

Zur theoretischen und psycholinguistischen Grundlegung der Computerlinguistik.

1. Einleitung

Die Frage nach dem Sprachverständnis des Computers läßt sich auf zweierlei Weisen auf-fassen: Einmal, nach herkömmlicher Art, wie Menschen mit Computern kommunizieren und in welcher Sprache sie dies tun. In diesem Sinne ist die Frage bereits von der Informatik beantwortet. Die etablierten Verständigungs-codes zwischen Mensch und Computer sind Programmiersprachen. Diese sind gekennzeichnet durch abgeschlossene, bekannte Systeme syntaktischer Regeln und semantischer Interpretation .

Es läßt sich aber auch nach dem Sprachtyp fragen, den Computer überhaupt verstehen können, also insbesondere danach, ob Computer natürliche Sprachen verstehen können. Mit natürlicher Sprache ist der Sprachtyp gemeint, den Kinder mühelos in einer sprachlichen Umgebung erwerben und den wir als kompetente Sprecher mühelos in unterschiedlichsten Situationen verwenden.

Lassen sich also Computer so programmieren, daß sie auch natürliche Sprache 'verstehen'? Diese Frage ist nicht trivial.

Zum einen deswegen, weil sich natürliche Sprachen hinsichtlich des Bekanntheits-grades von syntaktischer und semantischer Struktur gravierend von Programmiersprachen (formalen Sprachen) unterscheiden. Formale Sprachen werden durch elementare Vokabularen und Produktionensysteme generiert, die wohlbekannt sind. Für natürliche Sprachen versucht die Grammatiktheorie, diese erst zu finden.

Zum anderen, weil das Verstehen natürlicher Sprache stets durch situative Kontexte und konzeptuelles Wissen beeinflusst ist. Diese Eigenschaften besitzen formale Sprachen nicht.

Um über das Sprachverständnis des Computers in diesem Sinne reden zu können, soll zunächst angedeutet werden, welche Informationen Menschen verarbeiten, wenn sie Sprache verstehen. Daran anschließend soll am Beispiel der syntaktischen Analyse eines Satzes durch den Computer gezeigt werden, wie uns eine derartige Modellierung zu einem besseren Ver-ständnis der entsprechenden Prozesse führen kann, die beim menschlichen Sprachverständnis anzunehmen sind.

Zum Verständnis einer Äußerung wie (1)

(1) Wir haben gestern ein Faß aufgemacht.

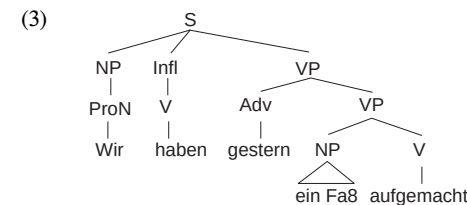
sind verschiedene Analyseschritte notwendig. (' signalisiert hervorgehobene Betonung.)

Stellen wir uns zunächst vor, daß -im physikalischen Sinne- Schallwellen unser Ohr errei-chen. Diese müssen -unabhängig von Stimmhöhe und Modulation- als Laute der deutschen Sprache identifiziert und in bestimmte Lautsequenzen zerlegt werden, die ihrerseits mit den entsprechenden lexikalischen Einheiten unseres mentalen Lexikons assoziiert werden. Diese Dekodierung vollzieht sich -linguistisch gesehen- auf der phonologisch-morphologischen Ebene.

Dieser erste, für das Verständnis von (1) elementare Schritt ist bisher weder in der psycholinguistischen Forschung noch in der Modellierung durch den Computer völlig geklärt. Ein in dieser Grobskizze zweiter Schritt ist die syntaktische Analyse. Da wir diesen Aspekt später noch ausführlicher darstellen, sei hier nur das folgende bemerkt: (1) wird offenbar so verstanden, daß ein Subjekt an einem Objekt eine adverbial spezifizierte verbale Operation durchführt. Dies läßt sich wie (2) darstellen

(2) (SU (Adv (Prät (V (Obj)))))

oder als Strukturbaum (3)



Daß eine derartige Analyse auch psycholinguistisch plausibel ist, geht daraus hervor, daß wir auch einen 'nonsense'-Satz wie (4), dem wir keine Bedeutung zuordnen können, in gewissem Sinne als 'grammatisch' erkennen und ihm eine entsprechende Struktur zuordnen.

(4) Die Binken haben uker ein Fretz gemorkelt.

Es liegt nahe, anzunehmen, daß die bisher unterschiedenen Analyseebenen niemandem während des Verstehensprozesses bewußt sind. Dennoch müssen wir diese Berechnungsschritte zugrunde legen, um das wahrgenommene Schallereignis in seine Bedeutung zu übersetzen; in informationstheoretischen Worten heißt dies, daß die ursprüngliche Information (physikalische Schallwellen) über verschiedene Berechnungen so manipuliert (verarbeitet) werden muß, daß sie die zugehörige Bedeutung repräsentiert. Zum Verständnis von (1) gehört natürlich mehr als mit (3) ausgedrückt wurde. So müssen wir auch eine semantisch-referentielle Interpretation leisten, etwa so: Das Subjekt 'wir' referiert auf eine nicht näher spezifizierte aber kontextuell identifizierbare

Menge von Individuen, deren eines der Sprecher der Äußerung ist, und die den Hörer enthalten kann, aber nicht muß. Die damit kongruente Verbform 'haben' verweist in Beziehung zu der partizipialen Perfektform 'aufgemacht' auf einen nicht näher spezifizierten Zeitraum in der Vergangenheit usw.

Damit kennen wir die wörtliche Bedeutung von (1). Darüberhinaus wissen wir aber, daß mit (1) in übertragener Bedeutung auch so etwas wie (5) gemeint sein kann:

(5) Wir haben gestern eine Party gegeben/ein Fest arrangiert.

Zudem hebt die Intonation von (1) zwei Elemente besonders hervor. Dies wird gewöhnlich in dem Sinne als Kontrast gedeutet, daß (1) gegenüber den denkbaren Alternativen ausgezeichnet wird, die bis auf Variation der hervorgehobenen Elemente identisch sind. So steht (1) etwa im Kontrast zu (6) bis (8):

(6) ihr habt heute ...

(7) sie haben yorgestern ...

(8) ich habe letzte Woche ...

Die Funktion dieser kontrastiven Hervorhebung ermitteln wir kontext- und situationsabhängig (pragmatisch) und verstehen daher (1) u.U. auch als Aufforderung an jemanden, doch seinerseits für das nächste Fest geradezustehen.

Die bisher isolierten Analyseebenen und das sprachliche Wissen, welches die Analysen, die für ein adäquates Verständnis von (1) notwendig sind, lassen sich schematisch wie (9) darstellen:

(9) verfügbares Wissen: Parsingprozesse:

phonologische Regeln	phonologisch
morphologische Regeln	morphologisch
mentales Lexikon	lexikalisch
grammatische Prinzipien	syntaktisch
semantische Regeln	semantisch
Weltwissen	pragmatisch

Die Interaktion der hier dargestellten Kenntnissysteme, die unser sprachliches Wissen repräsentieren, ist noch weitgehend ungeklärt. Manche der hier beteiligten Prozesse haben offensichtlich mit 'Weltwissen' zu tun (Pragmatik, kontextuelle Semantik) oder beziehen sich auf Ergebnisse anderer psychischer Erkenntnisprozesse. Derartige Prozesse sind nicht autonom. Demgegenüber lassen neuere psycholinguistische Überlegungen vermuten, daß insbesondere die syntaktische Analyse ein für die Sprachverarbeitung zentraler autonomer mentaler Prozess ist, dessen Prinzipien sich nicht aus anderen Kenntnissbereichen erklären lassen. Im folgenden sollen zunächst einige wesentliche Eigenschaften eines solchen autonomen mentalen Moduls der Sprachverarbeitung dargestellt werden.

Dieses Modul wird sodann im Rahmen eines universalgrammatischen Ansatzes in ein Modell integriert, welches die phylo-, onto- und aktualgenetischen Zusammenhänge ausdrückt. Schließlich soll gezeigt werden, daß die Entwicklung computergesteuerter Simulationen zu einem adäquaten Verständnis von Sprachverarbeitung führen kann, und wie ein solches Simulationsmodell in erster Annäherung zu konzipieren ist.

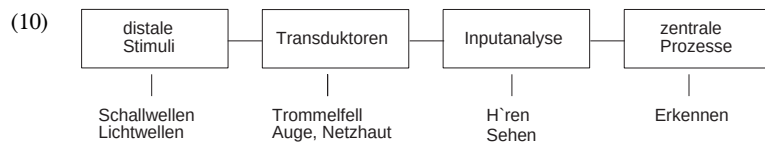
2. Künstliche Intelligenz, Sprachverarbeitung und Modularität.

"Künstliche Intelligenz (KI) ist die Beschäftigung mit den Methoden, die es dem Computer ermöglichen, Aufgaben zu lösen, zu deren Lösung Intelligenz notwendig ist, wenn sie vom Menschen durchgeführt werden." Diese von Minsky (1968) vorgeschlagene Charakterisierung dessen, was Gegenstand der KI-Forschung ist, bezieht sich auf ein intuitives Verständnis von menschlicher Intelligenz. Eine Explikation von menschlicher im Gegensatz zu künstlicher Intelligenz erscheint hingegen nur extensional möglich, d.h. durch Aufzählung derjenigen psychischen Prozesse, die in irgendeinem Sinne Intelligenz vermuten lassen. Die Möglichkeit einer intensionalen Definition wird ebenso wie die Möglichkeit einer Theorie intelligenten Verhaltens in der neueren kognitionspsychologischen Forschung (etwa Fodor 1983) generell angezweifelt. Demnach ist der menschliche Geist ein offenes Berechnungssystem, das ständig im informationellen Austausch mit seiner Umgebung steht. Distale Stimuli aus der den Menschen umgebenden Umwelt werden über verschiedene Wahrnehmungssysteme verarbeitet und in gewissem Maße im Gedächtnis gespeichert. Demgegenüber sind Computer typischerweise keine offenen Systeme. Sie lassen sich als universelle Turingmaschinen modellieren, deren interne Aktionen sich auf wenige wohldefinierte Teiloperationen beschränken, die während ihres Operierens keine externen Informationen aufnehmen können: Bewege das Band, füge hinzu, speichere, lies, drucke usw. Diese Informationen sind als automateninterne Kenntnisse zu beschreiben.

Traditionell unterschiedene Wahrnehmungssinne, wie Sehen, Schmecken, Hören

usw. lassen sich nun als informationsverarbeitende Module auffassen, die in sehr spezifischer Weise sehr spezifische Informationen verarbeiten. So verarbeiten wir etwa visuell nur Lichtwellen, gustatorisch nur gewisse Molekülverbände, auditorisch nur Schallereignisse. Um diese Informationen aber zum Gegenstand zentraler Prozesse (Intelligenz, Denken) machen zu können, müssen sie aufbereitet werden, m.a.W. sie müssen in eine zentrale Verarbeitungssprache übersetzt werden.

Ganz analog gilt für Rechenmaschinen, daß es keine Compiler für arbiträre Codes gibt. Jede Programmiersprache muß von einem spezifischen Compiler in die zentrale Verarbeitungssprache (den binären Code) des Computers übersetzt werden. Aus diesen Überlegungen läßt sich das folgende funktionale Modell ableiten:



Die Transduktoren-Analyse verarbeitet distale Stimuli und übersetzt diese in eine Repräsentation, die als Eingabe für die sogen. Inputanalyse dient, die ihrerseits diese Information so manipuliert, daß sie von zentralen Prozessen, wie etwa Problemlösen und allgemeines Nachdenken, weiterverarbeitet werden kann.

Die jeweils spezifischen Inputsysteme stellen autonome signalverarbeitende mentale Module dar (Sehen geschieht unabhängig von Schmecken und Hören). Solche Module haben gewisse typische gemeinsame Eigenschaften, die für zentrale Prozesse (Denken) nicht gelten. In unserem Zusammenhang ist nun interessant, daß man auch die Sprachverarbeitung in wesentlichen Aspekten als ein derartiges Modul ansehen kann. Welche Eigenschaften zeichnen nun solche Module aus?

(i) Inputsysteme sind domänenspezifisch: Der Bereich der von Inputsystemen zu lösenden Probleme ist scharf begrenzt. Unser Sprachverarbeitungssystem taugt z.B. weder dazu das Quietschen einer Tür noch dazu, die Syntax einer formalen Sprache automatisch und in der gleichen Schnelligkeit zu analysieren wie eine weitaus komplexere, natürliche Sprache.

(ii) Inputsysteme operieren zwangsläufig: Unabhängig von unserem Willen und Wünschen können wir nicht anders als Sprache als Sprache zu verstehen. (Versuchen Sie einmal wegzuhören, wenn ihr Gegenüber ihnen etwas erzählt.) In diesem Sinne verläuft Inputanalyse gewissermaßen reflexhaft.

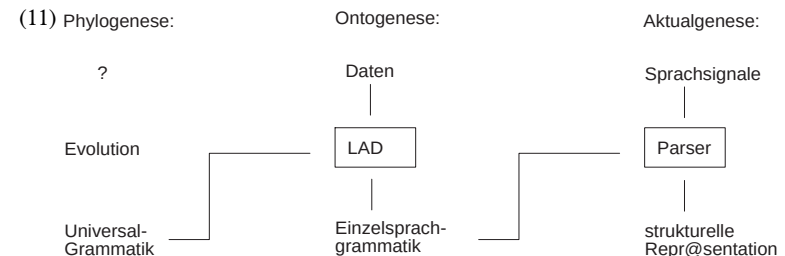
(iii) Inputsysteme sind schnell: Sprachverstehen gehört mit zu den schnellsten psychischen Prozessen. Obwohl die Komplexität der Teiloperationen zum Lösen einer einfachen Multiplikationsaufgabe wie etwa $3,14 \times 4,31$ wesentlich geringer ist als beim Verstehen eines gesprochenen Satzes, benötigen wir doch erheblich mehr Zeit, um das Ergebnis der Multiplikation zu berechnen.

(iv) Inputsysteme sind informationell abgeschlossen: Unabhängig von unserem Hintergrundwissen operieren Inputsysteme rein datenorientiert. Damit hängt die folgende Eigenschaft zusammen.

(v) Inputsysteme liefern einen seichten Output: Sie sind aufgrund ihrer Schnelligkeit und informationellen Abgeschlossenheit relativ unintelligent, d.h. verarbeiten nur minimale Mengen an Information.¹

Die genannten Eigenschaften gelten für zentrale Prozesse nicht. Man mache sich z.B. klar, wie lange sich über praktische, künstlerische oder wissenschaftliche Probleme nachdenken läßt, und wieviel Information dabei konsultiert wird. Derartige Überlegungen machen klar, daß allgemeine Problemlösungsprozesse (zentrale Prozesse) strikt von autonomen, modularen Prozessen wie etwa der Sprachverarbeitung zu trennen sind.

Die Sprachverarbeitung ist uns Menschen offenbar im Gegensatz zu anderen Säugetieren von Geburt an gegeben und entwickelt sich in der relativ kurzen Zeit des sog. 'Spracherwerbs'. Ein derartiger 'nativistischer' Ansatz nimmt an, daß sich in der Evolution bei menschlichen Organismen eine allgemeine Fähigkeit, zu sprechen, entwickelt. Diese sog. UG (= Universalgrammatik) -beim Individuum als angeborener LAD realisiert- führt ontogenetisch aufgrund relativ weniger Sprachdaten zum Erwerb einer muttersprachlichen Grammatik, die ihrerseits einen mentalen Prozessor -genannt Parser- steuert, der aktualgenetisch Sprachsignale in strukturelle Repräsentationen übersetzt, die die Voraussetzung für die Berechnung der Bedeutung von Äußerungen darstellen. Die folgende Graphik soll diese Zusammenhänge verdeutlichen.



Die Aufgabe des Linguisten besteht nun darin, die Struktur des in einem derart verstandenen Sprachverarbeitungsmoduls zugrundeliegenden Wissens und die Organisation der Verarbeitungsprozesse zu explizieren.

3. Linguistische Theorien und Computersimulationen.

Linguistische Theorien, wie etwa die generative Grammatik, stellen solche Explika-

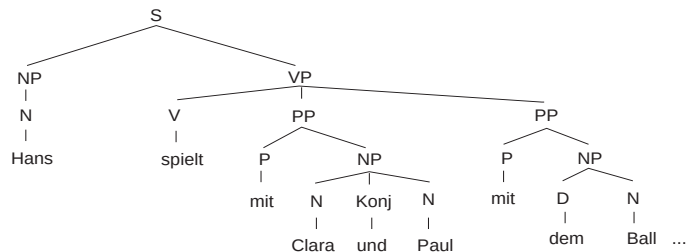
tionsversuche dar. Um nun zu zeigen, wie entsprechend solcher Theorien etwa syntaktische Analysen ablaufen könnten, bietet sich der Computer sowohl als Forschungsinstrument aber auch als Verständnishilfe geradezu an.

Im folgenden wollen wir deshalb ein Simulationsmodell für einen natürlichsprachlichen Parser vorstellen. Dabei ist entscheidend, daß ein derartiges Parsingmodell den psychischen Vorgang möglichst angemessen abbilden soll. Das läßt sich am besten dadurch demonstrieren, daß der computergesteuerte Prozess die gleichen Fehler macht wie sie beim menschlichen Sprachverstehen ebenfalls auftreten. Dazu bieten sich die aus der psycholinguistischen Forschung bekannten Holzweg-Sätze ('garden path') an:

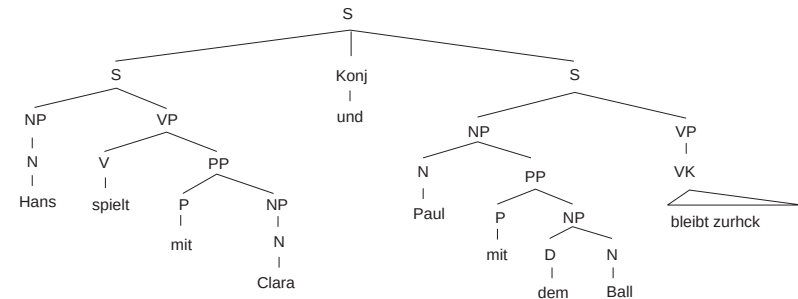
(12) Hans spielt mit Clara und Paul mit dem Ball bleibt zurück.

An der Stelle, wo das 'und' auftaucht, müssen wir eine syntaktische Entscheidung treffen: koordiniert das 'und' zwei Nominalphrasen, also 'Clara' und 'Paul', oder koordiniert es die beiden Sätze? Die Lesart, bei der das 'und' Nominalphrasen koordiniert, wird aufgrund des Verbs 'spielen' vorgezogen, so daß eine entsprechende syntaktische Struktur aufgebaut wird. Diese erweist sich jedoch dann als falsch, wenn im folgenden 'bleibt zurück' gelesen wird. Die aufgebaute Erwartung hinsichtlich der syntaktischen Struktur muß also zugunsten einer Satzkoordination geändert werden. Den unterschiedlichen syntaktischen Strukturen entsprechen die folgenden Graphen:

(13) (i) die Holzweg-Analyse:

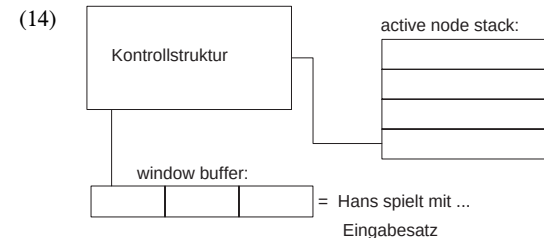


(ii) die richtige Analyse:



Die Strukturen unter (13)(i) sind fakultativ. Die PP 'mit dem Ball' kann entweder Adverbial oder Präpositionalattribut sein. In beiden Fällen ist die Struktur hinsichtlich des weiteren Satzverlaufs nicht adäquat. Da derartige Prozesse nicht bewußt sind, haben wir prima facie gute Gründe anzunehmen, daß sie autonom stattfinden.

Das folgende Modell (14) eines sog. deterministischen Kellerautomaten zeigt, wie sich derartige Phänomene auf dem Computer simulieren lassen. Der Parser liest den zu analysierenden Satz sukzessive von links nach rechts ein (window buffer) und baut gleichzeitig syntaktische Strukturen auf, die im sog. 'active node stack' gespeichert werden. Das entspricht auf einer hinreichend abstrakten Berechnungsebene in gewissen Hinsichten der menschlichen Sprachverarbeitung:



Ein deterministischer Kellerautomat kann aufgefaßt werden als 'on line'-Turingmaschine mit zwei Bändern, die bzgl. ihrer Schreib- und Lesemöglichkeiten beschränkt ist. So ist der Zugriff im Kellerspeicher (active node stack) nur auf das zuletzt abgelegte Kellersymbol möglich, während das Eingabeband (window buffer) von beschränkter Länge, dafür aber mit beliebigem Zugriff vorgesehen ist.

Eine Konfiguration des Automaten während der Analyse mag dies verdeutlichen. Der Kellerautomat speichert die Konstituenten, die bis zum Zeitpunkt, an dem das 'und' gelesen wird, unvollständig sind. Das Eingabeband liest den Satz von links nach rechts

und enthält höchstens drei Konstituenten. Der Schnappschuß unter (15) zeigt die aufgebauten Teilkonstituenten:

(15) active node stack:

(S (NP (N Hans)))
(VP (V spielt))
(PP (P mit))
(NP (N Clara))

window buffer:

(Konj und)		
------------	--	--

bisher ungesehen: Paul mit dem Ball bleibt zurhck

Im Verlauf des Parsings muß an dieser Stelle eine Entscheidung darüber getroffen werden, ob das 'und' zwei NPs oder zwei S-Knoten koordiniert. Dies kann allerdings in einem 'on line'-Prozeß nicht eindeutig entschieden werden; auch dann nicht, wenn der Parser den Lesespeicher füllen würde, um aus nachfolgendem lexikalischen Material weitere Informationen über den Satzverlauf zu gewinnen. Die lokale Ambiguität wird erst wesentlich später aufgelöst, so daß es scheint, als würde die Entscheidung hier arbiträr gefällt. Wenn dies so wäre, sollte man erwarten, daß verschiedene Leser in gleicher Verteilung beide Strukturen aufbauen, d.h. für manche Leser würde dieser Satz einen 'garden path' darstellen und für andere nicht. Dies ist aber offenbar nicht der Fall, so daß es gerechtfertigt scheint, gewisse Präferenzen des syntaktischen Operierens anzunehmen. Eine solche Parsingstrategie wäre etwa das 'minimal attachment' (Frazier 1978), d.h., daß bei syntaktischer Ambiguität, die auch semantisch für einen lokalen Bereich nicht entscheidbar ist, minimale Adjunktion vor zusätzlicher Strukturierung präferiert wird. Genau dies trifft im vorliegenden Fall zu. Der Parser koordiniert zwei NPs, weil dies im syntaktischen Sinne die einfachste der möglichen Alternativen darstellt. Zur Satzkoordination wären zusätzliche Teiloperationen nötig (etwa CREATE S, ATTACH NP usw.). Daß diese Parsingstrategie nicht zum Erfolg führt, klärt sich erst zu dem Zeitpunkt, an dem 'bleibt zurück' im Lesespeicher sichtbar wird. Der Parser hat bis dahin eine Struktur aufgebaut, die mit dem nun folgenden lexikalischen Material nicht mehr kompatibel ist. Ein weiterer Schnappschuß mag dies verdeutlichen:

(16) active node stack:

(S (NP (N Hans))
(VP (V spielt)
(PP (P mit)
(NP (NP (N Clara))
(Konj und)
(NP (N Paul))))
(PP (P mit)
(NP (D dem)
(N Ball))))))

window buffer:

(V bleibt)	(Adv zurhck)	
------------	--------------	--

Für diese Konfiguration gibt es kein grammatisches Prinzip, welches weitere Strukturierung erlauben würde, so daß der computergesteuerte Parser auf dergleichen strukturellen Holzweg geraten ist wie der menschliche Leser dieses Satzes.

Dies ist nicht etwa auf die Insuffizienz des Algorithmus zurückzuführen, sondern gerade im Sinne einer adäquaten Simulation menschlicher Sprachverarbeitung beabsichtigt. Es ist sicherlich möglich, Algorithmen zu entwickeln, die mit dem gezeigten Phänomen keinerlei Schwierigkeiten haben, doch ist ein solcher Algorithmus im Hinblick auf mentale Sprachverarbeitungsprozesse wenig aufschlußreich. Eine generelle Abgrenzung der Parsingstrategien (nicht nur) hinsichtlich der Behandlung von 'garden path'-Sätzen besteht in der Unterscheidung von sequentiell und parallelem Prozessieren. Derartige Parser (etwa ATN-Parser mit Zurücksetzen für sequentielle Abarbeitung oder sog. 'chart-Parser' für paralleles Prozessieren) sind keine guten Kandidaten für Simulationszwecke, weil ihr Berechnungsmodus zur Erklärung des 'garden path'-Phänomens prinzipiell nichts beiträgt. So arbeiten sequentielle Algorithmen stets nach dem Prinzip des Zurücksetzens, sagen also 'garden path'-Analysen voraus, die kein Leser feststellt. Parallel arbeitende Algorithmen verfolgen stets alle gleichzeitig möglichen Alternativen und verwerfen diejenigen, die sich im Analyseverlauf als falsch erweisen, d.h. aber, daß für diese Parser überhaupt kein 'garden path'-Satz existiert. Dieses kurze Beispiel vermag vielleicht einen Eindruck davon zu vermitteln, wie Computersimulationen zu einem adäquateren Verständnis der menschlichen Sprachfähigkeit führen können. Daß eine syntaktische Analyse diejenigen Berechnungsprozesse, die zum Verständnis von kontextuell gebundenen Äußerungen erforderlich sind, nicht voll erfaßt, haben die eingangs erwähnten Erläuterungen gezeigt. Daß wir aber gute Gründe haben, autonome modulare Berechnungssysteme in der Sprachverarbeitung anzunehmen, das lehrt nicht nur die Psycholinguistik, sondern das fordert die Computerlinguistik geradezu, und die Entwicklung von Parsern, die linguistische Theorien in Verarbeitungsprozesse umsetzen, erweitert und verschärft unser

Verständnis der den entsprechenden psychischen Verarbeitungsprozessen zugrundeliegenden Prinzipien. So hilft uns also das 'Sprachverständnis' des Computers, unser eigenes Sprachverständnis besser zu verstehen.

Literatur:

- Berwick, Robert & A. Weinberg, 1984. *The Grammatical Basis of Linguistic Performance - Language Use and Acquisition*. Cambridge, Mass.: MIT-Press.
- Chomsky, Noam, 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam, 1986: *Barriers*. Cambridge, Mass.: MIT-Press (= Linguistic Inquiry Monographs 13).
- Fanselow, Gisbert & S. Felix, 1987. *Sprachtheorie (Bände I & II) Eine Einführung in die generative Grammatik*. Tübingen: Francke (= UTB 1441 & UTB 1442).
- Frazier, Lyn, 1978. *On Comprehending Sentences: syntactic parsing strategies*. Bloomington: Indiana University Linguistic Club.
- Fodor, Jerry, 1983. *The Modularity of Mind*. Cambridge, Mass.: MIT-Press.
- Marcus, Mitchell, 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge, Mass.: MIT-Press.
- Winograd, Terry, 1983. *Language as a Cognitive Process*. Reading, Mass: Addison Wesley.