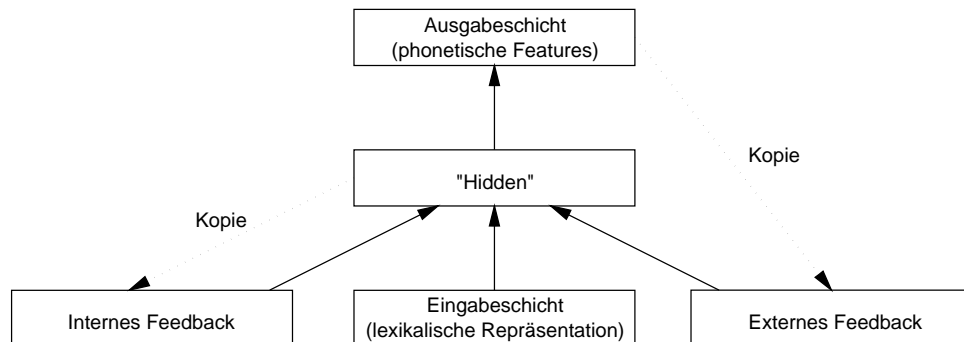


Abbildung 4.2.: Netzwerkarchitektur mit interner und externer Kontextschicht. Jedes Kästchen repräsentiert eine Menge von Knoten.

### 4.3. Das Modell

Grundlage des Modells bildet ein sogenanntes *einfaches rekurrentes Netzwerk* („SRN“) (Jordan 1986; Elman 1990). SRNs sind „*Error-Backpropagation*“-Netze (Rumelhart, Hinton & Williams 1984), die zur Behandlung sequentiell in der Zeit ablaufender Phänomene um eine Art „Zustandsgedächtnis“ erweitert sind, das zusammen mit der Eingabeschicht das Netz mit Informationen versorgt. Die Aufgabe des Netzes besteht darin, anhand einer Lexemrepräsentation eine entsprechende Sequenz von Phonemrepräsentationen zu generieren. Dazu wird zunächst an der Eingabe-Schicht (siehe Abb. 4.2) eine Repräsentation des zu produzierenden Wortes, kodiert als Aktivierungsmuster (0'en und 1'en), angelegt. Die Aktivierung wird dann durch gewichtete Leitungen zur „Hidden“-Schicht und schließlich zu den Knoten der Ausgabe-Schicht geleitet. Diese Knoten repräsentieren jeweils ein phonologisches Features und zusammen das Ausgabe-Phonem des Netzes. Um nun eine Sequenz von Ausgabe-Phonemen zu ermöglichen, ist es nötig, daß das Netz „weiß“, an welcher Stelle im Wort sich die Produktion gerade befindet. Dazu wird (entsprechend den Vorschlägen von Jordan (1986) bzw. Elman (1990)) nach jedem Zyklus — also nach der Produktion eines Phonems — das Aktivationsmuster der Ausgabe- und Hidden-Schicht in die entsprechenden *Feedback-Schichten* kopiert (siehe Abb. 4.2) eine Repräsentation des zu produzierenden Wortes, kodiert als Aktivierung). Im darauffolgenden Zyklus dient die in diesen Schichten kodierte Status-Information zusammen mit der Lexem-Repräsentation, die erhalten bleibt, als Input des Netzes.<sup>1</sup> Das Training des Netzwerks, also die Einstellung der Leitungsstärken, so daß aus der Lexem-Repräsentation die entsprechende Phonemsequenz erzeugt wird, erfolgt nach der

<sup>1</sup>Am Anfang der Produktion sind die Knoten der Feedback-Schichten mit neutralen Werten initialisiert.



Standard-Backpropagation-Methode.<sup>2</sup>

## 4.4. Simulationsstudien

### 4.4.1. Studie 1: Variation der Simulationsfaktoren

Studie 1 diente dazu, zunächst einen Einblick in den Zusammenhang zwischen verschiedenen Simulationsparametern und dem Verhalten des Modells zu gewinnen. Dazu wurden, wie in Tabelle 4.2 dargestellt, die Faktoren (A) Trainingsvokabular, (B) Art der Inputrepräsentation und (C) Art der Statusrepräsentation variiert. Jede der sich so ergebenden acht Bedingungen wurde mit drei Initialleistungsstärken getestet. Das Training erfolgte jeweils so lange, bis 90% der Segmente korrekt produziert wurden, wobei in jeder Trainingsepoche die Präsentationsreihenfolge der Items zufällig variiert wurde und ein Segment dann als korrekt gewertet wurde, wenn der euklidische Abstand vom Ausgabevektor zur Repräsentation des Ziel-Segments kleiner war als zu jedem anderen Segment.

Das durchschnittliche Fehlerprofil, erzeugt von den 24 trainierten Netzen, entsprach bezüglich den vier Frame-Constraints weitgehend den Standard-Versprecherdaten. Lediglich bei phonotaktischer Regularität gab es eine größere Abweichung von 92.6% in den Simulationen zu einem Standard von 99% in einer natürlichen Fehlersammlung (Stemberger 1983). Dell et al. führen den hohen Standard allerdings auf eine mögliche perzeptuelle Neigung der Fehler-Sammler in Richtung phonotaktisch korrekter Äußerungen zurück.

Der Vergleich unter den verschiedenen Simulationsbedingungen ergab, daß die Wahl des Trainingsvokabulars einen starken Effekt auf das Fehlerprofil der Simulation ausübt. Die Fehler der mit den häufigen Wörtern trainierten Netze entsprachen sowohl besser den Constraints der phonotaktischen Regularität, der CV-Kategorie-Erhaltung als auch dem Silbische-Konstituenten-Effekt. Die Variation der beiden anderen Faktoren resultierte im wesentlichen in einer unterschiedlichen Allgemein-Performanz.

Das Ergebnis der 1. Studie zeigt, daß das Modell relativ unabhängig von den variierten Faktoren ein adäquates Versprecherprofil produziert. Dell et al. führen zwei Hauptgründe für die gute Performanz an. Zum einen ist es sehr wahrscheinlich, daß wenn das Modell ein falsches Segment ausgibt, dieses dem Ziel-Segment phonetisch *ähnlich* ist, da ja als Ausgabe-Repräsentation phonetische Features gewählt wurden. Der euklidische Abstand zwischen einem prototypischen /b/ und einem prototypischen /p/ beträgt z.B. 1, während der Abstand von /b/ zu /d/  $\sqrt{3}$  und der Abstand von /b/ zu /a/  $\sqrt{6}$  beträgt. Der größere mittlere Abstand zwischen Vokalen und Konsonanten als zwischen Vokalen und Konsonanten untereinander erklärt auch den Kategorie-Erhaltungs-Effekt. Der zweite wichtige Faktor ist die Tendenz des Modells, „gewohnte“ Sequenzen zu produzieren. Wenn z.B. zuletzt ein Stop-Konsonant produziert wurde, befindet sich in der externen Feedback-Schicht dessen Kopie. In der Trainingsphase sollte das Netz „gelernt“ haben, daß auf einen Stop

---

<sup>2</sup>Es wurden *Bias*-Knoten verwandt; der *Momentum*-Faktor lag bei 0.5, die *Lernrate* zwischen 0.25 und 1.0.

4. Ein PDP-Ansatz zur Erklärung von „Frame-Constraints“

<b>Faktor / Bedingung</b>	<b>Motivation / Charakteristika</b>
A <i>Vokabular</i>	Feststellung des Zusammenhangs zwischen Aufbau des Trainingsvokabulars und erzeugtem Fehlerprofil unter der Hypothese, daß Frame-Constraints lediglich die Haupteigenschaften des gespeicherten Vokabulars reflektieren.
1. frequent	Die 50 häufigsten 3-buchstabigen, 3-segmentigen Wörter (Kuçera & Francis 1967). Diese machen insgesamt 88.9% der verwendeten Token dieser Klasse aus.
2. infrequent	Das Vokabular repräsentiert 8% der im Englischen verwendeten Token.
B <i>Eingabe-Repräsent.</i>	Wichtigkeit der Input-Repräsentation für Netzverhalten (Lachter & Bever 1988; Pinker & Prince 1988).
1. zufällig	Lexeme werden als zufällig ausgewähltes 30-stelliges 0-1-Muster repräsentiert. Aufgrund der fehlenden Korrelation zur Wortform kann diese Enkodierungsform als semantische Repräsentation und die Abbildung des Netzes als lexikalischer Zugriff interpretiert werden.
2. korreliert	Orthographische Enkodierung der Eingabe, wobei jeder Buchstabe durch 5 Bit dargestellt ist. Die Aufgabe des Netzes kann also als lautes Lesen verstanden werden.
C <i>Status-Repräsent.</i>	Feststellung des Einflusses externen Feedbacks.
1. intern	Es wird wie bei Elman (1990) nur internes Feedback verwendet (siehe Abb. 4.2).
2. intern-extern	Zusätzlich zum internen wird externes Feedback verwendet (siehe Jordan 1986). Externes Feedback allein reicht zur Lösung der Sequentialisierungsaufgabe nicht aus, da sich die Statusrepräsentation z.B. beim Wort <i>did</i> (/d/ /I/ /d/ null) zum Zeitpunkt des ersten /d/ nicht von der zum Zeitpunkt des zweiten /d/ unterscheidet. Externes Feedback sollte also lokale Zusammenhänge im Wort stärker gewichten.

Tabelle 4.2.: In Simulationsstudie 1 variierte Faktoren

nie ein weiterer Stop folgt. Der SRN-Mechanismus extrahiert gewissermaßen die stochastischen Haupteigenschaften des Trainingsmaterials. So ist die adäquatere Performanz in der Bedingung mit frequentem Vokabular auch darauf zurückzuführen, daß dieses die Lautstruktur der englischen Sprache besser reflektiert als das Trainingsset mit weniger häufigen Wörtern. Der deutlichere Silben-Konstituenten-Effekt in den Bedingungen mit frequenten Vokabular liegt z.B. darin begründet, daß im frequenten Vokabular, wie auch im Englischen im allgemeinen, VC-Sequenzen redundanter sind als CV-Sequenzen. Das heißt es gibt mehr Wörter (bzw. Token), die die gleiche VC-Sequenz beinhalten als Wörter, die die gleiche CV-Sequenz beinhalten. Aus diesem Grund ist die Wahrscheinlichkeit eines Fehlers beim Übergang von C auf V größer als beim Übergang von V auf C. Die Reproduktion des Initialeffekts kann damit begründet werden, daß zu Beginn der Produktion eines Wortes das Netzwerk weniger Informationen hat, da die Kontext-Schichten noch auf neutrale Werte initialisiert sind.

#### 4.4.2. Studie 2: Fehler-Generierung durch Rauschen

Das Problem in der Analogie zwischen den in Studie 1 erzeugten Fehlern und tatsächlichen Versprechern besteht darin, daß erstere auf einen Mangel an „Wissen“ zurückzuführen sind, während letztere eher in einer vorübergehenden Unfähigkeit, dies zu nutzen, begründet liegen. Das Problem wurde in Studie 2 gelöst, indem die Netze bis zu einer fehlerfreien Performanz trainiert wurden und dann in der Testphase Rauschen auf das Netz gegeben wurde, um so eine degradierte Performanz zu erreichen. Im einzelnen wurde normalverteiltes Rauschen zu den Hidden-Ausgabe-Verbindungen und — zum Vergleich — zu den Kontext-Hidden-Verbindungen addiert. Das Ergebnis war, daß bei gestörten Hidden-Output-Verbindungen das erzeugte Fehlerprofil wiederum sehr gut dem Standard entsprach. In der Bedingung mit gestörten Kontext-Hidden-Leitungen wurde dagegen das Standard-Fehlerprofil nicht reproduziert. Dies wurde als Bestätigung dafür gewertet, daß nicht jedes Modell, ab einem bestimmten Prozentsatz korrekter Produktionen ein adäquates Fehlerprofil erzeugt.

#### 4.4.3. Studie 3: Größeres Vokabular

Zuletzt wurde untersucht, wie sich das Modell bei einem größeren Vokabular (alle 308 3-Buchstaben, 3-Segment-Wörter aus dem Corpus von Kuçera und Francis (1967)) verhält. In dieser Bedingung konnten Netze nur nach Erhöhung der Anzahl der Hidden-Knoten auf 40 und nach Anwendung einiger Tricks (Entfernung einiger untypischer Wörter und Ausstattung der Kontexteinheiten mit einer Gedächtniskomponente (siehe Jordan 1986)) auf eine genügend gute Performanz (durchschn. 99.1%) trainiert werden. Die übrigen Fehler entsprachen weitgehend den Frame-Constraints. Lediglich der Silbische-Konstituenten-Effekt konnte nicht überprüft werden, da die übriggebliebenen Fehler jeweils nur ein Segment betrafen.

In einer letzten Simulation wurde das Trainingsset um 1- und 2-Segment-Wörter auf 408 Elemente erweitert. Die erzeugten Fehler waren nur zu 89.3% phonotaktisch regulär, das übrige Fehlerprofil jedoch adäquat.

## 4.5. Diskussion

Im großen und ganzen reproduzieren die Simulationen die vier Versprecher-Constraints, phonotaktische Regularität, Kategorieerhaltung, Silbenkonstituenz und Initialkonsonanten-Anfälligkeit sehr zufriedenstellend. Dieses Ergebnis bestätigt die Hypothese, daß diese Effekte nicht, wie bisher angenommen, die Existenz von Struktur-Regeln, bzw. Frames und separaten Repräsentationen von phonologischen Segmenten implizieren. Das Modell zeigt, daß die Effekte durch einen Mechanismus erzeugt werden können, der keine „a priori“-Annahmen über eine interne Repräsentation von Struktur und Inhalt enthält.

### 4.5.1. Linguistik vs. Psycholinguistik

Abgesehen von den zentralen Fragestellungen liefert das Modell, bzw. die Ergebnisse der Simulationsstudien, eine gute Illustration für die in Abschnitt 1.4 geschilderte Problematik der leichtfertigen Vermischung von Erklärungsebenen. Speziell in der Psycholinguistik liegt der Fehler nahe, von aus Sprachdaten extrahierten Generalisierungen — hier den Frame-Constraints — direkt auf einen analogen psychologischen Mechanismus — hier auf einen Frame-Mechanismus — zu schließen. Dies soll nicht bedeuten, daß solche Generalisierungen in psychologischen Theorien generell zu vermeiden sind, sondern, daß es notwendig ist, sich des Unterschieds zwischen einer genuin linguistischen Kompetenz-Theorie und einer genuin psychologischen Performanz-Theorie bewußt zu sein. Das Ergebnis der Simulationen suggeriert, daß Versprechereffekte (zumindest die Frame-Effekte) letztlich auf die Kovarianzstruktur des Vokabulars der zugrundeliegenden Sprache zurückgeführt werden können (siehe §4.4.1). Auf die Frage, *warum* das Vokabular diese Struktur hat und nicht eine andere, liefert das Modell keine Antwort. Dies zu beantworten, ist — so Dell et al. (1993, S. 180) — Aufgabe einer phonologischen Kompetenz-Theorie. Diese Sichtweise legt den Verdacht nahe, einen *circulus viciosus* folgender Art zu postulieren:

„Die Lautstruktur einer Sprache reflektiert die Struktur des Sprachproduktionssystems, wobei die Struktur des Sprachproduktionssystems die Lautstruktur der Sprache reflektiert.“

Klassische Theorien setzen vermutlich auch daher angeborene Regeln und Prädispositionen voraus. Hierzu besteht — jedenfalls aus diesem Grund — keine Notwendigkeit, da, wie auch Dell et al. (1993, S. 180) bemerken, die Lautstruktur einer Sprache nur auf einer breiten Grundlage psychologischer Erkenntnisse über Perzeption, Produktion, Kognition,

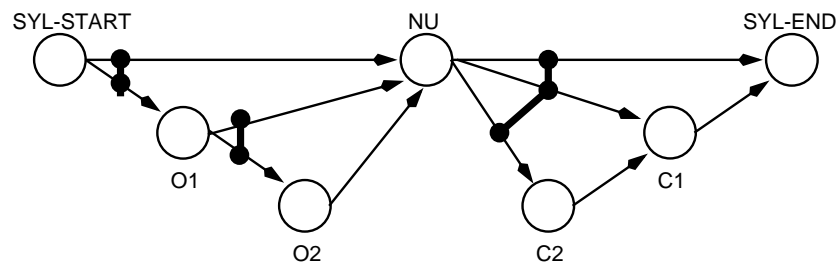


Abbildung 4.3.: Ein Kontrollknotennetzwerk für die Produktion von Silben (Eikmeyer et al. 1995, S. 22). Kreise repräsentieren Kontrollknoten, schwarze Punkte Ventilknoten, Pfeile exzitatorische Verbindungen und fette Linien gegenseitige Inhibition.

biomechanischer Erkenntnisse über Audition und Artikulation, sowie linguistischer Distributionsanalysen zu erklären ist.

Dell et al. sehen ihren Beitrag zur aktuellen Debatte in der Phonologie vor allem darin, daß sie demonstriert haben, daß eine Erklärung des Silben-Konstituenten-Effekts ohne die Annahme einer Onset/Rhyme-Trennung möglich ist. Außerdem suggerieren sie die Möglichkeit, aus dem Modell, bzw. aus der Struktur der vom Modell erzeugten Fehler, Vorhersagen darüber zu machen, welche zur Zeit nicht existierenden Phonemsequenzen in der Sprache zukünftig auftauchen können.

#### 4.5.2. Ein Vergleich mit alternativen Ansätzen

In erster Linie liefern Dell et al. (1993), wie gesagt, einen neuen Ansatz zur Erklärung von Frame-Constraints bei Versprechern. Der Vergleich mit anderen Sprachproduktionsmodellen sollte sich deshalb zunächst auf deren Erklärungsweise dieser Effekte richten. In Schades (1992) lokalem konnektionistischen Modell, z.B. wird die Sequentialisierung von Segmenten durch einen Kontrollknotenketten- bzw. durch einen *Kontrollknotennetz-Mechanismus* (Eikmeyer, Schade & Kupietz 1995; Eikmeyer et al. 1995) geleistet. Kontroll-Netze, wie in Abb. 4.3 dargestellt, bestehen aus einer Anzahl von *Kontrollknoten*, die im Falle des Silben-Netzes (derselbe Mechanismus wird auch für Syntax und Morphologie verwandt) die phonologischen Kategorien *Onset1*, *Onset2*, *Nucleus*, *Coda1* und *Coda2* repräsentieren. Jeder dieser Knoten ist mit gerichteten exzitatorischen Leitungen zu Phonemen seiner Kategorie verbunden. Während der Produktion einer Silbe ist immer genau einer der Kontrollknoten aktiv. Nach einer gewissen Zeit, wird das Phonem mit der höchsten Aktivierung zur Produktion ausgewählt. Dies ist wahrscheinlich das richtige Phonem, da nur dies sowohl von dem Kontrollknoten als auch vom Knoten der zu äußernden Silbe aktiviert wird. Nach der Selektion leitet der Kontrollknoten seine Aktiva-

tion an einen Nachfolger weiter. Welcher dies ist, hängt von den sogenannten „Gating“- oder *Ventilknoten* ab. Die Ventilknoten beziehen ihre Aktivierung aus der Phonemschicht, so wird z.B. das Ventil in Richtung *Onset2* von allen Phonemen aktiviert, die im Prinzip innerhalb eines Onsetclusters an zweiter Stelle stehen können ( $/l/O_2, /r/O_2$ )<sup>3</sup>. Außerdem sorgen unterschiedlich Ruheniveaus und eine starke laterale Inhibition zwischen den Ventilknoten dafür, daß in der Regel der richtige Ventilknoten maximal aktiviert ist und damit der richtige Kontrollknoten ausgewählt wird.

Die vier Frame-Constraints sind innerhalb des Modells von Schade (hypothetisch — besser wäre sicherlich eine Simulation) folgendermaßen zu erklären:

1. *Phonotaktische Regularität* ist zum Teil im Kontrollknoten-Mechanismus präkodiert. Silben ohne Nucleus sind aufgrund der Netzstruktur „a-priori“ ausgeschlossen; illegale Konsonanten-Kombinationen werden teilweise durch die getrennte Repräsentation von Konsonanten vermieden. Irreguläre Onset-Cluster, wie z.B.  $/dl/$ , sind im Prinzip möglich, aber unwahrscheinlich, da daß im Netzwerk vorhandene interlaterale Feedback sowohl ähnliche Fehler-Wörter/-Silben wie ähnliche Fehler-Phoneme begünstigt. Es ist also wahrscheinlicher, daß ein  $/kl/$ -Onset durch  $/gl/$  als durch  $/dl/$  ersetzt wird, da  $/dl/$  im Gegensatz zu  $/gl/$  („lexical Bias“) kein Feedback von der Lexem- bzw. Silben-Ebene bekommt und zweitens  $/g/$  wegen seiner phonetisch größeren Ähnlichkeit zu  $/k/$  mehr Feedback von der Ebene der phonetischen Features bekommt.
2. *Kategorie-Erhaltung* ist analog zur phonotaktischen Regularität zu erklären.
3. Der *Silbenkonstituenz-Effekt* ist ähnlich wie bei Dell et al. auf die stochastischen Eigenschaften der Struktur des Vokabulars zurückzuführen. In lokal konnektionistischen Modellen sind Wortfrequenzen meist durch Ruhewerte von Lemma-Knoten modelliert (vgl. Dell 1990, S. 331ff). Häufig zusammen auftretende Silbenkonstituenten geben einander proportional mehr Feedback. Daher ist der Übergang zwischen ihnen weniger fehleranfällig.
4. Der *Initialeffekt* ist ähnlich wie bei Dells Modell Ausdruck einer geringeren Präsenz von „Information“ im Netz. Der Mangel betrifft allerdings weniger Sequenzinformationen, sondern eine noch nicht völlig ausgebreitete und „gesetzte“ Aktivierung im Inhaltsnetz.

Zusammenfassend läßt sich feststellen, daß auch bei dem in Eikmeyer et al. (1995) beschriebenen Ansatz die Reproduktion von Frame-Constraints im wesentlichen nicht „a-priori“-Modellannahmen reflektiert, sondern auf die Sensitivität des Netzwerks für das zugrundeliegende Vokabular zurückzuführen ist. Der wesentliche qualitative Unterschiede zwischen den Ansätzen in bezug auf die Erklärung von Frame-Constraints besteht darin,

---

<sup>3</sup>Das Modell trennt zwischen Konsonanten an verschiedenen Silbenpositionen. Für das Phonem  $/t/$  gibt es z.B. im Netz die Knoten  $/t/O_1, /t/O_2, /t/C_1$  und  $/t/C_2$ .

daß erstens in dem lokalen Modell mehr zu motivierende Annahmen über Struktur und Repräsentation des Vokabulars gemacht werden und es zweitens dort keinen uniformen Mechanismus für Paradigmatik und Syntagmatik gibt. Allerdings läßt sich anhand dieser Unterschiede aufgrund der extremen Begrenztheit des erklärten Phänomens noch kein qualitativer Vergleich der Ansätze vornehmen, zumal Schades Modell eine wesentlich größere Datendomäne erklärt.

Wie bereits im ersten Teil dieser Arbeit mehrfach angedeutet (siehe z.B. §1.6.3), macht es wenig Sinn sogenannte „*Toy-Size*“-Modelle zu vergleichen, da sich die Größe der Menge gleichwertiger Modelle antiproportional zur Größe der Menge der erklärten Daten verhält. Deshalb ist zunächst zu überprüfen, ob z.B. die Uniformität des distribuierten Ansatzes nicht mit einer geringeren Allgemeingültigkeit erkauft wird. Im speziellen muß daher z.B. untersucht werden, ob mit diesem uniformen Mechanismus auch Versprecher mit Segmentbewegungen zu reproduzieren sind und ob der Mechanismus wirklich in der Lage ist, eine ausreichende Strukturierung des Vokabulars vorzunehmen. Kurz gesagt heißt das, daß überprüft werden muß, bis zu welcher Grenze sich der Ansatz erweitern läßt. Diese Untersuchung der Erweiterbarkeit des Ansatzes wird Gegenstand des nächsten Kapitels sein.

## 5. PDP–Modellierung der Sprachproduktion

### 5.1. Segmentbewegungen

Eine der größten Einschränkungen des Ansatzes von Dell et al. ist, daß nur solche Versprecher erklärt werden, die keine Segmentbewegungen beinhalten. Diese Einschränkung ist besonders kritisch, da diese nicht-kontextuellen, sogenannten „Non–Movement“-Fehler nur etwa 31% aller Fehler ausmachen (Garnham, Shillcock, Brown, Mill & Cutler 1981; Dell et al. 1993, S. 184). Dell et al. geben als möglichen Lösungsansatz eine Sichtweise an, in der kontextuelle Einflüsse ebenso wie andere fehlerverursachende Einflüsse, z.B. mangelnde Aufmerksamkeit, generell als Teil des Rauschens betrachtet werden, der sich auf das Netz auswirkt.

Um kontextuelle Einflüsse systematisch zu modellieren, wird vorgeschlagen, das Eingabe–Muster eines Zielwortes mit Repräsentationen zuvor produzierter und später zu produzierenden Wörtern zu kontaminieren (siehe Dell et al. 1993, S. 184). Im Falle einer Äußerung „York library“ entspricht z.B. jeder Eingabe–Knoten zuerst zu 80% seinem Soll–Wert für „York“ und zu 20% seinem Soll–Wert für „library“. Mit diesem Verfahren sollte es im Prinzip möglich sein, Antizipations- und Perseverationsfehler zu reproduzieren. Vertauschungsfehler sind mit diesem Mechanismus allerdings nicht erklärbar.

Eine solche Erweiterung des Modells um Kontaminationen der Input–Repräsentationen kann innerhalb einer Theorie der Sprachproduktion zum Beispiel als eine Störung des Lemma–Zugriffs–Prozesses oder eine Störung der, für die Sequentialisierung von Wörtern zuständigen, syntaktischen Prozesse aufgefaßt werden. Bevor theoretische Motivationen und Konsequenzen dieses Schrittes weiter diskutiert werden, soll allerdings durch Simulationen überprüft werden, wie sich das Fehler–Profil durch Mischung von Input–Repräsentationen tatsächlich ändert.



## 5.2. Simulationsstudien

### 5.2.1. Studie 1: Reproduktion

#### Methode

Um eine gesicherte Vergleichsbasis zu schaffen, habe ich zunächst eine Simulation des Modells implementiert, die der von Dell et al. in Simulationsstudie 2 verwendeten weitgehend entspricht. Die Netzwerk-Parameter unterschieden sich von denen des Originals allein darin, daß ich statt der korrelierten Eingabe-Repräsentation eine zufällige gewählt habe. Der Grund für diese Abweichung war, daß mir innerhalb eines allgemeinen Sprachproduktionsmodells eine auf Orthographie basierende Eingaberepräsentation ungerechtfertigt erscheint. Im Sinne eines Verzichts auf möglichst viele a-priori-Annahmen, war also die unkorrelierte Repräsentation vorzuziehen, zumal Dell et al.'s Studie 1 ergeben hatte, daß die Wahl der Repräsentation keinen Einfluß auf die Befolgung der Frame-Constraints ausübt.

Um zu überprüfen, ob meine Implementation die Effekte des Originals reproduziert, habe ich fünf von unterschiedlichen Initialleitungsstärken und unterschiedlichen Eingaberepräsentation<sup>1</sup> ausgehende Netze solange trainiert bis sie die 50 Wörter des frequenten Vokabulars fehlerfrei produzieren konnten. Die Netze brauchten alle weniger als 500 Trainings-Epochen.<sup>2</sup> Die fünf trainierten Netze wurden jeweils 10 mal mit jeweils 5 verschiedenen Rausch-Niveaus getestet. Rauschen wurde, wie bei Dell et al., durch Addition von normalverteilten Zufallswerten ( $\mu = 0$ ;  $\sigma \in \{0.2, 0.15, 0.1, 0.075, 0.05\}$ ) auf die Gewichte der Leitungen zwischen Hidden- und Output-Schicht simuliert, wobei die Gewichte der durch Rauschen gestörten Leitungen nach folgender Formel berechnet wurden:

$$w_{\sigma} = w + \text{Zufall}(\sigma) * |w| \quad (5.1)$$

Das Rauschen, bzw. die Leitungsstärken, wurden innerhalb der Produktion eines Wortes konstant gehalten. Insgesamt wurden so pro Rauschbedingung 10000 Segmente (einschließlich des Stop-Segments) produziert.

#### Ergebnisse

Vergleicht man das in Tabelle 5.1 dargestellte Fehlerprofil mit den Ergebnissen von Dell et al. bzw. mit den Standards aus der Versprecherliteratur, fällt auf, daß ein adäquater *Initialkonsonanten-Effekt* nicht reproduziert werden konnte (siehe Tabelle 5.2). Zwar ergab ein Mittelwertsvergleichstest (t-Test für gepaarte Variablen), daß Onset-Fehler signifikant häufiger sind als Konsonantenfehler an anderen Positionen ( $P < .01$ ) und daß es ein signifikantes Gefälle für Fehlerwahrscheinlichkeiten abhängig von der Segment-Position im Wort gibt ( $P < .05$ ), die durchschnittliche relative Häufigkeit von Onset-Fehlern

<sup>1</sup>Die Hamming-Distanz zwischen zwei Eingabe-Repräsentationen einer Simulation betrug mindestens 12.

<sup>2</sup>Wenn in Epoche 500 das Trainingsmaterial noch nicht beherrscht wurde, wurde das Training abgebrochen und mit neuen Initialleitungsstärken gestartet.

5. PDP-Modellierung der Sprachproduktion

Rauschen ( $\sigma$ )	Fehler	% phonot. regulär	% Kat.- erhaltend	% VC	% CV	Onset
.200	409 (4.09%)	95.26	98.29	2.44	0.49	39.75
.150	161 (1.61%)	97.39	98.76	2.48	0.00	36.67
.100	33 (0.33%)	96.88	100.00	0.00	0.00	39.02
.075	15 (0.15%)	100.00	100.00	0.00	0.00	33.33
.050	5 (0.05%)	100.00	100.00	0.00	0.00	40.00
insges.	623 (1.25%)	96.05	98.55	2.24	0.32	38.70

Tabelle 5.1.: Fehlerprofil bei verschiedenen Rausch-Niveaus

Fehler	Replikation	Dell et al.	Standards
<b>Anteil in %</b>	4.09	4.0	0.3
<b>% phonotakt. regulär</b>	95.26	96.1	99
<b>% kategorieerhaltend</b>	98.29	96.9	99.5
<b>Initialkonsonant</b>	39.75	77.6	62
<b>% VC</b>	2.44	5.4	6
<b>% CV</b>	0.49	3.9	2

Tabelle 5.2.: Ein Vergleich des in der Replikationsstudie (Studie 1) erzeugten Fehlerprofils mit dem aus Studie 2 von Dell et. al und den Standards aus der Versprecherliteratur.

( $\bar{e}_1 = .387$ ) lag jedoch nur wenig über der für eine zufällige Verteilung erwarteten ( $\mu_0 = 1/3$ ) und weit unter dem Standard von 62%.

Ein möglicher Grund für das Ausbleiben eines stärkeren Effektes ist der im Vergleich zur Original-Simulation möglicherweise größere Mindest-Abstand zwischen den Eingabe-Repräsentationen, der dazu führen konnte, daß sich das Netz stärker auf die Input-Schicht als auf die Information aus den Kontext-Schichten „verläßt“. Um diese Hypothese zu testen, wurde eine weitere Simulation mit 15 Input-Knoten (vorher 30) und einem Mindestabstand von 1 zwischen den Vektoren 25 mal getestet. Die Konsonantenfehler-Verteilung unterschied sich in dieser Simulation nicht signifikant von der Simulation mit 30 Eingabe-Knoten; die relative Häufigkeit eines Onset-Konsonanten-Fehlers betrug  $0.3975$  ( $\sigma = .0244$ ). Der Abstand der Eingabevektoren hat also keinen Effekt auf Onsetfehler-Häufigkeiten. Ein adäquater Initialkonsonanten-Effekt konnte somit nicht reproduziert werden.

### 5.2.2. Studie 2: Bewegungsfehler durch Kontamination der Input-Schicht

Das Ziel von Studie 2 war zu überprüfen, ob sich der Erklärungsbereich des Modells von Dell et al. mittels des oben beschriebenen Ansatzes der Kontamination von Eingabe-Mustern um Bewegungsfehler erweitern läßt. Es wurden dazu analog zu Studie 1 zunächst 15 Netze mit unterschiedlichen Anfangsgewichten so lange trainiert, bis sie das Trainingsmaterial fehlerfrei reproduzierten. In der Testphase wurde jedem Netz jeweils jedes der 50 trainierten Wörter zu 15% kontaminiert mit wiederum jedem der 50 trainierten Wörter als Eingabe präsentiert. Es ergaben sich also insgesamt  $15 * 50 * 50 = 37500$  zu produzierende Wörter. Die Kontamination einer Wortrepräsentation  $R$  durch die Repräsentation eines Stör-Wortes  $D$  im Eingabeknoten  $I_i$  erfolgte dabei nach der Formel:

$$I_i = 0.85 * R_i + 0.15 * D_i \quad (5.2)$$

Als Bewegungsfehler wurde jedes produzierte fehlerhafte Segment gewertet, das auch im Störwort vorkam. Um die Ergebnisse interpretieren zu können, wurden Kontrollsimulationen mit denselben Netzen gemacht. In den Kontrollsimulationen wurden die Eingabe-Repräsentation zu 0% kontaminiert, dafür wurde — wie in Studie 1 — zu den Gewichten der Leitungen zwischen Hidden- und Output-Schicht ein normalverteiltes Rauschen addiert ( $\sigma = .2$ ).

Die Kontamination der Eingabe-Schicht erzeugte 5.21% fehlerhafte Segmente von denen 21.76% als Bewegungsfehler klassifiziert wurden (siehe Tab. 5.3). Ein Mittelwertsvergleich (t-Test für unabhängige Stichproben) ergab keine signifikant gesteigerte relative Häufigkeit von Bewegungsfehlern gegenüber den Kontrollsimulationen. Verstöße gegen den Kategorie-Erhaltungs-Constraint waren signifikant häufiger ( $P < .01$ ) als in der Kontrollsimulation. Die geringe Anzahl der Simulationen ließ keinen gesicherten Vergleich zwischen den Proportionen phonotaktisch regulärer Versprecher zu.

Durch die Kontamination von Input-Vektoren mit einer Kontextwort-Repräsentation konnte also kein starker Effekt in Richtung der adäquaten Bewegungsfehler-Proportion von ca. 69% erreicht werden. Eine Ursache für das Ausbleiben eines stärkeren Effektes war vermutlich die hinsichtlich der phonologischen Gestalt der zu produzierenden Wörter arbiträre Eingabe-Repräsentation. Bei einer korrelierten Eingaberepräsentation, wären vermutlich mehr Bewegungsfehler entstanden. Wie bereits weiter oben dargelegt, halte ich aber eine korrelierte Repräsentation für nicht zu rechtfertigen. Ausgehend von einer solchen Repräsentation wäre die Aufgabe des Netzes auf eine Transformation „räumlich“ geordneter Information in eine zeitliche Ordnung beschränkt. Ein durch Mischen von Eingaben erzielter Effekt wäre trivial.

Fehler	Kontamination der Input-Schicht (15%)	Rauschen auf $H \rightarrow O$ ( $\sigma = .2$ )
Anzahl	7820	5566
Anteil in %	5.21	3.71
% phonotakt. regulär	90.48	94.59
% kategorieverhaltend	93.29	98.62
Initialkonsonant	34.73	34.53
% VC	5.93	2.80
% CV	3.29	2.17
% Bewegungsfehler	21.76	13.67

Tabelle 5.3.: Fehlerprofile bei einer zu 15% kontaminierten Input-Schicht und in der Kontrollbedingungen mit normalverteiltem Rauschen ( $\sigma = .2$ ) auf den Leitungen von der Hidden- zur Output-Schicht.

### 5.2.3. Studie 3: Bewegungsfehler durch Kontamination der Hidden-Schicht

Eine bessere erscheinende Methode ist, nicht Inputs zu mischen, sondern die durch das Netz selbst strukturierten Hidden-Repräsentationen. Dies wurde in einer weiteren Simulationstudie folgendermaßen realisiert: Es wurden dieselben 15 Netze wie in Studie 2 verwendet. In der Testphase wurde bei jedem zu produzierenden Segment jeweils die Repräsentation des Zielworts und die Repräsentation eines Störwortes als Eingabe verwendet und die beiden resultierenden Hidden-Vektoren im Verhältnis 80:20 gemischt.

Zwar ergaben die Simulationen eine signifikant gesteigerte Rate von Bewegungsfehlern ( $P < .01$ ), der erzielte Wert von 22.01% lag jedoch wiederum weit unterhalb des Standards von 69% aus der Versprecherliteratur. Außerdem wurde eine signifikant größere Proportion phonotaktisch nicht regulärer Wörter produziert als in der Kontrollsimulation ( $P < .05$ ; siehe Tab. 5.4).

### 5.2.4. Studie 4: Größeres Vokabular, unterschiedliche Segmentanzahl

In einer zusätzlichen Simulationsstudie sollte überprüft werden, wie sich das Netz beim Training mit einem größeren Vokabular mit Wörtern unterschiedlicher Segmentanzahl verhält. Als Trainingsmaterial dienten 197 häufige deutsche Wörter.<sup>3</sup> Ursprünglich sollte das Netz bis zu einer vollständig korrekten Performanz trainiert werden. Da diese aller-

<sup>3</sup>Es wurden die 200 häufigsten Wortformen nach (Kaeding 1898) ausgewählt. Drei Formen mußten gestrichen werden, da sie aussprachegleich mit anderen Formen waren. Es handelt sich zwar um einen sehr alten Schriftsprachekorpus, die ausgewählten Wörter sind aber auch in der Häufigkeitsliste umgangssprachlicher Wörter von H.-H. Waengler (1963) zu finden.

Fehler	Kontamination der Hidden-Schicht (20%)	Rauschen auf H → O ( $\sigma = .2$ )
<b>Anzahl</b>	8942	5566
<b>Anteil in %</b>	5.96	3.71
<b>% phonotakt. regulär</b>	86.44	94.59
<b>% kategorieerhaltent</b>	95.97	98.62
<b>Initialkonsonant</b>	36.91	34.53
<b>% VC</b>	4.66	2.80
<b>% CV</b>	1.41	2.17
<b>% Bewegungsfehler</b>	22.01	13.67

Tabelle 5.4.: Fehlerprofile bei einer zu 20% kontaminierten Hidden-Schicht und in der Kontrollbedingungen mit normalverteiltem Rauschen ( $\sigma = .2$ ) auf den Leitungen von der Hidden- zur Output-Schicht.

dings bei einer Konfiguration von 40 Eingabe- und 60 Hidden-Knoten bei mehreren Tests mit unterschiedlichen Initialleitungsstärken nicht erreicht werden konnte, wurde das Training bei einer weiteren Simulation nach Erreichen von 96% korrekt produzierten Segmenten (im Zyklus 700) abgebrochen. Die in diesem Stadium noch fehlerhaft reproduzierten Wörter sind in Tabelle 5.5 dargestellt.

Die Ergebnisse müssen mit Vorsicht beurteilt werden, da erstens nur eine Simulation mit insgesamt 35 Fehlern untersucht wurde und, da es zweitens unklar ist, wie die Performanz eines unvollständig trainierten Netzes zu interpretieren ist. Es fällt jedoch auf, daß der Anteil phonotaktisch regulärer Fehler wesentlich geringer ausfällt als in den vorausgegangenen Simulationen. Die Ursache für das Ausbleiben des „Phonotactic-Regularity“-Effekt ist, daß es in der Menge der deutschen Trainingswörter im Gegensatz zu der der englischen einige ähnliche Phoneme betreffende Subregularitäten gibt, die der Backpropagation-Algorithmus offensichtlich nicht extrahiert hat. Bezüglich der Kategorie der Vokale wird die Regel, daß ein Wort — ausgenommen sind Funktionswörter — eine Silbe enthalten muß, die einen Akzent tragen kann, 8 mal durch eine Vokal-Ersetzung durch /ə/ verletzt. Dies ist sicherlich noch durch die unzureichende Modellierung von Akzent zu erklären. Die zusätzliche Nichtbeachtung der unterschiedlichen Distribution der ähnlich kodierten Phoneme /ç/ und /x/ deutet jedoch darauf hin, daß das Netz sich zu sehr auf die Ähnlichkeit von Ausgaberepräsentationen stützt und dagegen kontextuelle Abhängigkeiten zu wenig berücksichtigt.

Möglicherweise sind diese ungewünschten Effekte durch eine adäquatere Enkodierung der Ausgabe und durch eine adäquatere Präsentation des Trainingsmaterials zu vermeiden. Z.B. könnten zuerst einfachere Wörter gelernt werden („*phased training*“, siehe Elman 1991; Clark 1993, Kap. 7). Zumindest wird durch diese Studie jedoch die Frage aufgeworfen, inwieweit phontaktische Regularität und Kategorie-Erhaltung von einander unabhängig sind. Zwar mag ein starker Zusammenhang der beiden Effekte psychologisch

plausibel klingen, die durch das Modell erklärte Datendomäne reduzierte sich allerdings stark.

## 5.3. Diskussion

### 5.3.1. Modellbewertung

Die Simulationsstudien haben gezeigt, daß sich der Erklärungsbereich des Modells von Dell et al. — zumindest auf die vorgeschlagene Art — nicht in ausreichendem Umfang um Bewegungsfehler erweitern läßt. Der Anteil der erzeugten „kontext“-bedingten „Versprecher“ lag jeweils weit unterhalb des bei realen Sprechern beobachteten Anteils. Darüber hinaus konnte der Initialkonsonanten-Effekt — aus unbekanntem Gründen — in keiner der Simulationen repliziert werden. Die adäquat beschriebene Datendomäne reduziert sich also auf die Effekte phonotaktische Regularität, Kategorieerhaltung und die größere Fehlerwahrscheinlichkeit bei VC-Clustern. Eine Perspektive für Erweiterungen ist nicht offensichtlich.

Abgesehen von den empirischen Befunden sprechen auch theoretische Überlegungen gegen die Erweiterbarkeit des Ansatzes zu einem allgemeinen Sprachproduktionsmodell. Die von Dell et al. vorgeschlagene Kontamination von Inputs ist z.B. schwer zu motivieren, da Antizipations-Fehler auf einen Mechanismus zurückzuführen sein sollten, der flüssiges Sprechen unter normalen Umständen erleichtert (eine ausführliche Diskussion findet sich bei Dell et al. 1995). Im Sprachproduktionsmodell von Schade (1992) sind z.B. das Vorwärmen von zu produzierenden Einheiten und Antizipationsfehler auf einen uniformen Mechanismus zurückzuführen. Innerhalb des vorgestellten Ansatzes gibt es keinen Mechanismus, der unter normalen Umständen von der Kontamination der Eingabe-Schicht durch zukünftig zu produzierende Wörter profitieren könnte. Der Erweiterungsvorschlag ist also gewissermaßen „ad hoc“. Darüber hinaus könnte ein hypothetisch auf diesem basierend „funktionierendes“ Modell zwar Antizipations- und Perseverationsfehler, aber keine Vertauschungsfehler erklären, da innerhalb des rekurrenten Netzes eine Segment-Antizipation die Wahrscheinlichkeit für eine nachfolgende Produktion des Zielsegmentes und damit die Wahrscheinlichkeit für eine Vertauschung nicht beeinflußt. Das dem Modell zugrundeliegende Sequentialisierungsverfahren ist inhärent für Zukunft und Vergangenheit „blind“. Der Versuch diese Einschränkung durch Eingabe-Überblendungen abzuschwächen, bedeutet die Einführung eines zweiten, unabhängigen Mechanismus und damit den Verzicht auf die zuvor positiv bewertete Uniformität des Ansatzes (siehe §4.5.2). Die damit verbundene Komplexitätssteigerung des Modells ist erstens theoretisch schwer zu motivieren und steht zweitens in keinem Verhältnis zu der begrenzten prinzipiell erreichbaren Allgemeinheit.

Ein weiterer Nachteil des Sequentialisierungsmechanismus ist, daß er nicht zuläßt, den zeitlichen Ablauf von Prozessen während, vor und nach der Produktion von Phonemen zu modellieren. Die minimale zeitliche Granularität, also ein Simulationszyklus, ist

5. PDP-Modellierung der Sprachproduktion

Eingabe	Ziel	Ausgabe	Reg.	Kat.
mit	mit #	mə*t#	-	+
menschen	mɛnʃən#	mə*ns*ən#	-	+
über	y:bə#	u:*be#	+	+
macht	maxt #	maç*t#	-	+
sie	zi:#	zi:z*#	-	
einmal	ainma:l#	ain#*o:*l#	-	
dich	dɪç#	də*ç#	-	+
du	du:#	b*u:#	+	+
und	ʊnt#	ʊne*#	+	-
ihres	i:rəs #	ə*rəs#	-	+
zwar	tsva:#	tsvə*#	-	+
zum	tsʊm#	tsʊn*#	+	+
großen	gro:sən#	j*re:*sən#	-	+
gemacht	gəmaxt#	gən*aç*t#	+	+
gemacht	gəmaxt#	gən*aç*t#	-	+
im	ɪm#	ɪn*#	+	+
mir	mi:e#	me:*e#	+	+
hier	hi:e#	he:*e#	+	+
kann	kan#	ə*an#	-	-
leben	le:bən#	lə*bən#	-	+
immer	ɪme#	ɪn*e#	+	+
als	als#	alsə*#	+	
welches	vɛlçəs#	və*lçəs#	-	+
zu	tsu:#	tsʊ*#	-	+
gewesen	gəve:zən#	gəvə*zən#	-	+
werde	vɛvɛdə#	və*ɛt*ə#	-	+
werde	vɛvɛdə#	və*ɛt*ə#	+	+
andere	andərə#	ant*ə#*	+	+
recht	rɛçt#	rɛçtz*#	-	
also	alzɔ:#	alzə*#	+	+
viel	fi:l#	fe:*l#	+	+
			45%	97%

Tabelle 5.5.: Die in Simulationsstudie 4 nach 700 Zyklen Training erzeugten Fehler. Doppelkreuze repräsentieren das Stop-Segment, also ein Wort-Ende. Fehler-Segmente sind mit einem Asterisk (\*) markiert. Ein Plus in der Spalte „Reg.“ bedeutet, daß das erzeugte Fehlerwort syntaktisch regulär ist. Kategorieerhaltende Segmentersetzungen sind in der Spalte „Kat.“ mit einem Plus markiert.

eben genau die Produktion eines Phonems. Dies schränkt die Erklärungsfähigkeit aller auf diesem Mechanismus basierenden Modelle erheblich ein. Außerdem ist dies auch ein Nachteil gegenüber lokal konnektionistischen Modellen, in denen in der Regel eine beliebig feine zeitliche Auflösung möglich ist. In Schades Modell z.B. breitet sich Aktivierung über die Zeit aus während laterale Inhibition für die Deaktivierung irrelevanter Einheiten sorgt (vgl. Dell 1986; Berg & Schade 1992; Schade & Berg 1992). „Zwingt“ man ein solches Modell zu einer schnellen Abfolge von Selektionen bzw. Produktionen entstehen spezifische Vorhersagen über das Fehlerprofil unter Zeitdruck gesprochener Äußerungen. Außerdem liefern solche Modelle Erklärungshypothesen über den *Entstehungsprozeß* von Verprechern. Demzufolge muß konstatiert werden, daß auch bezüglich der Erklärung von Dynamik zur Zeit lokal konnektionistische Modelle überlegen sind.

Es stellt sich also die Frage, welche Funktion innerhalb der psycholinguistischen Forschung Dell et al.'s Ansatz haben kann. Angesichts der Simulationsergebnisse scheint es zweifelhaft, ob die geringe Größe der erklärten Daten-Domäne (Frame-Constraints) lediglich auf die geringe Elaboriertheit des Ansatzes zurückzuführen ist. Anders formuliert, liegt der Eindruck nahe, daß das Modell, angesichts der wenigen nachgebildeten Effekte, trivial ist. In diesem Zusammenhang müssen die bei Dell et al. dargestellten und aus dem Modell (bzw. den Simulationen) gewonnenen Erkenntnisse neu diskutiert werden.

Das Hauptanliegen der Arbeit von Dell et al. war zu zeigen, daß bei der Sprachproduktion eine Struktur-Inhalt-Trennung auf der phonologischen Ebene nicht notwendigerweise psychologisch real ist. Dieses Ziel wird nicht auf zufriedenstellende Weise erreicht, da der Verzicht auf diese Trennung möglicherweise die Nicht-Erklärbarkeit von in diesem Zusammenhang äußerst relevanten Bewegungsfehlern impliziert. Die, im Sinne eines Verzichts auf Strukturrepräsentationen, relative Einfachheit der Erklärung von Versprechern ohne Segmentbewegungen wird durch eine sehr viel größere Komplexität bei der Erklärung von Segmentbewegungen, bzw. deren Nicht-Erklärbarkeit, erkaufte. Der Grund dafür ist, daß die Ausgabe eines einfach rekurrenten Netzwerks auf einer innerhalb der Trainingsphase bestimmten Projektion des Eingabe-/Zustandsraums auf den Ausgaberaum basiert. Es wird sozusagen das Element ausgegeben, das ausgehend von Eingabe- und Zustandsrepräsentation in bezug auf das Trainingsmaterial am wahrscheinlichsten ist. Eine „korrekte“ Sequentialisierung kann also durch Training erzielt werden. Sehr spezifische Ausgabefehler, also solche, die sich dem Betrachter als Sequentialisierungsfehler, bzw. als Antizipations-, Perseverations- und Bewegungsfehler darstellen, können — in weit überzufälliger Proportion — nur dann entstehen, wenn das Netz darauf trainiert ist, eben diese „Fehler“ zu machen. Dies scheint allerdings genauso wenig sinnvoll, wie ein symbolisches Sprachproduktionsmodell um „Fehlerregeln“ zu erweitern. Was ist also mit Dell et al.'s Ansatz gewonnen?



### 5.3.2. Möglichkeiten des psychologischen Erkenntnisgewinns

Eine interessante Eigenschaft eines PDP-Ansatzes zur Modellierung der Sprachproduktion besteht darin, daß er die Möglichkeit bietet — in Richtung eines in Kapitel 2 skizzierten Idealbilds — weitgehend auf „a-priori“-Annahmen über Repräsentationen zu verzichten. Dell et al. setzen dieses Idealbild zumindest partiell in die Realität um. Sie verwenden eine arbiträre Kodierung lexikalischer Einheiten als Eingabe, eine phonologisch motivierte Kodierung als Ausgabe und überlassen die interne Repräsentation in der Hidden-Schicht der Selbstorganisation durch den Trainingsalgorithmus. Während die Wahl der Eingaberepräsentation — in Anbetracht dessen, daß nur ein Teilphänomen der Sprachproduktion erklärt werden soll — nicht besser zu leisten ist, ist die Abstraktion von motorischem Verhalten in der Ausgabe möglicherweise problematisch. Ein Verzicht auf die Modellierung, bzw. die Simulation des Artikulationsapparates impliziert z.B., daß durch diesen wirksame Restriktionen physischer Art, in den Bereich interner Repräsentationen verschoben werden. Eine Diskussion weiterer Implikationen und die Darstellung einer alternativen, artikulatorisch phonologischen Herangehensweise mittels dynamischer Systeme findet sich bei Browman und Goldstein (1995) (siehe auch §5.3.4). Was die Wahl von Ein- und Ausgaberepräsentationen betrifft, unterscheidet sich das PDP-Modell nicht qualitativ von den meisten lokal konnektionistischen Modellen. Die als interne Repräsentation interpretierbare Hidden-Schicht dagegen hebt sich von denen lokaler Netzwerke ab.

#### 5.3.2.1. Interne Repräsentationen bei lokalen und distribuierten Netzen

Qua Definitionem ist das wesentliche Unterscheidungsmerkmal zwischen den beiden Klassen von Netzwerkmodellen, daß im lokalen Netzwerk jede modellierte Einheit in genau einem Knoten *lokalisierbar* ist. Im Gegensatz dazu sind in distribuierten Netzwerken zu modellierende Einheiten durch Aktivationsverteilungen über eine Menge von Knoten dargestellt („[...] items can be represented [...] by a pattern of activity of large sets of units with each unit encoding a microfeature of the item“, Rumelhart & McClelland 1986, S. 108; siehe aber vor allem van Gelder 1991). Verteilte Repräsentationen haben vor allem den Vorteil, daß sie in Analogie zum Gehirn, relativ unempfindlich gegen den Verlust eines Teils ihrer Knoten oder Verbindungen sind. Die Performanz eines Netzes verschlechtert sich allmählich mit zunehmendem Schaden („*graceful degradation*“). Wesentlich für die Konnektionismus-Symbolismus-Debatte war allerdings, daß sich aus distribuierten Repräsentationen keine *Symbole* extrahieren lassen. So ist in einem PDP-Netz keine Einheit lokalisierbar, die permanent als für eine bestimmte andere Entität stehend interpretiert werden kann. Innerhalb der theoretischen Diskussion fungierte dies als Argument gegen den Vorwurf, daß Konnektionismus gegenüber Symbolismus keinen wirklich neuen Ansatz darstellte (siehe Fodor & Pylyshyn 1988; Smolensky 1988). Die Tatsache, daß sich lokal konnektionistische Ansätze in dieser Hinsicht nicht in dem

Maße von traditionellen unterscheiden, da in ihnen symbolartige Einheiten lokalisierbar sind, ist hauptverantwortlich für ihre geringe Präsenz in der theoretischen Konnektionismus-Literatur. Der Unterschied zwischen den internen Repräsentationen lokaler und distribuerter Netze bedarf allerdings einer genaueren Betrachtung. In lokalen konnektionistischen Netzwerken sind zwar die Knoten mit Labels versehen, jedoch spielen diese Labels für die *Funktion* des Netzes keine Rolle. Sie erleichtern lediglich die Erstellung und die *Interpretation* eines Modells. Weiterhin ist auch eine traditionelle symbolische Interpretation des Zustands eines lokalen Netzes nur *approximativ* möglich, da jeweils mehrere Knoten einer Ebene, bzw. einer Kategorie gleichzeitig aktiv sein können. Eine getreue Interpretation der Phonemschicht eines Netzes zu einem Zeitpunkt *t* muß also auf der Aktivationsverteilung über alle (aktiven) Knoten dieser Schicht beruhen. In dieser Hinsicht ist auch die Information in einem lokalen Netz *distribuiert*.

### 5.3.2.2. Explanatorische Inversion

Der Unterschied zu Backpropagation-PDP-Netzen besteht also hauptsächlich darin, daß die Netzwerkarchitektur, das heißt die Verbindungen interner Knoten untereinander und zu den Ein- und Ausgabeknoten, von Hand festgelegt werden. Der so erstellte Repräsentationsrahmen ist ein Teil der durch das Modell aufgestellten Hypothesen. Die durch das Modell gelieferten Erklärungen beruhen dementsprechend auf den postulierten Knoten (bzw. deren Labels) und den postulierten Verbindungen zwischen diesen. Im Falle von auf Backpropagation basierenden PDP-Modellen scheint der Gewinn und die Formulierung psychologischer Erkenntnisse über die Struktur interner Repräsentationen und die Dynamik der ablaufenden Prozesse zunächst schwieriger. Aufgrund der nicht festgelegten Bedeutung einzelner interner Knoten und der stark verteilten Information über Knoten und Leitungen waren Erkenntnisse, darüber hinaus, daß gewisse psychologische Phänomene sich mit diesem radikal anderem Ansatz auch reproduzieren lassen, nicht offensichtlich. Demzufolge waren Vorwürfe wie:

„Ein konnektionistisches Modell ist *mutatis mutandis* genau so praktisch wie ein Stadtplan im Maßstab 1:1.“ (Levelt 1991, S. 68)

zunächst nicht leicht zu entkräften. Um psychologische Erkenntnisse zu gewinnen, war es nötig, den explanatorischen Prozeß umzukehren („*explanatory inversion*“, Clark 1993, S. 49): Ausgehend von einem „funktionierenden“ Programm — also quasi einer Level-3 Theorie — mußte untersucht werden, welche Eigenschaften des Programms für seine Funktion hauptsächlich verantwortlich sind, um so ein eine Erklärung auf einer höheren Ebene zu ermöglichen.

### 5.3.2.3. Untersuchung statischer Strukturen

Ende der achtziger Jahre wurden dann vermehrt statistische Verfahren zur *post-hoc-Analyse* eingesetzt. Um einen Einblick in die Repräsentationsstruktur in der Hidden-Schicht zu

gewinnen, wurden z.B. durch sogenannte *hierarchische Cluster-Analysen* die aus den verschiedenen Eingaben erzeugten Hidden-Vektoren ihrem euklidischen Abstand nach paarweise kombiniert (siehe z.B. Rosenberg & Sejnowski 1987). Die so gewonnene Baumstruktur ermöglicht zu interpretieren, in welche Klassen das Netz seine Eingaben unterteilt, und damit, Hypothesen über die Struktur interner Repräsentationen abzuleiten.

Das Dendrogramm in Abbildung 5.1 zeigt das Ergebnis einer Cluster-Analyse über die gemittelten Hidden-Repräsentationen aller Phoneme in einem Netzwerk aus der Replikationsstudie. Je weiter links zwei Phoneme miteinander verbunden sind, desto ähnlicher sind ihre Repräsentationen. Faßt man den Repräsentationsraum des Phoneminventars des Netzes als Vektorraum auf, stellt das Diagramm die Nähe-Relation zwischen den Phonem-Prototypen dar. Es fällt auf, daß die Phoneme im wesentlichen ihrer Ähnlichkeit nach geclustert sind. Daß diese Clusterung allerdings nicht nur nach der Ausgaberepräsentation — die ja auf phonologischen Features beruht — erfolgt ist, beweist die mit „Seltene Phoneme“ bezeichnete Klasse. Jedes der 4 Phoneme in diesem Cluster kommt jeweils nur einmal im Trainingmaterial vor. Offenbar hat sich das Netz einen eigenen Subraum für weniger wichtige Elemente geschaffen. Daß dies allerdings nicht notwendig zur Lösung der Aufgabe ist, zeigt der große Abstand zwischen den Repräsentationen von /j/ und /u/ in Abbildung 5.2. Mittels einer Hauptkomponentenanalyse wurde dort der Phonemraum eines anderen Netzes aus der Replikationsstudie auf zwei Dimensionen reduziert. In diesem Diagramm fällt außerdem auf, daß /E/, /i/ und /I/ sehr weit von den restlichen Vokalen entfernt liegen.

Anhand von Abbildung 5.2 läßt sich auch erklären, warum eine Kontamination der Eingabeschicht nicht im gewünschten Maße zu Bewegungsfehlern führt. Eine leichte Verschiebung des Eingabevektors in Richtung eines Kontext-Wortes sollte eine entsprechende Verschiebung der resultierenden Hiddenvektoren im Zustandsraum des Netzwerks bewirken. So sollten auftauchende Fehler-Segmente immer in enger Nachbarschaft zum Ziel-Segment liegen. Nur wenn im Kontaminationswort auftauchende Segmente in *direkter* Nachbarschaft von Segmenten des Ziel-Worts liegen, sollten Bewegungsfehler beobachtbar sein. Falls Phonemrepräsentationen zwischen Ziel- und Kontaminationssegmenten liegen, sollte es wahrscheinlicher sein, daß diese geäußert werden.

Das Cluster der „seltene Phoneme“ macht wiederum deutlich, wie entscheidend die Auswahl des Trainingsmaterials für das Verhalten von Backpropagation-Netzwerken ist. Eine große Rolle für die Strukturierung des Repräsentationsraums eines Netzes spielt auch die Reihenfolge in der die Trainingsbeispiele präsentiert werden (siehe Pinker & Prince 1988; Marcus et al. 1990; Plunkett & Marchman 1991b, 1991a; Stoet 1993).

### 5.3.2.4. Dynamik

Mit dem Aufkommen rekurrenter Netze wurde es nötig, die dynamischen Aspekte der Modelle über die Zeit zu analysieren. Als besonders fruchtbar erwiesen sich dazu die Methoden und das Beschreibungs-Vokabular aus der Theorie dynamischer Systeme (siehe §2.3; Port & van Gelder 1995). Die Prozeßdynamik eines Netzes wird dabei als Trajekto-

5. PDP-Modellierung der Sprachproduktion

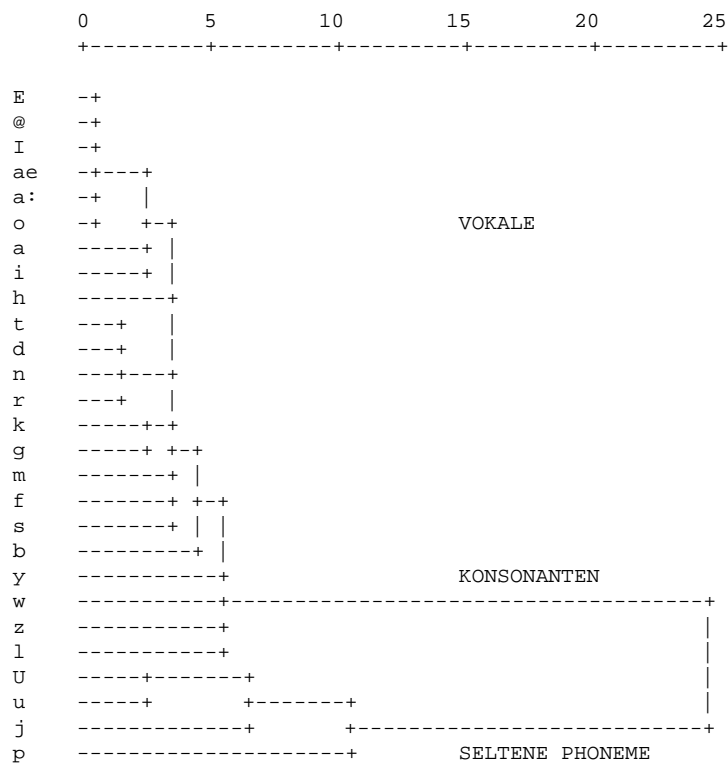


Abbildung 5.1.: Hierarchisches Cluster-Diagramm der Hidden-Knoten-Aktivierungen

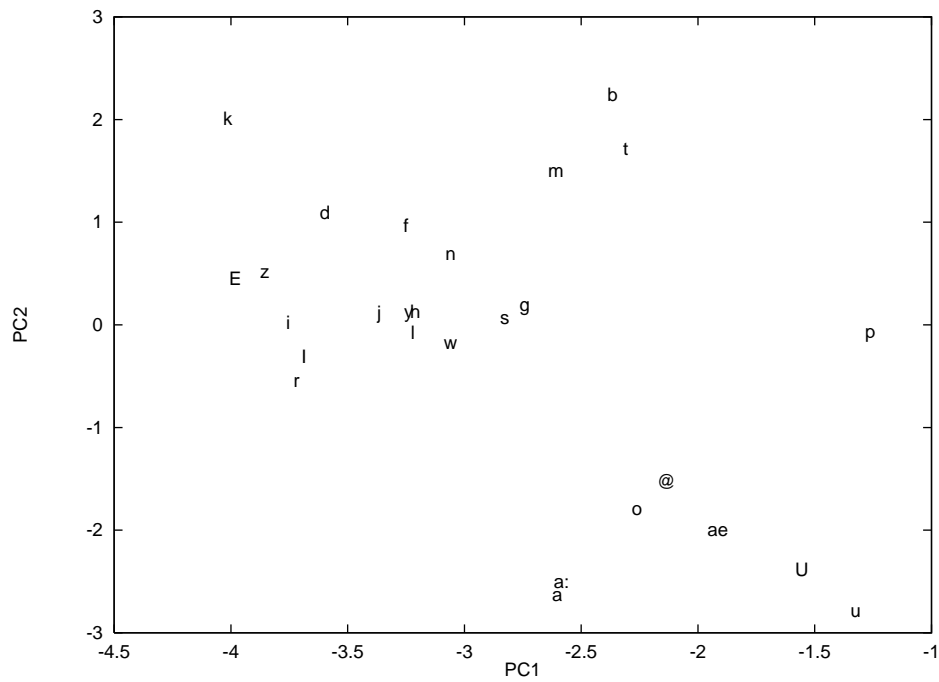


Abbildung 5.2.: Durch Hauptkomponentenanalyse auf 2 Dimensionen reduzierter Hidden-Raum eines Netzwerks aus der Replikationsstudie (PC1 erklärt 62% der Varianz, PC2 13.8%). Dargestellt sind die Projektionen der gemittelten Hidden-Repräsentationen aller Phoneme aus dem Trainingsmaterial. Die Phoneme sind in SAMPA-Notation dargestellt (ESPRIT project 2589 (SAM) 1992).

rie durch seinen Zustandsraum beschrieben. Die Regeln, nach denen Prozesse im Netz ablaufen, werden anstatt als symbolische Spezifikationen als Dynamik des Systems konzeptualisiert, die den Zustand des Systems in bestimmte Richtungen wahrscheinlicher lenkt als in andere (Elman 1995, S. 195). Zur Veranschaulichung des entsprechend der Anzahl der Hidden-Knoten hochdimensionalen Zustandsraum des Netzwerks wird in der Regel eine Hauptkomponentenanalyse durchgeführt. Abbildung 5.3 zeigt die durch die Produktion einiger Wörter erzeugten Trajektorien eines Netzes aus der Replikationsstudie projiziert auf die durch die beiden ersten Hauptkomponenten der Hidden-Repräsentationen aufgespannte Fläche. Die erste Hauptkomponente kann interpretiert werden, als repräsentiere sie — für Vokale und Konsonanten getrennt — die Position des Produktionssystems im Wort. Darüber hinaus ist das Diagramm allerdings wenig aufschlußreich, zum einen wegen des geringen Anteils aufgeklärter Varianz (34.52%) und zum anderen wegen der durch den SRN-Mechanismus bedingten groben zeitlichen Auflösung.

### 5.3.3. Zusammenfassung

Auch die durch statistische post-hoc Analysen der Netze gewonnenen Daten liefern wenig brauchbare psychologische Erkenntnisse. Dies ist allerdings nicht verwunderlich. Wie bereits vorher dargelegt, war die Ausgaberepräsentation des Netzes *a priori* auf phonologische Features festgelegt. Darüber hinaus schien das Sequentialisierungsverfahren nicht adäquat zu sein, da es keine Kontext-Effekte bei Versprechern erklären kann und keinen Aufschluß über den zeitlichen Ablauf dynamischer Prozesse geben kann.

Die beiden wesentlichen Faktoren für die Organisation der internen Repräsentationen waren also bereits vorgegeben bzw. inadäquat. Die Domäne eines möglichen Erkenntnisgewinns war damit von vornherein auf das Trainingsmaterial eingeschränkt. Angesichts der geringen psychologischen Beschreibungsadäquatheit und der geringen Allgemeinheit des Modells wäre es jedoch unzulässig gewesen, den durch den Faktor Trainingsmaterial erzielten Effekten, z.B. der Strukturierung des Hidden-Raums, direkt psychologische Realität beizumessen. Abgesehen davon war das englische Trainingsmaterial auch ökologisch wenig valide (nur 3-Segmentwörter, Unterrepräsentation von mehrsilbigen Wörtern), während die Simulation mit dem repräsentativeren deutschen Vokabular sehr inadäquate Versprecher erzeugte. Im Falle des englischen Vokabulars sind deshalb Rückschlüsse von der Strukturierung des Hidden-Raums auf die englische Sprache im allgemeinen nur begrenzt zulässig. Im Falle der Simulationsstudie mit dem deutschen Trainingsmaterial ist es offensichtlich, daß das Netz relevante phonologische Regularitäten der deutschen Sprache nicht extrahiert hat. Selbst wenn man also hypothetisch eine starke Abhängigkeit zwischen den stochastischen Eigenschaften einer Sprache und internen Repräsentationen annimmt, sind die durch die Simulationen gewonnenen Daten wenig weiterführend, da sie nicht die stochastischen Haupt-Eigenschaften der Struktur der Wörter der beiden Sprachen erfassen.

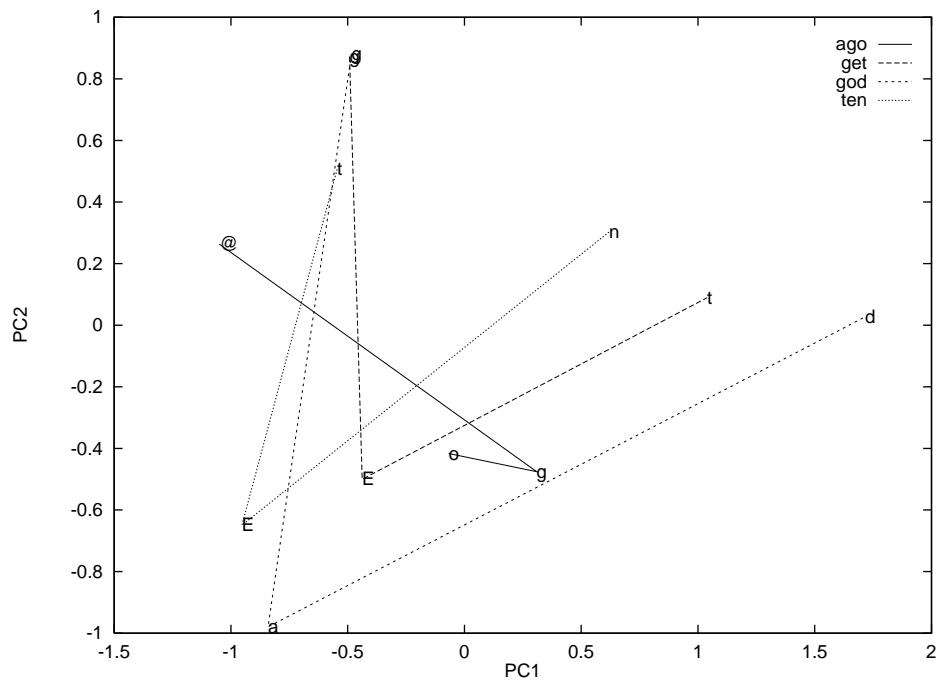


Abbildung 5.3.: Trajektorien durch den Zustandsraum (Hidden-Raum) eines Netzes aus der Replikationsstudie bei der Produktion der Wörter „ago“, „get“, „god“ und „ten“. Der 20-dimensionale Zustandsraum ist auf die durch die beiden ersten Hauptkomponenten aufgespannte Fläche projiziert. Die Hauptkomponenten wurden anhand aller Hidden-Repräsentationen errechnet, die das Netzwerk bei der Produktion der Wörter des Trainingsmaterial erzeugt. Hauptkomponente 1 (PC1) deckt 18.73% der Varianz ab, Hauptkomponente 2 (PC2) 15.79%.

#### 5.3.4. Schlußfolgerungen und Perspektiven

Die oben dargestellten Ergebnisse zeigen, daß es zur Zeit unklar ist, wie mit einem PDP-Ansatz Sprachproduktion auf der phonologischen Ebene modelliert werden kann. Im speziellen hat sich erwiesen, daß ein SRN („Simple Recurrent Network“) Mechanismus zur Sequentialisierung keine Alternative zu den klassischen, Struktur und Inhalt trennenden Slot-Filler-Mechanismen darstellt. Eine mögliche Schlußfolgerung ist, daß die Forschung auf dem Gebiet der Sprachproduktion weiterhin mit Hilfe von lokal konnektionistischen Modellen betrieben werden sollte. Die Perspektiven innerhalb dieses Paradigmas sind z.B. die Verbesserung der Modellierung von Syntax (siehe z.B. Eikmeyer et al. 1995) und eine Verbesserung der Modellierung psychologischen Entsprechungen von Semantik und vor allem Pragmatik und deren Anbindung an phonologische Prozesse (siehe z.B. Eikmeyer et al. 1995). Sicherlich besteht innerhalb dieses Paradigmas auch die Perspektive einer schrittweisen Ersetzung lokaler Repräsentationen durch distribuierte, z.B. auf der Konzeptebene (siehe Kessler 1995).

Ein ganz anderer Ansatzpunkt, der sich als kognitive, artikulatorische Phonologie klassifizieren läßt, besteht darin, den Sprachproduktionsprozeß ausgehend von einem mechanischen Artikulationsmodell zu untersuchen (Saltzman 1986; Browman & Goldstein 1990, 1995). Browman und Goldstein (1995) stellen ein System vor, in dem Äußerungen als Gruppe oder *Konstellationen* von Einheiten artikulatorischer Aktionen, sogenannter *Gesten*, organisiert sind. Jede Geste ist dabei als dynamisches System beschrieben, das die Formation und Auflösung von lokalen Konstriktionen im Vokaltrakt charakterisiert. Die Aufgabe der ersten Geste in der Produktion eines /b/ wäre also, die Lippen zu schließen. Ausgehend von diesem Ansatz wäre es interessant zu untersuchen, inwieweit phonologische Kategorien eine kausale Rolle im Produktions- bzw. Artikulationsprozeß spielen und ob anatomische Restriktionen bestimmte interne Repräsentationen überflüssig machen. Der Ausgangspunkt entspricht durch seinen Verzicht auf „a-priori“-Annahmen bezüglich Repräsentationen und Prozeßarchitektur und durch seine Einbeziehung von Motorik sicherlich am ehesten dem in Kapitel 2 skizzierten Idealbild. Allerdings ist noch unklar, wie eine solche Modellierung vorwiegend mechanischer Prozesse um die Erklärung psychologischer Prozesse erweitert werden kann, ohne lediglich die Grenze zwischen Geist und physischer Welt weiter nach innen zu schieben.



## 6. Zusammenfassung und Schlußfolgerungen

Im ersten Kapitel dieser Arbeit wurde gezeigt, daß der traditionelle symbolische Ansatz psychologischer Modell- und Theorie-Bildung wesentliche Unzulänglichkeiten aufweist. Hauptsächlich wurde kritisiert, daß von sehr wahrscheinlich falschen Prämissen, nämlich *optimalem Design*, *starker Modularität*, *starkem Nativismus*, „*Implementationsunabhängigkeit*“ und einer *Language-Of-Thought* ausgegangen wird. Weiterhin wurde demonstriert, daß die *Voraussetzung* dieser Prämissen — unabhängig von ihrer psychologischen Realität — nur sehr oberflächliche Erklärungen zuläßt, die Kognitionswissenschaft von relevanten Erkenntnissen aus benachbarten Wissenschaftsgebieten wie der Biologie isoliert und die Domäne von Vergleichsdaten ungerechtfertigt auf Kompetenzphänomene einschränkt. Aus diesen Gründen scheint ein Fortschritt für das Verständnis des menschlichen Geistes innerhalb des symbolischen Paradigmas sehr unwahrscheinlich. Im zweiten Kapitel wurde dann skizziert, wie eine Einbeziehung von Wissen aus der Neurobiologie, den Evolutionswissenschaften und der Verhaltensforschung zusammen mit einer Abkehr von einer rein rationalistischen und linguistisch geprägten Konzeption von Geist der Kognitionswissenschaft neue Perspektiven eröffnet.

Im zweiten Teil dieser Arbeit wurde schließlich untersucht, inwieweit die im ersten Teil skizzierten Prinzipien und Methoden einer *guten* Kognitionswissenschaft auf die aktuelle Forschung auf dem Gebiet der Sprachproduktion applizierbar sind. Dazu wurde zunächst die Entwicklungsgeschichte der Modelle und Theorien in diesem Forschungsgebiet mit der Entwicklung von dem in Kapitel 1 beschriebenen Prototypen eines klassischen, „schlechten“ Modells hin zu dem Prototypen eines in Kapitel 2 skizzierten guten, aber „utopischen“ Modells in Beziehung gesetzt. Dazu wurden die verschiedenen Modellierungsansätze anhand einiger der im 1. Teil dieser Arbeit thematisierten Prinzipien, wie z.B. „repräsentationelle Vorurteilsfreiheit“ und „Modellrestriktion durch Vergleichsdaten“ verglichen.

An einem in Hinblick auf die skizzierte Entwicklung „modernen“ PDP-Ansatz zur Modellierung der Sprachproduktion wurde zuletzt demonstriert, wie einige der dargestellten Methoden und Prinzipien sich praktisch auf die Konstruktion, die Analyse, die Interpretation und die Bewertung eines konkreten psychologischen Modells anwenden lassen. Dabei zeigten sich vor allem die Probleme, die sich durch die Wahl eines paral-

lel-distribuiert-konnektionistischen Ansatzes ergeben.

Im einzelnen erwiesen Simulationsstudien und eine theoretische Analyse, daß ein SRN kein psychologisch adäquater Mechanismus zur Linearisierung von Phonemen ist. Da dieser aber zur Zeit das einzige netz-inhärente Verfahren zur Sequentialisierung innerhalb des PDP-Ansatzes darstellt, liegt damit die Schlußfolgerung nahe, daß eine wesentliche Voraussetzung für die Produktion von Sprache innerhalb eines PDP-Ansatzes zur Zeit nicht zu modellieren ist. Eine nachfolgende *post-hoc* Analyse der Netze aus den Simulationsstudien warf weitere teilweise verallgemeinerbare Probleme einer PDP-Modellierung auf.

Die Quelle des Erkenntnisgewinns im PDP-Paradigma ist die Strukturierung der Information in der Hidden-Schicht, die vom Netz selbst erzeugt wird. Die Parameter, die diese Strukturierung im wesentlichen bestimmen, sind Ein- und Ausgaberepräsentation und Trainingsmaterial. Es zeigte sich, daß bei der Auswahl dieser Parameter genauso unfundierte Annahmen gemacht werden müssen, wie bei der Wahl interner Repräsentationen in lokal konnektionistischen Modellen. Die durch *post-hoc* Analysen von Hidden-Repräsentationen gewonnen Resultate können jedenfalls nicht unabhängig von diesem Annahmen interpretiert werden. Im Hinblick auf „repräsentationelle Vorurteile“ unterscheidet sich also ein PDP-Modell zunächst nur insofern von einem lokal konnektionistischen Modell, daß die „Vorurteile“ in der Regel nicht als Modellhypothesen explizit gemacht werden. Der PDP-Ansatz ist in dieser Hinsicht erst dann im Vorteil, wenn wenigstens die Ein- oder die Ausgabeschicht nicht lediglich der *Repräsentation* dienen, sondern — wie in Kapitel 2 beschrieben — respektive als Puffer für sensorischen Input oder motorischen Output. Für die Modellierung der Sprachproduktion heißt das, daß Ausgabeknoten nicht zur Repräsentation phonetischer Features dienen sollten, sondern zur Steuerung des Artikulationsapparats.

Die Anbindung von Ausgabeschichten an die physische Welt ist allerdings nur ein kleiner Schritt auf dem Weg zu einer Kognitionswissenschaft, die Kognition nicht lediglich als rein mentales Phänomen betrachtet. Kognition ist, was auch immer man darunter verstehen mag, in jeder Beziehung, durch die Anforderungen der Interaktion mit der Umwelt determiniert. Das größte Erkenntnispotential sehe ich deshalb in der Untersuchung bzw. Berücksichtigungen dieser Interaktionen. Eine mögliche Methode dies zu tun ist, z.B. auf Basis gesicherter biologischer Erkenntnisse zunächst *Animats* zu konstruieren. Eine ebenso legitime Herangehensweise ist allerdings, auf Basis von biologischem Wissen und „fundierter Hypothesen“ „humanoide“ situierte Agenten zu konstruieren und zu untersuchen, inwieweit die verbale und non-verbale Interaktion mit einer möglicherweise vereinfachten Umwelt ihre kognitive Architektur restringiert.

# Literaturverzeichnis

- Abraham, R. H. & Shaw, C. D. (1992). *Dynamics - The Geometry of Behavior* (2. Aufl.). Redwood City, CA: Addison-Wesley.
- Arbib, M. A. (1987). Levels of modelling of mechanisms of visually guided behavior. *Behavioral and Brain Sciences*, 10, 407–465.
- Bechtel, W. & Abrahamsen, A. (1991). *Connectionism and the Mind*. Blackwell.
- Beer, R. (1992). *A dynamical systems perspective on autonomous agents* (Tech. Rep. 92-11). Cleveland, OH: Case Western Reserve University.
- Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell*. Tübingen: Niemeyer.
- Berg, T. & Schade, U. (1992). The role of inhibition in a spreading–activation model of language production. Part I. The psycholinguistic perspective. *Journal of Psycholinguistic Research*, 21(6), 405–434.
- Boden, M. (1991). Horses of a different color? In W. Ramsey, S. P. Stich & D. E. Rumelhart (Hrsg.), *Philosophy and Connectionist Theory* (S. 3–19). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2, 14–23.
- Brooks, R. A. (1989). A robot that walks: Emergent behaviors from a carefully evolved network. *Neural Computation*, 1, 253–262.
- Brooks, R. A. (1990). Elephants don't play chess. In P. Maes (Hrsg.), *Designing Autonomous Agents*. Cambridge, MA: MIT Press / A Bradford Book.
- Brooks, R. A. (1991a). Intelligence without reason. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (S. 569–595). San Mateo, CA: Morgan Kaufman.
- Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence*, 47, 139–160.

---

LITERATURVERZEICHNIS

---

- Brooks, R. A. (1992). Artificial life and real robots. In F. J. Varela & P. Bourguine (Hrsg.), *Toward a Practice of Autonomous Agents: Proceedings of the first European Conference on Artificial Life* (S. 3–10). Cambridge, MA: MIT Press / A Bradford Book.
- Brooks, R. A. (1994). Coherent behavior from many adaptive processes. In D. Cliff, P. Husbands, J. A. Meyer & S. Wilson (Hrsg.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press / A Bradford Book.
- Brooks, R. A. & Stein, A. (1993). *Building brains for bodies* (Memo 1439). MIT AI Laboratory.
- Browman, C. P. & Goldstein, L. (1990). Representation and reality: physical systems and phonological structure. *Journal of Phonetics*, 18, 411-424.
- Browman, C. P. & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port & T. van Gelder (Hrsg.), *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.
- Bruce, D. (1994). Lashley and the problem of serial order. *American Psychologist*, 49, 95–103.
- Carey, S. (1990). Cognitive development. In D. Osherson & E. Smith (Hrsg.), *Thinking: An Invitation to Cognitive Science*. Cambridge, MA: MIT Press.
- Cheney, D. L. & Seyfarth, R. M. (1990). *How Monkeys See the World*. Chicago, IL: The University of Chicago Press.
- Chomsky, N. (1957). *Syntactic Structures*. Den Haag: Mouton.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3, 1–61.
- Chomsky, N. (1986). *Knowledge of Language: It's Nature, Origin and Use*. New York: Praeger.
- Church, A. (1956). *Introduction to Mathematical Logic*. Princeton: Princeton University Press.
- Churchland, P. (1981). Eliminative materialism and propositional attitudes. *Journal of Philosophy*, 78(2), 67–90.
- Churchland, P. M. (1984). *Matter and Consciousness*. Cambridge, MA: MIT Press.
- Clark, A. (1987). From folk-psychology to naive psychology. *Cognitive Science*, 11(2), 139–154.

---

LITERATURVERZEICHNIS

---

- Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press.
- Clark, A. (1993). *Associative Engines: Connectionism, Concepts, and Representational Change*. Cambridge, MA: MIT Press.
- Cleermans, A., Servan-Schreiber, D. & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation*, 1, 372–381.
- Cliff, D. (1990). Computational neuroethology: A provisional manifesto. In J. A. Meyer & S. Wilson (Hrsg.), *From Animals to Animats 1: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (S. 29–39). Cambridge, MA: MIT Press / A Bradford Book.
- Cliff, D. (1992). Neural networks for visual tracking in an artificial fly. In F. J. Varela & P. Bourguine (Hrsg.), *Toward a Practice of Autonomous Agents: Proceedings of the first European Conference on Artificial Life* (S. 78–87). Cambridge, MA: MIT Press / A Bradford Book.
- Cliff, D. (1995). Neuroethology, computational. In M. A. Arbib (Hrsg.), *Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press.
- Cliff, D., Harvey, I. & Husbands, P. (1993). Explorations in evolutionary robotics. *Adaptive Behavior*, 2, 73–110.
- Cliff, D., Husbands, P. & Harvey, I. (1993). Analysis of evolved sensory motor controllers. In *Proceedings of the Second European Conference on Artificial Life* (S. 192–204).
- Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M. I. Posner & O. S. M. Marin (Hrsg.), *Attention and Performance* (Vol. XI, S. 3–37). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.
- Dawkins, R. (1986). *The Blind Watchmaker*. UK: Longman.
- De Las Heras-Mathews, V. (1995). *Is the Cause auf Psychology Advanced by Dynamical Systems Theory?* Unpublished master's thesis, COGS, University of Sussex, UK.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of prediction from a connectionist model. *Journal of Memory and Language*, 27, 124–142.

---

LITERATURVERZEICHNIS

---

- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5(4), 313–349.
- Dell, G. S., Burger, L. K. & Svec, W. R. (1995). *Language production and serial order: A functional analysis and a model* (Tech. Rep. UIUC-BI-CS-95-04). Urbana, IL: The Beckman Institute, University of Illinois.
- Dell, G. S., Juliano, C. & Govindjee, A. (1993). Structure and content in language production. *Cognitive Science*, 17, 149–195.
- Dell, G. S. & Reich, P. A. (1980). Toward a unified model of slips of the tongue. In V. A. Fromkin (Hrsg.), *Errors in Linguistic Performance* (S. 273–286). New York: Academic Press.
- Dennett, D. C. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.
- Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Hrsg.), *Minds, Machines and Evolution: Philosophical Studies* (S. 129–151). Cambridge, MA: Cambridge University Press.
- Dennett, D. C. (1986). The logical geography of computational approaches: A view from the East Pole. In M. Harnish & M. Brand (Hrsg.), *Problems in the Representation of Knowledge* (S. 59–79). Tucson, AZ: University of Arizona Press.
- Dennett, D. C. (1988). Review of J. Fodor, *Psychosemantics*. *Journal of Philosophy*, 85(7), 384–389.
- Dennett, D. C. (1991a). *Consciousness Explained*. Little, Brown & Co.
- Dennett, D. C. (1991b). Mother nature versus the walking encyclopedia: A western drama. In W. Ramsey, S. P. Stich & D. E. Rumelhart (Hrsg.), *Philosophy and Connectionist Theory* (S. 21–30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Devaney, R. L. (1986). *An Introduction to chaotic dynamical systems*. Menlo Park, CA: Benjamin/Cummings.
- Eikmeyer, H.-J. & Schade, U. (1991). Sequentialization in connectionist language-production models. *Cognitive Systems*, 3(2), 128–138.
- Eikmeyer, H.-J. & Schade, U. (1993). The role of computersimulation in neurolinguistics. *Nordic Journal of Linguistics*, 16, 153–169.
- Eikmeyer, H.-J., Schade, U. & Kupietz, M. (1995). Ein konnektionistisches Modell für die Produktion von Objektbenennungen. *Kognitionswissenschaft*, 5, 108–117.

---

LITERATURVERZEICHNIS

---

- Eikmeyer, H.-J., Schade, U., Kupietz, M. & Laubenstein, U. (1995). Connectionist syntax and planning in the production of object specifications. In R. Meyer-Klabunde & C. Von Sutterheim (Hrsg.), *Conceptual and Semantic Knowledge in Language Production. Proceedings of a Workshop of the Special Collaborative Research Programm 245 "Language and Situation"* (Vol. 92, S. 18–30). Heidelberg/Mannheim.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1991). *Incremental Learning, or The Importance of Starting Small* (Tech. Rep. 9101). Center for Research in Language, University of California, San Diego.
- Elman, J. L. (1995). Language as a dynamical system. In R. Port & T. van Gelder (Hrsg.), *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.
- Ernst, P. (1994). *Meaning and Mind in Monkeys*. Unpublished master's thesis, Department of Philosophy, Kings College, University of London.
- ESPRIT project 2589 (SAM). (1992). *Multi-lingual speech input/output assessment, methodology and standardization* (Final Report 1 III 91 - 28 II 1992). SAM-UCL-G004.
- Feldman, J. A. (1981). A connectionist model of visual memory. In G. E. Hinton & J. A. Anderson (Hrsg.), *Parallel Models of Associative Memory* (S. 49–81). Hillsdale, NJ: Lawrence Erlbaum.
- Feldman, J. A. & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science* 6, 205–254.
- Fodor, J. A. (1968a). *Psychological Explanation*. Random House.
- Fodor, J. A. (1968b). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy*, 65, 627–640.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1985). Précis of the modularity of mind. *Behavioral and Brain Sciences*, 8, 1–42.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1991). Replies. In B. Loewer & G. Rey (Hrsg.), *Meaning in Mind: Jerry Fodor and His Critics*. Blackwell.

---

LITERATURVERZEICHNIS

---

- Fodor, J. A. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Franceschini, N., Pichon, J.-M. & Blanes, C. (1992). From insect vision to robot vision. *Philosophical Transactions of the Royal Society of London*, 337, 283–294.
- Galton, A. (1990). The church-turing thesis: Its nature and status. *AISB Quarterly*, 74, 9–19.
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D. & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, 19, 805–817.
- Garrett, M. F. (1976). Syntactic processes in sentence production. In R. Wales & E. Walker (Hrsg.), *New Approaches to Language Mechanisms*. Amsterdam: North-Holland.
- Garrett, M. F. (1980). Levels of processing in speech production. In B. Butterworth (Hrsg.), *Language Production*, 1 (S. 177–220). London: Academic Press.
- Garrett, M. F. (1988). Processes in language production. In F. Newmeyer (Hrsg.), *Linguistics: The Cambridge Survey* (Vol. III, S. 69–96). Cambridge: Cambridge University Press.
- Graubard, S. (1988). *The Artificial Intelligence Debate: False Starts, Real Foundations*. Cambridge, MA: MIT Press / Bradford Books.
- Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8, 191–219.
- Harley, T. A. (1990). Paragrammatisms: Syntactic disturbance or failure of control? *Cognition*, 34, 191–219.
- Harley, T. A. (1993). Connectionist approaches to language disorders. *Aphasiology*, 7, 221–249.
- Harnad, S. (1989). Minds, machines and Searle. *Journal of Theoretical and Experimental Artificial Intelligence*, 1, 5–25.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335–346.
- Harnad, S. (1992). Connecting object to symbol in modeling cognition. In A. Clark & R. Lutz (Hrsg.), *Connectionism in Context*. Springer Verlag.
- Harnad, S. (1994). Computation is just interpretable symbol manipulation; cognition isn't. *Minds and Machines*, 4, 379–390. (Special Issue on “What Is Computation?”)



---

LITERATURVERZEICHNIS

---

- Harnad, S. (1995). Grounding symbolic capacity in robotic capacity. In L. Steels & R. Brooks (Hrsg.), *The "Artificial Life" Route to "Artificial Intelligence."* *Building Situated Agents* (S. 276–286). Hillsdale, NJ: Lawrence Erlbaum.
- Hart, J., Berndt, R. S. & Caramazza, A. (1985). Category-specific naming deficit following cerebral infarction. *Nature*, 316, 439-440.
- Haugeland, J. (1985). *Artificial Intelligence: The very idea*. Cambridge, MA: MIT Press / Bradford Books.
- Husbands, P., Harvey, I. & Cliff, D. (1995). Circle in the round: State space attractors for evolved sighted robots. In *Proceedings of the N.A.T.O. Advanced Study Institute Workshop on the Biology and Technology of Intelligent Autonomous Agents*.
- Jacobs, A. M. & Grainger, G. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1311–1334.
- Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach* (Institute for Cognitive Science Report 8604). San Diego, CA: University of California.
- Kaeding, F. W. (1898). *Häufigkeitswörterbuch der deutschen Sprache*. Stieglitz b. Berlin: Selbstverl.
- Katz, J. (1964). Mentalism in linguistics. *Language*, 40, 124–137.
- Kessler, K. (1995). *(Konnektionistische) Simulation als wissenschaftliche Methode am Beispiel der Rezeption von „vor“ und „hinter“*. Unveröffentlichte Diplomarbeit, Lehrstuhl für Psychologie III, Universität Mannheim.
- Kučera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Lachter, J. & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning — A constructive critique of some connectionist learning models. *Cognition*, 28, 195–247.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Hrsg.), *Cerebral Mechanisms in Behavior* (S. 112–136). New York: Wiley.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1991). Die konnektionistische Mode. *Sprache und Kognition*, 10, 61–72.

---

LITERATURVERZEICHNIS

---

- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–142.
- MacKay, D. G. (1987). *The Organization of Perception and Action*. New York: Springer.
- Marcus, G., Ullman, M., Pinker, S., Hollander, M., Rosen, T. & Xu, F. (1990). *Overregularization* (Occasional Paper 41). Center for Cognitive Science, MIT.
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.
- Maturana, H. R. & Varela, F. J. (1988). *The Tree of Knowledge*. Boston: Shambala.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception: Part I. An account of the basic findings. *Psychological Review*, 88, 375–407.
- McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Hrsg.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1, S. 3–44). Cambridge, MA: MIT Press.
- Meringer, R. (1908). *Aus dem Leben der Sprache*. Berlin: Behr's Verlag.
- Meringer, R. & Mayer, K. (1895). *Versprechen und Verlesen: Eine psychologische Studie*. Stuttgart: Göschen'sche Verlagshandlung.
- Meyer, A. S. (1990). The phonological encoding of successive syllables. *Journal of Memory and Language*, 29, 524–545.
- Meyering, T. C. (1994). Fodor's modularity: A new name for an old dilemma. *Philosophical Psychology*, 7(1), 39–62.
- Miller, G. F. (1993). *Evolution of the human brain through runaway sexual selection: The mind as a protean courtship device*. Unpublizierte Promotionschrift, Stanford University, Psychology Department.
- Miller, G. F. (1996). A review of sexual selection and human evolution: How mate choice shaped human nature. In C. Crawford & D. Krebs (Hrsg.), *Evolution and Human Behavior: Ideas, Issues, and Applications*. Hillsdale, NJ: Lawrence Erlbaum.
- Minsky, M. (1985). *The Society of Mind*. New York: Simon & Schuster.
- Morton, J. (1981). The status of information processing models of language. In *Proceedings of the Royal Society of London* (S. 387–396).
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135–183.

---

LITERATURVERZEICHNIS

---

- Newell, A. & Simon, H. (1976). Computer science as empirical inquiry. In J. Haugeland (Hrsg.), *Mind Design*. Cambridge, MA: MIT Press.
- Noble, J. & Cliff, D. (1996). *On Simulating the Evolution of Communication* (Cognitive Science Research Paper 420). University of Sussex, Brighton, UK: School for Cognitive and Computing Sciences.
- Nolfi, S. & Parisi, D. (1991). *Auto-teaching: networks that develop their own teaching input* (Tech. Rep. PCIA-91-03). 15. Viale Marx, Rome: Department of Cognitive Processes and Artificial Intelligence.
- Nooteboom, S. G. (1969). The tongue slips into patterns. In A. J. Sciarone, A. J. Van Es-sen & A. A. Van Raad (Hrsg.), *Leyden Studies in Linguistics and Phonetics*. Den Haag: Mouton.
- Norman, D. (1986). Reflections on cognition and parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Hrsg.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, S. 110–146). Cambridge, MA: MIT Press.
- Norten, A. (1995). Dynamics: An introduction. In R. Port & T. van Gelder (Hrsg.), *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.
- Patterson, K. E. & Marcel, A. J. (1977). Aphasia, dyslexia and the phonological encoding of written words. *Quarterly Journal of Experimental Psychology*, 29, 307–318.
- Peacocke, C. (1986). Explanation in computational psychology: Language perception and level 1.5. *Mind and Language*, 1(2), 101–123.
- Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural nets. In *Proceedings of the International Joint Conference on Neural Networks*. Washington, DC.
- Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73–193.
- Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 17(2), 291–321.
- Plunkett, K. & Marchman, V. (1991a). Connectionism and developmental theory. *Psikologisk Skriftserie Aarhus*, 16, 1–77. (Auch publiziert in: *British Journal of Developmental Psychology* 10 (1992), S. 209–254)

---

LITERATURVERZEICHNIS

---

- Plunkett, K. & Marchman, V. (1991b). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition*, 38, 1–60.
- Plunkett, K. & Sinha, C. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, 10, 209–254.
- Port, R. & van Gelder, T. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.
- Putnam, H. (1960). Minds and machines. In S. Hook (Hrsg.), *Dimensions of Mind*. Collier Books.
- Putnam, H. (1967). The mental life of some machines. In H.-N. Castañeda (Hrsg.), *Intentionality, Minds, and Perception*. Wayne State University Press.
- Putnam, H. (1975). *Mind, Language, and Reality*. Cambridge, MA: Cambridge University Press.
- Pylyshyn, Z. (1986). *Computation and Cognition*. Cambridge, MA: MIT Press.
- Ramsey, W., Stich, S. P. & Garon, J. (1991). Connectionism, eliminativism, and the future of folk psychology. In W. Ramsey, S. P. Stich & D. E. Rumelhart (Hrsg.), *Philosophy and Connectionist Theory* (S. 199–228). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rickheit, G. & Strohner, H. (1993). *Grundlagen der kognitiven Sprachverarbeitung*. Tübingen: Francke Verlag.
- Rosenberg, C. & Sejnowski, J. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, 1, 145–168.
- Roth, G. (1986). Selbstorganisation – Selbsterhaltung – Selbstreferentialität: Prinzipien der Organisation der Lebewesen und ihre Folgen für die Beziehung zwischen Organismus und Umwelt. In A. Dress, H. Hendrichs & G. Küppers (Hrsg.), *Die Entstehung von Ordnung in Natur und Gesellschaft* (S. 149ff.). München: Piper.
- Roth, G. (1987). Autopoiese und Kognition: Die Theorie H. R. Maturanas und die Notwendigkeit ihrer Weiterentwicklung. In S. I. Schmidt (Hrsg.), *Der Diskurs des Radikalen Konstruktivismus* (S. 256–286). Frankfurt: Suhrkamp.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1984). Learning internal representations by error backpropagation. In D. E. Rumelhart & J. L. McClelland (Hrsg.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1, S. 318–362). Cambridge, MA: MIT Press.

## LITERATURVERZEICHNIS

---

- Rumelhart, D. E. & McClelland, J. L. (1986). On learning past tenses of English verbs. In D. E. Rumelhart & J. L. McClelland (Hrsg.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, S. 216–271). Cambridge, MA: MIT Press.
- Rutkowska, J. C. (1995). Can development be designed? what we may learn from the cog project. In *Proceedings of the Third European Conference on Artificial Life*. Springer Verlag.
- Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Hrsg.), *Experimental brain research* (S. 129–144). New York: Springer-Verlag.
- Schade, U. (1992). *Konnektionismus — Zur Modellierung der Sprachproduktion*. Opladen, Germany: Westdeutscher Verlag.
- Schade, U. (1995). *Konnektionistische Sprachproduktion*. Unveröffentlichte Habilitationsschrift, Fakultät für Linguistik und Literaturwissenschaften, Universität Bielefeld. (Erscheint 1997 bei: Opladen: Westdeutscher Verlag)
- Schade, U. & Berg, T. (1992). The role of inhibition in a spreading–activation model of language production. Part II. The simulational perspective. *Journal of Psycholinguistic Research*, 21(6), 435–462.
- Schank, R. & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schriefers, H., Meyer, A. & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture word interference studies. *Journal of Memory and Language*, 29, 86–102.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–458.
- Searle, J. R. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. R. (1984). Intentionality and its place in nature. *Synthese*, 61, 3–16.
- Searle, J. R. (1990). Is the brain's mind a computer program? *Scientific American*, 262, 20–25.
- Searle, J. R. (1993). The failures of computationalism. *Think*, 2, 68–73.
- Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology*, 1, 183–211.

## LITERATURVERZEICHNIS

---

- Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Hrsg.), *Sentence Processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In *The Production of Speech* (S. 109–136). New York: Springer Verlag.
- Skarda, C. A. & Freeman, W. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10, 161–195.
- Skinner, B. F. (1933). *Science and Human Behavior*. Macmillan.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–74.
- Stemberger, J. P. (1982). *The Lexicon in a Model of Language Production*. Promotionschrift, University of California, San Diego. (Veröffentlicht bei: New York: Garland, 1985)
- Stemberger, J. P. (1983). *Speech Errors and Theoretical Phonology: A Review*. Bloomington, IN: Indiana University Linguistics Club.
- Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, 1, 281–313.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Hrsg.), *Progress in the Psychology of Language* (Vol. 1, S. 143–186). London: Erlbaum.
- Stemberger, J. P. (1989). Speech errors in early language production. *Journal of Memory and Language*, 28, 164–188.
- Stemberger, J. P. (1992). A connectionist view of child phonology. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Hrsg.), *Phonological Development* (S. 165–189). Timonium, MD: York Press.
- Sterelny, K. (1990). *The Representational Theory of Mind*. Cambridge, MA: Basil Blackwell.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.

---

LITERATURVERZEICHNIS

---

- Stoet, G. (1993). *On the proper treatment of back-propagation* (Unveröffentlichtes Papier). Fakultät für Linguistik und Literaturwissenschaften, Universität Bielefeld.
- Teuber, H. L. (1955). Physiological psychology. *Annual Review of Psychology*, 9, 267–296.
- Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge, MA: MIT Press.
- Touretzky, D. S. (1990). BoltzCONS: Dynamic symbol structures in a connectionist network. *Artificial Intelligence*, 46, 5–46.
- Turing, A. (1937). On computable numbers with an application to the Entscheidungs problem. In *Proceedings of the London Mathematical Society* (Vol. 42, S. 230–265).
- Turing, A. M. (1964). Computing machinery and intelligence. In A. R. Anderson (Hrsg.), *Minds and Machines*. Englewood Cliffs NJ: Prentice Hall.
- van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14, 355–384.
- van Gelder, T. (1991). What is the “D” in PDP? A survey of the concept of distribution. In W. Ramsey, S. P. Stich & D. E. Rumelhart (Hrsg.), *Philosophy and Connectionist Theory* (S. 33–59). Hillsdale, NJ: Lawrence Erlbaum Associates.
- van Gelder, T. (1992). *What might cognition be if not computation?* (Tech. Rep. 75). Indiana University: Department of Cognitive Sciences.
- Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology*, 72, 175–196.
- Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, 3, 9–30.
- Wheeler, M. (1994a). *For whom the bell tolls? The role of representation and computation in the study of situated agents*. (Cognitive Science Research Paper 320). University of Sussex, Brighton, UK: School for Cognitive and Computing Sciences.
- Wheeler, M. (1994b, März). *The Presence of a Mind* [Unveröffentlichter Vortrag im Rahmen eines “Philosophy Society”-Symposiums]. University of Sussex, Brighton, UK.
- Wängler, H.-H. (1963). *Rangwörterbuch hochdeutscher Umgangssprache*. Marburg: N. G. Elwert Verlag.