Diplomarbeit im Fach Informatik

# Repräsentation und Verarbeitung sprachlicher Indikatoren für kognitive Ressourcenbeschränkungen

André Berthold
Universität des Saarlandes

Lehrstuhl Prof. Dr. Wolfgang Wahlster

# Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, daß ich diese Diplomarbeit ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe.

Saarbrücken, den 31.12.1998

André Berthold

André Berthold
Wöllnerstraße 10
04229 Leipzig
(E-mail: berthold@coli.uni-sb.de)

Diese Diplomarbeit wurde am Lehrstuhl für Künstliche Intelligenz von Prof. Dr. Wolfgang Wahlster an der Universität des Saarlandes im Rahmen des Projektes READY des SFB 378 der DFG angefertigt. Ich möchte mich bei allen Projektmitarbeitern für ihre Unterstützung bedanken, vor allem bei Dr. Anthony Jameson für seine Mühe und Geduld beim Probelesen und seine zahlreichen hilfreichen Anregungen und Korrekturen in vielen individuellen Besprechungen. Für Hilfe und Ideen für und bei der Implementation danke ich vor allem Thomas Weis, aber auch Ralph Schäfer und Thomas Weyrath. Wojciech Skut und Thorsten Brants vom Projekt NEGRA sei dafür gedankt, daß sie mir ihren Chunker zur Verfügung stellten und Wojciech besonders für die Geduld bei der Umsetzung meiner Modifikationswünsche. Dank für die gute Zusammenarbeit mit dem Projekt LISA gebührt besonders Claire Gardent und Stephan Thater.

Ein Dankeschön für's ausgiebige Korrekturlesen erhalten Markus Becker, Christian Braun, Jürgen Trouvain, Thomas Weis und Sascha Brawer. Flooije bin ich für alles dankbar, was nicht aufgezählt werden kann; besonders dafür, daß sie meine kognitive Belastung und meinen Zeitdruck geduldig ertragen hat.

# Inhaltsverzeichnis

# Abbildungsverzeichnis

# Tabellenverzeichnis

# Kapitel 1

# Einleitung

## 1.1 Motivation

Wenn man sich an ein Auskunftssystem der Deutschen Bahn wendet, erwartet man, daß die gewünschte Information strukturiert und verständlich dargestellt wird und irrelevante Informationen vernachlässigt werden. Diktieren die Damen und Herren aus der Telefonzentrale des Saarbrücker Hauptbahnhofs einem Anrufer die gewünschten Anschlußzeiten, geschieht das mit einer Geschwindigkeit, die an die Möglichkeiten des Anrufers angepaßt ist, diese Information auch zu notieren oder zu behalten. Unverständliche Information wird auf Nachfrage wiederholt und die Information, daß es auch einen billigeren Interregio nach Hamburg gibt, wenn der Anrufer mit gehetzter Stimme nach der nächsten IC-Verbindung gefragt hat, wird gar nicht erst erteilt.

Einen Kommilitonen, den man wenige Tage vor dem Abgabetermin seiner Magister- oder Diplomarbeit in der Bibliothek trifft, wo er noch dringend nach letzten Seitenangaben für genaue Referenzen in seinem Literaturteil sucht sucht, wird man nicht lange mit dem Bericht über den neuesten Kinofilm aufhalten, zu den man ihn nach Abgabe der Arbeit einladen möchte. Die Gedanken des Dialogpartners werden dafür noch zu stark beim bevorstehenden Abgabe- oder Prüfungstermin weilen als beim aktuellen Gespräch.

Diese Anpassungen der Informationspräsentation an den Kommunikationspartner beruhen nicht zuletzt auf einer Einschätzung seiner kognitiven Fähigkeiten der Informationsverarbeitung und seiner verfügbaren Zeit. Im zuerst geschilderten Fall ergibt sich zusätzlich die Problematik, daß weder visuelle Informationen noch irgendeine andere Form des Wissens über den Dialogpartner zur Verfügung stehen, sondern einzige Informationsquelle lediglich Form und Inhalt seiner sprachlichen Äußerungen sind.

Beschränkungen der kognitiven Verarbeitung und der verfügbaren Zeit lassen sich auch bei der Interaktion von Menschen mit Maschinen und Softwaresystemen erwarten. Einem Entwurf der Schnittstelle zwischen Mensch und Maschine, der diese Beschränkungen berücksichtigen kann, kommt daher eine herausragende Rolle zu: Nicht nur in kritischen Umgebungen, wie z.B. bei der Steuerung von Kernkraftwerken, Flug- und Fahrzeugen kann durch eine ergonomische Gestaltung der Schnittstelle der Umgang des Menschen mit der Maschine erheblich erleichtert und der Benutzer bei der Realisierung seiner Ziele unterstützt werden. Grundgedanke muß dabei stets sein, daß die Maschine sich an ihren menschlichen Benutzer anpaßt und nicht der umgekehrte Fall erzwungen wird.

Die vorliegende Arbeit soll in diesem Zusammenhang einen Beitrag dazu leisten, Symptome für die beiden spezifischen situativen Einschränkungen Zeitdruck und kognitive Belastung in den sprachlichen Äußerungen eines Menschen zu identifizieren, ihren diagnostischen Wert für die Verwendung in der Benutzermodellierungskomponente eines Dialogsystems zu beurteilen und ihre Verarbeitungsmöglichkeiten aufzuzeigen.

## 1.2 Übergeordnete Fragestellungen

**Ergonomie und KI** Die Berücksichtigung von spezifischen Eigenschaften eines Benutzers in komplexen Softwaresystemen hat eine wichtige Stellung in der Künstlichen Intelligenz (KI) und der ergonomischen Forschung inne. Letztgenannte beschäftigt sich vor allem mit den physischen, physiologischen und kognitiven Voraussetzungen des Entwurfs von Benutzerschnittstellen. Dagegen versucht man auf dem Gebiet der Benutzermodellierung in der KI vor allem Methoden zu entwickeln, die einerseits die Erkennung der relevanten Eigenschaften eines Benutzers möglich machen und andererseits eine korrekte dynamische Anpassung von Softwaresystemen an die Veränderlichkeit dieser Eigenschaften absichern. Diese Techniken sind mittlerweile soweit entwickelt, daß sie bereits auch in kommerziellen Kontexten eingesetzt werden.

**Modellierung von Benutzereigenschaften** Zu den Benutzereigenschaften, die in diesem Rahmen bisher untersucht wurden, zählen u.a. der Wissensstand sowie Ziele und Pläne eines Benutzers bei der Interaktion mit einem Softwaresystem. Einen Überblick hierzu gibt z.B. der Band Jameson, Paris und Tasso (1997), ältere Ansätze sind in Kobsa und Wahlster (1989) zu finden. Im Kontext einer gestiegenen Mobilität und der Verbreitung der Computertechnologie in der heutigen Welt ist zu erwarten, daß die Kommunikationssituationen mit Softwaresystemen nicht mehr auf die weitgehend statische Athmosphäre einer Laborsitzung beschränkt bleiben, sondern mehr und mehr auch unterschiedliche Umweltbedingungen die Interaktion entscheidend prägen. Das bleibt nicht ohne Einfluß auf das Verhalten des Benutzers, da er neben der Aufgabe der Kommunikation mit dem System ständig gezwungen ist, auf wechselnde Kontextbedingungen zu reagieren. Dadurch wird für diesen Benutzer auch die Verfügbarkeit unterschiedlicher Ressourcen beeinträchtigt, wozu nicht zuletzt seine Konzentrationsfähigkeit und die für die Interaktion zur Verfügung stehende Zeit gehören.

**Spezifische Eingabemodalitäten** Soll sich ein Softwaresystem langsam an unterschiedliche Eigenschaften seines Benutzers anpassen, ist die erste Voraussetzung dafür, daß es die aktuellen Ausprägungen dieser Eigenschaften entweder durch direkte Beobachtung erkennen oder aus anderen Beobachtungen ableiten kann. Für Beobachtungen steht dem System, sofern es nicht über andere Mechanismen und Sensoren verfügt, ausschließlich die Benutzereingabe zur Verfügung. Diese Eingabe kann in verschiedenen Modalitäten vorliegen: Sprache, Bedienung von Schaltelementen oder die Auswahl von Menüoptionen etc. Bei der Spracheingabe ist außerdem zu unterscheiden zwischen akustischer und grafischer, d.h. visueller Eingabe, spontaner Sprache und geplanter oder vorstrukturierter Sprache, Hochsprache und Dialekt.

**Psychologische Zusammenhänge** Die Beurteilung kognitiver Ressourcen eines Sprechers steht in engem Zusammenhang zu den Erkenntnissen über das kognitive System des Menschen und sein Anpassungsverhalten an wechselnde Umweltbedingungen. Sprache als kognitiver Prozeß ist aus diesem System nicht herauslösbar. Daher werden Einflüsse kognitiver Belastung und des Zeitdrucks, die sich in einer Verhaltensänderung manifestieren, nicht ohne Auswirkungen auf die Sprachproduktion bleiben und liefern damit Erkenntnisse über andere, nicht sprachproduktionsbezogene Prozesse.

**Empirisches Fundament** Die Bewertung, welche Verhaltensänderungen für die Einschätzung spezifischer Benutzereigenschaften relevant sind, muß objektiv sein und unter klar definierten Kontextbedingungen gelten. Der Systementwurf bedarf daher einer empirischen Grundlage, die sich z.B. aus der experimentellen psychologischen Forschung und eigenen Korpusanalysen ergeben kann. Große Sorgfalt sollte der Verwendung dieser Beobachtungen für die Anpassung des Systemverhaltens zukommen: Das Anpassungsverhalten der Maschine sollte für den Benutzer möglichst nachvollziehbar bleiben, was impliziert, daß ein Benutzer auf die Ursachen für eine Änderung des Systemverhaltens zurückschließen können sollte. Das setzt voraus, daß sich ein Softwaresystem bei seinen Anpassungen auch auf diejenigen Symptome stützt, die auch menschliche Dialogpartner – bewußt oder unbewußt – zu einer ähnlichen Anpassung führen würde.

**Inferenzen unter Unsicherheit** Im allgemeinen wird sich die Anpassung an einen Dialogpartner auf Information aus verschiedenen Quellen stützen und nicht nur auf die Kontrolle eines evtl. vorliegenden Triggersignals. Sowohl die zugehörigen Beobachtungen als auch die Inferenzen auf den Beobachtungen zeichnen sich durch ein hohes Maß an Unsicherheit aus. Durch verschiedene mögliche Erklärungen für ein und dasselbe Symptom, Probleme bei der Informationsübermittlung (z.B. Kanalstörungen), individuell unterschiedliche Auswirkungen von Belastung und Zeitdruck, besondere kommunikative Gegebenheiten der Situation oder Probleme der automatischen Analyse sind die Inferenzen bei der Modellierung der Benutzereigenschaften mit wahrscheinlichkeitstheoretischen Ansätzen besser zu erfassen als etwa mit zweiwertigen Logiken. Das muß sich in der Wahl eines geeigneten Inferenzmechanismus widerspiegeln, der Schließen unter Unsicherheit erlaubt.

## 1.3 Einbettung der Arbeit in das Projekt READY

READY ist ein Projekt im Sonderforschungsbereich (SFB) 378 'Ressourcenadaptive kognitive Prozesse' der Deutschen Forschungsgemeinschaft, dessen Ziel die Untersuchung von Methoden der ressourcenadaptiven Dialogführung ist (siehe Wahlster, Jameson, Ndiaye, Schäfer & Weis, 1995b; Schäfer, Weis, Weyrath & Jameson, 1997). READY steht dabei als Akronym für **Re**ssourcen**a**daptives **D**ialogs**y**stem und bezeichnet gleichermaßen das Projekt und den dort entwickelten Prototypen.

Das Projekt baut auf Vorarbeiten aus der experimentellen psychologischen und sprachpsychologischen Forschung und aus der Künstlichen Intelligenz auf. Die Arbeiten aus dem Bereich der KI beschäftigen sich vor allem mit Techniken der Modellierung und Verwaltung von

Systemeinschätzungen und Fragen der Dialogplanung. Als Domäne für den zu implementierenden Prototypen wurde ein Pannenhilfeszenario ausgewählt, das später näher beschrieben wird. Grundgedanke ist, daß ein Anrufer, der mit seinem Fahrzeug auf einer einsamen Landstraße liegengeblieben ist, per Mobiltelefon eine telefonische Pannenhilfe vom System READY anfordern kann. Die optimale Unterstützung des Anrufers bei der Reparatur seines Fahrzeugs soll vor allem auch auf einer an seine Ressourcenbeschränkungen angepaßten Form der Informationspräsentation und der Dialogführung beruhen. Zwei spezielle Formen der Ressourcenbeschränkung wurden für READY zur genaueren Untersuchung ausgewählt: Beschränkungen des Arbeitsgedächtnisses des Benutzers und seiner verfügbaren Zeit.

Bei der Untersuchung von Möglichkeiten der Einschätzung von Ressourcen eines Benutzers stehen in READY dabei zwei Fragestellungen besonders im Vordergrund:

- Wie können in einem natürlichsprachlichen Dialog die Ressourcenbeschränkungen eines Dialogpartners erkannt werden?
- Wie kann das System sein Dialogverhalten an die erkannten Ressourcenbeschränkungen seines Partners anpassen?

Die vorliegende Arbeit entstand im Rahmen des Projektes READY und widmet sich vor allem der ersten Fragestellung. Eine der wichtigen Grundlagen dafür ist die Kenntnis und korrekte Interpretation der jeweiligen Symptome einer Ressourcenbeschränkung. Im gewählten Szenario handelt es sich wegen der telefonischen und damit rein akustischen Verbindung vor allem auch um Symptome der Sprachproduktion. Die vorliegende Arbeit möchte einen Beitrag zur Kenntnis und Interpretation dieser Symptome leisten, indem die psychologisch erforschten Parameter identifiziert, ihre Relevanz für die Beurteilung von Ressourcenbeschränkungen bewertet und Ansätze zu ihrer Verarbeitung aufgezeigt werden.

## 1.4 Überblick über die Arbeit

Die vorliegende Arbeit gliedert sich in drei Teile.

Der **erste Teil** beschäftigt sich mit den *theoretischen* Grundlagen für die Erkennung von kognitiven Verarbeitungsbeschränkungen und Zeitdruck. Das umfaßt eine Vorstellung und Diskussion der relevanten theoretischen Vorarbeiten und Modelle in *Kapitel 2*. Dabei wird sowohl auf allgemeine kognitive Grundlagen als auch auf den speziellen Prozeß der Sprachproduktion eingegangen.

Der **zweite Teil** widmet sich ausführlich den *empirischen* Grundlagen, da vergleichbare integrative Arbeiten für die gewählte Themenstellung bisher nicht existieren. Insbesondere Erkenntnisse zum Zusammenhang von Sprache, kognitiver Belastung und Zeitdruck lagen bei Beginn dieser Arbeit nicht in einem umfassenden Kompendium mit einheitlichem theoretischen Rahmen vor, sondern lediglich als Vielzahl experimenteller Studien, die jeweils individuell unterschiedliche theoretische Annahmen mithilfe verschiedener Methodik überprüften und ebenso unterschiedliche Formen der Auswertung ihrer Ergebnisse lieferten. Eine unmittelbare Übernahme empirischer Ergebnisse und Schlußfolgerungen für den Entwurf einer Benutzermodellierungskomponente war dadurch nicht möglich.

Im *Kapitel 3* werden daher die verfügbaren experimentellen Vorarbeiten zu Sprachproduktion unter Belastung und Zeitdruck einer ausführlichen Diskussion unterzogen. Um die durch unterschiedliche Ansätze der experimentellen Studien erreichten Ergebnisse vergleichbar zu machen, werden diese Resultate und die sich daraus ergebenden Schlußfolgerungen im Kontext eines einheitlichen Sprachproduktionsmodells integriert. Ziel ist es, experimentell und damit empirisch fundierte Symptome für Charakteristika der Sprachproduktion unter den beiden speziellen Ressourcenbeschränkungen *kognitive Belastung* und *Zeitdruck* zu identifizieren und ihren diagnostischen Wert für die Verwendung in einem Dialogsystem abzuschätzen.

*Kapitel 4* liefert die Ergebnisse einer Korpusanalyse, die einerseits die Arbeit an der Grundlagenforschung in Kapitel 3 begleitete, indem daraus hilfreiche Hinweise auf mögliche Symptome bezogen wurden. Andererseits unterstützt sie die Modellbildung des Kapitels 6 durch eine domänenabhängige Eingrenzung der Erkenntnisse aus Kapitel 3. Zu den Resultaten zählen empirische Verteilungen der relevanten Symptome in der Domäne des Systems READY und eine Annotationsmethode, die diese Symptome kennzeichnet und für eine automatische Verarbeitung auf dem derzeitigen Forschungsstand zur Verfügung stellt.

Im **dritten Teil** der Arbeit, der durch die *Kapitel 5* und *6* gebildet wird, wird eine aktuelle Anwendung der Erkenntnisse in einem ressourcenadaptiven Softwaresystem beschrieben.

In *Kapitel 5* werden zunächst die aktuellen Möglichkeiten zur automatischen Erkennung von Symptomen in Softwaresystemen diskutiert und die Eingabeanalyse im Prototypen des Projekts READY vorgestellt. Bestandteil dieses Kapitels ist auch die Beschreibung einer beispielhaften Zusammenarbeit mit dem Projekt LISA innerhalb des genannten SFB 378 im Bereich der Eingabeanalyse.

*Kapitel 6* stellt den sprachlichen Teil der Benutzermodellierungskomponente des Systems READY vor und erläutert die Schritte bei der Modellbildung für den zugehörigen Prototypen des Systems. Die empirischen Ergebnisse aus den Kapiteln 3 und 4 werden hierbei berücksichtigt. Der aktuelle Stand der Implementation wird vorgestellt, und es wird ein Ansatze für Anpassungen des aktuellen Modells an neuere Erkenntnisse skizziert.

*Schlußfolgerungen und Ausblick* (Kapitel 7) beschließen die Arbeit. Der *Anhang* enthält einige technische Details und Tabellen, die der Kohärenz der Darstellung halber nicht in den zugehörigen Kapiteln dargestellt werden.

# Kapitel 2

# Theoretische Grundlagen aus Psychologie und Psycholinguistik

In diesem Kapitel werden Begriffe und Mechanismen beschrieben, welche den theoretischen Hintergrund dieser Arbeit bilden. Da die empirische Grundlage der späteren Kapitel durch eine umfangreiche psychologische Forschung in unterschiedlichen situativen Kontexten gebildet wird, sollen die Aspekte kognitiver Verarbeitung im allgemeinen und der Sprachproduktion im besonderen näher untersucht werden, die den Begriff der *Verarbeitungsbeschränkung* bzw. *Ressourcenbeschränkung* charakterisieren helfen. Damit wird gleichzeitig klar gemacht, wie deren Auswirkungen umgekehrt zur Einschätzung der kognitiven Verarbeitungsleistung beitragen können. Ein spezifischerer Rahmen wird für die in dieser Arbeit näher betrachtete Anwendung auf den Prozeß der Sprachproduktion vorbereitet, indem ein psychologisches Modell der Sprachproduktion vorgestellt wird.

## 2.1 Ressourcen und Prozesse

Aus alltäglichen Beobachtungen des eigenen Verhaltens und des Verhaltens von Mitmenschen – nicht nur in Streßsituationen – läßt sich vermuten, daß das menschliche kognitive System nicht beliebig belastet werden kann. So können die meisten Menschen nur eine begrenzte Anzahl von Aufgaben gleichzeitig ausführen, ohne daß die Leistung in den einzelnen Aufgaben beeinträchtigt wird. Diese Beobachtung führte zu dem Schluß, daß für die Verarbeitung von Informationen und die Steuerung von Handlungen dem Menschen nur endlich viele kognitive 'Arbeitsmittel' zur Verfügung stehen bzw. daß die *Kapazität* dieser Arbeitsmittel begrenzt ist. Mechanismen und Objekte, die für die Realisierung von Aufgaben zum Einsatz kommen, können unter dem Begriff *Ressourcen* zusammengefaßt werden. Diese Grobklassifikation bedarf jedoch einer Verfeinerung, wofür im folgenden die Entwicklung des psychologischen Ressourcenbegriffs kurz umrissen und seine Anwendbarkeit auf den Begriff der *kognitiven Belastung* untersucht werden soll.

### 2.1.1 Theorien kognitiver Ressourcen

**Single Resource Theory** Die beobachtbaren Leistungseinbußen bei der Handlungsausführung sind abhängig von der Anzahl der parallel auszuführenden Aufgaben: Je mehr Aufgaben ein Mensch parallel ausführt, desto schlechter wird gewöhnlich seine Leistung in den Einzelhandlungen. Da diese Beobachtung allgemeingültig zu sein schien, wurde bei der Erforschung des menschlichen kognitiven Systems die Existenz einer einzelnen zentralen kognitiven Ressource angenommen (Kahneman, 1973): Die Qualität bei der Ausführung von Handlungen ist von der Qualität unterschiedlicher kognitiver Mechanismen abhängig, wie z.B. Reizaufnahme, Informationsverarbeitung, motorische Steuerung. Diese kognitiven Prozesse konkurrieren nach der sogenannten *Single Resource Theory* um diese eine Ressource und können nur dann ihre volle Leistungsfähigkeit erreichen, wenn sie in ausreichender Menge zur Verfügung steht. Die Einheitsressource hat verschiedene Bezeichnungen erhalten; im Englischen wird jedoch meist der Begriff *attentional resource* (Aufmerksamkeit, Konzentration) verwendet.

**Anpassungen der Single Resource Theory** In Experimenten, in denen Probanden mehrere Aufgaben parallel auszuführen hatten, wurden jedoch bei verschiedenen Kombinationen von Versuchsaufgaben unterschiedliche Interaktionseffekte gemessen: Einige Aufgaben konnten sogar unabhängig von der Art der Nebenaufgabe unter allen Bedingungen unbeeinträchtigt ausgeführt werden. Daraus ergab sich eine Unterteilung kognitiver Prozesse in *aufmerksamkeitsintensive* und *automatische* oder *automatisierte* Prozesse. Im Gegensatz zu aufmerksamkeitsintensiven Prozessen beanspruchen automatisierte Prozesse die kognitive Ressource *Aufmerksamkeit* nicht und sind damit von deren verbleibender Kapazität vollkommen unabhängig. Aus diesem Grund sind sie parallel zu jeder Art von Aufgabe ausführbar.

**Multiple Resource Theory** Auch die Unterscheidung von automatisierten und aufmerksamkeitsintensiven Prozessen kann verschiedene Experimentalergebnisse der kognitiven Psychologie noch nicht ausreichend erklären: Die Leistung einzelner, als aufmerksamkeitsintensiv angenommener Aufgaben wurde nur von ganz spezifischen Zusatzaufgaben beeinflußt. Andere Zusatzaufgaben, deren Ausführungsqualität sich experimentell als ebenfalls abhängig von der zentralen Ressource Aufmerksamkeit erwiesen hatte, verursachten nur eine geringe oder gar keine Interaktion. Die Annahme, es existiere *ein einzelner* Pool an Aufmerksamkeit im kognitiven System des Menschen war daher zu revidieren: An ihre Stelle trat die (in verschiedensten Varianten diskutierte) Auffassung, daß Menschen spezialisierte kognitive Fähigkeiten besitzen, die für ganz bestimmte Aufgabentypen eingesetzt werden können, und daß diese Fähigkeiten im kognitiven System unterschiedlich miteinander interferieren (siehe z.B. Neumann, 1992; Wickens, 1992). Diese Sichtweise wird als *Multiple Resource Theory* bezeichnet. Zu ihren Charakteristika gehört auch, daß die binäre Sicht auf den unterschiedlichen Automatisierungsgrad verschiedener kognitiver Prozesse (*automatisiert* vs. *nicht automatisiert*) durch die flexiblere Idee eines Kontinuums zwischen zwei Polen abgelöst wurde (Neumann, 1992).

**Probleme der Multiple Resource Theory** Weitgehend anerkannte psychologische Modelle der beschriebenen Idee multipler Ressourcen sind zum heutigen Zeitpunkt kaum verfügbar, da die Theorie erst seit vergleichsweise kurzer Zeit zu Prominenz in der Forschungslandschaft gekommen ist und deshalb erst wenige experimentelle Studien Hypothesen unter dieser Sichtweise untersucht haben. Verantwortlich dafür ist sicherlich auch die Komplexität der Interaktion zwischen verschiedenen möglichen Einzelressourcen, die eine sehr feinkörnige Unterteilung der kognitiven Ressourcen notwendig zu machen scheint. Ansätze dafür liefert z.B. McLeod (1977).

### 2.1.2 Eigenschaften von Ressourcen

Jameson und Buchholz (1998) arbeiten charakteristische Eigenschaften eines allgemeinen Ressourcenbegriffs heraus[1]. Einige dieser Charakteristika sind in bezug auf die Modellierung kognitiver Ressourcen eines Dialogpartners besonders wichtig:

- Sofern die Gesamtkapazität einer Ressource auf verschiedene Prozesse aufgeteilt werden kann, kann einzelnen Prozessen ein unterschiedlicher Anteil an der Ressource zugeteilt werden. Im Zusammenhang mit der Zuteilung der Ressource steht meist die Qualität bei der Ausführung der zugehörigen Aufgabe. Diese Qualität läßt sich mit einer *Erfolgsmetrik* beschreiben: In einem *Erfolgsprofil* oder *Performanzprofil* (Wickens, 1992, siehe auch Abb. 2.1) kann der Zusammenhang zwischen Intensität des Ressourceneinsatzes und Qualität des Ergebnisses dargestellt werden.

- Der Ressourceneinsatz ist mit Kosten verbunden. Z.B. können zeitliche Kosten entstehen, d.h. eine Handlung dauert möglicherweise länger, wenn dafür eine Ressource weniger intensiv genutzt werden kann. Kosten können gleichfalls als Ausnutzung einer unbegrenzten Ressource gesehen und daher ebenfalls durch eine Metrik angegeben werden.

- In bestimmten Fällen können Prozesse, die auf eine bestimmte Ressource angewiesen sind, ihre Verarbeitung an die Verfügbarkeit dieser Ressource anpassen. Verschiedene Formen solchen *ressourcenadaptiven* Verhaltens werden von Jameson und Buchholz (1998) detaillierter betrachtet. Entscheidungen über die Ressourcenverteilung an verschiedene konkurrente Prozesse können unter Verwendung von Optimalitätskriterien aus Kosten-Nutzen-Verhältnissen abgeleitet werden, die sich z.B. aus Erfolgsprofilen ergeben.[2]

### 2.1.3 Effekte von Ressourcenbeschränkungen auf kognitive Prozesse

**Qualität** Kapazitiv begrenzte Systeme lassen sich durch eine maximale Kapazität und einen aktuellen Auslastungsgrad dieser Kapazität charakterisieren. Für das kognitive System des Menschen kann durch die Anwendung des Ressourcenbegriffes auf seine unterschiedlichen Teilkomponenten deren Auslastungsgrad und damit die kognitive Belastung charakterisiert werden.

[1] Der in Jameson und Buchholz (1998) beschriebene Ressourcenbegriff dient als Leitlinie im Sonderforschungsbereich 378 der DFG.

[2] Carletta, Caley und Isard (1995) verwenden für die Simulation einer Teilphase der Sprachproduktion (siehe Abschnitt 2.4.2) z.B. *Zeit* als Kosten- und *Informativität* als Nutzen-Funktion. Vergleichbare Techniken werden im System READY im Dialogmanagement eingesetzt (siehe z.B. Weis, 1999).

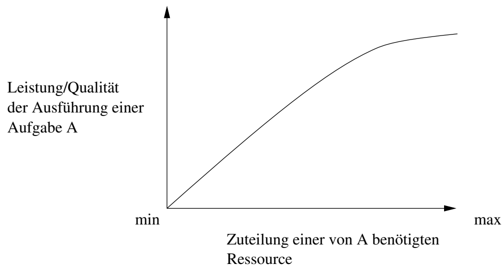


Abbildung 2.1: Performanzprofil nach Wickens (1992)

Konkurrieren mehrere verschiedene Prozesse zu ein und demselben Zeitpunkt um die gleiche Ressource, wird ein Zuteilungsmechanismus für diese Ressource benötigt. Jameson (1997) präzisiert diese Sicht um Kategoriensysteme für unterschiedliche Arten von Ressourcen und zugehörige Allokationsmodelle. Ein Allokationsmodell für die Ressource *Arbeitsgedächtnis* wird auch im Modell READY eingesetzt und später vorgestellt (Kapitel 6).

Ein solcher Zuteilungsmechanismus bleibt konsequenterweise nicht ohne Folgen für die Leistungsfähigkeit der Einzelprozesse: Wie bereits oben gesagt, lassen sich für metrisierbare Ressourcen Erfolgsprofile erstellen, die den Zusammenhang von Ressourcenausnutzung und Erfolg einer auszuführenden Aufgabe veranschaulichen. Ist die Kennlinie eines Erfolgsprofils monoton steigend, kann aus dem Erfolgsprofil entnommen werden, welchen Beeinträchtigungen der zugehörige Prozeß durch mangelnde Ressourcenzuteilung unterliegen kann. Wenn unterschiedliche Prozesse gleichzeitig auf ein und dieselbe Ressource zugreifen wollen, erhöht sich ihr Auslastungsgrad entsprechend den Anforderungen dieser Prozesse nur solange, bis die maximale Kapazität erreicht ist. Sobald das der Fall ist, bestehen nur noch zwei Möglichkeiten:

- Es kann nicht mehr allen Prozessen die von ihnen geforderte Menge der Ressource zugeteilt werden oder
- Einzelne Prozesse müssen vom Zugriff auf die Ressource ausgeschlossen werden, um für andere die Allokationsanforderungen weiter befriedigen zu können.

In Abbildung 2.2 sind links die Performanzprofile zweier von der gleichen Ressource abhängiger Prozesse gegenübergestellt. P(X) gibt ein Maß für die Ausführungsqualität der Aufgabe X an. Die Leistung in einer der beiden angegebenen Aufgaben nimmt jeweils mit der Zuteilung der Ressource an den zugehörigen Prozeß zu, bei Zuteilung an den konkurrenten Prozeß nimmt sie ab. Die Stärke des Effektes ist vom Zuteilungsverhältnis abhängig. Der rechte Teil der Abbildung 2.2 zeigt eine andere Form der Darstellung für den Effekt des Zuteilungsverhältnisses: Am Schnittpunkt der beiden gestrichelten Linien würden beide Prozesse ihre volle Leistung erreichen, was jedoch aufgrund der Interaktion und der begrenzten Kapazität nicht möglich ist.[3]

[3]Die Grafik wurde Rummer (1996a) entnommen.

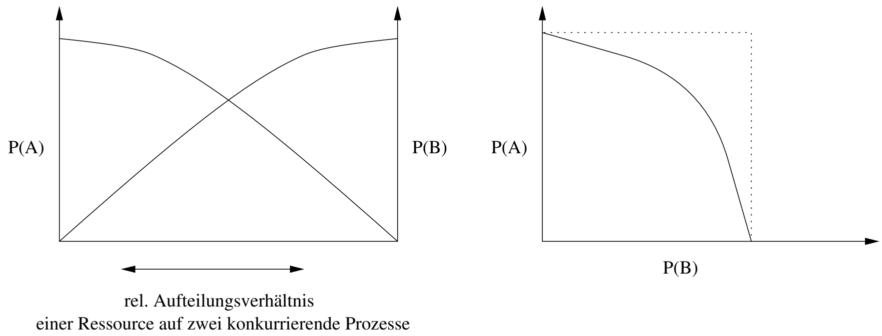


Abbildung 2.2: Interaktion der Performanzprofile zweier Aufgaben, die vom Zuteilungsverhältnis der gleichen Ressource abhängig sind

**Geschwindigkeit** Zwei konkurrierende Prozesse können durch ein geeignetes zeitliches Verhalten erreichen, daß eine von beiden benötigte, beschränkte Ressource jedem einzelnen entsprechend seinen Anforderungen zur Verfügung steht. Dafür müssen sie zeitversetzt auf die Ressource zuzugreifen und sie nach der Nutzung wieder freigeben. Dieses Vorgehen führt jedoch konsequenterweise zu einer Geschwindigkeitsreduktion beim Erreichen des gleichen quantitativen Ergebnisses wie im Fall ohne Ressourcenbeschränkung, da die Prozesse nicht mehr parallel ausgeführt werden. Dieses Vorgehen entspricht einem *Time-Sharing* auf der Ressource. Sofern ein Prozeß jedoch eine qualitativ inkrementelle Verbesserung seines Ergebnisses erlaubt, resultiert diese Form der Adaptivität in einer schlechteren Qualität des Ergebnisses innerhalb der gleichen Zeitspanne gegenüber voller Ressourcenzuteilung. Auch im Falle zeitlicher Anpassungen an den Ressourcenbedarf ergeben sich daher durch beschränkte Kapazität beobachtbare Effekte.

### 2.1.4 Anpassung von Agenten an eigene Ressourcenbeschränkungen

Ressourcenadaptive Prozesse spielen eine wichtige Rolle im kognitiven System des Menschen: Kann ein Prozeß sich nicht an eine wechselnde Ressourcenlage anpassen, so impliziert das, daß er nur mit voller Zuteilung der benötigten Ressourcen überhaupt ein Ergebnis liefern kann und in allen anderen Fällen ergebnislos bleibt. Die kognitiven Prozesse des Menschen müssen schon deshalb ressourcenadaptiv sein, damit sie bei wechselnden Umweltbedingungen mit unterschiedlich komplexen Reaktions- und Handlungsanforderungen weitgehend parallel ausgeführt werden können und in den meisten Fällen zu ausreichend guten Resultaten kommen. Drastisch formuliert ist Ressourcenadaptivität eine Voraussetzung für die Überlebensfähigkeit.

Dieser Aspekt kognitiver Prozesse kann in einem Dialogsystem ausgenutzt werden, das Aufschluß über die Belastung seines Benutzers erlangen möchte. Da ressourcenadaptive Prozesse (bis auf Extremfälle) immer eine Form der Ausgabe liefern, kann anhand der Qualität und Geschwindigkeit dieser Ausgabe die Leistungsfähigkeit des zugehörigen kognitiven Prozesses beurteilt und damit auf die Auslastung der benötigten Ressourcen zurückgeschlossen werden.

Die nachfolgenden Betrachtungen einzelner Ressourcen fokussieren auf diejenigen, die im Projekt READY im Mittelpunkt des Interesses stehen.

## 2.2 Arbeitsgedächtnisbelastung und Zeitdruck als Ressourcenbeschränkungen

### 2.2.1 Arbeitsgedächtnisbelastung

Zu den Differenzierungen der *Multiple Resource Theory* gehört die Einsicht, daß das menschliche Gedächtnis in eine Langzeitkomponente zur dauerhaften Speicherung von Information und eine für die Informationsverarbeitung in Situationen menschlichen Handelns eingesetzte Kurzzeitkomponente unterteilt ist. Für die Kurzzeitkomponente wird wegen ihrer Veränderlichkeit innerhalb kurzer Zeitspannen, des situationsgesteuerten Einsatzes und ihrer zentralen Bedeutung der Begriff *Arbeitsgedächtnis* (*Working Memory*, siehe Baddeley, 1986) verwendet.

Zur zentralen Stellung des Arbeitsgedächtnisses in der menschlichen Informationsverarbeitung trägt u.a. seine Brückenfunktion zwischen aus der Umwelt aufgenommener, im Langzeitgedächtnis gespeicherter und an die Umwelt abzugebender Information sowie die Aufgabe der Steuerung komplexer Handlungsabläufe bei. Eine Speicherung von Informationen ist im Arbeitsgedächtnis nur vorübergehend möglich.

In der psychologischen Forschung wurden mehrere Modelle des Arbeitsgedächtnisses entworfen (z.B. Baddeley, 1986; Shallice, 1982)[4]. Obwohl die unterschiedlichen Modelle im Detail voneinander abweichen, sind ihnen wesentliche Aspekte gemeinsam. Danach läßt sich das Arbeitsgedächtnis unterteilen in

- eine akustisch-phonologische Teilkomponente für sprachlich repräsentierte Information,
- eine Teilkomponente für die Verarbeitung visuell-räumlicher Information und
- eine Komponente der zentralen Verarbeitung, die diese Teilsysteme ansteuert.

Diese Differenzierung steht in engem Zusammenhang mit der Ressourcendiskussion. Die Trennung in den visuell-räumlichen und den akustisch-phonologischen Teil geht auf eine ähnliche Klasse von Experimenten zurück, wie sie auch der Revision der *Single Resource Theory* zugrunde lag: Die Variation verschiedener Primär-und Zusatzaufgaben in Experimenten zum Arbeitsgedächtnis ergab je nach Kombination der zugehörigen Eingabe- (visuell/akustisch) und Ausgabe-Modalitäten (motorisch/sprachlich[5]) unterschiedliche Beeinträchtigungen der Handlungsresultate.

Weiterhin wird eine Trennung in einen prozeduralen, verarbeitenden Teil (*Kontrollinstanz*) und ihre deklarative, speichernde Subkomponenten (*Puffer oder Speicher*) angenommen. Daran läßt sich die Unterteilung in automatisierte und kontrollierte Prozesse im Zusammenhang mit

[4]Ein spezielles Modell für die Sprachproduktion stellen Rummer, Grabowski und Vorwerg (1995) vor.

[5]Natürlich erfordert auch Sprachproduktion motorische Steuerung bei der Artikulation. Die Unterscheidung bezieht sich auf die auszuführende Handlung.

der Verwendung des Arbeitsgedächtnisses klarmachen: Während automatisierte Prozesse z.B. durch den Rückgriff auf abgespeicherte Schemata von der zentralen Kontrollinstanz weitgehend unabhängig sind, laufen kontrollierte Prozesse vor allem im zentralen, prozeduralen Teil des Arbeitsgedächtnisses ab. Sie werden direkt und bewußt gesteuert und können daher nicht beliebig parallel ausgeführt werden.

### Das Arbeitsgedächtnismodell von Baddeley (1986)

Im Modell von Baddeley (1986), das im Projekt READY als Referenzmodell dient, kontrolliert die **Zentrale Exekutive** den Informationsfluß, sorgt für den Abruf von Informationen aus dem Langzeitgedächtnis und verarbeitet diese Informationen aktiv. Sie steuert die beiden Subsysteme **Visuell-räumliches Sketchpad** und **Phonologische Schleife** an, die für die temporäre Speicherung visueller bzw. akustisch-phonologischer Information dienen. Sowohl die Verarbeitungskapazität der Zentralen Exekutive als auch die Speicherkapazität der Subsysteme ist begrenzt. Diese Begrenzungen sind allerdings nicht in absoluten Werten meßbar, wie noch Miller (1956) vermutete (siehe z.B. Baddeley, 1994; Shiffrin & Nosofsky, 1994). Sie hängen vielmehr stark von der Enkodierung und Organisation der zu speichernden und zu verarbeitenden Informationen ab.[6]

Die Phonologische Schleife im Baddeley-Modell besteht aus einem phonologischen Kurzzeitspeicher und einem subvokalen Auffrischungsmechanismus für die darin abgelegte Information. Die Auffrischung ist notwendig, da die Information im Speicher einem Vergessensprozeß von wenigen Sekunden unterliegt. Sprachliche Eingaben werden vom Kurzzeitspeicher direkt aufgenommen. Der Auffrischungsprozeß übernimmt ggf. zusätzlich die Umkodierung nichtsprachlicher Information, wenn diese in der Phonologischen Schleife gespeichert werden soll, um ihren 'Datentyp' an die Speicherkomponente anzupassen.

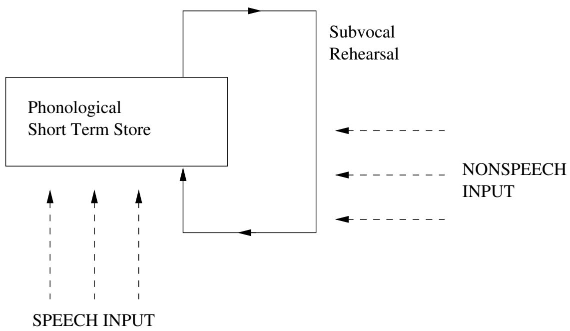


Abbildung 2.3: Funktionalität der Phonologischen Schleife nach Gathercole und Baddeley (1993)

Neue sprachliche oder nichtsprachliche Eingaben erfordern ein selektives Ersetzen der Information im Kurzzeitspeicher. Die möglicherweise daraus entstehenden Konflikte zwischen noch zu verarbeitender und neu zu speichernder Information und die Möglichkeiten einer Um-

[6]Diese Aspekte der Informationsverarbeitung werden oft als *Chunking* bezeichnet. Einen Überblick über die Idee der *informational chunks* und ihre Auswirkungen auf die Verarbeitungsleistung gibt Wickens (1992, Kap. 6).

kodierung von Informationen in Belastungssituationen (Gathercole & Baddeley, 1993) sind sowohl für die Sprachproduktion als auch im Bereich des Sprachverstehens relevant.

Da die Funktionalität und Bedeutung der Phonologischen Schleife in der psychologischen Forschung derzeit noch diskutiert wird (siehe z.B. Baddeley, Gathercole & Constanza, 1998), ist eine Differenzierung des Arbeitsgedächtnisses in Teilkomponenten und eine Zuordnung von Subprozessen der Informationsverarbeitung für Zwecke der Modellierung kognitiver Prozesse noch mit einem zu hohen Maß an Unsicherheit behaftet. Im folgenden wird diese Aufspaltung deshalb nur für klar erwiesene Zusammenhänge explizit herangezogen und das Hauptaugenmerk auf die Kapazität der Zentralen Exekutive des Arbeitsgedächtnisses gelegt. Das gilt auch für den Zusammenhang von Sprachproduktion und Arbeitsgedächtnis.

### 2.2.2 Zeitdruck

Zeitdruck wird in der Literatur vor allem als ein Phänomen der Wahrnehmung beschrieben. Bei Marx (1984) bezieht sich dies auf das Empfinden einer Diskrepanz zwischen zur Verfügung stehender und benötigter Zeit zur Ausführung einer Aufgabe. Dieser Begriff ist jedoch sehr allgemein gehalten und erfordert aufgrund unterschiedlicher Verwendung des Begriffes in der experimentellen Literatur (Kapitel 3) eine genauere Differenzierung.

Menschlichen Handlungen zugrundeliegende Ziele können meist in Subziele aufgespalten werden, denen wiederum Teilpläne für die Handlungsausführung und damit charakteristische Abschnitte der übergeordneten Handlung zugeordnet werden können. Auf jedem beliebigen dieser Handlungsniveaus kann die Zeit für die Ausführung der Handlung begrenzt sein. Speziell bei der Sprachproduktion im Sinne einer Handlung sind daher verschiedene Niveaus von Zeitdruck zu unterscheiden.

**Mikro-Zeitdruck:** Einerseits existiert je nach Kommunikationssituation für den Sprecher ein Druck, eine Äußerung innerhalb einer gewissen Zeitspanne zu *beginnen*, weil dies den Konventionen der Kommunikation entspricht oder durch eine spezielle Situation besonders gefordert sein kann. Dazu zählt z.B. Zeitdruck in einer Prüfungssituation, wo (starkes) Zögern bei der Beantwortung einer Frage vom Prüfer als Zeichen mangelnder Kenntnis interpretiert werden kann. Im Kontext von Telefongesprächen wird die Präsenz in der Kommunikationssituation für den Hörer sichergestellt, da dieser meist über keine anderen Indikatoren für die Aufrechterhaltung der Verbindung verfügt. Andererseits existiert für einen Sprecher gemäß diesen Konventionen auch der Druck, eine einmal begonnene Äußerung oder einen Satz innerhalb einer gewissen Zeitspanne *fertigzustellen*. Diese beiden lediglich lokalen Zeitbegrenzungen sollen im folgenden als *Mikro-Zeitdruck* bezeichnet werden.

**Makro-Zeitduck** Zeitdruck entsteht auch durch die Limitierung der Zeit für die Ausführung eines übergeordneten Zieles, d.h. für die Gesamtdauer einer komplexeren Folge von Handlungen[7]. Beispiele dafür sind die Begrenzung von Redezeit in öffentlichen Diskussionen, womit die Zeit begrenzt wird, innerhalb derer der eigene Standpunkt klargemacht und das Publikum überzeugt werden kann. Auch ein bevorstehender Termin kann die verfügbare

[7] Im Kontext des Dialogsystems READY ist ein komplexes Ziel des Anrufers die Reparatur seines Fahrzeugs.

Zeit für die Umsetzung eines komplexen Zieles beeinflussen. Solche Einschränkungen der Dauer für die Ausführung komplexer Handlungen induzieren eine Form des globalen Zeitdrucks und sollen daher als *Makro-Zeitdruck* bezeichnet werden.

Ein intelligentes Softwaresystem kann sich an die verfügbare Gesamtlänge von Planungs- und Handlungszeiträumen anpassen, indem es versucht, den Zeitdruck des Benutzers zu erkennen. In diesem Zusammenhang ist jedoch lediglich der Makro-Zeitdruck interessant. Die Untersuchung des Zeitdrucks auf Mikroniveau kann in einigen Fällen zu Erkenntnissen über den Makro-Zeitdruck beitragen. Meist ist Mikro-Zeitdruck aber bereits durch die Dialogsituation und damit unabhängig vom Makro-Zeitdruck gegeben. Er sollte keinen Einfluß auf die Systemeinschätzungen über den globalen Zeitdruck des Benutzers haben.

### 2.2.3 Interaktion von Belastung und Zeitdruck

Baber und Mellor (1996) sehen Belastung als Zeitdruck und umgekehrt: Danach ist Belastung mit der Gesamtmenge an 'Aktivität' gleichzusetzen, die pro Zeiteinheit ausgeführt werden muß. Ähnliche Gedanken finden sich bei Rummer et al. (1995) und Wegner und Erber (1992).

Zeitdruck ist in bezug auf die Informationsverarbeitung daher zumindest *vergleichbar* mit kognitiver Belastung: Sowohl als Effekt von Belastung als auch von Zeitdruck werden z.B. weniger Informationen abgerufen, als für eine Handlungsausführung notwendig wären (Niederberger, Engemann & Radtke, 1987; Marx, 1984). Wenn sich durch Verzögerungen der Ressourcenzuteilung aufgrund kognitiver Belastung Wartezeiten für einzelne Prozesse ergeben, ihre Gesamtdauer jedoch an ein zeitliches Limit gebunden ist, verringert sich dieses Limit weiter und erzeugt dadurch auch einen höheren Zeitdruck.[8]

Andererseits kann der Gedanke an ein Zeitlimit auch eine höhere Belastung nach sich ziehen, da alle Teilprozesse unter der Vorgabe des Zeitlimits ablaufen müssen. Dadurch werden ggf. mehr kognitive Ressourcen je Zeiteinheit gefordert.

Die Identifikation der genauen Zusammenhänge zwischen Zeitdruck und kognitiver Belastung muß der psychologischen Forschung überlassen werden. Wichtig für die Interpretation experimenteller Ergebnisse und die Modellbildung ist allein die Einsicht, daß die beiden Aspekte nicht unabhängig voneinander sind.

## 2.3 Möglichkeiten der Erkennung von Verarbeitungsbeschränkungen eines Dialogpartners

### 2.3.1 Umkehrung eines psychologischen Experimentalparadigmas

Eine bestimmte Klasse psychologischer Experimentalanordnungen nutzt den ressourcenadaptiven Aspekt menschlichen Verhaltens aus: Diese Experimente basieren auf dem sogenannten

[8] Da Sprachproduktion als Handlung zu verstehen ist, kann diese Einsicht auf die Sprachproduktion übertragen werden.

*Doppelaufgabenparadigma*. Das Verhalten von Probanden bei der parallelen Ausführung zweier Aufgaben wird in diesem Paradigma für Rückschlüsse darüber herangezogen, ob zwei Aufgaben auf gleiche kognitive Ressourcen zurückgreifen bzw. ob sie überhaupt ressourcenintensiv sind. Diese Eigenschaft wird danach beurteilt, ob es bei der parallelen Handlungsausführung zu Interferenzen kommt. Interferenzen sind an Leistungseinbußen in den Einzelaufgaben gegenüber geeigneten Kontrollbedingungen ohne Doppelaufgabe erkennbar. Voraussetzung für die Korrektheit der Beobachtungen ist allerdings, daß der maximale Auslastungsgrad der entsprechenden Ressourcen im Experiment erreicht wird.

Anhand der Performanz von Personen bei der Ausführung zweier Teilaufgaben, für die die Abhängigkeit von der gleichen Ressource experimentell abgesichert ist, kann auf die Auslastung dieser Ressource zurückgeschlossen werden. Diese Umkehrung des Doppelaufgabenparadigmas ist im Kontext eines Dialogsystems einsetzbar, das Inferenzen über kognitive Ressourcen seines Benutzers ableiten möchte.

### 2.3.2 Erkennung von Beschränkungen des Arbeitsgedächtnisses

Informationsverarbeitungsprozesse, insbesondere Planungsprozesse, die der Ausführung von Aufgaben vorangehen, müssen auf die Ressource *Arbeitsgedächtnis* zurückgreifen (Baddeley, 1986). Die Zugriffsmöglichkeiten werden sowohl durch dessen kapazitiv begrenzte (deklarative) Speicherfunktion als auch seine limitierten prozeduralen Fähigkeiten eingeschränkt. Ist ein Prozeß von der Ressource *Arbeitsgedächtnis* abhängig, erlauben Indikatoren beeinträchtigter Verarbeitung einen Rückschluß auf deren Auslastung. Für automatisierte Prozesse (die per Definition von der Kontrollinstanz des Arbeitsgedächtnisses unabhängig sind) ist dagegen kein Leistungsunterschied aufgrund von Engpässen der Ressource Arbeitsgedächtnis zu erwarten. Sie können daher nicht direkt zu einer Einschätzung der verfügbaren Arbeitsgedächtniskapazität beitragen. Da ihre Verarbeitung jedoch ggf. von den Ausgabedaten kontrollierter Prozesse abhängt, können sie als indirekte Indikatoren für diese Einschätzung herangezogen werden.

Durch die bisherige Diskussion von Arbeitsgedächtnismodellen und des Standes der Forschung ist klar, daß genaue Metriken für dieses spezielle begrenzte kognitive Teilsystem nicht angegeben werden können. Derartige Einschätzungen sind durch verschiedene Möglichkeiten der Informationsverarbeitung (z.B. durch Chunking[9]) situationsabhängig. Das Arbeitsgedächtnis stellt jedoch kein für alle denkbaren Situationen getestetes oder testbares System dar, da seine detaillierte Struktur noch weitgehend unbekannt ist. Die Angabe einer maximalen Kapazität wird dadurch unmöglich.

In vielen Bereichen sozialer Interaktion können Menschen jedoch relativ gut die kognitive Belastung eines Dialogpartners beurteilen und ihr eigenes Verhalten in kooperativer Weise entsprechend anpassen, ohne die speziellen Aspekte des Arbeitsgedächtnisses zu kennen. Eine heuristische, d.h. auf Erfahrung basierende Einschätzung der weiteren Belastungsfähigkeit ersetzt die Einschätzung einer Gesamtkapazität. Niemand wird beispielsweise von einem Autofahrer sagen können, welche Aufgaben in seinem Arbeitsgedächtnis im Moment gerade ausgeführt werden bzw. zu wieviel Prozent seine Kapazität ausgelastet ist. Durch Beobachtung

[9] Wickens (1992) beschreibt u.a. auch, wie der Umfang eines 'Chunks', der als Speichereinheit angenommen wird, von Übung, der Rekodierbarkeit von Information und verschiedenen Möglichkeiten der Rekodierung beeinflußt wird.

seines Verhaltens wird jedoch meist klar, ob seine aktuelle Konzentration auf den Verkehr parallele Aufgaben wie die Diskussion komplexer Gesprächsthemen erlauben könnte bzw. ob dabei qualtitativ hochwertige Antworten zu erwarten sind.

Interaktive Systeme sollten Inferenzen über die Anpassung ihres Verhaltens und die dafür relevanten Begründungen für den Anwender transparent machen, um den Anwender nicht erst zum Erlernen neuer Interaktionstechniken zu zwingen. Ein Dialogsystem sollte sich daher die Vorgehensweise von menschlichen Agenten in ähnlichen Situationen zunutze machen. Das bedeutet, daß die Einschätzung kognitiver Belastung vor allem feststellen helfen soll, inwiefern ein Mensch mit weiteren Aufgaben belastet werden kann und ob er zu ihrer reibungslosen oder qualititativ ausreichend guten Ausführung in der Lage sein wird.

Für die Beurteilung der Verarbeitungsfähigkeit eines Dialogpartners werden von Menschen nicht nur die *Folgen*, d.h. die *Symptome* einer möglichen Belastungssituation herangezogen, sondern aus eigenen Erfahrungswerten heraus auch Merkmale der Situation für die Vorhersage möglicher Auswirkungen berücksichtigt. Diese situativen Merkmale sind als *Ursachen* von Belastung zu sehen. Beide Aspekte werden in READY in den Modellierungsprozeß einbezogen[10] – im Rahmen dieser Arbeit stehen jedoch ausschließlich die Symptome im Vordergrund.

### 2.3.3 Einschränkung: Einflüsse von Übung

Verschiedene experimentelle Arbeiten weisen darauf hin, daß Übung die Effekte kognitiver Belastung neutralisieren kann. In einigen Untersuchungen entsprachen die Werte belasteter geübter Gruppen viel stärker den Werten aus Kontrollbedingungen, als die Werte belasteter, ungeübter Gruppen (Roßnagel, 1995a). Diskutiert wird dieses Phänomen u.a. auch von Greene (1984), Butterworth (1980a), Roßnagel (1995b).

Aus diesen Beobachtungen ergab sich die Schlußfolgerung, daß durch Übung Prozesse soweit automatisiert werden können, daß sie weniger Kapazität der relevanten Ressourcen für ihre Ausführung benötigen oder aber sogar ganz von diesen unabhängig werden können. Bourdin und Fayol (1994) geben ein eindrucksvolles Beispiel für die Sprachproduktion anhand eines Vergleiches von Schreiben und Sprechen bei Erwachsenen und Kindern: Je mehr Teilaspekte einer Handlung automatisiert werden, wie z.B. die motorischen Aktivitäten beim Schreiben, desto weniger Belastung verursacht die Handlung und desto mehr Kapazität steht für parallel auszuführende Handlungen zur Verfügung.

Dieser Effekt stellt zwar ein Problem für experimentelle Untersuchungen dar, in denen sich aufgrund der Geübtheit von Probanden weniger Leistungseinbußen unter Doppelaufgaben einstellen und damit möglicherweise falsche Schlußfolgerungen über die Interaktion verschiedener Aufgabentypen nach sich ziehen. Für die Einschätzung der kognitiven Belastung anhand von Symptomen in einem Dialogsystem ist er jedoch irrelevant: Ein Dialogsystem will sich an die *tatsächliche* und nicht an eine hypothetische, durch Doppelaufgabenexperimente vorhergesagte Belastung anpassen. Ist ein Dialogteilnehmer in der Ausführung einer Aufgabe geübt und zeigt daher weniger Interaktion dieser Doppelaufgabe mit der Sprachproduktion, steht er auch weniger unter Belastung und erfordert damit eine geringere Anpassungsleistung des Systems.

[10]Jameson (1997) beschreibt die Beziehungen zwischen diesen beiden Typen beobachtbarer Variablen näher.

## 2.4 Sprachproduktion

### 2.4.1 Einschränkung des Gegenstandsbereiches: Spontansprache

Sprachliche Äußerungen können in zwei verschiedenen Modalitäten auftreten. *Mündliche* und *schriftliche* Sprachproduktion unterscheiden sich vor allem bezüglich ihrer Anforderungen an kognitive und nicht-kognitive Ressourcen: Schriftliche Sprachproduktion erfolgt meist zeitlich getrennt von der Rezeption und ohne zeitliche Einschränkungen auf Mikroniveau, da sie von den Erwartungen des Lesers unabhängig ist. Bei mündlicher Sprachproduktion dagegen findet die Produktion der Äußerungen sowie deren Rezeption quasi zur gleichen Zeit statt. Verschiedene Kommunikationskonventionen verlangen dabei, daß ein Sprecher seinen Beitrag innerhalb einer limitierten Zeitspanne generiert. Daher ist ein Sprecher vor allem auch darauf bedacht, seine Rede flüssig zu halten und damit bei der Generierung seiner Äußerungen die zeitlichen Beschränkungen auf Mikroniveau zu beachten.

Die mündliche Sprachproduktion muß jedoch noch weiter unterschieden werden in eine Vermündlichung schriftlicher Inhalte, bei der im Moment der Äußerung keine Planung der Inhalte und Strukturen erfolgen muß[11], und die *Spontansprache*, deren Produkte unmittelbar vor der Artikulation geplant und formuliert werden. Mögliche Vorstrukturierungen, wie etwa in Situationen vorbereiteter Vorträge, tragen dazu bei, daß es sich hierbei eher um ein Kontinuum denn um eine Polarisierung handelt.

Schriftliche und nicht spontan geplante Sprachproduktion sind für ein Dialogsystem, für das eine mündliche Eingabemodalität vorgesehen ist, nicht weiter von Interesse. Der Rest der Arbeit konzentriert sich daher auf den Bereich des spontanen Sprechens. Im folgenden soll kurz motiviert werden, warum die Verwendung eines Sprachproduktionsmodells in diesem Zusammenhang wünschenswert und sinnvoll ist. Ein konkretes Referenzmodell wird vorgestellt und seine Berührungspunkte mit der Aufgabenstellung dieser Arbeit untersucht.

### 2.4.2 Psychologische Modelle der Sprachproduktion

**Motivation** Wie weiter oben bereits dargestellt wurde, ist das Ziel dieser Arbeit innerhalb des Projektes READY nicht die detaillierte Modellierung der menschlichen Produktion spontaner Sprache, sondern u.a. lediglich die Untersuchung von Veränderungen der Sprachproduktion unter kognitiver Belastung und Zeitdruck. Ein Sprachproduktionsmodell kann dabei die Erklärbarkeit bestimmter Merkmale von Äußerungen sichern, wie sie unter der kapazitiven Begrenzung dafür nötiger Ressourcen auftreten. Das Modell bietet weiterhin Unterstützung, wenn die Bedeutung einzelner Charakteristika von Äußerungen für die Benutzermodellierung einzuschätzen ist. Damit wird sichergestellt, daß nur tatsächlich kausal zusammenhängende Faktoren und Charakteristika der Sprachproduktion für Schlußfolgerungen über die kognitiven Eigenschaften eines Sprechers einbezogen werden.

Auch hier wird das Vorgehen psychologischer Modellbildung umgekehrt: Experimentelle Studien wollen aus beobachteten Charakteristika menschlichen Verhaltens Aufschluß über die Organisation des kognitiven Apparates gewinnen – hier soll dagegen ein bestimmtes, durch

[11] z.B. lautes Lesen, ausformulierte Vorträge oder Reden

empirische Studien bereits gestütztes Modell zugrundegelegt und es sollen aus ihm heraus Vorhersagen über Charakteristika der Sprachproduktion in Dialogsituationen gemacht werden.

**Generelle Einsichten über die Sprachproduktion** Ein einheitliches Modell, das von der gesamten sprachpsychologischen Forschung als Grundlage angenommen wird, existiert zum gegenwärtigen Zeitpunkt nicht. Einige grundlegende Aspekte des Sprachproduktionsprozesses werden jedoch weitgehend von einer Vielzahl der sprachpsychologischen Forscher akzeptiert, da sie mehrfach experimentell bestätigt werden konnten. Dazu gehören Charakteristika wie eine inkrementelle Arbeitsweise, eine Gliederung in verschiedene Repräsentationsebenen und ein unterschiedlicher Automatisierungsgrad der auf den verschiedenen Ebenen operierenden Prozesse (siehe z.B. Bourdin & Fayol, 1994).

Die Repräsentationsebenen, die sich aus der empirischen Forschung für die Sprachproduktion ergeben haben, lassen sich folgendermaßen beschreiben: In Abhängigkeit von der zu vermittelnden Information und seinen im aktuellen Diskursverlauf geltenden Zielen entwirft ein Sprecher über die Teilprozesse der Informationsselektion und der -linearisierung eine konzeptuelle Struktur der geplanten Äußerung. Diese quasi rein semantische Repräsentation wird durch die Auswahl passender Lemmata und deren Flexion in syntaktischen Strukturen gebunden. Motorische Pläne für die Ansteuerung der Muskulatur des Sprachproduktionsapparates bilden die letzte Ebene der Sprachproduktion, bevor diese Pläne in einen Lautstrom umgesetzt werden.

### 2.4.3 Das Sprachproduktionsmodell von Levelt (1989)

Levelt (1989) hat die allgemeinen Gedanken aus dem vorangehenden Abschnitt sowie Modelle von Teilsystemen der Sprachproduktion in einem einheitlichen Rahmen gebunden, der auf theoretischen Vorarbeiten aus den 70er Jahren basiert. Die aus diesen Arbeiten hervorgehenden Vorhersagen wurden neueren und älteren experimentellen Arbeiten gegenübergestellt und daraus ein detailliertes Bild der Sprachproduktion abgeleitet.

Ein wichtiges Merkmal der Sprachproduktion ist, daß sie als inkrementeller Prozeß organisiert ist. Diese Sichtweise wird nicht nur von Levelt (1989) vertreten, sondern ist auch in der gesamten vorangehenden Forschung verbreitet (siehe z.B. auch die Diskussion von Primärliteratur in Finkler, 1997, S. 46). Nur durch inkrementelle Verarbeitung wird eine massive Parallelverarbeitung und damit ein schneller und flexibler Ablauf bei der 'Berechnung' und Ausgabe mit unterschiedlichen Ein- und Ausgabeinkrementen möglich. Inkrementelle Verarbeitung beinhaltet zum einen, daß bereits generierte Zwischenergebnisse auf einer Ebene nach und nach **qualitativ** und **quantitativ** verfeinert werden können, andererseits aber auch, daß höhere Ebenen eine noch unvollständige Eingabe aus niederen Ebenen bereits zur Verarbeitung übernehmen können.[12]

In Levelts Modell finden sich die unterschiedlichen Repräsentationsebenen dieser inkrementellen Verarbeitung als Phasen eines *sequentiellen* Produktionsprozesses wieder. Die Phasen der **Konzeptualisierung**, **Formulierung** und **Artikulation** stehen durch unterschiedliche Rückkopplungskanäle miteinander in Verbindung. Einzelne Ausgaberepräsentationen werden

[12]Levelt (1989) bezeichnet dieses Funktionsprinzip als 'Wundts Prinzip'.

durch eine zentrale **Monitoring**-Komponente überwacht.[13] Abbildung 2.4 verdeutlicht grafisch die Zusammenhänge zwischen den unterschiedlichen Phasen, die in den folgenden Abschnitten vorgestellt werden.

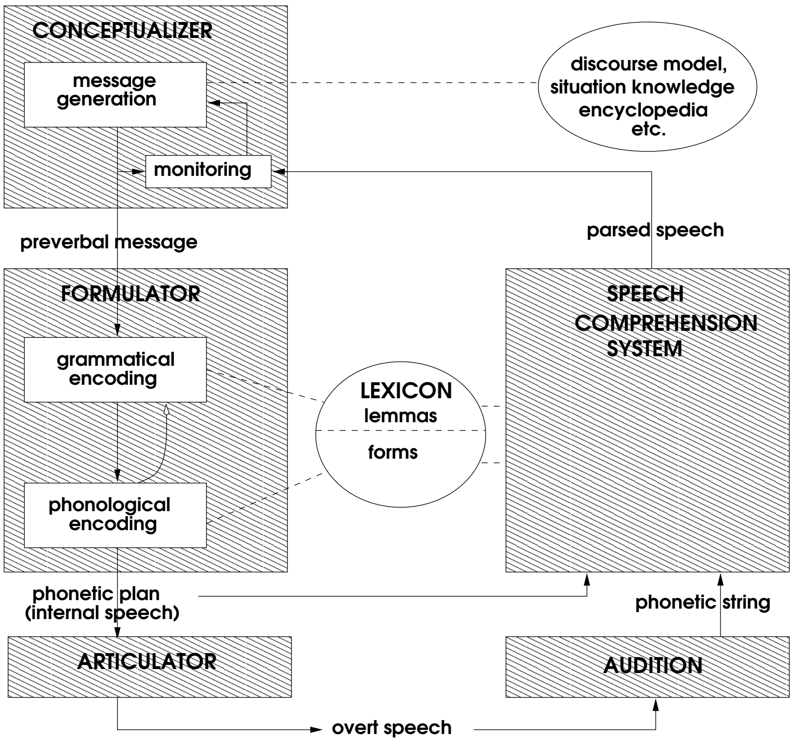


Abbildung 2.4: Schematische Darstellung des Sprachproduktionsmodells nach Levelt (1989)

Den unterschiedlichen Phasen der Sprachproduktion können unterschiedliche kognitive Prozesse zugeordnet werden. Daher unterscheiden sie sich vor allem auch bezüglich ihres Ressourceneinsatzes, ihres Automatisiertheitsgrades und ihrer Ressourcenadaptivität. Wenn Arbeitsgedächtnisbelastung und Zeitdruck eines Sprechers zu beurteilen sind, ist es besonders wichtig festzustellen,

- welche Merkmale der endgültigen Äußerung in welcher Verarbeitungsphase entstehen und
- wie stark diese Verarbeitungsphasen von der zur Verfügung stehenden Zeit und der Verfügbarkeit kognitiver Ressourcen abhängig sind.

Wie im Abschnitt 2.3 deutlich gemacht wurde, sind automatisierte Prozesse für Störungen durch Mehrfachbelastung nicht anfällig, da sie nicht von den attentionalen kognitiven Ressourcen abhängen. Auswirkungen kognitiver Belastung auf automatisierte Phasen der Sprachproduktion werden daher entweder gar nicht oder nur in geringem Maße auftreten. Deshalb sind

[13] Vielfach wird an der Architektur von Levelts Modell eine zu starke Modularität kritisiert (Rummer, 1996a; Pechmann, 1991). Levelt lehnt selbst jedoch den Begriff des *Moduls* aus den verschiedensten Gründen ab und läßt in seiner Theorie genügend Raum für alternative, parallele, konnektionistische Ansätze (Levelt, 1989, Kap. 1).

Beeinflussungen der Merkmale einer Äußerung, die aus einer stark automatisierten Phase resultieren, für die Modellierung der kognitiven Ressource Arbeitsgedächtnis nur von untergeordneter Bedeutung. Gleiches gilt auch für den Zeitdruck, da eine zeitliche Beeinflussung von automatischen Handlungsabläufen eine Steuerbarkeit der zugehörigen Prozesse erfordern würde und damit von Ressourcen der Zentralen Exekutive abhängig wäre.

#### 2.4.3.1 Konzeptualisierung

In der Konzeptualisierungsphase werden vom Sprecher die Inhalte seiner Äußerung festgelegt. Dafür muß der Sprecher

- die Intention der Äußerung aufgrund seiner Diskursziele planen,
- die zugehörigen Informationseinheiten auf einem prälinguistischen Niveau auswählen **(Selektion)**,
- Strukturen generieren, die die wichtigen Aspekte der intendierten Äußerung durch Spezifikation der zugehörigen Parameter festlegen **(Konstruktion)** und
- die einzelnen Informationseinheiten und Strukturen in eine zeitliche Reihenfolge bringen **(Linearisierung)**.

Ausgabe der Konzeptualisierungsphase sind präverbale Strukturen, die von Levelt als *messages* bezeichnet werden.[14] Sie legen durch spezielle Markierungen die für den Inhalt einer Äußerung nötigen thematischen und semantischen Parameter fest. Außerdem spezifizieren sie deren Modus (Frage, Aussage etc.), die Thema-Rhema-Struktur – bei Levelt als 'Perspektive' bezeichnet – und alle deiktischen und aspektuellen Eigenschaften. Dabei werden auch Referenten bezüglich ihrer Verfügbarkeit im Diskurskontext und ihrer Prominenz in der Diskursgeschichte gekennzeichnet. Die Repräsentation erfolgt, ohne bereits echtes sprachliches Material (Lemmata oder grammatische Strukturen) zu kodieren. Die Konzeptualisierungsphase stellt daher die Schnittstelle zwischen Denken und Sprache dar.

Levelt legt für die Organisation der präverbalen Darstellung propositionale Strukturen zugrunde (siehe z.B. Grabowski, 1991): Die Einheiten der konzeptuellen Ebene werden dabei in Relationen als Funktoren und Argumente sowie durch thematische Rollen gekennzeichnet.

Während der Konzeptualisierung werden zwei Ebenen der Planung unterschieden: In der sprachlichen **Makroplanung** werden durch eine Zerlegung der kommunikativen Ziele des Sprechers in Teilziele und die Selektion der zu vermittelnden Information *intendierte Sprechakte* abgeleitet. Während der Phase der **Mikroplanung** erfolgt die Festlegung einer Informationsstruktur für diese Sprechakte, indem Referenten und der Fokus der Äußerung markiert werden und sie zu propositionalen Strukturen vervollständigt werden (siehe z.B. auch Rummer & Grabowski, 1996; Hermann & Grabowski, 1994; Butterworth, 1980a).

Bereits wegen der Forderung hoher Flexibilität an die Phase der Konzeptualisierung können keine automatisierten Schemata für ihren Ablauf verantwortlich sein. Die Planungsvorgänge

[14] Da eine sinnvolle deutsche Übersetzung nicht gefunden werden konnte, soll im weiteren der Begriff der *präverbalen semantischen Strukturen* weiterverwendet werden.

Selektion und Linearisierung erfordern eine bewußte Kontrolle und Steuerung. Daher liegt während der Konzeptualisierung eine hohe Beanspruchung der zentralen kognitiven Ressourcen des Arbeitsgedächtnisses[15] vor. Die in dieser Phase ablaufenden Teilprozesse sind somit wesentlich von der Verfügbarkeit dieser Ressourcen und der Konkurrenz durch andere Prozesse abhängig.

#### 2.4.3.2 Formulierung

Sobald in der Konzeptualisierungsphase erste semantische Teilstrukturen generiert worden sind, können diese bereits von der Formulierungsphase weiterverarbeitet werden. Diese Verarbeitung gliedert sich in zwei Subphasen: Während der **grammatischen Enkodierung** werden für die Konzepte der präverbalen Strukturen und die zwischen ihnen spezifizierten Beziehungen passende Lemmata aus dem Langzeitgedächtnis abgerufen bzw. aktiviert. Diese müssen mit den Markierungen der konzeptuellen Struktur vereinbar sein. Die einzelnen Lemmata legen eigene Anforderungen an syntaktische Strukturen fest: Z.B. bestimmen Verben geeignete syntaktische Argumente als Füller für ihre thematischen Rollen. Unter diesen Einschränkungen werden Satzrahmen ausgewählt und die grammatischen Funktionen der vorhandenen Referenten festgelegt. Gesteuert wird die Generierung der zugehörigen syntaktischen Oberflächenstruktur sowohl durch die Funktions-Argument-Strukturen der konzeptuellen Ebene als auch durch die syntaktischen Constraints der Lemmata und Informationen über Diskursfokus und -perspektive.

In der resultierenden Oberflächenstruktur sind die einzelnen Lemmata durch den umgebenden syntaktischen Kontext bereits mit Flexionsinformation annotiert. Per Zugriff auf die lexikalische Form wird während der **phonologischen Enkodierung** ein artikulatorischer Plan aufgebaut, der auch prosodische Information über Markierung von Referenten im Diskurskontext enthält. Das Ergebnis dieser Phase bezeichnet Levelt als *internal speech*.

Eine absolute Klassifikation der grammatischen und phonologischen Enkodierung bezüglich ihrer Beanspruchung attentionaler Ressourcen scheint nicht möglich. Sowohl Levelt (1989) als auch Rummer (1997) und Power (1985) gehen davon aus, daß die gesamte Phase der Formulierung weitgehend automatisiert und nichtintentional ist. Besonders gilt dies für die phonologische Enkodierung, in der keine weitere Selektion, sondern lediglich eine Verarbeitung bereits bestehender und annotierter Strukturen vorgenommen wird. Einige Aspekte der Steuerbarkeit implizieren jedoch, daß die Phase der Formulierung zumindest bei Bedarf unter attentionale Kontrolle gestellt werden kann (z.B. Rummer & Grabowski, 1996; Postma, Kolk & Povel, 1990) .

#### 2.4.3.3 Artikulation

Während der abschließenden Artikulationsphase werden die Anweisungen des artikulatorischen Planes in Muskelbewegungen umgesetzt. Wie experimentell nachgewiesen werden konnte, ist dieser Prozeß selbst vollständig automatisiert.[16] Allerdings stehen bestimmte Aspekte dieser

[15] Rummer (1996a) setzt diese Ressourcen lediglich mit der Zentralen Exekutive gleich. Levelt (1989, S. 27) assoziiert sie mit dem gesamten Arbeitsgedächtnis.

[16] Eine bewußte Ansteuerung der rund 100 Muskeln (Levelt, 1989), die für die Artikulation eingesetzt werden, schließt sich von selbst aus.

Phase unter bewußter Kontrolle, wie sich z.B. durch bewußt eingesetzte Sprechpausen belegen läßt (Levelt, 1989). Power (1986) argumentiert aufgrund seiner Experimentalergebnisse sogar, daß während der Artikulation generell eine kognitive Beanspruchung besteht.

Die phonologischen Pläne, die die Eingabe für diese Phase darstellen, spezifizieren ausschließlich kontextirrelevante Aspekte der Artikulation, d.h. solche, die von den speziellen Gegebenheiten des Vokaltraktes unabhängig sind. Die spezielle Umsetzung zur Generierung des Lautstromes wird von Levelt der Tätigkeit eines selbstregulierenden Systems zugeschrieben, das mit diesen Parametern angesteuert wird.

Die Geschwindigkeit der Artikulation kann von der Geschwindigkeit abweichen, mit der die Formulierungskomponente ihre Ausgabe liefert. Für diese Zwecke wird bei Levelt die Existenz eines artikulatorischen Puffers angenommen, in dem die generierten phonologischen Phrasen zwischengespeichert werden. In Levelts Modell dient der Artikulatorische Puffer dafür, die Ausgabe der Formulierungskomponente fortführen zu können, ohne daß auf eine Abarbeitung der produzierten Ausgabe durch die Artikulationskomponente gewartet werden muß. Ein Überlauf des Puffers wird durch geeignete interne Mechanismen aufgefangen. Der Puffer ist nicht mit der Phonologischen Schleife identisch!

#### 2.4.3.4 Monitoring

Eine besondere Stellung im Sprachproduktionsmodell nimmt die Monitoring-Komponente ein, da sie keine generierende Funktion wahrnimmt und damit an der Konstruktion einer Äußerung nicht direkt beteiligt ist. Sie überwacht die Ausgaben der Konzeptualisierungsphase sowie der Formulierungs- und Artikulationsphase der Sprachproduktion und vergleicht diese mit den Zielen und Plänen des Sprechers, Anforderungen der Dialogsituation, einem Diskursmodell, einem Partnermodell und Diskurskonventionen und -maximen.

Während für den Monitor die Resultate der Konzeptualisierung direkt zugänglich sind, müssen die Ausgabestrukturen der Formulierungsphase (*internal speech*) und der Artikulationsphase (*overt speech*) erst in das interne Format der *parsed speech* umgewandelt werden. Die *parsed speech* enthält alle beim Sprachverstehen benötigte morphologische, syntaktische und semantische Information und liegt wieder in propositionaler Form vor. Der Monitoring-Prozeß ist in Abbildung 2.4 der Phase der Konzeptualisierung zugeordnet, da er nur auf konzeptuellen Strukturen operiert, die aus der *internal* und *overt speech* erst durch Analyseschritte gewonnen werden müssen. Verantwortlich dafür zeichnet das komplette Sprachverstehenssystem (Levelt, 1983, 1989), so daß eine Identität dieses Systems mit der Monitoring-Komponente von Levelt angenommen wird.

Der Monitoring-Prozeß bedient sich des Arbeitsgedächtnisses, um die Ausgaberepräsentationen des Sprachverstehens zwischenzuspeichern (Levelt, 1989, Kap.1 und 12). Er prüft nicht nur, ob die zu äußernde oder geäußerte Information mit der in der Konzeptualisierungsphase intendierten Information übereinstimmt, sondern auch, ob die Äußerung linguistisch wohlgeformt ist. Da auch das Sprachverstehen ein weitgehend von Aufmerksamkeit abhängiger Prozeß ist, gehört der Monitoring-Prozeß zu den von kognitiven Ressourcen abhängigen Subprozessen der Sprachproduktion.

### 2.4.4 Arbeitsgedächtnis und Sprachproduktion

In Abschnitt 2.2.1 wurden verschiedene Bestandteile des Baddeley'schen Arbeitsgedächtnismodells erläutert. Es liegt zwar nahe, spezielle Aspekte dieses Modells wie die Phonologische Schleife explizit im Zusammenhang mit einer Einschätzung der Arbeitsgedächtniskapazität zu modellieren. Zahlreiche negative experimentelle Evidenz (Gathercole & Baddeley, 1993) hat jedoch dazu geführt, daß die Bedeutung dieses Subsystems bei der Sprach*produktion* als gering eingeschätzt wird.[17] Ebensowenig können für die genaue Verbindung der einzelnen Sprachproduktionsphasen zur Zentralen Kontrolle derzeit detaillierte Aussage getroffen werden: Die Zentrale Kontrolle als Kontroll- und Koordinationsinstanz und ihr Zusammenhang mit anderen kognitiven Prozessen ist in Baddeleys Modell für derartige Aussagen noch zu unspezifiziert. Da auch auch im Sprachproduktionsmodell von Levelt (1989) kaum näher auf die genaue Verbindung von Sprachproduktionsprozessen zu den Subkomponenten des Arbeitsgedächtnisses eingegangen wird, muß die Position einer Gesamtsicht auf die kognitive Ressource Arbeitsgedächtnis, wie sie für die Einschätzung der Arbeitsgedächtniskapazität im Projekt READY (siehe Abschnitt 2.2.1) vertreten wird, als für den derzeitigen Forschungsstand maximal angesehen werden.

## 2.5 Einschätzung von Belastung und Zeitdruck mit Hilfe von Indikatoren aus der Sprachproduktion

### 2.5.1 Integration von Ergebnissen experimenteller Studien

Arbeitsgedächtnisbelastung und Zeitdruck eines Dialogteilnehmers sind keine direkt beobachtbaren Benutzereigenschaften. Daher ist es für ihre Einschätzung notwendig, die kausalen Zusammenhänge dieser beiden Phänomene mit beobachtbaren Ereignissen und Zuständen bzw. Verhaltensweisen des Benutzers zu kennen. Eine Unterscheidung von Ursachen und Wirkungen der beiden variablen Größen ist dabei erforderlich: Aus der Beobachtung von Ursachen kann eine Veränderung der nicht beobachtbaren Größe *vorhergesagt* werden, aus der Beobachtung von Wirkungen kann die Veränderung der Variablen *diagnostiziert* werden.[18] Nicht beobachtbare Größen werden in der psychologischen Forschung als *interne Variablen* bezeichnet.[19]

Die psychologische Forschung weist eine Vielzahl von Arbeiten auf, in denen Zusammenhänge zwischen dem Zustand des Arbeitsgedächtnisses bzw. Zeitdruck einerseits und der Veränderung zeitlicher und qualitativer Parameter des Verhaltens von Versuchspersonen andererseits experimentell untersucht wurden. Um eine konsistente Beurteilung der sprachlichen Eingabe eines Dialogsystems zu ermöglichen, wird in Kapitel 3 versucht, die experimtentel-

[17] Die Einsicht geht vor allem auf Versuchspersonen in neurolinguistischen Experimenten zurück, die in den für die Phonologische Schleife klassischen Tests der Behaltensspanne (Zahlen oder Nichtwörter) durch neurologische Beeinträchtigungen eine nur geringe Leistung zeigten, während sie parallel dazu jedoch eine normale Sprachproduktion aufwiesen (siehe Baddeley et al., 1998)

[18] Da entsprechende Inferenzen stark mit Unsicherheit behaftet sind, werden bei der Modellierung keine binärlogik-basierten Mechanismen eingesetzt. Siehe dazu Kapitel 6.

[19] Jameson (1997) zeigt, wie eine Klassifikation von Variablen in Ursachen und Wirkungen für sprachunabhängige Phänomene erfolgen kann.

len Studien im Lichte des vorgestellten Sprachproduktionsmodells zu interpretieren: Während bei Levelt (1989) ressourcenrelevante Fragestellungen lediglich am Rande untersucht werden, beeinflussen in den verschiedenen experimentellen Arbeiten unterschiedliche theoretische Annahmen über die Natur des menschlichen Sprachproduktionssystems sowohl das experimentelle Design als auch die Interpretation der Ergebnisse. Für die Einschätzung der Verwendbarkeit unterschiedlicher sprachlicher Parameter in einem natürlichsprachlichen Dialogsystem muß daher überprüft werden, ob die Vielzahl der experimentellen Ergebnisse auch mit dem verwendeten Sprachproduktionsmodell vereinbar ist.

Die dabei identifizierten kausalen Zusammenhänge können durch Graphen dargestellt werden. Knoten entsprechen dabei psychologischen Variablen. Die Richtung der Kanten zwischen zwei Knoten verdeutlicht die Beziehung von Ursache und Wirkung als gerichteten Kausalzusammenhang. Externe Ursachen (z.B. ein bevorstehender Termin) beeinflussen den Zustand der internen Variablen (z.B. Zeitdruck). Diese Zustandsveränderung manifestiert sich in einer Verhaltensveränderung (z.B. ständiger Blick zur Uhr, siehe Abb. 2.5).

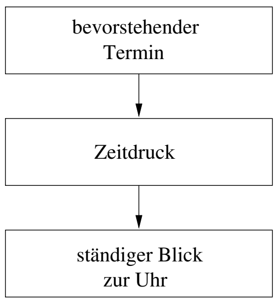


Abbildung 2.5: Darstellung eines kausalen Graphen

Veränderungen der Sprachproduktion aufgrund von Zeitdruck und beschränkter Arbeitsgedächtniskapazität werden im folgenden als **sprachliche Symptome bzw. Indikatoren für Zeitdruck und kognitive Belastung** bezeichnet.

### 2.5.2 Vorüberlegungen zu möglichen Symptomen

Levelts Modell ist ein kognitives Modell der Sprachproduktion. Daher können die allgemeinen Betrachtungen zu kognitiven Ressourcen und Prozessen aus den Abschnitten 2.1 bis 2.3.1 auf die Sprachproduktion übertragen werden. Welche konkreten Auswirkungen können Ressourcenbeschränkungen unter dieser Annahme auf die inkrementelle Verarbeitung haben und welche Auswirkungen sind daher für die Sprachproduktion zu erwarten?

- Ist in einer Verarbeitungsphase der Zugriff auf Ressourcen zur Verarbeitung der Eingabeinkremente verwehrt, bzw. wird der Zugriff wegen Beschränkungen der zur Verfügung stehenden Zeit nicht durchgeführt, erfolgt eine inhaltliche Modifikation der Zwischenergebnisse. Folge ist ein *qualitativ* beeinträchtigtes Endresultat.

- Die einzelnen Teilprozesse der Sprachproduktion sind stets wenigstens von der Ressource *Eingabedaten* abhängig. Muß eine Phase aus verschiedenen Gründen auf die Eingabeinkremente warten und kann sie die dadurch verlorene Zeit nicht durch eine Beschleunigung ihrer Verarbeitung kompensieren, wird die Geschwindigkeit des gesamten Sprachproduktionsprozesses reduziert. Damit reduziert sich die *Quantität* der innerhalb einer Zeitspanne generierten Augabe. Als Gründe für verzögerte Ausgabe kommen sowohl Interferenzen mit sprachproduktionsinternen als auch mit außerhalb der Sprachproduktion ablaufenden kognitiven Prozessen in Frage:

  - Durch die gleichzeitige Beanspruchung einer Ressource kann es zu einer unterschiedlichen Zuteilung an die konkurrierenden Prozesse kommen. Einer der beteiligten Prozesse kann gezwungen sein, auf die Freigabe der Ressource zu warten.
  - Durch die Kontrollinstanz des Monitorings können qualitativ inadäquate Zwischendarstellungen zu einer Korrektur zurückgewiesen werden. Die beanstandete Zwischendarstellung muß zumindest in Teilen zunächst erneut generiert werden und steht daher für Nachfolgephasen nicht zur Verfügung.

- Kontrollinstanzen sind im allgemeinen zusätzliche Werkzeuge, die zur Überprüfung der korrekten Funktionsweise anderer Komponenten eines Systems eingesetzt werden. Unter zunehmendem Zeitdruck (der innerhalb des Sprachproduktionsprozesses auch durch die Verlangsamung von Teilprozessen entstehen kann) kann es dazu kommen, daß diese Kontrolle abgeschaltet wird. Dadurch werden Zwischenresultate seltener überprüft, enthaltene Fehler bleiben unentdeckt. Daraus resultiert wiederum eine verminderte Qualität.

- Zu einem frühen Zeitpunkt getroffene Entscheidungen sind bei einer inkrementellen Verarbeitungsweise bereits nach einer Teilausgabe nicht mehr intern revidierbar. Verstärkt auftretende offene Korrekturen und abgebrochene Äußerungen als weiteres Merkmal verminderter Qualität sind die Folge.

## 2.6 Zusammenfassung

In den vorangegangenen Abschnitten wurden wichtige psychologische und psycholinguistische Grundlagen für die Erkennung von kognitiven Verarbeitungsbeschränkungen und Zeitdruck vorgestellt. Durch die Verwendung eines Sprachproduktionsmodells und eine differenzierte Sichtweise auf kognitive Ressourcen wird es möglich, experimentelle Arbeiten zum Zusammenhang von Belastung und Zeitdruck mit Charakteristika der Sprachproduktion zu beurteilen und die Konsistenz der Beobachtungen abzusichern. Einige Vorüberlegungen zu möglichen Effekten von Beschränkungen der Ressourcen *Arbeitsgedächtnis* und *Zeit* wurden diskutiert, die sich bereits theoretisch aus den beschriebenen Modellen ergeben. Im folgenden Kapitel sollen experimentelle Arbeiten zu konkreten Aspekten der Sprachproduktion unter Anwendung der erläuterten Begriffe und Mechanismen für die Identifikation *relevanter* sprachlicher Symptome für Verarbeitungsbeschränkungen herangezogen werden, die die Grundlage für den sprachbezogenen Teil einer Benutzermodellierungskomponente bilden können.

# Kapitel 3

# Psychologische Grundlagen: Empirie

> The purpose of theory is to generate new experimental questions and synthesise new links among old ones. Cognitive computational models sometimes fall short of these goals by reproducing a set of experimental results in a representational language which allows little contact with other cognitive systems and experimental methods.
>
> Malloch, Shapiro, Hitch, Culpin und Towse (1995)

Dieses Kapitel diskutiert und evaluiert empirische und experimentelle Untersuchungen aus der Psychologie unter den folgenden Gesichtspunkten:

- Welche Parameter wurden im Zusammenhang mit Experimenten zu kognitiver Belastung und Zeitdruck gemessen?
- Welche Veränderungen der Parameter wurden festgestellt?
- Sind die Parameter für die Verwendung in der Benutzermodellierungskomponente eines Dialogsystems relevant?

Das Kapitel liefert außerdem einige Hinweise auf mögliche Symptome für Verarbeitungsbeschränkungen, die sich aus der Anwendung psycholinguistischer Theorie[1] auf eine erste informelle Durchsicht der Daten aus der Voruntersuchung des Projektes READY ergaben[2]. Während Phänomene, die unter die erste Kategorie fallen, über eine empirische Grundlage verfügen, steht für die Phänome der zweiten Kategorie eine Überprüfung aus. Ihre Aufnahme in dieses Kapitel erfolgte, um mit der Einbindung in einen systematischen Rahmen die Datenanalyse und Modellierung für alle Symptome gleichermaßen zu fundieren. Anders formuliert ist das Ziel des Kapitels eine Darstellung charakteristischer Aspekte der Sprachproduktion unter Zeitdruck und kognitiver Belastung. Die angegebenen Beispiele wurden, so weit möglich, dem Korpus der genannten Vorstudie entnommen.

[1] Damit sind Schlußfolgerungen aus den experimentellen Arbeiten und dem Sprachproduktionsmodell gemeint.
[2] In Kapitel 4 sind nähere Angaben zu dieser Studie zu finden.

# 3.1 Vorgehensweise

## 3.1.1 Evaluationskriterien

Wenn sprachliche Parameter zur Einschätzung von Arbeitsgedächtnisbelastung und Zeitdruck eines Benutzers verwendet werden sollen, sind einige qualitative Merkmale der Zusammenhänge zwischen diesen beiden Benutzereigenschaften und ihren potentiellen Symptomen von besonderer Bedeutung. Sie beeinflussen vor allem die Entscheidung, ob und wie das entsprechende Symptom in verschiedenen Typen von Dialogsystemen Verwendung finden kann. Tabelle 3.22 im Abschnitt 3.9 am Ende dieses Kapitels gibt eine Übersicht über die Bewertung dieser Merkmale für die unterschiedlichen sprachlichen Phänomene. Die Merkmale im einzelnen sind:

**Kontextabhängigkeit:** Sind die Zusammenhänge in beliebigen bzw. zumindest möglichst vielen allgemeinen Kontexten gültig oder erfordern sie bestimmte Parameter der Sprechsituation? Besonderes Interesse gilt dabei den Fragen der *Meßbarkeit* und der *Richtung des experimentell festgestellten Effektes*.

- **Individuen:** Wie stark ist der Einfluß von individuellen Sprechercharakteristika auf Veränderungen der Sprachproduktion?
- **Textsorte:** Welche sozialen Parameter und Konventionen in der Sprechsituation beeinflussen das Verhalten eines Sprechers?
- **Natürlichkeit:** Sind die beobachteten Effekte künstlich experimentell induziert oder treten sie auch in realen Situationen auf?

**Weitere natürliche Ursachen:** Gibt es für das beobachtete Symptom neben kognitiven Verarbeitungsproblemen und Zeitdruck auch andere Ursachen, die bei der Modellbildung und Inferenz berücksichtigt werden müssen?

**Modellkonsistenz:** Sind die in psychologischen Experimenten beobachteten Zusammenhänge mit dem verwendeten Sprachproduktionsmodell und seinem Bezug zu kognitiver Belastung und Zeitdruck konsistent?

**Stärke des empirischen Nachweises:** Ist das Symptom lediglich theoretisch aus dem Sprachproduktionsmodell abgeleitet, oder bereits experimentell bestätigt worden? Gibt es ausreichend viele deutliche, oder nur wenige, evtl. widersprüchliche experimentelle Ergebnisse?[3]

**Notwendigkeit von Basiswerten:** Tritt das Phänomen bei Spontansprache immer auf und erfährt unter kognitiver Belastung und Zeitdruck lediglich eine *quantitative Veränderung* oder tritt es nur unter Belastung auf und ist damit Teil *qualitativer* Veränderungen der Sprachproduktion? Für die Parameter der erstgenannten Kategorie ist die Relativierung an einem Basiswert erforderlich.[4]

[3] Symptome, für die nur ein schwacher empirischer Nachweis besteht, sollten nur unter dem Vorbehalt weiterer empirischer Bestätigung bei der Modellierung kognitiver Benutzereigenschaften berücksichtigt werden.

[4] Insbesondere bei der spontanen Sprachproduktion besteht wegen der Spontaneität der Planung eine kognitive Grundbelastung. In der Schriftsprache nicht auftretende Symptome gehören zur Normalität. Wenn diese Phänome-

**Formalisierbarkeit:** Ist mit derzeit verfügbaren Werkzeugen der Linguistik, Informatik und Computerlinguistik eine formalisierte Darstellung und Verarbeitung möglich?[5]

### 3.1.2 Methodik der Darstellung

#### 3.1.2.1 Reihenfolge der Parameterdarstellung

In Kapitel 5 wird untersucht, inwiefern computerlinguistische Verarbeitungsphasen als Nebeneffekt Aufschluß über die für die Modellierung von Verarbeitungsbeschränkungen wichtigen Symptome liefern können. Da für die unterschiedlichen linguistischen Ebenen verschiedene computerlinguistische Techniken eingesetzt werden, ist es sinnvoll, eine Zuordnung der beobachtbaren Symptome zu den Ebenen der linguistischen Verarbeitung vorzunehmen. Diese Sicht orientiert sich an den Möglichkeiten der *Erhebung* der Symptomdaten. In Kapitel 2 wurde aber auch angedeutet, daß die Ursachen für das Auftreten der Parameter einzelnen Phasen des Sprachproduktionsprozesses zugeordnet werden können. Diese Unterteilung orientiert sich an der *Entstehung* der Symptome und ist nützlich unter dem Gesichtspunkt der Modellierung kognitiver Prozesse des Menschen. Sie bildet das Fundament für Kapitel 6. Da das Ziel der Arbeit die Verwendung der Ergebnisse in einem natürlichsprachlichen Dialogsystem ist, wird für die Darstellung in den folgenden Abschnitten die an der Reihenfolge der Verarbeitung sprachlicher Eingaben orientierte erstgenannte Sichtweise bevorzugt.

#### 3.1.2.2 Tabellen

In den nachfolgenden Abschnitten sind Tabellen zu finden, die Angaben über Methodik, qualitative und quantitative Ergebnisse experimenteller Arbeiten für das im jeweiligen Abschnitt behandelte sprachliche Phänomen zusammenfassen. Die Arbeiten zu kognitiver Belastung sind im oberen Teil der Tabelle angegeben; sofern Arbeiten zur Wirkung von Zeitdruck existieren, finden sich deren Ergebnisse im unteren Teil wieder. Die Ordnung erfolgt alphabetisch nach den Namen der Autoren. Arbeiten, die in einer Tabelle bereits zum wiederholten Mal zitiert werden, sind mit einem hochgestellten '=' gekennzeichnet.

Die Untersuchungsergebnisse werden in den verschiedenen Arbeiten mit unterschiedlichen, vom Versuchsaufbau abhängigen Tests auf signifikante Veränderungen geprüft. Die Resultate dieser Tests sind in Spalte *Signifikanzniveau* angegeben. Das Signifikanzniveau ($p$) gibt dabei die Wahrscheinlichkeit an, daß die Untersuchungsergebnisse rein zufällig entstanden sein könnten. In der psychologischen Forschung wird davon ausgegangen, daß ein Effekt mit einer anderen Stichprobe reproduzierbar ist, wenn $p < 0.05$. Bis zu $p < 0.1$ wird das Ergebnis oft als schwacher Nachweis der Reproduzierbarkeit gewertet. Falls die Autoren keine Zahlenangaben geliefert haben, ist in der Spalte 'k.A.' vermerkt.

Die Spalte *Ten.* gibt die Richtung (Tendenz) an, in die sich der Wert des untersuchten Parameters durch Belastung bzw. Zeitdruck verändert. Sie kann drei verschiedene Werte enthalten: '$+$' steht für eine Erhöhung des jeweiligen Parameters unter kognitiver Belastung/Zeitdruck,

ne im Zusammenhang mit Belastung und Zeitdruck untersucht werden, geht es nicht darum, *ob die entsprechenden Phänomene auftreten*, sondern *wie sie von zusätzlichen Aufgaben verändert werden*.

[5] Diese Fragestellung wird in Kapitel 5 untersucht.

'−' für eine Verringerung, '0' für keine Veränderung. Sind die Werte der Ergebnis-Spalte in Klammern gesetzt, wurde für die entsprechende Arbeit entweder keine statistische Auswertung angegeben[6] oder es wurde lediglich ein nichtsignifikanter Trend der Werteveränderung beobachtet.

### 3.1.3 Klassen von Phänomenen

Aus der Literatur ergeben sich drei große Klassen von Indikatoren für kognitive Belastung, deren Verhältnis zueinander in Abbildung 3.1 dargestellt ist.

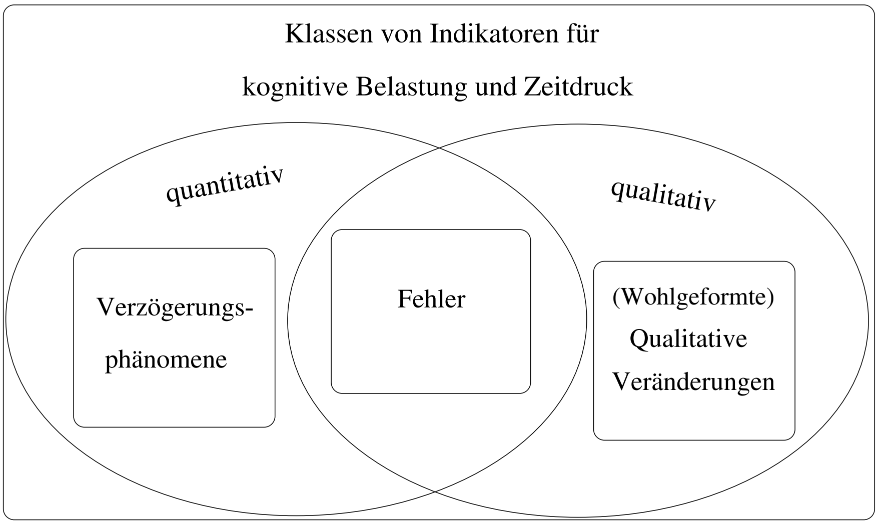


Abbildung 3.1: Klassen von Indikatoren für kognitive Belastung und Zeitdruck

**Klassen von Fehlern** Im Abschnitt 3.3.4 wird untersucht, inwiefern die Korrektur von Fehlern in einer Äußerung Aufschluß über die kognitive Belastung geben kann. Die Klassifizierung der sogenannten *Error-* Korrekturen (Levelt, 1983) liefert eine Klassifizierung der zugehörigen Sprechfehler. Das Auftreten der unterschiedlichen Fehlertypen hat durch ihre Zuordnung zu unterschiedlichen Ebenen der Sprachproduktion verschiedene Implikationen für die Einschätzung kognitiver Belastung. Da die einzelnen Fehlerkategorien auch unterschiedlichen linguistischen Ebenen zuzuordnen sind, sollen sie hier kurz zusammenfassend dargestellt werden:

Levelt unterscheidet **lexikalische, syntaktische und phonologische**[7] Fehler. Marx (1984) erweitert diese Klassifikation noch um **flexionsmorphologische** Fehler. Die einzelnen Klassen werden in verschiedenen Abschnitten behandelt: Phonologische Fehler im Abschnitt 3.2.8, lexikalische Fehler in 3.5.2, syntaktische und flexionsmorphologische Fehler im Abschnitt 3.3.

[6] Eine Entscheidung der Signifikanz auf Rohdaten oder Durchschnittswerten ist nicht möglich.

[7] Bei Levelt werden phonologische als 'phonetische Fehler' bezeichnet. Dieser Begriff ist jedoch m.E. zu eng gefaßt, zumal Levelt ihnen u.a. auch die Bezeichnung 'sublexikalische Fehler' zuordnet. Siehe auch Abschnitt 3.2.8.

## 3.2 Merkmale der Prosodie und Phonetik/Phonologie

Prosodische Merkmale einer Äußerung umfassen die zeitliche Organisation des Sprechens sowie Melodie und Akzent (Bußmann, 1990). Um den Gebieten Syntax und Semantik ausreichend Platz widmen zu können, beschränkt sich dieses Unterkapitel auf die zeitliche Dimension der Prosodie. Abschnitt 3.7 gibt u.a. einen Überblick über mögliche Symptome aus den anderen Bereichen.

### 3.2.1 Artikulationsgeschwindigkeit

#### 3.2.1.1 Abgrenzung des Phänomens: Artikulationsgeschwindigkeit

(3.1) Die **Artikulationsgeschwindigkeit** $G_A$ eines Sprechers ist der Quotient aus der Gesamtdauer aller *Artikulation*sphasen seiner Äußerung und der Anzahl der in diesen Phasen artikulierten sprachlichen Einheiten.

$$G_A = \frac{\text{Anzahl artikulierter sprachlicher Einheiten}}{\text{Gesamtdauer artikulierter sprachlicher Einheiten}}$$

Die Dauer der Artikulationsphasen läßt sich als Gesamtdauer der Äußerung abzüglich aller stillen Pausen und nichtartikulatorischen Geräusche berechnen. Als Basiseinheit für die Anzahl sprachlicher Einheiten kommen sprachliche Segmente unterschiedlicher Komplexität in Frage: Worte und Silben sind die bevorzugten Einheiten in den meisten psychologischen Untersuchungen. Silben sind jedoch den Worten vorzuziehen.[8]

(3.2) **Me010, Dialog 3:**
<**äh**> <**P**> Im <**D**> <**P**> Innenraum unter'm Lenkrad <**P**> links <**P**> ist ein Sicherungskasten.

*Dauer artikulierter sprachlicher Einheiten*

*Gesamtsprechdauer*

Die Gesamtsprechdauer im Beispiel beträgt 5.387 s, die (fettgedruckten) Verzögerungsphänomene[9] dauern insgesamt 2.187 s. 17 Silben wurden artikuliert. Die Artikulationsgeschwindigkeit beträgt daher 5.312 Silben/s.

#### 3.2.1.2 Artikulationsgeschwindigkeit und kognitive Belastung

Wenn ein Autofahrer in einer ihm unbekannten Ortschaft Zeit für die Orientierung benötigt, um seine weitere Fahrtroute zu bestimmen, wird er dafür seine Reisegeschwindigkeit vermindern. Hier liegt eine allgemeine Strategie der Handlungssteuerung bei überlappender Ausführung

[8] Gründe dafür sind, daß a) eine Definition des Begriffes Wortes problematisch ist (z.B. Komposita), b) die Wortlänge ohnehin eine zu starke Varianz aufweist, c) verschiedene Sprachen unterschiedliche durchschnittliche Wortlängen besitzen, d) die Silbe die Grundeinheit der prosodischen Forschung ist und e) Silben am häufigsten für die Messung der Artikulationsgeschwindigkeit eingesetzt werden. Siehe dazu auch Kowal (1991).

[9] <P> = stille Pause, <D> = Silbendehnung, <äh> = gefüllte Pause

von Handlung und Planung vor: Die Ausführungsgeschwindigkeit einer Aufgabe kann innerhalb eines gewissen Intervalls an die Geschwindigkeit der Planung angepaßt werden, wenn ein zum Ziel führender Zustand (z.B. die Bewegung des Fahrzeugs) eine Geschwindigkeit besitzt oder die Geschwindigkeit der zeitlichen Aufeinanderfolge von Teilaufgaben veränderlich ist. Dadurch wird während der Handlungsausführung durch eine Verringerung der Ausführungsgeschwindigkeit zusätzliche Zeit für die Planung und die Suche nach planungsunterstützenden Informationen gewonnen.

Eine Veränderung der Ausführungsgeschwindigkeit ist auch bei der Sprachproduktion, insbesondere in der Artikulationsphase möglich, wenn das Tempo der Generierung von Eingabeinkrementen für diese und die vorangehenden Phasen des Sprachproduktionsprozesses nicht ausreichend hoch ist. Verschiedene psychologische Experimente haben diesen Effekt kognitiver Belastung nachgewiesen. Eine Übersicht ist in Tabelle 3.1 angegeben.[10] In fast allen angegebenen Arbeiten wird eine Reduktion der Artikulationsgeschwindigkeit aufgrund zusätzlicher Belastung durch Doppelaufgaben oder eine höhere Zahl für die Konzeptualisierung notwendiger Inferenzen beobachtet.

#### 3.2.1.3 Problem Nr. 1: Individuelle Unterschiede

Die zitierten Arbeiten belegen, daß im Gegensatz zu den Überlegungen von Goldman-Eisler (1968) die Artikulationsgeschwindigkeit nicht als *unveränderliches* Persönlichkeitsmerkmal angesehen werden kann. Vielmehr stellen sie fest, daß verschiedene Individuen Sprache mit unterschiedlicher *Basis*geschwindigkeit artikulieren.

Um Veränderungen ausgewählter Sprechparameter in Situationen der Belastung oder des Zeitdrucks erkennen zu können, ist die Kenntnis von individuellen Basiswerten für diese Parameter nötig. Das Problem ist jedoch allgemeiner Art und auf weitere Symptome anwendbar. Da die psychologischen Untersuchungen diesen Punkt lediglich mit unterschiedlichen Methoden der statistischen Auswertung berücksichtigen und individuelle Werte in den Experimentalergebnissen nur selten angegeben werden, kann auf konkrete Unterschiede nicht näher eingegangen werden. Im Abschnitt 3.8 am Ende dieses Kapitels wird auf das Problem zurückzukommen sein.

#### 3.2.1.4 Problem Nr. 2: Kommunikative Funktion

Levelt (1989) weist auf eine möglicherweise lokal kommunikative Funktion der Artikulationsgeschwindigkeit hin, die dem Hörer vor allem am Ende von Sätzen anzeigen soll, daß eine Fortführung der Äußerung geplant ist. Deese (1980) stellt erhöhte Silbenraten nicht nur am Ende von Sätzen und am Beginn von Folgesätzen fest, sondern auch in sogenannten parenthetischen Kontexten[11]. Auch Deese interpretiert diese Befunde als Signal an den Hörer, daß der Sprecher noch keinen Abschluß seiner Äußerung beabsichtigt, sondern damit fortfahren möchte. Diese lokalen Veränderungen sind jedoch nicht wesentlich, da sie nur über wenige Silben hinweg verlaufen, während kognitive Belastung eine globalere Wirkung hat. Wenn die

[10] Zur Erläuterung des Tabellenaufbaus siehe Abschnitt 3.1.2

[11] Das sind Wiederholungen bereits geäußerter Fakten oder Äußerungsbestandteile, die lediglich zusätzliche Erklärungen liefern, jedoch für den Dialogverlauf unwichtig sind.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Duez (1982) | Quasiexperimentell: sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorbereitete Interviews | Unerwartete Interviews | 5.7 Silben/s[a] | 5.2 Silben/s[a] | $p = 0.05$ | – |
| Grosjean und Deschamps (1973) | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Interviews und Beschreibungen | 30/20[b] | Radio-Interviews | Cartoon-Beschreibungen | 5.29 Silben/s,[c] $(SD = 0.42)$ | 4.45 Silben/s,[c] $(SD = 0.4)$ | $p < 0.001$, $(z = 5.03)$ | – |
| Kowal und O'Connell (1987) | Sprachlich und sprachlich-visuell: Wiedergeben des Inhalts eines Films | 39 | Wiedergabe des Inhalts nach Ende des Films | Wiedergabe des Inhalts während des Films | 5.40 Silben/s, $(SD = 0.53)$ | 4.73 Silben/s, $(SD = 0.59)$ | $p < 0.01$, $(t(37) = 3.70)$ | – |
| Kowal, Wiese und O'Connell (1983) | Zusammenfassung experimenteller Arbeiten | – | Arbeiten zu Interviews (3) | Arbeiten zum Nacherzählen (11) | 5.26 Silben/s, $(SD = 0.14)$ | 5.17 Silben/s, $(SD = 0.53)$ | – | (–) |
| Lazarus-Mainka und Arnold (1987) | Nachsprechen von Worten, Sätzen und Geschichten bei gleichzeitiger Sortieraufgabe | 72 | Sortierung von Plasteelementen nach Form | Sortierung von Plasteelementen nach Form und Farbe | k.A. | k.A. | $p < 0.05$, $(F = 11.25)$ (Worte) bzw. $p < 0.05$, $(F = 6.13)$ (Sätze) | – |
| Wiese (1983), Bed.1 | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Sprechen in der Muttersprache | Sprechen in einer Fremdsprache | 4.57 Silben/s | 3.67 Silben/s | $p < 0.01$, $(F(1, 24) = 20.21)$ | – |
| Wiese (1983), Bed.2 | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Nacherzählung mit verfügbarem Stimulusmaterial | Nacherzählung ohne Zugang zum Stimulusmaterial | 4.4 Silben/s | 3.9 Silben/s | $p < 0.05$, $(F(1, 24) = 5.85)$ | – |

Tabelle 3.1: Untersuchungen zur Artikulationsgeschwindigkeit

[a] Duez schließt aus der zugrundegelegten Artikulationszeit auch gefüllte Pausen, Wiederholungen und Satzfragmente aus.

[b] Kontrollgruppe/Experimentalgruppe

[c] Die Originaldaten sind in Silben/min angegeben.

Artikulationsgeschwindigkeit für eine vollständige Äußerung erhoben wird, können sie daher vernachlässigt werden.

#### 3.2.1.5 Fazit

Die Artikulationsgeschwindigkeit ist ein guter Indikator für kognitive Belastung. Für den Zeitdruck konnten keine Evidenzen gefunden werden. Wichtig ist ihre Relativierung an einem individuellen Basiswert. Abbildung 3.2 gibt einen Überblick über mögliche Einflüsse auf die Veränderung der Artikulationsgeschwindigkeit.[12]

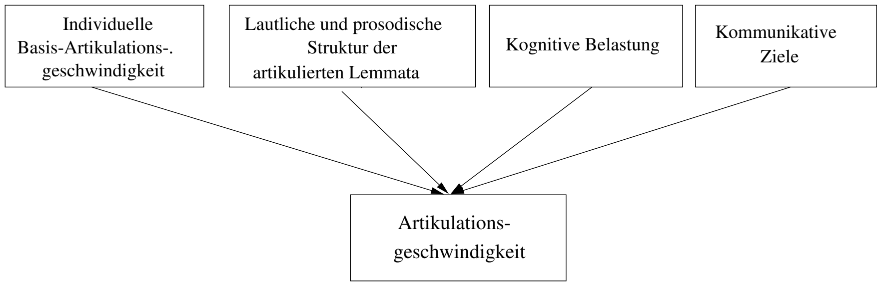


Abbildung 3.2: Einflüsse auf den Wert der Artikulationsgeschwindigkeit

### 3.2.2 Pausen

Neben der Geschwindigkeitsverringerung bei paralleler Handlungsausführung und -planung kann für den Gewinn zusätzlicher Planungszeit ein zweites Prinzip menschlicher Handlungssteuerung beobachtet werden: die vorübergehende Unterbrechung. Eine Unterbrechung kann automatisch entstehen, wenn aufgrund bisher vorliegender Planungsergebnisse keine Handlungsfortführung möglich ist.

Das Beispiel des Autofahrers macht auch diese Strategie klar: Kann die Entscheidung über die Fortführung der Fahrtroute nicht bis zu einem bestimmten Zeitpunkt getroffen werden (etwa bis unmittelbar vor dem Befahren einer Kreuzung), kann der Fahrer sein Fahrzeug vorübergehend stoppen.

Verzögerungen der Planung äußern sich somit als Pausen in den Handlungen, die die Pläne in Aktionen umsetzen. Bei der Sprachproduktion entstehen entsprechend Sprechpausen. Eine erste vorläufige Definition des Begriffes kann wie folgt lauten:

(3.3) **Sprechpausen** sind in den Strom der Sprache eingebettete Perioden vokaler Inaktivität. (Hieke, Kowal & O'Connell, 1983).

[12]Verschiedentlich wurde ich darauf hingewiesen, daß auch emotionale Belastung die Artikulationsgeschwindigkeit beeinflußt. Innerhalb des READY-Projektes wird jedoch emotionale Belastung als indirekte Ursache gesehen, die ihrerseits Einfluß auf kognitive Belastung und Zeitdruck ausübt. Weitere Angaben dazu finden sich in Weyrath (1998).

#### 3.2.2.1 Vorüberlegungen zu Sprechpausen und kognitiver Belastung

Frühe psychologische Untersuchungen zum Zusammenhang von Sprechpausen und kognitiven Planungsprozessen führte Goldman-Eisler (1968) aus. In ihren Experimenten beobachtete sie eine Korrelation von Pausenphänomenen mit dem Grad der kognitiven Belastung durch eine sprachliche Aufgabe und legte damit den Grundstein für eine rege Forschungsarbeit. Ihre Experimente bestätigten die Hypothese, daß unterschiedliche Formen der Sprachproduktion verschiedenen Planungsaufwand erfordern und die zeitlichen Effekte dieses Planungsaufwandes sich in der Sprachproduktion selbst niederschlagen. Sprechen kann daher als implizite Doppelaufgabe angesehen werden (Rummer, 1996a). Unterschiedliches Pausenverhalten in der Spontansprache unter Belastung wurde aber auch für Bedingungen mit nicht-sprachlichen Doppelaufgaben gegenüber Kontrollbedingungen nachgewiesen (Bromme & Wehner, 1987; Roßnagel, 1995a).

Die zwischen Konzeptualisierung und Artikulation liegende Phase der Formulierung ist weitgehend automatisiert und wird daher von zusätzlicher kognitiver Belastung kaum beeinflußt. Gleiches gilt auch für die Artikulationsphase selbst. Am stärksten wird das kognitive System nach Levelts Modell dagegen während der Konzeptualisierungsphase beansprucht. Bei einer kaskadierten, inkrementellen Verarbeitung, wie sie bei der Sprachproduktion zur Anwendung kommt, kann durch die Unterbrechung einer initialen Phase auch die Unterbrechung nachfolgender Phasen herbeigeführt werden, sofern diese Phasen nicht über genügend Reserven an noch nicht verarbeiteter Eingabeinformation verfügen (siehe z.B. Zitat von Fry in Butterworth, 1980a). Effekte erhöhter Planungsanforderungen oder reduzierter Konzeptualisierungsgeschwindigkeit lassen sich daher erst in der Artikulationsphase beobachten, obwohl sie nicht dort entstanden sind.

#### 3.2.2.2 Abgrenzung und Erweiterung des Pausenbegriffs

Die Argumentation des vorangegangenen Abschnitts setzt Pausen gleich mit einer unbewußten und willentlich nicht beeinflußbaren Unterbrechung der Artikulation. Zwei wichtige Punkte sind dabei jedoch außer acht gelassen worden:

- Sprechpausen können (auch kontrolliert) in weiteren Funktionen eingesetzt werden. Mehrere Einzelfunktionen können dabei gleichzeitig in einer einzelnen Sprechpause realisiert werden. Sprechpausen sind daher *multifunktional* (Kowal, 1991). Insgesamt lassen sich fünf wesentliche Funktionen identifizieren (siehe dazu u.a. Grosjean & Deschamps, 1973; Butterworth, 1980a; Gee & Grosjean, 1983; Wiese, 1983; Kowal, 1991):

  **Markierung** von Passagen einer Äußerung, um die Aufmerksamkeit des Hörers in Übereinstimmung mit den Zielen des Sprechers zu fokussieren,

  **Artikulation spezieller Phonemfolgen,** z.B. für den Druckaufbau bei der Artikulation von Plosivlauten,

  **Atmung,**

  **Unterbrechung** der Sprachproduktion für die Ausführung einer nicht-sprachlichen Handlung (Die Sprachproduktion wird erst nach Beendigung der Handlung fortgeführt.).

**Verzögerung** der Sprachproduktion, um Kapazität und/oder Zeit für dialogbegleitende Handlungen zu gewinnen.[13] Es wird versucht, die Sprachproduktion während der Handlung weiterzuführen.

Die Pausen, die der Verzögerung der Sprachproduktion zum Zwecke der Planung dienen, werden im folgenden als **kognitive Pausen** bezeichnet. Pausen, die die Markierungsfunktion umsetzen, heißen **kommunikative Pausen**[14].

- Der Begriff *Sprech*pause als Phase der Unterbrechung der Artikulation ist im Hinblick auf die Identifikation von verzögernden Phänomenen in der Sprachproduktion zu eng gefaßt:
  - Im allgemeinen ist es einem Dialogteilnehmer nicht möglich, über eine längere Phase der Stille hinweg die Sprecherrolle im Dialog zu behalten. Eine solche Periode könnte von seinem Dialogpartner als Zeichen mißverstanden werden, daß der Sprecher beabsichtigt, den *Dialogturn* abzugeben. Sprecher setzen aus diesem Grund bei der Sprachproduktion auch nicht-bedeutungstragende, aber artikulierte Füllaute wie 'äh', 'ähm' ein, um diese Phasen zu überbrücken.
  - Da sich Sprache durch ihre bedeutungstragende Funktion definiert und die zugrundeliegende semantische Struktur in der Konzeptualisierungsphase generiert wird, müssen auch diejenigen Abschnitte einer Signalsequenz als Pausen gezählt werden, in denen keine solche Bedeutung vermittelt wird (d.h. für die keine (konzeptuelle) Planung stattfinden muß). Dazu gehören stereotype oder floskelhafte Phrasen, die jedoch wegen der Kontextabhängigkeit ihrer Interpretation in Abschnitt 3.5 behandelt werden.

Für signallose Pausen wird im folgenden in Übereinstimmung mit der aufgeführten psychologischen Literatur der Begriff **stille Pause** verwendet. Füllaute werden als **gefüllte Pause**n bezeichnet.

#### 3.2.2.3 Pausen und kognitive Belastung: Experimentelle Ergebnisse

Sprechpausen treten bei jeder Form des Sprechens (nicht nur beim spontanen Sprechen, sondern etwa auch beim lauten Lesen) auf. Um ausschließlich diejenigen Pausen für die Einschätzung der Arbeitsgedächtnisbelastung des Sprechers heranzuziehen, die vorrangig aufgrund kognitiver Verarbeitungsprobleme entstehen, müssen die in einer Äußerung auftretenden Pausen geeignet klassifiziert werden. Durch eine solche Filterung wird sichergestellt, daß die Einschätzung der kognitiven Belastung z.B. nicht durch kommunikative Ziele des Sprechers verfälscht werden kann. Aufgrund der Multifunktionalität von Sprechpausen muß jedoch jede Einzelpause individuell unterschiedlich eingeschätzt werden, da jede Pause die verschiedenen Funktionen in unterschiedlichem Maße realisieren kann.[15]

[13] Unter Experimentalbedingungen werden entweder Aufgaben eingesetzt, die selbst sprachliches Material verwenden und daher unmittelbar mit der Sprachproduktion inferieren (z.B. die Wiederholungen von Zahlenreihen) oder solche, die die Konzentration auf sprachunabhängige Aufgaben erfordern (z.B. Tracking-Aufgaben).

[14] Die beiden Bezeichnungen gehen auf Kowal (1991) zurück.

[15] Kowal (1991) behauptet, daß Pausen *vor allem* durch ihre kommunikative Funktion charakterisiert würden und die Zuordnung kognitiver Ursachen in der psychologischen Forschung überbewertet worden sei. In dem von

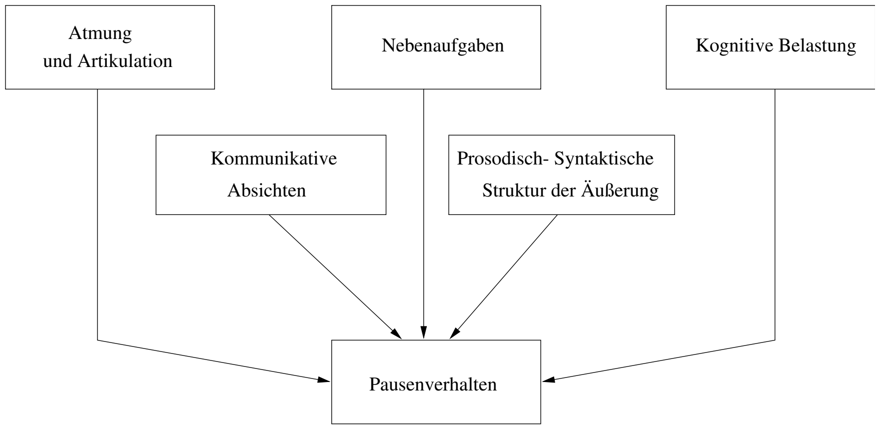


Abbildung 3.3: Einflüsse auf die Realisierung von Sprechpausen

Die Bewertung der pausenbezogenen Eigenschaften von Äußerungen umfaßt in den meisten experimentellen Arbeiten drei Bereiche: *Gesamtdauer*, *individuelle Dauer* und *Häufigkeit*. Um die kognitive Belastung eines Dialogpartners mithilfe seiner Pausenverteilung beurteilen zu können, reichen diese Dimensionen allein jedoch nicht aus. Die psychologischen Untersuchungen in den Tabellen 3.2 bis 3.6 trennen Pausen im Sinne der oben angegebenen Abgrenzung auch nach ihren lautlichen Aspekten in gefüllte und stille Pausen. Der nächste Abschnitt wird zeigen, daß eine weitere Differenzierung notwendig ist.

Als Ergebnis der angegebenen Arbeiten kann vorerst zusammengefaßt werden, daß unter kognitiver Belastung der zeitliche Gesamtanteil stiller und gefüllter Pausen an der Sprechdauer zunimmt. Das ist zum einen auf eine Zunahme der Anzahl stiller und gefüllter Pausen zurückzuführen, zum anderen auf die Verlängerung der durchschnittlichen Dauer von Einzelpausen.

#### 3.2.2.4 Vereinfachungsmöglichkeiten bei der Modellierung von Sprechpausen

Goldman-Eisler (1968), Kap.8 und Levelt (1989) stellen übereinstimmend fest, daß sich die Atmung an das Sprechverhalten anpaßt und nicht umgekehrt. Kowal (1991), S. 34 und Butterworth (1980a) heben außerdem hervor, daß *reine* Atempausen[16] nur selten auftreten: im allgemeinen werden Positionen einer Äußerung für Atmung genutzt, in denen ohnehin eine Pause in einer der anderen genannten Funktionen realisiert wird (siehe auch Breznitz, 1989; Grosjean & Deschamps, 1973): Es besteht eine Subsumptionsbeziehung zwischen Atempausen und den durch die vier anderen Funktionen gebildeten Klassen. Zusätzlich liegen derzeit fast keine Untersuchungsergebnisse vor, die den Zusammenhang von Atmung und Pausenverteilung beim spontanen Sprechen unter kognitiver Belastung zum Gegenstand haben (Kowal, 1991). Aus die-

ihr analysierten Sprachmaterial (Reden und Interviews verschiedener Politiker) mag dieses Argument Gültigkeit besitzen. Zahlreiche der hier zitierten Untersuchungen treten durch die Veränderungen der Pausenmaße aufgrund kognitiver Belastung jedoch den Gegenbeweis an.

[16]D.h. Pausen, die lediglich zum Luftholen dienen, jedoch an keiner grammatischen Grenze stehen.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte**[a] | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Duez (1982)[=] | Quasiexperimentell: sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorbereitete Interviews | Unerwartete Interviews | 25.3 % der GSD | 22.3 % der GSD | k.A. | $(-)$ |
| Goldman-Eisler (1968) | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Cartoonbeschreibung und -interpretation | 9 | Beschreibung von Cartoons | Interpretation von Cartoons | 55.4% der GSD[b] | 69.8% der GSD[b] | $p < 0.005$ $(t = 3.6$, one-tailed) | + |
| Greene (1984), Exp. 1 | Rein sprachliche Aufgabe mit unterschiedlichem Planungsaufwand: Darlegung eines Problems und seiner Lösung | 39 | Vorbereitetes 4- Schritte-Konzept auf Makroebene (2 Gruppen)[c] | Kein vorbereitetes Konzept zur Verfügung | 33.6% der GSD, $(SD = 13.1)$(Gr.1); 41.4% der GSD, $(SD = 9.8)$(Gr.2) | 45.8% der GSD, $(SD = 14.6)$ | Gruppe1: sign. (k.A.), Gruppe2: n.s. | + (+) |
| Greene (1984), Exp. 2[d] | Rein sprachliche Aufgabe mit unterschiedlichem Planungsaufwand: Darlegung eines Problems und seiner Lösungen | 30 | Diskussion eines Themas mit Kenntnis eines 4-Schritte- Konzepts und nach einer (Gr.1) bzw. zwei Übungssitzungen (Gr.2) | Keine vorherige Übung, kein Konzept zur Verfügung | 25.3% der GSD, $(SD = 10.5)$ (Gr.1); 23.9 % der GSD, $(SD = 0.100)$ (Gr.2) | 30.3 % der GSD, $(SD = 13.2)$ | n.s. | (+) |
| Greene, Lindsey und Hawn (1990) | Rein sprachliche Aufgaben mit unterschiedlichen Zielen: Charakterisierung einer Person | 70 | Klare und deutliche Darlegung einer Personenbeschreibung | Zusätzlich Beachtung sozialer Rücksichtnahme | 18.6 % der GSD[e] | 25.9 % der GSD[e] | $p = 0.064$ (n.s.), $(F(1,64) = 3.53)$ | (+) |
| Grosjean und Deschamps (1973)[=] | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Interviews und Beschreibungen | 30/20 | Radio-Interviews | Cartoon-Beschreibungen | 15.5 % der GSD[f], $(SD = 11.0)$ | 41.3 % der GSD[f], $(SD = 4.0)$ | $p < 0.001$, $(z = 5.94)$ | + |

Tabelle 3.2: Untersuchungen zur Dauer stiller Pausen

[a] GSD = Gesamtsprechdauer
[b] Originaldaten: pause/speech ratio $= 1.24$ (unbelastet) bzw. $2.31$ (belastet)
[c] Bei Gruppe 2 bestand eine leichte Grundbelastung, da die Reihenfolge der Punkte im Konzept gegensätzlich zur geforderten Präsentationsreihenfolge war.
[d] Die Messung der Pausendauer begann nach Präsentation der Aufgabe. Sie enthält daher die Einsatzlatenz.
[e] Originaldaten: pause/phonation ratio $= 0.229$ (unbelastet) bzw. $0.350$ (belastet)
[f] Originaldaten: Verhältnis Artikulationszeit/Gesamtsprechzeit $= 84.45$ (unbelastet) bzw. $58.71$ (belastet)

| Autoren | Methode | VPn | Aufgabe | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Kowal und O'Connell (1987)$^{=}$ | Sprachlich und sprachlich-visuell: Wiedergeben des Inhalts eines Films | 39 | Wiedergabe des Inhalts nach Ende des Films | Wiedergabe des Inhalts während des Films | 32.7 % der GSD, $(SD = 8.0)$ | 61.9 % der GSD, $(SD = 12.0)$ | sign. (k.A.) | + |
| Kowal et al. (1983)$^{=}$ | Zusammenfassung experimenteller Arbeiten | – | Arbeiten zu Interviews (3) | Arbeiten zum Nacherzählen (11) | 17.2 % der GSD, $(SD = 2.7)$ | 33.3 % der GSD, $(SD = 4.2)$ | – | (+) |
| Roßnagel (1995a) | Sprachliche Instruktion von Karatekas durch *ungeübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | 24.45 % der GSD | 35.12 % der GSD | $p < 0.001$, $(F(1,39) = 46.81)$ | + |
| Wiese (1983)$^{=}$ | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Sprechen in der Muttersprache | Sprechen in einer Fremdsprache | durchschnittliche Dauer: 0.77 s | durchschnittliche Dauer: 1.1 s | $p < 0.001$, $(F(1,24) = 13.38)$ | + |

Tabelle 3.3: Untersuchungen zur Dauer stiller Pausen (Fortsetzung)

| Autoren | Methode | VPn | Aufgabe | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Bromme und Wehner (1987) | Rein sprachliche Aufgaben mit unterschiedlicher außersprachlicher Planung | 12 | Zusammenstellung einer Speisekarte (Aufgabe) | Denksportaufgabe (Problem) | $\bar{x} = 10.4^a$, $(SD = 13.6)$ | $\bar{x} = 13.4^a$, $(SD = 19.6)$ | n.s. | (+) |
| Deese (1980) | Sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorherige Planung der Makrostruktur möglich (referatartiges Konzept) | Keine vorherige Planung der Makrostruktur | 7.36 Pausen/100 Worte | 7.22 Pausen/100 Worte | sign. (k.A.)[b] | – |
| Jou und Harris (1992)[c] | Vorlesen eines Zeitungsartikels mit anschließender Inhaltswiedergabe | 52 | Einfache Inhaltswiedergabe | Inhaltswiedergabe, begleitet von Kopfrechenaufgaben (mental arithmetic) | 0.47[d] | 2.53[d] | $p < 0.001$, $(F(1, 50) = 102.10)$ | + |
| Rummer (1996a), Exp. 1[e] | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer diskontinuierlichen Doppelaufgabe (Bewertung aufeinanderfolgender Lichtsignale auf farbliche Gleichheit) | 0.0321 /s (Bericht)[f], 0.0053 /s (Erzählen) | 0.0640 /s (Bericht)[f], 0.0515 /s (Erzählen) | $p < 0.01$, $(F(1, 54) = 14.57)$ | + |
| Rummer (1996a), Exp. 2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer kontinuierlichen Doppelaufgabe (Tracking- Aufgabe) | 0.032 /s (Bericht)[f], 0.005 /s (Erzählen) | 0.147 /s (Bericht)[f], 0.141 /s (Erzählen) | $p < 0.001$, $(F(1, 54) = 80.44)$ | + |

Tabelle 3.4: Untersuchungen zur Häufigkeit stiller Pausen

[a] Relativierte Anzahl stiller 5- Sekunden Intervalle. Eine Basiseinheit für die Relativierung ist nicht angegeben.

[b] Angabe nur für Sammelindex 'Type I disfluencies' (stille und gefüllte Pausen)

[c] Im Experiment wurden nur Pausen berücksichtigt, die länger waren als 5 s.

[d] Relativierte Anzahl von Pausen. Eine Basiseinheit für die Relativierung ist nicht angegeben.

[e] Im Experiment wurden nur Pausen berücksichtigt, die länger waren als 1 s.

[f] Es ist nicht eindeutig, wie Rummer sein Maß erhoben hat: die von ihm angegebene Tabelle ist mit 'Häufigkeit' überschrieben, im Text spricht er vom 'Anteil an der Redezeit'. Die Angaben erfolgen 'pro Zeiteinheit', daher vermutlich in Pausen/Sekunde.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Greene et al. (1990)[=] | Rein sprachliche Aufgaben mit unterschiedlichen Zielen: Charakterisierung einer Person | 70 | Klare und deutliche Darlegung einer Personenbeschreibung | Zusätzlich Beachtung sozialer Rücksichtnahme | k.A. | k.A. | $p = 0.168$ (n.s.), $(F(1,64) = 1.94)$ | 0 |
| Grosjean und Deschamps (1973)[=] | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Interviews und Beschreibungen | 30/20[a] | Radio-Interviews | Cartoon-Beschreibungen | $0.52$ s ; [b] | $1.32$ s ; [c] | $p < 0.001$, $(z = 5.94)$ | + |

Tabelle 3.5: Untersuchungen zur Dauer gefüllter Pausen

[a] Kontrollgruppe/Experimentalgruppe
[b] alle gefüllten Pausen $< 2$ s, 94.51 % unter 1 s
[c] 15 % der gefüllten Pausen zwischen 2 und 6 s

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Deese (1980)$^{=}$ | Sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorherige Planung der Makrostruktur möglich (referatartiges Konzept) | Keine vorherige Planung der Makrostruktur | 3.82 Pausen/100 Worte | 2.55 Pausen/100 Worte | sign. (k.A.)[a] | – |
| Duez (1982)$^{=}$ | Quasiexperimentell: sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorbereitete Interviews | Unerwartete Interviews | 0.077 Pausen/s[b] | 0.182 Pausen/s[b] | k.A. | (+) |
| Greene (1984)$^{=}$, Exp. 1 | Rein sprachliche Aufgabe mit unterschiedlichem Planungsaufwand: Darlegung eines Problems und seiner Lösung | 39 | Vorbereitetes 4- Schritte-Konzept auf Makroebene (2 Gruppen)[c] | Kein vorbereitetes Konzept zur Verfügung | 0.06 Pausen/s, $(SD = 0.051)$ (Gr.1) 0.113 Pausen/s $(SD = 0.135)$ (Gr.2) | 0.092 Pausen/s, $(SD = 0.083)$ | n.s. | (+) (–) |
| Rummer (1996a)$^{=}$, Exp. 1 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer diskontinuierlichen Doppelaufgabe (Bewertung aufeinanderfolgender Lichtsignale auf farbliche Gleichheit) | 0.0996 /s (Bericht)[d], 0.0322 /s (Erzählen) | 0.1523 /s (Bericht)[d], 0.1247 /s (Erzählen) | $p < 0.01$, $(F(1, 54) = 8.04)$ | + |
| Rummer (1996a)$^{=}$, Exp. 2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer kontinuierlichen Doppelaufgabe (Tracking- Aufgabe) | 0.020 /s (Bericht)[d], 0.032 /s (Erzählen) | 0.188 /s (Bericht)[d], 0.185 /s (Erzählen) | n.s. | (+)[e] |
| Wiese (1983)$^{=}$ | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Sprechen in der Muttersprache | Sprechen in einer Fremdsprache | 2.44/100 Silben, $(SD = 1.77)$ | 5.67/100 Silben, $(SD = 3.63)$ | $p < 0.001$ | + |

Tabelle 3.6: Untersuchungen zur Häufigkeit gefüllter Pausen

[a] Angabe nur für Sammelindex 'Type I disfluencies' (stille und gefüllte Pausen)

[b] Duez gibt Rohdaten an: In 22'20" wurden in der unbelasteten Bedingung 103 gefüllte Pausen gezählt, in der belasteten in 13'08" dagegen 143.

[c] Bei Gruppe 2 bestand eine leichte Grundbelastung, da die Reihenfolge der Punkte im Konzept gegensätzlich zur geforderten Präsentationsreihenfolge war.

[d] Es ist nicht eindeutig, wie Rummer sein Maß erhoben hat: die von ihm angegebene Tabelle ist mit 'Häufigkeit' überschrieben, im Text spricht er vom 'Anteil an der Redezeit'. Die Angaben erfolgen 'pro Zeiteinheit', daher vermutlich in Pausen/Sekunde.

[e] Im Text und einer grafischen Darstellung weist Rummer (Rummer, 1996a, S.119f.) auf einen '*Rückgang [gefüllter Pausen] in der Berichtsituation*' hin. Diese Angabe widerspricht jedoch den Werten in seiner Tabelle 9.4. Übernommen wurden die Daten der Tabelle.

sen Gründen wird vereinfachend für diese Arbeit die Annahme gemacht, daß Atempausen als gesonderte Klasse vernachlässigt werden können und davon auszugehen ist, daß sie mit den anderen Pausenklassen erfaßt sind.

Eine zweite Vereinfachung betrifft die Analyse prosodischer Merkmale von Pausen in ihrem Kontext. Deese (1980) unterscheidet *hesitation pauses* und *linguistically intentional pauses*: kognitive Pausen entstehen demnach nach den Bedürfnissen des Sprechers und sind nicht Bestandteil der Äußerungsplanung. Kommunikative Pausen dagegen sind intentional und werden gezielt im Diskurs plaziert. Daher entstehen sie nicht zufällig in der Artikulationsphase, sondern sind bereits Bestandteil der phonologischen Pläne der Formulierungsphase. Entsprechend zeigt Deese (1980), daß diese Pausen sich in den prosodischen Verlauf einer Äußerung einpassen. Prosodische Merkmale könnten daher bei der Zuordnung von Pausen zu den Kategorien *kognitiv* und *kommunikativ* eine höhere Präzision erreichen helfen.

#### 3.2.2.5 Pausen und kognitive Belastung: Notwendigkeit differenzierterer Kriterien

Die Darstellung in den Tabellen 3.2 bis 3.6 faßt die experimentellen Arbeiten unter den allgemeingültigsten Subklassen zusammen, die für alle Arbeiten gültig sind. Zahlreiche Autoren weisen jedoch auf die Notwendigkeit einer weiteren Differenzierung der Pausendaten hin, vor allem im Hinblick auf ihre Funktion. Unter dem Aspekt der Modellierung (Kapitel 6) ist eine formale Klassifizierungsmöglichkeit für Einzelpausen nach ihrer Zugehörigkeit zu den einzelnen Funktionen ebenfalls wünschenswert. In den folgenden Abschnitten werden einige zusätzliche Dimensionen der Pausenrealisierung diskutiert, die sich für diese Zwecke als hilfreich herausgestellt haben. Abbildung 3.4 gibt einen Überblick über diese Kriterien.

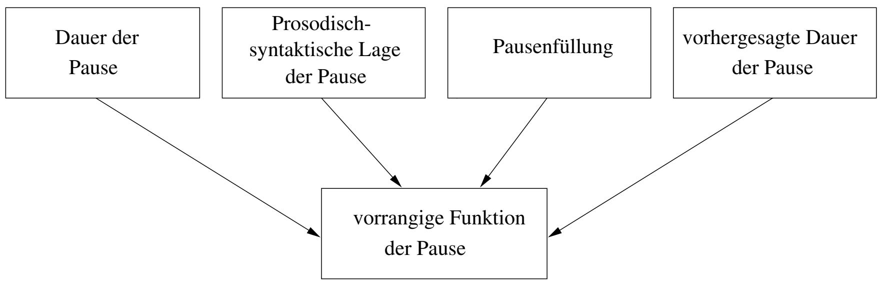


Abbildung 3.4: Hilfreiche Kriterien zur Feststellung der Pausenfunktion

**Pausenfüllung** Stille Pausen können prinzipiell jede der oben genannten Funktionen realisieren. Gefüllte Pausen dagegen erfordern vokale Artikulation und sind daher nicht parallel zur Verschlußbildung einsetzbar. Wegen der bei der Artikulation notwendigen Expiration kommen gefüllte Pausen auch für Atmung nicht in Frage.

Genausowenig wird offenbar die kommunikative Funktion der Markierung mit gefüllten Pausen realisiert: Besonders Politiker messen der kommunikativen Funktion von Sprechpausen hohe Bedeutung bei (Kowal, 1991). Entsprechend sind in der Sprachproduktion von Politikern außergewöhnlich viele kommunikative Pausen zu erwarten. Duez (1982) untersucht spontansprachliche Äußerungen von Politikern für drei verschiedene Situationstypen (Reden, vorbereitete sowie zufällige Interviews). Sie stellt fest, daß bei den Reden gefüllte Pausen fast nicht auftreten. Der kommunikative Wert gefüllter Pausen muß daher von den Sprechern als äußerst gering eingeschätzt werden. Damit verbleibt als einzige Funktion gefüllter Pausen offenbar die Verzögerung.

Auch die Form der Kommunikationssituation hat einen Einfluß auf das Auftreten von gefüllten Pausen: Greene (1984) weist nach, daß die Zahl gefüllter Pausen in monologischen Situationen von kognitiver Belastung kaum beeinflußt wird. Verzögerungseffekte werden von seinen Probanden lediglich mit einer höheren Anzahl stiller Pausen erreicht. Daraus folgert Greene, daß gefüllte Pausen als Verzögerungsphänomen für Monologe nicht zur Anwendung kommen. Sie sind in Dialogen daher offenbar vor allem auf die Anwesenheit eines zweiten Sprechers zurückzuführen und treten daher in einer Doppelfunktion auf: Einerseits verzögern sie die Sprachproduktion, andererseits signalisieren sie dem Dialogpartner, daß eine Fortsetzung der begonnenen Äußerung beabsichtigt und eine Unterbrechung nicht erwünscht ist (Mitchell, Hoit & Watson, 1996; Greene, 1984).[17]

Da in Dialogen auch stille Pausen als Mittel der Verzögerung eingesetzt werden können, droht eine Unterbrechung durch den Dialogpartner offenbar nur dann, wenn die Dauer der Verzögerung eine gewisse Grenze überschreitet. Gefüllte Pausen sorgen dafür, daß Verzögerungsdauern erreicht werden können, die durch stille Pausen allein nicht umzusetzen sind, d.h. sie deuten auf einen besonders hohen Planungsbedarf hin. Gefüllte Pausen sind daher für die Einschätzung kognitiver Belastung besonders wichtig.

**Pausenlage** Pausen können (theoretisch) an beliebigen Stellen zwischen den Wörtern einer Äußerung auftreten. Der rein theoretische Wert dieser Aussage wird anhand folgender Argumentation klar, die gleichzeitig zu einer Differenzierung der Pausenfunktion entsprechend ihrer syntaktischen Lage führt:

Sprechparameter in Bedingungen des *lauten Lesens* werden als Standardwert für das individuelle Sprechverhalten gewertet, da eine starke Konzentration auf die Artikulationsphase der Sprachproduktion erfolgen kann (Bromme & Wehner, 1987). Insbesondere entsteht beim Lesen keine kognitive Belastung durch Konzeptualisierungsarbeit (Anderson, 1985). Daher werden unter Lesebedingungen im allgemeinen keine für Planung (oder Nebenhandlungen) verzögernden Pausen erwartet.

Auffällig ist, daß sich beim Vergleich spontansprachlicher Äußerungen mit Äußerungen aus Kontrollbedingungen des lauten Lesens verschiedene Pausenverteilungen ergeben. Es konnte nachgewiesen werden, daß die Verteilung der beim lauten Lesen markierend/strukturierend eingesetzten Pausen wohldefinierten Beschränkungen durch die syntaktische Struktur der Äußerungen folgt (Gee & Grosjean, 1983; Deese, 1980). Butterworth (1980a) identifiziert beim lauten Lesen Pausen verstärkt an Phrasengrenzen.

Eine syntaktische Klassifikation der Pausenlage kann somit Vorhersagen über erwartete Pausendauern aufgrund kommunikativer Ziele unterstützen. Folgende Positionsklassen haben sich

[17] Diese Form kommunikativer Funktion ist von der oben angegebenen kommunikativen Funktion verschieden und setzt die Notwendigkeit einer Verzögerung voraus.

als relevant erwiesen:

- **Äußerungsgrenzen**[18],
- **Satzgrenzen** innerhalb einer Äußerung,
- **Phrasengrenzen** innerhalb eines Satzes und
- Positionen **innerhalb von Phrasen**.[19]

Vertreter der letztgenannten Gruppe werden als *intraphrasale Pausen* bezeichnet. Jou und Harris (1992) weisen darauf hin, daß unter unbelasteten Sprechbedingungen normalerweise versucht wird, die Integrität einer Phrase zu erhalten, indem sie nicht von intraphrasalen Pausen durchbrochen wird. Daher werden vor allem intraphrasale Pausen häufig mit Planungsvorgängen assoziiert: Duez (1982) stellt einen erhöhten intraphrasalen Pausenanteil in den von ihr untersuchten Korpora vor allem in den Sprechsituationen mit höheren Planungsanforderungen fest. Jou und Harris (1992) messen unter Belastung eine signifikante Zunahme ($p < 0.01$) der Anzahl intraphrasaler Pausen. Grosjean und Deschamps (1973) beobachten unter Belastung weniger eine zahlenmäßige Zunahme als vielmehr eine Verlängerung der durchschnittlichen Dauer intraphrasaler Pausen. Butterworth (1980a) berichtet von einem unpublizierten Experiment, bei dem ein reduzierter Anteil an intraphrasalen Pausen unter Wiederholungs- und Übungsbedingungen festgestellt wurde.[20] Ähnliche Ergebnisse erzielt auch Roßnagel (1995a): Der Anteil intraphrasaler Pausen einer geübten Versuchsgruppe steigt bei Belastung durch eine Zusatzaufgabe signifikant an ($p < 0.001$). Außerdem liegt der Anteil intraphrasaler Pausen am Gesamtpausenanteil bei der geübten Gruppe immer unter dem einer Gruppe mit den gleichen Aufgaben, die aus in der Aufgabe ungeübten Sprechern besteht.[21]

Eine Zunahme oder ein extrem hoher Anteil intraphrasaler Pausen kann demnach als deutliches Symptom kognitiver Belastung gewertet werden.

Roßnagel (1995b) weist jedoch ein gegensätzliches Ergebnis vor: er registriert einen Rückgang des prozentualen Anteils intraphrasaler Pausen an der Gesamtpausendauer zugunsten extraphrasaler Pausen. Als Erklärung nimmt Roßnagel an, daß Phrasen die kleinsten Planungseinheiten bei der Sprachproduktion sind und ein Kontextwechsel von der Sprach- zur Doppelaufgabe nur in den Planungspausen, also extraphrasal erfolgen konnte. Außerdem assoziiert er intraphrasale Pausen mit einer besonders aufmerksamen Wortwahl.Dieses Untersuchungsergebnis weist darauf hin, daß verschiedene Strategien der Anpassung der Sprachproduktion in unterschiedlichen Situationen möglich sind.

[18]Diese Pausen werden unter dem Begriff der *Einsatzlatenz* behandelt: siehe Abschnitt 3.2.6.

[19]Als *Phrase* wird eine syntaktisch zusammengehörige Wortgruppe ohne finite Verbform bezeichnet, als *(Teil)satz* eine syntaktische Konstruktionen mit finiter Verbform (Bußmann, 1990).

[20]Die Belastung wird dadurch reduziert, daß Formulierungs- und Artikulationsphase der Sprachproduktion von Wiederholung und Übung einer Aufgabe weitgehend unangetastet bleiben, in der Konzeptualisierungsphase jedoch Ergebnisse früherer Planungsvorgänge nicht neu generiert werden müssen, sondern noch weitgehend im Arbeitsgedächtnis verfügbar sind.

[21]Aufgrund der sehr heterogen präsentierten Meßwerte wurde auf die Darstellung von Ergebnissen zu Häufigkeit und Dauer intraphrasaler Pausen in einer Tabelle verzichtet. Die einzelnen Arbeiten sind in anderen Tabellen mehrfach genannt.

**Pausendauer** Experimentell nachgewiesen wurden auch Zusammenhänge zwischen der Dauer von Pausen und realisierter Funktion:

- Das Intervall, in dem kognitive stille Pausen betrachtet werden, kann **nach unten begrenzt** werden: Unterhalb einer bestimmten Minimaldauer sind Pausen ausschließlich sprecher- jedoch nicht situationsabhängig verteilt. Diese Pausen realisieren ausschließlich die motorisch-artikulatorische Funktion der Verschlußbildung und haben keine kognitiven oder kommunikativen Ursachen. Um zu vermeiden, daß derartige artikulationsbedingte Pausen Aufnahme in die gemessene Gesamtpausendauer finden, wird die Dauer der gemessenen Pausen in experimentellen Untersuchungen nach unten begrenzt. Goldman-Eisler (1968, S.36) setzte 250 ms als Grenze ein, wies jedoch gleichzeitig auf den dadurch möglichen Informationsverlust hin.[22] Kowal (1991), und Roßnagel (1995b) verwenden 270 ms[23]. Grosjean und Deschamps (1973), Greene (1984) und Mitchell et al. (1996) arbeiten ebenfalls mit 250 ms, Butterworth (1980a), Beattie (1983) und Boomer (1965) mit 200 ms. Duez (1982) geht bis auf 180 ms herunter, Hieke et al. (1983) schlagen sogar 130 ms vor.[24] Das Festlegen einer Minimaldauer ist somit auch bei der Pausenanalyse zum Zwecke der Einschätzung kognitiver Belastung zulässig. Kowal et al. (1983) geben eine ausführlichere Übersicht über die in einer Sammlung von Untersuchungen aus dem Bereich des Erzählens verwendeten Minimalwerte.

- Bereits erwähnt wurde, daß Konventionen der Dialogführung es einem Sprecher nicht erlauben, beliebig lange (still) zu pausieren. Lange stille Pausen führen dazu, daß der Dialogpartner den Dialogturn übernehmen möchte. Die Dauer stiller Pausen ist daher in Dialogen auch **nach oben begrenzt**. Butterworth (1980a) gibt dafür in Dialogen 2.5 s als Maximalwert an, Duez (1982) 2 s.[25] Für Pausen, die in Dialogsituationen den Wert von 2 - 2.5 überschreiten, kann angenommen werden, daß der Sprecher keine Fortsetzung seiner Sprachproduktion beabsichtigt. Ausnahmen sind Situationen, in denen ein Dialogteilnehmer die Beantwortung einer Frage erwartet: Sofern dafür von seinem Partner zunächst außersprachliche Handlungen auszuführen sind , können Pausen über den genannten Maximalwert hinausgehen. Lange Pausen können demnach im allgemeinen Nebenaufgaben mit Unterbrechung der Sprachproduktion zugeschrieben werden.

- Ähnliche Längenbeschränkungen gelten auch für gefüllte Pausen: In den Untersuchungen von Grosjean und Deschamps (1973) liegen die meisten gefüllten Pausen im Bereich unterhalb 2s, in ihrer Dialog-Bedingung 'Interview' sogar alle.

- Kombinationen von stillen und gefüllten Pausen können die angegebenen Maximallängen überschreiten. Leider nimmt keine der vorgestellten experimentellen Arbeiten zu diesen Kombinationen Stellung. Durch die Zusammenfassung gefüllter und stiller Pausen in Gesamtwerten wird die lokale Kombination der unterschiedlichen Pausen vernachlässigt.In

[22] In der Realität sind Verschlußphasen nur selten länger als 100 ms (Jürgen Trouvain, persönliches Gespräch).

[23] Tatsächlich gibt Roßnagel (1995a) 27 ms an. Diese Angabe ist höchstwahrscheinlich fehlerhaft abgedruckt.

[24] Deese (1980) interpretiert die willkürlich erscheinenden Grenzen als Ausdruck der Unsicherheit, mit der Pausen bestimmter Dauer der artikulatorisch bedingten Klasse zugeordnet werden.

[25] Untersuchungen zu Monologen (z.B. Bromme & Wehner, 1987; Jou & Harris, 1992) zeigen, daß diese Längenbeschränkungen nicht für beliebige Situationen der Sprachproduktion gelten, sondern in Monologen durchaus mehr als 5 s Unterbrechung auftreten können.

den Voruntersuchungen zum Projekt READY treten jedoch Pausen auf, die durch die Kombination von Einzelpausen, die den genannten Längenbeschränkungen entsprechen, in ihrer Gesamtdauer die angegebenen Maximalwerte überschreiten. Dadurch erreicht ein Sprecher offenbar eine Verlängerung der maximalen Verzögerungsdauer ohne Freigabe des Dialogturns. Derartige Extremverzögerungen können nur durch einen hohen Verzögerungsbedarf, jedoch nicht mehr durch kommunikative Ziele erklärt werden und sind daher besondere Indikatoren kognitiver Belastung.

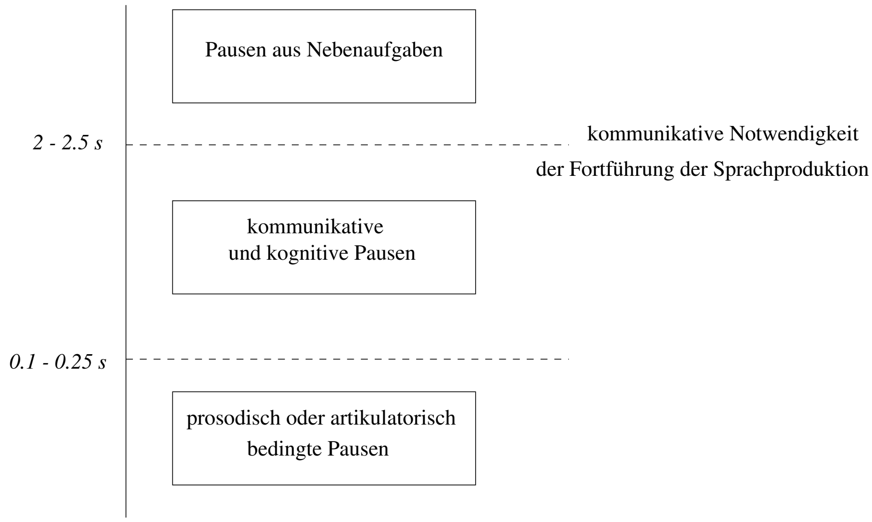


Abbildung 3.5: Grobklassifizierung der Pausenfunktion nach ihrer Dauer

Aus der bisherigen Analyse der anderen Pausencharakteristika ergab sich, daß Pausen an verschiedenen syntaktischen Positionen einer Äußerung unterschiedliche Funktionen mit unterschiedlicher Wahrscheinlichkeit realisieren. Auch die Funktion beeinflußt die Pausendauer: Pausen zwischen Sätzen sind länger als solche zwischen Phrasen; diese wiederum sind länger als phraseninterne Pausen (Duez, 1982; Grosjean & Deschamps, 1973; Butcher, 1980).

Gee und Grosjean (1983) haben aus einer Untersuchung der Pausenzeiten beim lauten Lesen ein Modell abgeleitet, mit dem Vorhersagen über die Realisierung kommunikativer Pausen in planungsfreien bzw. wenig planungsintensiven Zusammenhängen gemacht werden können. Die Vorhersagewerte aus ihrem PHI-Algorithmus können helfen, die kommunikativ relevanten Bestandteile von Pausen in spontansprachlichen Äußerungen zu isolieren bzw. eine (sprecherabhängige) Standardverteilung kommunikativer Pausen zu definieren. In dem Modell werden die Vorhersagen anhand eines syntaktischen Komplexitätsindex berechnet, der auch von intonatorischen Charakteristika der umgebenden Phrasen beeinflußt wird. Die Modellierung sagt (vereinfacht) aus, daß Pausen umso länger werden, je komplexer die angrenzenden Phrasen sind, wobei eine Balancierung der Phrasenlängen erfolgt. Eine genauere Darstellung dieses Modells erfolgt im Kapitel 5.

Unterstützung für das Modell von Gee und Grosjean liefern auch verschiedene andere Arbeiten: Deese (1980) ließ eine Stichprobe von Pausen durch Versuchspersonen nach ihrer subjektiv wahrgenommenen Länge klassifizieren. Dabei stellte sich heraus, daß die durchschnittliche Dauer der den Kategorien *kurz*, *mittel* und *lang* zugeordneten Pausen für Pausen an Satzgrenzen höher ist, als die durchschnittliche Dauer der Pausen in den gleichen Kategorien für solche, die innerhalb von Sätzen auftraten. Die Toleranzgrenze, oberhalb derer Pausen von Hörern als lang empfunden werden, liegt daher offenbar zwischen Sätzen viel höher; Pausen werden dort mit einer bestimmten Mindestlänge bereits erwartet. Entsprechend müssen Inferenzen über kognitive Belastung bei einer gleich langen Pause zwischen Sätzen und innerhalb von Phrasen zu unterschiedlichen Ergebnissen führen.[26]

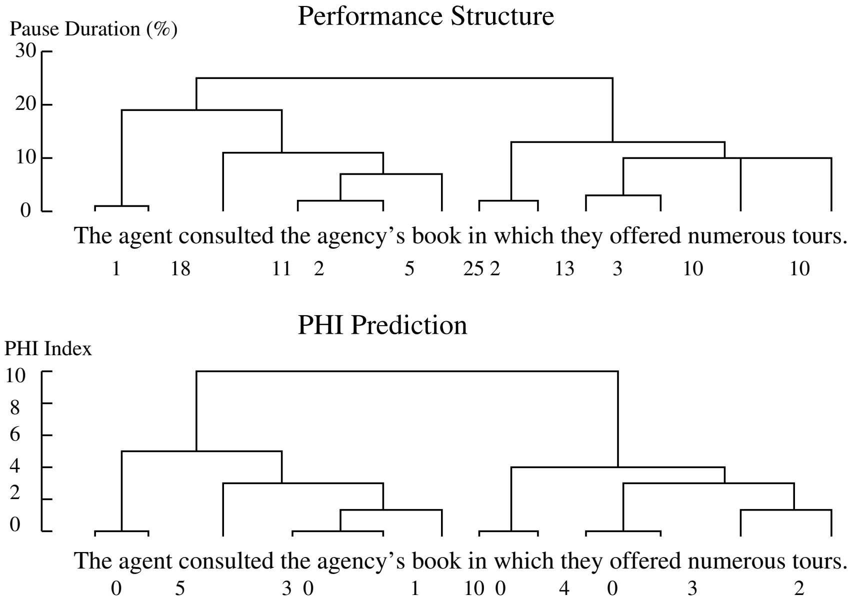


Abbildung 3.6: Beispielvorhersage für die Pausendauer (nach Gee und Grosjean 1983): oben das Pausenmuster aus einem Leseexperiment, unten die Vorhersage des Algorithmus

Ähnliche Effekte beschreibt Butcher (1980): In einer seiner Untersuchungen nahmen die Versuchspersonen in 75 % der Fälle Pausen innerhalb von Phrasen bereits ab 200 ms Länge bewußt wahr. Diese Grenze stieg auf 500 ms Dauer, wenn die Pausen an Satzgrenzen auftraten. Butcher nennt dieses Phänomen '*what is expected is not noticed*' . Pausen, die erwartet, also auch geplant werden, sind daher weniger nützlich für die Einschätzung kognitiver Belastung: Erwartungen beschreiben den Normalfall, aus dem man nichts Außergewöhnliches schließen kann. Auch diese Interpretation verstärkt die Bedeutung intraphrasaler Pausen bei der Untersuchung kognitiver Belastung.

Levelt (1989) weist darauf hin, daß die Makroplanung nachfolgender Äußerungen vor allem an Phrasengrenzen durchgeführt wird. Die an diesen Positionen ohnehin eingesetzten kommunikativen Pausen stellen bereits zusätzliche Planungszeit zur Verfügung (Deese, 1980). Reicht diese Zeit nicht aus, um die Planung für die weitere Artikulation ausreichend fortzusetzen, müssen die ohnehin vorhandenen kommunikativen Pausen verlängert werden. Eine Verlängerung von Pausen an Phrasen- und Satzgrenzen gegenüber ihren positionsbedingten Vorhersagewerten ist daher mit einer starken Relevanz für die Veränderung der Makroplanung durch

[26]Ein Nebeneffekt dieser Beobachtung ist, daß ein Sprecher innerhalb von Phrasen weniger Zeit für stille Pausen hat, wenn er keine Unterbrechung durch den Dialogpartner herbeiführen möchte.

kognitive Belastung behaftet. Auch Grosjean und Deschamps (1973) weisen nach, daß unter Belastung längere (jedoch nicht mehr) Pausen an Phrasenenden enstehen.

**Einschränkungen** Die in den vorangegangenen Abschnitten angegebenen Kriterien sind nicht uneingeschränkt anwendbar. Zusätzliche Einflußfaktoren können nach Butterworth (1980a) auf die Dauer und Verteilung von Pausen wirken:

- Die Länge von Pausen kann innerhalb einer Äußerung durch Eigenschaften des nachfolgenden Lemmas beeinflußt werden:
  - semantische Markiertheit (Anzahl der Präsuppositionen, die die Verwendung des Lemmas im aktuellen Kontext trägt),
  - Häufigkeit seiner Benutzung im allgemeinen Sprachgebrauch: Pausen treten beim spontanen Sprechen häufiger vor Inhaltswörtern (Nomen, Adjektive, Verben) als vor Funktionswörtern (Präpositionen, Artikel, Konjunktionen) auf,[27]
  - Vorhersagbarkeit des Folgewortes im Kontext: Je mehr Möglichkeiten im begonnenen Kontext für die Fortsetzung bestehen, desto häufiger und desto längere Sprechpausen entstehen vor diesen Worten.

  Die Ursachen dieser Effekte sind in Levelts Modell in der Phase des lexikalischen Zugriffs zu vermuten. Eine Erklärung ist jedoch schwierig: Levelt favorisiert keine der angegebenen Theorien des lexikalischen Zugriffs, da alle in irgendeinem Aspekt problematisch sind. Daher besteht derzeit wenig Klarheit über die Vorgänge des Lemma-Abrufs. Die Formulierungsphase ist jedoch weitgehend automatisiert, weshalb eine Veränderung der Abrufgeschwindigkeit unter Arbeitsgedächtnisbelastung weitgehend auszuschließen ist. Die genannten Kontextfaktoren dürfen daher als stabil angesehen und vernachlässigt werden.

- Bisher wurden Aspekte der Makro- und Mikroplanung nicht explizit unterschieden. Diese globale Sicht auf die Äußerungsplanung wird auch in vielen psychologischen Arbeiten eingenommen. Die beiden Formen der Planung erfordern jedoch ggf. kognitive Kapazität in unterschiedlichem Maße und werden daher auch von der Verfügbarkeit solcher Ressourcen ggf. unterschiedlich beeinflußt. Für die Modellierung kognitiver Ressourcen sollen die beiden Planungsprozesse aus den folgenden Gründen vorläufig nicht unterschieden werden:
  - Verantwortlich für eine unterschiedliche Beeinflussung der Mikro- und Makroplanung sind vor allem verschiedene Charakteristika von Aufgabentypen (Strube & Pentenrieder, 1987). Eine Berücksichtigung des Aufgabentyps wäre aber nur dann sinnvoll, wenn dessen Eigenschaften bezüglich kognitiver Verarbeitung geklärt sind und die ausgeführte Aufgabe bekannt ist. Bisher bestehen in der psychologischen Forschung lediglich Ansätze für eine derartige Klassifikation, die nicht allgemeingültig einsetzbar sind.

[27] Die Lokalisation vor Funktionswörtern entspricht im Deutschen und Englischen einer Lage an Phrasengrenzen, da Funktionswörter meist phraseneinleitend plaziert werden.

| | Satzgrenze | | Phrasengrenze | | phrasenintern | |
|---|---|---|---|---|---|---|
| | lang | kurz | lang | kurz | lang | kurz |
| stille | + | – | ++ | + | +++ | ++ |
| gefüllt | +++ | ++ | ++++ | +++ | +++++ | ++++ |

Tabelle 3.7: Pausenphänomene und vermutliche Realisierung der Verzögerungsfunktion

  – Die Zuordnung von verschiedenen Pausenrealisierungen zu Makro- und Mikroplanung ist, von den Pausen an Phrasenenden abgesehen, ungeklärt und erlaubt eine solche getrennte Analyse nicht.

- Pausen bei der Sprachproduktion können nicht nur von direkten Beeinflussungen der Konzeptualisierung herrühren, sondern auch durch die Monitoring-Komponente des Sprachproduktionssystems verursacht werden: Werden bereits geplante Äußerungsbestandteile als inadäquat oder fehlerhaft eingeschätzt, können sie in die entsprechenden Planungsphasen zurückverwiesen werden, bevor eine Artikulation erfolgt. Da die 3 Phasen *Planung - Rückweisung - erneute Planung* sequentiell aufeinanderfolgen, wird die Weitergabe von Eingabeinkrementen an Nachfolgephasen (Formulierung, Artikulation) verzögert. Die Funktionsfähigkeit des Monitoring-Prozesses wird ebenfalls von kognitiver Belastung beeinflußt (siehe Abschnitt 3.6). Demzufolge kann verändertes Monitoring ebenfalls die Veränderung von Pausenverteilungen bewirken.

#### 3.2.2.6 Fazit: Zusammenfassung kognitiv relevanter Pausenphänomene

Zusammenfassend ergibt sich, daß folgende *Veränderungen* des Pausenverhaltens für die Bewertung des kognitiven Planungsaufwands im Kontext eines Dialogsystems relevant sind:

- die Verlängerung stiller Pausen in allen syntaktischen Positionen gegenüber kommunikativ motivierten Vorhersagewerten für Bedingungen ohne Belastung,
- das Auftreten von stillen Pausen an Positionen, für die keine Pausen vorhergesagt werden, insbesondere im Inneren von Phrasen und
- das Auftreten gefüllter Pausen.

Tabelle 3.7 gibt einen vereinfachten Überblick über die Relevanz unterschiedlicher Pausenphänomene für die Modellierung kognitiver Belastung.

#### 3.2.2.7 Pausen und Zeitdruck

Erstaunlicherweise liegen kaum Arbeiten zum Zusammenhang von Zeitdruck mit dem Pausenverhalten vor. Baber und Mellor (1996) untersuchen im Zusammenhang mit Problemen bei der

Spracherkennung die Wirkung von Zeitdruck auf einige Sprechparameter. Leider hat die Untersuchung nur eine quasi-experimentelle Bedeutung. Es sind lediglich Daten für zwei Versuchspersonen angegeben, die außerdem stark unterschiedliche Reaktionen auf Zeitdruck zeigen. Baber und Mellor identifizierten vor allem eine Reduktion der Vokallänge und des Raumes zwischen zwei Worten gegenüber normalen Sprechbedingungen als Ursachen für eine verschlechterte Spracherkennung. Untersuchungen zu Effekten von Makro-Zeitdruck auf das Pausenverhalten liegen dem Autor nicht vor. Auch Ergebnisse aus Experimenten zum schnellen Lesen wie Breznitz (1989) behandeln lediglich Mikro-Zeitdruck und sind daher nicht verwendbar.

## 3.2.3 Silbendehnungen

### 3.2.3.1 Abgrenzung: Silbendehnungen

Statt eine Unterbrechung der Sprachproduktion durch eine Pause herbeizuführen, können lokale Verzögerungen der Sprachproduktion in der Artikulationsphase schon durch die Dehnung wortfinaler Silben erfolgen.[28] Gedehnte Silben sind Silben, deren Dauer über das unter Berücksichtigung der aktuellen Artikulationsgeschwindigkeit notwendige Maß gedehnt werden. Die notwendige Silbendauer ergibt sich aus den phonemischen Eigenschaften der einzelnen Silbenbestandteile, die aufgrund von Bedeutungsdifferenzierungen für ein spezielles Lemma lexikalisch vorgegeben sind.

### 3.2.3.2 Silbendehnungen und kognitive Belastung

Ein Zusammenhang dieses Phänomens mit kognitiver Belastung wurde von Grosjean und Deschamps (1973) für die Bedingung 'Beschreibung' nachgewiesen: Silbendehnungen und gefüllte Pausen teilten fast gleichmäßig unter sich die Anzahl der *gefüllten* Pausen auf, wobei Silbendehnungen einen Anteil von 47% stellen. Dieser Wert ist möglicherweise für die untersuchte Sprache (Französisch) spezifisch - er drückt jedoch aus, welche wichtige Bedeutung Silbendehnungen bei der Verzögerung der Sprachproduktion haben können. In Kapitel 4 wird diese Feststellung durch die Auswertung der Voruntersuchung zusätzlich illustriert werden.

Auch Deese (1980) macht die Parallele zwischen Silbendehnungen und Sprechpausen deutlich: Seine Versuchspersonen sollten Pausen in Äußerungen nach der subjektiv empfundenen Länge bewerten. Die Probanden klassifizierten häufig auch Äußerungsabschnitte als Pausen, in denen keine Stilleperiode auftrat, sondern lediglich eine Silbendehnung. Hörer können daher offenbar von der Art der Realisierung eines Verzögerungseffektes abstrahieren und nehmen nur die Verzögerung selbst wahr, d.h. die Funktion statt der Form. Der Effekt in Deeses Untersuchung beschränkte sich allerdings auf solche Äußerungspositionen, an denen eine Verzögerung nicht erwartet wird (speziell innerhalb von Phrasen). Obwohl nicht alle überlangen Silben entsprechend bewertet werden können, deutet sich an, daß gedehnte Silben der Klasse der Pausenphänomene zugeordnet werden müssen.

Die Klasse der Worte, die Silbendehnungen unterzogen wurden, ist nach der Analyse von Grosjean und Deschamps relativ abgeschlossen: 94 % der Silbendehnungen in ihren Beschrei-

[28] Wortinterne Silben werden normalerweise nicht selbständig gedehnt, sondern höchstens zusammen mit wortfinalen Silben. Dies kann aber auch als eine Verringerung der Artikulationsgeschwindigkeit interpretiert werden.

bungen (89 % in den Interviews) entfielen auf Funktionsworte, wobei die einsilbigen wiederum besonders bevorzugt werden. Ähnliche Dehnungsverteilungen beobachtet auch Butterworth (1980b). Aus der eigenen empirischen Untersuchung (Abschnitt 4.3) wird deutlich, daß Silbendehnungen häufig mit Pausen kombiniert werden und insbesondere im Zusammenhang mit Wiederholungen und Fehlansätzen auftreten (siehe auch Grosjean & Deschamps, 1973).

#### 3.2.3.3 Erfassung von Silbendehnungen

Die Länge einer Silbe ist abhängig von ihrem phonemischen Aufbau und ihrer Position innerhalb der morphologischen Struktur eines Wortes: Für die Messung der Dehnung einer Silbe ist daher ein Vergleich mit einem an der entsprechenden Wortposition erwarteten Basismaß notwendig. Für dieses Basismaß sind außerdem sprecherspezifische Charakteristika sowie prosodische und syntaktische Constraints verantwortlich, deren Untersuchung für diese Arbeit nicht vorgesehen ist.

Da gedehnte Silben einerseits die Artikulationsgeschwindigkeit herabsetzen, andererseits aber auch eine Füllfunktion für Pausen der Planung realisieren, ist eine ausschließliche Zuordnung zu einem der beiden Phänomene nicht möglich. Da die genaue Verlängerung einer Silbe gegenüber einem Basismaß aus den soeben erläuterten Gründen für die vorliegende Arbeit nicht berechnet werden kann, soll auch im Hinblick auf die Modellierung in Kapitel 6 eine alternative Vorgehensweise vorgeschlagen werden: Das Häufigkeitsmaß (relative Anzahl), ist vom Ausmaß der Verlängerung unabhängig und kann daher direkt erhoben werden. Wegen der den gefüllten Pausen ähnlichen Funktion von Silbendehnungen kann ihre Anzahl zur Zahl der gefüllten Pausen hinzusummiert werden und erhöht damit auch den Häufigkeitsindex kognitiver Pausen. Der Einfluß unterschiedlicher Dauer von Silbendehnungen kann (mit einem vergleichsweise geringen Informationsverlust) mit dem Maß der Artikulationsgeschwindigkeit erfaßt werden: Die Verringerung der Artikulationsgeschwindigkeit wurde bereits als Indikator kognitiver Belastung identifiziert. Da Silbendehnungen ebenfalls verringernd auf die Artikulationsgeschwindigkeit wirken, stellt dieses Vorgehen eine angemessene Vereinfachung dar.

### 3.2.4 Sprechgeschwindigkeit

#### 3.2.4.1 Abgrenzung des Phänomens: Sprechgeschwindigkeit

Fast alle psychologischen Arbeiten, in denen Sprechparameter untersucht werden, machen Aussagen über die beobachtete Sprechgeschwindigkeit der Versuchspersonen (Baber & Mellor, 1996; Grosjean & Deschamps, 1973; Bromme & Wehner, 1987; Marx, 1984; Wiese, 1983). Grund dafür ist offenbar, daß die Sprechgeschwindigkeit ein integratives Maß über die verbale Produktivität innerhalb einer bestimmten Zeitspanne darzustellen scheint. Im Gegensatz zur Artikulationsgeschwindigkeit verwendet die Sprechgeschwindigkeit $G_S$ jedoch die Gesamtsprechdauer (GSD) einer Äußerung als Basisgröße, ohne Teilsequenzen auszuschließen (siehe Def. 3.4). Die Sprechgeschwindigkeit wird wie die Artikulationsgeschwindigkeit in Silben/s angegeben.

(3.4) Die **Sprechgeschwindigkeit** $G_S$ eines Sprechers berechnet sich als Quotient aus der

Gesamtdauer einer Äußerung und der Anzahl der in dieser Zeit artikulierten sprachlichen Einheiten.

$$G_S = \frac{\text{Anzahl artikulierter sprachlicher Einheiten}}{\text{Gesamtsprechdauer}}$$

(3.5) **Me010, Dialog 3:**
<**äh**> <**P**> Im <**D**> <**P**> Innenraum unter'm Lenkrad <**P**> links <**P**> ist ein Sicherungskasten.

*Gesamtsprechdauer*

Die Gesamtsprechdauer von Beispiel 3.5 beträgt 5.387 s, 17 Silben wurden artikuliert. Die Sprechgeschwindigkeit beträgt daher 3.156 Silben/s. Von der Gesamtsprechdauer entfallen jedoch 2.187 s auf die (fettgedruckten) Verzögerungsphänomene. (Siehe zum Vergleich Beispiel 3.2.)

#### 3.2.4.2 Sprechgeschwindigkeit und Zeitdruck

Von verschiedenen Versuchspersonen in einer Studie von Weyrath (1998) wurde die Beschleunigung der Sprechgeschwindigkeit als Symptom für Zeitdruck gewertet. Dieses Ergebnis entspricht zum Teil empirischen Befunden (Marx, 1984), die in dialogischen Situationen einen umgekehrt u-förmige Zusammenhang zwischen subjektiv empfundenen Zeitdruck und Sprechgeschwindigkeit feststellte: Ihre Gruppe mit niedrigem Zeitdruck hatte eine signifikant höhere Sprechgeschwindigkeit als die Kontrollgruppe; für die Gruppe mit hohem Zeitdruck war dieser Unterschied jedoch nicht signifikant. Nach Marx' Erkenntnissen darf daher gerade bei hohem Zeitdruck keine Veränderung der Sprechgeschwindigkeit erwartet werden. Problematisch an der Untersuchung von Marx ist jedoch u.a., daß die Variable 'Zeitdruck' zwischen den beiden Zeitdruckgruppen nicht experimentell kontrolliert wurde, sondern ihr Ausmaß lediglich auf einer postexperimentellen Selbsteinschätzung der Versuchspersonen beruht. Aufgrund subjektiver Unterschiede in der Interpretation des Terminus *hoher Zeitdruck* ist die Aussagekraft der Untersuchungsergebnisse von Marx zweifelhaft.

#### 3.2.4.3 Sprechgeschwindigkeit und kognitive Belastung

Zahlreiche Arbeiten stellen einen Rückgang der Sprechgeschwindigkeit unter kognitiver Belastung fest. Tabelle 3.8 gibt einen Überblick. Die Sprechgeschwindigkeit weist jedoch eine Anzahl unangenehmer Eigenschaften auf, die im folgenden besprochen werden.

#### 3.2.4.4 Problematische Eigenschaften der Sprechgeschwindigkeit

Beim Literaturstudium fiel auf, daß ein terminologisches Problem der Unterscheidung von Artikulations- und Sprechgeschwindigkeit im alltäglichen Sprachgebrauch besteht. Die Versuchspersonen in der Studie von Weyrath (1998), in der eine erhöhte Sprechgeschwindigkeit als Symptom für Zeitdruck gewertet wurde, könnten ebenfalls Opfer dieses Problems geworden sein: Als Einschätzung für die Sprecheräußerungen wurde jeweils angegeben 'spricht sehr schnell'. Damit kann jedoch auch eine schnelle Artikulation gemeint sein.

Die Tendenz der Untersuchungsergebnisse in Tabelle 3.8 zeigt deutlich in Richtung einer verringerten Sprechgeschwindigkeit unter Belastung. Durch ihre Berechnung sind im Maß der

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Bromme und Wehner (1987)$^{=}$ | Rein sprachliche Aufgaben mit unterschiedlicher außersprachlicher Planung | 12 | Zusammenstellung einer Speisekarte (Aufgabe) | Denksportaufgabe (Problem) | 2.16 Silben/s[a], $(SD = 0.82)$ | 1.96 Silben/s[a], $(SD = 0.94)$ | $p = 0.05$, $(t = 2.2, df = 11)$ | – |
| Bromme und Wehner (1987), Vorexperiment | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Interview, Erzählung | 12 | Interview | Nacherzählung | 3.52 Silben/s[b], $(SD = 0.82)$ | 2.72 Silben/s[b], $(SD = 0.8)$ | k.A. | – |
| Duez (1982)$^{=}$ | Quasiexperimentell: sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorbereitete Interviews | Unerwartete Interviews | 4.26 Silben/s | 4.06 Silben/s | k.A. | (–) |
| Grosjean und Deschamps (1973)$^{=}$ | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Interviews und Beschreibungen | 30/20[c] | Radio-Interviews | Cartoon-Beschreibungen | 4.40 Silben/s[d], $(SD = 0.50)$ | 2.55 Silben/s[d], $(SD = 0.41)$ | $p < 0.001$, $(z = 5.92)$ | – |
| Kowal und O'Connell (1987)$^{=}$ | Sprachlich und sprachlich-visuell: Wiedergeben des Inhalts eines Films | 39 | Wiedergabe des Inhalts nach Ende des Films | Wiedergabe des Inhalts während des Films | 3.62 Silben/s, $(SD = 0.46)$ | 1.78 Silben/s, $(SD = 0.57)$ | sign. (k.A.) | – |
| Kowal et al. (1983)$^{=}$ | Zusammenfassung experimenteller Arbeiten | – | Arbeiten zu Interviews (3) | Arbeiten zum Nacherzählen (11) | 4.31 Silben/s, $(SD = 0.1)$ | 3.43 Silben/s, $(SD = 0.43)$ | – | (–) |
| Wiese (1983)$^{=}$ | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Sprechen in der Muttersprache | Sprechen in einer Fremdsprache | 3.25 Silben/s, $(SD = 0.51)$ | 1.59 Silben/s, $(SD = 0.46)$ | k.A. | (–) |

Tabelle 3.8: Untersuchungen zur Sprechgeschwindigkeit

[a] Originaldaten: 10.8 Silben/ 5 s (unbelastet) bzw. 9.8 Silben/ 5 s (belastet)
[b] Originaldaten: 17.6 Silben/ 5 s (unbelastet) bzw. 13.6 Silben/ 5 s (belastet)
[c] Kontrollgruppe/Experimentalgruppe
[d] Die Originaldaten sind in Silben/min angegeben.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Zeitdruck** | | | | | | | | |
| Kelly und Stone (1982) | Inteviews mit Typ-A- und Typ-B- Personen über situationsspezifisches soziales 'Wohlbefinden' | 24 | Interview beliebig lang | Interview unter Zeitbeschränkung (10 min) | 2.72 Worte/s, $(SD = 0.36)$(Typ A); 2.04 Worte/s , $(SD = 0.46)$(Typ B) | 2.96 Worte/s, $(SD = 0.40)$(Typ A); 2.65 Worte/s , $(SD = 0.19)$(Typ B) | $p < 0.01$, $(F(1,20) = 8.25)$ | + |
| Marx (1984) | Führen eines Gespräches mit Ziel der Überzeugung eines Partners | 70 | Ohne Zeitdruck | Zeitdruck durch Wiederholung einer bereits ohne Zeitdruck ausgeführten Tätigkeit mit Zeitbegrenzung | 2.94 Silben/s, $(SD = 1.35)$ | 3.42 Silben/s, $(SD = 2.16)$; [2.87 Silben/s, $(SD = 0.68)$(NZD)[a]; 3.70 Silben/s, $(SD = 3.15)$(HZD)] | n.s. | (+) |
| Marx (1984) | Lautes Denken über ein Problem | 70 | Ohne Zeitdruck | Zeitdruck durch Wiederholung einer bereits ohne Zeitdruck ausgeführten Tätigkeit mit Zeitbegrenzung | 2.36 Silben/s, $(SD = 0.84)$ | 2.87 Silben/s, $(SD = 0.86)$; [3.14 Silben/s, $(SD = 0.66)$(NZD)[a]; 2.72 Silben/s, $(SD = 0.92)$] | $p = 0.019$, $(t = 2.41)$(HZD); [$p = 0.006$, $(t = 2.88)$(NZD); n.s., $(t = 1.55)$(HZD)] | + (u)[b] |

Tabelle 3.9: Untersuchungen zur Sprechgeschwindigkeit (Fortsetzung)

[a]Postexperimentell erfolgte eine Zuordnung der Probanden zu den Gruppen 'hohes Zeitdruckempfinden' (HZD) bzw. 'niedriges Zeitdruckempfinden' (NZD) durch Selbsteinschätzung.

[b]Umgekehrt u-förmiger Zusammenhang: höher unter niedrigem Zeitdruckempfinden, gleich unter hohem Zeitdruckempfinden

Sprechgeschwindigkeit jedoch stille Pausen verschiedenster Funktion (auch rein kommunikativ motivierte oder Pausen aus Nebenhandlungen), gefüllte Pausen, sowie gedehnte Silben und die Artikulationsgeschwindigkeit einbezogen. Kognitiv bedingte Ursachen für die Veränderung dieser Parameter wirken sich gemeinsam mit nicht-kognitiven Ursachen gleichberechtigt und undifferenziert auf die Sprechgeschwindigkeit aus. Zumindest in den Arbeiten aus Tabelle 3.8, die die genannten störenden Parameter ebenfalls erheben, können Effekte der Reduktion der Sprechgeschwindigkeit unter kognitiver Belastung oder Zeitdruck durch die Veränderung eines oder mehrerer ihrer Basisparameter (Artikulationsgeschwindigkeit, Einsatzlatenz und Pausendauer) erklärt werden (Marx, 1984; Wiese, 1983; Levelt, 1989; Grosjean & Deschamps, 1973). In der Untersuchung von Kelly und Stone (1982) ergab es sich sogar, daß für eine Versuchsgruppe eine *Erhöhung* der Sprechgeschwindigkeit mit einer *Verringerung* des Anteils an Verzögerungsphänomenen an der Gesamtsprechzeit einhergeht.[29] Bei der Berechnung der Sprechgeschwindigkeit könnten sich auch komplementäre Veränderungen von Einzelparametern gegenseitig aufheben und dadurch unveränderte Meßergebnisse erzeugen. Diese würden suggerieren, daß keine Zustandsänderung stattgefunden hat.

#### 3.2.4.5 Fazit

Obwohl z.T. recht deutliche Nachweise über die Veränderung der Sprechgeschwindigkeit unter Zeitdruck und kognitiver Belastung geführt worden sind, ist die Sprechgeschwindigkeit *kein Basisparameter sprachlicher Äußerungen*. Unterschiedliche Einflüsse führen undifferenziert zu Veränderungen des Parameters. Aus diesen Gründen sollte bei der Modellierung kognitiver Belastung die Sprechgeschwindigkeit nicht berücksichtigt, sondern einer detaillierten Analyse ihrer Einzelparameter der Vorzug gegeben werden.

### 3.2.5 Kognitive Rhythmen bei der Sprachproduktion

**Abgrenzung des Phänomens: Kognitive Rhythmen**

Sprachliche Planung gliedert sich in eine Planung auf Makroniveau und Mikroniveau (siehe Abschnitt 2.4.3). Diese Planungsprozesse laufen nicht gleichberechtigt nebeneinander ab: in verschiedenen Phasen eines Dialoges oder Monologes ist eine größere Anzahl an Makroplanungsschritten zur Auswahl der Information für die nächsten Äußerungsschritte notwendig. Da während der Makroplanungszeit zuvor geplante Äußerungsbestandteile meist auch formuliert und artikuliert werden und dafür eine Fortführung der Mikroplanung aufrechterhalten werden muß, bleibt diese implizite Doppelaufgabe nicht ohne Einfluß auf diese beiden Phasen der Sprachproduktion. Henderson, Goldman-Eisler und Skarbek (1966) registrierten erstmals, daß in Phasen der Makroplanung Sprecher die Geschwindigkeit in der Artikulationsphase reduzieren. Aus einer höheren Anzahl stiller Pausen und geringerer Artikulationsgeschwindigkeit resultiert eine reduzierte Sprechgeschwindigkeit (siehe auch Goldman-Eisler, 1968; Beattie, 1983).

[29] Es handelt sich um die Gruppe der Typ-A-Personen. Werte für die Artikulationsgeschwindigkeit sind nicht angegeben.

Phasen verstärkter Makroplanung sind nach Henderson et al. (1966) in der Länge offenbar relativ konstant. Trägt man die auflaufenden Artikulations- und Pausendauern einer Äußerung auf den Achsen eines Koordinatensystems ab, so ergeben sich durch die periodische Wiederkehr der Makroplanung typische Graphen, von Levelt (1989) als 'Henderson-Graphen' bezeichnet (Abb. 3.7). Eine Verlängerung des Graphen parallel zur y-Achse erfolgte, wenn der Sprecher seine Artikulation vorübergehend unterbrach, d.h. stille Pausen einsetzte. Auf der x-Achse werden die Einzeldauern der Artikulationssegmente abgetragen. Eine treppenförmige Bewegung des Graphen um eine vom Koordinatenursprung ausgehende *Gerade* stünde für eine konstante Sprechgeschwindigkeit.[30] Da die echte Sprechgeschwindigkeit sich beim spontanen Sprechen jedoch um diese Gerade in Form sinusförmiger Schwingungen bewegt und dadurch direkt den Ablauf kognitiver Operationen anzuzeigen scheint, werden derartige Muster als **kognitive Rhythmen** bezeichnet.

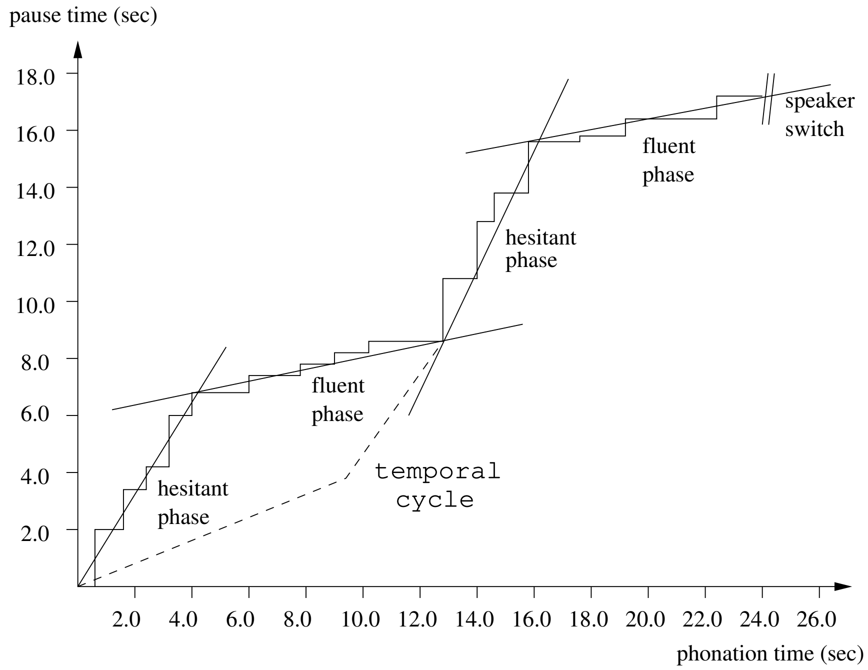


Abbildung 3.7: Henderson-Graph: flüssige und verzögerte Phasen alternieren bei der spontanen Sprachproduktion (nach Beattie, 1983)

#### 3.2.5.1 Kognitive Rhythmen und Arbeitsgedächtnisbelastung

In Abhängigkeit vom Planungsaufwand varriiert die Schwingungslänge und die Amplitude der kognitiven Rhythmen möglicherweise erheblich. Können die Schwingungsparameter für eine vergleichbare sprachliche Aufgabe als Kontrollbedingungen festgelegt werden, so könnte eine Abweichung der Sprachproduktion vom Kontrollmuster ein weiterer möglicher Indikator für kognitive Belastung sein: Wenn die Sprechgeschwindigkeit sich längere Zeit auf einem niedrigen Niveau bewegt, deutet das an, daß u.a. möglicherweise mehr Verzögerungselemente für die

[30] Ein solcher Kurvenverlauf wird von Goldman-Eisler (1968, S.82) z.B. beim lauten Lesen identifiziert.

Artikulation angewendet werden. Das kann auf eine zeitliche Beeinflussung der Planungsphase (quantitative Veränderung der inkrementellen Verarbeitung) zurückgeführt werden. Ein solcher Effekt ist mit einem gestörten Performanzprofil der Konzeptualisierung erklärbar und zwar sowohl durch erhöhte Planungsanforderungen (etwa durch ein geringeres Wissensniveau) als auch durch eine Geschwindigkeitsreduktion der Planung wegen nicht verfügbarer Ressourcen.

#### 3.2.5.2 Gründe für den Ausschluß kognitiver Rhythmen

Drei Gründe sind verantwortlich dafür, daß die Verwendung kognitiver Rhythmen als Indikator kognitiver Belastung derzeit nicht empfohlen werden kann:

- Der Zusammenhang von kognitiven Rhythmen mit kognitiver Beanspruchung ist nur wenig empirisch untersucht.
- Die Existenz kognitiver Rhythmen ist (sicherlich auch aufgrund der spärlichen Datenlage) in der Literatur z.T. heftig umstritten (siehe die Diskussion in Beattie, 1983; Power, 1983; Beattie, 1984; Power, 1984) und kann daher nicht als endgültig abgesichert angesehen werden.
- Die Beobachtbarkeit von Rhythmen bedingt eine bestimmte Mindestlänge sprachlicher Äußerungen: Wie aus der Grafik 3.7 ersichtlich ist, beträgt die durchschnittliche 'Schwingungsdauer' einer Phase mehr als 10 s. Um eine Stabilität der Beobachtung sicherzustellen, ist evtl. sogar ein Vielfaches dieser Dauer für *eine* Äußerung Voraussetzung.[31]

Auch Levelt (1989, Kap.4) warnt mit einer eher vorsichtigen Einschätzung des Phänomens vor Verallgemeinerungen. Zusätzliche Argumente finden sich bei Kowal (1991), die daher ebenfalls zu einer kritischen Einschätzung des Konzeptes der kognitiven Rhythmen kommt. Bromme und Wehner (1987) haben Schwierigkeiten, in ihrem Korpus Rhythmen überhaupt zu identifizieren.

### 3.2.6 Einsatzlatenz

#### 3.2.6.1 Abgrenzung des Phänomens: Einsatzlatenz

Vor Beginn eines Dialogturns haben Formulierungsphase und Artikulationsphase noch keine Eingabe abzuarbeiten. Zuerst ist ein zeitlicher Vorlauf der Planungsphase notwendig, um erste Eingabeinkremente für die nachgeordneten Phasen der Sprachproduktion zu generieren. Die Artikulation setzt für diese Planungseinheiten wegen der zeitlich versetzten Verarbeitung erst zu einem späteren Zeitpunkt ein. Man bezeichnet diesen Vorlauf der Planungsphase daher als *Einsatzlatenz.*

Die Einsatzlatenz in bezug auf eine Partneräußerung ist mit der Reaktionszeit in Experimenten zur Perzeption vergleichbar. Die beiden Größen können jedoch nicht gleichgesetzt werden: Zwar besteht für einen Hörer nach der Beendigung der Partneräußerung ein kommunikativer Zwang, nunmehr seine eigene Äußerung zu beginnen. Andererseits kann er diese Äußerung

[31] Nicht zufällig beziehen sich Beobachtungen kognitiver Rhythmen daher auf monologische Situationen.

nicht 'irgendwie' starten, sondern möchte meist eine adäquate Erwiderung auf die Partneräußerung generieren, wofür sowohl für die Analyse der Partneräußerung als auch für die Generierung der eigenen Äußerung tiefere semantische und pragmatische Verarbeitung notwendig ist. Aus Abbildung 3.8 kann folgende operationale Definition der Einsatzlatenz abgeleitet werden[32]:

(3.6) Die **Einsatzlatenz** $E_P$ eines Sprechers berechnet sich aus der zeitlichen Differenz zwischen dem Ende des Signals der Partneräußerung $t_S$ und dem Zeitpunkt des Einsetzens seiner eigenen Artikulation $t_P$.[33]

$$E_P = t_P - t_S$$

#### 3.2.6.2 Einsatzlatenz und kognitive Belastung

Für die Bestimmung der Einsatzlatenz eines Dialogteilnehmers $P$ muß der entsprechenden Äußerung eine Instruktion oder Partneräußerung vorangehen, auf die ein Sprecher mit seiner Äußerung Bezug nimmt.

Vor Beginn seiner eigenen Äußerungsplanung muß $P$ die Äußerung von $S$ soweit vollständig verarbeitet haben, wie das für den Beginn der Planung seines eigenen Dialogbeitrags notwendig ist. Da auch das Sprachverstehenssystem inkrementell arbeitet, wird nicht erst nach Beendigung einer Äußerung durch den Sprecher $S$ mit der Verarbeitung bei $P$ begonnen, sondern Artikulation von $S$ und Verarbeitung bei $P$ laufen weitgehend parallel ab. Mit welcher Qualität und Geschwindigkeit die Verarbeitung jedoch stattfinden kann und wie weit sie zum Zeitpunkt $t_S$ unvollständig ist, ist unter anderem abhängig von der syntaktischen Komplexität der Partneräußerung , der Schwierigkeit ihrer Interpretation im Diskurskontext und der Menge an Inferenzen, die ggf. über ihrer Semantik auszuführen sind. Insofern spielt auch die Unterstützung der Verarbeitung bei $P$ durch eine geeignete Partnerorientierung von $S$ eine Rolle (siehe Abschnitt 3.6).

Neben diesen auf die Äußerung von $P$ bezogenen Faktoren der Einsatzlatenz tritt die davon unabhängige zusätzliche kognitive Belastung des Hörers selbst (z.B. durch Doppelaufgaben), welche seinen Sprachverstehensprozeß und den Beginn oder die Dauer seines Planungsprozesses beeinflussen können (Lazarus-Mainka & Arnold, 1987).

Psychologische Arbeiten, die einen Effekt kognitiver Belastung auf die Einsatzlatenz nachweisen, sind in Tabelle 3.10 aufgeführt.[34] Bei kognitiver Belastung ergab sich in fast allen Arbeiten eine signifikante Verlängerung der Einsatzlatenz gegenüber Kontrollbedingungen. Als Erklärung für die Ergebnisse von Power (1985) ergibt sich eine zusätzliche Einflußgröße bei der Modellierung kognitiver Belastung.

[32] Im folgenden werden für die beiden Sprecher die Kürzel $P$ und $S$ verwendet. $S$ steht für *System*, $P$ für (Dialog)*Partner*. Die Verwendung entspricht auch der im Projekt READY.

[33] Bei Überlappung entstehen negative Werte. Diese werden hier nicht berücksichtigt, da sie auch in den experimentellen Studien keine Rolle spielten.

[34] Einige Arbeiten geben ihre Ergebnisse in ms an. Der Vergleichbarkeit halber wurden sie in der Tabelle auf s umgerechnet.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Gerrig und Littman (1990), Exp. 2[a] | Entscheidung, ob Bedeutung einer ambigen Partneräußerung einer vorgegebenen Interpretation entspricht (durch Knopfdruck)[b] | 28 | Partneräußerung von Partner mit gemeinsamem Hintergrundwissen | Partneräußerung von Partner ohne gemeinsames Hintergrundwissen | 3.38 s (Gr.1), 3.87 s (Gr.2) | 4.31 s (Gr.1), 4.56 s (Gr.2) | $p < 0.05$, $(F(1,44) = 4.89)$ | + |
| Greene (1984)[=], Exp. 2 | Rein sprachliche Aufgabe mit unterschiedlichem Planungsaufwand: Darlegung eines Problems und seiner Lösungen | 30 | Diskussion eines Themas mit Kenntnis eines 4-Schritte- Konzepts und nach einer (Gr.1) bzw. zwei Übungssitzungen (Gr.2) | Keine vorherige Übung, kein Konzept zur Verfügung | 7.05 s, $(SD = 4.54)$ (Gr.1); 5.63 s, $(SD = 5.26)$ (Gr.2) | 14.65 s, $(SD = 7.095)$ | $p = 0.002$, $(F(1,27) = 12.39)$ | + |
| Greene et al. (1990)[=] | Rein sprachliche Aufgaben mit unterschiedlichen Zielen: Charakterisierung einer Person | 70 | Klare und deutliche Darlegung einer Personenbeschreibung | Zusätzlich Beachtung sozialer Rücksichtnahme | 3.94 s, $(SD = 1.78)$ (Gr.1); 2.74 s, $(SD = 1.24)$ (Gr.2) | 6.76 s, $(SD = 5.78)$ (Gr.1); 4.52 s, $(SD = 2.17)$ (Gr.2) | $p = 0.009$, $(F(1,49) = 7.32)$ | + |
| Horton und Keysar (1996), Exp.-Bed.1[c] | Identifikation von Objekten anhand von Beschreibungen des Sprechers | 24 | Gemeinsamer Kontext für Beschreibungen (d.h. gleiche Kontextobjekte) | Kein gemeinsamer Kontext | 2.392 s (Gr.1), 1.665 s (Gr.2) | 2.935 s (Gr.1), 1.338 s (Gr.2) | $p < 0.04$ (Gr.1), $(F(1,20) = 5.2)$ , k.A. (Gr.2) | + <br> $(-)$ |
| Power (1983), Exp. 1 | Generieren von Sätzen für vorgegebene Nomenpaare | 36 | Generierung von Sätzen (2 Gruppen) | Generierung unter zusätzlichem Memorieren einer Zahlenfolge von 3 bzw. 6 Ziffern (Gr. A und B) | 2.32 s (Gr.1); <br> 2.67 s (Gr.2) | 1.99 s (Gr.1A), 1.87 s (Gr.1B); 2.35 s (Gr.2A), 2.06 s (Gr.2B) | $p < 0.005$, $(F(2,84) = 5.87)$ | – |
| Power (1983), Exp. 1 | Generieren von Sätzen für vorgegebene Nomenpaare | 36 | Generierung für semantisch verwandte Nomenpaare | Generierung für semantisch entfernte Nomenpaare | 2.32 s (Gr.1), 1.99 s (Gr.2), 1.87 s (Gr.3) | 2.67 s (Gr.1), 2.35 s (Gr.2), 2.06 s (Gr.3) | $p < 0.05$, $(F(1,25) = 4.49)$ | + |
| Roßnagel (1995a) | Sprachliche Instruktion von Karatekas durch *geübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | 0.415 s | 0.863 s | $p < 0.02$, $(F(1,39) = 3.89)$ | + |

Tabelle 3.10: Untersuchungen zur Einsatzlatenz

[a] In diesem Experiment wurde nur die *Verarbeitungs*phase der Einsatzlatenz gemessen. Es war keine Sprachgenerierung nötig.

[b] In Gruppe 1 wurde die dominante Interpretation vorgegeben, in Gruppe 2 die nicht dominante Interpretation.

[c] Die Daten von (Horton & Keysar, 1996) sind aufgrund des 2 x 2 Experimentaldesigns (Zeitdruck vs. Belastung) sowohl für Zeitdruck als auch für kognitive Belastung verwendbar (Gr.1 ist die Gruppe ohne Zeitdruck, Gr.2 mit Zeitdruck).

| Autoren | Methode | VPn | Aufgabe | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Roßnagel (1995a)[=] | Sprachliche Instruktion von Karatekas durch *ungeübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | 1.701 s | 2.422 s | $p < 0.001$, $(F(1,39) = 34.32)$ | + |
| Roßnagel (1995b) | Beschreibung eines Maschinenmodells gegenüber 2 versch. Hörergruppen (A= Kind/B= Student) | 45 | Präsenz des Modells während der Beschreibung | Abwesenheit des Modells während der Beschreibung (Gr.1) bzw. Zweitaufgabe Memorieren einer Zahlenfolge (Gr.2) | 1.443 s (Gr. A),<br>1.450 s (Gr. B) | 2.551 s (Gr.1A) bzw. 3.035 s (Gr.2A);<br>2.564 s (Gr.1B)<br>3.026 s (Gr.2B) | $p < 0.01$, $(F(1,33) = 14.97)$ | + |
| Wegner und Erber (1992), Exp. 2[a] | Manuelle Aufgabe mit sprachlichem Stimulusmaterial: Entscheidung, über die Farbe eines visuell präsentierten Wortes mit Variation des Stroop- Effekts[bc] | 46 | Zusätzliches Memorieren einer einstelligen Zahl | Zusätzliches Memorieren einer neunstelligen Zahl | 0.485 s | 0.549 s | $p < 0.03$, $(F(1,20) = 6.28)$ | + |
| **Zeitdruck** | | | | | | | | |
| Horton und Keysar (1996), Exp.-Bed.2[d] | Identifikation von Objekten mit gemeinsamem (Gr.A) bzw. unterschiedlichem Kontextwissen (Gr.B) anhand von Beschreibungen des Partners | 24 | Kein Zeitdruck | Zeitdruck | 2.392 s (Gr.A),<br>2.935 s (Gr.B) | 1.665 s (Gr.A),<br>1.338 s (Gr.B) | k.A. | –<br>– |
| Wegner und Erber (1992), Exp. 1[e] | Nennen von assoziierten Konzepten zu vorgegebenen Worten (permanente Zusatzaufgabe: (Nicht-) Nachdenken über ein Konzept) | 56 | Milder Zeitdruck | Starker Zeitdruck | 3.80 s | 2.37 s | $p < 0.0001$, $(F(1,54) = 31.30)$ | – |

Tabelle 3.11: Untersuchungen zur Einsatzlatenz (Fortsetzung)

[a] Die gemessene Einsatzlatenz betrifft nur die Verarbeitungsphase. Sprachliche Äußerungen waren nicht auszuführen.
[b] Der Stroop- Effekt entsteht durch die Interferenz von Wortfarbe und Wortbedeutung: z.B. wird das Wort 'rot' in blauer Schrift meist als 'blau' erkannt.
[c] Als permanente Zusatzaufgabe war von den Probanden Nachdenken bzw. Nicht-Nachdenken über ein bestimmtes semantisches Konzept gefordert.
[d] Die Daten von (Horton & Keysar, 1996) sind aufgrund des 2 x 2 Experimentaldesigns sowohl für Zeitdruck als auch für kognitive Belastung verwendbar.
[e] Die gemessene Einsatzlatenz betrifft sowohl Verarbeitungs- als auch Generierungsphase.

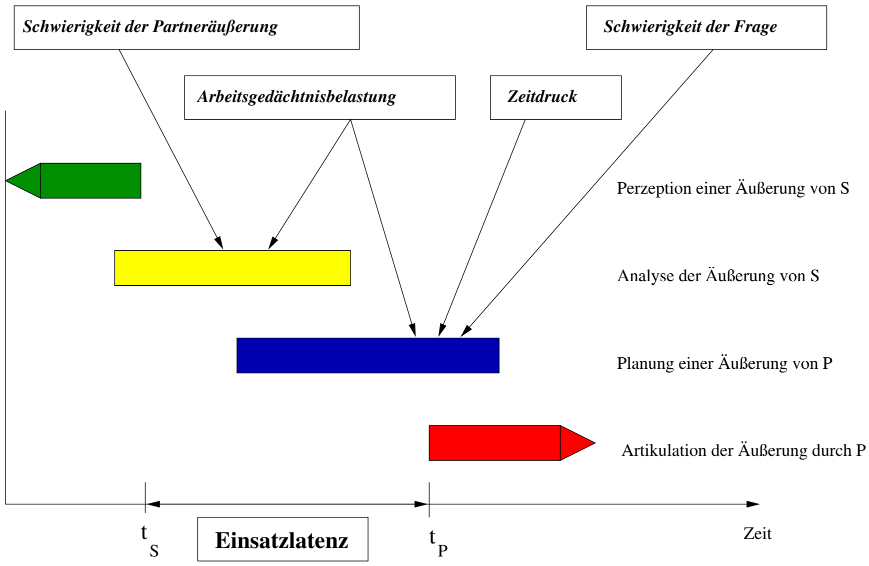


Abbildung 3.8: Zeitliche Organisation und Einflüsse auf die Einsatzlatenz bei einer Abfolge zweier Dialogturns

#### 3.2.6.3 Problem Nr. 3: Strategien der Belastungsreduktion

Die Ergebnisse von Power (1985) zur Einsatzlatenz (Tabelle 3.10) zeigen, daß unter unterschiedlichen Belastungsbedingungen für bestimmte Parameter völlig gegensätzliche Effekte auftreten können. Seine Versuchspersonen hatten so rasch wie möglich Sätze zu vorgegebenen Nomenpaaren zu bilden. Die Experimentalgruppe wurde mit einer zusätzlichen Memorierungsaufgabe unter kognitive Belastung gesetzt. Die kürzere Einsatzlatenz in dieser Gruppe gegenüber der Kontrollgruppe erklärt Power damit, daß die Versuchspersonen stereotyperes Material bei der Generierung der Beispielsätze einsetzten.[35]

Powers Probanden erreichten durch eine Reduktion der konzeptuellen Planung die Reservierung von zusätzlicher Kapazität für die Nebenaufgabe. Unabhängig davon, ob das geschilderte Vorgehen als *Regulierung* der kognitiven Belastung oder als ihre *Folge* gesehen werden muß, deutet die Beobachtung darauf hin, daß Menschen in speziellen Fällen Strategien der Belastungsreduktion anwenden können. Diese Strategien erreichen in Doppelaufgabensituationen durch eine geringere Qualität bei der Ausführung der Nebenaufgabe, daß es in keiner der parallel ausgeführten Aufgaben zu einem Zusammenbruch der Verarbeitung kommt.

Ähnliche Formen aufgaben- und damit auch belastungsabhängiger qualititativer Veränderung der Sprachproduktion untersuchen Rummer (1996a) anhand der Verwendung unterschiedlicher Steuerungsschemata bei der sprachlichen Makroplanung und Strube und Pentenrieder (1987) für den Einsatz unterschiedlicher Verarbeitungsstrategien bei Beurteilungs- und Schätz-

[35]Insbesondere ist hier die Verwendung von prototypischen Relationen für die Nomen gemeint: z.B. wird für das Nomenpaar 'Bauer - Feld' die Relation 'bearbeiten' statt etwa 'zum Verkauf ausschreiben' verwendet.

aufgaben.

Rummer (1996a) deutet an, daß die Problematik beim Entwurf experimenteller Untersuchungen bereits bekannt ist: Extreme Experimentalbedingungen können die abhängigen Variablen des Experiments in einen Zustand bringen, der die Versuchspersonen zur Wahl einer vom normalen Vorgehen unterschiedlichen Handlungsstrategie zwingt. Überraschende Versuchsergebnisse, die allen Erwartungen widersprechen, sind die Folge. Für die Modellierung der kognitiven Belastung ergibt sich als Schlußfolgerung daraus, daß die Belastung mit den Aussagen der vorliegenden Parameter *innerhalb eines bestimmten Belastungsintervalls* erkannt werden kann. Fälle extremer Belastung machen die Beobachtungen aus den zitierten Experimenten und daraus abgeleiteten Schlußfolgerungen ggf. vollkommen wertlos und bedingen eine Modellierung, die sich auf die Ergebnisse von Arbeiten stützt, die unter solchen Extrembedingungen gewonnen worden sind.

#### 3.2.6.4 Einsatzlatenz und Zeitdruck

Horton und Keysar (1996) weisen unter Zeitdruck eine Verringerung der Einsatzlatenz nach. Analog zu Powers Erkenntnissen ist auch nach der Interpretation von Horton und Keysar eine Reduktion der semantischen Verarbeitung für diesen Effekt verantwortlich, die vor allem zu einem Rückgang des Partnerbezugs der Äußerungen führt. Horton und Keysar (1996) zeigen außerdem, daß bei der Äußerungsgenerierung eine initiale Äußerungsplanung durchgeführt wird, die anschließend an Erfordernisse der Dialogsituation angepaßt wird. Zu diesen Erfordernissen zählt auch die Orientierung auf den Gesprächspartner.

Verantwortlich für das Einleiten einer Äußerungsanpassung ist der Monitoringprozeß, der die Information in der initialen Äußerung mit dem Hörermodell des Sprechers vergleicht. Unter Zeitdruck wird die Aktivität des Monitoring-Prozesses jedoch offenbar reduziert, weniger partneradaptierte Äußerungen sind die Folge. Durch den dadurch erreichten Gewinn an Verarbeitungszeit kann die Einsatzlatenz entscheidend reduziert werden.

#### 3.2.6.5 Probleme bei der Messung der Einsatzlatenz

Per Definition ist sowohl in der Verarbeitungs- als auch in der Planungsphase artikulierte Sprachproduktion ausgeschlossen. Eine Präzisierung kann jedoch dahingehend erfolgen, daß der Beginn der Artikulation erst ab dem Beginn der ersten 'substantiellen' Äußerung gewertet wird. Diese Messung übergeht in Übereinstimmung mit Greene et al. (1990) sowohl äußerungsinitiale gefüllte Pausen als auch ausschließlich der Verzögerung dienende stereotype Äußerungssequenzen (siehe Abschnitte 3.2.2 und 3.4.3).

Zur Interpretation der Ergebnisse in Tabelle 3.10 ist eine wichtige Unterscheidung notwendig: In den psychologischen Experimenten, die die Einsatzlatenz für die sprachliche Bearbeitung einer Aufgabenstellung messen, wird meist vom Sprecher ein explizites Signal[36] gefordert, sobald dieser meint, die Aufgabe verstanden oder den Stimulus verarbeitet zu haben. Damit kennzeichnet ein Sprecher den (weitgehenden) Abschluß seiner Verarbeitungsphase. Die in Tabelle 3.10 im Überblick dargestellten experimentellen Untersuchungen beziehen sich z.T. nur

[36] Das ist z.B. ein Mausklick oder das Drücken einer Taste.

auf diese Teilphase der Einsatzlatenz bis zum Ende der Verarbeitung (siehe Fußnoten).

Für ein Dialogsystem ist es unrealistisch, eine solche Kennzeichnung durch den Benutzer zu erwarten. Dadurch wird eine Trennung der beiden Teilphasen in den meisten Fällen unmöglich (Power, 1985). Die Definition der Einsatzlatenz muß daher aus praktischen Gründen die Form von Definition 3.6 beibehalten. Da die Wirkungen auf den zeitlichen Verlauf der beiden einzelnen Abschnitte nicht getrennt erfaßt werden können, müssen die entsprechenden Ursachen für die Veränderung einer einzelnen Phase als gemeinsame Ursachen für beide Phasen modelliert werden.

Die Einsatzlatenz muß ebenso wie einige andere Sprechparameter für eine geeignete Beurteilung an einem Basiswert relativiert werden. Im Gegensatz zu z.B. Sprechpausen und Artikulationsgeschwindigkeit ist dieser Basiswert jedoch stark aufgabenspezifisch, d.h. von der 'globalen' Schwierigkeit der Beantwortung (Weis, 1999, spricht von 'intrinsischer Schwierigkeit') der Partneräußerung und der Form der Dialogsituation (Dialog oder Monolog) abhängig. Das läßt sich auch an den stark unterschiedlichen Werten für die Einsatzlatenz in den verschiedenen in Tabelle 3.10 angegebenen Experimenten erkennen (zwischen 0.414 s bei (Roßnagel, 1995a) und 14.65 s bei (Greene, 1984)).

Unabhängig von der Schwierigkeit einer Aufgabe steht die Einsatzlatenz nach Power (1985) auch in Korrelation zur Menge des sprachlichen Materials, die zu äußern ist: Je mehr Worte die generierten Sätze bei Power enthielten, desto höher wurde auch der Wert der Einsatzlatenz. Indirekt besteht daher auch ein Zusammenhang zwischen der Einsatzlatenz und der Menge sprachlich zu vermittelnder Information.

### 3.2.7 Durchschnittliche Länge von Artikulationssequenzen

(3.7) Eine **Artikulationssequenz** ist eine Phase ununterbrochener Artikulation, d.h. die Zusammenfassung aller zwischen zwei stillen Pausen artikulierten sprachlichen Einheiten. Ihre Länge[37] wird durch die Zahl der in dieser Phase artikulierten sprachlichen Einheiten angegeben. Die **durchschnittliche Länge von Artikulationssequenzen** berechnet sich aus der Anzahl aller artikulierter sprachlicher Einheiten einer Äußerung und der Anzahl stiller Pausen.

Als Einheit dienen Silben(/Pause). Dieses Maß liefert eine Aussage darüber, wie häufig eine Äußerung durch Pausen unterbrochen wurde.

(3.8) **Me010, Dialog 3:**
<**äh**> <**P**> Im <**D**> <**P**> Innenraum unter'm Lenkrad <**P**> links <**P**> ist ein Sicherungskasten.

*Gesamtsprechdauer*

In Beispiel 3.5, hier wiedergegeben als 3.8, sind 4 stille Pausen enthalten. Insgesamt wurden (unter Ausschluß der gefüllten Pausen) 17 Silben artikuliert. Durchschnittlich waren die Artikulationssequenzen in diesem Beispiel daher 4.25 Silben (je Pause) lang.

[37] Motivation für den Vorzug der *Länge* über die *Dauer* ist die dadurch erreichte Unabhängigkeit von der individuellen Artikulationsgeschwindigkeit.

Grosjean und Deschamps (1973), Duez (1982), Wiese (1983), Kowal et al. (1983) und Kowal und O'Connell (1987) weisen darauf hin, daß die Länge von Artikulationssequenzen durch kognitive Belastung beeinflußt wird.

Die Bezeichnung dieses Parameters täuscht jedoch darüber hinweg, daß er keine neue Information über die Äußerungsstruktur liefert. Zusätzlich weist er einige Nachteile auf:

- Die Länge von Artikulationssequenzen ist abhängig von der Anzahl stiller Pausen. Für stille Pausen wurde jedoch herausgearbeitet, daß sie nicht nur in kognitiver Funktionen auftreten, sondern auch kommunikativ. Setzt ein Sprecher ggf. zusätzliche kommunikative Pausen, verringert sich die Länge der Artikulationssequenzen. Entsprechend den Ergebnissen den in Abschnitt 3.2.7 angegebenen Arbeiten würde das fälschlicherweise als Indikator kognitiver Belastung gewertet werden müssen.

- Der Parameter liefert die gleichen Informationen wie ein Parameter *Häufigkeit stiller Pausen* (siehe Grosjean & Deschamps, 1973), da er als dessen Kehrwert berechnet werden kann:

  $$\text{durchschnittliche Länge von Artikulationssequenzen} = \frac{1}{\text{Häufigkeit stiller Pausen}}$$

- Der Parameter ist undefiniert, wenn in einer Äußerung keine stille Pause auftritt[38]

- Gefüllte Pausen, die die Artikulation sinntragender sprachlicher Einheiten ebenfalls unterbrechen, werden nicht berücksichtigt.

- Der Wert kurzer Äußerungen wird von einer zusätzlichen Pause ungleich stärker betroffen als der einer langen Äußerung.

- Wie bei den kognitiven Rhythmen (Abschnitt 3.2.5) ist für die sinnvolle Messung dieses Sprechparameters eine gewisse Mindestlänge einer Äußerung erforderlich. Diese Mindestdauer ist lediglich in den monologischen Situationen der genannten experimentellen Untersuchungen des spontanen Sprechens gegeben.

Insgesamt sollte dem Parameter *Häufigkeit kognitiver Pausen* der Vorzug gegenüber der Länge von Artikulationssequenzen gegeben werden. Er gibt an, wie viele Pausen mit vorwiegend kognitiver Funktion relativ zur Äußerungslänge aufgetreten sind und ist daher nur dann nicht definiert, wenn auch keine Berechnung der Artikulationsgeschwindigkeit möglich ist. Das ist der Fall, wenn eine Äußerung lediglich aus stillen und gefüllten Pausen besteht.

### 3.2.8 Phonologische Sprechfehler

#### 3.2.8.1 Abgrenzung des Phänomens: Phonologische Sprechfehler

Phonologische Fehler entstehen nach Marx (1984), wenn '*auf Lautebene Elemente vertauscht oder weggelassen, spätere Silben antizipiert oder perseveriert werden*'.[39] Phonologische Fehler

[38]Das Problem stellt sich aber genauso für die Häufigkeit von Pausen, nur unter anderen Voraussetzungen.

[39]Marx spricht von phonetischen Fehlern, was m.E. für die intendierte Phänomenklasse bereits deshalb nicht korrekt ist, da in der Definition auf Silben bezug genommen wird. Silben sind jedoch keine Einheiten der Phonetik, sondern der Phonologie.

führen oft zum Entstehen von nichtlexikalischen Lemmata, d.h. Wortformen für die es keinen Eintrag im Lexikon der zugehörigen Sprache gibt. In anderen Fällen ist die Unterscheidung von phonologischen und lexikalischen Fehlern nur durch semantische Inferenzen möglich: Dann nämlich, wenn durch die Silbenmanipulation eine ebenfalls nicht intendierte, aber zufällig wieder lexikalische Wortform entsteht. An der Form eines phonologischen Fehlers ist das eigentlich bei der lexikalischen Enkodierung beabsichtigte Item durch Silben oder Silbenfolgen des generierten Nicht-Wortes erkennbar.

Levelt (1989) führt das Entstehen phonologischer Fehler z. B. auf Interferenzen bei der lexikalischen Enkodierung zurück, die von vorangehendem oder nachfolgendem sprachlichen Material herrühren können (z.B. als Verschmelzung von lexikalischen Formen oder Vertauschung von Morphemen über Wortgrenzen hinweg).[40]

Die Fehlerrate kann auf unterschiedliche Weise gemessen werden: Bei Bromme und Wehner (1987) werden Fehlerhäufigkeiten in Intervallen einer festen Zeitdauer erhoben und sind damit zeitlich basiert. Nachteil dieses Vorgehens ist die zusätzliche Abhängigkeit der Sprechfehlerrate von der Artikulationsgeschwindigkeit. Ein aussagekräftigeres Maß kann durch die Relativierung der Fehlerzahl an der Silbenzahl der zugehörigen Äußerung erzielt werden. Die Sprechfehlerhäufigkeit wird bei einer solchen Vorgehensweise nicht durch die Ausprägung anderer kognitiv relevanter Parameter verfälscht. Es erfolgt eine Gegenüberstellung zur Menge des *artikulierten* Materials, in dem Fehler ja auch überhaupt nur auftreten können.

#### 3.2.8.2 Phonologische Sprechfehler und kognitive Belastung

Levelt (1989, Kap.11) geht davon aus, daß für die Generierung der artikulatorischen Pläne für die verwendeten Lemmata (die bereits mit flexionsmorphologischer Information annotiert sind) gepackte phonetische Beschreibungen abgerufen und expandiert werden. Der Ablauf des dafür verwendeten Mechanismus' ist weitgehend ungeklärt (siehe Levelt, 1989, S.318ff.). Wegen der hohen Anforderungen an die Geschwindigkeit des Prozesses geht Levelt jedoch davon aus, daß sein Ablauf hochgradig automatisiert ist. Er kann somit nicht von kognitiver Belastung beeinflußt werden.

Diese Einschätzung wird durch experimentelle Arbeiten bestätigt: Die von Bromme und Wehner (1987) analysierten Einzelfehler (Versprecher, Neologismen und Stottern) traten nur außerordentlich selten auf und wurden daher in einem Wortfehlerindex zusammengefaßt. Auch für diesen Gesamtindex der Häufigkeiten von phonologischen Fehlern fanden Bromme und Wehner keine signifikante Veränderung unter kognitiver Belastung. Für die Einschätzung der Variablen 'Belastung' können daher die phonologischen Fehler an sich als irrelevant angesehen werden.

[40] Ein bekanntes Beispiel für einen phonologischen Fehler im Deutschen ist die Transformation von 'h-Moll-Messe' zu 'h-Mess-Molle', anschließend 'h-Moss-Melle' durch einen Radiosprecher. Levelt (1983) gibt als Beispiel 3.9, der tatsächlich auch als phonetischer Fehler gelten kann.

(3.9) A *unut*- unit from the yellow dot.

#### 3.2.8.3 Phonologische Sprechfehler und Zeitdruck

Gleiches kann u.a. aufgrund der Untersuchungsergebnisse von Marx (1984) für den Zusammenhang von phonologischen Sprechfehlern und Zeitdruck ausgesagt werden. Auch hier konnten keine signifikanten Veränderungen festgestellt werden.

Postma et al. (1990) beobachten zwar einen Effekt von Zeitdruck auf die Menge von Sprechfehlern, der vor allem auf phonologische Fehler zurückzuführen ist.[41] Allerdings bezieht sich auch diese Untersuchung lediglich auf Mikro-Zeitdruck, wobei die Versuchspersonen zusätzlich auch keine Konzeptualisierungsarbeit auszuführen hatten.

#### 3.2.8.4 Fazit

Phonologische Fehler erscheinen als Indikatoren für kognitive Belastung und Zeitdruck nicht geeignet, da sie außerordentlich selten auftreten und ihre Entstehung auf automatisierte Phasen der Sprachproduktion zurückzuführen ist. Ihr Zusammenhang mit Zeitdruck ist nur wenig untersucht und kann daher nicht beurteilt werden.

## 3.3 Merkmale der Syntax

### 3.3.1 Vorbemerkungen

#### 3.3.1.1 Syntaktische Operationen bei der Sprachgenerierung

Für die syntaktische Struktur einer Äußerung direkt relevante Operationen während der Sprachproduktion sind in Levelts Modell

- der lemmagesteuerte Aufbau syntaktischer Beschränkungen,
- die darauf basierende Auswahl von Satzrahmen,
- die Kombination der Lemmata in Phrasen einer Surface Structure sowie
- die Annotation der Lemmata mit Flexionsinformation.

Alle diese Operationen sind in der Formulierungsphase lokalisiert. Nach Levelts Modell ist die Formulierungsphase jedoch hochgradig automatisiert und von kognitiver Belastung nicht *direkt* beeinflußbar. Die hauptsächlichen Ursachen für Symptome aus der Formulierungsphase liegen demzufolge aufgrund der Inkrementalitätseigenschaft der Sprachproduktion in den zu dieser benachbarten Phasen der Konzeptualisierung und des Monitorings.

Einerseits kommen Fehlfunktionen der Konzeptualisierungsoperationen selbst als Ursachen syntaktischer Probleme in Frage, da durch sie mit der Auswahl konzeptueller Relationen die unmittelbaren Grundlagen für die Operationen der Formulierungsphase gelegt werden. Aber

[41] Kontrollgruppen 350 bzw. 815 Fehler, Zeitdruckgruppen 498 bzw. 1001 Fehler; ($p < 0.001, F(1,30) = 162.51$)

auch erst während der Artikulation verfügbare Information (und damit eine Fehlfunktion außersprachlicher Prozesse) kann zur Veränderung syntaktischer Strukturen führen. Eine wichtige Funktion dabei kommt dem Monitoring-Prozeß der Sprachproduktion zu.

#### 3.3.1.2 Einflüsse des Monitorings auf die syntaktische Struktur

Ausgaberepräsentation der unterschiedlichen Stufen der Formulierungsphase sind Pläne und Strukturen, die die Basis für die spätere Artikulation darstellen. Die Form dieser Strukturen ist nicht *direkt* zugänglich. Nur wenn das gesamte in der Formulierungsphase generierte Material an die Artikulationsphase weitergegeben wird, kann man über diese Strukturen und Pläne *indirekt* Aufschluß erhalten.

Der Monitoring-Prozeß der Sprachproduktion bewertet die Strukturen möglichst bereits vor der Artikulation und blockiert im allgemeinen[42] die Artikulation nicht wohlgeformter Strukturen. Wenn die Weitergabe von fehlerhaften oder inadäquaten Strukturen durch das Monitoring des Sprachproduktionssystems eingeschränkt wird, müssen diese Strukturen neu generiert oder nachgebessert werden. Die inkrementelle Verarbeitung wird dadurch verzögert und in den Strukturen der Formulierungsphase enthaltene Symptome für Belastung oder Zeitdruck äußern sich lediglich durch die zusätzlich benötigte Zeit, d.h. wiederum in Form von Verzögerungen der Artikulation (siehe Abschnitt 3.2).

In zahlreichen Fällen kann der Monitoring-Prozeß die Bewertung nicht vor der Artikulation abschließen, sondern führt sie erst auf einer parallel zur Artikulation ablaufenden Analyse der eigenen Äußerung aus. Wenn Strukturen der Formulierungsphase bereits artikulatorisch umgesetzt worden sind und erst dann vom Monitoring-Prozeß als verbesserungsbedürftig eingestuft werden, muß ein Sprecher offen zu artikulierende Selbstkorrekturen ausführen (Abschnitt 3.3.4). Diese Korrekturen stören ebenfalls die syntaktische Struktur einer Äußerung. Selbstkorrekturen sind daher als syntaktisches Phänomen zu sehen, obwohl die korrigierten Aspekte der zugrundeliegenden Äußerungsbestandteile anderen linguistischen Ebenen zugeordnet werden können.[43]

#### 3.3.1.3 Veränderungen syntaktischer Strukturen durch Belastung oder Zeitdruck

Resultat von kognitiver Belastung und Zeitdruck können Strukturen sein, die mit weniger Ressourcenaufwand entstehen. Der Abschnitt zu den syntaktischen Symptomen kognitiver Belastung läßt sich daher in *irreguläre Veränderungen* der Syntax und *Anpassungen durch präferierten Einsatz charakteristischer Strukturen* unterteilen. Beide Phänomene können inhaltliche Konsequenzen haben (siehe Abschnitt 3.5).

Bromme und Wehner (1987) und Jou und Harris (1992) weisen darauf hin, daß beim spontanen Sprechen eine dem schriftlichen ähnliche 'fehlerfreie' Sprachproduktion angestrebt wird (siehe dazu auch Clark & Clark, 1977). Diese kommunikative Notwendigkeit, fehlerfrei zu sprechen, kann aber in den seltensten Fällen umgesetzt werden. Analog zu verschiedenen Sprech-

[42] Beeinträchtigungen des Monitorings werden im Abschnitt 3.6 untersucht.

[43] Das Problem der Interaktion von Syntax und Semantik ist allgemeiner Art: Die beiden linguistischen Ebenen können kaum voneinander getrennt werden, da syntaktische Veränderungen semantisch/pragmatische Konsequenzen haben und umgekehrt semantische Entscheidungen die Syntaxstruktur beeinflussen.

parametern der prosodischen Ebene sind daher auch auf syntaktischer Ebene vor allem die *Veränderungen* ohnehin vorhandener syntaktischer Eigenschaften durch *zusätzliche* kognitive Belastung zu untersuchen. Entsprechend ist zu erwarten, daß für die meisten Parameter aus diesem Abschnitt Basiswerte herangezogen werden müssen.

Zur Beurteilung der Fehlerhaftigkeit ist ein Vergleich entstandener Strukturen mit Forderungen von Regelwerken auszuführen. Derartige Regeln können von unterschiedlicher Härte und entsprechend mehr oder weniger 'verletzungstolerant' sein. Weiche linguistische Constraints (z.B. der Pragmatik) stehen harten linguistischen Constraints, wie sie hier für die syntaktische Struktur vorgegeben sind, gegenüber.

## 3.3.2 Flexionsmorphologische Sprechfehler

### 3.3.2.1 Abgrenzung des Phänomens: Flexionsmorphologische Sprechfehler

Im Deutschen muß bei der Flexion von Lemmata auf fünf verschiedene Merkmale zurückgegriffen werden: Genus, Kasus, Numerus für Nomen und Adjektive sowie Numerus, Aspekt und Tempus für Verben.

(3.10) Ein **flexionsmorphologischer Fehler** liegt vor, wenn Kongruenzbedingungen auf den Kategorien Kasus, Genus, Numerus zwischen Verb und Nomen, Nomen und Adjektiven, Artikeln und Nomen, Präpositionen und Nominalphrasen etc. verletzt werden bzw. ungrammatische Verbalkonstruktionen ausgeführt werden.

Tempus und Aspekt weisen nicht nur grammatische, sondern vor allem inhaltliche Merkmale auf. Die Wahl eines fehlerhaften Tempus oder Aspektes ist nur mit einer wahrheitsfunktionalen Semantik zu beurteilen und daher als semantisches Merkmal einzustufen. Entsprechend werden Fehler des Tempus und Aspekt nicht als Flexionsfehler betrachtet.

Flexionsmorphologische Fehler definieren sich durch die Verletzung harter linguistischer Constraints. Daher sind sie im Gegensatz zu den meisten inhaltlichen Inadäquatheiten – siehe Abschnitt 3.5.2 und 3.4.1 – auch ohne eine Korrektur des Sprechers erkennbar. Ein Beispiel von (Marx, 1984) verdeutlicht den Begriff:

(3.11) Ich kenn' Dich jetzt schon seit *zwei Jahre* ...

### 3.3.2.2 Flexionsmorphologische Fehler und kognitive Belastung

Zum Zusammenhang flexionsmorphologischer Fehler mit Arbeitsgedächtnisbelastung sind nur wenige Untersuchungen bekannt. Oft fließen flexionsmorphologische Fehler lediglich in einen Gesamt-Sprechfehlerindex ein.

In den Experimenten von Jou und Harris (1992) waren parallel zum Wiedergeben von Information aus einem zuvor gehörten Zeitungsartikel Kopfrechenaufgaben (Summenbildung) auszuführen. Einzelne Summanden wurden über die Periode der Artikelpräsentation hinweg in bestimmten Abständen über Kopfhörer vorgegeben und dadurch eine dauerhafte kognitive Belastung (Behalten der Zwischensumme) mit einer lokal ansteigenden Belastung (Präsentation eines Summanden) kombiniert. Jou und Harris weisen auf den hohen Automatisierungsgrad der Flexion hin und stellen übereinstimmend damit auch in ihrer belasteten Gruppe kein einziges Auftreten flexionsmorphologischer Fehler fest.

Bei der Klassifikation von Selbstkorrekturen von Levelt (1983) werden flexionsmorphologische Fehler unter sublexikalischen Fehlern subsumiert. Korrekturen für diesen Fehlertyp traten in der Gesamtmenge aller Korrekturen bei Levelt in nur sehr geringem Maße auf (siehe Abschnitt 3.3.4).[44]

Obwohl die Flexion selbst automatisiert ist, muß ggf. jedoch beachtet werden, daß Flexionsfehler auch durch die Änderung einer konzeptuellen Entscheidung entstehen könnten: Die Änderung von Parametern der ausgewählten semantischen Relation einer umzusetzenden Proposition (z.B. der Perspektive) während der Artikulation kann sich in einer entsprechenden Korrektur äußern (Bsp. 3.12). Darauf wird im Abschnitt 3.3.4 näher einzugehen sein.

(3.12) Der Benzinfilter -, den mal festziehen.

#### 3.3.2.3 Flexionsmorphologische Fehler und Zeitdruck

Marx (1984) analysierte im Zusammenhang mit Zeitdruck einen Gesamtindex an Sprechfehlern, der sowohl flexionsmorphologische als auch syntaktische, lexikalische und phonologische ('phonetische') Fehler beinhaltet. Sie konnte für den Gesamtindex keinen signifikanten Unterschied bezüglich der Variablen Zeitdruck nachweisen. Aus den von Marx angegebenen Rohdaten für die flexionsmorphologischen Fehler ist zwar nicht ersichtlich, inwiefern ein Effekt bei einer getrennten Analyse vorliegen würde. Die geringe Anzahl an aufgetretenen flexionsmorphologischen Fehlern bestätigt jedoch, daß es sich bei Flexionsfehlern um ein sehr seltenes Phänomen handelt.

#### 3.3.2.4 Fazit

Flexionsmorphologische Sprechfehler erweisen sich aufgrund ihres seltenen Auftretens und der hochgradigen Automatisierung der Flexion bei der Sprachproduktion für die Einschätzung der kognitiven Belastung und des Zeitdrucks eines Sprechers als nicht geeignet.

### 3.3.3 Syntaktische Strukturfehler: Unvollständigkeit und Anakoluthe

#### 3.3.3.1 Abgrenzung des Phänomens: Syntaktische Strukturfehler

Wenn syntaktische Strukturen generiert werden, die nach den grammatischen Regeln der Sprache nicht wohlgeformt sind, entstehen syntaktische Fehler. Unter die verletzbaren Regeln fallen positions- und reihenfolgebezogene Constraints, die syntaktisch wohlgeformte Realisierung obligatorischer Konstituenten von Phrasen, die Anwesenheit obligatorischer Konstituenten als Einhaltung der Verbvalenz bzw. überhaupt die Forderung der Anwesenheit eines Verbs im

[44] Leider macht Levelt keine Angaben zur Anzahl unkorrigierter Fehler.

Satz.[45] Valenzbezogene Forderungen lassen sich auch als *Vollständigkeit*sforderungen zusammenfassen. Einige Beispiele für syntaktische Strukturfehler sind in den Beispielen 3.13 bis 3.16 angegeben. Wiederholungen von Worten oder Wortfolgen werden hier explizit *nicht* als syntaktische Strukturfehler betrachtet, sondern in einem eigenen Abschnitt untersucht (3.3.5)

(3.13) Do you hear with which ear? (Levelt, 1989)
(Anakoluth)

(3.14) Also nicht nur ich rausfinden, sondern Du mußt ihn auch beheben. (Me008, Dialog1)
(Unvollständige syntaktische Struktur: fehlendes Auxiliar, fehlendes direktes Objekt)

(3.15) Innenraum riecht's nach Benzin. (Me010, Dialog 1)
(Unvollständige syntaktische Struktur: fehlende Präposition und Artikel)

(3.16) Die Leitung, warum hört denn die – (Me016, Dialog 1)
(Satzabbruch)

Strukturen wie 3.13 entsprechen keinen syntaktischen Regeln der Satzbildung, da einerseits wichtige Konstituenten fehlen, zusätzlich vorhandene Teilstrukturen jedoch nicht die Anforderungen an diese Konstituenten erfüllen. Derartige Strukturen werden **Anakoluthe** genannt. 3.14, 3.15, und 3.16 stellen **unvollständige Strukturen** dar.

Stemberger (1982) analysiert verschiedene Typen syntaktischer Sprechfehler aus einem englischen Korpus. Er hebt hervor, daß in spontansprachlichen Korpora echte syntaktische Fehler nur sehr selten auftreten, bezieht sich damit jedoch vor allem auf Probleme der syntaktischen Vereinbarkeit von Konstituenten und weniger auf unvollständige Realisierungen. Die Bewertung syntaktischer Strukturen wird von Faktoren wie Dialekt, mangelnder Abgrenzung zu Phänomenen der Wort- und Morphemselektion sowie Unsicherheit und Subjektivität von Einschätzungen der 'Ungrammatikalität' erschwert. Der Übergang zwischen 'grammatischen' und 'ungrammatischen' Konstruktionen ist nicht klar abgegrenzt, sondern besteht aus einer Grauzone.

#### 3.3.3.2 Formen der Unvollständigkeit

Syntaktische Strukturen können auf verschiedene Art unvollständig sein.

(3.17) Ein unvollständiger Satz $S$ ist **intern unvollständig**, wenn keine Fortsetzung $F$ existiert, die den Satz zu einem wohlgeformten sprachlichen Zeichen macht. Ein Satz ist dagegen **extern unvollständig**, wenn er vor Ende seiner Generierung abgebrochen wird. In diesem Falle existiert zumindest eine *syntaktische* Fortsetzung $F$, die $S$ zu einem wohlgeformten Satz $SF$ vervollständigt.[46]

Zur Klasse intern unvollständiger Strukturen gehören Sätze mit fehlenden syntaktischen Argumenten im Satzinneren oder fehlendem Verb (3.14 und 3.15) an der durch die Satzstruktur vorgegebenen Position, wenn gleichzeitig nachfolgenden Konstituenten realisiert wurden.

[45]Ausnahmen siehe Abschnitt 3.4.2

[46]Für 3.16 wäre dies z.B. die Fortsetzung '... *auf zu funktionieren*'

Im Falle externer Unvollständigkeit wurde eine vom Sprecher geplante Fortsetzung entweder durch den Monitoring-Prozeß verworfen oder die Fortsetzung konnte mit der verbleibenden Information der noch nicht geäußerten konzeptuellen Teilstruktur nicht adäquat ausgeführt werden. In bestimmten Fällen kann auch ein mangelhafter Zugriff auf lexikalische Informationen Grund für einen Abbruch sein (Jou & Harris, 1992). Extern unvollständige Strukturen werden in einem eigenen Abschnitt behandelt (Abschnitt 3.3.6).

Nicht alle intern unvollständigen syntaktischen Strukturen sind jedoch fehlerhaft: Die Bildung von **Ellipsen** ist ein adäquates Mittel, Sprachproduktion hochgradig effizient zu halten: Dabei werden ausgezeichnete Argumente oder Relationen aus dem vorangehenden Diskurskontext nicht sprachlich realisiert, da sie vom Hörer leicht inferiert werden können. Die Bildung von Ellipsen unterliegt jedoch Regelmäßigkeiten und kann daher von unregelmäßiger Unvollständigkeit abgegrenzt werden.

Im Deutschen sind nur zwei Formen der Elision zugelassen: *Verbelision* in Koordinations-Strukturen oder Frage-Antwort-Sequenzen und *Lexikalische Elision*, die den Wegfall eines prototypischen oder generischen Arguments bei bestimmten Verben gestattet (Bußmann, 1990). Die Löschung eines Subjekts oder Objekts ist nicht erlaubt. Ellipsen, die sich auf den vorangehenden Diskurskontext beziehen, können jedoch mehr oder weniger schwer verarbeitbar sein. Die Bewertung einer Ellipse ist daher von den semantischen Eigenschaften der beteiligten Konstituenten und der vorangehenden Äußerungen abhängig. Für elliptische Äußerungen kann die Bewertung der Wohlgeformtheit daher nicht von einer Bewertung des Erfolges semantischer und pragmatischer Inferenzen getrennt werden. Deshalb werden Ellipsen gesondert im Abschnitt 3.4.2 behandelt.

#### 3.3.3.3 Anakoluthe, interne Unvollständigkeit und kognitive Belastung

Die Weitergabe von Teilen der konzeptuellen Struktur an die Formulierungsphase kann bereits erfolgen, wenn noch nicht der gesamte propositionale Gehalt für die Äußerung auf dem Niveau der Mikroplanung feststeht (Levelt, 1989, S. 161). Während des zeitlichen Voranschreitens (und entsprechend erfolgender Artikulation) *einer* Äußerung können in der Konzeptualisierungsphase durch neue Planungsentscheidungen Substrukturen generiert werden, die mit bereits formulierten oder artikulierten Äußerungsteilen nicht syntaktisch kombinierbar sind. Gründe dafür können sein: veränderte Perzeption, revidierte Folgerungen, revidierte Ziele, Verfügbarkeit neuer Information (siehe z.B. Levelt, 1983). Derartige Revisionen der konzeptuellen Struktur können in Verletzungen der grammatischen Regeln einer Sprache resultieren.

Jou und Harris (1992) beobachten unter Belastung eine signifikante Zunahme an nicht wohlgeformten syntaktischen Strukturen. Diese Untersuchung ist jedoch einzigartig, da in anderen experimentellen Studien lediglich Gesamtindices für „Sprechfehler"angegeben werden, die zwischen unterschiedlichen Arten von Fehlern nicht differenzieren (z.B. Bromme & Wehner, 1987) oder syntaktische Fehler nicht als Sprechfehler erhoben wurden (z.B. Postma et al., 1990). Die unterschiedlichen Zusammensetzungen dieser Gesamtindices erschweren es enorm, über Einzelphänomene wie syntaktische Fehler eine Aussage zu treffen. Die Zusammenfassungen sorgen vor allem dafür, daß Ergebnisse aus unterschiedlichen Untersuchungen nicht mehr vergleichbar sind. Die geringe Häufigkeit an syntaktischen Strukturfehlern, die Stemberger (1982) für spontane Sprache feststellte, zeigt sich allerdings auch in der Untersuchung von Jou und

Harris (1992): Fehlerhafte Strukturen stellten lediglich einen Anteil von 2 % an allen untersuchten Sprachdefekten.

#### 3.3.3.4 Ausnahmen für die Spontansprache: Extrapositionen

Einige syntaktische Veränderungen, die in der Schriftsprache nicht akzeptabel sind, werden bei der spontanen Sprache möglicherweise akzeptiert, da sie trotz ihrer Ungrammatikalität das Verständnis erleichtern. Dazu zählen z.B. Extrapositionen von Modifikatoren oder Argumenten in der in 3.18 und 3.19 angegebenen Form. Die Akzeptanz solcher Phänomene zeigt sich vor allem in der geringen Neigung von Sprechern, diese ungrammatischen Äußerungen zu korrigieren.

Uszkoreit et al. (1998) untersuchen syntaktisch korrekte Phänomene der Extraposition am Beispiel von Relativsätzen. Ihre Daten wurden aus schriftsprachlichen Korpora erhoben. Uszkoreit et al. (1998) stellen fest, daß verschiedene Einzelparameter wie *Gesamtlänge des Relativsatzes*, *Position der Landestelle im Satz* und *Menge des der Landestelle folgenden sprachlichen Materials* bestimmend dafür sind, ob eine Extraposition ausgeführt wird. Extrapositionen wie die in den beiden folgenden Beispielen können daher möglicherweise auf grundlegende Verarbeitungsbegrenzungen des kognitiven Apparates zurückgeführt werden. Im Zusammenhang mit kognitiver Belastung stellt sich z.B. die Frage, ob unter Belastung ggf. bereits unter schwächeren Constraints extraponiert wird und ob ausreichend starke Beschränkungen möglicherweise auch zu nicht-wohlgeformten Strukturen führen können.

(3.18) Ich hab noch 'ne Frage, überhaupt. (Me006, Dialog 1)
(Rechtsextraposition eines Modifikators)

(3.19) An den Behälter, da geht unten ein Kabel dran. (Me023, Dialog 5)
(Fokussierung eines Präpositionalobjekts)

(3.20) ... oder ob da ein Stein drin ist oder ein Nagel. (Me031, Dialog 2)
(Rechtsextraposition eines NP-Konjunkts)

(3.21) Es wird nicht schneller, das Auto. (Fr004, Dialog 4)
(Rechtsextraposition eines Arguments: Subjekt)

Die vorgezogene (3.19) bzw. nachgeschobene (3.21) Generierung eines Referenten bzw. einer Lokalisation, die in der Realisierung der sie verbindenden Relation lediglich anaphorisch (bzw. kataphorisch) aufgenommen werden, sind Beispiele für Extrapositionserscheinungen, wie sie in der Schriftsprache normalerweise nicht zu finden sind. Im folgenden wird auf diese Phänomene mit den Begriffen **Linksextraposition** bzw. **Rechtsspezialisierung**[47] verwiesen.

[47] Auch Worm, Egg und Winter (1997) bevorzugen den Begriff **Rechtsextraposition**.

#### 3.3.3.5 Ausnahmen der Spontansprache und kognitive Belastung

Für die Phänomene, die in den Beispielen 3.18 bis 3.21 dargestellt wurden, stellt sich die Frage, ob einerseits der Sprecher durch derartige Mechanismen seine eigene Generierungsarbeit verringern (bzw. die Gesamtanforderungen zeitlich aufteilen) kann und ob dadurch sogar eine Anpassung an die bereits bestehende eigene kognitive Belastung erfolgt. Andererseits vereinfacht der Sprecher durch eine derartige Fokussierung auf die Referenten einer Proposition bzw. auf die Relation durch deren vorzeitige Realisierung auch die Verarbeitung des Hörers. Im Falle eines bewußten Einsatzes kommen daher auch kommunikative, d.h. hörerorientierte Ziele für die Anwendung der Konstruktionen in Frage.

Phänomene der Verarbeitungsbeschränkung bei der Sprachverarbeitung, wie sie beispielsweise in Sätzen mit mehrfacher Zentraleinbettung auftreten, werden in der Psycholinguistik seit langem untersucht (z.B. Gibson, 1991). Dabei hat sich auch herausgestellt, daß entsprechende Verarbeitungsbeschränkungen bei der spontansprachlichen Generierung auftreten, während sie bei der schriftlichen Formulierung normalerweise weniger Probleme bereiten. Inwiefern diese Beschränkungen sich unter zusätzlicher Belastung verstärken, ist bisher offenbar nicht untersucht worden.[48]

#### 3.3.3.6 Fazit

Syntaktische Strukturfehler in Form von Anakoluthen und interner Unvollständigkeit treten nach den Erkenntnissen aus der Literatur in spontansprachlichen Korpora nur selten auf. Unter Belastung erfolgt jedoch eine häufigere Generierung solcher Strukturen. Sie stellen insgesamt ein empirisch wenig untersuchtes Symptom dar. Sie können als schwaches Symptom kognitiver Belastung gewertet werden, wenn man die Notwendigkeit, fehlerfrei zu sprechen, auch bei der spontanen Sprachproduktion anerkennt. Diese Entscheidung sollte einer Revision unterzogen werden, sobald genauere Erkenntnisse über den exakten Zusammenhang von kognitiver Verarbeitungsfähigkeit und strukturellen Fehlern vorliegen.

Zahlreiche Phänomene aus dem Grenzbereich zwischen regulären und fehlerhaften Konstruktionen, insbesondere in bezug auf Extrapositionen lassen eine genauere Untersuchung interessant erscheinen, da einige Schlußfolgerungen aus der Analyse dieser Strukturen mit Vorhersagen aus dem Arbeitsgedächtnismodell vereinbar sind.

Über den Zusammenhang zum Zeitdruck kann aufgrund fehlender experimenteller Evidenz nichts ausgesagt werden.

[48]Es deutet sich hier ein interessantes Forschungsgebiet an, auf dem zahlreiche potentielle Indikatoren kognitiver Belastung vermutet werden dürfen. Modelle kognitiver Verarbeitung wie Gibson (1991) oder das u.a. in Lewis (1993) vorgestellte NL-SOAR können dabei unterstützend eingesetzt werden, da ihre prozedurale Definition explizit auf die kognitive Kapazität eines Sprechers (bzw. Hörers) Bezug nimmt.

### 3.3.4 Selbstkorrekturen

#### 3.3.4.1 Abgrenzung des Phänomens: Selbstkorrekturen

Nur gering zeitversetzt zur Artikulation des Sprechers baut der Hörer syntaktische und semantische Strukturen des Gehörten auf. Da der Monitoring-Prozeß bei Levelt durch das Sprachverstehenssystem übernommen wird, ist der Sprecher in der Lage, seine eigene Äußerung auf die Implikation möglicherweise falscher Schlußfolgerungen hin zu untersuchen. Eine Rücknahme artikulierter Strukturen (im Sinne einer Löschung) durch den Sprecher ist nicht mehr möglich. Daher setzen Sprecher im Falle einer fehlerhaften Äußerung oder problematischer Äußerungsbestandteile spezielle Mechanismen ein, die dem Hörer signalisieren, in welcher Form er bereits bestehende Strukturen zu revidieren und durch andere zu ersetzen hat. Das resultierende Phänomen der Selbstkorrekturen ist u.a. von Levelt (1983) umfangreich untersucht worden.

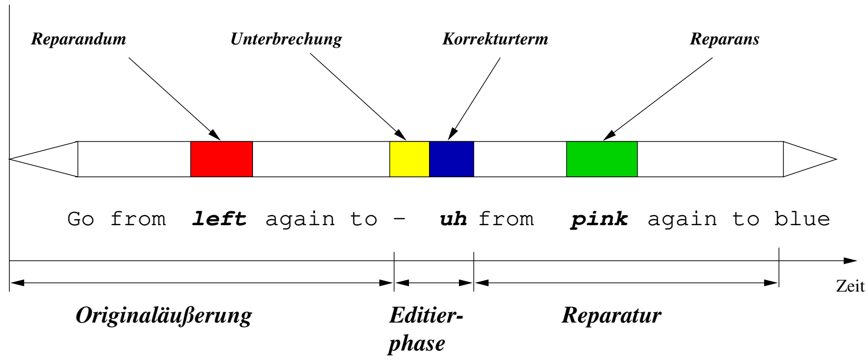


Abbildung 3.9: Phasen und zeitliche Organisation einer Selbstkorrektur an einem Beispiel von Levelt (1983)

Selbstkorrekturen sind ausschließlich für die Spontansprache charakteristisch und stehen daher in direktem Zusammenhang mit der kognitiven Belastung durch die Äußerungsplanung *während* des Sprechens. Levelt (1983) arbeitet die in Abb. 3.9 angegebene Struktur einer Selbstkorrektur heraus.

Der Sprecher äußert mit der **Originaläußerung** eine Konstruktion oder ein Lexem, die sein Monitoring-Prozeß als fehlerhaft oder inadäquat einschätzt. Dieser Bestandteil der Originaläußerung wird *Reparandum* genannt. Der Alarm des Monitors führt (ggf. mit einer gewissen Verzögerung) zu einer **Unterbrechung** der Artikulation, die von einem **Korrekturterm** gefolgt sein kann. Der Korrekturterm gibt dem Hörer möglicherweise nähere Informationen darüber, wie die anschließend erfolgende **Reparatur** in die bisherige Äußerung zu integrieren ist. Der Reparatur-Abschnitt enthält das *Reparans*, welches das Reparandum ersetzt. Information über die Ansatzstelle des Reparaturstrings in der Originaläußerung ist auch implizit in der letzten Phase der Korrektur enthalten, indem Teile der Originaläußerung wieder aufgenommen werden, die dem Reparandum vorangehen. Im Beispiel von Abbildung 3.9 erfolgt diese Wiederaufnahme vor der Präposition 'from', die eine Präpositionalphrase einleitet.

#### 3.3.4.2 Klassen von Selbstkorrekturen

Nicht immer handelt es sich beim Reparandum um eine fehlerhafte Konstruktion im Sinne linguistischer Regelverletzungen. Als korrigierbedürftig werden etwa auch (potentielle) Ambiguitäten eingeschätzt oder die Wahl einer unpassenden Terminologie. Levelt (1989) bezeichnet die beiden korrigierten Ebenen als *intention and meaning* und *linguistic deviancy*. Levelt (1983) nimmt folgende Klassifikation von Selbstkorrekturen vor:

**Error-Repairs:** Korrekturen der Formulierung: Die konzeptuelle Struktur wird nicht geändert. Die korrigierten Fehler gliedern sich in drei Klassen:

- Lexikalische Fehler durch Abruf eines unzutreffenden Lemmas, die zu einer falschen Aussage führen würden (z.B.: 'links' statt 'rechts'). Diese Klasse stellt bei Levelt (1983) 38% aller Korrekturen und wird mit $EL$ bezeichnet.
- Syntaktische Fehler, d.h. syntaktische Realisierungen von Referenten oder Relationen, für die mit dem noch nicht artikulierten Teil der konzeptuellen Struktur keine Fortsetzung möglich ist (sogenannte *deadlocks*) (z.B. 'And black ... - uh and *from* black to white.') ($ES$, 2%)
- Phonetische/Phonologische Fehler[49] ($EF$ 1%)

**Difference-Repairs:** Korrekturen der konzeptuellen Struktur ($D$ 1%)

**Appropriateness-Repairs:** Korrektur von Teilen der konzeptuellen Struktur in Hinblick auf Angemessenheit von Begriffen bezüglich Hörer, Sprecherzielen und Dialogsituation.[50] Es existieren drei Klassen:

- Ambiguität im Kontext, z.B. von Referenzen (z.B. 'in the middle of the paper' statt 'in the middle') (AA, 5%)
- unangepaßter Terminologie-Level, z.B. durch Unterspezifikation (z.B. 'the blue spot' statt 'the blue disc')(AL, 13 %)
- Inkohärenz mit bereits im Diskurs verwendeten Referenten und Relationen (AC, 5%)[50 51]
- Zweifelsfälle zwischen den Klassen AL und AC (ALC, 7%)

**Covert Repairs:** In der Äußerung ist kein Reparandum und kein Reparans identifizierbar. Diese Klasse wird durch zwei unterschiedliche Phänomene vertreten:

- Realisierung von Unterbrechung und Korrekturterm, anschließend normale Fortführung der Äußerung (17%)
- Wiederholungen von Worten oder Wortgruppen (7 %).

[49] Zur Terminologie siehe Abschnitt 3.2.8.

[50] Levelt hebt hervor, daß es sich bei Verletzungen der 'Angemessenheit' nicht um Fehler wie bei den Error-Korrekturen handelt, d.h. nicht korrigierte Appropriateness-Fehler führen nicht zu falschen Aussagen.

[51] Levelt gibt als charakteristisches Beispiel die Kohärenz durch Beschränkung auf Bewegungsverben an, z.B. wird als Fortsetzung zu „Wenn Du einen Schritt weitergehst..."möglicherweise statt des statischen Verbs 'sein' in „... dann *bist* Du bei X."lieber ein Bewegungsverb verwendet, wie in „... dann *kommst* Du zu X.".

**Rest Category:** Alle nicht anderweitig klassifizierbaren Korrekturen (R, 2.5%)

Drei Korrekturklassen werden in der Literatur auch unter anderen Namen analysiert und daher in getrennten Abschnitten genauer behandelt: *Difference-* und eine Untermenge der *Syntactic Error-* Korrekturen werden als *Fehlansätze* bezeichnet (Abschnitt 3.4.1). *Covert Repairs* sind *Wiederholungen* (3.3.5) und *stille oder gefüllte Pausen* (3.2.2).

#### 3.3.4.3 Häufigkeit von Selbstkorrekturen, kognitive Belastung und Zeitdruck

Ein Sprecher führt Selbstkorrekturen nur dann aus, wenn sein Monitoring-Prozeß problematische Äußerungsbestandteile nach Ausgabe der Formulierungsphase (internal speech) bzw. nach Artikulation (overt speech) erkennt *und* eine Korrektur veranlaßt. Nicht erkannte Inadäquatheiten und Fehler sind daher nicht von erkannten, aber nicht korrigierten Inadäquatheiten zu unterscheiden. Von vornherein korrekte Strukturen können lediglich durch zeitliche Verzögerungen von intern korrigierten Strukturen unterschieden werden. Diese Verzögerungen sind auf die nötige Zeit für die Nachbesserung der ursprünglichen Planung zurückzuführen. Verringertes Monitoring läßt sich daher vor allem an einer Änderung des Verhältnisses von offen korrigierten zu nicht korrigierten Sprechfehlern erkennen. Abbildung 3.10 zeigt die Funktionsweise und Position des Monitoring- Prozesses in der Sprachproduktion und die zur Überwachung der inneren und äußeren Rede dienenden Rückkopplungsschleifen.

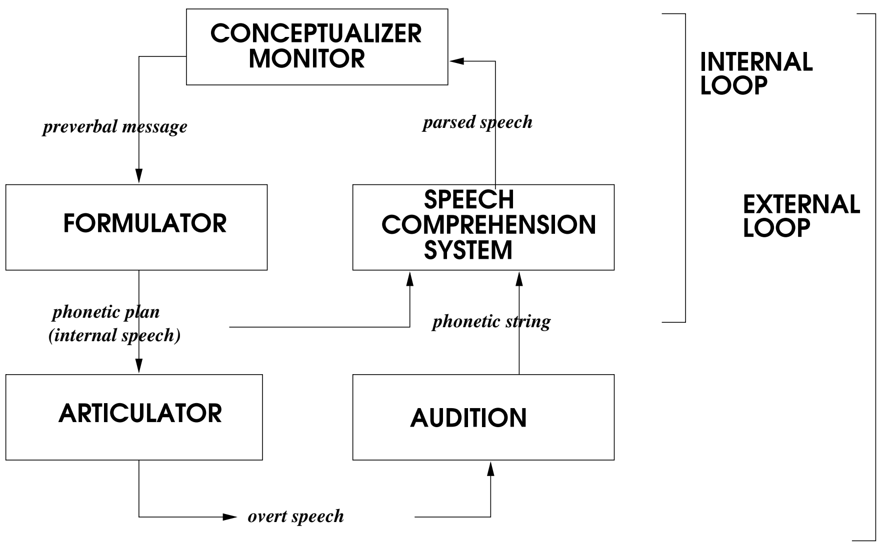


Abbildung 3.10: Die *Perceptual Loop Theory* des Monitorings nach Levelt (1989)

Die Effekte kognitiver Belastung auf die Häufigkeit von Selbstkorrekturen sind widersprüchlich, wie ein Blick auf Tabelle 3.12 klarmacht. Die Tabelle enthält allerdings lediglich jene Arbeiten, die einen Gesamtindex an Selbstkorrekturen angeben. Eine Einzelanalyse der Häufigkeiten für die im vorangegangenen Abschnitt vorgestellten Ursachenkategorien ergibt in einigen Fällen ein eindeutigeres Bild (siehe Abschnitte 3.2.2, 3.3.5, 3.4.1).

Die Korrektur von Äußerungen ist ein Prozeß, bei dem dem Hörer eine korrekte Modifikation der bisher aufgebauten Strukturen für die Partneräußerung ermöglicht werden muß. Wenn

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Bromme und Wehner (1987)[=][a] | Rein sprachliche Aufgaben mit unterschiedlicher außersprachlicher Planung | 12 | Zusammenstellung einer Speisekarte (Aufgabe) | Denksportaufgabe (Problem) | 4.5[b], ($SD = 2.5$) | 5.1[b], ($SD = 3.4$) | n.s. | (+/−)[c] |
| Deese (1980)[=] | Sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorherige Planung der Makrostruktur möglich (referatartiges Konzept) | Keine vorherige Planung der Makrostruktur | 1.66/100 Worte | 1.65/100 Worte | n.s. | 0 |
| Jou und Harris (1992)[=] | Vorlesen eines Zeitungsartikels mit anschließender Inhaltswiedergabe | 52 | Einfache Inhaltswiedergabe | Inhaltswiedergabe, begleitet von Kopfrechenaufgaben (*mental arithmetic*) | 1.38[d] | 1.34[d] | n.s. | 0 |
| Oviatt (1995), Exp. 1 | Sprachliche Eingabe von Anforderungen für das Mieten eines Fahrzeugs | 44 | Vorstrukturierte Eingabe | Unstrukturierte Eingabe | 0.78/100 Silben | 1.78/100 Silben | $p < 0.15$, $t_+(14) = 88$, one-tailed | (+) |
| Rummer (1996a)[=], Exp. 1 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer diskontinuierlichen Doppelaufgabe (Bewertung aufeinanderfolgender Lichtsignale auf farbliche Gleichheit) | 0.0041 /s (Bericht), 0.0018 /s (Erzählen) | 0.0013 /s (Bericht), 0.0004 /s (Erzählen) | $p < 0.1$, ($F(1,54) = 2.5$) | − |
| Rummer (1996a)[=], Exp. 2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer kontinuierlichen Doppelaufgabe (Tracking- Aufgabe) | 0.004 /s (Bericht), 0.002 /s (Erzählen) | 0.003 /s (Bericht), 0.000 /s (Erzählen) | n.s. | 0 |
| Wiese (1983)[=] | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Sprechen in der Muttersprache | Sprechen in einer Fremdsprache | 0.68/100 Silben, ($SD = 0.67$) | 1.68/100 Silben, ($SD = 1.60$) | $p < 0.05$ | + |

Tabelle 3.12: Untersuchungen zu Selbstkorrekturen

[a]Fehlansätze und Satzfragmente werden zusammengefaßt.

[b]Der Wert ist '*an der Gesamtverbalisationsmenge relativiert*'. Eine Verbalisationseinheit ist nicht angegeben.

[c]Die Anpassung erfolgte abhängig vom Individuum in unterschiedliche Richtungen.

[d]Relativierte Anzahl von Selbstkorrekturen. Eine Basiseinheit für die Relativierung ist nicht angegeben.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Zeitdruck** | | | | | | | | |
| Marx (1984)[=] | Führen eines Gespräches mit Ziel der Überzeugung eines Partners | 70 | Ohne Zeitdruck | Zeitdruck durch Wiederholung einer bereits ohne Zeitdruck ausgeführten Tätigkeit mit Zeitbegrenzung | 1.10/100 Silben, $(SD = 1.07)$ | 1.53/100 Silben, $(SD = 0.74)$; [1.18/100 Silben, $(SD = 0.56)$(NZD)[a]; 1.72/100 Silben, $(SD = 0.76)$(HZD)] | $p > 0.058$(n.s.), $(t = 1.93)$; [n.s. (NZD); $p = 0.022$, $(t = 2.37)$(HZD)] | (+) |
| Marx (1984)[=] | Lautes Denken über ein Problem | 70 | Ohne Zeitdruck | Zeitdruck durch Wiederholung einer bereits ohne Zeitdruck ausgeführten Tätigkeit mit Zeitbegrenzung | 0.76/100 Silben, $(SD = 0.87)$ | 0.62/100 Silben, $(SD = 0.83)$; [0.82/100 Silben, $(SD = 1.01)$(NZD)[a]; 0.51 /100 Silben, $(SD = 0.72)$(HZD)] | n.s. | 0 |

Tabelle 3.13: Untersuchungen zu Selbstkorrekturen (Fortsetzung)

[a] Postexperimentell erfolgte eine Zuordnung der Probanden zu den Gruppen 'hohes Zeitdruckempfinden' (HZD) bzw. 'niedriges Zeitdruckempfinden' (NZD) durch Selbsteinschätzung.

ein Hörer über genügend Information verfügt, kann er diese Modifikation aufgrund seines linguistischen, Kontext- und Weltwissens ggf. auch selbst ableiten. Insofern ist die Ausführung einer Selbstkorrektur ein Zugeständnis an den Hörer. Evidenz dafür gibt der Unterschied des Ausmaßes an Selbstkorrekturen in monologischen Situationen ohne Hörer (lautes Denken), gegenüber Situationen mit Zuhörer bzw. in Dialogen. So liegen z.B. bei Marx (1984) die für das Laute Denken erhobene Werte unter denen des spontanen Sprechens gegenüber einem Partner.

Verschiedene Begründungen kommen für die widersprüchlichen Ergebnisse in Tabelle 3.12 in Frage: Abschnitt 3.6 zeigt, daß die Aktivität des Monitoringprozesses von kognitiver Belastung und Zeitdruck reduziert werden kann. Eine Erhöhung der Selbstkorrekturrate läßt sich einerseits auf eine höhere Rate an Fehlern und Ungenauigkeiten in der Konzeptualisierungsphase zurückführen, andererseits jedoch auch auf einen niedrigeren Anteil davon bereits in der *internal speech* erkannter Fehler. Dadurch werden verstärkt externe Korrekturen notwendig. Eine Senkung der Selbstkorrekturrate ist dagegen nicht allein mit einer niedrigeren Fehlerrate zu erklären (die Zahl verschiedener Fehler steigt ja unter Belastung an). Sie kann in diesen Fällen nur auf eine reduzierte Korrektureinleitung zurückgeführt werden.

Offenbar sind die Zusammenhänge von Selbstkorrekturen mit den semantischen Aspekten der sie enthaltenden Äußerungen komplexer als bisher angenommen. Eine pauschale Messung der Häufigkeit aller Klassen von Selbstkorrekturen ist dem möglicherweise nicht angemessen. Um zu aussagefähigen Ergebnissen zu kommen, müssen daher neue Wege der Analyse eingeschlagen werden.

Zu beachten ist im Zusammenhang mit der Betrachtung des Verhältnisses von Selbstkorrekturen zu unkorrigierten Fehlern jedoch, daß bei der Spontansprache nicht für alle auftretenden Fehler (E-Kategorie) Korrekturen erwartet werden dürfen. Grundlage dafür kann, wie auch schon für die Phänomene aus Abschnitt 3.3.3 eine Einschätzung der Kosten für die Korrektur im Verhältnis zum dadurch erreichbaren Nutzen sein. Verletzungen syntaktischer Constraints sind beim spontanen Sprechen daher mit einem höheren Akzeptanzgrad versehen als bei schriftsprachlichen Äußerungen.[52] Auslassungen von Artikeln (Bsp. 3.22), oder Links- und Rechtsextrapositionen von Konstituenten und Modifikatoren (Bsp. 3.18 und 3.19) sind als Beispiele für im Diskurs weitgehend akzeptierte Phänomene anzusehen.

(3.22) Hast Du Handy oder sowas?

#### 3.3.4.4 Spezialfall: Häufigkeit von ES-Korrekturen und Zeitdruck

Die Anzahl der Korrekturen syntaktischer Fehler nimmt laut Marx (1984) unter Zeitdruck zu. Als Erklärung dafür nimmt Marx an, daß ein Sprecher unter Zeitdruck seine Handlungen nur ungenau plant und daher möglicherweise von vornherein bereits mehr Fehler generiert. Dem widersprechend zeigen allerdings bei ihr die anderen Kategorien von Fehlern keine signifikante Veränderung unter Zeitdruck. (Ein Sprecher würde unter Zeitdruck also vor allem nur mehr syntaktische Fehler generieren.) Ein daraus folgendes Problem ist, daß die Funktion des Monitoring-Prozesses von Zeitdruck unbeeinträchtigt bleiben muß, um die Entdeckung der Fehler zu gewährleisten. Verschiedene Arbeiten zur Hörerorientierung von z.B. Roßnagel

[52] Kapitel 4 und Worm et al. (1997) geben einen Überblick über linguistische Klassifikation außergrammatischer Phänomene, die in den Korpusdaten von READY auftreten.

(siehe Abschnitt 3.6) legen allerdings eine Reduzierung des Monitorings unter kognitiver Belastung/Zeitdruck nahe. Eine Relativierung der von Marx angegebenen absoluten Werte für Korrekturen an der Fehlermenge (wie sie bereits oben vorgeschlagen wurde) könnte mehr Klarheit bringen.

#### 3.3.4.5 Unterbrechungsstelle bei Selbstkorrekturen und kognitive Belastung

Levelt stellte bei seiner Analyse Regelmäßigkeiten für die Form der Realisierung einer Selbstkorrektur über alle Klassen von Selbstkorrekturen hinweg fest. Unter kognitiver Belastung können diese Aspekte Veränderungen erfahren. Das bezieht sich insbesondere auf den Unterbrechungszeitpunkt, für den nach Levelts Korpusanalyse folgende Regel gilt:

(3.23) **Main Interruption Rule:** Stop the flow of speech immediately upon detecting a trouble.

Unterbrechungen erfolgen bei über 50% der Daten von Levelt (1983) unmittelbar (d.h. mit 0 Silben Abstand) hinter dem fehlerhaften Item. Die Ursachen für einen verspäteten Abbruch können verschiedener Natur sein: entweder wird ein Abbruchsignal vom Monitoring-Prozeß verzögert gegeben (d.h. der Fehler wird erst nach der Artikulation entdeckt). Verantwortlich dafür kann ein vorübergehendes Suspendieren des Monitorings sein. Eine zweite mögliche Erklärung wäre, daß eine bereits initiierte Unterbrechung erst verspätet erfolgt (das Signal liegt vor, wird aber eine Zeitlang ignoriert). Wie bereits beschrieben, können auch die automatisierten Phasen der Sprachproduktion unter gewissen Bedingungen unter exekutive Kontrolle gestellt werden, was insbesondere für einen Abbruch notwendig ist. Eine verspätete Unterbrechung könnte dann als dritte Ursache haben, daß die exekutive Kontrolle eine Zeitlang nicht realisierbar war. Beide Phänomene lassen sich im Baddeleyschen Modell durch hohe Belastung des Arbeitsgedächtnisses erklären. Eine bewußte Zurückstellung des Monitoralarms, etwa um eine begonnene komplexe Phrase zunächst zu beenden, schließt Levelt mit der Regel 3.23 explizit aus.

Eine wichtige Frage für die Erklärung der Ursachen verspäteter Unterbrechung ist, *wie* das Monitoring erfolgt: ist es ein kontinuierlicher Prozeß oder kann auch Monitoring erst auf einer gewisse Menge an Eingabeinkrementen starten? Aufschluß darüber geben Analysen des Sprachverstehenssystems. So stellt Levelt (1989) anhand eigener Daten fest, daß Monitoring vor allem gegen Ende von Phrasen stattfindet, während die Aufmerksamkeit innerhalb von Phrasen vor allem für die Planung eingesetzt wird. Diese Beobachtung ist bei der Beurteilung des Unterbrechungszeitpunktes in Betracht zu ziehen, da bei Inaktivität des Monitoring-Prozesses während der Planung keine Unterbrechung erwartet werden darf.

Daß die Funktion des Monitoring-Prozesses ganz allgemein durch kognitive Belastung oder Zeitdruck reduziert werden kann, wird im Abschnitt 3.6 deutlich gemacht. Wenn die gleichen Mechanismen für die Überwachung der Hörerorientierung verantwortlich sind, wie für die Kontrolle der linguistischen Adäquatheit von Äußerungen und das Monitoring prinzipiell zu jeder Zeit eine Korrektur einleiten kann, dann stellt die verspätete Einleitung einer Selbstkorrektur ein zusätzliches Symptom für Belastung dar. Dieser Zusammenhang muß genauer untersucht werden, ehe der Parameter für den Einsatz in Dialogsystemen geeignet sein wird.

#### 3.3.4.6 Struktur von Selbstkorrekturen und kognitive Belastung

Im Zusammenhang mit der Hörerorientierung eines Sprechers steht die äußere Form von Selbstkorrekturen. Es wurde bereits darauf verwiesen, daß sowohl Korrekturterm als auch Wiedereinsatzstelle Indikatoren dafür sein können, wie der Hörer die bereits aufgebauten sprachlichen Strukturen zu modifizieren hat. Für einen großen Teil der Selbstkorrekturen in seinen Korpusdaten stellt Levelt fest, daß ihre strukturelle syntaktische Realisierung eine Wohlgeformtheitsregel (3.24) erfüllt. Grundaussage dieser Regel ist, daß Originaläußerung und Reparatur eine grammatische Konstruktion ergeben, wenn man sie durch eine Konjunktion verbindet.

(3.24) **Wellformedness Rule:** An original utterance $O$ plus repair $R$, $< OR >$, is well formed if and only if there is a string $C$ such that the string $< OC\ or\ R >$ is well formed, where $C$ is a completion of the constituent directly dominating the last element of $O$ (*or* is to be deleted if that last element is itself a connective such as *or* or *and*).

Mit dieser Regel ordnet Levelt die Ausführung von Selbstkorrekturen als ein syntaktisches Phänomen mit einem Regelwerk ein, das dem Status der grammatischen Regeln einer Sprache gleichkommt. Eine Verletzung der Wohlgeformtheitsregel entspricht somit einem syntaktischen Fehler. Levelt (1989, S.487) hebt hervor, daß kognitive Belastung oder Zeitdruck Gründe für die Verletzung von Regel 3.24 sein können. Untersuchungen, die diesen Aspekt systematisch überprüfen, fehlen zum gegenwärtigen Zeitpunkt.

#### 3.3.4.7 Korrekturterme und kognitive Belastung

Die Verwendung eines Korrekturterms signalisiert dem Hörer zum einen, daß eine Korrektur stattfindet (und verhindert, daß dieser das selbst inferieren muß) und gibt ggf. auch Auskunft darüber, welches Element der Originaläußerung repariert werden soll bzw. welchen Typs die Selbstkorrektur ist (Levelt, 1989, S. 483). So traten in Levelts Daten in 62 % der Fälle Korrekturterme auf, wenn eine Korrektur des E-Typs erfolgte, jedoch nur in 28 % der Fälle, wenn es sich um eine A-Korrektur handelte. Auch die Mengen verwendeter Korrekturterme waren für unterschiedliche Korrekturtypen verschieden. Levelt gibt für sein englisches Korpus z.B. eine starke Präferenz für 'er', 'or', 'no' bei E-Korrekturen, 'thus' für A-Korrekturen an. Bei E-Korrekturen wurde 'sorry' oft als Entschuldigung verwendet, aber fast nie für Reparaturen der A-Klasse.

Da diese Eigenschaft einer Selbstkorrektur stark hörerbezogene Aspekte aufweist (der Sprecher weiß ja, was er korrigiert) und die allgemeine Hörerorientierung eines Sprechers unter kognitiver Belastung und Zeitdruck nachweislich abnimmt (siehe Abschnitt 3.6), könnten Untersuchungen des Gebrauchs von Korrekturtermen (Häufigkeiten und Arten) im Zusammenhang mit kognitiver Belastung und Zeitdruck den theoretisch vorhersagbaren Einfluß der beiden Variablen überprüfen helfen.

Besonders der von Levelt als universal herausgestellte Korrekturterm 'er' (zu 'übersetzen' wohl am ehesten mit 'äh') steht in engem Zusammenhang zu gefüllten Pausen: Er begleitete in Levelts Daten 84 % der Fälle von C-Korrekturen. Sein Einsatz erfolgt umso häufiger, je näher die Unterbrechungsstelle am Reparandum liegt. Möglichereise wird dieser Korrekturterm also in einer Doppelfunktion eingesetzt. Für die Klassifikation von 'äh' als gefüllte Pause stellt dies jedoch kein Problem dar.

#### 3.3.4.8 Fazit

Die Häufigkeit des Auftretens von Selbstkorrekturen kann nur dann sinnvoll in Zusammenhang mit kognitiver Belastung und Zeitdruck gebracht werden, wenn die unterschiedlichen Klassen von Korrekturen getrennt analysiert werden. Diese Unterscheidung erfolgte bisher in keiner experimentellen Arbeit. Relevant für die Einschätzung der kognitiven Belastung sind folgende wichtige Aspekte von Selbstkorrekturen, die jedoch alle weiterer experimenteller Bestätigung bedürfen:

- Wird eine Selbstkorrektur bei einem aufgetretenen Fehler überhaupt ausgeführt, d.h. wird der Fehler erkannt? Wenn er erkannt wird: Wie lange dauert es, bis der Alarm des Monitors gegeben wird? (Unterbrechungszeitpunkt)
- Wie stark unterstützt der Sprecher mit der Realisierung seiner Korrektur beim Hörer die Revision der bereits ausgeführten Analyse? (Editierterm, Wohlgeformtheit)

### 3.3.5 Wiederholungen

#### 3.3.5.1 Abgrenzung des Phänomens: Wiederholungen

Der Begriff der Wiederholung ist bereits intuitiv klar: Bestimmte Bestandteile einer Äußerung werden mehrfach artikuliert. Unter einen weiten Begriff der Wiederholung fallen inhaltliche Wiederholungen (etwa durch Paraphrasierung), die wörtliche Wiederholung von Äußerungsteilen der vorangehenden Partneräußerung und wörtliche Wiederholungen innerhalb eines Satzes ein und desselben Sprechers. Paraphrasierungen und Wiederholungen von Material aus der Partneräußerung müssen der textsemantischen Ebene[53] zugeordnet werden und sollen daher mit der Darstellung der semantischen Merkmale behandelt werden (Abschnitte 3.4.3 bzw. 3.5.5). Der Begriff *Wiederholung* wird daher hier auf das syntaktische Phänomen der *wörtlichen Wiederholungen innerhalb eines Satzes* eingegrenzt. Die folgende Definition orientiert sich an Wiese (1983):

(3.25) Eine **Wiederholung** ist ein unverändert reproduziertes Textsegment von beliebiger Länge. Zwischen erstem und zweitem Vorkommen des Segments befindet sich höchstens eine stille und/oder gefüllte Pause. Die Reproduktion dient nicht der Bedeutungsdifferenzierung.

Wiederholungen können demnach unterschiedlich komplex sein: Wiederholungen von Silben (Bsp. 3.26), Fragmenten und Worten (3.27) treten ebenso auf wie Wiederholungen von Wortfolgen (3.28, 3.29).

[53] Äußerungs- und damit Satzgrenzen werden überschritten.

(3.26) Fol- <P> folgende Panne. (Ma001, Dialog3)

(3.27) Ja, alle <P> alle durchschauen. (Me034, Dialog3)

(3.28) Ich hab' noch <P> ich hab' noch 'ne Frage überhaupt. (Me006, Dialog1)

(3.29) Überprüf mal das mit 'm <D> mit 'm Telefon. (Me056, Dialog1)

#### 3.3.5.2 Wiederholungen und kognitive Belastung

Bromme und Wehner (1987) weisen auf die verzögernde Funktion hin, die Wiederholungen innerhalb einer Äußerung haben können. Sie können damit funktional gleichwertig zu stillen und gefüllten Pausen eingesetzt werden. Jou und Harris (1992) stellen fest, daß Wiederholungen von Wortfolgen eine Art Wiederaufnahme darstellen, deren Zweck es ist, für die Fortführung einer durch eine Zusatzaufgabe unterbrochenen Äußerung 'den Faden wieder aufzunehmen'[54]. Dieses Vorgehen werten sie als ein Indiz für vorübergehend außerordentlich hohe kognitive Belastung.

Eine dritte Erklärung für die Entstehung von Wiederholungen gibt Levelt (1983) an: Wiederholungen können demnach dann entstehen, wenn vom Sprecher eine Selbstkorrektur eingeleitet, ihre Ausführung jedoch anschließend wieder verworfen wurde. Demnach wären Wiederholungen vor allem ein Symptom für die Funktion des Monitoring-Prozesses. Wiederholungen zeugen in diesem Fall für die Unsicherheit des Sprechers bezüglich seiner Planungsresultate. Levelts Sichtweise schließen sich auch Postma et al. (1990) an. Nach ihren Untersuchungsergebnissen ist die Häufigkeit von Wiederholungen (als Untergruppe der Verzögerungsphänomene) mit der Häufigkeit von Selbstkorrekturen positiv korreliert.

Weitere Arbeiten, in denen Wiederholungen im Zusammenhang mit kognitiver Belastung als abhängige Variable erhoben wurden, sind in Tabelle 3.14 aufgeführt.[55] Unter Belastung wird meist ein signifikanter Anstieg der Anzahl von Wiederholungen registriert.

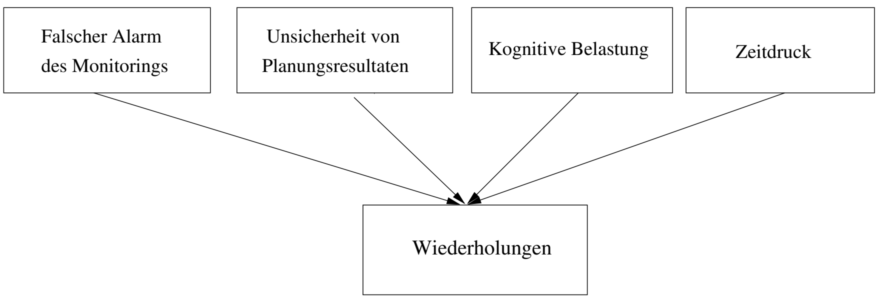


Abbildung 3.11: Einflüsse für das Entstehen von Wiederholungen

[54]Bei Jou und Harris heißen diese Wiederholungen *retracings*.

[55]Die Untersuchung von Postma et al. (1990) wurde trotz ihrer methodischen und detaillierten Analyse nicht in die Tabelle aufgenommen, da es sich in ihrer Experimentalaufgabe lediglich um eine Nachsprechaufgabe handelte.

| Autoren | Methode | VPn | Aufgabe | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Bromme und Wehner (1987)[=] | Rein sprachliche Aufgaben mit unterschiedlicher außersprachlicher Planung | 12 | Zusammenstellung einer Speisekarte (Aufgabe) | Denksportaufgabe (Problem) | 11.7[a], $(SD = 5.11)$ | 12.0[a] $(SD = 8.1)$ | n.s. | (+) |
| Deese (1980)[=] | Sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorherige Planung der Makrostruktur möglich (referatartiges Konzept) | Keine vorherige Planung der Makrostruktur | 2.62/100 Worte | 3.67/100 Worte | sign. (k.A.)[b] | + |
| Duez (1982)[=] | Quasiexperimentell: sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorbereitete Interviews | Unerwartete Interviews | 0.013 Wdh./s[c] | 0.023 Wdh./s[c] | k.A. | (+) |
| Grosjean und Deschamps (1973)[=] | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Interviews und Beschreibungen | 30/20[d] | Radio-Interviews | Cartoon-Beschreibungen | 6.84 % aller gefüllten Pausen | 7.97 % aller gefüllten Pausen | k.A. | (+) |
| Jou und Harris (1992)[=] | Vorlesen eines Zeitungsartikels mit anschließender Inhaltswiedergabe | 52 | Einfache Inhaltswiedergabe | Inhaltswiedergabe, begleitet von Kopfrechenaufgaben (mental arithmetic) | 1.29[e] | 2.29[e] | $p < 0.01$ | + |
| Wiese (1983)[=] | Rein sprachliche Aufgaben mit unterschiedlichem Planungsaufwand: Nacherzählung von Cartoons | 32 | Sprechen in der Muttersprache | Sprechen in einer Fremdsprache | 0.74/100 Silben, $(SD = 0.92)$ | 2.2/100 Silben, $(SD = 2.37)$ | $p < 0.05$ | + |

Tabelle 3.14: Untersuchungen zu Wiederholungen

[a] Der Wert ist '*an der Gesamtverbalisationsmenge relativiert*'. Eine Verbalisationseinheit ist nicht angegeben.

[b] Angabe nur für Sammelindex 'Type II disfluencies' (Wiederholungen, Korrekturen, Fehlansätze)

[c] Duez gibt Rohdaten an: In 22'20" wurden in der unbelasteten Bedingung 18 Wiederholungen gezählt, in der belasteten in 13'08" ebenfalls 18.

[d] Kontrollgruppe/Experimentalgruppe

[e] Relativierte Anzahl von Wiederholungen. Eine Basiseinheit für die Relativierung ist nicht angegeben.

| **Autoren** | **Methode** | **VPn** | **Aufgabe** | | **Werte** | | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Zeitdruck** | | | | | | | | |
| Marx (1984)[=] | Führen eines Gespräches mit Ziel der Überzeugung eines Partners | 70 | Ohne Zeitdruck | Zeitdruck durch Wiederholung einer bereits ohne Zeitdruck ausgeführten Tätigkeit mit Zeitbegrenzung | 0.49/100 Silben, ($SD = 0.54$) | 1.13/100 Silben, ($SD = 1.34$) ; [0.58/100 Silben, ($SD = 0.68$)(NZD)[a]; 1.34/100 Silben, ($SD = 1.51$)(HZD)] | $p = 0.017$, ($t = 2.45$); [n.s. (NZD); $p = 0.005$, ($t = 3.14$) (HZD)] | + |
| Marx (1984)[=] | Lautes Denken über ein Problem | 70 | Ohne Zeitdruck | Zeitdruck durch Wiederholung einer bereits ohne Zeitdruck ausgeführten Tätigkeit mit Zeitbegrenzung | 0.61/100 Silben, ($SD = 0.88$) | 0.72/100 Silben, ($SD = 1.02$) ; [0.91/100 Silben, ($SD = 1.13$)(NZD)[a]; 0.62/100 Silben, ($SD = 0.97$)(HZD)] | n.s. | n.s. |

Tabelle 3.15: Untersuchungen zu Wiederholungen (Fortsetzung)

[a] Postexperimentell erfolgte eine Zuordnung der Probanden zu den Gruppen 'hohes Zeitdruckempfinden' (HZD) bzw. 'niedriges Zeitdruckempfinden' (NZD) durch Selbsteinschätzung.

#### 3.3.5.3 Wiederholungen und Zeitdruck

Die Untersuchung von Marx (1984) liefert als Ergebnis eine höhere Anzahl von Wiederholungen unter Zeitdruck. Dieser Effekt ist konsistent mit der von Levelt verwendeten Sichtweise von Wiederholungen als Selbstkorrekturen: Wenn ein Sprecher zunächst dazu ansetzt, seine Äußerung zu korrigieren, die Korrektur schließlich aber doch wieder verwirft, deutet das zumindest darauf hin, daß der Sprecher Unsicherheit bezüglich der Planungsresultate seiner Sprechplanung empfindet. Marx verweist darauf, daß gerade unter Zeitdruck nur suboptimale Planungserfolge erzielt werden, da Sprecher mehr 'ad hoc' planen und nicht alle Bedingungen für die Gültigkeit und Anwendbarkeit ihrer Planungsergebnisse überprüfen. Diese Überprüfungsfunktion müßte für Resultate der Sprachproduktion vom Monitoring-Prozeß übernommen werden. Führt er die Überprüfung zu einem späteren Zeitpunkt aus (während die konzeptualisierten Einheiten bereits von der Artikulationsphase bearbeitet werden), könnte sich die Überprüfung und erneute Freigabe in Form von Wiederholungen äußern.

Marx favorisiert jedoch eine ähnliche Erklärungshypothese für Wiederholungen an Satz- und Teilsatzgrenzen, wie Jou und Harris (1992): Sie dienen dazu, eine Verknüpfung zwischen 'Gesagtem und noch zu Sagendem' herzustellen und die Information des 'Gesagten' nicht zu verlieren. Insofern würden sie einer Intensivierung der Informationsspeicherung dienen, die notwendig wird, wenn Zeitdruck die Informationsspeicherung stört. Diese Rückführung auf Kapazitätsprobleme ist ein erneuter Beweis dafür, in welch großer Nähe Zeitdruck und kognitive Belastung zueinander stehen.

#### 3.3.5.4 Fazit

Wiederholungen stellen einen außerordentlich gut untersuchten sprachlichen Parameter mit einheitlicher Veränderungstendenz unter kognitiver Belastung dar. Das empirische Fundament für den Zusammenhang von Wiederholungen mit Zeitdruck ist hingegen nur sehr dünn. Wiederholungen sollten daher vor allem als Parameter bei der Modellierung kognitiver Belastung Berücksichtigung finden. Für den Zusammenhang mit Zeitdruck muß, vor allem auch wegen der widersprüchlichen Ergebnisse der angegebenen empirischen Untersuchung auf die Notwendigkeit weiterer Forschungsarbeit hingewiesen werden. Ein Verzicht auf den Einsatz des Parameters ist vorerst naheliegend.

### 3.3.6 Satzfragmente

#### 3.3.6.1 Abgrenzung des Phänomens: Satzfragmente

Satzfragmente stellen eine besondere Ausprägung unvollständiger syntaktischer Strukturen dar: Begonnene Sätze werden während der Artikulation abgebrochen und nicht vervollständigt. In speziellen Fällen erfolgt eine Korrektur der Semantik des abgebrochenen Satzes durch den nachfolgenden Satz (siehe Abschnitte 3.4.1 und 3.3.4). Ein Beispiel für ein Satzfragment ist in 3.31 angegeben zusammen mit einer nach Definition 3.30 möglichen Fortsetzung in Klammern.

(3.30) Ein **Satzfragment** ist eine unvollständige syntaktische Struktur $U$, für die eine syntaktische Fortsetzung $F$ existiert, so daß $UF$ einen wohlgeformten Satz bildet. Nach $U$ endet entweder die gesamte Äußerung oder der Sprecher beginnt einen neuen Satz.[56]

(3.31) Die Leitung, warum hört denn die ... (Me016, Dialog 1)
... auf zu funktionieren.

#### 3.3.6.2 Satzfragmente und kognitive Belastung

Satzfragmente verletzen die kommunikative Forderung, wohlgeformte Äußerungen zu bilden, und das entsprechende Bestreben eines Sprechers, fehlerfrei zu sprechen (Clark & Clark, 1977), auf besonders deutliche Art und Weise. Bricht ein Sprecher einen Satz mitten in dessen Generierung ab, können dafür mehrere Gründe verantwortlich sein: Zum einen besteht die Möglichkeit, daß der Sprecher die noch nicht geäußerte, aber bereits geplante sprachliche Information, die den Satz vervollständigt hätte, als nicht mehr adäquat erachtet. In diesem Fall ist eine Korrektur der Äußerung zu erwarten und es entsteht die komplexere syntaktische Konstruktion eines Fehlansatzes, die weiter unten untersucht wird (3.4.1). Im Gegensatz dazu stehen Unterbrechungen des gesamten Sprachproduktionsprozesses durch einen erfolglosen Versuch, ein den Satz fortführendes Lemma abzurufen (Jou & Harris, 1992). Auch eine plötzliche Allokation der für die Sprachproduktion eingesetzten kognitiven Ressourcen für eine Nebenaufgabe kann der Anlaß für einen Abbruch sein (Beispiel 3.31).

In den Experimenten in Tabelle 3.16 wurde unter Belastung eine erhöhte Anzahl von Satzfragmenten gemessen.

#### 3.3.6.3 Fazit

Wegen der außerordentlich schwerwiegenden Eingriffe in den Sprachproduktionsprozeß, die Anlaß für den Abbruch eines Satzes sein können, sollten Satzfragmente als außerordentlich wichtige Symptome kognitiver Belastung dienen können. Ein Zusammenhang mit Zeitdruck konnte aus der psychologischen Literatur nicht entnommen werden.

## 3.4 Merkmale von Syntax und Semantik

Dieser Abschnitt nimmt Phänomenbereiche auf, die weder rein syntaktischer noch rein semantischer Natur sind. Nun haben semantische Entscheidungen *immer* auch irgendeine Auswirkung auf die Syntax und umgekehrt. Hier sollen jedoch vor allem solche Merkmale von Äußerungen behandelt werden, bei denen Realisierungen der syntaktischen Ebene immer einer Effektklasse auf der semantischen Seite zugeordnet werden können und umgekehrt. Z.B. ist die Menge syntaktischer Ausdrücke, die anaphorische Referenzen realisieren, recht klar abgegrenzt. Ihre Entsprechung auf der Seite der Semantik sind unterspezifizierte Referenten. Diese hinterlassen nach der Semantikkonstruktion Leerstellen in der semantischen Struktur, welche allein mit dem semantischen Wissen des Satzes nicht gefüllt werden können.

[56] Diese Bedingung unterscheidet Satzfragmente von der Originaläußerung einer Selbstkorrektur.

| Autoren | Methode | VPn | Aufgabe | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|
| | | | Kontrollgruppe | Experimentalgruppe | Kontrollgruppe | Experimentalgruppe | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Bromme und Wehner (1987)$^{=}$[a] | Rein sprachliche Aufgaben mit unterschiedlicher außersprachlicher Planung | 12 | Zusammenstellung einer Speisekarte (Aufgabe) | Denksportaufgabe (Problem) | 4.5[b] | 5.1[b] | n.s. | (+) |
| Jou und Harris (1992)$^{=}$ | Vorlesen eines Zeitungsartikels mit anschließender Inhaltswiedergabe | 52 | Einfache Inhaltswiedergabe | Inhaltswiedergabe, begleitet von Kopfrechenaufgaben (mental arithmetic) | 0.4[c] | 1.5[c] | $p < 0.01$ | + |
| Rummer (1996a)$^{=}$, Exp. 1 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer diskontinuierlichen Doppelaufgabe (Bewertung aufeinanderfolgender Lichtsignale auf farbliche Gleichheit) | 0.0037 /s (Polizist), 0.0051 /s (Nachbar) | 0.0100 /s (Polizist), 0.0142 /s (Nachbar) | $p < 0.01$, ($F(1, 54) = 9.58$) | + |
| Rummer (1996a)$^{=}$, Exp. 2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer kontinuierlichen Doppelaufgabe (Tracking- Aufgabe) | 0.004 /s (Polizist), 0.005 /s (Nachbar) | 0.022 /s (Polizist), 0.019 /s (Nachbar) | $p < 0.001$, ($F(1, 54) = 16.2$) | + |

Tabelle 3.16: Untersuchungen zu Satzfragmenten

[a] Fehlansätze werden mit Satzfragmenten zusammengefaßt.
[b] Der Wert ist '*an der Gesamtverbalisationsmenge relativiert*'. Eine Verbalisationseinheit ist nicht angegeben.
[c] Relativierte Anzahl von Fragmenten. Eine Basiseinheit für die Relativierung ist nicht angegeben.

## 3.4.1 Fehlansätze auf Satzebene

### 3.4.1.1 Abgrenzung des Phänomens: Fehlansatz

Fehlansätze bilden eine Unterklasse der Selbstkorrekturen, werden jedoch wegen der zahlreichen, speziell auch dieses Phänomen untersuchenden experimentellen Arbeiten hier in einem Extra-Abschnitt behandelt.

Der Terminus *Fehlansatz* und sein englisches Pendant *false start* werden in der psychologischen Literatur auf verwirrende Weise gebraucht: Wiese (1983), Roßnagel (1995a) und Kowal (1991) beziehen sich damit lediglich auf den fehlerhaften Teil einer Selbstkorrektur beliebiger Klasse, d.h. auf die Originaläußerung mit dem Reparandum. Roßnagel (1995b) dagegen gibt als Definition[57] '*abgebrochene Äußerungsansätze, die in keiner syntaktisch oder semantisch definierbaren Relation zu den anderen Äußerungen des Textes stehen*'. Wenn Roßnagel damit meint, daß nicht versucht wird, die konzeptuelle Struktur des Ansatzes in einer alternativen syntaktischen Realisierung neu zu formulieren (sondern daß sie verworfen wird), dann entsprechen Fehlansätze einer Untermenge der D-Repairs bei Levelt (1983)[58]: '*The speaker realizes that another idea than the current one has to be expressed first and interrupts his speech to start anew.*'. Auch Jou und Harris (1992) vertreten diese Sichtweise indirekt über die Abgrenzung von Satzfragmenten: '*Fragments can be easily distinguished from false starts. In the case of false starts, a prompt and forceful correction was usually made; this did not occur in the case of fragments.*'. Für Deese (1980) fallen die beiden Klassen Fehlansatz und Satzfragment sogar zusammen: er unterscheidet lediglich *false starts on sentences* im Gegensatz zu *corrections internal to sentences*. Duez (1982) definiert Fehlansätze als '*any sequence of segments that is intended to start the next utterance but is interrupted and replaced by another that will get completed.*' .

Die in 3.3.4 eingeführte Terminologie ist ausreichend, um die von Wiese (1983), Roßnagel (1995a) und Kowal (1991) verwendete Definition abzudecken. Da außerdem für Satzfragmente bereits eine eigene Definition gegeben wurde (Abschnitt 3.3.6), wird der Begriff *Fehlansatz* hier mit der Bedeutung von Definition 3.32 definiert, die mit Roßnagel (1995b), Levelt (1983) und Jou und Harris (1992) vereinbar ist.

(3.32) Ein **Fehlansatz** ist eine Sequenz

$<$ Satzfragment, Korrektursignal, Reparatur $>$,

wobei als Korrektursignal Korrekturterme beliebiger Komplexität[59] in Frage kommen. Reparatur-Segmente können nur vollständige Sätze oder weitere Fehlansätze sein. Die Forderung der Satz-Eigenschaft für die Reparatur unterscheidet Fehlansätze von den restlichen Selbstkorrekturen.

Fehlansätze *bedingen* ein vorausgehendes Satzfragment. Die Beurteilung, ob der neu begonnene Satz eine erneute Realisierung der gleichen Inhalte darstellt oder völlig neue Ideen

[57] Diese Definition entspräche nach Roßnagels Angaben angeblich auch Kowal (1991); Kowals Definition deckt jedoch m.E. mehr Phänomene ab.

[58] Siehe Abschnitt 3.3.4.

[59] Das bedeutet, daß von prosodischen Merkmalen des nachfolgenden Materials bis zu kompletten Sätzen ('Ach nee, halt, ich erzähl' Quatsch.') beliebige Konstruktionen möglich sind.

vermittelt, ist schwierig. Jou und Harris (1992) sehen als Symptom für die Absicht eines erneuten Versuches mit gleicher Semantik vor allem die Länge der Pause, die dem Satzfragment folgt: '*Subjects often paused in an 'uh...' or 'uhm' for a long time, [...] finally giving up the unfinished sentence and starting a new sentence.*' Sie führen dieses Verhalten darauf zurück, daß der Sprecher Probleme beim lexikalischen Abruf hat und daher weniger den begonnenen Satz nicht beenden *will*, sondern ihn nicht beenden *kann*.

Einige Beispiele sollen den Unterschied von Fehlansätzen und einfachen Satzfragmenten deutlicher machen[60]:

(3.33) In der Mitte ist ein großes rundes Ding, so 'n ... Ach nee, halt. Weiter rechts ist 'ne orangene Kappe.
(Fehlansatz: Löschung der Semantik des Fragments beabsichtigt)

(3.34) In der Mitte ist ein großes rundes Ding, so 'n ... Gucken Sie mal, ob sie links den Ölmeßstab sehen.
(Fehlansatz mit erneutem Versuch, aus einer 'lexikalischen Sackgasse' herauszukommen.)

(3.35) In der Mitte ist ein großes rundes Ding, so 'n ... – Die Verbindung ist schon wieder weg.
(Satzfragment: Fortführung des Dialoges ohne Löschung der Semantik des Fragments)

An Fehlansätzen als Vertreter der D-Kategorie der Selbstkorrekturen wird der Unterschied zwischen Makro- und Mikroplanung besonders deutlich: Bei einer D-Korrektur wird die *gesamte* konzeptuelle Struktur verworfen und neu geplant, bei allen anderen Formen der Selbstkorrektur dagegen werden *nur Bestandteile* einer bereits bestehenden syntaktischen oder semantischen Struktur eines Satzes ausgetauscht.

Problemfälle für die Unterscheidung zwischen Fehlansätzen und den anderen Klassen von Selbstkorrekturen sind Korrekturen innerhalb der ersten Phrase eines Satzes: Bei Korrekturen wird aufgrund der Wohlgeformtheitsregel maximal bis zur zuletzt begonnenen Phrase zurückgegangen, d.h. es erfolgt keine Wiederaufnahme *vor* dem Beginn der zuletzt begonnenen Phrase. Die Möglichkeit der Koordination, wie in 3.24 gefordert, wäre ansonsten nicht mehr gegeben. Da der Beginn des Satzes unkorrigiert bleibt, ist erkennbar, daß es sich nur um die Korrektur einer Teilkonstruktion handelt. Erfolgt der Abbruch jedoch innerhalb der ersten Phrase eines Satzes, ist ein solcher unveränderter Satzbeginn noch nicht verfügbar, da der Satzbeginn gleichzeitig die korrigierte Originaläußerung darstellt. Eine Zuordnung ist in diesen Fällen meist offen. Die charakteristische Form bestimmter Korrekturterme, die lediglich im Zusammenhang mit Fehlansätzen eingesetzt werden, könnte einer der möglichen Indikatoren sein, die den Unterschied feststellen helfen.

#### 3.4.1.2 Fehlansätze und kognitive Belastung

Levelt bezieht sich mit dem Namen *Difference-* Repairs explizit auf die Veränderung der *konzeptuellen* Äußerungsgrundlage. Ihr Entstehen wird vor allem auf Probleme bei der Linearisierung sprachlichen Materials auf Makroniveau zurückgeführt. Eine andere, von Levelt nicht

[60]Die Beispiele sind künstlich, da vergleichbare Äußerungen für alle drei Ebenen in der Voruntersuchung zu READY nicht zu finden waren.

erwähnte Ursache können jedoch auch Revisionen der für die Äußerung selegierten Informationsmenge sein. Sowohl die Informationsselektion als auch ihre Linearisierung sind Subprozesse der Makro- und Mikroplanung. Sie werden daher in der Phase der Konzeptualisierung ausgeführt. Eine optimale Informationsselektion und -linearisierung der Makro- und Mikropropositionen kann in der Konzeptualisierungsphase nur dann erfolgen, wenn genügend Zeit und kognitive Kapazität dafür zur Verfügung steht. Wieviel von diesen beiden Ressourcen dabei jeweils benötigt wird, ist u.a. von der Gesamtmenge an selegierbaren Informationen und dem Typ der sprachlichen Äußerung abhängig. Unter kognitiver Belastung bzw. Zeitdruck steht jedoch die notwendige Planungszeit bzw. -kapazität dafür ggf. nicht zur Verfügung.

Diese Hypothese wird durch die in Tabelle 3.17 angegebenen Arbeiten bestätigt, da sie unter beiden Bedingungen eine Erhöhung der Zahl an Fehlansätzen beobachten.

Roßnagel (1995b) stellt im Gegensatz zu allen anderen Arbeiten in Tabelle 3.17 eine geringere Zahl an Fehlansätzen unter kognitiver Belastung fest. Er erklärt das dadurch, daß unter Belastung nicht mehr die Möglichkeit besteht, fortlaufend die kommunikative Angemessenheit von Äußerungen zu überprüfen. Das bedeutet, daß die Funktion des Monitoring-Prozesses unter kognitiver Belastung verringert wird und weniger Korrekturen eingeleitet werden. Gleichzeitig stellt Roßnagel aber auch fest, daß ein hoher Anteil von Fehlansätzen generell typisch sei für spontanes Sprechen mit hohem Planungsanteil, d.h. daß erhöhte Planungsanforderungen *mehr* Fehlansätze verursachen. In der belasteten Gruppe seines Experimentes ist aber eine solche höhere Belastung durch mehr konzeptuelle Planung gegeben. Außerdem widerspricht Roßnagels Erklärungsversuch für die geringere Anzahl von Fehlansätzen auch seinen Ergebnissen für Selbstkorrekturen, deren Zahl sich unter Belastung erhöht hat. Auch die Ausführung von Selbstkorrekturen setzt die Beurteilung kommunikativer Angemessenheit voraus und es ist fraglich, warum ein kognitiver Prozeß, der diese Beurteilung durchführt, das ausschließlich für Konstituenten unterhalb des Satzniveaus tun sollte. Die Erklärung einer verringerten Anzahl von Fehlansätzen durch verringertes Monitoring erscheint daher nicht überzeugend. Möglicherweise sind auch hier vielmehr Strategien der kognitiven Verarbeitung ausschlaggebend für den Experimentaleffekt.

#### 3.4.1.3 Fehlansätze und Zeitdruck

Marx (1984) weist darauf hin, daß unter Zeitdruck die Handlungsplanung beschleunigt und daher nur ungenau ausgeführt wird. Anwendbarkeitsbedingungen von Operatoren bzw. ihre Nebenwirkungen können dann nicht ausreichend und früh genug berücksichtigt werden. Die entstehenden Resultate sind entsprechend ungenügend und bei offen auftretenden Konflikten (die bei der Sprachproduktion vom Monitoring-Prozeß erkannt werden müssen) einer Revision zu unterziehen. Marx erhebt zwar Fehlansätze nicht als abhängige Variable, jedoch lassen sich die Vorhersagen, die sie für das Ansteigen von Sprechfehlern und Selbstkorrekturen im allgemeinen Fall macht, zumindest theoretisch auch auf die Klasse der Fehlansätze übertragen.

#### 3.4.1.4 Fazit

Eine Verwendung von Fehlansätzen als Indikator außerordentlich hoher kognitiver Belastung sollte trotz Roßnagels widersprüchlicher Untersuchung möglich sein. Gründe für die Unterbre-

| **Autoren** | **Methode** | **VPn** | **Aufgabe** Kontrollgruppe | Experimentalgruppe | **Werte** Kontrollgruppe | Experimentalgruppe | **Signifikanzniveau** | **Ten.** |
|---|---|---|---|---|---|---|---|---|
| **Arbeitsgedächtnisbelastung** | | | | | | | | |
| Bromme und Wehner (1987)$^{=}$[a] | Rein sprachliche Aufgaben mit unterschiedlicher außersprachlicher Planung | 12 | Zusammenstellung einer Speisekarte (Aufgabe) | Denksportaufgabe (Problem) | 4.5[b], $(SD = 2.5)$ | 5.1[b], $(SD = 3.4)$ | n.s. | $(+/+)$[c] |
| Deese (1980)$^{=}$ | Sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorherige Planung der Makrostruktur möglich (referatartiges Konzept) | Keine vorherige Planung der Makrostruktur | 1.47/100 Worte | 2.24/100 Worte | sign. (k.A.) | + |
| Duez (1982)$^{=}$ | Quasiexperimentell: sprachliche Aufgaben mit unterschiedlichem Planungsaufwand | 10 | Vorbereitete Interviews | Unerwartete Interviews | 0.012 Fehlansätze/s[d] | 0.023 Fehlansätze/s[d] | k.A. | (+) |
| Roßnagel (1995a)$^{=}$ | Sprachliche Instruktion von Karatekas durch *ungeübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | 5.0 % aller Äußerungseinheiten | 10.7 % aller Äußerungseinheiten | $p < 0.001$, $(F(1,39) = 15.24)$ | + |
| Roßnagel (1995b)$^{=}$ | Beschreibung eines Maschinenmodells gegenüber 2 versch. Hörergruppen (A= Kind/B= Student) | 45 | Präsenz des Modells | Abwesenheit des Modells während der Beschreibung (Gr.1) bzw. Zweitaufgabe Memorieren einer Zahlenfolge (Gr.2) | 15.9 % aller Äußerungseinheiten (Gr. A); 15.3 % (Gr. B) | 6.1 % aller Äußerungseinheiten (Gr.1A), 6.5 % (Gr.2A); 5.7 % (Gr.1B), 5.9 % (Gr.2B) | $p < 0.001$, $(F(1,33) = 56.21)$ | – |

Tabelle 3.17: Untersuchungen zu Fehlansätzen

[a] Fehlansätze und Satzfragmente werden zusammengefaßt.
[b] Der Wert ist 'an der Gesamtverbalisationsmenge relativiert'. Eine Verbalisationseinheit ist nicht angegeben.
[c] Die Anpassung erfolgte abhängig vom Individuum in unterschiedliche Richtungen.
[d] Duez gibt Rohdaten an: In 22'20" wurden in der unbelasteten Bedingung 16 Fehlansätze gezählt, in der belasteten in 13'08" dagegen 17.

chung und Korrektur kompletter Sätze wurden bereits in Abschnitt 3.3.6 aufgezählt. Auch hier stellt sich jedoch die Frage, ob nicht differenzierende Situationsparameter helfen könnten, die unterschiedliche Formen kognitiver Belastung zu unterscheiden und dadurch die verschiedenen Richtungen der experimentellen Effekte zu klären. Zusammenhänge von Fehlansätzen mit Zeitdruck ließen sich lediglich theoretisch ableiten.

### 3.4.2 Ellipsen und Anaphern

Sprecher beziehen sich im Diskurs auf zuvor bereits realisiertes sprachliches Material mithilfe geeigneter sprachlicher Referenzmechanismen. Die Möglichkeiten der Referenz gehen von Wiederholungen über Paraphrasierungen oder gekürzte Paraphrasierungen bis zu Anaphern und Ellipsen. Motivation für die Auswahl eines geeigneten Referenzmechanismus können neben dem Ziel effizienter Sprachproduktion verschiedene Sprecherziele wie *Verdeutlichung der thematischen Fortsetzung, rhetorische Erwiderung* oder *Höflichkeit* sein. Jedoch auch arbeitsgedächtnisbezogene Gründe kommen dafür in Frage: Die Verwendung eines Lemmas oder einer Konstruktion aufgrund hoher Aktivierung im Arbeitsgedächtnis reduziert die Generierungsanforderungen durch Verwendung soeben geparsten (und damit noch aktivierten oder anderweitig verfügbaren) sprachlichen Materials.

Der generierungsseitigen Effizienz durch Wiederverwendung soeben geparster Strukturen beim Sprecher steht die Forderung nach Verarbeitungseffizienz beim Hörer (und damit auch die Gricesche Quantitätsmaxime) entgegen: Was bereits bekannt ist, muß nicht wiederholt werden. Als Resultat der Verfügbarkeit durch hohe Aktivierung im Arbeitsgedächtnis sowohl des Hörers *als auch des Sprechers* könnten Referenten und Relationen als inferierbar angesehen und daher vom Hörer in seiner eigenen Äußerung ausgespart werden oder nur mit einer Form der Kurzreferenz erneut verwendet werden. Die Bildung von Ellipsen und Anaphern wäre die Folge.

#### 3.4.2.1 Experimentell identifizierte Bedingungen für Anaphern- und Ellipsenbildung

Die Untersuchung von Levelt und Kelter (1982) zielt vorrangig darauf ab, die Bedingungen zur Wiederaufnahme syntaktischen Materials des Sprechers durch den Hörer zu untersuchen. Da auf diesem Gebiet bisher nur wenige Autoren gearbeitet haben und daher nur wenige Vorarbeiten vorliegen, beziehen sich die von Levelt und Kelter ausgeführten Experimente lediglich auf die Wortebene: Untersucht wurde die Entsprechung von Frage und Antwort bezüglich einer in der Frage vorhandenen oder nicht vorhandenen Präposition. Ein interessanter Nebeneffekt dieser Arbeit betrifft die Bewertung der Elliptizität von Äußerungen unter den gleichen zwei Bedingungen.

Die Versuchspersonen mußten Referenten in der thematischen Rolle des Instruments oder des Patiens einer in der Frage enthaltenen Relation auf vorgegebenen Bildern identifizieren.[61] Eines der Experimente zeigte, daß sich die Anzahl der nicht elliptischen Äußerungen unter Bedingungen zusätzlicher Belastung durch zwischen Frage und Antwort liegendes weiteres sprachliches Material gegenüber der Kontrollbedingung um etwa die Hälfte verringerte. Die

[61]Eine Frage lautete z.B. 'Wen läßt Peter seine Violine sehen?'

Antworten wurden jedoch auch insgesamt kürzer.

Ein weiteres Experiment führten Levelt und Kelter in einer dem System READY verwandten Experimentalanordnung aus: Die Inhaber von 228 Läden in der Nijmegener Innenstadt wurden telefonisch nach den Öffnungszeiten ihres Geschäfts am Samstag befragt. Bei der Hälfte der Anrufe wurde durch Präsentation zusätzlichen sprachlichen Materials unmittelbar nach der Frage wiederum eine Abblenkung herbeigeführt. In den meisten Antworten der Angerufenen, die die Präposition enthielten, wurde von den Probanden auch das Hauptverb realisiert und umgekehrt. Levelt und Kelter bezeichnen diesen Trend als 'Tendenz zu globaler Kürze'. Der Effekt war jedoch von der Menge ablenkenden Materials in der Frage unabhängig.

#### 3.4.2.2 Ellipsen und kognitive Belastung

Eine wichtige Erkenntnis aus den Experimenten von Levelt und Kelter ist folgende: Wenn zwischen einer Relation, die aufgrund eines vorhandenen Antezedenten eine Ellipsenbildung erlaubt, und diesem Antezedenten zu viel das Arbeitsgedächtnis belastendes sprachliches Material liegt, wird keine Ellipse gebildet, sondern ein vollständiger Satz formuliert. Die Umkehrung dieser Beobachtung läßt die Hypothese zu, daß bei einer möglichen, aber nicht realisierten Ellipsenbildung Faktoren vorliegen, die die Zugänglichkeit des Antezedenten im Diskurskontext reduzieren. Offenbleiben muß, ob der Verzicht auf die Ellipsenbildung durch den Sprecher hörerorientierte oder sprecherbezogene Gründe hat. Im ersten Fall würde der Sprecher dem Hörer durch volle sprachliche Realisierung der Relation die Interpretation seiner Äußerung erleichtern, im zweiten Fall wäre die naheliegendere Interpretation, daß der Sprecher die Ellipse nur mit erhöhtem kognitiven Aufwand generieren könnte und daher darauf verzichtet. Verschiedene Erkenntnisse über die Hörerorientierung unter kognitiver Belastung (Abschnitt 3.6) implizieren, daß die hörerbezogene Erklärung nicht haltbar ist.

Die Verwendung von Anaphern ist ein der Ellipsenbildung verwandter Prozeß. Solange keine gegenteiligen Erkenntnisse vorliegen, ist es daher adäquat anzunehmen, daß die gleichen Annahmen auch für die Anaphernbildung gemacht werden dürfen.

Die Verwendung voller sprachlicher Realisierungen von Relationen und Referenten in Kontexten, in denen Ellipsen- und Anaphernbildung möglich wäre, kann daher nicht uneingeschränkt als ein Symptom kognitiver Belastung gewertet werden.

#### 3.4.2.3 Ellipsen und Zeitdruck

In den Abschnitten 3.5.4 und 3.5.3 wird gezeigt, daß unter Zeitdruck bei Sprechern die Tendenz besteht, ihre Äußerungen knapp zu halten und nur die absolut notwendige Information zu liefern. Ob der Trend der globalen Kürze, den Levelt und Kelter beobachten, auch unter Bedingungen des Zeitdrucks gilt, ist nicht bekannt. Verwandte Effekte der Kürze wie das Auslassen zusätzlicher Information unter Zeitdruck machen jedoch diese Annahme wahrscheinlich. Zusätzlich wird unter Zeitdruck die Hörerorientierung eines Sprechers reduziert (3.6), was einen geringeren Einsatz von Mechanismen der Höflichkeit erwarten läßt. Die ausführliche Formulierung und Wiederverwendung von Partnermaterial, die von Levelt und Kelter (1982) als ein solcher Mechanismus identifiziert wurden, würde damit auch nur noch eingeschränkt zur Anwendung kommen. Da Untersuchungen zur Interaktion von Zeitdruck und Ellipsenbildung jedoch völlig fehlen und in den genannten Experimenten *kein* Zeitdruck vorlag, dürfen *fehlende Ellipsen* (also hohe Ausführlichkeit) vorerst lediglich als *Indikator fehlenden Zeitdrucks* interpretiert werden.

#### 3.4.2.4 Fazit

Ellipsen und Anaphern lassen derzeit vor allem aufgrund der fehlenden experimentellen Untersuchungen keine Aussagen über den Zusammenhang mit Zeitdruck und kognitiver Belastung zu. Die theoretisch abgeleiteten Vorhersagen, die den Einsatz von Ellipsen und Anaphern in einem Dialogsystem auf vorläufiger Basis möglich machen sollen, beruhen auf einem dünnen Fundament. Die Modellierung sollte daher entsprechend nur mit der gebotenen Vorsicht die zugehörigen Einflüsse berücksichtigen.

### 3.4.3 Echo-Äußerungen

#### 3.4.3.1 Abgrenzung des Phänomens: Echo-Äußerungen

Beun (1985) analysiert für Informationsdialoge (Frage-Antwort-Sequenzen) die Realisierung von Dialogakten, um die Ziele eines Dialogpartners abzuleiten.

(3.36) Ein **Dialogakt** $D$ ist ein Paar der Form

$$D = <\text{semantischer Inhalt, kommunikative Funktion}>$$

Wiederholungen von Konstituenten und Worten im Inneren von Sätzen und Phrasen *eines* Sprechers als rein syntaktisches Phänomen wurden in Abschnitt 3.3.5 untersucht. Ein Sprecher kann jedoch auch semantische Inhalte seiner eigenen Äußerung oder einer Partneräußerung wiederholen, indem er sie paraphrasiert oder syntaktisch ähnlich realisiert. Diese Wiederholungsphänomene wurden bisher nicht erfaßt, da sie mit der Sichtweise von Wiederholungen als Selbstkorrekturen nicht vereinbar sind. Greene et al. (1990) bezeichnet Äußerungen, die bereits vermittelte Informationen erneut versprachlichen, als *ideational repetitions*. Stammen die Äußerungen vom gleichen Sprecher innerhalb eines Dialogturns, führen sie zu Redundanz, die in einem eigenen Abschnitt untersucht wird (Abschnitt 3.5.5). Eine Klassifikation der unterschiedlichen Phänomene ist in Tabelle 3.18 dargestellt.

| **Sprecher/Lage** | **innerhalb eines Satzes** | | **außerhalb eines Satzes** | |
|---|---|---|---|---|
| | wörtlich | paraphrasiert | wörtlich | paraphrasiert |
| gleich | Wiederholung | Redundanz | Redundanz | |
| verschieden | – | | Echo-Äußerung | ? |

Tabelle 3.18: Abgrenzung der Echo-Äußerungen von anderen Wiederholungsphänomenen

Dialogakte, deren semantischer Inhalt identisch ist mit dem semantischen Inhalt der vorangehenden *Partner*äußerung (bzw. bei Partneräußerungen, die aus mehreren Propositionen bestehen, mit einer dieser Propositionen) sind ein Spezialfall inhaltlicher Wiederholungen. Eine besondere Form liegt dann vor, wenn ihre syntaktische Realisierung der syntaktischen Konstruktion der Partneräußerung (ggf. mit geringfügigen Anpassungen) entspricht. Dieses Phänomen soll im folgenden als *Echo-Äußerungen* bezeichnet werden.

(3.37) Die syntaktische Realisierung eines Dialogaktes $D_1$ heißt **Echo-Äußerung**, wenn $D_1$ ein Dialogakt $D_0$ des Dialogpartners unmittelbar vorangeht, so daß gilt:
$I_0 = Inhalt(D_0)$, $I_1 = Inhalt(D_1)$, $I_0 = I_1$ und die syntaktischen Realisierungen von $D_0$ und $D_1$ sind ähnlich.

(3.38) Dialog 1:
**Me039:** Die ist eigentlich nicht lose . Da kann nix rauskommen .
**Me040:** Is - <D> Da kann nichts rauskommen. Dann ...

(3.39) Dialog 5:
**Me013:** Der Waschwasserbehälter [ist ein] weisser durchsichtiger Behälter mit 'm roten Deckel obendrauf.
**Fr014b:** Durchsichtiger %Behälter mit %rotem Deckel oben drauf .

#### 3.4.3.2 Echo-Äußerungen und kognitive Belastung

Rummer et al. (1995) arbeiten vier Steuerungsmechanismen für die sprachliche Makroplanung heraus. Diese werden von Sprechern vor allem als Anpassung an besondere situationelle Parameter eingesetzt. Durch den Einsatz bestimmter Steuerungsmechanismen kann aber auch eine Anpassung an die eigene kognitive Belastung erfolgen: Besonders die Reizsteuerung als flexibler, kognitiv wenig beanspruchender Mechanismus ist in vielen Situationen zu beobachten, wenn die eigene kognitive Belastung die Makroplanung in sehr starkem Maße behindert. Bei diesem Steuerungsmechanismus werden Reize aus dem Sprachsignal des Dialogpartners zur Generierung eigener Gesprächs'beiträge' verwendet.[62] Die Sprachgenerierung erfolgt stark *schematisiert*. In Situationen extrem hoher Belastung kann durch Einsatz der Reizsteuerung zumindest formal ein Gespräch aufrechterhalten werden, indem der Erwartung der *Erwiderung* im Dialog nachgekommen wird. Rummer et al. (1995) vergleichen diesen Modus der Dialogführung mit der, die im System ELIZA (Weizenbaum, 1966) umgesetzt wird: Die Reaktion erfolgt lediglich auf dem vorgegebenen Sprachsignal, aber nicht auf einer inhaltlichen Analyse. Ein deutliches Beispiel für den Einsatz der Reizsteuerung ist die Sprachproduktion eines Autofahrers im Dialog mit seinem Beifahrer, wenn er bei starkem Verkehr eine unübersichtliche Kreuzung mit ausgefallener Ampel befährt und sich daher stark auf außersprachliche Handlungen konzentrieren muß.

Echo-Äußerungen können daher als Symptom für den Steuerungsmechanismus der Reizsteuerung angesehen werden, der bei extremer kognitiver Belastung eingesetzt wird. Diese vorläufige Einschätzung berücksichtigt jedoch nicht, daß noch weitere Ursachen für die Wiederholung einer Partneräußerung in Frage kommen können.

[62] Rummer et al. (1995) sehen als Grundlage für diesen Steuerungsmechanismus ein Wortprotokoll der letzten Partneräußerung, das im Arbeitsgedächtnis gespeichert wird.

#### 3.4.3.3 Echo-Äußerungen und Sprecherziele

Beun (1985) untersucht sogenannte *repetitions* in Dialogen, die jedoch Definition 3.37 und nicht Definition 3.25 (Wiederholungen) entsprechen.Er weist darauf hin, daß mit einer Echo-Äußerung verschiedene kommunikative Ziele verwirklicht werden können:

**Erinnerung:** Der Sprecher möchte die Speicherung der Information in seinem eigenen Langzeit- oder Arbeitsgedächtnis unterstützen.

**Pausieren:** Der Sprecher signalisiert seinem Partner, daß er mit (kognitiver) Verarbeitung (z.B. Generierung einer eigenen Äußerung) beschäftigt ist.

**Rückmeldung:** Der Sprecher signalisiert, daß die Partneräußerung verstanden wurde.

**Nachfrage:** Der Sprecher signalisiert, daß er nicht sicher ist, die Partneräußerung richtig verstanden zu haben.

*Erinnerung* und *Pausieren* sind vorrangig sprecherbezogene Ziele mit kognitiven Ursachen; *Rückmeldung* und *Nachfrage* setzen dagegen den Dialog fort, indem sie implizit den Dialogturn an den Hörer abgeben. Ein einzelner Dialogakt kann mehrere Funktionen gleichzeitig realisieren. Z.B. kann durch eine scheinbare Rückmeldung auf die Partneräußerung, die jedoch ggf. gar nicht verlangt oder erwartet wird, auch eine Pause für die Planung eigenen Handelns gewonnen werden.

Um festzustellen, unter welchen Bedingungen ein Sprecher mit einer Echo-Äußerung eine der vier Funktionen verfolgt, hatten bei Beun (1985) Versuchsteilnehmer telefonische Informationsdialoge mit einem Partner unter unterschiedlich komplexen äußeren Bedingungen zu führen. Unterschiedliche Ausprägungen der Parameter Duplex/Nicht-Duplex-Gespräch sowie Sprechen-auf-Knopfdruck/Freies-Sprechen wurden miteinander kombiniert. Die unterschiedliche Komplexität der Kombinationen induzierten sekundär auch eine unterschiedliche zusätzliche Belastung bei der Dialogführung, da äußere Dialogbedingungen gegenüber dem freien Sprechen variiert wurden.

Es ergab sich, daß die verschiedenen Kommunikationssituationen sich bezüglich der absoluten Anzahl von Echo-Äußerungen unterscheiden, die mit der Komplexität der Situation zurückging. Beun führt das auf eine erhöhte Konzentration der Versuchsteilnehmer zurück, was jedoch bedeuten würde, daß Echo-Äußerungen vor allem durch schlechte Konzentration entstehen.

Wenn Sprecher auf eine Echo-Äußerung keine Antwort erhalten (und auch nicht darauf bestehen) oder nach der Wiederholung selbst das Gespräch fortsetzen, ist die Wahrscheinlichkeit für die Realisierung der Nachfragefunktion sehr gering. Beun kann für mindestens siebzig Prozent aller Wiederholungen daher diese Funktion ausschließen und schließt auch sonst aus seinen Daten, daß diese Funktion äußerst selten einer Echo-Äußerung zugrundeliegt. Die wenigen Fälle, für die Beun die Realisierung der Erinnerungsfunktion annimmt (9%) gingen mit einer Reduktion von Artikulationsgeschwindigkeit und Lautstärke einher.[63]

Die Funktionen *Pausieren* und *Rückmeldung* sind in ihrer Realisierung weit schwerer auseinanderzuhalten: Beun spekuliert, daß ein Sinken der Artikulationsgeschwindigkeit auch beim

[63] Als Symptom diente vor allem, daß die Sprecher den Sprechknopf nicht benutzten, also vom Partner nicht gehört werden konnten.

Pausieren herbeigeführt wird. Insgesamt wurden jedoch etwa 70 % aller Wiederholungen in dieser Untersuchung vor allem zur Rückmeldung eingesetzt. Beun schließt das daraus, daß nur der wichtigere Teil der Sprecheräußerung wiederholt wurde, z.T. Präpositionen hinzugefügt wurden und durch Drücken des Sprechknopfes auch der Partner in die Lage versetzt wurde, die Äußerung zu hören, jedoch der Sprechkanal nicht für den Partner freigegeben wurde. Da die gleichen Verhaltensweisen auch mit der Realisierung der Pausenfunktion vereinbar sind[64], kann angenommen werden, daß die Ziele der Rückmeldung und des Pausierens oft gemeinsam auftreten.

#### 3.4.3.4 Fazit

Die in den Experimenten von Beun (1985) von den Versuchspersonen abgefragten Informationen waren sehr kurz, aber nur bei vollständigem Verständnis für den Fragenden brauchbar (z.B. Flugziele, Abflugszeiten, Ankunftszeiten). Es kann daher angenommen werden, daß das Aufteilungsverhältnis zwischen Pausen und Rückmeldungen in diesem Experiment zugunsten der Rückmeldungen verschoben ist. Beun bestätigt die Verwendung von Echo-Äußerungen vor allem auch zum Zwecke der Verzögerung. Beuns Argumentation für die Realisierung der Rückmeldefunktion ist auch mit einer Argumentation für die Pausenfunktion vereinbar.

Echo-Äußerungen können daher als eines der starken Symptome für kognitive Belastung gewertet werden. Über den Zusammenhang von Echo-Äußerungen und Zeitdruck kann derzeit nichts ausgesagt werden. In Kapitel 6 wird gezeigt, wie die geringe Sicherheit bei der Entscheidung zwischen den kommunikativen Zielen hinter einer Echo-Äußerung in einem Dialogsystem umgesetzt und berücksichtigt werden kann.

## 3.5 Merkmale der Semantik und Pragmatik

Da auf konzeptueller Ebene die inhaltlichen Aspekte der Äußerung geplant werden, lassen sich Beschränkungen der Planungsprozesse im Sinne der in Kapitel 2 vorgestellten Performanzprofile natürlich besonders in der semantischen und pragmatischen Struktur von Äußerungen erwarten: '*Die Äußerungsplanung auf der konzeptuellen Ebene ist der entscheidende Faktor für die Belastung des kognitiven System beim Sprechen.*' (Rummer & Grabowski, 1996).

Entspechend Levelts Theorie unterliegt nicht nur der Inhalt, sondern auch die Anordnung sprachlich zu vermittelnder Informationen Einschränkungen bei der kognitiven Verarbeitung (Levelt, 1989, S.159):

> There are general restrictions on working memory that induce a speaker to prefer one linearization over another. These restrictions are [...] well defined and quite general in nature.

Levelt meint damit sowohl inhaltsbezogene als auch prozeßbezogene Restriktionen des Generierungsprozesses. In Kapitel 4 von Levelt (1989) sind einige Beispiele für derartige Beschränkungen in Form von Prinzipien formuliert. Dazu zählen z.B. das *Prinzip der natürlichen Ordnung*, das mit Skripten und Frames aus der Textsemantik verglichen werden kann, ein

[64] *Wenn* jemand pausieren möchte, soll sein Partner natürlich davon Kenntnis erhalten.

*Prinzip der Konnektivität* von Äußerungen, das *Stack-Prinzip* bei der Wiederaufnahme früherer Gesprächsthemen[65] und das *Minimal-load-Prinzip*. Diese Prinzipien haben vor allem zwei Funktionen: Sie erleichtern im Sinne der Griceschen Konversationsmaximen (Grice, 1975) dem Hörer das Verständnis (da er annehmen darf, der Sprecher hält sie ein). Andererseits halten sie den Verarbeitungsaufwand bei der Generierung auch für den Sprecher gering, da er sich an den Prinzipien orientieren kann. Schlägt die Umsetzung der Prinzipien fehl, kommt dafür mit Sicherheit als Ursache auch kognitive Belastung in Frage. Zumindest das Stack- und das Konnektivitäts-Prinzip können auch durch Diskursobligationen (siehe Abschnitt 3.5.7) beschrieben werden.

Semantische Merkmale einer Äußerung könnten sehr reichhaltige Informationen über Aspekte kognitiver Belastung liefern. Arnold (1997) gibt ein praktisches Anwendungsbeispiel: Um die Leistungsfähigkeit russischer Kosmonauten im Orbit einzuschätzen, werden von der Bodenstation nicht nur Stimmhöhe und Sprechgeschwindigkeit, sondern auch lexikalischer Gebrauch und Sprachniveau analysiert.

**Probleme bei der Darstellung von Merkmalen der Semantik und Pragmatik** Die Wirkungen von kognitiver Belastung und Zeitdruck auf semantische und pragmatische Merkmale sind in weit weniger Arbeiten untersucht worden als Aspekte der Prosodie und Syntax. Das liegt mit Sicherheit auch an den wenig formalisierten Analysemethoden und der damit verbundenen aufwendigen manuellen Datenauswertung.

In einigen Arbeiten werden Texte aus Monologen oder Dialogen in sogenannte *propositionale Strukturen* überführt (Grabowski, 1991). Propositionale Strukturen bestehen aus Prädikaten und deren Argumenten in thematischen Argumentrollen. Zusammenhänge zwischen Propositionen können hierarchisch angeordnet und durch einen Kohärenzgraph dargestellt werden. Monolog- oder Dialogbeiträge in Propositionsformat überführen für ihre Auswertungen u.a. Jou und Harris (1992), Roßnagel (1995b), Roßnagel (1995a).

Auch für die genannten Prinzipien besteht das Formalisierungsproblem, da sie abstrakte Beschreibungen über Sequenzen semantischer Strukturen darstellen und daher auf diese Strukturen zurückgreifen müßten. Kriterien für die Bewertung inhaltlicher und ordnungsbezogener Aspekte werden bei der Auswertung von Experimenten daher meist nur informell angegeben.[66] Entsprechend ist die Einschätzung der für die Datenerhebung verwendeten Kriterien oft vom (subjektiven) Begriffsverständnis der Auswerter in der untersuchten Domäne abhängig. Die Ergebnisse der experimentellen Untersuchungen sind daher besonders schwer verallgemeinerbar.

Bei der Generierung der Grundlagen für die Semantik einer Äußerung sind kompliziertere Prozesse betroffen als bei Artikulation oder Formulierung. Diese Prozesse führen aus dem Bereich der Sprachproduktion heraus, da sie generelle Aspekte menschlichen Denkens und Schlußfolgerns berühren. Informationsauswahl und -präsentation in der Konzeptualisierungsphase der Sprachproduktion sind nur eine Instanz allgemeiner menschlicher Informationsverarbeitung. Wissensniveau, analytische Denkfähigkeit, etc. sind einige der Faktoren, die hier

[65]Die zuletzt erwähnten Gesprächsthemen werden zuerst wieder aufgenommen oder gar nicht mehr.

[66]Z.B. verwendet Marx (1984) Maße der 'Diffusität' oder 'Floskelhaftigkeit', die noch nicht einmal durch Rückführung auf Verletzungen bestimmter textsemantischer Prinzipien definiert sind. Für eine Definition in Frage kämen z.B. das Konnektivitäts- oder Stackprinzip von Levelt (1989, Kap.4).

einen starken Einfluß haben. Entsprechend kann die experimentelle Forschung auch weit weniger Zusammenhänge zwischen semantischen Eigenschaften von Äußerungen und kognitiven Prozessen liefern, als für die Symptome der vorangegangenen Kapitel.

**Vorgehensweise** Im folgenden sollen zum gegenwärtigen Zeitpunkt vorliegende Ergebnisse vorgestellt werden, für die z.T. bereits Möglichkeiten der Modellierung und Erkennung bestehen. In fast allen Fällen ist für die weitere Untersuchung der einzelnen Phänomene jedoch eine sehr viel breitere empirische Basis erforderlich. Voraussetzung dafür ist sicherlich zunächst eine bessere Kenntnis kognitiver Informationsverarbeitung.

Einige experimentell noch nicht untersuchte inhaltliche Aspekte wurden auch innerhalb der Forschungsgruppe im Projekt READY häufig diskutiert. Soweit zumindest eine 'intuitive' Übereinstimmung dahingehend bestand, daß sie als Anpassung an Zeitdruck oder Belastung in Frage kommen könnten, werden sie hier ebenfalls aufgeführt.

Als Resultat dieses Abschnittes ergibt sich somit vor allem eine große Auswahl an Fragestellungen für die weitere kognitionspsychologische Forschung.

## 3.5.1 Überblick

Trotz der geringeren empirischen Basis sollten die Ergebnisse aus Arbeiten zu semantischen und pragmatischen Merkmalen auf die gleiche Art und Weise dargestellt werden, wie in den Tabellen der Abschnitte 3.2 und 3.3. Aus Platzgründen werden die Untersuchungsergebnisse für alle in diesem Abschnitt 3.5 besprochenen Arbeiten bereits hier in Tabelle 3.19 zusammengefaßt. Erläuterungen zu den entsprechenden Arbeiten sind in den zum jeweiligen Symptom gehörigen Abschnitten zu finden.

## 3.5.2 Lexikalische Sprechfehler

### 3.5.2.1 Abgrenzung des Phänomens: Lexikalische Sprechfehler

Die Realisierung einzelner Lexeme in einer sprachlichen Äußerung ist von der intendierten Bedeutung des gesamten Satzes abhängig und damit natürlich von der Information, die der Sprecher vermitteln möchte. Referenten und Relationen werden auf der konzeptuellen Ebene mit ihren semantischen Eigenschaften festgelegt. Diese werden durch die Umsetzung in entsprechende Lexeme in die Lemmata einer Sprache und durch diese in eine syntaktische Struktur überführt und werden innerhalb dieser ggf. durch Modifikatoren ergänzt.

Lexikalische Fehler entstehen dann, wenn die sprachliche Realisierung einer Relation, eines Referenten oder eines Modifikators nicht der intendierten Bedeutung aus der konzeptuellen Struktur entspricht. Marx (1984) sieht daher lexikalische Fehler als „Auswahlfehler innerhalb einer Wortkategorie". Ob Übereinstimmung artikulierter Sequenzen mit der intendierten konzeptuellen Struktur besteht, kann in den meisten Fällen ausschließlich vom Sprecher selbst beurteilt werden. Lexikalische Fehler in einer Partneräußerung sind daher nur schwer feststellbar. Ausnahmefälle sind diejenigen lexikalischen Fehler, die durch eine Selbstkorrektur explizit korrigiert werden.

| Merkmal | Autoren | Methode | VPn | Bedingung | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|---|
| | | | | ohne Belastung | mit Belastung | ohne Belastung | mit Belastung | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | | |
| Detailliertheit | Roßnagel (1995a)$^{=}$ geübte Vpn. | Sprachliche Instruktion von Karatekas durch *geübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | k.A. | 21.0 % aller Äußerungseinheiten auf gröbstem Auflösungsniveau | n.s. | 0 |
| | Roßnagel (1995a)$^{=}$ ungeübte Vpn. | Sprachliche Instruktion von Karatekas durch *ungeübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | 11.6 % aller Äußerungseinheiten auf gröbsten Auflösungsniveau | 42.1 % aller Äußerungseinheiten auf gröbsten Auflösungsniveau | $p < 0.001$, $(F(1,39) = 159.26)$ | – |
| | Roßnagel (1995b)$^{=}$ | Beschreibung eines Maschinenmodells gegenüber 2 versch. Hörergruppen (A= Kind/B= Student) | 45 | Präsenz des Modells | Abwesenheit des Modells während der Beschreibung (Gr.1) bzw. Zweitaufgabe (Memorieren einer Zahlenfolge, Gr.2) | 0.67 unspez. Äußerungseinheiten (Gr. A), 3.93 (Gr. B) | 9.8 unspez. Äußerungseinheiten (Gr.1A); 10.41 (Gr.2A), 10.4 (Gr.1B), 10.93 (Gr.2B) | $p < 0.001$, $(F(1,33) = 62.59)$ | – |
| | Rummer (1996a)$^{=}$, Exp. 1 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer diskontinuierlichen Doppelaufgabe (Bewertung aufeinanderfolgender Lichtsignale auf farbliche Gleichheit) | ca. 0.02 Detaillierungen/s (Bericht)[a], ca. 0.031 (Erzählen) | ca. 0.015 /s (Bericht)[a], ca. 0.01 /s (Erzählen) | $p < 0.01$, $(F(1,54) = 8.71)$ | – |
| | Rummer (1996a)$^{=}$, Exp. 2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer kontinuierlichen Doppelaufgabe (Tracking-Aufgabe) | 0.019 Detaillierungen/s (Bericht), 0.031 /s (Erzählen) | 0.016 /s (Bericht), 0.013 /s (Erzählen) | $p < 0.05$, $(F(1,54) = 5.04)$ | – |
| Redundanz | Greene et al. (1990)$^{=}$ | Rein sprachliche Aufgaben mit unterschiedlichen Zielen: Charakterisierung einer Person | 70 | Klare und deutliche Darlegung einer Personenbeschreibung | Zusätzlich Beachtung sozialer Rücksichtnahme | 0.008 inhaltl. Wdh./s Artikulationsdauer | 0.021 inhaltl. Wdh./s Artikulationsdauer | $p < 0.014$, $(F(1,65) = 6.33)$ | + |
| | Jou und Harris (1992)$^{=}$ | Vorlesen eines Zeitungsartikels mit anschließender Inhaltswiedergabe | 52 | Einfache Inhaltswiedergabe | Inhaltswiedergabe, begleitet von Kopfrechenaufgaben (mental arithmetic) | 1.29[b] | 2.29[b] | $p < 0.001$ | + |

Tabelle 3.19: Untersuchungen zu semantischen und pragmatischen Merkmalen

[a]Rummer gibt nur grafische Darstellungen der relativierten Mittelwerte. Im seinem Anhang sind unrelativierte und daher nicht vergleichbare Mittelwerte angegeben.

[b]Relativierte Anzahl von Wiederaufnahmen. Eine Basiseinheit für die Relativierung ist nicht angegeben.

| Merkmal | Autoren | Methode | VPn | Bedingung | | Werte | | Signifikanzniveau | Ten. |
|---|---|---|---|---|---|---|---|---|---|
| | | | | ohne Belastung | mit Belastung | ohne Belastung | mit Belastung | | |
| **Arbeitsgedächtnisbelastung** | | | | | | | | | |
| Füllphrasen | Greene et al. (1990) | Rein sprachliche Aufgabe mit unterschiedlichen Zielen: Charakterisierung einer Person | 70 | Klare und deutliche Darlegung einer Personenbeschreibung | Zusätzlich Beachtung sozialer Rücksichtnahme | 0.014 soziozentr. Sequenzen/s Artikulationsdauer | 0.046 soziozentr. Sequenzen/s Artikulationsdauer | $p = 0.001$, $(F(1,65) = 11.87)$ | + |
| Zusätzliche Information | Roßnagel (1995a), geübte Vpn. | Sprachliche Instruktion von Karatekas durch *geübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | k.A. | 39.0 % aller Äußerungseinheiten | n.s. | 0 |
| | Roßnagel (1995a)[=], ungeübte Vpn. | Sprachliche Instruktion von Karatekas durch *ungeübte* Karatelehrer mit Zusatzaufgabe | 40 | Instruktion | Zusätzliches Memorieren einer Zahlenfolge | 68.5 % aller Äußerungseinheiten | 15.3 % aller Äußerungseinheiten | $p < 0.001$, $(F(1,39) = 46.99)$ | – |
| | Rummer (1996a), Exp. 2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 55 | Sprechen ohne zusätzliche Bedingungen | Sprechen mit einer kontinuierlichen Doppelaufgabe (Tracking-Aufgabe) | 0.056 Prop./s (Bericht), 0.063 Prop./s (Erzählen) | 0.029 Prop./s (Bericht), 0.045 Prop./s (Erzählen) | $p < 0.05$, $(F(1,54) = 4.99)$ | – |
| **Zeitdruck** | | | | | | | | | |
| Detailliertheit | Rummer et al. (1995), Vergleich Exp. 1+2, Berichtsbed. | Sprechen über Ereignisse in Situationen des Berichtens | 20 | Kein Zeitdruck | Milder Zeitdruck | 13.8 Episoden[a] $(SD = 8.4)$, 1.8 Makroprop. $(SD = 1.1)$, 1.4 Detaillierungen, $(SD = 1.8)$ | 10.4 Episoden[a] $(SD = 6.4)$, 4.8 Makroprop. $(SD = 1.3)$, 1.5 Detaillierungen, $(SD = 1.6)$ | k.A.[b] | – |
| | Rummer et al. (1995), Vergleich Exp. 1+2, Erzählbed. | Sprechen über Ereignisse in Situationen des Erzählens | 20 | Kein Zeitdruck | Milder Zeitdruck | 18.1 Episoden[a], $(SD = 9.2)$, 2.6 Makroprop. $(SD = 2.9)$, 2.0 Detaillierungen, $(SD = 1.7)$ | 7.3 Episoden[a], $(SD = 4.1)$, 5.1 Makroprop. $(SD = 2.8)$, 0.3 Detaillierungen, $(SD = 0.5)$ | k.A.[b] | – |
| Zusätzliche Information | Rummer et al. (1995), Vergleich Exp. 1+2 | Sprechen über Ereignisse in Situationen des Berichtens und des Erzählens | 20 | Kein Zeitdruck | Milder Zeitdruck | 5.2 Prop. (Bericht)[a], $(SD = 8.6)$, 12.2 (Erzählen), $(SD = 10.3)$ | 4.1 Prop. (Bericht)[a], $(SD = 3.2)$, 6.9 (Erzählen), $(SD = 7.5)$ | k.A.[b] | – |

Tabelle 3.20: Untersuchungen zu semantischen und pragmatischen Merkmalen (Fortsetzung)

[a] '*Intervallskalierte Daten (Auftretenshäufigkeit je Versuchpersonenäußerung)*' (Rummer et al., 1995, S.38).

[b] Es sind lediglich Signifikanzmaße aus Interaktionstests zwischen den Variablen Situation und Zeitdruck angegeben.

#### 3.5.2.2 Lexikalische Fehler und kognitive Belastung

Nach Levelt (1983) gehen lexikalische Fehler entweder auf eine fehlerhafte Realisierung eines Teils der konzeptuellen Struktur in der Phase der Formulierung zurück oder auf Fehler der Perzeption. In seinem Korpus, der Beschreibungen zahlreicher farbiger Objekte enthält, kommt es z.B. vor, daß ein Sprecher 'rot' sagt, jedoch aufgrund des Kontextes klar ist, daß er 'blau' meint. Da diese Fehler auf der automatisierten Ebene der lexikalischen Enkodierung stattfinden, und die Aktivierung lexikalischer Formen aufgrund der Geschwindigkeitsanforderungen an diesen Prozeß hochgradig parallel erfolgen muß (Levelt, 1989), kommen für lexikalische Fehler lediglich Ursachen der fehlerhaften Aktivationsausbreitung in Frage, die keine arbeitsgedächtnisrelevanten und damit keine durch kognitive Belastung beeinflußbaren Aspekte aufweisen. Im Falle fehlerhafter Perzeption liegt die Ursache für einen lexikalischen Fehler außerhalb des Sprachproduktionssystems auf einer Ebene der Reizverarbeitung. Eine Untersuchung des Einflusses von kognitiver Belastung auf Reizverarbeitung würde hier zu weit führen.

Zu lexikalischen Fehlern wurde in keiner der bisher diskutierten experimentellen Arbeiten eine eigene Analyse ausgeführt. Lediglich Bromme und Wehner (1987) geben einen Wortfehlerindex an, aus dem jedoch aufgrund der Vermischung mit phonetischen/phonologischen und flexionsmorphologischen Fehlern keine Angaben über die Menge lexikalischer Fehler gezogen werden können.

#### 3.5.2.3 Fazit

Die Menge lexikalischer Fehler ist höchstwahrscheinlich nicht von kognitiver Belastung beeinflußbar. Ein empirischer Nachweis dafür ist jedoch bisher nur in geringem Maße geführt worden. Aus diesem Grund sollte von einer Verwendung lexikalischer Fehler als Symptom für Belastung oder Zeitdruck abgesehen werden.

### 3.5.3 Zusätzliche Information

#### 3.5.3.1 Abgrenzung zusätzlicher Information

Sprachproduktion unterliegt einer Steuerung durch Ziele des Sprechers und Hörers. Mögliche im folgenden betrachtete Ziele sind *gemeinsame Lösungssuche für ein Problem*, *Information* oder *Unterhaltung des Gesprächspartners*. Für die Umsetzung seiner Ziele muß ein Sprecher Information vermitteln, die sich für jedes Sprecherziel danach klassifizieren läßt, inwiefern sie zum Erreichen des Zieles beiträgt.

Hilfestellung bei der Bewertung dieser Eigenschaft geben Pläne, die zur Realisierung des Zieles führen können. Ob eine Informationseinheit 'relevant' ist, wird dadurch bestimmt, ob ihre Kenntnis die Durchführung eines der nächsten Planungsschritte vorbereitet, selbst einen Planungsschritt realisiert oder für keines von beidem eine Rolle spielt. Jede Information, die nicht direkt zur Umsetzung des Planes beiträgt, kann als zusätzliche Information klassifiziert werden. Die Klassifikation erfolgt für die einzelnen Propositionen der Sprecheräußerung.

Rummer et al. (1995) und Rummer (1996a) untersuchen diese Eigenschaften von Informationen beim Berichten (Sprecherziel: Information) und Erzählen (Sprecherziel: Unterhalten) über Ereignisse. Die verwendeten Propositionen werden nach ihrem zeitlichen Bezug zum Ereigniszeitpunkt unterschieden. Es existieren fünf Klassen: *vor dem Ereignis liegend*, *ereignisbezogen*, *auf den Sprechzeitpunkt bezogen* (metalinguistisch), *zukunftsbezogen* und *ohne temporalen Bezug*. Rummer und Grabowski (1996) (Exp.1) weisen nach, daß beim Berichten eines Ereignisses relativiert an der Gesamtäußerung mehr sachliche Information vermittelt wird (und damit hauptsächlich Information, die zur Klasse *ereignisbezogen* gehört) als beim Erzählen.

An diesem Beispiel wird klar, daß in Abhängigkeit von den Sprecherzielen in unterschiedlichen Situationen die gleiche Information Zusatzinformation oder relevante Information sein kann. Möglicherweise ist diese Eigenschaft nicht einmal binär, sondern es müssen Grade der Relevanz unterschieden werden.

Zusätzliche Information kann jedoch nicht nur in eigenen Propositionen vermittelt werden, sondern auch durch eine (unnötig) detaillierte Realisierung von Referenten oder Relationen. Die vorgestellten experimentellen Untersuchungen zu zusätzlicher Information beziehen sich ausschließlich auf das propositionale Niveau. Die Detailliertheit von Konstituenten in bezug auf die notwendige Information innerhalb einer Proposition wird im Abschnitt 3.5.4 untersucht.[67]

Ein spezieller Fall zusätzlicher Information ergibt sich bei der Beantwortung von Entscheidungsfragen. Normalerweise genügt es bei einer bejahenden Antwort, lediglich 'Ja' zu äußern. In Abhängigkeit von der logischen Struktur der Frage kann es bei einer verneinenden Antwort jedoch mehrere mögliche Skopierungen des Negators geben, so daß ein Sprecher diesen Skopus explizit spezifizieren muß. Im Falle der bejahenden Antwort und bei lediglich einer Skopierungsmöglichkeit bei verneinenden Antworten kann die erneute Aufnahme von Material aus der Frage somit als zusätzliche Information betrachtet werden, da sie keinen neuen semantischen Beitrag liefert. Umgekehrt liegt bei der negierenden Beantwortung mit mehreren möglichen Skopierungen eine Unterspezifikation vor, die als fehlende Detailliertheit klassifiziert werden muß (siehe 3.5.4)

#### 3.5.3.2 Zusätzliche Information, Zeitdruck und kognitive Belastung

Rummer et al. (1995) untersuchen die Veränderung der Verteilungen von ereignis- und nicht ereignisbezogener Information unter charakteristischen Situationsparametern wie u.a. kognitiver Belastung. Damit die Sprecher vorrangig jeweils das der Textsorte entsprechende Ziel *Informieren* bzw. *Unterhalten* verfolgen, wurden für das Experiment Gesprächssituationen mit einem Polizisten bzw. einem Nachbarn nachgestellt. Unter Zeitdruck/kognitiver Belastung fokussieren Sprecher offenbar stärker auf die Informationen, die vor allem der Vermittlung von Wissen über Situationen oder Ereignisse dienen: Beim Vergleich der Daten aus dem ersten und zweiten in Rummer et al. (1995) dargestellten Experiment wurde für die Belastungsbedingung ein allgemeiner Rückgang der nicht-ereignisbezogenen Information für die beiden Sprechsituationen Erzählen und Berichten festgestellt. Der Effekt zeigte sich jedoch in der Erzählsituation stärker; vor allem auch deshalb, da beim Erzählen von vornherein mehr derartige Informationen vermittelt und daher unter Belastung auch weggelassen werden können. Beim Darstellen eines Ereignisses, an dem der Sprecher selbst nur als Beobachter teilnahm, können auch selbstreferentielle Propositionen als zusätzliche Information gewertet werden. Ihre Anzahl ging bei der

[67] Detailliertheit ist in diesem Sinne als Form notwendiger Information von zusätzlicher Information zu unterscheiden. Die Grenze ist hier jedoch nur schwer klar zu ziehen.

Experimentalgruppe ebenfalls signifikant zurück ($p < 0.05$).

Das Experiment von Rummer und Grabowski (1996) ist so angelegt, daß Zeitdruck mit kognitiver Belastung durch eine Doppelaufgabe gleichgesetzt werden kann[68]. Daher können die Experimentalergebnisse für beide Variablen als relevant erachtet werden: Aus einer Kontrollsituation wurden Basiswerte für die Verteilungen erhoben, um die anschließenden Veränderungen unter kognitiver Belastung bzw. Zeitdruck feststellen zu können. Zwischen den beiden Situationen bestand kein signifikanter Unterschied in der Anzahl der ereignisbezogenen Informationsmenge, jedoch wurde ein höherer Anteil an nicht-ereignisbezogener Information ($p < 0.05$) in der Erzählbedingung gemessen. Nicht-ereignisbezogene Information ist in den genannten Situationen auch jede Information, die sich auf den Sprecher selbst bezieht, da er nicht Teilnehmer, sondern nur Beobachter des Ereignisses war. Die Erzählsituation ist gegenüber der Berichtssituation gekennzeichnet durch einen höheren Anteil an Propositionen mit expliziter Sprecherreferenz (selbst-referentielle Äußerungen, $p < 0.05$) und einen höheren Anteil an Propositionen, die inferiertes Wissen versprachlichen ($p < 0.05$).

Ein Rückgang der Menge an Zusatzinformation darf nicht mit einer inhaltlichen 'Verbesserung' der Sprachproduktion gleichgesetzt werden. Ob Zusatzinformation erwünscht ist, ist vor allem von den Zielen der Sprecher im Dialog abhängig.[69] Außerdem verändern sich parallel meist auch andere inhaltliche Parameter (z.B. der Auflösungsgrad der Darstellung - siehe Abschnitt 3.5.4), die negativ auf die inhaltliche Gesamtqualität der Sprecheräußerungen wirken.

Roßnagel (1995b) analysierte die Menge an Zusatzinformationen, die geübte und ungeübte Sprecher (Karatelehrer) unter verschiedenen Belastungsbedingungen bei der Instruktion eines Handlungsablaufs (Karatetechniken) vermittelten. Alle Propositionen, deren Information für eine Ausführung der instruierten Handlungen als nicht unbedingt notwendig klassifiziert werden konnte, fielen unter den Begriff der Zusatzinformation. Dazu gehörten z.B. Motivationen, warum bei verschiedenen alternativen Plänen für die Ausführung einer Teilhandlung ein spezieller ausgewählt wurde. Unter Belastung durch eine Zusatzaufgabe (Memorieren von Zahlenreihen) wurde für ungeübte Sprecher ein signifikanter Rückgang der Menge an Zusatzinformation gemessen. Die Werte für geübte Sprecher änderten sich unter Belastung nicht.

#### 3.5.3.3 Fazit

Die genannten Arbeiten legen den Schluß nahe, daß unter Zeitdruck und kognitiver Belastung eine reduzierte Informationsauswahl erfolgt, der zuerst die im Diskurskontext nicht unbedingt notwendigen Informationen zum Opfer fallen. Es ist im allgemeinen schwierig einzuschätzen, ob ein Sprecher die Menge der von ihm vermittelten zusätzlichen Information unter Bedingungen ohne Belastung tatsächlich um zusätzliche Information erweitern würde. Daher sollte aus dem *Fehlen* von Zusatzinformation nichts über den Belastungszustand eines Sprechers abgeleitet werden. Im Gegensatz dazu darf jedoch die *Anwesenheit* von zusätzlicher Information als ein Symptom für *fehlende* kognitive Belastung und *mangelnden* Zeitdruck angesehen werden.

[68]Siehe dazu auch Abschnitt 2.2.3.

[69]Beim Erzählen hat Zusatzinformation aufgrund ihres Unterhaltungswertes eine andere Bedeutung als beim sachlichen Berichten.

### 3.5.4 Detailliertheit

Beim Sprechen über Ereignisse und Zustände muß ein Sprecher aus einer Fülle von Informationen eine Selektion treffen. Neben seinen kommunikativen Zielen sind dabei Kontextfaktoren wie z.B. Position von Sprecher und Hörer im Raum, Gemeinsamkeit des äußeren Kontextes (z.B.visuell wahrnehmbare Objekte), Diskursgeschichte und -kontext oder der Wissensstand des Partners ausschlaggebend . Die sprachliche Realisierung von Relationen (Zustände oder Ereignisse) und deren Argumenten (Referenten) kann in Abhängigkeit von diesen Faktoren mehr oder weniger detailliert erfolgen. Levelt (1989, Kap.4) stellt es bildlich dar:

> Grice's quantity maxims require the speaker to sail between the Scylla of underdetermination or ambiguity and the Charybdis of overspecification or redundancy.

#### 3.5.4.1 Detailliertheit von Ereignis- und Handlungsdarstellungen

**Analysemethoden** Rummer und Grabowski (1996) untersuchen den Auflösungsgrad der Ereignisdarstellung unter kognitiver Belastung durch eine Zusatzaufgabe. Grundlage dafür ist die Zerlegung eines (den Auswertern vorher bereits bekannten) Ereignisses in Subereignisse und Zustände. Die Versuchspersonen bekommen von diesem Ereignis durch einen Film Kenntnis und werden anschließend aufgefordert, das Ereignis zu beschreiben.[70] Die bei der Ereignisdarstellung verwendeten Propositionen wurden nach drei verschiedenen Auflösungsgraden klassifiziert: *Episoden*, die ein einzelnes Subereignis oder einen Zustand beschreiben, *Detaillierungen*, die ein Subereignis oder einen Zustand feinkörniger beschreiben, als entsprechend der Zerlegung naheliegt, und *Makropropositionen*, die mehrere Subereignisse oder Zustände zusammenfassen.

Bei Roßnagel (1995a) wurden die Instruktionen geübter und ungeübter Karatelehrer untersucht. Für die Bewertung des Detailliertheitsgrades wurden die zu instruierenden Karatetechniken in konstitutive Teilhandlungen zerlegt, welche z.T. wiederum zerlegbar waren. Dadurch entstand ein Planungsgerüst mit 6 verschiedenen Ebenen. Die Propositionen in den Äußerungen der Versuchspersonen wurden dann einer der 6 Ebenen zugeordnet.

Roßnagel (1995b) zerlegt zur Analyse des Auflösungsgrades von Instruktionen eine Bauanleitung für die Anleitung eines Partners zum Aufbau eines Maschinenmodells in einzelne Bauschritte. Die Propositionen der Instruktionen durch die Versuchspersonen wurden dann danach klassifiziert, ob sie Ansatzspezifikation (*Wo muß das Bauteil befestigt werden?*) und/oder Zustandsspezifikation (*Wo befindet sich das Bauteil im fertigen Modell?*) für den einzelnen Bauschritt oder keines von beidem vermitteln (unspezifizierte Propositionen).

#### 3.5.4.2 Ereignis- und Handlungsdarstellungen und kognitive Belastung

Bei Roßnagel (1995a) wurden die Werte von 4 Versuchsgruppen analysiert, die sich auf den Dimensionen geübte/ungeübte Sprecher und Belastung/keine Belastung unterschieden. Es zeigten sich unter Belastung keine Unterschiede in der Darstellungsauflösung bei der geübten Gruppe.

[70] Inhalt des Films ist ein Diebstahl in einem Optikfachgeschäft.

Für die ungeübte Gruppe verschob sich die Verteilung der Informationen aus den unterschiedlichen Planungsebenen signifikant zugunsten der gröberen Ebene, d.h. der Auflösungsgrad der Darstellung ging zurück: Für die unteren Ebenen nahm die Anzahl der Äußerungseinheiten ab bzw. für die oberen Ebenen zu.

Roßnagel (1995b) analysiert die Werte von 3 Versuchsgruppen: eine unbelastete und zwei belastete Gruppen. Gruppe 1 wurde durch eine Doppelaufgabe belastet, Gruppe 2 dadurch, daß während der Erklärung das physikalische Modell nicht als Hilfestellung verfügbar war. Bei den belasteten Gruppen stellt Roßnagel eine signifikant geringere Auflösung der Bauschritte fest. Grund dafür war vor allem eine höhere Anzahl an Bauschritten, die in den Dimensionen Ansatz- und Zustandsspezifikation unspezifiziert blieben.

Rummer (1996a) weist eine ähnliche Tendenz unter kognitiver Belastung durch eine Doppelaufgabe nach. Parallel zur Ereignisdarstellung hatten seine Versuchspersonen die farbliche Gleichheit jeweils zweier aufeinanderfolgender Lichtreize durch einen passenden Tastendruck zu bewerten. Die an der Äußerungsdauer relativierte Anzahl an Details ging signifikant zurück ($p < 0.01$), es war jedoch kein statistisch bedeutsamer Effekt für die Menge verwendeter Makropropositionen und Episoden nachweisbar.

#### 3.5.4.3 Ereignis- und Handlungsdarstellungen und Zeitdruck

In Experiment 2 von Rummer et al. (1995)[71] gaben die Probanden unter leichtem Zeitdruck eine vergröberte Ereignisdarstellung, was durch einen signifikanten Rückgang der Menge an Details ($p < 0.05$) und Episoden ($p < 0.01$) zugunsten eines Anstiegs der Anzahl an Makropropositionen ($p < 0.01$) gekennzeichnet wurde.

Widersprechende Ergebnisse liefern Kelly und Stone (1982): Sie bewerten die *Konkretheit*, mit der Gefühle und Erfahrungen in einem Interview unter Zeitdruck dargestellt werden. Die Bewertung erfolgte im Intervall [0,1] von *sehr konkret* (1) bis *sehr unkonkret* (0). Der Effekt von Zeitdruck zeigte sich in Richtung auf eine genauere Darstellung. Leider ist weder eine Formalisierung des *specificity* genannten Maßes noch Beispiele dafür angegeben. Je nach Definition kann es sich bei sehr spezifischer Information auch um Redundanz oder zusätzliche Information handeln.

#### 3.5.4.4 Detailliertheit von Referenzen

**Unterspezifikation** Den extremsten Fall von referentieller Unterspezifikation stellen sogenannte *Dummy-* Phrasen (Clark & Wilkes-Gibbs, 1986) dar. Dazu zählen allgemeine Konzepte wie 'Ding' oder 'Teil' bzw. 'Wie-heißt-es-noch' oder 'Wie-soll-ich-es-nennen'. Clark und Wilkes-Gibbs (1986) weisen jedoch darauf hin, daß Dummy-NPs selten als endgültige Objektreferenz verwendet werden und meist zu einem späteren Zeitpunkt durch eine Korrektur präzisiert werden.[72] Ob unter Belastungsbedingungen auch häufiger auf tieferen Ebenen einer konzeptuellen Hierarchie liegende, jedoch trotzdem nicht ausreichend spezielle Konzepte zur Anwendung kommen, ist bisher offenbar nicht untersucht worden.

[71] Untersucht wurden auch hier die Textsorten 'Erzählen' und 'Bericht'.

[72] Mit Levelts Selbstkorrektur-Klassifikationen würden diese Korrekturen Korrekturen des Terminologie-Levels entsprechen und somit zur Klasse AL gehören.

**Überspezifikation** Fachbegriffe stellen in einer konzeptuellen Hierarchie eine der speziellsten Form der Objektbenennung dar. Wenn man davon ausgeht, daß für die Auswahl eines Fachbegriffes die meisten Informationen zu verarbeiten sind, um die Anwendbarkeit des Begriffes festzustellen, scheint umgekehrt auch der Schluß nahezuliegen, daß unter Belastungs- und Zeitdruckbedingungen weniger Fachbegriffe zur Anwendung kommen. Dieses Argument ist jedoch nicht haltbar, da die Verwendung von Fachbegriffen auch von der Prominenz der Begriffe für den Sprecher, d.h. dem Grad seiner Geübtheit in der Verwendung der Begriffe abhängig ist.

#### 3.5.4.5 Detailliertheit von Referenzen, kognitive Belastung und Zeitdruck

Für die Realisierung von Referenzen hatten die Versuchspersonen bei Roßnagel (1995b) generell drei Möglichkeiten: Fachbegriffe zu verwenden, diese zu umschreiben oder Fachbegriff und Umschreibung gemeinsam zu benutzen. Unter Belastung wurden mehr Fachbegriffe ohne Beschreibung verwendet als in der unbelasteten Bedingung ($p < 0.01$ bei $F(1, 33) = 11.72$). Dies scheint der oben angegebenen Argumentation zu widersprechen, daß für die Auswahl von Fachbegriffen eine besonders große Informationsmenge verarbeitet werden muß. Die häufige Verwendung von Fachbegriffen kann durch Übung geschult werden. Im Extremfall kann das soweit führen, daß Fachbegriffe immer bevorzugt ausgewählt werden, da die anderen Benennungen für den Sprecher nunmehr als weniger prototypisch erscheinen.

Die Versuchspersonen in Roßnagels Experimentalaufbau waren fachliche Laien in bezug auf den Aufbau von Maschinenmodellen. Allerdings wurden die Fachbegriffe für bestimmte Bauteile in einer dem Experiment unmittelbar vorangehenden Trainigsphase intensiv erlernt. Ihre Benutzung ist deshalb eher auf eine noch starke Aktivierung der Begriffe als einen gezielten Abruf spezieller Konzepte zurückzuführen. So gesehen sind Fachbegriffe in Roßnagels Experiment lediglich die *knappste* mögliche Form der Objektreferenz. Aus Roßnagels Experiment kann daher auch abgeleitet werden, daß Objektreferenzen unter Belastung besonders knapp und mit einem prominenten Begriff realisiert werden.

Jou und Harris (1992) stellen unter Zusatzaufgaben (allerdings ohne statistische Analyse) eine erhöhte Anzahl von Dummy-Elementen in den sprachlichen Äußerungen fest, die den bei Clark und Wilkes-Gibbs (1986) vorgestellten Dummy-NPs zu entsprechen scheinen. Ob diese NPs einer Korrektur unterzogen wurden, ist nicht berichtet. Im natürlichsprachlichen Generierungssystem, das von Carletta et al. (1995) beschrieben wird, gehören Dummy-NPs zum Reservoir am Konstruktionen, die bei der Generierung unter Zeitbeschränkungen zur Anwendung kommen können, später jedoch passend korrigiert werden müssen.

#### 3.5.4.6 Fazit

Detailliertheit der Darstellung von Ereignissen, Handlungen und Referenzen ist ein Merkmal sprachlicher Äußerungen, das in engem Zusammenhang zu kognitiver Belastung und Zeitdruck steht. Die Bewertung des Einflusses dieser beiden Variablen wird vor allem dadurch erschwert, daß zahlreiche zusätzliche Parameter der Sprechsituation und des Dialogzustandes den Detailliertheitsgrad außerordentlich stark beeinflussen. Eine Bewertung müßte entsprechend an den aktuellen Werten dieser Situationsparameter relativiert werden. Die bisherigen Ergebnisse können lediglich als Ansatz verstanden werden und weitere experimentelle und theoretische Forschung hin zu einer besseren Strukturierung des Gebietes ist wünschenswert.

Die wenigen vorliegenden Arbeiten zeigen einen einheitlichen Trend in Richtung auf einen geringeren Auflösungsgrad bei der Ereignis- und Handlungsdarstellung und eine weniger präzise Kennzeichnung von Referenten unter Belastung. Der Parameter sollte in einem Dialogsystem trotz der genannten Einschränkungen Berücksichtigung finden können, indem zunächst situationsunabhängige Extremwerte als erste Annäherung eingesetzt werden.

## 3.5.5 Redundanz

### 3.5.5.1 Abgrenzung des Begriffes: Redundanz

Redundante Information ist Information, die ein Sprecher in seinen Äußerungen mehrfach vermittelt. Diese mehrfache Informationsvermittlung kann als simple Wiederholung von Teilen sprachlicher Äußerungen oder als deren Paraphrasierung realisiert sein. Prinzipiell fallen auch Wiederholungen im Inneren von Sätzen unter den Begriff redundanter Information. Da Wiederholungen jedoch als eigenständiges Phänomen im Rahmen der Selbstkorrekturen identifiziert wurden, sollen im folgenden nur solche Propositionen als redundant klassifiziert werden, die entweder einen eigenen Satz bilden oder deren erste und wiederholte sprachlichen Realisierung durch weitere Informationen voneinander getrennt sind, d.h. die nicht direkt aufeinanderfolgen. Die kursiv gedruckten Teile von Bsp. 3.40 geben einen Eindruck von redundantem Material.[73]

(3.40) **ME065, Dialog 1:**
Ja. Nein. Es ist manchmal 's Gespräch ganz weg, wenn ich - , wenn du mit mir sprichst oder so. *Auf einmal ist das Gespräch weg.*

### 3.5.5.2 Redundanz und kognitive Belastung

Jou und Harris (1992) stellen unter kognitiver Belastung eine erhöhte Redundanz in den Sprecheräußerungen fest: '*Increased redundancy in speech is a sign of strained attentional resources.*' Greene et al. (1990) erzeugten bei ihren Versuchspersonen Belastung durch die Aufgabe, gleichzeitig eine höhere Anzahl kommunikativer Ziele zu verwirklichen. In einem Rollenspiel mußten die Teilnehmer der unbelasteten Gruppe als Personalchefs lediglich informativ Auskunft geben über einen ihrer Angestellten, während in der belasteten Gruppe zusätzlich Rücksichtnahme auf die Gefühle der beschriebenen Person gefordert war. Es ergab sich ein höherer Anteil an inhaltlichen Wiederholungen für die belastete Gruppe ($p < 0.014, F(1, 65) = 6.33$), relativiert an der Artikulationszeit. Redundanz kann jedoch auch durch spezielle kommunikative Ziele hervorgerufen werden, da Information durch jede Form der Wiederholung als besonders bedeutsam markiert werden kann.

[73] Siehe auch Tabelle 3.18.

#### 3.5.5.3 Fazit

Redundanz ist ein Phänomen, dem mit Erklärungen durch kognitive Belastung allein kaum beizukommen ist. Einerseits gilt das in Abschnitt 3.5.4 wiedergegebene Zitat von Levelt über das Verhältnis von Redundanz und Unterspezifikation, andererseits können Wiederholungen von bereits geäußerten Phrasen auch einen Korrekturzweck verfolgen. Redundanz ist daher vor allem kontextbezogen zu beurteilen, indem die lokalen Ziele des Sprechers für die Einschätzung herangezogen werden.

### 3.5.6 Füllphrasen

#### 3.5.6.1 Begriff: Füllphrasen

Sprecher verwenden in ihren Dialogbeiträgen u.a. auch Phrasen wie z.B. 'wie soll ich sagen?', 'verstehst Du?' etc. Merkmale dieser Sequenzen sind:

- Sie liefern keinen semantischen Beitrag zum Dialogverlauf.
- Sie signalisieren lediglich (pragmatisch), daß der Sprecher im Moment seine Äußerung nicht fortsetzen kann, aber mit weiterer Planung beschäftigt ist.

Der letztgenannte Punkt wird vor allem auch daran deutlich, daß der Sprecher, obwohl es sich um Fragen handelt, keine Partnerreaktion erwartet.

Greene et al. (1990) bezeichnen diese floskelhaften und stereotypen Phrasen als 'soziozentrische Sequenzen'. Die von ihnen angegebene Definition als '*non-substantive, routinised words and/or phrases*' scheint eher auf Intuition oder Erfahrung zu basieren, denn auf semantischer oder syntaktischer Analyse und ist daher für eine Operationalisierung nicht ohne weiteres verwendbar. Abgesehen davon, daß auch hier das Problem der fehlenden Möglichkeiten zu einer formalen semantischen Analyse besteht, tritt bei der Behandlung von Füllphrasen ein ähnliches Problem wie bereits bei den inhaltlichen Wiederholungen (Abschnitt 3.4.3) auf: Ihre Bedeutungslosigkeit im Dialog ist nur anhand einer Analyse der Sprecherziele feststellbar und gehört daher in den Bereich der Pragmatik. Z.B. ist es durchaus möglich, daß ein Sprecher nach einer Erläuterung eines Sachverhalts durch die Phrase 'verstehst Du?' den Hörer um ein explizites Rückmeldesignal des Verständnisses bitten möchte, weil davon die weitere Planung seines Diskursbeitrages abhängt. Batliner, Burger und Kießling (1994) sprechen im Zusammenhang mit Füllphrasen auch von *formelhaften Parenthesen*.

#### 3.5.6.2 Füllphrasen und kognitive Belastung

Greene et al. (1990) untersuchten experimentell, ob unter Belastungsbedingungen soziozentrische Sequenzen häufiger angewendet werden, etwa um den Dialogverlauf trotz hoher Planungsanforderungen fortführen zu können. Die Untersuchung ergab einen signifikanten Unterschied in der Anzahl der Füllphrasen zwischen einer Gruppe mit einem einzelnen kommunikativen Ziel und einer Gruppe, deren Dialogverhalten von mehreren verschiedenen Zielen gesteuert werden sollte. In der belasteten Gruppe traten fast dreimal so viele Sequenzen für die gleiche Artikulationsspanne auf, wie in der unbelasteten.

Die Beeinträchtigung der kognitiven Verarbeitung, die sich unter Belastungsbedingungen ergibt, sorgt vor allem dafür, daß ein Sprecher mehr Zeit für die Planung seiner Äußerung benötigt und diese Zeit im Dialogverlauf gewinnen muß, indem er Verzögerungselemente in seiner Sprachproduktion einsetzt. Füllphrasen erfüllen, da sie keinen semantischen Beitrag zum Dialog liefern, offenbar ebenfalls diesen Zweck und stehen daher in engem Zusammenhang mit gefüllten Pausen. Sie können deshalb als ein ähnlich wirkungsvoller Indikator kognitiver Belastung gesehen werden.

#### 3.5.6.3 Füllphrasen und Zeitdruck

Marx (1984) bewertet in ihrer Arbeit die Floskelhaftigkeit von Äußerungen unter Zeitdruck. Als floskelhafte Phrasen bezeichnet sie „stereotype, vorgeformte Aussagen ohne kreatives Planen". Effekte für die Veränderung dieser qualitativen Eigenschaft von Äußerungen stellt sie allerdings nur für ihre Dialogsituation fest und nicht für die Bedingung des lauten Denkens. Diese Parallele zum Auftreten von gefüllten Pausen (siehe 3.2.2) weist darauf hin, daß es sich bei Füllphrasen um hörerorientierte Phänomene handelt: Auch gefüllte Pausen werden in monologischen Situationen nur selten beobachtet, da ein Sprecher in solchen Situationen beliebig lang seine Sprachproduktion unterbrechen kann.

Die Beantwortung der Frage, warum unter Zeitdruck mehr floskelhafte Äußerungen auftreten, ist im vorliegenden Sprachproduktionsmodell nicht eindeutig möglich. Das Entstehen von Füllphrasen kann (wie die Entstehung anderer Verzögerungsphänomene auch) darauf zurückzuführen sein, daß unter Zeitdruck lediglich suboptimale Äußerungen generiert werden, die von der Monitoring-Komponente als nicht vereinbar mit den Benutzerzielen abgelehnt und zurückverwiesen werden. Die Zeit bis zur Revision der entsprechenden Äußerungsbestandteile muß ein Sprecher dann anderweitig füllen.

#### 3.5.6.4 Fazit

Füllphrasen weisen ähnliche Eigenschaften auf wie gefüllte Pausen. Im Gegensatz zu diesen ist es jedoch problematisch, die Menge der Füllphrasen eindeutig zu definieren. Ihre Bedeutungslosigkeit kann nur im Kontext festgestellt werden und ist keine Eigenschaft bestimmter sprachlicher Konstruktionen per se. Da die Veränderungen des Parameters *Füllphrasen* unter Belastung und Zeitdruck in die gleiche Richtung weisen wie für gefüllte Pausen, wäre ihre Behandlung im Zusammenhang mit der Pausenanalyse sinnvoll.

### 3.5.7 Diskursobligationen

#### 3.5.7.1 Abgrenzung des Begriffes 'Diskursobligation'

Das Verhalten von Dialogteilnehmern ist auch soziales Verhalten: bestimmte Konventionen der Dialoggestaltung gelten, deren Einhaltung vom jeweiligen Dialogpartner erwartet wird. Traum und Allen (1994) verwenden das Konzept der *Diskursobligation*, um Verpflichtungen des Dialogpartners für seine möglichen sprachlichen Reaktionen auf eine sprachliche Äußerung des Partners zu beschreiben. Z.B. ist eine Antwort im allgemeinen eine Obligation, die sich aus einer Frage ergibt. Eine Diskursobligation entsteht aus Regeln für die sprachliche Interaktion. Sie kann als Markierung für den Dialogakt repräsentiert werden, den ein Sprecher mit seiner nächsten Äußerung erfüllen muß.[74] Einerseits können sich aus einem einzelnen Dialogbeitrag verschiedene Diskursobligationen ergeben, andererseits können auch mit einer Äußerung mehrere Obligationen erfüllt werden. Tabelle 3.21 gibt Beispiele für mögliche Diskursobligationen, wobei $S_1$ und $S_2$ die beiden Dialogteilnehmer sind.

| **Obligationsquelle** | **Obligation** |
|---|---|
| $S_1$ Request $A$ | $S_2$ accept $A$ or reject $A$ |
| $S_1$ Yes-No-Question whether $P$ | $S_2$ Answer-if $P$ |
| $S_1$ Wh-Question $P(x)$ | $S_2$ Inform-ref $x$ |
| $S_1$ utterance not understood | $S_2$ repair utterance |

Tabelle 3.21: Beispiele für Diskursobligationen nach Traum und Allen (1994)

Eine wichtige Auswirkung auf die Planung der kommunikativen Beiträge eines Sprechers haben jedoch auch diejenigen seiner Ziele, die den entstandenen Diskursobligationen widersprechen können. Daher haben Diskursobligationen einen ähnlichen Status wie etwa die Konversationsmaximen von Grice für den Inhalt von Einzeläußerungen im Dialog: Sie können vom Sprecher dann mißachtet werden, wenn sie seinen eigenen Dialogzielen hinderlich sind.

Indirekt gehört zu jeder Diskursobligation eine Zeitspanne, innerhalb derer ihre Erfüllung erwartet wird. Diese Dauer soll als *Lebensdauer* bezeichnet werden. Zum Beispiel können bestimmte Obligationen durch Dialogbeiträge mit aufschiebender Wirkung für eine gewisse Zeitspanne zurückgestellt werden. Obligationen 'altern' mit dem Fortschreiten des Dialoges und werden mit steigendem Alter für den Sprecher weniger verpflichtend, da neue Obligationen hinzukommen. Ist eine Obligation innerhalb ihrer Lebensdauer nicht erfüllt, kann sie daher als mißachtet betrachtet werden.

#### 3.5.7.2 Diskursobligationen, kognitive Belastung und Zeitdruck

Untersuchungen zu Diskursobligationen und kognitiver Belastung liegen nicht vor. Einige theoretische Betrachtungen im Zusammenhang mit Erkenntnissen aus der Auswertung der READY-Voruntersuchung können jedoch zur Bildung von Hypothesen beitragen, die durch experimentelle Untersuchungen bestätigt werden müssen:

Der Grad der Kooperativität zweier Dialogpartner läßt sich anhand der Menge ihrer gemeinsamen Ziele und der zu ihrer Verwirklichung eingesetzten Pläne messen. Je kooperativer ein Dialogpartner ist, desto seltener tritt ein Konflikt zwischen Diskursobligationen und Zielen auf. In einer kooperativen Umgebung ist es daher umso wahrscheinlicher, daß ein Sprecher die Diskursobligationen aus dem vorangehenden Dialogbeitrag seines Dialogpartners erfüllen möchte. Ist eine maximal kooperative Umgebung gegeben und ein Sprecher kommt seinen Diskursobligationen trotzdem nicht nach, müssen andere Gründe als widersprüchliche Ziele dafür

[74]Zur Klassifikation von Dialogakten siehe z.B. Klesen (1997).

vorliegen: Insbesondere wird ein Sprecher Obligationen dann nicht erfüllen, wenn er dazu nicht in der Lage ist. Das ist beispielsweise der Fall, wenn wichtige Voraussetzungen für die Generierung einer entsprechenden Äußerung nicht erfüllt sind, wie z.B. die Verfügbarkeit von Information für die Beantwortung einer Frage.

Unter den verschiedenen Obligationen, die sich aus dem Dialogbeitrag eines Partners ergeben, kann ein Hörer eine Rangfolge bilden, die die Bedeutsamkeit einer Erfüllung der jeweiligen Obligationen repräsentiert. Unter Zeitdruck kann ein Sprecher ggf. nicht allen Obligationen nachkommen und wird den als weniger bedeutsam eingeschätzten Obligationen möglicherweise in geringerem Maße nachkommen. Er erfüllt dann nur einen Teil aller Obligationen und weist ggf. einige sogar explizit zurück.

Eine spezielle Diskursobligation ist die Forderung einer Antwort auf eine Frage. Als Ursache für das Fehlschlagen der Antwortgenerierung in einer kooperativen Umgebung könnten einerseits Beschränkungen des Arbeitsgedächtnisses in Frage kommen: Fehlen relevanter Information kann sowohl auf einen erfolglosen Zugriff auf die entsprechenden Informationen als auch auf ungenügende Verarbeitung zurückzuführen sein. Andererseits kann die Informationsverarbeitung auch unvollständig ausfallen, wenn eine nur ungenügende Handlungsplanung bzw. fehlende Überprüfung der Anwendbarkeit von Operatoren unter Zeitdruck (siehe Marx, 1984) erfolgt.

#### 3.5.7.3 Fazit

Die Verletzung von Diskursobligationen könnte ein Indikator für kognitive Belastung oder Zeitdruck sein. Experimentell wurde diese Möglichkeit bisher nicht untersucht. Der Begriff der Obligation bietet aber eine gute Grundlage für die Formalisierung des Problems und eine entsprechend präzise Auswertung.

## 3.6 Aspekte der Hörerorientierung

### Abgrenzung des Begriffes: Hörerorientierung

Sprecher sind grundsätzlich daran interessiert, daß der Hörer ihre Äußerungen verstehen kann - sonst wäre jede Art der Kommunikation sinnlos. Zahlreiche Aspekte von Äußerungen auf den drei bisher betrachteten linguistischen Ebenen lassen sich auch danach klassifizieren, wie sehr sie die Verständnisfähigkeit des Hörers unterstützen.

Dazu zählen einerseits so triviale Merkmale wie eine ausreichende Lautstärke oder die Verteilung strukturierender Pausen auf geeignete Positionen in der syntaktischen Struktur einer Äußerung. Zu syntaktischen Merkmalen gehört, daß Korrekturen fehlerhafter Konstruktionen ausgeführt werden und durch Korrekturterme gekennzeichnet werden. Malloch et al. (1995) weisen darauf hin, daß auch mit der Grammatikalität einer Äußerung eine gewisse Hörerorientierung einhergeht, weil ungrammatische Strukturen höhere Anforderungen an die Verarbeitung in der phonologischen Komponente des Arbeitsgedächtnisses stellen. Syntaktische Wohlgeformtheit ist daher auch im Interesse des Hörers wünschenswert. Entsprechend muß auch die Wiederaufnahmeposition der Reparaturphase einer Selbstkorrektur im Hinblick auf ihre Hörerorientierung beurteilt werden. Unterschiede im Pausenverhalten haben dabei einen geringeren Einfluß auf die Verständnisfähigkeit als Korrekturen, Fehlansätze und Wiederholungen (Deese, 1980).

Insgesamt beschreiben diese Aspekte der Hörerorientierung die Anpassung an einen *generischen Hörer* und können unabhängig vom jeweiligen Dialogpartner beurteilt werden. Andere Aspekte der Hörerorientierung sind dagegen an speziellen Eigenschaften des Dialogpartners zu relativieren. Dazu zählen Äußerungsbestandteile, die gemeinsames Wissen oder ein bestimmtes Wissensniveau voraussetzen, wie z.B. Fachbegriffe oder Referenzen auf den räumlichen Kontext des Dialoges.

Ein Sprecher überprüft die Qualität seiner eigenen Äußerungen mit Hilfe des Monitoring-Mechanismus der Sprachproduktion sowohl in bezug auf linguistische Korrektheit als auch auf die Umsetzung seiner Ziele durch die inhaltlichen Aspekte der geplanten sprachlichen Konstruktionen. Untersuchungen zur Partneradaptivität von Äußerungen brachten in diesem Zusammenhang vor allem zwei wesentliche Ergebnisse:

- Die Äußerungsplanung eines Sprechers erfolgt zuerst egozentrisch; eine Adaption an den Gesprächspartner wird erst in einer späteren Phase der Äußerungsplanung aufgrund eines Vergleiches von intendierter Äußerung und Partnermodell ausgeführt. (Horton & Keysar, 1996; Dell & Brown, 1991).

- Kognitive Belastung bzw. Zeitdruck beeinflussen die Hörerorientierung von Äußerungen negativ (Roßnagel, 1995b, 1995a). Das wird vor allem mit einem Rückgang der Aktivität des Monitoring-Prozesses erklärt, der als einzige Komponente des Sprachproduktionssystems die Vergleichsfunktion übernehmen kann.

Sprachproduktion ist daher als ein Zwei-Stufen-Prozeß zu sehen: Zunächst wird eine Äußerung mit Orientierung an einem generischen Hörer konzeptualisiert. Die Monitoring-Komponente vergleicht das Resultat mit dem vorliegenden Partnermodell und veranlaßt Modifikationen dieser vorläufigen Äußerung. Die Funktion des Monitoringprozesses spielt daher für die Partneradaptivität eine große Rolle.

## Hörerorientierung, kognitive Belastung und Zeitdruck

Das von Dell und Brown (1991) beschriebene Zwei-Phasen-Modell der Generierung von Äußerungen, schließt in der ersten Phase der Generierung einen Bezug auf irgendeine Form des Partnermodells aus. Stehen für die anschließende Anpassung der Äußerung nicht genügend kognitive oder zeitliche Ressourcen zur Verfügung, wird die Anpassung der ursprünglichen Konzeptualisierung an den Dialogpartner nicht oder nur ungenügend ausgeführt. Auch diese Schlußfolgerung ergibt sich aus den Arbeiten von Dell und Brown (1991) und Horton und Keysar (1996).

**Explizitheit**
Levelt und Kelter (1982) interpretieren Ausführlichkeit bei der Fragebeantwortung als Mechanismus, mit dem ein Sprecher Aufmerksamkeit signalisiert. Dieses Signal kann ausschließlich

für den Dialogpartner bestimmt sein und stellt daher eine Variante der Orientierung auf den Hörer dar. Wenn unter Belastung oder Zeitdruck ein Sprecher nicht mehr alle seine kommunikativen Ziele verwirklichen kann, werden unwichtigere Ziele zuerst vernachlässigt (Ziesche & Klein, 1997). Wie wichtig dabei Höflichkeit und Signalisierung der Aufmerksamkeit eingeschätzt werden, ist sicherlich vom individuellen Sprecher abhängig. Sie treten jedoch in jedem Fall hinter das Ziel, Information zu vermitteln zurück und würden bei ausreichend starker Belastung irgendwann doch zur Disposition stehen. Stark ausführliche Beantwortung von Fragen als Zeichen der Aufmerksamkeit kann daher als ein Indikator für *fehlende oder geringe* Belastung und *fehlenden oder geringen* Zeitdruck interpretiert werden.

**Fachbegriffe**

Roßnagel (1995b) sieht unerläuterte Fachbegriffe als Nachweis geringer Hörerorientierung an, da durch die Verwendung von Fachbegriffen dem Hörer eine stärkere eigene Inferenzleistung zugemutet wird. Unter Belastung durch eine Doppelaufgabe geht einerseits der Anteil der verwendeten Fachbegriffe bei seinen Probanden signifikant in die Höhe. Gleichzeitig werden weniger davon erläutert oder umschrieben. Ein ähnlicher Effekt verringerter Hörerorientierung unter Belastung zeigte sich bei Roßnagel (1995a): Der Anteil an Fachbegriffen nahm bei einer Gruppe ungeübter Sprecher relativiert an der Gesamtäußerungsmenge signifikant zu, wenn die Versuchspersonen eine Zusatzaufgabe auszuführen hatten. Gleichzeitig wurden auch hier weniger Begriffe erläutert.

Wichtig ist in diesem Zusammenhang ein Urteil über die kognitive Kompetenz des Dialogpartners, das sozusagen den Basiswert für die Beurteilung der Hörerorientierung liefert. Bei Roßnagel wurden die Unterschiede im Verhalten gegenüber einem Erwachsener bzw. einem Kind untersucht. Die kognitive Kompetenz kann u.a. durch den Wissensstand eines Hörers beschrieben werden. Der Wissensstand des Partners als Parameter für die Hörerorientierung spielt in READY für die Erkennung von Verarbeitungsbeschränkungen jedoch nur eine untergeordnete Rolle, da im Szenario der Anrufer das Dialogsystem als vollständigen Experten ansehen kann. Eine Modellierung der Einschätzung des System-Wissensstandes durch den Benutzer wird nur in sehr speziellen Kontexten überhaupt sinnvoll sein – dann sollte die kognitive Kompetenz als modulierender Faktor bei der Beurteilung der Hörerorientierung jedoch mit in Betracht gezogen werden.

**(Orts-)Deixis**

Deiktische Äusdrücke sind sprachliche Ausdrücke mit impliziten Bezügen auf den Diskurskontext. Sie sind nur unter Zuhilfenahme von Kontextinformation interpretierbar. Levinson (1990) unterscheidet Orts-, Zeit-, Diskurs-, Personen- und Sozialdeixis.[75]. Ein Hörer kann Referenzen auf die Dialogsituation nur dann korrekt auflösen, wenn die von der Äußerung betroffenen Aspekte dieser Situation sowohl Sprecher *als auch* Hörer bekannt sind. Ein kooperativer Sprecher sollte das bei seiner Äußerungsplanung berücksichtigen, um ein optimales Sprachverständnis zu erzielen.

In einem telefonischen Dialog stellt insbesondere der Bezug auf die räumlichen Parameter der Dialogsituation eine mögliche Quelle für Verständnisprobleme dar: Der Hörer hat gewöhn-

[75]Siehe auch Pinkal (1985).

lich keine Kenntnis von der räumlichen Umgebung seines Dialogpartners bzw. von dessen Position in dieser Umgebung zum Sprechzeitpunkt. Ein Sprecher muß bei räumlichen Bezügen zumindest notwendige Kontextinformation für eine eindeutige Interpretation bereitstellen. Sie wird damit implizit zum Bestandteil des Diskurswissens. Räumliche Bezüge der Lage oder Richtung können in ein Koordinatensystem eingeordnet werden, von dem der Sprecher implizit oder explizit Gebrauch macht. Levelt (1989) unterscheidet sprecher- und objektzentrierte Koordinatensysteme, die weiter nach der Direktheit ihres Bezuges differenziert werden können (direkt vs. indirekt). Es ergeben sich jedoch nicht vier, sondern nur drei sinnvolle Klassen:

**direkt sprecherzentriert:** „Vor mir ist der Benzinfilter.",

**indirekt sprecherzentriert:** „Der Benzinfilter ist links neben dem gelben Kasten." und

**direkt objektzentriert:** „Der Benzinfilter befindet sich im vorderen Teil des Motorraumes".

Damit eine Lokalisation für beide Dialogpartner ausreichend verständlich ist, müssen Sprecher und Hörer entweder ein gemeinsames Referenzsystem zu ihrer Verfügung haben, oder der Hörer muß in der Lage sein, Referenzen in das Koordinatensystem des Partners einzuordnen. Levelt (1989), S. 50 nennt dieses Problem das *Koordinationsproblem*. Das Koordinatensystem 'Auto' ist im READY-Szenario beiden Dialogpartnern weitgehend bekannt. Daher stellen lediglich die erstgenannten, sprecherzentrierten deiktischen Referenzen mögliche Problemfälle dar.

Hermann und Grabowski (1994) beschreiben verschiedene Experimente zum Zusammenhang von Partnerbezug einer räumlichen Beschreibung und dafür notwendigem kognitiven Aufwand. Deutlichstes Ergebnis ist, daß die Reaktionszeiten für die Produktion von komplexen Objektreferenzen sich als abhängig von der Unterschiedlichkeit der Positionen der Dialogpartner im Raum erweisen. Sprecher und Hörer müssen eine größere Menge an Inferenzen ausführen, die für die Anpassung der Lokalisation an den Partner notwendig sind. Außerdem muß ggf. zusätzliche Information über die Position des Sprechers im Raum vermittelt werden. Der erforderliche Planungsaufwand für die Lokalisationsgenerierung erzeugt daher bei stark unterschiedlichen Referenzsystemen eine höhere kognitive Belastung.

Im READY-Szenario sind nicht nur stark unterschiedliche Positionen von Sprecher und Hörer im Raum gegeben, sondern (zumindest anfangs) auch völlige gegenseitige Unkenntnis über die Position des anderen Teilnehmers. Der Anrufer ist daher von vornherein bei der Generierung von Ortsbezügen eine grundlegenden Belastung ausgesetzt. Ist er durch Doppelaufgaben oder andere planungsbezogene kognitive Operationen zusätzlich belastet, kann dieser Prozeß stark beeinträchtigt werden. Die zweistufige Äußerungsplanung und die Schlußfolgerungen der genannten Untersuchungen lassen erwarten, daß die Beschreibungsqualität in bezug auf Partnerorientierung mit zunehmender Belastung sinkt.

Sprecherzentrierte Äußerungen sind bei telefonischer Interaktion nicht generell ein Problem: Wenn dem Hörer zuvor die Position des Sprechers im Raum mitgeteilt worden ist, hat dieser ein entsprechendes Referenzsystem zur Verfügung. Ortsdeiktische Äußerungen *ohne* Angabe des Kontextes deuten aber auf eine mangelnde partnerorientierte Modifikation der ursprünglichen Äußerungsplanung in Richtung auf stärker objektzentrierte Referenzen hin. Wie bereits beschrieben, wird die Verantwortung für das Einleiten dieser Modifikationen dem Monitoring-Prozesses zugerechnet, der durch Zeitdruck und kognitive Belastung beeinflußbar ist.

### Fazit

Unter Belastung und Zeitdruck nimmt die Hörerorientierung eines Sprechers generell ab: Äußerungen werden nicht mehr in Hinblick auf ihre Anpassung an den Dialogpartner überprüft. Mangelnde Hörerorientierung kann sich in vielen verschiedenen Aspekten von Äußerungen manifestieren. Auch die Umsetzung von Zielen der Höflichkeit und Anzeige von Aufmerksamkeit wird unter Zeitdruck und Belastung reduziert. Resultat sind verstärkte Ellipsenbildung und weniger ausführliche Beantwortung von Fragen.

Die Bewertung der Hörerorientierung liefert ein außerordentlich wichtiges Symptom für die Einschätzung von kognitiver Belastung und Zeitdruck. Problematisch ist, daß sich dieses Charakteristikum sprachlicher Äußerungen aus vielen einzelnen Aspekten zusammensetzt, die von verschiedenen zusätzlichen, z.T. bisher wenig untersuchten Kontextfaktoren abhängig sind. In vielen Fällen muß die Hörerorientierung eines Sprechers in Abhängigkeit von individuellen Eigenschaften des Hörers bewertet werden. Z.B. sind Fachbegriffe als Symptom fehlender Hörerorientierung nur adäquat, wenn der Dialogpartner tatsächlich in der jeweiligen Domäne einen nur sehr geringen Wissensstand hat. Übungseffekte können die Anpassung an den Hörer vermindern und sind deshalb zusätzlich in Betracht zu ziehen.

## 3.7 Weitere potentielle Indikatoren

### Weitere phonetische und phonologische Merkmale

Grundsätzlich sind lassen sich die phonetischen Merkmale von Sprache in den drei Klassen

1. Zeitlicher Verlauf
2. Frequenz-Charakteristika, wobei unterschieden werden
   (a) Verlauf der Stimmtonhöhe (F0)
   (b) Umfang der Stimmtonhöhe (F0)
   (c) Frequenzaufteilung (Spektrum)
3. Amplitude (Lautstärke)

Melodie und Akzent sind im Zusammenhang mit kognitiver Belastung und Zeitdruck lediglich als unterstützende Kriterien für die Erkennung einzelner anderer Symptomklassen (z.B. Erkennung von Selbstkorrekturen im Sprachsignal, siehe Abschnitt 5.2.2) untersucht worden, jedoch selten als eigenständige Symptome für Belastung oder Zeitdruck.

Baber und Mellor (1996) zeigen, daß z.B. die Geschwindigkeit der phonetischen Realisierung von Vokalen oder Konsonanten durch Zeitdruck beeinflußt werden kann. Dazu gehören z.B. kürzere Vokale. Kürzere Pausen zwischen Worten, erhöhte glottale Schwingungsrate und höhere Atemfrequenz sowie eine geringere Lautstärke sind weitere von ihnen angegebene Charakteristika.

Jou und Harris (1992) weisen darauf hin, daß ihre Versuchspersonen unter Belastung eine unpräzisere Aussprache aufwiesen und mit einer weicheren (leiseren) Stimme artikulierten. Guggisberg, Laederach und Adler (1981) stellen Korrelationen zwischen kognitiver Belastung und dem Druck, mit dem Plosive artikuliert werden, sowie mit Schwankungen der Lautstärke fest. Oviatt und VanGent (1996) und Oviatt, Levov, MacEachern und Kuhn (1996) deuten an, daß diese Charakteristika auch durch emotionale und zielorientierte Parameter der Sprechsituation beeinflußbar sind, die in Zusammenhang mit Belastung stehen können.

Eine vollständige Aufarbeitung dieser Subklasse prosodischer Merkmale stellt ein eigenes Forschungsgebiet dar, in dem vor allem der Einfluß persönlichkeitsbezogener Parameter eine große Rolle spielen muß. Da bereits die Darstellung der formalen Grundlagen für die phonetische Analyse den ohnehin umfangreichen Rahmen dieser Arbeit sprengen würde, müssen diese Sprechcharakteristika hier unberücksichtigt bleiben.

### Weitere Merkmale der Syntax und Semantik

Einige weitere Merkmale der Sprachproduktion unter Belastung existieren, die in die bisherige Aufstellung keine Aufnahme fanden. Folgende Gründe für den Ausschluß kommen in Frage:

1. Die entsprechenden Beobachtungen wurden von den Autoren der entsprechenden Arbeiten notiert, aber keiner statistischen Auswertung unterworfen. - *oder* -

2. Die Symptome wurden nur auf der Basis undefinierter Klassifikationskriterien bewertet. - *oder* -

3. Die Anzahl der Versuchspersonen in der betreffenden Studie war zu gering ($n < 3$).

Zu den Symptomen und Schlußfolgerungen zählen:

- Arbeitsgedächtnisbelastung
  - *poorer organization of information:* Jou und Harris (1992)
  - schlechtere Speicherung und/oder Verarbeitung präsentierter Information (poor recall): Jou und Harris (1992), Bourdin und Fayol (1994)
  - mangelnde Fähigkeit zur Auflösung von Ambiguitäten: Niederberger et al. (1987)
  - verringerter vorbereitender Informationsabruf: Niederberger et al. (1987)
  - wiederholtes Abfragen der gleichen Information: Niederberger et al. (1987) S.90
- Zeitdruck
  - stärkere Diffusität: Marx (1984)

## 3.8 Probleme: Einflüsse auf das Sprechverhalten

In den vorangegangenen Abschnitten wurden einige Einflüsse auf das Sprechverhalten dokumentiert, die die Allgemeingültigkeit der festgestellten Tendenzen stark einschränken. Zu diesen Einflüssen gehören:

**Individuelle Basisparameter:** Auf die Bedeutung individueller Unterschiede im Sprechverhalten wird von verschiedenen Autoren hingewiesen. Einen deutlichen Nachweis führen Bromme und Wehner (1987). Wenn ein Dialogsystem kein personengebundenes a-priori-Wissen über den aktuellen Benutzer besitzt und zwei Sprecher mit unterschiedlichen Basiswerten (z.B. der Artikulationsgeschwindigkeit) nacheinander mit dem System interagieren, gelangt das System bei gleichen prozentualen Abweichungen des Parameters für die beiden Sprecher zu unterschiedlichen Einschätzungen. Ein Sprecher mit hoher Basis-Artikulationsgeschwindigkeit würde bei einer mittleren Geschwindigkeit der aktuellen Äußerung genauso beurteilt, wie ein Sprecher, bei dem die mittlere Geschwindigkeit der Normalwert ist.

**Kommunikative Absichten:** Nicht nur kognitive Ursachen verändern das menschliche Sprechverhalten, sondern auch kommunikative Ziele, die der Sprecher in der Dialogsituation verfolgt.

**Übungseffekte:** Ein Sprecher, der in einer bestimmten Dialogsituation sehr geübt ist (z.B. weil er sich häufig in einer solchen Sprechsituation befindet) zeigt unter der gleichen Belastung weitaus geringere Effekte auf seine Sprachproduktion als ein ungeübter Sprecher.

**Verschiedene Strategien zur Belastungsreduktion:** Statt einen Belastungseffekt auf die Sprachproduktion entstehen zu lassen, der einzelne Parameter in der experimentell festgestellten Richtung verändert, können Sprecher belastungsmindernde Strategien einsetzen.

Während die beiden erstgenannten Probleme, individuelle Basiswerte und kommunikative Absichten, eine ernsthafte Hürde für ein Dialogsystem darstellen können, sind die beiden letztgenannten Punkte, Übungseffekte und Anpassungsstrategien, für unsere Ziele weniger problematisch: Dialogsysteme, die wie READY die *tatsächliche Belastung* eines Benutzers in der Sprechsituation einschätzen möchten, brauchen nichts über die objektive Belastung durch die entsprechende Sprechaufgabe auf einer neutralen, sprecherunspezifischen Skala zu wissen[76]. Wenn Übungseffekte oder unterschiedliche Strategien der Belastungsanpassung dem Sprecher helfen, seinen tatsächlichen Belastungszustand (oder seinen Zeitdruck) zu reduzieren, dann darf auch das Dialogsystem diese freigesetzten Ressourcen erkennen und berücksichtigen. Das bedeutet: Es muß sich lediglich an einen weniger belasteten Sprecher anpassen.

Kommunikative Absichten können in vielen Fällen durch eine Beurteilung der Dialogsituation und Ableitung der generischen Sprecherziele eingegrenzt werden. Einige Effekte können, wie am Beispiel der Pausen gezeigt wurde, während der Auswertung der Sprechparameter eliminiert werden. Situationscharakterisierungen unterstützen den Systementwurf. Rummer (1996a),

[76]Ausnahme ist ggf. die Bewertung der Einsatzlatenz.

Rummer et al. (1995), Rummer und Grabowski (1996), Rummer (1996b) liefern mit dem in Abbildung 3.12 dargestellten Klassifikationssystem von Dialogsituationen ein Beispiel, das für die Charakterisierung der Dialogsituation in READY nützlich war.

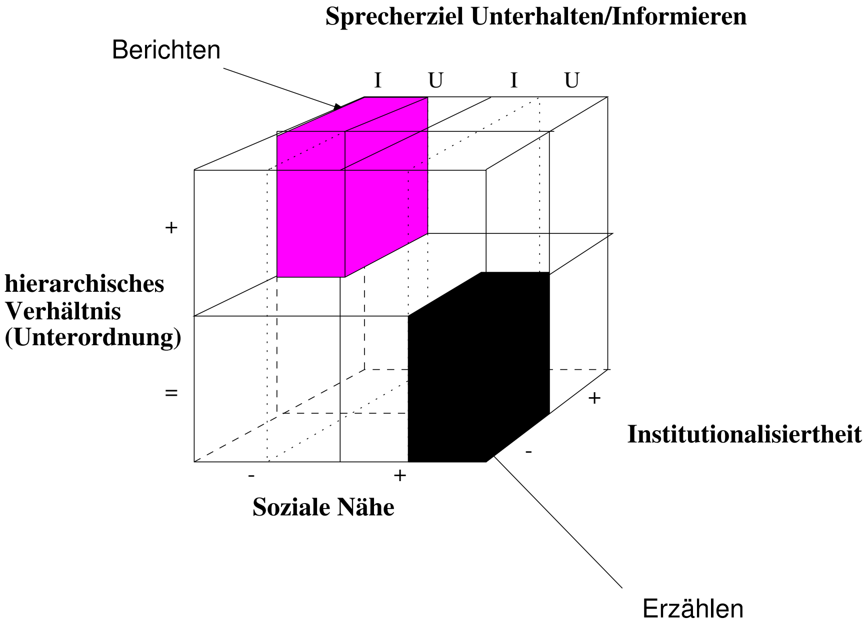


Abbildung 3.12: Klassifikationskriterien für Sprechsituationen (Rummer und Grabowski, 1996)

Die Kenntnis individueller Basisparameter würde dazu dienen, das Ziel der korrekten Interpretation auch für die individuell unterschiedlichen Ausprägungen von verschiedenen Symptomen zu erreichen. Kenntnis über diese Basiswerte hat ein System jedoch nur dann, wenn der gleiche Benutzer zu einer wiederholten Sitzung mit dem System antritt und die Daten aus den vorangegangenen Interaktionen noch verfügbar sind. Alternativ ließe sich vor Beginn der Interaktion ein Test auf individuelle Besonderheiten durchführen, wie er etwa von sprecherabhängigen Spracherkennern anhand von Beispielsätzen durchgeführt wird. Beide Varianten sind für ein Hilfesystem, das für jedermann ohne vorheriges Training verfügbar sein soll und mit sehr vielen unterschiedlichen Benutzern konfrontiert wird, nicht verwendbar. Wenn das System mit Durchschnittswerten für alle bisher bekannten Sprecher arbeitet oder generische Werte aus einer Datenbasis zugrundelegt, wird das Problem lediglich verschoben, jedoch nicht gelöst.

Dagegen können Korrelationen zwischen Benutzereigenschaften, wie z.B. Behaltensspanne und Artikulationsgeschwindigkeit, dabei helfen, unbekannte Parameter aus beobachteten Parametern abzuleiten. Die Häufung einzelner Werte über einen längeren Zeitraum hinweg (der ggf. von stark abweichenden Werten unterbrochen wird) erhöht die Wahrscheinlichkeit, daß es sich bei den häufig auftretenden Werten um einen Basiswert handelt. In Kapitel 6 wird daher eine Variante dieses wahrscheinlich einzig möglichen Weges vorgestellt, der aus dem Dilemma der unbekannten individuellen Basiswerte herausführen kann: Das System muß *während der Interaktion* mit dem Benutzer versuchen, seine Basisparameter zu *lernen*.

## 3.9 Zusammenfassung: Evaluationskriterien und potentielle Symptome

In Abschnitt 3.1.1 wurde Kriterien angegeben, deren Werte für die Beurteilung der Relevanz potentieller Symptome von kognitiver Belastung und Zeitdruck als Hilfestellung dienen können, wenn über ihren Einsatz zur Modellierung dieser Benutzereigenschaften in einem Dialogsystem entschieden werden soll. Tabelle 3.22 gibt eine Zusammenfassung für die in den vorangegangenen Abschnitten herausgearbeitete Bewertung der einzelnen Kriterien.[77]

Folgende Spalten sind in der Tabelle enthalten:

**Symptom:** die in diesem Kapitel behandelten Phänomene der Sprachproduktion.

**Tendenz:** Entwicklung des Symptoms unter kognitiver Belastung (d.h. Arbeitsgedächtnisbelastung, AGB) und Zeitdruck (ZD), wobei '+' für ein Ansteigen und '−' für eine Verringerung des jeweiligen Parameters steht. (Für viele Symptome des Zeitdrucks können keine Angaben gemacht werden, da die vorliegende experimentelle Evidenz für eine Bewertung nicht ausreichend war.)

**Basiswert:** gibt an, ob ein Basiswert für den jeweiligen Parameter erforderlich ist. Ist das der Fall, ist ein 'x' eingetragen, sonst ein '−', bei speziellen Basiswerten ein erläutertes '*'.

**Anforderungen an den Text:** experimentell identifizierte Erfordernisse bezüglich bestimmter Parameter der Sprechsituation, um den jeweiligen Parameter zuverlässig beobachten zu können.

**Zusätzliche Ursachen:** all jene zusätzlichen Einflüsse, die außer kognitiver Belastung, Zeitdruck und individuellen Unterschieden (siehe Spalte 'Basiswert') nach den vorliegenden Experimenten zu einer Veränderung des Parameters führen können.

**Nachweis (AGB):** Verhältnis der Anzahl vorliegender experimenteller Untersuchungen, welche die Tendenz für kognitive Belastung nachgewiesen haben, zur Zahl aller zum jeweiligen Symptom vorliegenden Untersuchungen. Dabei bedeutet ein Eintrag 'm/n', daß n Arbeiten zum Symptom vorlagen, von denen m den angegebenen Effekt nachgewiesen haben. Damit ist einerseits ein Maß über die Konsistenz der Untersuchungsergebnisse gegeben und andererseits ein Eindruck, wie gut das jeweilige Symptom untersucht ist.

**READY:** gibt an, ob der Parameter in der aktuellen Implementation des READY- Prototypen verwendet wird und stellt damit einen Vorausblick auf die Kapitel 4 bis 6 dar.

Die beiden Kriterien '*Natürlichkeit*' und '*Konsistenz im Sprachproduktionsmodell*' wurden wegen der einheitlichen Werte nicht als Spalten in die Tabelle aufgenommen: Alle untersuchten Symptome treten nicht nur in Laborbedingungen sondern auch in alltäglichen Sprechsituationen auf. Ungeklärt ist diese Behauptung allein für kognitive Rhythmen der Sprachproduktion. Für alle Symptome konnte außerdem eine Erklärung im vorliegenden Sprachproduktionsmodell gefunden werden, die sich entweder aus den Auswertungen der angegebenen experimentellen Arbeiten oder aus Angaben von Levelt (1989) ableiten ließen.

[77] Die Fußnoten zur Tabelle befinden sich auf der nachfolgenden Seite.

| **Symptom** | **Tendenz bei** | | **Kontextabhängigkeit** | | | **Nachweis (AGB)** | **READY** |
|---|---|---|---|---|---|---|---|
| | **AGB** | **ZD** | **Basiswert** | **Anforderungen an den Text** | **Zusätzliche Ursachen** | | |
| **Prosodische + Phonologische Merkmale** | | | | | | | |
| Artikulationsgeschwindigkeit | – | (+)[l] | x | Minimallänge | Komm. Ziele, pros.-syn. Struktur | 7/7 | x |
| Stille Pausen | | | | | | | |
| Häufigkeit | + | k.A. | *[a] | – | Komm. Ziele, Nebenaufgaben, Wortgebrauch, (Atmung, Artikulation) | 4/5 | x |
| Gesamtdauer | + | k.A. | *[a] | – | (s. Häufigk. stille Pausen) | 9/10[b] | x |
| Gefüllte Pausen | | | | | | | |
| Häufigkeit | + | k.A. | x | Typ: Dialog | – | 5/6 | x |
| Gesamtdauer | + | k.A. | x | Typ: Dialog | – | 1/2 | x |
| Intraphrasale Pausen | | | | | | | |
| Häufigkeit | | k.A. | – | Minimalstruktur | (s. Häufigk. stille Pausen) | 1 | x |
| Gesamtdauer | + | k.A. | – | Minimalstruktur | (s. Häufigk. stille Pausen) | 4/5 | x |
| Silbendehnungen (Anzahl) | + | k.A. | – | – | Prosodisch-syntaktische Struktur | | (x)[c] |
| Sprechgeschwindigkeit | – | + | x | Minimallänge | wie Pausen und Art.-geschw. | 7/7 | – |
| Kognitive Rhythmen | k.A. | k.A. | (?) | Minimallänge | (?) | – | – |
| Einsatzlatenz | +/– | – | *[e] | vorangehender Dialogakt Frage oder Aufgabe | Schwierigkeit der Partneräußerung, Textsorte (Dialog, Monolog) | 8/10[f] | (x) |
| Länge von Artikulationssequenzen | – | k.A. | – | Minimallänge | – | – | (x)[g] |
| Phonologische Sprechfehler (Anz.) | 0 | 0 | – | – | – | (2/2) | (x)[h] |
| **Syntaktische Merkmale** | | | | | | | |
| Flexionsmorpholog. Sprechfehler (Anz.) | 0 | 0 | – | – | – | (2/2) | (x)[h] |
| Syntaktische Fehler (Anz.) | + | 0 | – | – | – | 1 | (x)[h] |
| Selbstkorrekturen (Anz.) | +/–/0 | + | – | – | – | –[d] | (x)[i] |
| Wiederholungen (Anz.) | + | + | – | – | Unsicherheit von Planungsresultaten | 6/6 | x |
| Satzfragmente (Anz.) | + | k.A. | – | – | Änderung des Kontextes | 4/4 | x |
| **Syntaktisch-semantische Merkmale** | | | | | | | |
| Fehlansätze (Anz.) | + | k.A. | – | – | Änderung des Kontextes | 3/5 | x |
| Ellipsen und Anaphern (Anz.) | (k.A.) | (+)[j] | – | – | Diskursgeschichte, Komm. Ziele | – | x |
| Echo-Äußerungen (Anz.) | + | k.A. | – | – | Ziele der Rückmeldung, Rückfrage und des Memorierens | 1 | x |
| **Semantische Merkmale** | | | | | | | |
| Lexikalische Sprechfehler (Anz.) | (0) | k.A. | – | – | – | 1 | (x)[h] |
| Menge zusätzlicher Information | – | – | – | – | Situationstypen nach Abbildung 3.12– | 2/3 | x |
| Detailliertheit | – | – | – | – | Diskursgeschichte | 4/5 | (x)[k] |
| Redundanz | + | k.A. | – | – | komm. Ziele | 2/2 | – |
| Füllphrasen (Anz.) | + | + | (?) | – | – | 1/1 | (x) |
| Verletzung v. Diskursobligationen | + | (+)[r] | – | Typ des vorangehenden Dialogakts | komm. Ziele | – | – |
| Hörerorientierung | – | – | – | – | Kognitive Kompetenz des Partners | (2/2) | (x)[m] |

Tabelle 3.22: Zusammenfassung: Werte der Evaluationskriterien für alle Symptome

[a]Der Basiswert ist v.a. von der prosodisch-syntaktischen Struktur der Äußerung abhängig.

[b](Greene, 1984) wird für beide Versuchsgruppen gezählt.

[c]Der Parameter geht in Artikulationsgeschwindigkeit und Häufigkeit gefüllter Pausen auf.

[d]Wegen stark widersprüchlicher Ergebnisse wird keine Tendenz und keine Nachweisstärke angegeben.

[e]Der Basiswert ist aufgabenspezifisch.

[f]Der Nachweis ist sogar noch stärker, da verschiedene Experimente mit mehreren Gruppen arbeiteten.

[g]Verwendung der Häufigkeit von Pausen: Dieser Parameter liefert die gleiche Information.

[h]Nur zusammengefaßt als Gesamt-Sprechfehlerindex.

[i]Es werden andere Dimensionen berücksichtigt als die Anzahl. Siehe Abschnitt 5.

[j]Diese Tendenz ist lediglich eine Schlußfolgerung aus dem Rückgang der Detailliertheit und zusätzlicher Information. Siehe Abschnitt 3.4.2.

[k]Der Parameter wird im Zusammenhang mit Anaphern und Ellipsen in einem Maß der Explizitheit erhoben.

[l]Schwach belegte Schlußfolgerung aus der Tendenz der Sprechgeschwindigkeit.

[m]Nur spezifische Aspekte der Hörerorientierung, wie Ortsdeixis, und Form von Selbstkorrekturen werden berücksichtigt.

# Kapitel 4

# Analyse einer Feldstudie und ihre Ergebnisse: Annotation und Parameterwerte

## 4.1 Fragestellungen

In Kapitel 3 wurden mögliche Symptome für Arbeitsgedächtnisbelastung und Zeitdruck vorgestellt, die sich aus der psychologischen Literatur ergeben. In Tabelle 3.22 wurde zusammengefaßt, welche davon für die Verwendung in einem Dialogsystem *grundsätzlich* geeignet erscheinen. Bevor ein sinnvoller Einsatz dieser Symptome jedoch erfolgen kann, sind einige zusätzliche Überlegungen notwendig:

- Unterschiedliche Einflußfaktoren wie individuelle Basisparameter und individuelle Anpassungsleistung sowie Situationsabhängigkeit erfordern es, diese **allgemeine Einschätzung für spezielle Szenarien anzupassen**.
- Es ist wichtig, die **Aussagekraft eines gemessenen oder berechneten Parameters bei unterschiedlichen Datenmengen** beurteilen zu können. Das bezieht sich z.B. auf die Erhebung der Artikulationsgeschwindigkeit für sehr kurze Äußerungen (wie etwa ein einsilbiges „Ja").
- Um die Verarbeitung der einzelnen Parameter auf eine formale Grundlage zu stellen, ist außerdem eine **Repräsentationsform für die Parameter** zu finden, die von unterschiedlichen Komponenten computerlinguistischer Verarbeitung geliefert werden kann und direkt die Berechnung der für die Modellierung von kognitiver Belastung und Zeitdruck notwendigen Informationen ermöglicht.[1]

Es darf nicht erwartet werden, daß unterschiedliche belastete Benutzer auf allen Parametern gleiche Veränderungen zeigen. Der Einfluß der verschiedenen Symptome auf die Einschätzung

[1]Zur Erzeugung dieser Darstellung siehe Kapitel 5, zur Verarbeitung siehe Kapitel 6.

von Arbeitsgedächtnisbelastung und Zeitdruck muß daher weitgehend unabhängig voneinander bleiben.[2] Eine geeignete formale Sprache, die sowohl mit unvollständiger Information als auch mit verschiedenen Evidenzen für die Erklärung ein und desselben Sachverhalts umgehen kann, sind Bayessche Netze, deren Anwendung auf die vorliegende Problemstellung in Kapitel 6 vorgestellt wird. Die Datenstrukturen der Bayesschen Netze und zugehörigen Lösungsalgorithmen benötigen konkrete Parameterwerte, um korrekte und sinnvolle Einschätzungen des Benutzerverhaltens zu erzielen. Daher sollen in diesem Kapitel mittels **konkreter Werte für die Parameter aus einem Korpus** folgende Fragen geklärt werden:

- Treten die herausgearbeiteten Symptome in spontansprachlichen Dialogen des gewählten Szenarios (hier: READY) häufig genug und mit genügend Variabilität auf, um ihre Verwendung zu rechtfertigen?
- Welche konkreten Minimal- und Maximalwerte (d.h. Wertebereiche) der Variablen, welche A-priori Verteilungen bzw. Standardwerte für unbeeinflußte Dialogabschnitte sind zu erwarten?
- Welche Granularität bei der Verteilung gemessener Werte auf Intervalle innerhalb dieser Grenzen sollte für die Modellierung angewendet werden?

Alle genannten Punkte werden anhand der Datenauswertung aus der Feldstudie der READY-Voruntersuchung (siehe 4.3) beantwortet. Bei den Ausführungen dieses Kapitels sollte stets im Auge behalten werden, daß vor allem wegen der Situationsabhängigkeit konkreter Parameterausprägungen alle Ergebnisse bis auf die Repräsentation auf den Typ der READY-Situation beschränkt und domänenabhängig anzusehen sind. Außerdem muß warnend hinzugefügt werden, daß trotz Ähnlichkeiten der Datenerhebung aus der Feldstudie zu Datenerhebungen in psychologischen Untersuchungen alle statistischen Beschreibungen lediglich deskriptiven Charakter haben. Sie können z.T. als Voraussetzung für die Bildung neuer Hypothesen bei Untersuchungen mit inferentiellen Zielsetzungen verwendet werden. Es ist jedoch aufgrund der Vorbereitung und Durchführung der Feldstudie **nicht beabsichtigt und nicht möglich, psychologische Inferenzen aus den vorliegenden Daten abzuleiten. Das erklärt u.a. den Verzicht auf vom Leser ggf. erwartete statistische Tests über den Stichproben.**

## 4.2 Annotationsmethode

### 4.2.1 Zielsetzung

Mit der Annotation von Korpora werden verschiedene Ziele verfolgt. Getaggte Korpora unterstützen beispielsweise das maschinelle Lernen und die Verbesserung von Erkennungsleistungen. Dafür werden Eingabedaten, die eine Systemkomponente normalerweise verarbeiten soll, mit einer Annotation versehen und dieser annotierte Korpus einem Trainingsalgorithmus

[2] Aufgrund dieser individueller Anpassungen an kognitive Belastung und Zeitdruck kann daher kein Gesamtindex durch Zusammenfassung aller Paramter gebildet werden, der aussagekräftig genug wäre ohne zu simplifizieren.

übergeben. Dieser Algorithmus versucht dann Parameter für die echte Systemkomponente abzuleiten, die bei einem Lauf über unpräparierte Eingabedaten leistungsoptimierend wirken (z.B. Wahrscheinlichkeitsverteilungen für die Abfolge von Part-of-Speech-Tags für eine Vorstufe des Parsings). Darstellungen, die beim maschinellen Lernen zugrundegelegt werden, sind (meist) gleichzeitig auch diejenigen Strukturen, die bei der Analyse echter, nicht-annotierter Eingabedaten von den Systemkomponenten generiert werden.

Mit der Annotation spontansprachlicher Strukturen für das System READY werden vor allem die folgenden Anforderungen verbunden:

- Einheitliche und formale Darstellung von Information
- Orientierung der Verarbeitung an Strukturen, die von computerlinguistischen Verarbeitungskomponenten erzeugt werden können
- Rückgriff auf bereits existierende Ergebnisse der Sprachverarbeitung, ohne im Forschungsprototyp des Systems entsprechende Komponenten anschließen zu müssen (Simulation)

Die vorzustellende Annotationsmethode stellt sowohl ein Resultat als auch eine Voraussetzung für die Datenanalyse der READY-Voruntersuchung dar. (Diese Analyse ist im Abschnitt 4.3 zu finden). Um klarzumachen, auf welcher Basis die Erhebung der Daten in 4.3 erfolgte, wird das Tagset für die Annotation zuerst vorgestellt. Genauere Angaben zum Korpus sind in Abschnitt 4.3.1 und Anhang A zu finden.

### 4.2.2 Aufbau auf verwandten Arbeiten

Bei der Annotation ausschließlich spontan- und nicht schriftsprachlicher Phänomene ist es sinnvoll, sich an den Eingaberepräsentationen von Systemen mit spontansprachlicher Interaktion zu orientieren. Die Annotationsmethode des Projektes VERBMOBIL erwies sich für diese Zwecke als geeignet und leicht erweiterbar:

Vermobil hat die Konstruktion eines Systems zum Ziel, das auf die Unterstützung eines Verhandlungsdialoges zwischen zwei nicht-englischsprachigen Partnern ausgerichtet ist. Die Eingabe in VERBMOBIL ist spontansprachliche Information. Spontansprachliche Phänomene selbst sind jedoch für VERBMOBIL weitgehend uninteressant. Daher erfolgt die Annotation der Korpora vor allem mit dem Ziel, bei der syntaktischen und semantischen Verarbeitung die relevante Information aus der Eingabe extrahieren zu können. Eine simple Markierung und Löschung der außergrammatischen Passagen aus dem Eingabestring würde zwar zu einem wohlgeformten und syntaktisch verarbeitbaren String führen, jedoch ginge damit auch semantische und pragmatische Information[3] verloren, die für eine erfolgreiche Gesamtverarbeitung notwendig sein könnte. Die Annotation in VERBMOBIL erfolgt daher trotz des vordergründigen Zieles der Extraktion relevanter grammatischer Information sehr detailliert. Um die für READY relevanten, für VERBMOBIL jedoch uninteressanten Information extrahieren zu können, sind verschiedene

[3]Dazu gehören beispielsweise anaphorische Referenzen auf Bestandteile einer korrigierten Passage oder die pragmatische Implikation, daß der Inhalt korrigierter Aussagen vom Sprecher in irgendeiner Art als unzutreffend eingestuft wird.

Erweiterungen und/oder Anpassungen notwendig, die in den folgenden Abschnitten vorgestellt werden.

Batliner et al. (1994) fordern, daß eine Erweiterung des von ihnen vorgestellten Tagsets lediglich als Spezialisierung bereits existierender Annotationsklassen vorgenommen wird. Die Rückführbarkeit in ein Standardformat durch die Generalisierung der spezialisierten Darstellung mittels Bildung von Äquivalenzklassen soll dabei gewährleistet bleiben. Diese Forderung wurde bei den im folgenden vorgestellten Erweiterungen und Anpassungen berücksichtigt. Zusätzliche Annotationen (wie in Abschnitt 4.2.4 und 4.2.6), die nicht auf die VERBMOBIL-Annotationsklassen abbildbar sind, können vollständig gelöscht werden. Damit ist die Verarbeitbarkeit des gleichen Eingabestrings parallel zu READY auch in computerlinguistischen Modulen gewährleistet, die auf der VERBMOBIL-Annotation arbeiten.

### 4.2.3 Annotation zeitlicher Parameter

Zeitliche Information für einen Korpus kann erfaßt werden, indem der Korpus in Abschnitte sektioniert wird und für jeden einzelnen Abschnitt ein Attribut *Zeitliche Dauer* mit einem Wert versehen wird. Verbale Abschnitte eines Korpus können durch die Zuordnung ihrer schriftlichen Realisierung gekennzeichnet werden.

Aus der zeitlichen Information der einzelnen verbalen Bestandteile einer Äußerung wird in READY im Gegensatz zu stillen oder gefüllten Pausen keine weitere Information außer ihrer Gesamtartikulationsdauer benötigt. Daher ist eine explizite Annotation der einzelnen Worte mit zeitlicher Information nicht notwendig, wenn statt dessen ihre Komplemente (stille und gefüllte Pausen sowie die Gesamtdauer des zugehörigen Dialogturns) gemessen werden.

Die Annotation von *Dialogturns* erfolgt in VERBMOBIL mithilfe der Sprecherinitialen und einer direkt anschließenden dreistelligen Zahl, die mit der Aufeinanderfolge der Dialogturns gegenüber dem vorangehenden Turn um 1 erhöht wird. Führende Nullen werden mit dargestellt. Nonverbale und diejenigen verbalen Bestandteile einer Äußerung, die gefüllten Pausen entsprechen, werden mithilfe spitzer Klammern gekennzeichnet (siehe Kohler et al. (1994) für nähere Informationen zur allgemeinen Transliteration in VERBMOBIL). Tabelle 4.1 gibt Beispiele für die unterschiedlichen Phänomene in VERBMOBIL und die im folgenden erläuterten Anpassungen.

Um die Einheitlichkeit bei der Behandlung mit zeitlicher Information annotierter Komponenten eines Dialoges zu gewährleisten, wurden die Dialogturns für die Analyse der Voruntersuchungsdialoge ebenfalls in spitze Klammern eingefaßt. Zusätzlich wird eine Kategorie WE eingeführt, die die stille Pause während eines Sprecherwechsels bezeichnet. Alle drei Markierungen werden direkt nach der öffnenden spitzen Klammer eingetragen. Anschließend folgt in Schrägstriche eingefaßt die zeitliche Infomation als Gleitkommazahl mit drei Kommastellen und eine schließende Klammer. Bei einem Wert kleiner als 1 kann die Null weggelassen werden. Diese Zahl gibt die Dauer der entsprechend annotierten Komponente in Sekunden an. Die Angabe der zeitlichen Information erfolgt derzeit nicht für die Silbendehnungen, da deren „Über"länge, wie bereits beschrieben, nur aufwendig zu bestimmen ist. Atempausen werden nicht als eigene Klasse behandelt. Die Begründung dafür ist in Abschnitt 3.2.2 zu finden.

Nichtartikulatorische Geräusche werden der Klasse der stillen Pausen zugeordnet. Die von

| Phänomen | Verbmobil | READY |
|---|---|---|
| Dialogturn | ME001 | <ME001/3.123/> |
| stille Pause | <P> | <P/.233/> |
| Zögern (kurze stille Pause) | <Z> | <Z> |
| Atmen | <A> | <P> |
| Silbendehnung | ?? | <D> |
| gefüllte Pause | <ähm> | <ähm/.347/> |
| nichtartikulatorisches Geräusch | <#> | <#/.220/> |
| Unverständliches | <%> | <%/.220/> |
| Sprecherwechsel | – | <WE/2.130/> |

Tabelle 4.1: Annotation zeitlicher Information aufbauend auf der Verbmobil- Basistransliteration

ihnen vermittelte Information kann ggf. pragmatisch oder in bezug auf das Monitoring von Nebenhandlungen wichtig sein. Die differenzierte Annotation von VERBMOBIL wird daher beibehalten. Da die unterschiedlichen Geräuschklassen derzeit jedoch nicht für die Analyse des Benutzerverhaltens herangezogen werden, wird intern in READY eine Reklassifikation auf die Klasse 'NON-ART' vorgenommen.

Ein zusätzliches Problem ergibt sich durch die zahlreichen technischen Störungen in den Daten der Voruntersuchung, die mehrfach zur Unverständlichkeit ganzer Äußerungspassagen führten. Da auch in einem realen Kontext mit Information gerechnet werden muß, der vom Spracherkenner kein String mit Entsprechung im Lexikon der Eingabesprache zugeordnet werden kann, muß dieses Problem bei der Berechnung von Einzelparametern berücksichtigt werden: Z.B. für die Berechnung der Artikulationsgeschwindigkeit ist es notwendig, die Silbenzahl aller artikulierten Einheiten zu kennen. Für artikulierte, jedoch unverständliche Passagen ist diese Information im allgemeinen nicht bekannt. Diese Passagen müssen daher von der Analyse des Dialogturns vollkommen ausgeschlossen werden und ihre zeitliche Dauer ist entsprechend von der Gesamtdauer des Turns abzuziehen. Die Annotation unverständlicher Bestandteile erfolgt daher grundlegend wie in VERBMOBIL mit dem Zusatz der zeitlichen Information (wie oben angegeben).

Beispiel 4.1 gibt einen Beispielausschnitt aus einem der Dialoge.

(4.1) **Fr016, Dialog 5:**
<FR016/16.905/>
<äh/.344/> <P/.186/> Moment . <P/.721/> Ich hab' keine Taschenlampe .
<P/2.300/> <oleck/.437/> <P/.213/> Wo ist 'n das Ding .
<#Montagegeräusche/7.962/> <%/.676/> .

## 4.2.4 Annotation syntaktischer Parameter

### 4.2.4.1 Spezialisierung von VERBMOBIL-Tags

Bei der Annotation syntaktischer Phänomene wird die Annotation außergrammatischer Phänomene, wie sie in Batliner et al. (1994) angegeben ist, im Kern beibehalten. Diese Annotation beruht auf der Einteilung einer Korrektur, Wiederholung bzw. eines Satzfragmentes in drei Phasen.

| **Phänomen** | **Verbmobil/READY** | | | | |
|---|---|---|---|---|---|
| | Originaläußerung | | | Korrekturterm | Reparaturphase |
| | Start | | Ende | | |
| Korrektur | +/ | @M @R @D | /+ | <Korrekturterm> | @M @R @I |
| Wiederholung | =/ | @M | /= | | @M |
| Abbruch | –/ | @A | /– | | – |
| Spezifikation | +SR/ | @SR | /RS+ | – | @SR |
| Fokussierung | +SL/ | @SL | /LS+ | – | @SL |

Tabelle 4.2: Annotation außergrammatischer und spezifischer syntaktischer Information (Teil 1), aufbauend auf der Verbmobil- Basistransliteration

Eine entsprechende Darstellung wurde in Abbildung 3.9 gegeben und findet sich in der Darstellung der Annotationen in Tabelle 4.2 und den zugehörigen Beispielen (4.2 und 4.3) wieder. Batliner et al. annotieren in jedem Fall Informationen in der ersten Phase (Originaläußerung). Diese Phase wird in eine Klammerung eingeschlossen, die gleichzeitig die Phänomenklasse *Korrektur*, *Wiederholung* bzw. *Fragment* angibt. Die Elemente der zweiten Phase (Unterbrechung und Korrekturterm) werden (sofern sie überhaupt vorhanden sind) annotiert wie stille und gefüllte Pausen. Bei Korrekturen und Wiederholungen werden weiterhin zusätzlich die korrespondierenden Elemente aus Phase 1 und Phase 3 mit Wortmarkierungen versehen, die die wiederholten Teile (@M), korrigierten (@R), in der Reparaturphase gelöschten (@D) oder gegenüber der Originaläußerung hinzugefügten (@I) Worte durch Voranstellung und Konkatenation mit dem Wort markieren. Die Annotation erfolgt nur bis zum dem letzten Wort der Originaläußerung mit einer Entsprechung in der Reparaturphase. Alle nachfolgenden, die unterbrochene Konstituente vervollständigenden Worte werden nicht markiert. Tabelle 4.2 gibt einen Überblick.

(4.2) **Me0137, Dialog 5:**

<ME137/10.546/>
<+/> <@R1> Nein . </+> <P/.186/> <@R1> Ja . <P/.494/> Und <D> <Z> dann so ungefähr im Abstand <Z> von <Z> fünfzehn Zentimeter <P/.730/> <+/> <@M1> von <D> <Z> <@M2> dem <@R1> Frontblech </+> , <also/.328/> <D> <@M1> von <@M2> dem <@R1> Blech <@R1> wo <@R1> sie <@R1> dranstehen .

Wiederholungen und Korrekturen können in Korrekturen eingebettet sein, Korrekturen wiederum in Fragmente. Ist das der Fall, wird jedes einzelne Phänomen so gelabelt, als ob die eingebetteten ungrammatischen Phänomene bereits gelöscht wären, d.h. es erfolgt ggf. eine mehrfache Klammerung der gleichen Dialogstellen.

(4.3) **Me004, Dialog 8:**

<ME004/9.539/>
Was ist denn drauf? <P/1.270/> <äh/.231/> <Achso/.255/> <-/> A- </-> <=/> <@M1> es </=> <D> <P/.299/> <@I1> Ja . <=/> <@M1> e- </=> <@M1> es blinkt nur das Lämpchen . <P/1.410/> Oder ist da <+/> <+/> <=/> <@M1> irgend <@D1> ’ne </+> <D> <@M1> <@M2> irgend- </=> <P/.209/> <@M2> <@R1> irgendein </+> <Z> <@R1> irgendwas draufgemalt?

Für den „+/"-Tag aus Batliner et al. ergaben sich zwei adäquate Spezialisierungen aufgrund in der Voruntersuchung festgestellter syntaktischer Phänomene:

- Realisierung eines Referenten oder einer Lokalisation mit nachfolgender anaphorischer Aufnahme (von der syntaktischen Realisierung her eine Fokussierung, jedoch nicht von der inhaltlichen Intention her), d.h. eine Konstruktion, deren spezifischer Teil *links* vom unterspezifizierten steht und anstelle der Anapher eingesetzt werden kann.:
- Nachbesserung einer Konstituente (nicht direkt als Korrektur realisiert), bei der der spezifischere Teil *rechts* vom unterspezifizierten steht und an dessen Stelle eingesetzt werden kann (+/SR) .

Der spezialisierte Referent wurde in die Tags +SL/, /LS+ bzw. +SR/, /RS+ eingefaßt. Die Worte der Spezialisierung erhielten als Voranstellung die Markierung @SLx bzw. @SRx, wobei x durch eine einstellige Zahl für die eindeutige Markierung der Spezialisierung im Äußerungskontext ersetzt wurde.

(4.4) **Ma001, Dialog 1:**

[...] <@SL1> In <@SL1> meinem <@SL1> Auto , <P/.410/> <+SL/> da </LS+> riecht 's im Innenraum schon seit ner Viertelstunde nach <D> Benzin [...] .

(4.5) **Me183, Dialog 8:**

<So/.148/> Und jetzt mal wieder <+/> <@R1> e- </+> <äh/.231/> <@R1> <+SR/> reinstecken </RS+> und wieder rausziehen . <P/.353/> <@SR1> Ganz <@SR1> reinstecken .

#### 4.2.4.2 Neu eingeführte Tags

Zusätzliche Tags mußten eingeführt werden, um syntaktische Phänomene zu kennzeichnen, die in VERBMOBIL von den zugehörigen computerlinguistischen Verarbeitungskomponenten erkannt werden. Diese Erweiterung erfolgt vor allem im Hinblick auf die fehlenden Verarbeitungsmöglichkeiten in READY. Eine Erweiterung oder Weiterverwendung in Anwendungen, die maschinelles Lernen von Wahrscheinlichkeitsverteilungen auch für diese Phänomen zum Ziel haben, ist jedoch denkbar.

Unvollständige und daher ungrammatische Äußerungen wurden mit dem Tag '<//constit>' gekennzeichnet, indem für das (oder die) für die Bildung eines vollständigen Satzes notwendige(n) lexikalischen Argument(e) deren syntaktische Kategorie anstelle von 'constit' in den Tag eingesetzt und der Tag an der mutmaßlichen Position des fehlenden Elements annotiert wurde (bei einer angenommenen Standardwortstellung [Subjekt-Auxiliar/finites Verb-Objekte-(Partizip)] bei Verb-Zweit-Sätzen, [Konjunktion-Subjekt-Objekte-finites Verb/Auxiliar mit Partizip] bei Verb-End-Sätzen und [Auxiliar-Subjekt-Objekte-Partizip] bei Fragen). Für phrasale Konstituenten wurde die Fehlstelle durch die grammatische Funktion der Konstituente gekennzeichnet, da deren Realisierung in diesem Rahmen meist offen ist.

Die verwendeten Kategorien für lexikalische Elemente entsprechen den Tags des Stuttgart-Tübingen-Tagsets, das auch für die syntaktische Verarbeitung (siehe Abschnitt 5.3.2) verwendet wird. Die verwendeten grammatischen Kategorien sind an gebräuchliche Tags für grammatische Kategorien angelehnt (siehe z.B. Skut & Brants, 1998a, 1998b) [4]

(4.6) **Ma074, Dialog 8:**
<//SB> <//VAFIN> Nicht mal 'nen Deckel aufzuziehen .

Die Annotation von Anakoluthen ist weniger einfach, daher wurde vereinfachend die Markierung <//ANA> eingesetzt, die solche Stellen kennzeichnet.

(4.7) **Ma105, Dialog 6:**
<Ahso/.282/> Also es hört sich so an , ich hätt' genug ( Wasser @) <//ANA>

Unvollständige außergrammatische Phänomene sind, wie bereits erwähnt, unbedingt von ihren grammatischen Verwandten, den Ellipsen, zu trennen. Deren Markierung wird im folgenden Abschnitt beschrieben. Einige der unvollständigen Parameter fallen unter den sogenannten Telegrammstil: Sie können daher als elliptisch betrachtet werden, und werden somit erst dort behandelt. Allerdings erfolgte der Vollständigkeit halber die Annotation der Unvollständigkeit in diesen Fällen ebenfalls wie soeben beschrieben.

### 4.2.5 Annotation syntaktisch-semantischer Phänomene

Die Unterscheidung von Fehlansätzen und Satzfragmenten ist ein semantisches Problem. Daher ist keine gesonderte Annotation für Fehlansätze notwendig, wenn die im vorangegangenen Abschnitt vorgestellte Markierung von Satzfragmenten vorgenommen wird.

Elliptische Äußerungen zeichnen sich aus durch die Aussparung syntaktischer Elemente. Diese können von verschiedener linguistischer Komplexität sein. Daher entspricht der Entwurf der Ellipsenannotation einer zur Annotation ungrammatischer Äußerungen komplementären Vorgehensweise: Mit einem Tag <///constit> [5] werden elliptische Äußerungen gekennzeichnet, indem für 'constit' die syntaktische Kategorie des *vorhandenen* sprachlichen Materials eingesetzt wird, an das außerdem mit einem Bindestrich seine grammatische Funktion angehängt wird.

(4.8) **Me010, Dialog 8:**
Oder zwischen Lenkrad und Kupplungspedal <///PP-MO> .

Eine besondere Form der Ellipsenbildung durch die Elision obligatorischer syntaktischer Konstituenten besteht neben definiten Ellipsen beim sogenannten Telegrammstil. Bußmann (1990) rechnet unter diese Erscheinung jedoch nur solche Äußerungen, bei denen das Subjekt gelöscht wird. Definite Elisionen eines Objekts werden von ihr als relativ selten angesehen.

Die Datenauswertung von Abschnitt 4.3 zeigt, daß definite Elisionen[6] von Referenten auch ohne lexikalische Determiniertheit des Referenten als prototypischer Füller einer thematischen

[4]Eine Kurzaufstellung beider Tagsets ist im Anhang C in den Tabellen C.1 und C.2 zu finden.

[5]Zu beachten: statt zweier Querstriche bei den ungrammatischen Äußerungen werden hier drei Querstriche verwendet.

[6]Beispiele für reguläre lexikalische Elisionen sind:

Verbrolle auftreten können: Wenn Subjekt oder Objekt im Diskurskontext noch sehr präsent (salient) sind, übernehmen sie in diesem Zusammenhang offenbar die Prototypenfunktion und können ebenfalls gelöscht werden. Bußmanns Behauptung, daß definite Objektelisionen im Deutschen nur selten auftreten, ist zumindest für diese besondere Äußerungsform in der Spontansprache nicht ohne weiteres aufrechtzuerhalten.

Da Äußerungen mit Subjekt- oder Objektelision, wenn sie nicht unter den engen Begriff des Telegrammstils von Bußmann fallen, weder als rein elliptisch noch als rein ungrammatisch klassifiziert werden können, werden an ihnen in READY Information aus beiden Bereichen annotiert: Einerseits werden die fehlenden syntaktischen Argumente im oben beschriebenen Sinne gekennzeichnet, andererseits wird die gesamte Passage durch den Tag <///TTS> (für Teletype-Style) markiert, der auf die elliptischen Eigenschaften hinweist. Gleiches gilt für die Verkürzung von Äußerungen durch Auslassung eines Artikels, die ebenfalls häufig auftraten (Gapping).

(4.11) **Me027, Dialog 2:**

<///TTS> <//SB> Sind ganz locker?

(4.12) **Fr031a, Dialog 5:**

<///TTS> <//ART> Behälter ist leer .

Auf eine Annotation von Echo-Äußerungen wird in READY verzichtet, da die Feststellung der stereotypen Wiederholung von Partnermaterial durch den Anschluß einer Komponente der semantischen Verarbeitung gewährleistet ist (siehe Kapitel 5). Basis dafür ist lediglich die Annotation der Dialogakte, die im folgenden Abschnitt vorgestellt wird.

### 4.2.6 Annotation inhaltlicher Parameter

Die Kommunikation im READY-Szenario läuft stark schematisch ab: Nach der Begrüßung und Schilderung der Pannensymptome durch P tauschen S und P Informationen aus, bis S schließlich P eine Instruktion gibt. P versucht diese auszuführen und gibt eine entsprechende Rückmeldung. Eventuell müssen (vor allem im Falle fehlenden Erfolges bei der Umsetzung der Instruktion) weitere Informationen ausgetauscht werden. Diese Abfolge bleibt so lange bestehen, bis die Panne behoben ist und die Partner sich verabschieden.

Um die unterschiedlichen Äußerungen der Voruntersuchung für die Auswertung semantischer Parameter geeignet unterscheiden zu können, wurde die in Abbildung 4.1 angegebene Dialoggrammatik herausgearbeitet. Die unterschiedlichen Kanten sind entweder mit dem Namen eines Subnetzes markiert oder mit einem Dialogaktbezeichner (Letztere sind zu erkennen an ihrem ersten Zeichen „$”und den doppelten Bindestrichen im Wortinneren sowie der Schriftart Courier). Die Subnetzbezeichner sagen aus, daß bei Durchquerung der zugehörigen Kante das durch sie bezeichnete Subnetz abzuarbeiten ist. Die Subnetze 'Backchanneling' und

(4.9) Die Hühner legen (Eier). (definit)

(4.10) Peter ißt (etwas). (indefinit)

'Klärung' können an verschiedenen Stellen optional auftreten. Die Knoten, an denen ein Sprung in diese beiden Subnetze erfolgen kann, sind mit den Knotenzeichen der beiden Subnetze markiert. Alle verwendeten Dialogakte sind in Tabelle 4.3 dargestellt.

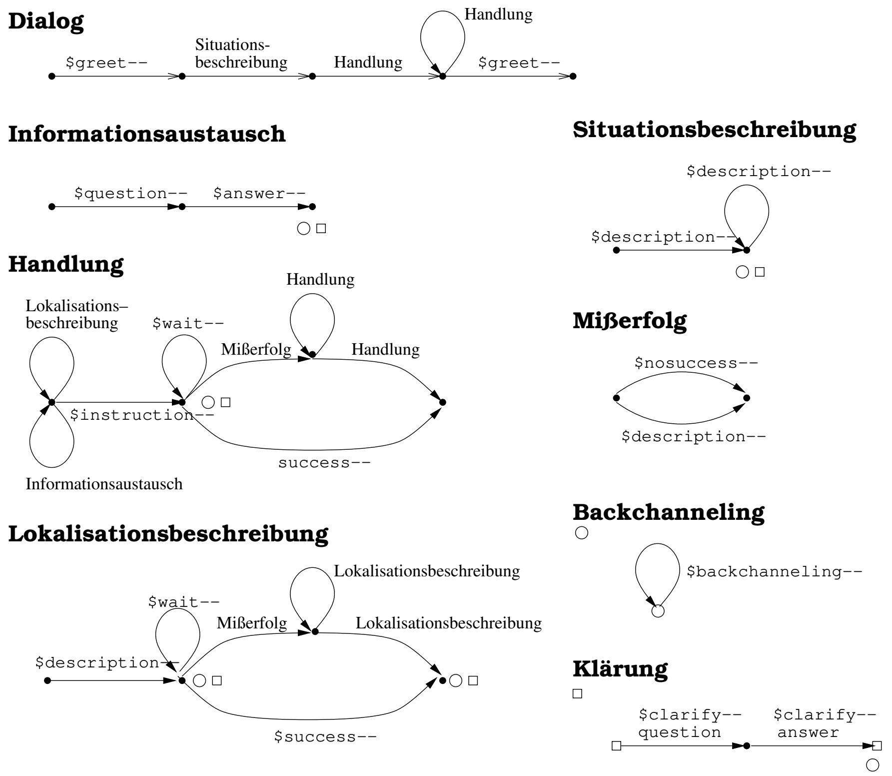


Abbildung 4.1: Dialognetz aus den Dialogen der Voruntersuchung

Auf die Annotation lexikalischer Sprechfehler wird aus den gleichen Gründen verzichtet, wie bereits auf die Annotation flexionsmorphologischer Fehler. Detailliertheit und Redundanz sind nur mit einer geeigneten semantischen Verabeitung erfaßbar und nicht mit einem Satz an Tags abzudecken, da ein formalisierter Vergleich mit der Diskursgeschichte bzw. dem semantischen Wert der aktuellen Äußerung erfolgen muß. Ähnliches gilt für Füllphrasen und unnötige Information, die teilweise durch die Annotation parenthetischer Äußerungsbestandteile abegedeckt werden, wie sie in VERBMOBIL erfolgt.

## 4.3 Feldstudie

### 4.3.1 Erhebungsmethode

Um sprachliche Daten aus der READY-Domäne zu erhalten, wurde die READY-Situation an einem Abend an einem Straßenrand inszeniert[7]. Abwechselnd wurde dafür an zwei PKW durch Manipulation der Funktionsfähigkeit der Fahrzeuge eine Panne simuliert. Die Fahrzeugführer

[7] Vorbereitung und Durchführung des Experiments erfolgte durch Anthony Jameson und Roger Knop.

| Dialogaktklasse | Subtyp | Beschreibung | Beispiel |
|---|---|---|---|
| Question | $question–yes-no | gut vorstrukturierte Frage, auf die mit 'Ja' oder 'Nein' geantwortet werden kann | Motorhaube offen? (D6, Me008) |
| | $question–information | Frage nach einer Information/einem Sachverhalt, auf die nicht mit 'Ja' oder 'Nein' geantwortet werden kann | Ungefähr in der Mitte oder weniger? (D8, Me194) |
| | $question–clarify | Nachfrage aufgrund akustischen Nichtverstehens oder zur Bestätigung | Hallo? (D5, Me013) |
| Answer | $answer–yes-no | 'Ja' oder 'Nein'; kann verstärkendes Material enthalten, das bei Verneinung den Skopus des Negators klärt, die entsprechende stattdessen vorliegende Situation beschreibt oder eine bejahte Klausel wiederholt | Ja . (D6, Ma009) |
| | $answer–information | Mitteilung von Information (Situationsbeschreibung als Antwort auf eine Frage) | Fast ganz unten . (D8, Ma195) |
| | $answer–clarify | Wiederholung der letzten Äußerung auf Zeichen des Nichtverstehens hin | Ich bin dran . (D5, Ma014) |
| Instruction | $instruction–direct | direkte Instruktion durch Verwendung eines Imperativs oder eines Infinitivs | Deckel wieder zudrehen . (D6, Me108) |
| | $instruction–indirect | Frage nach Information, die jedoch nur geliefert werden kann, wenn eine Handlung ausgeführt wird oder Beschreibung einer Lokalisation oder eines Referenten, für die der Partner eine Handlung ausführen muß (meist die Suche nach einem Objekt) | Ist da irgendwas undicht? (D6, Me045) |
| Success | $success–success | explizites Signal, daß die Handlung (erfolgreich) abgeschlossen wurde bzw. daß der zu lokalisierende Referent gefunden wurde | Deckel zu . (D6, Ma109) |
| | $success–nosuccess | explizites Signal, daß die Handlung nicht ausgeführt bzw. der zu lokalisierende Referent nicht gefunden werden kann | Ich find den Hebel nicht. (D8, Ma017) |
| Backchanneling | $backchanneling–wait | explizite Bitte um Geduld | Da unten? Moment . (D3, Ma019) |
| | $backchanneling– | allgemeines Signal, daß Partneräußerung gehört (und verstanden?) wurde | Alles klar . (D6, Me106) |
| Description | $description–situation | Beschreibung eines Sachverhalts, einer Situation | Die ist eigentlich nicht lose . Da kann nix rauskommen . (D1, Ma039) |
| | $description–localisation | Beschreibung einer Richtung, eines Ortes, eines Referenten | Da ist 'ne Leitung dran . (D1, Me027) |
| Soliloquy | $soliloquy– | Äußerungen während Nebenaufgabe, die nicht hörergerichtet sind | Was ist 'n das da jetzt? (D5, Fr048) |
| Ignore | $ignore–as-pause | Äußerungen von Versuchsleiter, Zusatzpersonen während Nebenaufgabe | – |
| | $ignore– | Durch Weitersprechen des Partners unterbrochene Äußerungsansätze | – |
| | $ignore–as-interrupt | Sequenzen der Verbindungsüberprüfung | – |
| other | $greet– | Begrüßung, Verabschiedung | Schuler. Guten Abend . (D8, Ma001) |

Tabelle 4.3: Dialogakte aus den Dialogen der Voruntersuchung

(P) sollten Unterstützung bei der Pannenbehebung erhalten, indem sie per Mobiltelefon Verbindung mit einem telefonischen Pannenhilfesystem aufnahmen. Die Rolle des Systems (S) wurde von einem PKW-Mechaniker übernommen. Die Dialoge wurden am Platz des Mechanikers auf Audio-Kassetten aufgezeichnet. Die Versuchspersonen erhielten eine Belohnung von 20 DM für ihre Teilnahme an der Untersuchung.

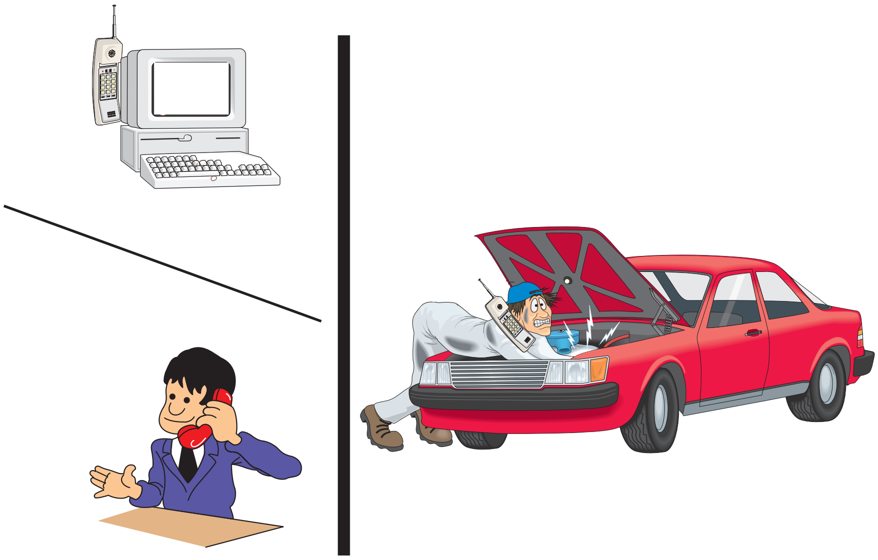

Abbildung 4.2: Simulation des Ready-Szenarios in der READY-Voruntersuchung

Insgesamt wurden 9 Dialoge mit 8 Versuchspersonen[8] geführt. Der Mechaniker war in jedem Dialog die gleiche Person. Bei 5 Dialogen wurde Zeitdruck erzeugt, indem für die erfolgreiche Behebung der Panne innerhalb einer vorgegebenen Zeitspanne eine zusätzliche Belohnung von weiteren 20 DM in Aussicht gestellt wurde. Die vorgegebene Zeitspanne war in verschiedenen Dialogen unterschiedlich[9]. Um die Erinnerung an das Zeitlimit aufrecht zu erhalten, gab einer der auf der Seite von P anwesenden Versuchsleiter in unregelmäßigen Abständen die verbleibende Zeit bekannt.

Der Wortlaut der Dialoge wurde transkribiert und mit Markierungen für die eindeutige Identifikation der Turns und die Zuordnung zu den beiden Dialogteilnehmern versehen. Zwei der Dialoge wurden in Anlehnung an die Transliterationsvorschriften im Projekt VERBMOBIL ausführlicher behandelt, indem Informationen zu Sprechpausen und nichtartikulatorischen Geräuschen sowie zu überlappenden Dialogstellen hinzugefügt wurden.[10] Wegen eines Versehens des Mechanikers wurde einer der Dialoge ohne Zeitdruck nicht aufgezeichnet. Ein Überblick über allgemeine Eigenschaften der verbleibenden acht Dialoge gibt Tabelle 4.4[11]. Im Anhang A sind Kurzbeschreibungen der Dialoge zu finden.

[8]Dialog 1 und Dialog 2 wurde mit der gleichen Person auf der Seite von P realisiert.

[9]Da die Leiter der Studie nach den ersten Zeitdruck-Dialogen vermuteten, daß das Zeitlimit für die Behebung der Panne zu groß gewählt wurde, um bei den Versuchspersonen den Eindruck von Zeit*druck* zu erzeugen, wurde die Bearbeitungsdauer bei den späteren Dialogen verkürzt.

[10]Diese Aufgaben wurde von Bernd-Paul Simon ausgeführt.

[11]Da teilweise Äußerungen der Versuchsleiter aus dem Hintergrund transkribiert wurden, addiert sich Anzahl der Turns in den Spalten *Mech.* und *Anrufer* nicht zu *ges.*.

| Dialog | Zeitdruck | Dauer | Turns | | | Sätze | Tokens |
|---|---|---|---|---|---|---|---|
| | | | ges. | Mech. | Anrufer | | |
| Dialog 1 | – | 3:23 | 66 | 33 | 32 | 116 | 488 |
| Dialog 2 | x | 2:44 | 44 | 19 | 23 | 66 | 290 |
| Dialog 3 | – | 2:41 | 44 | 23 | 21 | 57 | 176 |
| Dialog 4 | x | 2:09 | 35 | 18 | 17 | 62 | 243 |
| Dialog 5 | x | 3:06 | 67 | 26 | 35 | 94 | 262 |
| Dialog 6 | – | 5:16 | 122 | 58 | 60 | 190 | 585 |
| Dialog 7 | x | 2:06 | 37 | 18 | 17 | 59 | 175 |
| Dialog 8 | x | 13:24 | 232 | 122 | 106 | 346 | 1344 |
| gesamt | | 34:49 | 647 | 317 | 311 | 990 | 3563 |
| ohne Zeitdruck | (–) | | | 114 | 113 | | |
| Zeitdruck | (x) | | | 203 | 198 | | |

Tabelle 4.4: Eigenschaften der Dialoge aus der Voruntersuchung

Die Kommunikation in einigen der Dialoge war zum Teil erheblich erschwert durch Störungen der mobilen Telefonverbindung. P's Äußerungen sind stellenweise vollkommen unverständlich. Dadurch ergaben sich viele Turns, die lediglich der Verbindungsüberprüfung oder Diskussion der Verbindungsqualität dienen. Durch diese für die gewählte Situation nicht untypische ständige Unterbrechung und Wiederaufnahme wurde gewissermaßen eine zusätzliche kognitive Belastung sowohl beim Mechaniker als auch beim Anrufer induziert, deren Einschätzung in realen Situationen wichtig sein kann.

### 4.3.2 Analysemethode

Aufbauend auf der ursprünglichen wörtlichen Transliteration der Dialoge wurden verschiedene Anpassungen und Erweiterungen vorgenommen. Diese sind

- Transliteration der Pausen in den verbleibenden sechs Dialogen,
- Annotation zeitlicher Parameter für alle Pausen, Dialogturns, unverständlicher Passagen etc.,
- Berichtigung inhaltlicher Inkorrektheiten, die bei der auditiven Analyse entstanden waren,[12]
- Anpassung von dialektalen Besonderheiten, Orthographie und Zeichensetzung an die Bedürfnisse computerlinguistischer Standardlexika der deutschen Sprache,
- Markierung überlappender Dialogabschnitte,
- Annotation aller Klassen von Parametern mit den zugehörigen Tags.

[12] Dazu gehört die erneute Zuordnung von Dialogturns zu den unterschiedlichen Sprechern und eine Aufspaltung bzw. Zusammenfassung vorliegender Dialogturns, um die Integrität einzelner Äußerungen zu erhalten. Die Marken der Dialogturns folgen daher in den Dialogen nicht immer im Abstand von 1 aufeinander (z.B. folgt auf 'Me019' in Dialog 4 'Fr026'). Um die Konsistenz der Daten mit früheren Veröffentlichungen zu erhalten, wurden die Marken nicht überarbeitet.

Da die Literaturstudie nur verhältnismäßig wenige Parameter zum Zeitdruck ergab, liegt es nahe, die Werte aus den Zeitdruck-Dialogen mit denen der Dialoge ohne zusätzlichen Zeitdruck zu vergleichen. Die Idee, die deskriptive statistische Auswertung um Inferenzen zu erweitern, die diesen Unterschied betreffen, wurde jedoch fallengelassen: Die Feldstudie war nicht von vornherein für eine solche Auswertung entworfen worden, so daß eine echte Kontrolle der Versuchsbedingungen nicht erfolgen konnte. Außerdem wurden keine Hypothesen zur Wirkung des Zeitdrucks aufgestellt. Eine nachträgliche Hypothesenbildung ist jedoch nicht zulässig.

Bei Durchsicht der Daten zeigte sich zusätzlich, daß es kaum systematische Unterschiede zwischen beiden Bedingungen gab, die auf eine Wirkung des Faktors Zeitdruck zumindest hätten hindeuten können. Verschiedene Erklärungen kommen dafür in Frage:

- Der Zeitdruck ist nicht stark genug gewesen, um einen Effekt zu erzeugen, sodaß die Probanden sich nicht anpassen *mußten*.
- Marx (1984) gibt an, daß Zeitdruck vor allem auch dann empfunden wird, wenn eine Aufgabe, die schon einmal in unbegrenzter Zeit ausgeführt werden konnte, nunmehr in einem Zeitlimit ausgeführt werden muß. Das heißt, die Handelnden können nur dann einen Zeitdruck empfinden, wenn sie wissen, daß die zur Verfügung stehende Zeit knapp bemessen ist und entsprechend reagieren. Den Sprechern im vorliegenden Szenario könnte daher auch unklar gewesen sein, was ein Zeitlimit von mehreren Minuten für die Behebung ihrer Panne bedeutet, so daß sie sich nicht anpassen *konnten*.
- Der Handlungsablauf ist (auch bei den Sprechhandlungen) in der vorliegenden Situation möglicherweise bereits optimiert im Sinne einer besten Strategie. Wenn diese Strategie nicht von äußeren Faktoren beeinflußbar ist, heißt das, daß die Probanden sich nicht anpassen *wollten*.

### 4.3.3 Zeitliche Parameter

Um die zeitlichen Parameter der Äußerungen auswerten zu können, wurden die Dialoge von den Audiokassetten per Kassettendeck und einer Soundblaster-kompatiblen Audiokarte auf einem PC mit einer Samplingrate von 22 kHz bei 16 bit Darstellungsbreite (Mono) ins WAVE-Dateiformat überführt.

Die automatische Messung zeitlicher Parameter eines Sprachsignals ist mit verschiedenen Problemen behaftet (Klos & Ellgring, 1987). Da außerdem ein automatischer Pausendetektor nicht zur Verfügung stand, wurden die zeitlichen Parameter mit Unterstützung des Shareware-Programms Goldwave (Version 1.53 unter Windows95) manuell gemessen. Dieses Programm ermöglicht die Markierung und das getrennte Abspielen von einzelnen Abschnitten einer WAVE-Datei. Für die Markierung steht zusätzlich zur auditiven Information eine Darstellung der Schallwellen im Audiosignal in grafischer Form auf einer Zeitachse zur Verfügung. Zur Markierung der einzelnen Dialogturns wurde aus den verschiedenen Vergrößerungsfaktoren die Auflösung '1:100' gewählt. Damit wurde 1 s des Signals bei einer Auflösung von 800 x 600 Pixeln auf 6.25 cm Bildschirmbreite dargestellt. Der durchschnittliche Fehler bei der Darstellung der Markierungen durch Goldwave beträgt $\pm$ 0.005 s. Für die gewählte Auflösung wurde durch mehrfache Markierung der gleichen Dialogstelle ein manueller Meßfehler von $\pm$ 0.02 s festgestellt.

Gemessen und notiert wurden

- die **Gesamtdauer** der einzelnen Dialogturns als Beginn des ersten sprachlichen Signals eines Sprechers (Abweichung der Amplitude von 0) bis zum Ende des letzten sprachlichen Signals vor dem Wechsel des Dialogturns,
- alle **stillen Pausen** innerhalb eines Dialogturns, die länger sind als 200 ms,[13]
- alle **gefüllten Pausen** (*äh*, *ähm*, verschiedene Interjektionen) innerhalb eines Dialogturns vom Beginn des zugehörigen sprachlichen Signals bis zur nächsten stillen Pause bzw. bis zum nächsten, nicht zur gefüllten Pause gehörigen sprachlichen Signal,
- die **Einsatzlatenz** eines Sprechers als Dauer zwischen dem Ende des letzten sprachlichen Signals des Dialogpartners und dem Beginn des ersten sprachlichen Signals des Sprechers (Es wurden auch Einsatzlatenzen unter 200 ms notiert.),
- die **Dauer unverständlicher Passagen und nichtartikulatorischer Geräusche**, sowie von nicht-punktuellen Nebengeräuschen innerhalb der Dialogturns.

#### 4.3.3.1 Artikulationsgeschwindigkeit

Die Grafiken 4.3 und 4.4 geben einen Überblick über die Mittelwerte und Standardabweichungen der Artikulationsgeschwindigkeit je Dialogturn für die beiden Dialogrollen (Anrufer und Mechaniker) in den unterschiedlichen Dialogen. Für die Interpretation der Grafiken ist es wichtig, in Erinnerung zu rufen, daß der Mechaniker stets die gleiche Person war, während in der Rolle des Anrufers sieben verschiedene Personen agierten. Für beide Grafiken wurde die gleiche Skaleneinteilung gewählt, um den Vergleich zu erleichtern.

Im Abschnitt 3.2.2 wurde gezeigt, daß nicht nur stille, sondern auch gefüllte Pausen in Dialogen zur Verzögerung der Sprachproduktion eingesetzt werden können. Im Hinblick auf die spätere Modellierung (Kapitel 6) wurden bei der Auswertung der Feldstudie gefüllte Pausen von der Berechnung der Artikulationsgeschwindigkeit ausgeschlossen. Die Artikulationsgeschwindigkeit stellt damit ein bereinigtes Maß dar für die Geschwindigkeit, mit der *die lexikalische Bedeutung tragenden Bestandteile* einer Äußerung artikuliert werden. Diese Abgrenzung entspricht der Bedeutung des von Kowal (1991) verwendeten Begriffes „Verbales".[14]

Die Werte der Einzeldialoge sind nach Mittelwert und Standardabweichung geordnet. Deutlich wird daran zum einen, daß die Artikulationsgeschwindigkeit ein und desselben Sprechers auch über mehrere verschiedene Dialoge hinweg im Rahmen eines ausgezeichneten Mittelwertintervalls bleibt (SD von $\bar{x}$ = 0.373), während die Streuung der Mittelwerte für verschiedene Personen sich in einem größeren Intervall bewegt (SD von $\bar{x}$ = 0.603). Außerdem liegen die Standardabweichungen für alle Sprecher in allen Dialogen im Intervall [1.25,2.1], für den Mechaniker sogar nur im Intervall[1.4,1.9]. Diese Werte lassen einerseits die Interpretation zu,

[13] Zur Begründung dieses Wertes siehe Abschnitt 3.2.2.

[14] Kowal geht genauer auf unterschiedliche Vorgehensweisen bei der Partitionierung des Lautstromes in Pausen und Verbales ein, vor allem in Bezug auf die Vergleichbarkeit von Untersuchungsergebnissen.

daß die Artikulationsgeschwindigkeit eines Sprechers nur innerhalb eines wohldefinierten Intervalls begrenzter Ausdehnung um einen Mittelwert schwankt. Andererseits deuten sie darauf hin, daß individuelle Unterschiede eine große Rolle spielen und die Ableitung eines Basiswertes für den sprecherspezifischen Mittelwert notwendig ist. Nur so kann bewertet werden, ob die Artikulationsgeschwindigkeit eines Sprechers verhältnismäßig *hoch*, *normal* oder *niedrig* ist. Z.B. entspricht die niedrigste Artikulationsgeschwindigkeit des Anrufers in Dialog 2 noch nicht einmal dem Mittelwert der Anruferin in Dialog 4 (Abb. 4.4).


Abbildung 4.3: Vergleich von Mittelwerten und Standardabweichungen der Artikulationsgeschwindigkeiten des Mechanikers

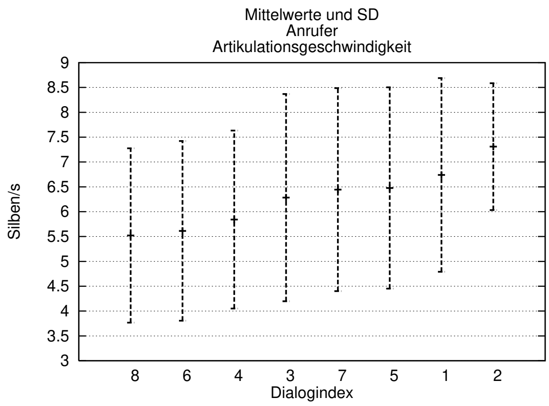


Abbildung 4.4: Vergleich von Mittelwerten und Standardabweichungen der Artikulationsgeschwindigkeiten der Anrufer

Tabelle 4.5 gibt Empfehlungen für das Wertespektrum und Verteilungsparameter für Standardverteilungen, die für die Artikulationsgeschwindigkeit Anwendung finden sollten.[15] Abbildung 4.5 zeigt, daß der Anteil von Äußerungen mit Werten der Artikulationsgeschwindigkeit

[15]Diese Werte müssen durch die Einbeziehung der Silbendehnungen in die Artikulationsgeschwindigkeit (siehe 3.2.3) möglicherweise noch leicht nach oben korrigiert werden, wenn eine getrennte Modellierung der Dauer gedehnter Silben erfolgen soll.

in unterschiedlichen Intervallen dieses Spektrums annähernd normalverteilt ist. Eine Granularität von 5, die das Intervall [0,10] in gleichmäßige Abschnitte teilt, erscheint sinnvoll, wobei im Gegensatz zur Abbildung die Werte oberhalb 10 Silben/s unter vergleichsweise geringem Informationsverlust im Teilintervall [8,10] erfaßt werden sollten. Das für die Modellierung interessante **Kernintervall** liegt zwischen 2 und 8 Silben/s; das Randintervall 8 bis 10 Silben/s dient als **Pufferzone** für die Erfassung sehr hoher Werte. Der Wert für die Granularität ist entsprechend so zu interpretieren, daß zur Aufteilung des angegebenen Intervalls [min,max] ein Pufferintervall oberhalb *max* und eines unterhalb *min* hinzukommt, das die hohen (niedrigen) Werte aufnehmen kann.

| **Artikulationsgeschwindigkeit - Erwartungswert -** (Silben/s) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 5.358 | 1.578 |
| Anrufer | 6.035 | 1.898 |

| **Artikulationsgeschwindigkeit - Intervalle -** (Silben/s) | | |
|---|---|---|
| **Min.** | **Max.** | **Granularität** |
| 2.0 | 8.0 | 1 + 3 + 1 |

| **Artikulationsgeschwindigkeit - Verteilungen -** | | | | | |
|---|---|---|---|---|---|
| **Sprecher** | **0.0 - 2.0** | **2.0 - 4.0** | **4.0 - 6.0** | **6.0 - 8.0** | **$>$ 8.0** |
| Mechaniker | 0.020 | 0167 | 0.469 | 0.313 | 0.031 |
| Anrufer | 0.011 | 0.139 | 0.314 | 0.405 | 0.132 |

Tabelle 4.5: Referenzwerte Artikulationsgeschwindigkeit

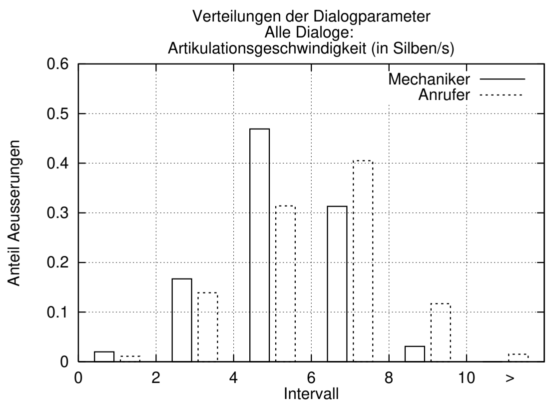


Abbildung 4.5: Verteilung der Äußerungen auf Intervalle mit charakteristischer Artikulationsgeschwindigkeit

Die für eine Äußerung erhobene Artikulationsgeschwindigkeit basiert auf den mittleren Silbendauern aller Silben der Äußerung. Die Länge von Silben mit phonetisch gleichem Onset, Nukleus und Coda ist aufgrund ihrer ggf. bedeutungsdifferenzierenden Funktion in der jeweiligen Sprache nicht beliebig wählbar. Vielmehr wird die Dauer der Silben durch die phonologischen Eigenschaften der lexikalischen Form des entsprechenden Wortes vorbestimmt. Bei einer großen Stichprobe, d.h. einer langen Äußerung, werden Unterschiede in der Dauer der Einzelsilben geglättet. Bei kurzen Äußerungen dagegen verschieben sich die Akzentuierungsverhältnisse zugunsten von Wort-Hauptakzenten. Z.B. haben einsilbige Worte lediglich einen

Hauptakzent, der jedoch stärker gedehnt wird als Nebenakzente. Dehnungen, die Pausen vorangehen, fallen ebenfalls stärker ins Gewicht.

Diese Unterschiede bleiben nicht ohne Wirkung auf die Artikulationsgeschwindigkeit: Aus den Grafiken 4.6 und 4.7 wird ersichtlich, daß kurze Äußerungen im Schnitt den niedrigeren Teil des Spektrums der Artikulationsgeschwindigkeit abdecken, während längere näher an den Mittelwerten liegen. Die Standardabweichungen sind jedoch für kurze Äußerungen größer, d.h. die Streuung der Werte breiter. Beim Vergleich von Äußerungen mit weniger als 3 gegenüber denen mit mehr als 7 Silben ergaben sich folgende Unterschiede für die Standardabweichungen: Mechaniker 1.630 vs. 1.019; Anrufer 1.867 vs. 1.331. Ein Verzicht auf die Erhebung der Artikulationsgeschwindigkeit bei sehr kurzen Äußerungen erscheint daher im Hinblick auf eine korrekte Modellierung der Benutzereigenschaften sinnvoll.

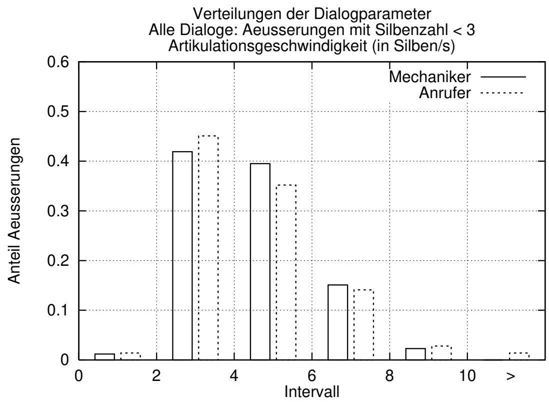


Abbildung 4.6: Verteilung der Äußerungen mit Silbenzahl $<3$ auf Intervalle mit charakteristischer Artikulationsgeschwindigkeit

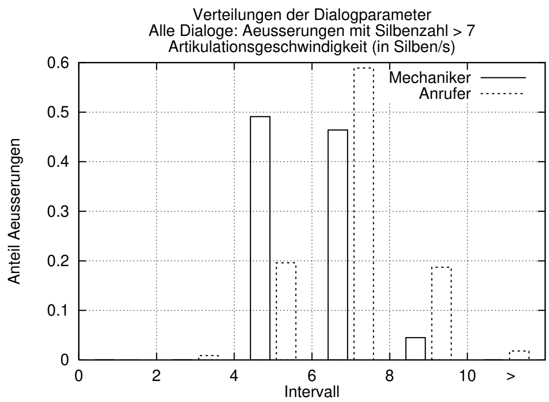


Abbildung 4.7: Verteilung der Äußerungen mit Silbenzahl $>7$ auf Intervalle mit charakteristischer Artikulationsgeschwindigkeit

#### 4.3.3.2 Pausen

**Pausendauer** Abbildung 4.8 gibt einen Überblick über den Anteil von Äußerungen mit einem charakteristischen Verhältnis der Gesamtpausendauer zur Gesamtäußerungsdauer an der Menge aller Äußerungen für die beiden Sprecher.[16] Zur Gesamtpausendauer werden dabei sowohl stille als auch gefüllte Pausen hinzugezählt. Aus der Abbildung wird deutlich, daß der Anteil ohne stille[17] oder mit nur kurzen gefüllten Pausen beim Anrufer deutlich höher ist als beim Mechaniker. Dieser Unterschied könnte u.a. dadurch verursacht worden sein, daß der Mechaniker über die gesamten Dialoge hinweg neben der Äußerungsplanung die z.T. komplexe Handlungsplanung für das weitere Vorgehen zur erfolgreichen Reparatur als Doppelaufgabe auszuführen hatte. Besonders sei darauf hingewiesen, daß Äußerungen mit einem Verhältnis von 0.8 bis 1 durchaus auftreten können, wobei Planungsanforderungen hierfür oft als Hauptursache identifizierbar sind. Die Äußerung des Mechanikers in einem charakteristischen Dialogausschnitt aus Dialog 1 (Bsp. 4.13) kann dafür als Beispiel dienen.

(4.13) <**MA001/7.884/**>
Waschbüsch . Hallo . Ich bin der Jens <Z> und ich steh' hier unten <P/.345/> in der Saarbrücker Straße und in meinem Auto , <P/.410/> da riecht 's im Innenraum schon seit ner Viertelstunde nach <D> Benzin <P/.596/> während der Fahrt .

<WE/.380/>

<**ME002/1.406/**>
<ahja/.408/> <P/.559/> <äh/.438/> .

[ Erneuter Sprecherwechsel ]

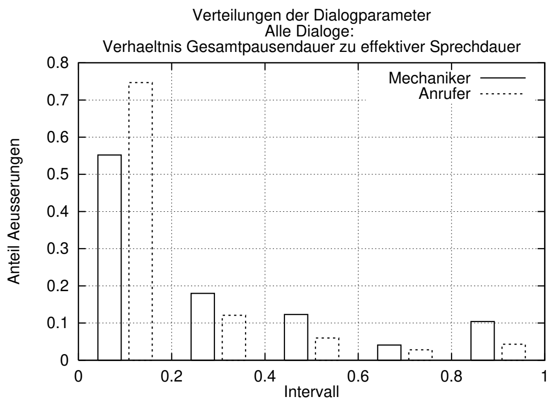


Abbildung 4.8: Verteilung der Äußerungen auf Intervalle mit charakteristischem Pausenanteil (alle Pausen)

[16]In Abbildungen, die Verteilungen von Werten auf Intervalle zeigen, werden stets nach oben halboffene Intervalle zugrunde gelegt. D.h., daß z.B. in Abbildung 4.8 im Intervall 0 - 0.2 alle Pausen enthalten sind, für deren Dauer $d$ gilt: $0 \leq d < 0.2$.

[17]Das Intervall 0 - 0.2 s entspricht pausen*freien* Äußerungen, da Pausen unter 200 ms nicht gemessen wurden; der Minimalanteil also 0.2 s beträgt.

**Pausenfüllung** Insgesamt wurden im Korpus 399 gefüllte und 479 stille Pausen[18] gezählt. Für die Interpretation des Aufteilungsverhältnisses der Gesamtpausendauer der Äußerungen auf stille und gefüllte Pausen bietet sich eine ähnliche Erklärung durch Unterschiede im Planungsaufwand an (Abbildung 4.9). Zwei Beobachtungen sind dabei wichtig:

- Der Mechaniker verwendete häufiger gefüllte Pausen als die einzelnen Anrufer.
- Gefüllte Pausen haben entweder einen sehr geringen Anteil an der Gesamtpausendauer oder einen sehr hohen. Mittelwerte sind weniger wahrscheinlich.

In Kapitel 3 wurde die Bedeutung der verzögernden Funktion gefüllter Pausen herausgehoben. Wenn diese Argumentation korrekt ist, läßt sich der hohe Anteil gefüllter Pausen beim Mechaniker u.a. auch wieder auf Belastung durch die Doppelaufgabe zurückführen. Es kommen zwar auch individuelle Unterschiede als Ursache in Frage, jedoch läßt ein Vergleich der Abbildungen 4.10 und 4.11 erkennen, daß die Mittelwerte des Mechanikers über den Mittelwerten fast aller Anrufer liegen (Gesamt-Mittelwert des Mechanikers: $\bar{x} = 0.515$, $SD = 0.374$). Diese Erkenntnis macht individuelle Unterschiede als alleinige Ursache für den höheren Anteil gefüllter Pausen beim Mechaniker weniger wahrscheinlich. Die Polarisierung des Anteils gefüllter Pausen am Gesamtpausenanteil kann ähnlich erklärt werden: Möglicherweise werden gefüllte Pausen nur als besonderes Mittel der Verzögerung eingesetzt, wenn dies unbedingt notwendig ist, dann jedoch massiv.

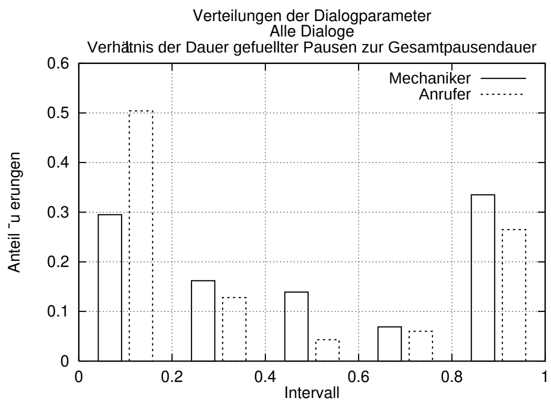


Abbildung 4.9: Verteilung der Äußerungen auf Intervalle mit charakteristischem Anteil gefüllter Pausen an der Gesamtpausendauer

Die Datenlage für intraphrasale Pausen (4.12) läßt als einzige Schlußfolgerung zu, daß intraphrasale Pausen in der Domäne ein seltenes Phänomen darstellen.[19]

[18] Zu den gefüllten Pausen zählen die Silbendehnungen. Pausen vor einem Sprecherwechsel gelten nicht als stille Pausen.

[19] Allerdings wurde die syntaktische Datenanalyse mithilfe eines partiellen Parsers (Chunker) vorgenommen, der auch in READY eingesetzt wird (siehe Abschnitt 5.3.2). Der Chunker erkennt NPs, APs, PPs und Konjunktionen, jedoch keine Verbalphrasen.

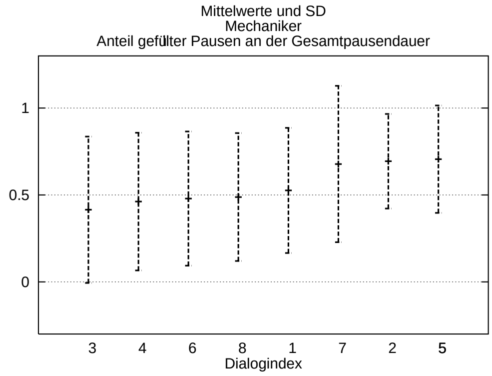


Abbildung 4.10: Vergleich von Mittelwerten und Standardabweichungen des Anteils gefüllter Pausen an der Gesamtpausendauer des Mechanikers

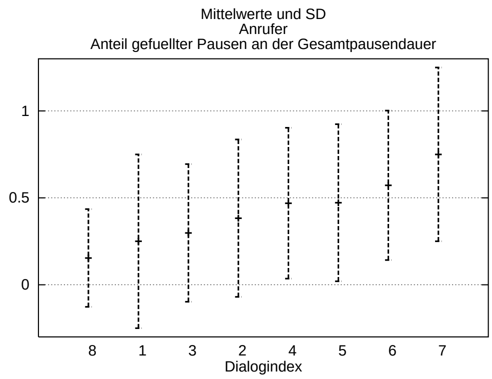


Abbildung 4.11: Vergleich von Mittelwerten und Standardabweichungen des Anteils gefüllter Pausen an der Gesamtpausendauer der Anrufer

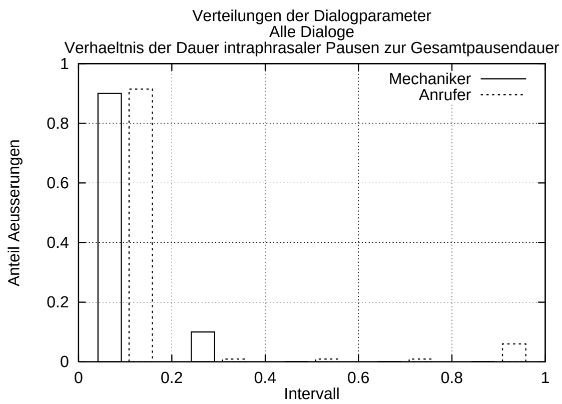


Abbildung 4.12: Verteilung der Äußerungen auf Intervalle mit charakteristischem Anteil intraphrasaler Pausen an der Gesamtpausendauer

Die gewählte Intervallgröße zur Erfassung des Gesamtpausenanteils von 0.2 und die zugehörige Granularität von 5 erscheinen für die Modellierung des Parameters *Gesamtpausenanteil* adäquat. Diese Unterteilung bietet sich auch für das Aufteilungsverhältnis der Pausendauer auf stille und gefüllte Pausen an, wobei hier die Obergrenze 1 zwangsläufig festliegt.

**Pausenhäufigkeit** Um die Häufigkeit von Pausen innerhalb einer Äußerung sinnvoll zu erfassen, wurde in Kapitel 3 vorgeschlagen, statt des Maßes 'Länge von Artikulationssequenzen' (Einheit: Silben/Pause) dessen Kehrwert 'Häufigkeit von Pausen' (Einheit: Pausen/Silbe) einzusetzen. Grund dafür ist vor allem, daß kein zusätzlicher, scheinbar besondere Informationen liefernder Sprechparameter angegeben werden muß, dessen unmittelbaren Zusammenhang mit Verzögerungsphänomenen nicht deutlich wird.

Für kurze Äußerungen wird dieser Parameter stärker von der Äußerungslänge beeinflußt, als von der Anzahl Pausen. Der Parameter ist daher ggf. nur sinnvoll zu erheben für Äußerungen mit einer minimalen Silbenzahl. Er ist undefiniert, wenn keine substantielle Äußerung erfolgte, sondern nur stille und gefüllte Pausen aneinandergereiht wurden. Der Parameter 'Verhältnis Pausendauer zur Gesamtsprechdauer' (der dann 1 ist) liefert in diesem Fall das entsprechende 'Ersatzmaß'. Die gewählte Granularität des Parameters von 7 liefert sinnvolle Ergebnisse. Tabelle 4.6 faßt die Daten für die Pausenparameter zusammen.[20]

| **Verhältnis GPD/GSD - Erwartungswert -** | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.252 | 0.314 |
| Anrufer | 0.144 | 0.243 |

| **Verhältnis GPD/GSD - Intervalle -** | | |
|---|---|---|
| **Min.** | **Max.** | **Granularität** |
| 0.0 | 1.0 | 5 |

| **Verhältnis GPD/GSD - Verteilungen -** | | | | | |
|---|---|---|---|---|---|
| **Sprecher** | **0.0 - 0.2** | **0.2 - 0.4** | **0.4 - 0.6** | **0.6 - 0.8** | **0.8 - 1.0** |
| Mechaniker | 0.020 | 0167 | 0.469 | 0.313 | 0.031 |
| Anrufer | 0.011 | 0.139 | 0.314 | 0.405 | 0.132 |

| **Relative Pausenzahl - Erwartungswert -** (Pausen/Silbe) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.180 | 0.276 |
| Anrufer | 0.116 | 0.302 |

| **Relative Pausenzahl - Intervalle -** (Pausen/Silbe) | | |
|---|---|---|
| **Min.** | **Max.** | **Granularität** |
| 0.0 | 0.6 | 6 + 1 |

| **Relative Pausenzahl - Verteilungen -** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Sprecher** | **0.0 - 0.1** | **0.1 - 0.2** | **0.2 - 0.3** | **0.3 - 0.4** | **0.4 - 0.5** | **0.5 - 0.6** | **> 0.6** |
| Mechaniker | 0.522 | 0.131 | 0.111 | 0.076 | 0.048 | 0.052 | 0.059 |
| Anrufer | 0.725 | 0.123 | 0.043 | 0.033 | 0.007 | 0.029 | 0.040 |

Tabelle 4.6: Referenzwerte Pausen

#### 4.3.3.3 Silbendehnungen

Silbendehnungen wurden entsprechend der im Kapitel 3 gegebenen Motivation nicht als eigenständiger Parameter erhoben. Ihre Anzahl ging in der Anzahl gefüllter Pausen auf, während

[20] GPD steht in dieser Tabelle für die Gesamtpausendauer, GSD für die Gesamtsprechdauer.

die erhöhte Silbendauer die Artikulationsgeschwindigkeit beeinflußte. Bei der Beurteilung von Silbendehnungen mußte rein perzeptiv vorgegangen werden, indem subjektiv als verlängert empfundenen Silben als Dehnungen charakterisiert werden. Die Probleme bei der Erhebung von Silbendehnungen wurde im Abschnitt 3.2.3 beschrieben.

#### 4.3.3.4 Einsatzlatenz

In Kapitel 3 wurde ausgeführt, daß gefüllte Pausen, gefolgt von weiteren gefüllten und stillen Pausen vor Äußerungsbeginn in die Berechnung der Einsatzlatenz einbezogen werden können. Um einen Vergleich der Korpusdaten mit den Werten der Arbeiten in Tabelle 3.10 zu ermöglichen, wurde die Einsatzlatenz bei der Datenauswertung jedoch nach Definition 3.6 zwischen dem Ende des Dialogturns des Dialogpartners und dem Beginn der ersten Artikulation eines Sprechers gemessen.

Die Verteilung der Einsatzlatenzen auf Intervalle unterschiedlicher Dauer zeigt Abbildung 4.13. Die durchschnittliche Einsatzlatenz bei der Beantwortung einer Partneräußerung in der vorliegenden Domäne liegt innerhalb des Intervalls [0 s; 1 s]. Der Anrufer hat einen höheren Anteil an Einsatzlatenzen, die über die Grenze von 2 s hinausgehen, da er z.T. Handlungen auszuführen hatte, die er ohne Rückmeldesignal direkt auf die Instruktion des Mechanikers hin begann.

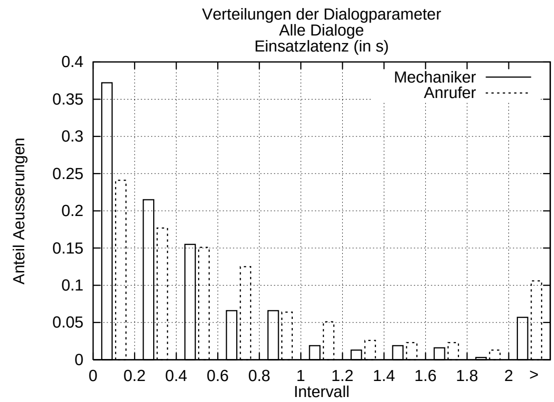


Abbildung 4.13: Verteilung der Einsatzlatenzen auf unterschiedliche Längenintervalle

Als Schlußfolgerung daraus ergibt sich u.a., daß ein System bereits nach Verstreichen einer Einsatzlatenz von etwa 1 s über einen starken Indikator für kognitive Belastung oder eine Nebenhandlung verfügt und erste Inferenzen möglich sind. In Abhängigkeit vom vorangehenden Dialogakt (Frage/Instruktion) kann eine weitgehend eindeutige Zuordnung zu einer dieser beiden Ursachen für die lange Einsatzlatenz ausgeführt werden. Abbildung 4.14 deutet an, daß eine Unterscheidung der sprachlichen Aufgabe, d.h. eine Differenzierung aufgrund des vorangehenden Dialogakts des Partners auch generell sinnvoll erscheint, wenn man die relative Dauer der Einsatzlatenz beurteilen möchte: Die Einsatzlatenz auf eine Instruktion hin war durchschnittlich stets länger als die Einsatzlatenz auf ein Frage. Verschiedene Basiswerte für die Beurteilung von *langer* oder *kurzer* Einsatzlatenz lassen sich schon durch diese Grobunterscheidung der Aufgabe erhalten. Aus der Abbildung geht auch hervor, daß es nicht notwendig ist, Einsatzlatenzen oberhalb der Grenze von 2 s zu differenzieren.

Es ergeben sich die in Tabelle 4.7 angegebenen Rahmenwerte. Die starke Standardabweichung für die Einsatzlatenz des Anrufers auf Dialogakte vom Typ `Instruktion` hin verstärkt die Bedeutung einer Differenzierung der Basiswerte für die Einsatzlatenz auf der Basis der Schwierigkeit der gestellten Aufgabe. Die geringen Einsatzlatenzen des Mechanikers kommen möglicherweise dadurch zustande, daß er sehr oft als Strategie des Äußerungsbeginns zunächst eine gefüllte Pause artikulierte.

| **Einsatzlatenz - Erwartungswert - (s)** | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.619 | 1.271 |
| Anrufer | 1.261 | 3.150 |
| Frage | 0.800 | 1.498 |
| Instruktion | 1.291 | 3.336 |

| **Einsatzlatenz - Intervalle - (s)** | | |
|---|---|---|
| **Min.** | **Max.** | **Granularität** |
| 0.0 | 1.6 | 8 + 1 |

| **Einsatzlatenz - Verteilungen -** | | | | |
|---|---|---|---|---|
| **Sprecher** | **0.0 - 0.4** | **0.4 - 0.8** | **0.8 - 1.2** | **> 1.2** |
| Mechaniker | 0.587 | 0.221 | 0.085 | 0.108 |
| Anrufer | 0.418 | 0.276 | 0.115 | 0.191 |
| Frage | 0.396 | 0.396 | 0.099 | 0.110 |
| Instruktion | 0.324 | 0.382 | 0.141 | 0.183 |

Tabelle 4.7: Referenzwerte Einsatzlatenz

In Abschnitt 3.2.6 wurde darauf hingewiesen, daß Überlappungen von Partneräußerungen, d.h. negative Einsatzlatenzen nicht berücksichtigt werden. Diese Unterbrechungen eines Sprechers können jedoch in vielen Kulturen als Mittel des Zeitgewinns gewertet werden, wenn sie sozialen Konventionen (wie Höflichkeit) zuwider laufen. Sie stellen damit ebenfalls ein mögliches Symptom für Zeitdruck dar. Unterbrechungen kamen im Korpus allerdings nur selten vor (siehe Abschnitt 4.3.5.1).

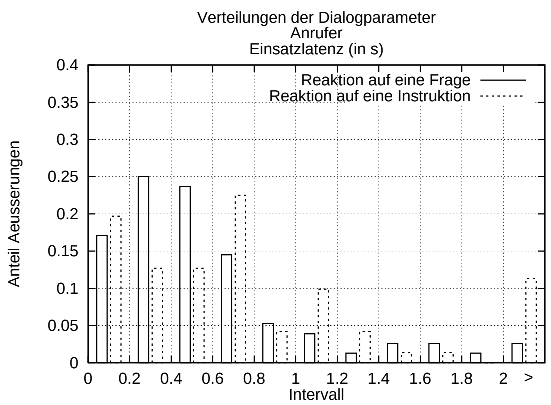


Abbildung 4.14: Einsatzlatenz des Anrufers bei unterschiedlichem vorangehendem Dialogakt des Mechanikers

### 4.3.4 Syntaktische Parameter

#### 4.3.4.1 Strukturprobleme: Anakoluthe, Unvollständigkeit und Extrapositionen

**Anakoluthe** Insgesamt traten lediglich 9 Fälle unkorrigierter ungrammatischer Äußerungen auf[21], die nicht unter den Unvollständigkeitsbegriff (Def. 3.17) fallen und nicht zu den in der Spontansprache akzeptierten Phänomenen gehören. Sie verteilen sich wie folgt auf Unterkategorien:

| | |
|---|---|
| - doppelt vorhandene Konstituenten | 3 |
| - fehlerhafte Verbstellung in Aussagen oder Subordinationen[22] | 3 |
| - andere | 3 |

In Übereinstimmung mit den Vorhersagen aus Kapitel 3 sind Linearisierungsprobleme für alle Fälle als mögliche Erklärung denkbar: Konstituenten treten an den falschen Positionen im Satz auf oder mußten doppelt realisiert werden, da die Entscheidung für die Satzstruktur während des Sprechens revidiert wurde.

Das sporadische Auftreten von Anakoluthen erlaubt keine neuen Schlußfolgerungen über die Wirkung von Zeitdruck oder Belastung. Da auch durch die fehlerhaften Konstruktionen der intendierte semantische Inhalt transportiert werden konnte, erscheint außerdem wahrscheinlich, daß die Sprecher den Aufwand für eine Korrektur gegenüber dem zu erreichenden Nutzen als zu hoch einschätzten und daher auf eine Revision der Äußerung verzichteten. Insofern kann in Übereinstimmung mit Kapitel 3 das Auftreten und die fehlende Korrektur als Symptom von Belastung Verwendung finden.

**Unvollständigkeit** Die Datenanalyse legt nahe, daß einige Verletzungen grammatischer Regeln im Sinne der in Kapitel 3 beschriebenen internen Unvollständigkeit - im Satz fehlende obligatorische Konstituenten - in der Spontansprache offenbar als Mittel der Verkürzung von Äußerungen eingesetzt werden. Aufgrund der Häufigkeit ihres Auftretens und der stets fehlenden Korrektur können sie schon fast als regelhaftes Phänomen angesehen werden. Bußmann (1990) erlaubt interne Unvollständigkeit nur beim sogenannten *Telegrammstil*. Da zu den im Korpus beobachteten Phänomenen neben der von Bußmann aufgeführten Subjektlöschung auch Objekt- und Artikellöschungen zählten, sollen diese im folgenden unter dem Begriff **erweiterter Telegrammstil** zusammengefaßt werden. Sie werden als elliptische Äußerungen annotiert und die entsprechenden Ergebnisse daher im Abschnitt 4.3.5.2 dargestellt.

Die Menge der verbleibenden unvollständigen Äußerungen, die nicht der erweiterten Telegrammstil-Kategorie zugeordnet werden konnten, war dagegen relativ klein: Insgesamt traten 12 Fälle auf.[23] Ursache dafür, daß der Sprecher den jeweiligen Fehler nicht korrigierte, kann wie bei den Anakoluthen die Einschätzung des Sprechers sein, daß der Aufwand der Korrektur in keinem Verhältnis zum dadurch erzielten Gewinn steht (siehe Bsp. 4.14. Andererseits ist es

[21] Korrigierte Anakoluthe wären ein Paradoxon. :-)

[22] Aussagesätze mit Verberststellung, Subordinationen mit Verbzweitstellung.

[23] Drei Fälle davon sind möglicherweise auf technische Probleme zurückzuführen, die sich lediglich in einer Verbindungsunterbrechung ohne Störgeräusch äußerten und daher nicht als solche erkennbar waren.

auch möglich, daß die Fehler von den Sprechern nicht als Fehler registriert wurden. In diesem Fall wäre ein fehlerhaftes Monitoring als Ursache zu identifizieren.

(4.14) **Me008, Dialog 1:**
Du mußt ihn beheben . <Z> <Aha/.288/> <Z> Gut . <Also/.191/> Nicht nur ich <//VAFIN> <//OA> rausfinden , sondern du mußt ihn auch beheben .

**Extrapositionen** Im Abschnitt 3.3.3 wurde auf den möglichen Zusammenhang von kognitiver Belastung mit Extrapositionsphänomenen verwiesen. Der Sprecher kann lokal seinen Generierungsaufwand verringern, indem er dem restlichen Satz voranstehend (komplexe) Referenzen in Form von Nominalphrasen oder Präpositionalphrasen generiert und im Kernsatz nur anaphorisch darauf bezug nimmt. Rechtsextrapositionen ergeben sich, wenn Referenten, Modifikatoren oder Konjunkte dem eigentlichen Satz nachgeordnet werden. Die Beispiele 3.18 bis 3.21 aus dem Korpus in Kapitel 3 verdeutlichen die Form solcher Äußerungen.

In den Dialogen der Feldstudie traten die genannten Extrapositions-Konstruktionen relativ häufig auf. Sie verteilen sich wie in Tabelle 4.8 angegeben. Bemerkenswert daran ist, daß der Mechaniker doppelt so viele Rechtsextrapositionen wie der Anrufer bei der gleichen Anzahl von Dialogturns verwendete. Um eine geeignete Einschätzung über die Extrapositionsaktivität der beiden Dialogteilnehmer zu erhalten, wurden die Linksextrapositionen an der Anzahl aller Referenzen in der jeweiligen Äußerung relativiert und die Erwartungswerte für Linksextrapositionen (Tab. 4.8) aufgrund dieses relativierten Maßes errechnet. Für Rechtsextrapositionen ist eine solche Relativierung nicht angebracht, da davon nicht nur Referenzen betroffen sind, sondern auch Komplemente und Modifikatoren nachgestellt werden können.

| | ges. | mit ZD | ohne ZD | Mech. | Anrufer |
|---|---|---|---|---|---|
| Rechtsspezialisierungen | 30 | 14 | 16 | 20 | 10 |
| Linksextrapositionen | 10 | 5 | 5 | 4 | 6 |

| **Rechtsextrapositionen - Erwartungswert -** (Häufigkeit) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.076 | 0.309 |
| Anrufer | 0.032 | 0.177 |
| gesamt | 0.054 | 0.253 |

| **Linksextrapositionen - Erwartungswert -** (rel. Anzahl von Referenten) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.033 | 0.167 |
| Anrufer | 0.014 | 0.077 |
| gesamt | 0.024 | 0.134 |

Tabelle 4.8: Referenzwerte Extrapositionen

### Fazit

Nicht akzeptierbare, nicht korrigierte unvollständige Strukturen treten äußerst selten auf. Sie sind als Symptom kognitiver Belastung in Übereinstimmung mit der Literatur von untergeordneter Bedeutung.

Im Abschnitt 3.3.3 wurden zwei hypothetische Erklärungen für das Entstehen von Extrapositionen formuliert. Das Zutreffen der Hypothesen ist experimentell noch zu klären. Bei der Verarbeitung in READY können jedoch aufgrund der Vorhersagen aus dem den Modellen kognitiver Ressourcen und dem Sprachproduktionsmodell Extrapositionen als Symptom für Belastung zur Anwendung kommen, solange gegenteilige Evidenz fehlt.

#### 4.3.4.2 Selbstkorrekturen

Insgesamt wurden im Korpus 47 Selbstkorrekturen im Leveltschen Sinne in 38 Äußerungen gezählt; das entspricht 6.1 % aller Äußerungen. Wiederholungen und Fragmente werden, wie bereits oben angedeutet, einzeln untersucht. Die Zusammenfassung aller Korrekturen, Wiederholungen und Fragmente ergibt 113 Äußerungen (18 %). Die einzelnen Korrekturen verteilten sich wie in Tabelle 4.9 angegeben auf die einzelnen Klassen. Zur Menge der *Difference*- Repairs kommen 22 Fehlansätze, die extra analysiert werden.

| Klasse | gesamt | Mech. | Anrufer | ohne ZD | mit ZD |
|---|---|---|---|---|---|
| *gesamt* | 47 | 32 | 15 | 11 | 36 |
| *Error* | 22 | 15 | 7 | 5 | 17 |
| Lexikalische | 7 | 6 | 1 | 2 | 5 |
| Syntaktische | 13 | 8 | 5 | 3 | 10 |
| Phonologische | 1 | 1 | 0 | 0 | 1 |
| Flexionsmorphologisch | 1 | 0 | 1 | 0 | 1 |
| *Appropriateness* | 20 | 13 | 7 | 5 | 15 |
| Ambiguität | 5 | 3 | 2 | 2 | 3 |
| Terminologie | 14 | 10 | 4 | 3 | 11 |
| Kohärenz | 1 | 0 | 1 | 0 | 1 |
| *Difference* | 4 | 4 | 0 | 1 | 3 |
| *Rest* | 1 | 0 | 1 | 0 | 1 |

| | gesamt | | ohne ZD | | mit ZD | |
|---|---|---|---|---|---|---|
| Kriterium | abs. | rel. | abs. | rel. | abs. | rel. |
| wohlgeformt nach Regel 3.24 | 45 | 95.8 | 10 | 90.9 | 34 | 94.4 |
| Korrekturterm vorhanden | 15 | 31.9 | 3 | 27.3 | 12 | 33.3 |
| stille Pause vorhanden | 23 | 48.9 | 7 | 63.6 | 16 | 44.4 |
| Pause und/oder Korrekturterm | 34 | 72.3 | 9 | 81.9 | 25 | 69.4 |

Tabelle 4.9: Verteilungen von Selbstkorrekturen auf die unterschiedlichen Klassen und ihre Eigenschaften

Die Abbildung 4.15 zeigt, daß Selbstkorrekturen beim Mechaniker fast doppelt so häufig waren, wie bei den Anrufern. Eine Interpretation dieses Ergebnisses ist schwierig: sowohl interindividuelle Unterschiede als auch Belastungsbedingungen und Abhängigkeit des Wertes von den Äußerungstypen, die in seiner Dialogrolle möglich sind, sind mögliche Erklärungen. Die Werte für die Zeitdruck-Bedingung waren mit den Werten aus der Bedingung ohne Zeitdruck fast identisch, daher werden sie hier nicht konkreter angegeben.

In Kapitel 5 werden verschiedene Arbeiten zur robusten Verarbeitung außergrammatischer Phänomene diskutiert. Einige von ihnen geben Zahlen an, die die Häufigkeit von Selbstkorrekturen in den jeweils untersuchten Korpora beschreiben. Während im READY- Korpus 18% aller Äußerungen Korrekturen enthielten (Wiederholungen und Fehlansätze der Vollständigkeit halber eingeschlossen), reicht die Spannbreite in anderen Korpora von 6.1% (Bear, Dowding & Shriberg, 1992) und 10% (Nakatani & Hirschberg, 1993) über 20% (Ruland, Rupp, Spilker, Weber & Worm, 1998) bis zu 25% (Heeman & Allen, 1994) bzw. 50% (Hindle, 1983). Levelt (1983) gibt 34% an. Diese Zahlen deuten auf die Korpusabhängigkeit der Häufigkeit von Selbstkorrekturen und damit erneut auf die Notwendigkeit einer domänen- und situationsspezifischen Analyse hin.

Bemerkenswert an den Terminologie-Korrekturen (AL) des Mechanikers ist folgendes: sie ersetzten in 6 der 10 Fälle ein gewähltes Lemma durch ein *allgemeineres* Konzept und nur in

4 Fällen durch ein *spezialisierteres*. Da es sich um Selbstkorrekturen handelt, ist klar, daß der Auslöser für die Reparatur ein Alarm des Monitorings sein muß. Offenbar korrigierte der Sprecher jedoch bei einer Spezialisierung Unangemessenheiten durch Unterspezifikation, die durch Verarbeitungsbeschränkungen während der Generierung der Originaläußerung entstanden sein könnten auch für sich selbst (Bsp. 4.16), während die Korrektur hin zu einem allgemeineren Konzept hörerbezogene Gründe hat (Bsp. 4.15).[24]

Die Werte in den Tabellen 4.10 können zusammengefaßt werden, um die Häufigkeit von Selbstkorrekturen als binäres Merkmal zu behandeln (d.h. Erwartung einer Selbstkorrektur in einer beliebig langen Äußerung). Diese Modellierung ist in Anbetracht der wenigen Daten adäquater. Kapitel 6 wird genauer darauf eingehen.

(4.15) **Me040, Dialog 3:**
Und <+/> <@M1> wenn <@M2> der <@R1> durchgebrannt <@M3> ist </+> <Z/.154/> , <@M1> wenn <@M2> der <@R1> nicht <@R1> mehr <@R1> durchgehend <@M3> ist , <Z/.163/> ist sie kaputt <#Schnarren/.550/> <P/4.987/> .

(4.16) **Me022, Dialog 8:**
Weiter links <Z> <+/> <@M1> <///TTS> <//APPR> Richtung <D> <P/.463/> <äh/.340/> <@R1> Tür </+> . <P/.621/> <@M1> <///TTS> <//APPR> Richtung <@R1> Fahrertür <///PP-MO> .

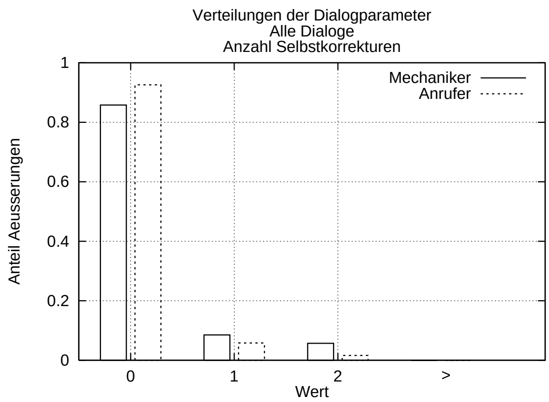


Abbildung 4.15: Verteilung der Äußerungen mit Selbstkorrekturen für beide Dialogteilnehmer

| **Häufigkeit von Selbstkorrekturen - Erwartungswert -** (rel. Zahl/Silbe) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.205 | 0.550 |
| Anrufer | 0.011 | 0.030 |

| **Häufigkeit von Selbstkorrekturen - Intervalle -** (rel. Zahl/Silbe) | | |
|---|---|---|
| **Min.** | **Max.** | **Granularität** |
| 0.05 | 0.15 | 1 + 1 + 1 |

| **Häufigkeit von Selbstkorrekturen - Verteilungen -** (rel. Zahl/Silbe) | | | |
|---|---|---|---|
| **Sprecher** | **< 0.05** | **0.05 - 0.15** | **> 0.15** |
| Mechaniker | 0.903 | 0.093 | 0.004 |
| Anrufer | 0.938 | 0.061 | 0.001 |

Tabelle 4.10: Referenzwerte Selbstkorrekturen

[24]Siehe dazu auch die Diskussion der Dummy-Phrasen in Abschnitt 3.5.4.

**Fazit**
Die aus dem Korpus abgeleiteten Erwartungswerte für das Auftreten von Selbstkorrekturen beschreiben zwar die Häufigkeit in domänenspezifischen Dialogen, können jedoch aufgrund der widersprüchlichen Ergebnisse der Arbeiten in Tabelle 3.12 für die Modellierung in Kapitel 6 derzeit leider keinen Nutzen bringen. Die differenziertere Betrachtung der AL-Korrekturen kann sich als hilfreich bei der Beurteilung der Hörerorientierung eines Sprechers erweisen.

Dagegen erwies es sich ganz deutlich, daß Selbstkorrekturen eine Ausnahme darstellen, die nicht der Wohlgeformtheitsregel 3.24 entsprechen. Ebenso darf bei Selbstkorrekturen die Anwesenheit eines Korrekturterms oder einer stillen Pause standardmäßig erwartet werden. Die Zusammenhänge der beiden Strukturregeln für Selbstkorrekturen mit kognitiver Belastung bleiben weiterhin jedoch hypothetisch und sollten nur unter dem Vorbehalt weiterer experimenteller Bestätigung Einsatz finden.

#### 4.3.4.3 Wiederholungen

Im Korpus traten insgesamt 33 Wiederholungen auf (in 30 Äußerungen). Sie verteilen sich fast gleichmäßig auf die unterschiedlichen Komplexitäts-Kategorien.

| Typ | Anzahl |
|---|---|
| Phoneme | 11 |
| Worte | 12 |
| Wortgruppen | 10 |

Ein Unterschied unter Zeitdruck besteht bei der Verteilung der Äußerungen mit Wiederholungen genausowenig wie für Selbstkorrekturen. Eine weitere interessante Parallele von Wiederholungen und Selbstkorrekturen, die zu Levelts Klassifikation von Wiederholungen *als* Korrekturen paßt, ist die fast doppelt so hohe Anzahl an Wiederholungen für den Mechaniker (Abb. 4.16).

Wie sich schon in der in Abschnitt 3.3.5 dargestellten Literatur andeutet, ist es notwendig, die Anzahl der Wiederholungen an der Gesamtäußerungsmenge zu relativieren. Eine Wiederholung in einem kurzen Satz wiegt dadurch schwerer als eine Wiederholung in einem längeren Satz.[25] Die in Tabelle 4.11 angegebenen Werte beziehen sich auf diese an der Silbenzahl relativierte Anzahl von Wiederholungen. Die gewählten Intervallgrenzwerte und die Granularität wurde durch Zusammenfassung zu kleiner Intervalle und Festlegung von Pufferintervallen für Extremwerte so festgelegt, daß möglichst keine Intervalle Null-Werte enthalten und daß die Werte aufeinanderfolgender Teilintervalle möglichst nur wenig schwanken. Auch Wiederholungen können jedoch als binäres Maß behandelt werden.

[25]Weiterhin läßt sich auch noch die Anzahl wiederholter bzw. korrigierter *Silben* an der Gesamtmenge artikulierter Silben einer Äußerung relativieren. Dieses zweite Relativierung verwischt jedoch das ohnehin intuitiv nicht sehr aussagekräftige Maß der *relativen Anzahl* an Wiederholungen weiter.

| Wiederholungen - Erwartungswert - (rel. Anzahl) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.006 | 0.026 |
| Anrufer | 0.004 | 0.025 |

| Wiederholungen - Intervalle - (rel. Anzahl) | | |
|---|---|---|
| **Min.**[a] | **Max.** | **Granularität** |
| 0.05 | 0.2 | 1 + 1 + 1 |

[a]Dieser Wert entspricht 5 Korrekturen auf 100 Silben.

| Anzahl Wiederholungen - Verteilungen - | | | |
|---|---|---|---|
| **Sprecher** | **$<$ 0.05** | **0.05 - 0.15** | **$>$ 0.15** |
| Mechaniker | 0.958 | 0.034 | 0.006 |
| Anrufer | 0.975 | 0.011 | 0.015 |

Tabelle 4.11: Referenzwerte Wiederholungen

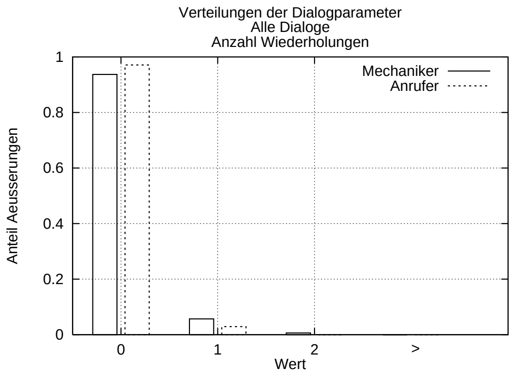


Abbildung 4.16: Verteilung der Äußerungen mit Wiederholungen für beide Dialogteilnehmer

## 4.3.5 Syntaktisch-semantische Parameter

### 4.3.5.1 Fehlansätze und Satzfragmente

Satzfragmente wurden als Voraussetzung für einen Fehlansatz klassifiziert. Im Kapitel 3 wurde bereits auf die Schwierigkeit verwiesen, Fehlansätze von reinen Satzfragmenten zu unterscheiden, wenn der Sprecher nach dem Fragment die Äußerung fortsetzt. Im allgemeinen kann dafür eine Beurteilung der Pausenlänge zwischen dem Ende des Fragments und dem Neuansatz sowie ein Vergleich des semantischen Wertes von Neuansatz und Fragment hilfreich sein. Aus den Daten der Feldstudie lassen sich keine Erkenntnisse über einen möglichen Minimalwert für Pausen ableiten, die Fragment und Neuansatz trennen. Dafür müssen größere Korpora herangezogen werden. Da außerdem vor allem zur Beurteilung der semantischen Ähnlichkeit keine geeigneten formalen Beurteilungskriterien zur Verfügung stehen, wird hier eine zusammenfassende Aufstellung von Fehlansätzen und Satzfragmenten vorgenommen. Diese Vereinfachung erfolgt bereits auch im Hinblick auf Möglichkeiten der automatischen Erkennung von Symptomen, die in Abschnitt 5 diskutiert werden.

Von den 628 Äußerungen von Mechaniker und Anrufer enthielten insgesamt 60 mindestens ein Satzfragment. Die unterschiedlichen Ursachen, die sich für die Entstehung von Satzfragmenten ergaben, sind zusammen mit den Häufigkeiten ihres Auftretens in Tabelle 4.12 dargestellt. Relativierte Werte sind bezogen auf die Gesamtäußerungszahl der Dialoge unter beiden

Bedingungen.[26] Bei *echten Fragmenten* besteht die Äußerung lediglich aus einem Fragment oder endet mit einem Fragment. Der Grund dafür ist i.a. nicht erkennbar. Bei *Fragmenten aufgrund geänderten Kontexts* wird die Äußerung durch während des Sprechens verfügbare neue Information oder Beobachtungen unterbrochen. Die Kategorien B, C und D der Tabelle stellen die Menge der Fehlansätze, die bei Levelt (1983) als Teilmenge der *Difference-Repairs* analysiert werden.

| Klasse | gesamt | ohne ZD | | mit ZD | |
|---|---|---|---|---|---|
| | | abs. | rel. | abs. | rel. |
| A) Echte Fragmente | 24 | 9 | 3.96 | 15 | 3.74 |
| B) Fragmente aufgrund geänderten Kontexts | 7 | 1 | 0.44 | 6 | 1.50 |
| C) Fragmente gefolgt von einem Satz mit unterschiedlicher Bedeutung | 6 | | | 6 | 1.50 |
| D) Fragmente gefolgt von einem Satz mit gleicher oder ähnlicher Bedeutung | 9 | 5 | 2.20 | 4 | 0.99 |
| E) Unterbrechungen durch den Dialogpartner | 7 | | | 7 | 1.75 |
| F) Technische Störungen | 8 | | | 8 | 1.99 |
| G) Vorstrukturierung für eine Frage | 1 | | | 1 | 0.25 |

Tabelle 4.12: Verteilung von Fragmenten auf unterschiedliche Ursachen

Wie bereits bei Wiederholungen und Selbstkorrekturen stellen auch die in Tabelle 4.13 angegebenen Werte die an der Silbenzahl relativierte Anzahl von Fragmenten dar. Für die Festlegung der Intervallgrenzen und Granularität kann ebenfalls parallel zu diesen beiden Symptomen vorgegangen werden. Auch hier kann aufgrund der niedrigen Werte in den verschiedenen Intervallen, die auf eine von Null verschiedenene Anzahl an Fragmenten hinweisen, eine binäre Modellierung adäquater sein. Unterschiede in der Häufigkeit von Fragmenten zwischen den beiden Dialogrollen ergaben sich aus der Feldstudie nicht. Dagegen deutet Abbildung 4.17 auf eine Zunahme der Anzahl an Fragmenten unter Zeitdruck hin. Diese Beobachtung sollte einer genaueren experimentellen Untersuchung unterzogen werden.

| **Fragmente und Fehlansätze - Erwartungswert -** (rel. Anzahl/Silbe) | | |
|---|---|---|
| **Sprecher** | $\bar{x}$ | $SD$ |
| Mechaniker | 0.026 | 0.113 |
| Anrufer | 0.023 | 0.107 |

| **Fragmente und Fehlansätze - Intervalle -** (rel. Anzahl/Silbe) | | |
|---|---|---|
| **Min.** | **Max.** | **Granularität** |
| 0.05 | 0.35 | 1 + 1 + 1 |

| **Fragmente und Fehlansätze - Verteilungen -** (rel. Anzahl/Silbe) | | | |
|---|---|---|---|
| **Sprecher** | **< 0.05** | **0.05 - 0.25** | **> 0.25** |
| Mechaniker | 0.920 | 0.042 | 0.038 |
| Anrufer | 0.913 | 0.054 | 0.032 |

Tabelle 4.13: Referenzwerte Fragmente und Fehlansätze

#### 4.3.5.2 Ellipsen

In den Dialogen konnten neben der normalen Ellipsenbildung[27] die beiden besonderen, in Abschnitt 4.3.4.1 bereits erwähnten Arten von Verbellipsen beobachtet werden:

- Verwendung eines Infinitivs für die Dialogakte der Instruktion (beim Mechniker), was einer Elision des Subjekts entspricht und

[26] Diese Aufstellung zeigt, daß bei der Erhebung von Fragment-Daten zunächst auf die Bedingungen Störung und Unterbrechung (durch den Dialogpartner) getestet werden, da diese einen nicht unwesentlichen Anteil an der Gesamtzahl stellen und damit nur Satzfragmente als Indikator kognitiver Belastung eingesetzt werden, bei denen der Sprecher die *technische* Möglichkeit hatte, seinen Satz zu beenden.

[27] Elision der Verbalphrase in Konjunktionen oder bei Erwiderungen auf Fragen.

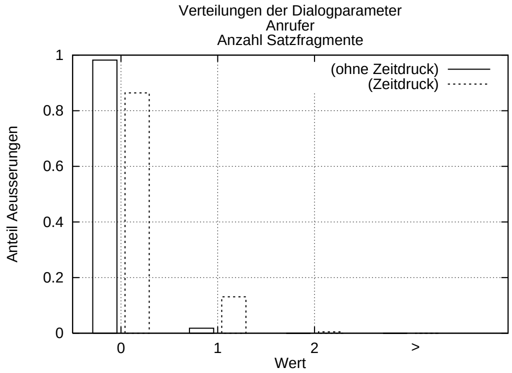


Abbildung 4.17: Verteilung der fragmentarischen Äußerungen für den Anrufer

- Verwendung eines Partizip unter Elision des Auxiliars und des Subjekts.

Die Korpusanalyse ergab außerdem zahlreiche Ellipsen, die mit den lexikalischen Ellipsen Bußmann (1990) vergleichbar sind und hier unter den *erweiterten Telegrammstil* gerechnet werden. Dabei wurde entweder das Subjekt oder das direkte Objekt eines Satzes ausgespart, wenn es im Diskurskontext besonders markiert, d.h. *salient* war. Salienz in der Diskursgeschichte ergibt sich durch eine Erwähnung des Referenten in der vorangehenden Äußerung bzw. die Fokussierung des gleichen Referenten über mehrere Äußerungen hinweg (z.B. bei der Instruktion verschiedener Handlungen an ein und demselben Bauteil, wie Öffnen, Auffüllen und Schließen des Kühlmittelbehälters). Kombinationen von beiden Ellipsenformen traten ebenfalls auf.

Im Korpus waren 195 Dialogturns mit elliptischen Äußerungen enthalten. Das sind 31.0% aller Dialogturns von Mechaniker oder Anrufer.[28] Einige Äußerungen enthielten mehrfache Ellipsen, die innerhalb ein und desselben Satzes auftreten konnten. Die elliptischen Äußerungen (inklusive Telegrammstil) verteilten sich prozentual fast identisch auf Dialoge unter Zeitdruck und ohne Zeitdruck: Von 261 Ellipsen wurden 163 unter Zeitdruck (124 Äußerungen = 30.92%) und 98 ohne Zeitdruck (71 Äußerungen = 31.28%) gebildet. In Abbildung 4.18, die die prozentuale Verteilung für beide Dialogteilnehmer zeigt, wurde für jede Äußerung die 'Anzahl elliptischer Äußerungsbestandteile 'zusammengefaßt. Beispiel 4.17 macht klar, wie diese Zusammenfassung erfolgte.

(4.17) **Me024, Dialog 6**
<äh/.571/> Mal <D> vorsichtig <///TTS> <//OA> aufdrehen <///VVINF> .
<P/1.704/> Nicht ganz <///ADV-MO> , nur mal <///TTS> <//OA> lösen
<///VVINF> .

Die Beispieläußerung besteht aus zwei Sätzen. Beide Sätze sind elliptisch durch die Verwendung eines Infinitivs (VVINF). Außerdem fehlt in beiden Sätzen das direkte Objekt (OA). Da aus dem Kontext deutlich ist, welches Bauteil gemeint ist, wird das jedoch unter den Telegrammstil eingeordnet. Zusätzlich

[28]Es verbleiben 150 Dialogturns, d.h. 23.9%, falls man den erweiterten Telegrammstil herausläßt.

wird im zweiten Satz das Verb des ersten Satzes mit einem Adverb modifiziert, ohne es erneut zu artikulieren. Die Äußerung erhält daher eine Bewertung von 5 für die Ellipsis.

Der Vergleich der Mittelwerte für die Elliptizität von Äußerungen zeigt, daß es im Korpus wahrscheinlich keinen Einfluß des Zeitdrucks auf die Ellipsenbildung gab. Beim Mechaniker lag die durchschnittliche Elliptizität ohne Zeitdruck bei 0.561 (SD = 0.912) und mit Zeitdruck bei 0.542 (SD = 0.816). Bei den Anrufern ergaben sich 0.301 (SD = 0.549) ohne Zeitdruck und 0.268 (SD = 0.508) mit Zeitdruck. Dieses Bild wird auch nicht differenzierter, wenn man die Verteilung der Elliptizität auf die unterschiedlichen Formen der Ellipsis untersucht (Tabelle 4.14). Entweder die geringen Mengen an Vertretern der jeweiligen Kategorie erlauben gar keine weiteren Schlüsse oder der prozentuale Anteil an der Gesamtäußerungsmenge ist für die jeweilige Kategorie unter den beiden Bedingungen fast gleich groß.

| Tag[a] | gesamt | ohne ZD | | mit ZD | |
|---|---|---|---|---|---|
| | | abs. | rel. | abs. | rel. |
| Komplemente | | | | | |
| NP-SB | 8 | 2 | 0.88 | 6 | 1.50 |
| ??-SB | 1 | 1 | 0.44 | | |
| NP-OA | 3 | | | 3 | 0.75 |
| NP[b] | 7 | 2 | 0.88 | 5 | 1.25 |
| NE | 1 | 1 | 0.44 | | |
| OC | 1 | 1 | 0.44 | | |
| OP | 1 | | | 1 | 0.25 |
| Modifikatoren | | | | | |
| PP-MO | 29 | 9 | 3.96 | 20 | 4.99 |
| ADV-MO | 26 | 11 | 4.85 | 15 | 3.74 |
| ??-MO | 4 | 2 | 0.88 | 2 | 0.50 |
| Teile der Verbalphrase | | | | | |
| VVINF | 55 | 22 | 9.69 | 33 | 8.23 |
| VVPP | 12 | 1 | 0.44 | 11 | 2.74 |
| andere | | | | | |
| TTS[c] | 91 | 36 | 15.86 | 55 | 13.72 |
| WH | 12 | 8 | 3.52 | 4 | 1.00 |
| Fortsetzungen[d] | 6 | | | 6 | 1.5 |
| KON | 1 | 1 | 0.44 | | |
| Zweifelsfälle | 2 | 1 | 0.44 | 1 | 0.25 |

| Tag[a] | gesamt | ohne ZD | | mit ZD | |
|---|---|---|---|---|---|
| | | abs. | rel. | abs. | rel. |
| Komplemente | | | | | |
| SB | 26 | 10 | 4.40 | 16 | 3.99 |
| OA | 28 | 17 | 7.49 | 11 | 2.74 |
| OC | 4 | 2 | 0.88 | 2 | 0.50 |
| OP | 3 | | | 3 | 0.75 |
| PD | 1 | | | 1 | 0.25 |
| Teile der Verbalphrase | | | | | |
| VAFIN | 7 | 3 | 1.32 | 4 | 1.00 |
| VVPP | 3 | 1 | 0.44 | | |
| VVIZU | 1 | 1 | 0.44 | | |
| andere | | | | | |
| ART | 21 | 7 | 3.08 | 14 | 3.49 |
| APPR | 4 | 1 | 0.44 | 3 | 0.75 |

Tabelle 4.14: Links: Verteilung elliptischer Äußerungen (*realisierte* Konstituententypen)
Rechts: Verteilung der Äußerungen im 'Telegrammstil' (*ausgelassene* Konstituententypen)

Für die die Auslassungen im erweiterten Telegrammstil (100 Auslassungen in 91 Äußerungen) läßt sich ebenfalls kein deutlicher Unterschied für verschiedene Ausprägungen von Zeitdruck erkennen: 42 Auslassungen (36 Äußerungen = 15.86 %) ohne Zeitdruck stehen 58 Auslassungen (55 Äußerungen = 13.71 %) mit Zeitdruck gegenüber. Außer für die Auslassung

[a] Siehe Tabellen C.1 und C.2.

[b] An der Rektion der NP war der Kasus nicht erkennbar.

[c] 86 Äußerungen bei 91 Auslassungen insgesamt. 45 Äußerungen waren nicht anderweitig elliptisch. Von einer Auslassung können mehrere Konstituenten betroffen sein.

[d] Damit werden Fortsetzungen von begonnenen Partneräußerungen wie in Bsp. 4.18 oder Wiederholungen von Teilen der eigenen vorangehenden Äußerung durch einen Sprecher wie in Bsp. 4.19 erfaßt.

(4.18) $\mathcal{S}$ : 'Dann wird ...?' – $\mathcal{P}$ : '... das Klackern weniger'

(4.19) $\mathcal{S}$ : 'Könnte gelb und rot markiert sein.' – $\mathcal{P}$ : 'Gelb und rot markiert.' – $\mathcal{S}$ 'Könnte !'

von direkten Objekten (Klasse OA) ergibt sich auch hier entweder kein Unterschied in der Auftretenshäufigkeit oder die vorliegenden Daten sind zu spärlich, um einen Schluß zu erlauben.

(4.20) **ME024, Dialog 8:**
<///TTS> <//ART> Hebel gefunden <///VVPP>

(4.21) **ME015, Dialog 5:**
<///TTS> <//OA> Gefunden <///VVPP>?

(4.22) **ME004, Dialog 7:**
<///TTS> <//ART> Motorhaube öffnen <///VVINF> .

(4.23) **ME032, Dialog 5:**
<#Lachen/.257/> OK . <///TTS> <//OA> Auffüllen <///VVINF> .

Abbildung 4.18 zeigt einen deutlichen Unterschied der Elliptizität von Äußerungen für die beiden Dialogteilnehmer. Als Erklärung dafür kommen jedoch weder Belastungsursachen noch individuelle Unterschiede in Frage: Vielmehr war die Ellipsenbildung durch die Rolle des Sprechers im Dialog beeinflußt. Durch die Möglichkeit, Instruktionen und Nachfragen durch Infinitivverwendung bzw. Subjekt-/Objekt-Elision stark zu verkürzen, standen dem Mechaniker mehr Möglichkeiten offen, die er ganz offensichtlich auch ausgenutzt hat. Der Anrufer hatte umfangreiche und korrekte Beschreibungen zu geben, die eine Ellipsenbildung bei ihm wahrscheinlich in den Hintergrund haben treten lassen.

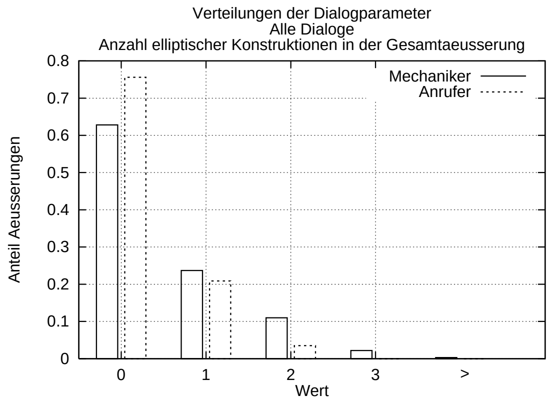


Abbildung 4.18: Verteilung der Ellipsenstärke auf unterschiedliche Intervalle für beide Dialogteilnehmer

**Fazit**
Es zeigte sich kein Unterschied in bezug auf die Elliptizität von Äußerungen unter Zeitdruck. Als Erklärung des unterschiedlich starken Einsatzes elliptischer Konstruktionen zwischen den

beiden Dialogteilnehmern können die unterschiedlichen *Möglichkeiten* zur Elision durch unterschiedliche sprachliche Aufgaben dienen. Ebenso ist die Möglichkeit, den erweiterten Telegrammstil anzuwenden, offenbar stark vom Ablauf des Dialoges und damit von der Dialoggeschichte geprägt. Weitere Untersuchungen zum Einfluß von Belastung und Zeitdruck auf das Ellipsenverhalten sollten diese Variable berücksichtigen.

In READY können Ellipsen allein noch kein Indikator für Belastung oder Zeitdruck sein. Aufgrund des engen Zusammenhanges von Ellipsen und Detailliertheit wird daher ein integrierender Ansatz verfolgt, der in Kapitel 5 vorgestellt wird.

#### 4.3.5.3 Echo-Äußerungen

Die Bewertung der Echo-Eigenschaft von Äußerungen ist eng an Dialogakte gebunden, die zur Klasse `Backchanneling` gehören. Es sind jedoch verschiedene Arten des Backchannelings möglich, die von Zeichen des Verständnisses bis zu expliziten Rückfragen reichen. Von den 157 Äußerungen des Typs `Backchanneling` enthielten 27 Äußerungen[29] (17.20%) eine Form der stereotypen Wiederholung, wie sie in 3.4.3 beschrieben wurde.

| vorangehender Dialogakt | mit nachfolgendem Backchanneling | | |
|---|---|---|---|
| | gesamt | davon Echo | relativ |
| DESCRIPTION | 77 | 15 | 19.5 |
| ANSWER | 35 | 6 | 17.1 |
| INSTRUCTION | 40 | 3 | 7.5 |

Tabelle 4.15: Verteilung stereotypen Backchannelings auf vorangehende Dialogakte

Tabelle 4.15 zeigt, welche Dialogakte mit einem Backchanneling-Signal beantwortet wurden und wieviele davon die beschriebene stereotype Form hatten. Es läßt sich einerseits erkennen, daß bei einer vorangehenden Frage die Wiederholung der Frage *keine* bevorzugte Dialogstrategie zu sein scheint. Andererseits zeigt sich, daß die Erkenntnisse von Beun (1985) den hohen Anteil stereotyper Wiederholungen auf Dialogakte vom Typ `Description` und `Answer` erklären können: `Answer` und `Description` sind in der vorliegenden Domäne diejenigen Dialogakte, die die meiste Information vermitteln, jedoch keinerlei Handlungspläne für die Verarbeitung dieser Information bereitstellen (wie etwa `Instruction`). Daher hat der Hörer in diesen Fällen entweder einen hohen Bedarf, die Informationen durch Wiederholung besser zu speichern oder er konnte durch Signalisierung des Verständnisses gleichzeitig Zeit gewinnen für seine eigene Handlungsplanung.[30]

Zur Beurteilung von Echo-Äußerungen im System READY brauchen nach der Datenlage die Antworten auf den Dialogakt `Question` des Systems nicht herangezogen werden. Die Rückmeldung auf den Dialogakt `Answer` würde bei der Modellierung vom System auszuführen sein, muß daher nicht beurteilt werden und entfällt somit. Wichtig wird die Bewertung der Stereotypie in READY aus diesem Grund vor allem für die Dialogakte, die vom Benutzer auf Systemäußerungen vom Typ `Description` folgen.

[29] 35 Äußerungen in Tabelle 4.15 kommen zustande, da einige Äußerungen mehrere Dialogakte enthielten.

[30] In Übereinstimmung mit Beun (1985) (siehe Kapitel 3) sind auch beide Ziele möglich.

In Übereinstimmung mit der Literatur hat sich gezeigt, daß stereotype Wiederholungen in den Dialogabschnitten, wo sie am meisten erwartet werden können, auch tatsächlich auftreten. Diese Form des Rückmeldesignals wird daher eine besondere Bedeutung für die Einschätzung der kognitiven Belastung in READY spielen.

### 4.3.6 Inhaltliche Parameter

#### 4.3.6.1 Zusätzliche Information

Im Kontext des Dialogsystems READY ist das vorrangige Ziel des Anrufers die Reparatur seines Fahrzeugs. Andere mögliche Ziele wie die Bildung sozialer Kontakte, Unterhalten des Mechanikers etc. dürfen mit Sicherheit als untergeordnet betrachtet werden. Entsprechend kann erwartet werden, daß die Informationsauswahl des Benutzers ähnlich wie z.B. in der Berichtssituation bei Rummer et al. (1995) sehr sachbezogen erfolgt. Ziel der Kommunikation ist es vor allem, dem Mechaniker alle notwendigen Informationen für die Reparatur des Fahrzeugs zu vermitteln.

Die Dialoge im Korpus liefen entsprechend dieser Hypothese tatsächlich sehr zielorientiert ab, so daß zusätzliche Informationen über nicht handlungsbezogene Ereignisse oder Situationen nur sehr selten auftrat. Einige nicht-ereignisbezogene Äußerungen überbrückten Passagen, in denen der Anrufer auf der Suche nach einem Bauteil oder mit der Erfüllung einer Instruktion des Mechanikers beschäftigt war. Es wurde bereits darauf verwiesen, daß zum gegenwärtigen Zeitpunkt kaum automatisch beurteilt werden kann, ob Information zusätzliche Information ist. Die Auswertung erfolgte daher nur informell, was sich auch in einer stark simplifizierten Modellierung niederschlagen wird.

Das weitgehende Fehlen von Zusatzinformation im Korpus verstärkt die Hypothese aus Kapitel 3, daß die Anwesenheit von Zusatzinformation als wichtiger Indikator für *fehlenden* Zeitdruck dienen kann: Wenn der Anrufer zusätzliche Informationen vermittelt, ist er entweder weniger an einer zielgerichteten Kommunikation interessiert oder (wie auch am Beispiel der ausführlichen 'Ja'-Antworten im folgenden Abschnitt deutlich wird) kann sich genügend Zeit für die Generierung einer stark hörerorientierten Äußerung lassen. Der Parameter *Zusatzinformation* wird im Abschnitt 4.3.7 erneut aufgegriffen.

#### 4.3.6.2 Detailliertheit

**Detailliertheit der Beantwortung von Fragen** Ein Aspekt der Detailliertheit, der in der gesichteten Literatur des Kapitels 3 nicht behandelt wurde, ist die Form der Beantwortung von Entscheidungsfragen: Eine Entscheidungsfrage fordert vom Dialogpartner die Bewertung des Wahrheitsgehaltes der Frage-Proposition. Wenn die Antwort positiv ist (d.h. vom Antwortenden die Wahrheit der Proposition behauptet wird), ist keine ergänzende Beschreibung dieser Proposition notwendig. Im Korpus wurde vom Antwortenden jedoch z.T. entweder die Frageproposition assertional wiederholt oder ergänzende Information vermittelt, die entweder die Gültigkeit des Wahrheitsgehaltes einschränkte oder die Frageproposition spezialisierte.[31]

[31] D.h., es wurde behauptet, daß die Proposition auch in einem stärker restringierten Kontext noch wahr ist.

Eine negative Beantwortung impliziert, daß der Antwortende die Proposition der Frage für falsch hält. In einer binären Logik, die in einer geschlossenen Welt interpretiert wird, entspricht das der Behauptung, daß das Gegenteil wahr ist, also die Negation der Proposition. Die Proposition der Frage wird in einer solchen Logik zum Argument eines Negationsoperators. Wenn mehrere Möglichkeiten der Skopierung in der entsprechenden Proposition bestehen, führen Negationsoperatoren jedoch Ambiguitäten ein. In solchen Fällen sollte ein kooperativer Sprecher daher den Skopus der Negation explizit angeben, was durch eine Formulierung der negierten Proposition erfolgen kann. Beispiel 4.24 aus Dialog 2 macht das Problem klarer.

(4.24) **Me009:** Hast du Radkappen drauf?
**Ma012:** Nein . Die hab' ich schon abgemacht .

Hier führt bei der verneinenden Antwort vor allem der zeitliche Aspekt der Negation zur Ambiguität: Waren nie Radkappen befestigt oder sind sie nur *nicht mehr* befestigt? Der Anrufer reagiert entsprechend, indem er die letztgenannte Variante explizit formuliert.

Im Korpus traten 42 Antworten vom Typ `$answer--yes-no` auf Entscheidungsfragen auf (d.h. Fragen vom Dialogakttyp `$question--yes-no`). Diese verteilten sich wie in Tabelle 4.16 angegeben auf die beiden Möglichkeiten der Beantwortung und unterschiedliche Explizitheitsniveaus. Die relativen Maße beziehen sich auf die Gesamtzahl der Antworten vom Typ `$answer--yes-no` für die entsprechende Antwortpolarität unter der jeweiligen Zeitdruckbedingung, die auf Fragen vom Typ `$question--yes-no` gegeben wurden.[32]

Unter Zeitdruck wurden mehr positive Antworten knapp erteilt, ohne die Frageproposition zu wiederholen. Lediglich Spezialisierungen wurden gegeben, wenn dies nötig war. Die geringe Menge an vorliegenden Daten erlaubt jedoch keine weitgehenden Schlußfolgerungen. Unter der Voraussetzung, daß diese Hypothese experimentell zu bestätigen ist, wird für die Modellierung daher lediglich angenommen, daß eine explizite positive Antwort als Zeichen für fehlenden Zeitdruck zu interpretieren ist.

| Antwortform | gesamt | ohne ZD | | mit ZD | |
|---|---|---|---|---|---|
| | | abs. | rel. | abs. | rel. |
| 'Ja' | 10 | 3 | 27.3 | 7 | 43.8 |
| 'Ja' mit Wiederholung | 6 | 4 | 36.4 | 2 | 12.5 |
| 'Ja' mit Spezialisierung | 11 | 4 | 36.4 | 7 | 43.8 |
| 'Nein' | 6 | 3 | 50.0 | 3 | 33.3 |
| 'Nein' mit Spezialisierung | 9 | 3 | 50.0 | 6 | 66.6 |

Tabelle 4.16: Ausführlichkeit der Dialogakte `$answer--yes-no`

**Detailliertheit der Ereignisdarstellung und Detailliertheit von Referenten** Um die Detailliertheit der Ereignisdarstellung beurteilen zu können, ist es notwendig, über eine ausreichend formalisierte Darstellung des Ereignisses zu verfügen, das beschrieben wird. Da zur Aufgabe

[32]Dabei wurden auch verspätete, d.h. erst nach dazwischenliegenden Dialogturns abgegebene Antworten ausgewertet. Entsprechend Levelt und Kelter (1982) können gerade solche Dialogturns auch eine verstärkt ausführliche Beantwortung nach sich ziehen. Es traten jedoch nur 5 solcher Äußerungen auf, daher wird diese Einschränkung nicht berücksichtigt.

eines Pannenhilfesystems wie READY jedoch gerade gehört, ein solches Ereignis zu erkennen, ist eine Beurteilung der Ereignisdarstellung des Benutzers während des Dialoges nicht möglich. READY verfügt erst nach dem Dialog über eine ausreichend gute Kenntnis des Pannenereignisses, d.h. die Daten könnten lediglich für eine nachträgliche erneute Auswertung verwendet werden. Maschinelles Lernen ist jedoch in der aktuellen Forschungsphase im READY-Prototypen nicht vorgesehen. Außerdem ist die Beurteilung der Detailliertheit von Referenten stark diskursgeschichtsabhängig. Die Implementation einer Diskursgeschichte ist jedoch ebenfalls nicht Gegenstand der aktuellen Forschungsphase im Projekt READY, da das Dialogmanagment hier über einen weniger differenzierten Mechanismus ausgeführt wird. Auf die Auswertung des Parameters wurde daher unter den gegebenen Umständen verzichtet.

#### 4.3.6.3 Diskursobligationen

Tabelle 4.17 zeigt die Abfolge der Dialogakte in den Dialogen des Korpus. Den auftretenden Dialogakten[33] in den Zeilen der Tabelle ist für jeden der anderen Dialogakte eine Zahl zugeordnet, die angibt, wie oft die entsprechenden Dialogakte direkt aufeinander folgten. Mehrfachnennungen sind dabei enthalten, da einzelne Äußerungen mit mehreren Dialogakten annotiert wurden. Grund dafür war entweder, daß die Äußerung mehrere Sätze enthielt, denen jeweils einzeln Dialogakte zugeordnet wurden oder daß mit einer Äußerung mehrere Dialogakte gleichzeitig realisiert wurden.[34] Dieses Vorgehen ist motiviert dadurch, daß in READY die Modellierung der Benutzereigenschaften diskretisiert wird auf sogenannte Zeitscheiben. Eine Zeitscheibe entspricht dabei einem Dialogturn. Verbunden damit ist die Konsequenz, daß innerhalb einer Zeitscheibe keine Differenzierung irgendwelcher Phänomene erfolgen kann. Tabelle 4.18 zeigt die aus Tabelle 3.21 abgeleiteten Diskursobligationen für READY und die Häufigkeiten, mit denen die entsprechenden Obligationen im Dialog direkt, d.h. in der nachfolgenden Äußerung erfüllt wurden.

<table>
<tr><th colspan="2" rowspan="2">Dialogakt/Folgedialogakt</th><th colspan="3">$question</th><th colspan="3">$answer</th><th colspan="2">$instr.</th><th colspan="2">$succ.</th><th colspan="2">$backch.</th><th colspan="2">$descr.</th></tr>
<tr><th>YN</th><th>I</th><th>C</th><th>YN</th><th>I</th><th>C</th><th>D</th><th>I</th><th>S</th><th>NS</th><th>W</th><th>-</th><th>S</th><th>L</th></tr>
<tr><td>$question–yes-no</td><td>(YN)</td><td colspan="3">7</td><td>42</td><td></td><td></td><td>2</td><td>4</td><td>6</td><td>1</td><td>7</td><td>3</td><td>18</td><td>3</td></tr>
<tr><td>$question–information</td><td>(I)</td><td colspan="3">4</td><td></td><td>11</td><td></td><td>3</td><td>4</td><td></td><td></td><td></td><td>3</td><td>3</td><td>7</td></tr>
<tr><td>$question–clarify</td><td>(C)</td><td colspan="3">6</td><td></td><td></td><td>45</td><td>3</td><td>1</td><td>8</td><td></td><td>3</td><td></td><td>14</td><td></td></tr>
<tr><td>$answer–yes-no</td><td>(YN)</td><td colspan="3">14</td><td colspan="3"></td><td>11</td><td>5</td><td>3</td><td>1</td><td></td><td>13</td><td>6</td><td>5</td></tr>
<tr><td>$answer–information</td><td>(I)</td><td colspan="3"></td><td colspan="3"></td><td></td><td></td><td>3</td><td></td><td></td><td>6</td><td>4</td><td>1</td></tr>
<tr><td>$answer–clarify</td><td>(C)</td><td colspan="3">14</td><td colspan="3">5</td><td>7</td><td>3</td><td>3</td><td>1</td><td></td><td>16</td><td>6</td><td>4</td></tr>
<tr><td>$instruction–direct</td><td>(D)</td><td colspan="3">13</td><td colspan="3"></td><td></td><td></td><td>18</td><td>1</td><td>6</td><td>13</td><td>14</td><td></td></tr>
<tr><td>$instruction–indirect</td><td>(I)</td><td colspan="3">7</td><td colspan="3">3</td><td>1</td><td></td><td>15</td><td>3</td><td>7</td><td>14</td><td>10</td><td></td></tr>
<tr><td>$success–success</td><td>(S)</td><td>10</td><td>1</td><td>12</td><td colspan="3">4</td><td>12</td><td>8</td><td></td><td></td><td></td><td>11</td><td>6</td><td>9</td></tr>
<tr><td>$success–nosuccess</td><td>(NS)</td><td>1</td><td></td><td></td><td colspan="3">1</td><td></td><td>4</td><td></td><td></td><td></td><td>3</td><td></td><td>5</td></tr>
<tr><td>$backchanneling–wait</td><td>(W)</td><td colspan="3">6</td><td colspan="3">1</td><td>1</td><td>1</td><td>5</td><td>1</td><td></td><td>5</td><td>4</td><td>2</td></tr>
<tr><td>$backchanneling–</td><td>(–)</td><td>15</td><td>9</td><td>8</td><td colspan="3">6</td><td>12</td><td>6</td><td>11</td><td>1</td><td>5</td><td>30</td><td>50</td><td>10</td></tr>
<tr><td>$description–situation</td><td>(S)</td><td>22</td><td>4</td><td>23</td><td>5</td><td>2</td><td>1</td><td>17</td><td>17</td><td>4</td><td></td><td></td><td>53</td><td>6</td><td>17</td></tr>
<tr><td>$description–localisation</td><td>(L)</td><td colspan="3">9</td><td colspan="3">2</td><td>1</td><td></td><td>9</td><td>2</td><td>4</td><td>20</td><td>7</td><td>2</td></tr>
</table>

Tabelle 4.17: Verteilung und Abfolge der wichtigsten Dialogakte

[33] Fettgedruckte Dialogakte sind verstärkt auf Mechanikerseite eingesetzt worden, alle anderen eher auf Anruferseite.

[34] Z.B. konnten einige Äußerungen vom Typ `$question--yes-no` meist auch gleichzeitig als `$instruction--indirect` gelabelt werden.

| Obligationsquelle | Obligation |
|---|---|
| Frage | Antwort |
| $question–yes-no | $answer–yes-no |
| $question–information | $answer–information |
| $question–clarify | $answer–clarify |
| Instruktion | Erfolgsmeldung |

| Dialogakt | Anzahl | Obligation erfüllt | |
|---|---|---|---|
| | | Antwort/Success | rel. |
| Mechaniker | | | |
| Question | 146 | 96 | 65.8 |
| $question–yes-no | 61 | 40 | 65.6 |
| $question–information | 25 | 11 | 44.0 |
| $question–clarify | 60 | 45 | 75.0 |
| Instruction | 92 | 37 | 40.2 |
| $instruction–direct | 54 | 19 | 35.2 |
| $instruction–indirect | 38 | 18 | 47.4 |

Tabelle 4.18: Diskursobligationen in READY und zugehörige Daten aus dem Korpus

Verletzungen von Diskursobligationen sind eher ein direktes Symptom für kognitive Belastung: Wenn der Sprecher weiß, daß er die entsprechende Obligation erfüllen sollte, das auch mit seinen eigenen Zielen vereinbaren kann, dann deutet eine Verletzung der Obligation darauf hin, daß er ihr nicht nachkommen *konnte*. Grund dafür können auch kognitive Verarbeitungsprobleme sein.

Spezielles Augenmerk wurde auf Dialogakte geworfen, die eine explizite Fortsetzung vom Mechaniker forderten und damit als Signal der Ungeduld interpretiert werden können. Dazu zählen z.B. Äußerungen wie 'Und jetzt?' oder 'Wie geht's weiter?'. Unter Zeitdruck wären ggf. mehr solche Signale zu erwarten. Diese Hypothese hat sich jedoch nicht bestätigt: von den 8 Äußerungen, die eine solche Aufforderung zur Fortsetzung enthielten, traten 5 in den Zeitdruck-Dialogen und 3 in den Dialogen ohne Zeitdruck auf. Das entspricht für beide Bedingungen in etwa einer ähnlichen Häufigkeit des Auftretens. Anzumerken ist dabei jedoch, daß die Äußerungen aus lediglich 4 Dialogen stammen, d.h. nur von der Hälfte der Dialogteilnehmer auf der Seite des Anrufers überhaupt eingesetzt wurden. Individuelle Neigungen zum Einsatz eines solchen Signals treten hier also möglicherweise vor den Einfluß der entsprechenden Dialogsituation.

### 4.3.7 Seltene Phänomene

Im folgenden sind alle Parameter aufgeführt, die sich zwar aus der psychologischen Literatur ergaben, die jedoch im untersuchten Korpus nur außerordentlich selten oder gar nicht auftraten. Bei einigen dieser Parameter ist dieser Effekt mit Sicherheit auf den Typ der Dialogsituation zurückzuführen, bei anderen kommt als Erklärung auch in Frage, daß sie insgesamt nur mit geringer Häufigkeit zu erwarten sind. Die Symptome dieses Abschnitts werden bei der Modellierung in READY mit Bayes'schen Netzen entweder entfallen müssen oder sie müssen zu Gesamtindices zusammengezogen werden.

#### 4.3.7.1 Kognitive Rhythmen

Längere Äußerungssequenzen ein und desselben Sprechers folgen in den Dialogen der READY-Domäne nur selten aufeinander. Die maximale Äußerungsdauer[35] aus allen Dialogen betrug 20.323 s (Me157, Dialog 8). Wie in Abschnitt 3.2.5 bereits erwähnt wurde, ist eine bestimmte Mindestdauer von Äußerungen für die Beurteilung kognitiver Rhythmen erforderlich. Ein Eindruck davon, ob kognitive Rhythmen in der vorliegenden Domäne existieren, war daher nicht zu erhalten. Diese Einsicht kommt zur bereits erwähnten Problematik des Konzepts hinzu. Kognitive Rhythmen werden daher für die Modellierung im System READY nicht herangezogen.

#### 4.3.7.2 Phonologische Sprechfehler

Ein einziger *möglicherweise* phonologischer Sprechfehler konnte im gesamten Korpus gefunden werden. Seine Zuordnung zur Klasse der lexikalischen Fehler ist jedoch ebenso wahrscheinlich. Schwierig wird die Entscheidung dadurch, daß die Gründe, die der Sprecher für die Korrektur hatte, nicht nachvollziehbar sind. Der Fehler wurde vom Sprecher korrigiert.

(4.25) **Me048, Dialog 8:**
<äh/.535/> <P/2.037/> <+/> Der Hebel </+> , <also/.173/> die Motorhaube ist <+/> <+/> *<@M1> noch <@M2> nicht </+> <Z> <@M2> <@M3> ni- <@M1> <@M4> no- </+> <Z> <@M1> <@M4> noch <@M2> <@M3> nicht* geöffnet?

Phonologische Sprechfehler sollten als Indikator für Zeitdruck oder kognitive Belastung in einem Dialogsystem auch aufgrund ihrer Überprüfung an realen Daten keine große Bedeutung haben. Die Artikulation korrekter phonologischer Strukturen können Sprecher ganz offensichtlich weitgehend aufrechterhalten. Diese Schlußfolgerung bedeutet nicht, daß phonologische Sprechfehler keine Indikatoren für außerordentlich große kognitive Belastung sein können. Eine Antwort darauf kann jedoch nur weitere empirische Forschungsarbeit geben. Phonologische Sprechfehler werden daher in READY mit anderen sporadischen Phänomenen zusammengefaßt.

#### 4.3.7.3 Flexionsmorphologische Sprechfehler

Auch flexionsmorphologische Sprechfehler traten im Korpus nicht in einer Form auf, die eine hundertprozentige Zuordnung nur zu dieser Klasse erlaubt. Die einzige Äußerung mit einem näherungsweise als flexionsmorphologisch zu kennzeichnenden Fehler ist in Beispiel 4.26 angegeben. Fraglich ist jedoch, ob die Wahl eines falschen Tempus Ursachen bei der Flexion oder schon in der konzeptuellen Auswahl der Perspektive auf das Ereignis hat. Flexionsmorphologische Sprechfehler werden in READY behandelt wie Phonologische Sprechfehler.

(4.26) **Ma001, Dialog 2:**
<mhm/.338/> An <D> der Saarbrücker Straße <P/.485/> , da ist mir beim Fahren aufgefallen , <P/.472/> daß <D> am linken Vorderreifen <P/.387/> , wenn ich fahr’

[35]Unter Ausschluß einer Äußerung mit einer 10 s - Nebentätigkeitspause.

<P/.172/><+/> <@*R1*> *fing* <@*M1*> *'s* <@*D1*> *plötzlich* </+> <@*R1*> *hat* <@*M1*> *'s* <Z> *angefangen* zu klackern .

#### 4.3.7.4 Lexikalische Sprechfehler

Im Korpus konnten keine unkorrigierten lexikalischen Sprechfehler gefunden werden. Wie bereits in Kapitel 3 angesprochen, sind lexikalische Sprechfehler ohne nachfolgende Korrektur auch prinzipiell nur dann erkennbar, wenn der Wahrheitsgehalt der zugehörigen Äußerung überprüft werden kann. Die sieben korrigierten lexikalischen Sprechfehler sind in der Aufstellung in Abschnitt 4.3.4.2 eingegangen. Obwohl die Anzahl lexikalischer Sprechfehler stark über der Anzahl flexionsmorphologischer und phonologischer Fehler liegt, muß sie doch als zu geringfügig betrachtet werden, um für die Beurteilung der kognitiven Belastung relevant zu werden. Die Behandlung in READY wird daher gemeinsam mit den anderen Fehlerkategorien erfolgen.

#### 4.3.7.5 Deixis

Es zeigte sich, daß ortsdeiktische Ausdrücke in den Dialogen des Korpus durchaus auftraten. Allerdings waren davon nur wenige für den Mechaniker im Kontext nicht interpretierbar. Meist konnten die deiktischen Äußerungen unter Zuhilfenahme der Diskursgeschichte aufgelöst werden. Wie bereits erwähnt, wurde die Auswertung der Diskursgeschichte in dieser Diplomarbeit zurückgestellt. Ortsdeiktische Äußerungen werden daher zwar als Symptom kognitiver Belastung Aufnahme im System READY finden. Die vereinfachte Modellierung ist jedoch als vorläufig und verbesserungsbedürftig anzusehen.

#### 4.3.7.6 Füllphrasen und Füllworte

Zu den Füllworten gehört u.a. das saarländische 'ai', welches vom Mechaniker sehr häufig eingesetzt wurde. Es ist jedoch nach Auskunft von Dialektsprechern nicht als gefüllte Pause klassifizierbar, da seine Verwendung syntaktischen Beschränkungen wie etwa das englische 'well' unterliegt.[36] Einige Interjektionen wie 'achso', 'also', 'tja' wurden im Korpus ebenfalls bereits als gefüllte Pausen gewertet, wenn sie satzeinleitend verwendet wurden. Sie finden sich daher im Pausenindex wieder.

Füllphrasen kamen im Korpus nur selten zur Anwendung. Lediglich sechsmal verzögerte der Mechaniker die Sprachproduktion mit Phrasen wie 'glaube ich', 'ja', 'ich weiß nicht', 'sagen wir mal'. Zwei der Parenthesen lagen syntaktisch zwischen einer Präposition und ihrem NP-Argument, zwei zwischen Konjunktionen und ihrem S-Komplement und jeweils eine zwischen Verb und Akkusativobjekt sowie zwischen Auxiliar und VP-Komplement. Das bedeutet, daß diese Konstruktionen im Korpus *ausschließlich für eine intraphrasale Verzögerung* verwendet wurden. Leider ist auch hier die Datenmenge zu gering, um weitere Schlußfolgerungen zu erlauben.

[36] Abweichend davon ist die vom Mechaniker ebenfalls häufig eingesetzte Zusammenziehung 'aiäh' als gefüllte Pause gewertet worden.

#### 4.3.7.7 Redundanz

Redundante Äußerungen zeichneten sich in den Dialogen des Korpus vor allem durch die Wiederholung von zuvor bereits geäußerten Phrasen und weniger durch deren Paraphrasierung aus. Insgesamt wurden jedoch lediglich 11 Äußerungen mit redundanten Bestandteilen festgestellt, von denen 3 aus den Dialogen ohne Zeitdruck und 8 aus den Dialogen mit Zeitdruck stammen (6 vom Anrufer, 5 vom Mechaniker). Diese spärliche Datenlage läßt kaum Schlüsse über den Zusammenhang zum Zeitdruck oder zu kognitiver Belastung zu. Die Modellierung sollte daher so erfolgen, wie die (ebenfalls wenig ausführliche) experimentelle Literatur es empfiehlt: Redundante Äußerungen werden mit einem binären Tag versehen. Auftreten von Redundanz kann dann als schwacher Indikator von kognitiver Belastung und ebenso als schwacher Indikator für Zeitdruck verwendet werden.

# Kapitel 5

# Erkennung von Symptomen in Ausgaberepräsentationen computerlinguistischer Module

## 5.1 Fragestellung

Aus der empirischen Forschung und der Analyse des Korpus haben sich zahlreiche sprachliche Symptome für Belastung und Zeitdruck ergeben. Um die Werte dieser Symptome für die Verarbeitung in einem Dialogsystem überhaupt zugänglich zu machen, müssen sie aus den verschiedenen Ausgaberepräsentationen der linguistischen Verarbeitungsebenen extrahiert werden können. Welche Form die jeweilige konkrete Repräsentation hat, ist von den einzelnen Verarbeitungskomponenten abhängig.

Die wenigsten der spontansprachlichen Symptome, die die Einschätzung von Belastung und Zeitdruck unterstützen, stehen während oder nach der linguistischen Verarbeitung für eine Benutzermodellierungskomponente direkt zur Verfügung. Einige der jeweiligen Sprechparameter sind lediglich in Zwischenstufen auf dem Weg zur semantischen Repräsentation notwendig und in dieser Abschlußrepräsentation irrelevant (z.B. die gesamte zeitliche Information). Andere Phänomene werden nur analysiert, weil sie Störgrößen in der Eingabe darstellen und für eine vollständige Analyse korrigiert oder eliminiert werden müssen (z.B. der Abschnitt der Originaläußerung bei Selbstkorrekturen oder Wiederholungen). Inhaltliche Aspekte einer Äußerung können erst durch einen zusätzlichen Inferenzschritt über den Ausgabestrukturen gewonnen werden (z.B. die Information über redundante oder zusätzliche Information).

Der erste Teil dieses Kapitels untersucht, inwieweit mit verbreiteten Verarbeitungsmechanismen bereits für die Erkennung von Verarbeitungsbeschränkungen verwendbare Repräsentationen erzeugt werden. Daraus ergeben sich vor allem Vorschläge, wie Verarbeitungskomponenten modifiziert werden müßten, um einige der gewünschten Informationen zugänglich zu machen.

Im zweiten Teil werden die Analyseschritte im System READY vorgestellt. Wegen der spezifischen Zielsetzung in diesem Projekt war es nicht beabsichtigt (und aufgrund des Forschungsstadiums der meisten dieser Komponenten auch nicht möglich), alle Schritte der Eingabeanalyse in READY durch den Anschluß echter Verarbeitungsmodule zu realisieren. In Abschnitt 4.2 wurde daher eine Annotationsform vorgestellt, die die Ergebnisse verschiedener Analyseschritte für die Eingabe im System READY zusammenfassen kann. Um Möglichkeiten der Interaktion zu demonstrieren und Lücken in der Annotation zu füllen, wird am Beispiel zweier echter computerlinguistischer Teilkomponenten gezeigt, wie ein beispielhafter Zusammenschluß im System READY realisiert wurde.

## 5.2 Ausgaberepräsentationen bestehender Verarbeitungsmechanismen

Störende Information und Zwischendarstellungen werden während der linguistischen Analyse einer sprachlichen Eingabe meist nicht an nachfolgende Analysekomponenten weitergegeben. Ein Teil der Symptome beruht jedoch gerade auf diesen beiden Sorten von Information. Um diese Information der Benutzermodellierungskomponente zugänglich zu machen, sollte sie nicht eliminiert, sondern lediglich in eine Darstellung überführt werden, die von nachfolgenden Verarbeitungskomponenten ignoriert werden kann. Dadurch kann die für die Benutzermodellierung relevante Information auch in einer einheitlichen Endrepräsentation kodiert werden, die Grundlage für die Erhebung der Symptome ist.

Im Hinblick auf die Verarbeitungsgeschwindigkeit und die Anpassung von Systemen auch an eigene Ressourcenbeschränkungen kann aber auch ein **Anytime-Verhalten** angestrebt werden: Die Benutzermodellierung kann damit bereits so früh wie möglich noch unvollständige oder präzisierbare Information in ihre Analyse integrieren und erste, noch ungenaue Inferenzen leisten. Durch diese Verwendung partieller Information kann der Rechenaufwand für nachfolgende Anpassungsschritte zumindest teilweise bereits nebenläufig zur linguistischen Analyse verteilt werden. Abbildung 5.1 zeigt, nach welchen Zwischenschritten der Analyse genügend Information zur Anytime-Verwendung spezieller Symptome zur Verfügung stehen sollte. Die Diskussion der folgenden Abschnitte stützt sich auf die Annahme der ebenfalls in 5.1 dargestellten verbreiteten Ein- und Ausgabestrukturen.

### 5.2.1 Prosodische Merkmale

Ziel der Spracherkennung ist es, ein akustisches Sprachsignal auf einen String abzubilden, der das Signal möglichst vollständig abdeckt und der Sequenz der tatsächlich geäußerten Worte entspricht. Die dafür eingesetzten statistischen Verfahren generieren auf der Basis beobachteter Regelmäßigkeiten Hypothesen über die tatsächliche Äußerung. Da einige Zweifelsfälle[1] nicht während der Spracherkennung, sondern erst mithilfe syntaktischer und semantischer oder pragmatischer Information aufgelöst werden können, ist die Ausgabe der Spracherkennung ein Wort*hypothesen*graph (WHG), der verschiedene Möglichkeiten der Analyse für das Eingabesignal simultan repräsentiert.

Um akustische Segmente Phonemen zuordnen zu können und eine Strukturierung des Eingabestromes zu erreichen, muß der Spracherkenner sich auf zeitliche Information stützen. Das

[1] z.B. aufgrund sprecherspezifischer Besonderheiten oder Störungen des Signals

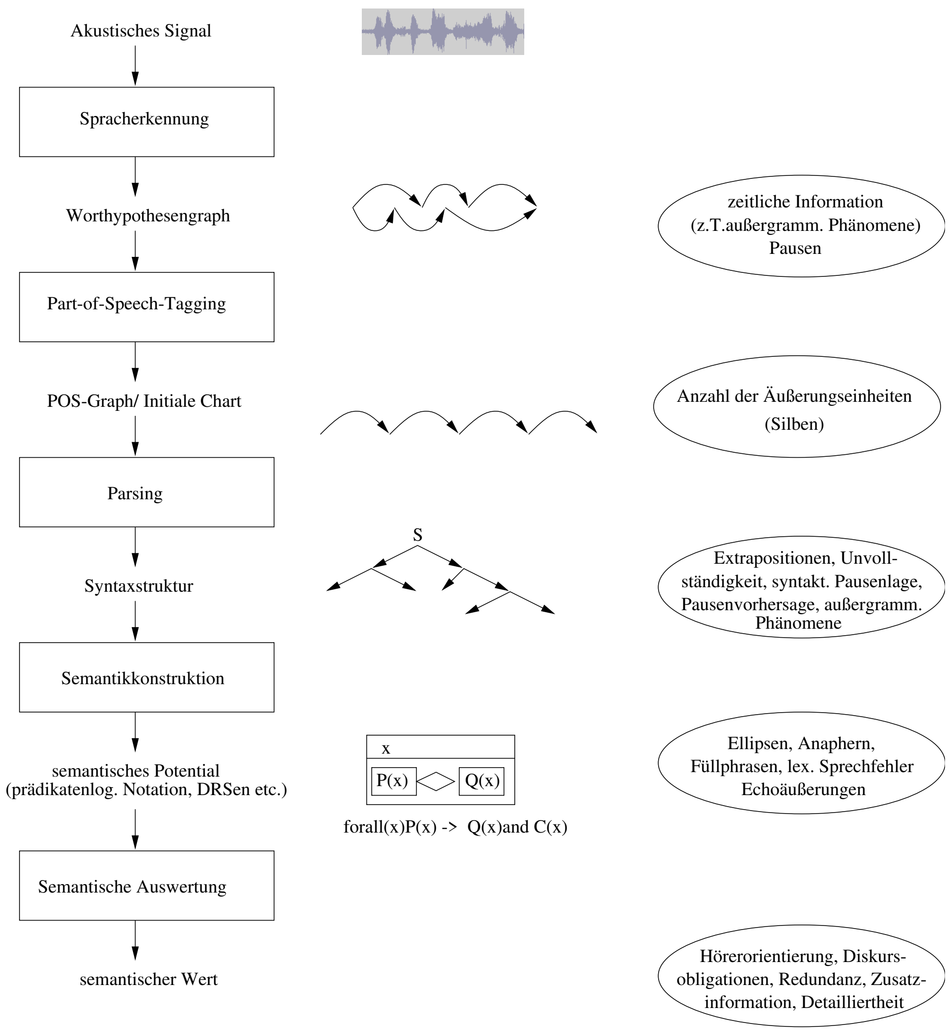


Abbildung 5.1: Gebräuchliche Eingabe- und Ausgaberepräsentationen linguistischer Verarbeitung und die theoretisch daraus entnehmbaren Symptome

Signal wird in zeitliche Fenster unterteilt, die 5 - 10 ms lang sind. Die Fenstergrenzen können als Knoten in einem Graphen dargestellt werden; das Fenster selbst durch eine Kante zwischen seinen beiden Begrenzungsknoten. Die Markierung der Kante beschreibt das im Fenster vorliegende Signal. Durch die Zusammenfassung nebeneinanderliegender Kanten (d.h. nebeneinanderliegender Zeitfenster) und Spektralanalysen gelangen Spracherkenner zu Phonhypothesen als Grundlage für die Worterkennung.[2]

Zeitliche Informationen, die für die Erkennung von Verarbeitungsbeschränkungen notwendig sind, können daher bereits *vor dem Parsing* auf dem WHG erhoben werden. Wenn der WHG in eine Chart für das Parsing oder die semantische Verarbeitung überführt wird (siehe z.B. Ruland et al., 1998; Worm, 1998), sollte die zeitliche Information für eine verfeinerte Analyse weiterhin zur Verfügung stehen: Besonders für Pausen wurde z.B. bereits darauf hingewiesen, daß ihre syntaktische Lage eine wichtige Rolle im Hinblick auf den Zusammenhang mit kognitiver Belastung darstellt.

### Dauer von Artikulationssequenzen

Die Zeitfenster des Spracherkenners besitzen eine feste Länge. Um die zeitliche Information über verschiedene Optimierungs- und Verarbeitungsschritte hinweg zur Verfügung zu halten, müssen Zusammenfassungen von Kanten gekennzeichnet werden. Jede Kante im WHG, die eine Projektion einzelner Zeitfenster darstellt, kann daher mit der Anzahl der durch die Projektion überdeckten Kanten markiert werden (Abb. 5.2). Die zeitliche Dauer der hypothetischen Worte, die diesen Projektionskanten zugeordnet werden, ist aufgrund der festen Länge der Zeitfenster durch Multiplikation der Dauer eines Zeitsegments mit der Anzahl der von der Projektion dominierten Kanten errechenbar.

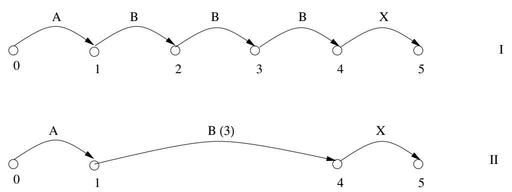


Abbildung 5.2: Verfügbarkeit zeitlicher Information während der Optimierung des WHG

### Stille Pausen

Bei diesem Vorgehen werden auch Abschnitte der Signalabwesenheit vorgefunden.[3] Sie sind im Worthypothesengraphen als $\varepsilon$-Kanten vertreten und werden vor dem Beginn des Parsings zwecks Optimierung eliminiert (Nederhof, Bouma, Koeling & van Noord, 1997). Die Zusammenfassung aufeinanderfolgender $\varepsilon$-Kanten liefert zusammen mit deren Gesamtdauer die not-

[2]Das gesamte Vorgehen ist von den zur Spracherkennung selbst eingesetzten Mechanismen wie Hidden-Markov-Modellen oder neuronalen Netzen unabhängig, da es sich hierbei um grundlegende Fragen der Spracherkennung handelt.

[3]Der Spracherkenner kann diese Abschnitte für die Segmentierung des Lautstromes nutzen.

wendige Information über die Dauer einer stillen Pause im Sprachsignal.[4] Kanten im WHG können statt einer Verschmelzung ihres Anfangs- mit ihrem Endknoten und Entfernung der dazwischenliegenden Kanten auch 'schonend' optimiert werden: $\varepsilon$-Kanten bleiben im Graphen einfach bestehen. Da nach wie vor ihr Startknoten mit ihrem Endknoten gleichgesetzt wird, können sie prinzipiell durch weitere Verarbeitungsschritte aufgrund dieser Eigenschaft 'Startknoten = Endknoten' ignoriert werden[5]. Diese Kanten werden im folgenden *Zirkel-Kanten* genannt.(Abb. 5.3).

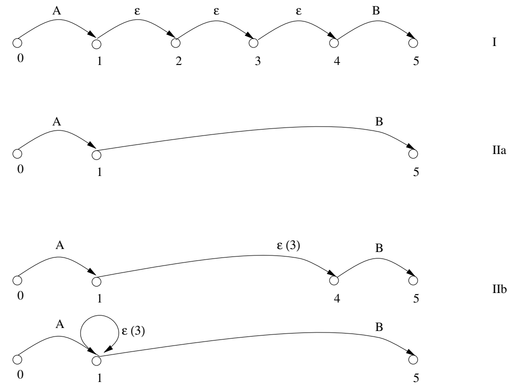


Abbildung 5.3: Optimierung eines Worthypothesengraphen (I) in der aktuellen Form (IIa: Eliminieren) vs. erforderliche Modifikation (IIb: Zusammenfassen und Kollabieren)

**Gefüllte Pausen**

Gefüllte Pausen sind vom Spracherkenner nicht zu identifizieren, obwohl sie im Sprachmodell als potentielle Worte enthalten sein müssen: Die Information, ob es sich bei erkannten Worten um nicht*lexikalische* sprachliche Elemente handelt, steht frühestens nach weiteren Vorverarbeitungsstufen des Parsings (wie Part-of-Speech[6]-Tagging) zur Verfügung. Batliner, Kießling, Burger und Nöth (1995) zeigen, daß gefüllte Pausen im Sprachsignal anhand akustischer Merkmale identifizierbar sind, jedoch kann erst mit POS-Tags und dem dadurch möglichen Zugriff auf ein Lexikon zuverlässig entschieden werden, ob eine gefüllte Pause vorliegt. Gefüllte Pausen können daher möglicherweise erst aufgrund lexikalischer Information nach dem POS-Tagging eliminiert werden. Eine schonende Elimination wie für stille Pausen ist dabei auch hier zu empfehlen.

**Artikulationsgeschwindigkeit**

Die Artikulationsgeschwindigkeit kann auf dem WHG erhoben werden, allerdings stehen dann möglicherweise wegen mehrerer möglicher überspannender Analysen noch deren verschiedene

[4] Siehe Definition von stillen Sprechpausen 3.3.

[5] In einigen Fällen (wie z.B. beim Chart-Parsing) müssen die Kanten ggf. gesondert markiert werden, um sie von gleichartigen Kanten in den eigenen Strukturen zu unterscheiden.

[6] Im folgenden wie allgemein üblich mit POS abgekürzt.

Werte in Konkurrenz zueinander. Die Anzahl der Vokale im Eingabesignal[7] kann ein von den Worthypothesen unabhängiges, approximatives Maß für die Silbenzahl liefern. Exakt kann die Silbenzahl einer Äußerung nach einem POS-Tagging und zugehörigem Lexikonzugriff errechnet werden, wenn im Lexikon Angaben über Silbengrenzen enthalten sind. Das Problem multipler Analysen bleibt dabei jedoch solange bestehen, bis sich eine Verarbeitungskomponente z.B. aufgrund von Information aus syntaktischer und semantischer Verarbeitung für eine der Analysen entschieden hat. Bei in frühen Stufen noch nicht verfügbarem Lexikon sind alternativ als approximative Vorstufe auch Silbentrennalgorithmen einsetzbar, wie sie in der Dokumentenverarbeitung gebräuchlich sind. Die Artikulationsdauer der zugehörigen Analyse des WHG wird durch die Addition der zeitlichen Projektionsannotationen aus dem WHG errechnet, ohne zirkuläre Kanten von Pausen einzubeziehen und entspricht damit Formel 3.1.

#### Einsatzlatenz

Die Einsatzlatenz eines Sprechers entspricht einer initialen stillen Pause. Ihre Dauer kann entsprechend berechnet werden. Wichtig ist dabei, daß mit der Aufzeichnung des Sprachsignals für den Sprecher *unmittelbar nach dem Ende des Sprachsignals seines Dialogpartners*[8] begonnen wird.

### 5.2.2 Syntaktische Merkmale

Obwohl Parsingverfahren meistens ausgehend von einem Eingabe*string* beschrieben werden (z.B. Naumann & Langer, 1994), steht in realen Systemen mit Eingabe gesprochener Sprache selten *ein eindeutiger* Eingabestring zur Verfügung: Vielmehr werden einem oder mehreren Parsern verschiedene, das Eingabesignal vollständig überspannende Analysepfade des WHG übergeben (Nederhof et al., 1997; Ruland et al., 1998) und die beste Analyse durch ein Ranking der Ergebnisse ermittelt.

#### Außergrammatische Phänomene

Für die Erkennung von Verarbeitungsbeschränkungen müssen vor allem außergrammatische Phänomene wie Wiederholungen, Selbstkorrekturen, Fehlansätze und unvollständige Strukturen identifiziert werden. Die Erkennung kann an drei Punkten der Verarbeitung erfolgen: *vor*, *während* und *nach* dem Parsing. Unterschiedliche Verfahren decken dabei jedoch unterschiedliche Bereiche der spontansprachlichen außergrammatischen Phänomene ab (s.u.). Endgültige Schlußfolgerungen über die Anzahl und Form außergrammatischer Phänomene steht daher meist erst nach mehreren Parsingversuchen und den zugehörigen Korrekturen der Eingabe bereit. Im Sinne des beschriebenen Anytime-Verfahrens können aus partieller Information nach der Spracherkennung und während des Parsings bereits Hinweise auf relevante Symptome abgeleitet werden.

[7] Dieses Vorgehen unter Berücksichtigung von Diphthongen und Vokalverdoppelungen wird im Projekt READY angewendet, da kein Lexikon zur Verfügung steht.

[8] Im Falle von READY handelt es sich dabei um das System selbst.

**Vor dem Parsing** Einige Verfahren erkennen außergrammatische Phänomene bereits auf dem WHG bzw. auf einem POS-Graphen (Ruland et al., 1998). Möglich ist das, da Wortgrenzen nach einer abgebrochenen Originaläußerung mit Reparandum prosodisch markiert sind und mit statistischen Verfahren beurteilt werden können. Gleiches gilt für Unterbrechungsstellen und Korrekturterme. Ruland et al. (1998) arbeiten dabei 'schonend' in der Form, wie sie oben beschrieben wurde: Es wird nie Information aus dem Graphen gelöscht, sondern lediglich neue hinzugefügt. Mehrere POS-Kanten, die die Originaläußerung darstellen, werden durch Kanten ergänzt, die von der Wiedereinsatzstelle in der Originaläußerung ausgehend die gesamte Korrektur oder Wiederholung überspannen und lediglich mit der Korrekturfortsetzung[9] markiert sind. Für die Beurteilung des Typs des außergrammatischen Phänomens (Wiederholung, Selbstkorrektur oder Fehlansatz) ist jedoch bei Ruland et al. (1998) eine partielle syntaktische Analyse der kollabierten Strukturen erforderlich.

Heeman und Allen (1994) setzen ein kombiniertes Verfahren aus Pattern Matching und statistischer Klassifikation ein, um außergrammatische Phänomene zu erkennen. Der zugehörige Algorithmus stützt sich auf sogenannte *detection cues* und arbeitet auf Sequenzen von POS-Tags. Anhand der Form eines gefundenen Patterns ist die Form des Phänomens erkennbar. Meist werden durch reine Mustererkennung noch zu viele 'falsche' Korrekturen geliefert, d.h. Abschnitte der Äußerung, die in Wirklichkeit wohlgeformt sind. Diese werden von der statistischen Bewertungskomponente aussortiert. Die Verwendung unbewerteter Ergebnisse des Pattern-Matchers ist für ein Anytime-Verfahren nicht zu empfehlen, da Ungenauigkeiten der Erkennung die approximative Genauigkeit der Benutzermodellierung weiter verschlechtern würden. Ein demjenigen von Heeman und Allen (1994) ähnelndes Verfahren beschreiben Bear et al. (1992).

Nakatani und Hirschberg (1993) setzen ein rein statistisches Verfahren auf Wortpaaren, Markierungen von Pausen und anderen Korrektureinleitern ein. Der Ansatz zeigt deutlich, daß Information aus der Signalanalyse auch nach der Spracherkennung noch zur Verfügung stehen kann. Die Erkennung des Phänomentyps wird ebenfalls durch das statistische Erkennungsverfahren vorgenommen.

Heeman und Allen (1997) integrieren die verschiedenen Phasen von der Signalerkennung bis zur Erkennung von speziellen Diskurselementen in einem einzigen Sprachmodell. Bereits während der Spracherkennung sollen dabei auch außergrammatische Erscheinungen erkannt werden. Ausgabe des Verfahrens ist ein POS-Graph, der bereits mit den erkannten Phänomenen annotiert ist. Problematisch an diesem Verfahren ist jedoch für die Zwecke der Benutzermodellierung, daß nach der Spracherkennung kein WHG zur Verfügung zu stehen scheint und damit die gesamte zeitliche Information, vor allem über stille Pausen, verlorengeht.

**Während des Parsings** Hindle (1983) erkennt Korrekturen und Wiederholungen durch eine verschränkte Verarbeitung während eines (deterministischen) Parsings. Neben der Idealisierung des Determinismus nimmt Hindle zusätzlich an, daß immer ein Korrektursignal zur Verfügung steht, das als Indikator für ein außergrammatisches Phänomen dient. Diese Forderung schränkt die Verwendbarkeit des Verfahrens stark ein. Der Ansatz selbst ist jedoch interessant: In einem Fenster festgelegter Länge werden von links nach rechts Strings miteinander verglichen und anhand der Gleichheit von Worten und Wortfolgen (Operation *Surface Copy:* Wiederholung),

[9] Zur Terminologie siehe 3.3.4.

Gleichheit lexikalischer Kategorien (*Category Copy:* Selbstkorrektur) und Unvollständigkeit von Konstituenten (*Stack Copy:* Fehlansatz oder Selbstkorrektur) vor und hinter einem Korrektursignal entschieden, ob der entsprechende Abschnitt des Signals ein außergrammatisches Phänomen umfaßt. Am Ende des Verfahrens steht nur noch eine Analyse des korrekten Strings zur Verfügung. Ein Protokoll der Löschoperationen könnte die für die Benutzermodellierung gewünschte Information bereitstellen.

Die Möglichkeit, außergrammatische Phänomene während des Parsings zu erkennen, ist wegen der Doppelaufgabe für den Parser jedoch problematisch: Der Parser muß, während er eine Analyse erstellt, diese Analyse auch bewerten (Worm, 1998). Verfahren vor oder nach dem Parsing wird daher im allgemeinen der Vorzug gegeben.

**Nach dem Parsing** Worm (1998) eliminiert Originaläußerungen einer Korrektur bzw. einer Wiederholung erst während der semantischen Verarbeitung auf der Chart eines Parsingresultats ohne überspannende Analyse. Ziel ist auch hier die Bildung von Projektionskanten, die jeweils eine gesamte Korrektur überdecken, jedoch nur mit der Analyse der Korrekturfortsetzung markiert sind. Um die Information aus solchen Projektionen der Benutzermodellierungskomponente zugänglich zu machen, können derartige Kanten speziell markiert werden. Eine Reanalyse aller Projektionskanten kann offenlegen, welche Kanten nicht durch typische Parsing-Operationen[10] entstanden sind. Eine Reanalyse ist in jedem Fall notwendig, um herauszufinden, ob es sich bei der Struktur unter einer Projektion um eine Korrektur, eine Wiederholung oder einen Fehlansatz handelt.

Im Gegensatz zu Worms Verfahren korrigieren Ansätze wie der von Rosé und Waibel (1996) die syntaktische Analyse nicht nach, um zu einer semantischen Analyse zu kommen. Die Gründe für eine fehlschlagende Syntaxanalyse stehen somit auch für die Benutzermodellierung nicht zur Verfügung. Ähnlich verhält es sich mit der Verarbeitung nach Nederhof et al. (1997): Eine weitgehend wahrscheinliche, vollständige syntaktische Analyse soll aus partiellen Analysen aufgebaut werden, indem eine möglichst gute Abdeckung des Originalstrings mit Sequenzen lokaler Projektionen gesucht wird. Auch dieses Verfahren liefert nicht die notwendigen Informationen für eine Benutzermodellierungskomponente: Das Hauptaugenmerk liegt lediglich auf einer Vervollständigung der Analyse, was einer Eliminierung der außergrammatischen Phänomene gleichkommt. Zumindest in Kombination mit Benutzermodellierungskomponenten sollten Ansätze wie diejenigen von Rosé und Waibel (1996) und Nederhof et al. (1997) daher nicht zur Anwendung kommen.

### Andere syntaktische Aspekte

Ist eine Korrektur durch eines der genannten Verfahren identifiziert worden, steht auch die Information darüber zur Verfügung, ob ein Korrekturterm eingesetzt wurde und von welchem Typ dieser war. Die Wohlgeformtheit der Korrektur müßte mit einem zusätzlichen Parsing-Schritt beurteilt werden, für den nach Regel 3.24 die Originaläußerung vervollständigt und mit einer Konjunktion mit der Korrekturfortsetzung zusammengezogen wird. Ergibt das Parsing eine er-

[10]Gemeint sind damit Operationen wie z.B. *Shift*, *Expand* und *Complete* beim Earley-Algorithmus bzw. allgemeine Regelanwendungen in anderen Verfahren.

folgreiche Analyse einer Konjunktion ist die Korrektur wohlgeformt.[11]

Inhaltliche Aspekte von Selbstkorrekturen wie die Zuordnung zu den verschiedenen Korrekturklassen *Error, Appropriateness, Difference* bzw. die Beurteilung von Fragmenten und Fehlansätzen können, wenn überhaupt, erst nach der semantischen Analyse erfaßt werden. Auch hier bietet sich eine anytime-artige Verwendung der bisher erreichten partiellen Information an.

### 5.2.3 Inhaltliche Merkmale

In den Arbeiten aus Kapitel 3, die inhaltliche Aspekte der Äußerungen von Versuchspersonen auswerten, erwies sich neben der geringen Zahl experimenteller Studien vor allem die fehlende Formalisierung der Bewertungskriterien als hinderlich für die Integration von Versuchsergebnissen verschiedener Experimente und für die Übernahme der Ergebnisse in die Modellierung von READY.

Von einigen Autoren wurden für die Darstellung der semantischen Information propositionale Strukturen verwendet. Propositionale Analysen sind jedoch insofern als schwach zu betrachten, da sie in zu starkem Maße von individuellen (insbesondere „intuitiven") Einschätzungen abhängig sind und keine formale Interpretationsmöglichkeit in einem Modell (als referentielle Seite einer zugrundeliegenden Semantik) zur Verfügung stellen. Konsistenzüberprüfungen der Strukturen sind aus den gleichen Gründen unmöglich. Die in einigen Studien verwendete Propositionsklassifikation (zurückgehend auf Grabowski, 1991) steht in der Tradition pseudoformaler semantischer Theorien, die bereits seit Mitte der 70er Jahre für die maschinelle Sprachverarbeitung als ungeeignet bekannt sind (siehe z.B. Pinkal, 1993).

Eine bessere Analysemöglichkeit geben Strukturen in Prädikatenlogik erster oder höherer Stufe. Durch eine kompositionale Semantik in der Form von Montague (1974) für Sätze bzw. von Kamp und Reyle (1993) für Diskurse könnten präzisere und vor allem miteinander vergleichbare Ergebnisse erzielt werden. Verschiedene Probleme der Semantikkonstruktion sind jedoch auch für Sätze und Texte noch nicht vollständig gelöst, so daß Psychologen damit leider immer noch kein vollständiges Analysewerkzeug zur Verfügung steht.

Diese Diskrepanz zwischen der Auswertung experimenteller Studien und der Formalisierung der erkannten Probleme in der computerlinguistischen Verarbeitung wirkt sich zwangsläufig auf den Prozeß einer automatischen Erkennung der zur Semantik/Pragmatik gehörigen Symptome aus. Im Projekt READY liegt der Schwerpunkt auf der Identifikation und Verarbeitung von Indikatoren für Verarbeitungsbeschränkungen. Der erhebliche Aufwand für die Integration *aller* in Abschnitt 3 identifizierten semantischen und pragmatischen Phänomene ist daher im Rahmen dieser Diplomarbeit nicht zu leisten. Für einige ausgewählte Probleme aus den Bereichen der Detailliertheit und der Stereotypie von Äußerungen wurden jedoch Lösungsansätze mithilfe bestehender semantischer Verarbeitungsmöglichkeiten erarbeitet und implementiert. Ein Beispiel dafür sind die Ergebnisse der Zusammenarbeit mit dem Projekt LISA (Pinkal & Siekmann, 1998) und die Integration von Analysen des zugehörigen Prototypen CHOLI[12].

Die verschiedenen im Einsatz befindlichen Semantikformalismen mit unterschiedlichen Repräsentationsformen lassen es zudem als wenig fruchtbar erscheinen, hier notwendige Modifi-

[11]In einigen Fällen ist dafür möglicherweise bereits ein Chunking ausreichend.

[12]Siehe Abschnitt 5.3.3.

kationen und Inferenzsysteme auf diesen Strukturen zu diskutieren. Aus diesem Grund werden semantische und pragmatische Merkmale bei der Verarbeitung in READY lediglich informell dargestellt, um Eingabedaten für die Benutzermodellierungskomponente zu erhalten. Zusätzliche experimentelle Untersuchungen, deren Auswertungsmethode sich an Erkenntnissen der computerlinguistischen Semantikforschung orientiert, müssen diese empirische und verarbeitungsbedingte Lücke in der Zukunft füllen helfen.

## 5.3 Eingabeanalyse in READY

### 5.3.1 Allgemeine Aspekte der Implementation

Die Basis für die Erhebung der Sprechparameter für einzelne Äußerungen stellt ein nach den Vorgaben von Abschnitt 4.2 mit zeitlicher und syntaktischer Information annotierter Eingabestring dar, in dem außerdem alle außergrammatischen Phänomene gekennzeichnet sind. Dieser String wird von speziellen Analysekomponenten nachbearbeitet (5.3.2) oder erhält eine explizite, strukturell vollständige syntaktische Repräsentation zugeordnet. Diese Spezifikation der Eingabe steht in der Eingabekomponente von READY nach einem differenzierbaren Auswahlschritt des Benutzers (5.3.5) zur Verfügung.

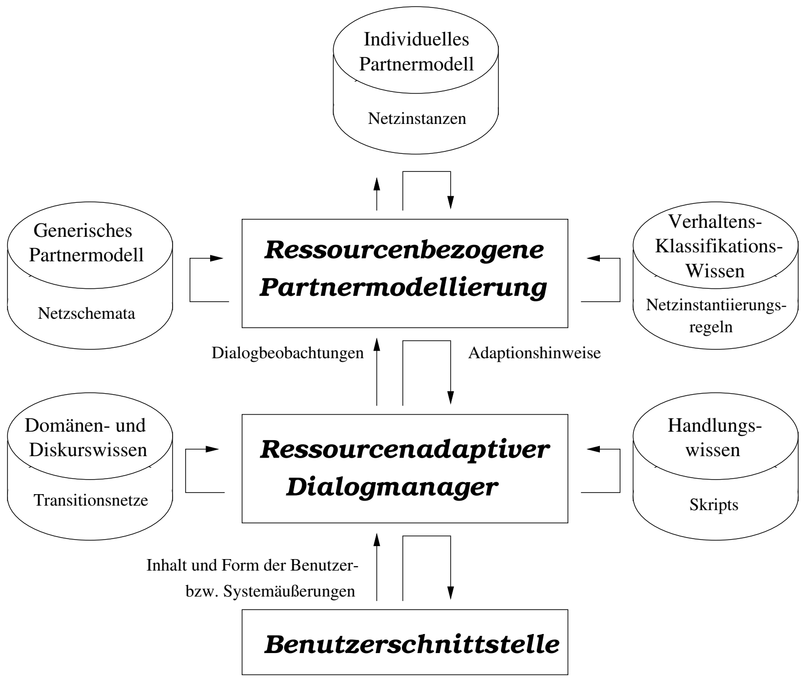


Abbildung 5.4: Architektur des READY-Prototypen

Abbildung 5.4 zeigt die Architektur des READY-Prototypen (Wahlster & Jameson, 1998b; Weis, 1999), Abbildung 5.5 gibt einen Einblick in den Datenfluß bei der Eingabeanalyse und damit auch über die Beiträge, die im Rahmen dieser Diplomarbeit zum READY-Prototypen geleistet wurden. Analysekomponenten sind durch Rechtecke dargestellt, Wissensbasen durch

Zylinder. Datenstrukturen, die für die Eingabedaten in READY aufgebaut werden, sind als abgerundete Kästchen dargestellt.

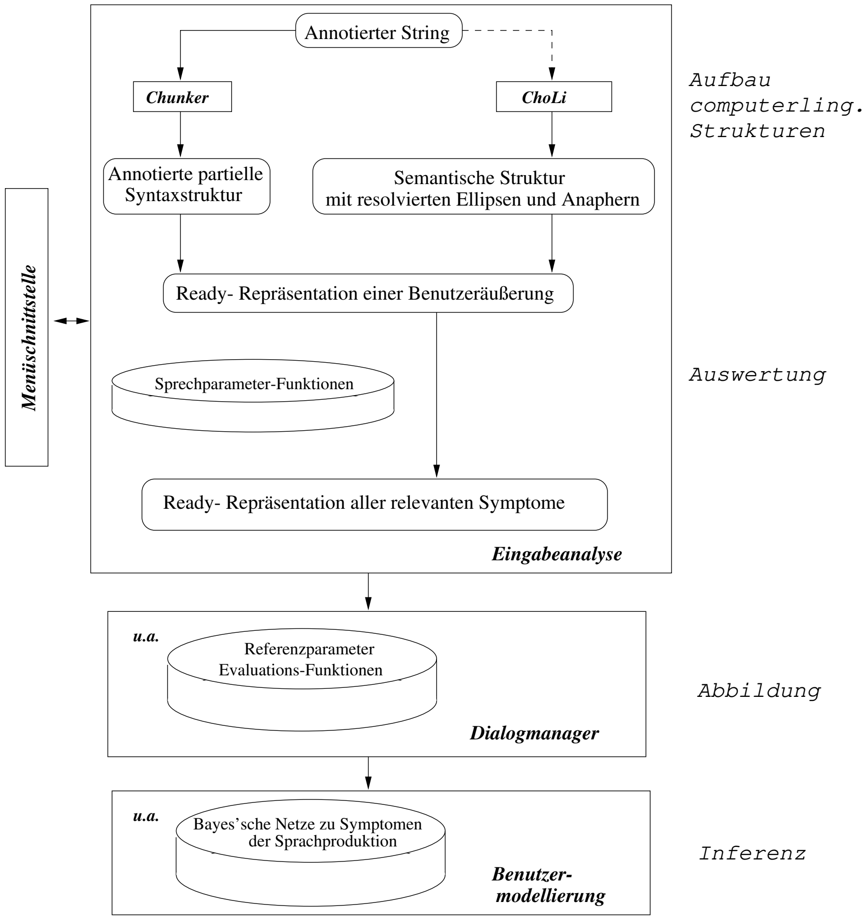


Abbildung 5.5: Datenfluß bei der Analyse einer Benutzereingabe in READY

Der Kern des Systems READY ist objektorientiert in LUCID Common LISP[13] implementiert (Weis, 1999).[14] Eine Äußerung des Benutzers wird repräsentiert durch eine Instanz der CLOS-Klasse `User-Utterance`. Spezifische Slots dieses Objekts nehmen Information über die syntaktische und semantische Repräsentation der zugehörigen Äußerung auf.

Kernstruktur für die Analyse der Sprechparameter ist ein Syntaxbaum[15], der im Sinne der integrierten Endrepräsentation (s.o.) auch Information über außergrammatische Phänomene, Pausen, Einsatzlatenz etc. enthält.

Auf dieser Struktur operiert eine Anzahl von *Sprechparameterfunktionen*, die die einzelnen Parameter so auswerten, daß sowohl für einfache Eigenschaften (wie Silbenzahl) als auch für Eigenschaften, die aus einfachen Eigenschaften berechnet werden (wie Artikulationsgeschwin-

[13]Version 4.2.2. unter Solaris auf Maschinen vom Typ SPARC 20 und SPARC Ultra 1

[14]Die Interaktion mit dem unten vorgestellten Menüinterface erfolgt über eine Lisp-Java-Schnittstelle, die die Eingabedaten jedoch ebenfalls als Lisp-Objekte zur Verfügung stellt.

[15]Im Slot `syn-structure` von `User-Utterance`.

digkeit) Absolutwerte der Äußerung zur Verfügung stehen. Einige davon werden durch Division durch den Wert der jeweiligen Basiseinheit für die Äußerungsmenge (Silbenzahl) in relative Maße überführt, sofern diese Grundlage der weiteren Analyse sind. Da für die Instantiierung der Variablen der Benutzermodellierungskomponente eine Abbildung der Absolutwerte in ein Normintervall notwendig ist[16], liegt für alle Sprechparameterfunktionen, deren Parameter einer Variablen zugeordnet ist, auch eine Abbildungsfunktion vor. Diese Funktionen sind als *Evaluationsfunktionen* zusammengefaßt und werden im Zusammenhang mit der Darstellung der Bayes'schen Netze zur Benutzermodellierung im Anhang kurz beschrieben.

Viele Sprechparameterfunktionen wurden als Methoden der Klasse `User-Utterance` implementiert, da sie auf Parameter zugreifen, die selbst Slots der zugehörigen Klasseninstanz sind. Um zu verhindern, daß bei jedem Zugriff auf einen solchen Parameter der Wert des Parameters neu berechnet wird, werden alle Absolutwerte nach ihrer Berechnung in den jeweiligen Slots der Instanz gespeichert.[17]

Als Sprechparameterfunktionen stehen derzeit zur Verfügung:

- Slot-Accessoren für Basisstrukturen:

```
(defclass User-Utterance (utterance)
  ((phon-structure ... :accessor phon-structure)
   (syn-structure ... :accessor syn-structure)
   (sem-info ... :accessor sem-info)
   ;;inherited from class 'utterance:
   (u-string ... :accessor u-string)
   (u-type ... :accessor u-type) ;; Dialogakt-Typ
   ...
   )
```

- Echte Methoden (auf der syntaktischen oder semantischen Struktur operierend, setzen den Wert der zugehörigen Slots):

```
(defmethod articulation-duration ((utt User-utterance)) ...)
(defmethod number-of-syllables ((utt User-utterance)) ...)
(defmethod all-pauses ((utt User-utterance)) ...)
(defmethod cognitive-pauses ((utt User-Utterance)) ...)
(defmethod number-of-false-starts ((utt User-Utterance)) ...)
(defmethod number-of-fragments ((utt User-Utterance))
(defmethod number-of-repetitions ((utt User-Utterance)) ...)
(defmethod number-of-errors ((utt User-Utterance)) ...)
(defmethod number-of-self-corrections ((utt User-Utterance)) ...)
(defmethod number-of-editing-terms-in-self-corrections ((utt User-Utterance))  ...)
(defmethod number-of-wellformed-self-corrections ((utt User-Utterance)) ...)
(defmethod rel-explicitness ((utt User-Utterance)) ...)
```

- Methoden, die auf Werten anderer Parameter-Slots operieren:

```
(defmethod articulation-rate ((utt User-Utterance)) ...)
(defmethod frequency-of-fragments-and-false-starts ((utt User-Utterance)) ...)
(defmethod frequency-of-repetitions ((utt User-Utterance)) ...)
(defmethod frequency-of-errors ((utt User-Utterance)) ...)
(defmethod frequency-of-self-corrections ((utt User-Utterance)) ...)
(defmethod quality-of-self-corrections ((utt User-Utterance)) ...)
```

[16] Zur Motivation siehe Kapitel 6.

[17] Diese Slots heißen so wie die Methoden, mit einem vorangestellten `abs`. Die durch die Slots erzeugte redundante Information dient lediglich einem laufzeitfreundlicheren Zugriff und besseren Debugging-Möglichkeiten. Im Prinzip kann darauf verzichtet werden.

- Slot-Accessoren für derzeit noch fest anzugebende, da nicht berechnete Parameter und Parameter des akustischen Signals:

```
(defclass User-Utterance (utterance)
   ...
   (improper-deixis ... :accessor improper-deixis)
   (quality-of-content ... :accessor quality-of-content)
   (superfluous-information ... :accessor superfluous-information)
   (anaphora-and-ellipsis ... :accessor anaphora-and-ellipsis)
   ;;inherited from class 'utterance:
   (duration ... :accessor duration)
   (speed ... :accessor speed)
   (level-of-details ... :accessor level-of-details)
   )
```

`anaphora-and-ellipsis` wird berechnet, wenn spezielle Dialogakt-Kombinationen vorliegen (siehe Abschnitt 5.3.3). Von den beiden Parametern `duration` und `speed` wird jeweils nur einer angegeben.

### 5.3.2 Berechnung einer partiellen syntaktischen Struktur: Chunker

Für die Pausenvorhersage und die Analyse syntaktischer Strukturen in bezug auf Symptome für Belastung und Zeitdruck muß sich READY auf eine syntaktische Analyse stützen können. Da als Forschungsziel von READY die Erkennung und Verarbeitung von Symptomen sowie eine Anpassung des Systemverhaltens im Vordergrund steht, steht in READY derzeit keine vollständige syntaktische Analysemöglichkeit zur Verfügung. Zusätzlich wird die syntaktische Verarbeitung im READY-Szenario durch die bekannten Probleme der Spontansprache zu einem nicht gerade trivialen Problem, dem sich komplette Forschungsprojekte widmen.

Partielle syntaktische Analysen sind ohne die Entwicklung einer Grammatik und eines detaillierten Lexikons mithilfe von statistischen Verfahren möglich. Der Grund dafür ist die relativ feste Wortfolge in bestimmten Konstituententypen wie Nominalphrasen (NPs), Präpositionalphrasen (PPs), Adverbialphrasen (APs)[18], aufgrund derer man für ein statistisches Verfahren durch Training auf einem passenden Korpus Wahrscheinlichkeitsparameter ableiten und für eine Analyse einsetzen kann.

Um in READY über diese Form der partiellen syntaktischen Analyse zu verfügen, die über simples Part-of-Speech-(POS) Tagging hinausgeht, wurde sowohl für die Voruntersuchung als auch im READY-Prototypen [19] ein partieller Parser eingesetzt, der im Projekt NEGRA[20] entwickelt wurde. Der Chunker basiert auf der Infrastruktur eines POS-Tagging-Algorithmus und stützt sich auf die POS-Annotationen des POS-Annotationstools TnT (Skut & Brants, 1998a, 1998b). Eine Beispielanalyse ist in 5.1 angegeben. Der Chunker liefert als Ausgabe eine geklammerte und mit Phrasenbezeichnungen versehene Analyse, bei der die Ausgabe des POS-Taggings beibehalten wird. Daher steht auf jeder Zeile ein Wort der Eingabe. Besonders sei auf

[18]Verbalphrasen eignen sich im Deutschen wegen der relativ freien Wortstellung in den topologischen Feldern der Satzstruktur nicht für diese Form der statistischen Verarbeitung. Dadurch werden vollständige Analysen von Sätzen mit rein statistischen Mitteln unmöglich.

[19]Der Prototyp arbeitet auf Strukturen, die durch das Chunking vorbearbeitet und z.T. für die Modellierung besonderer Phänomene von Hand erweitert wurden. Die Auswertung der Feldstudie beruhte in bezug auf intraphrasale Pausen komplett auf den Ausgabedaten des Chunkers.

[20]NEGRA ist ebenso wie READY im Sonderforschungsbereich 378 der DFG angesiedelt.

die Position von Pausen und Dehnungen hingewiesen, die sich an ihrer ursprünglichen Position im Eingabestring in die partielle Struktur eingliedern.

(5.1)
```
        (PP     (APPR   "An"    )
<D>
                (ART    "der"   )
                (ADJA   "Saarbruecker"   )
                (NN     "Strasse"        ))
<P/.485/>
        ($KOM   ","     )
        (ADV    "da"    )
        (VAFIN  "ist"   )
        (PPER   "mir"   )
        (PP     (APPRART "beim" )
                (NN      "Fahren"       ))
        (VVPP   "aufgefallen"   )
        ($KOM   ","     )
<P/.472/>
        (KOUS   "dass"   )
<D>
        (PP     (APPRART "am"   )
                (ADJA    "linken"       )
                (NN      "Vorderreifen" ))
<P/.387/>
        ($KOM   ","     )
        (KOUS   "wenn"  )
        (PPER   "ich"   )
        (KON    "fahr'" )
<P/.172/>
        (VVFIN  "fing"  )
        (PPER   "'s"    )
        (ADJD   "ploetzlich"     )
        (VAFIN  "hat"   )
        (PPER   "'s"    )
<Z>
        (VVPP   "angefangen"    )
        (VZ     (PTKZU  "zu"    )
                (VVINF  "klackern"      ))
        ($.     "."     )
```

### 5.3.3 Detailliertheit, Echo-Äußerungen und Ellipsen: CHOLI

In den Abschnitten 3.5.3 und 3.5.4 wurde auf die Bedeutung von zusätzlicher Information und der Explizitheit von Referenzen verwiesen. Im Abschnitt 4.3.6.2 der Feldstudienergebnisse wurde beschrieben, daß verschieden komplexe Antworten auf Entscheidungsfragen unterschiedliche Inferenzen über die kognitive Belastung oder den Zeitdruck eines Sprechers erlauben könnten. Echo-Äußerungen wurden als Symptome des Steuerungsmechanismus *Reizsteuerung* zu den besonders starken Indikatoren für kognitive Belastung gezählt.

Diese Eigenschaften von Benutzeräußerungen zählen zu den Merkmalen der syntaktisch-semantischen und der semantischen/pragmatischen Ebene und lassen sich für eine Äußerung daher lediglich durch spezielle Analysen und Inferenzen auf ihrer semantischen Struktur feststellen, die sie in den Diskurskontext, besser jedoch in Weltwissen einbetten kann.

Diese Probleme der semantischen Auswertung stellen aktuelle Forschungsgebiete in der computerlinguistischen Semantik dar. Das Projekt LISA[21] untersucht solche Fragestellungen vor allem unter dem Gesichtspunkt der Steuerung semantischer Auswertung und implementiert

[21] LISA ist so wie READY im Sonderforschungsbereich 378 der DFG angesiedelt.

zahlreiche der entsprechenden Verfahren. Dazu gehört beispielsweise die Resolution von Ellipsen anhand des Diskurskontextes. Der in LISA implementierte Prototyp CHOLI kann daher einige der Inferenzen liefern, die READY für seine Benutzereinschätzung benötigt.[22]

#### 5.3.3.1 Zielsetzungen von LISA, Verfahren und Funktionsweise von CHOLI

In einer vollständigen computerlinguistischen Analysekomponente stehen nach der Semantikkonstruktion auf Satz-und Textebene Strukturen zur Verfügung, die das semantische Potential der analysierten Äußerungen beschreiben. Für diesen Analyseschritt wird lediglich die semantische Information der bedeutungstragenden Einheiten selbst, jedoch kein Wissen über Elemente des vorangehenden Diskurses, pragmatische Inferenzen, oder spezielle Interpretationen im Kontext der Sprechsituation oder gar der außerhalb des Diskurses existierenden Welt herangezogen. Dadurch sind zahlreiche Informationen, die durch die lexikalische Semantik der realisierten sprachlichen Zeichen nur ambig oder unvollständig beschrieben wurden, nicht aufgelöst.[23]

Unterspezifizierte Darstellungen stellen eine Form der kompakten Darstellung solch mehrdeutiger Information dar, da sie zahlreiche unterschiedliche Interpretationen simultan repräsentieren (siehe z.B. Pinkal, 1996). Mittel zur Darstellung von unterspezifizierten Strukturen in einer logikbasierten Darstellung der Semantik sind z.B. ungebundene Variablen, wie sie auch durch die Semantikkonstruktion elliptischer Ausdrücke und anaphorischer Referenzen entstehen können.

Ellipsen, einige Typen anaphorischer Referenz und Ambiguitäten aufgrund lexikalischer Eigenschaften sowie Ambiguitäten durch Skopuseigenschaften gehören zu den in diesem Zusammenhang gut untersuchten Phänomenen. Die unterschiedlichen Lösungen für ihre unterspezifizierten Darstellungen liegen prinzipiell in einem sehr großen Suchraum, der durch Welt-und Diskurswissen vorgegeben wird. LISA untersucht Techniken, mit denen die Suche im Lösungsraum durch linguistische Information aus verschiedenen Verarbeitungsebenen gesteuert werden kann. Dazu gehören z.B. spezielle Phänomene der Diskursorganisation oder Diskursrelationen wie Parallelität und Kontrast.

Im Beispiel 5.2 (Gardent, 1997) liegt beispielsweise eine VP-Ellipse vor, die in der logischen Struktur des semantischen Potentials eine ungebundene Variable einführt. Diese Variable kann an viele verschiedene Lösungen gebunden werden. Zieht man jedoch in betracht, daß eine Parallelität zwischen beiden Sätzen besteht, die durch die Elemente 'John' und 'Peter' und die ähnlichen syntaktischen Konstruktionen bestimmt wird, so ergeben sich die prominentesten Lösungsvarianten 5.3 und 5.4. Andere mögliche Lösungen werden durch die Steuerung der Inferenz mithilfe der linguistischen Steuerungsinformation 'Parallelität' ausgeschlossen.[24]

(5.2) John likes his wife. Peter does too.

(5.3) Peter likes his wife.

[22] Die Zusammenarbeit der beiden Projekte erfolgte mit Claire Gardent vom Projekt LISA, der ich für zahlreiche Anregungen dankbar bin.

[23] Der Begriff unterschiedlicher *Lesarten* faßt unterschiedliche Typen dieser Mehrdeutigkeit zusammen.

[24] Auch unter dem Constraint der Parallelität werden noch linguistisch inadäquate Lösungen generiert. Der gesamte Prozeß ist komplizierter, als er hier auf knappem Platz dargestellt werden kann. Daher sei der interessierte Leser auf die in diesem Abschnitt angegebene Literatur verwiesen.

(5.4) Peter likes John's wife.

Im Laufe der Forschungsarbeiten in LISA wurde ein Prototyp CHOLI implementiert, der mithilfe von Constraintlösungstechniken Inferenzen auf unterspezifizierten Darstellungen durchführen kann, die Menschen bei der sprachlichen Analyse auf ganz natürliche Weise realisieren. Da die Prädikatenlogik erster Stufe für die Darstellung der natürlichen Sprache bekanntermaßen nicht ausreicht und daher Variablen höherer Stufe in den semantischen Repräsentationen auftreten, wird als Berechnungsformalismus für die Constraintlösung gefärbte Unifikation höherer Stufe, kurz HOCU, eingesetzt (siehe z.B. Gardent & Kohlhase, 1996). Die spezielle Form der Verwendung von Higher Order Unification (HOU) im Zusammenhang mit Constraintlösungstechniken in CHOLI wird auch als *Kontextunifikation* bezeichnet.[25] Siehe dazu Niehren, Pinkal und Ruhrberg (1997a, 1997b).

Der Darstellungsformalismus in CHOLI ist getypte Logik höherer Stufe. Durch die Typung verfügen alle Strukturen und Substrukturen der Darstellung über einen Typ aus der Menge der Basistypen e und t und den daraus zusammengesetzten Typen <a, b>[26].

Lösungen von Kontextconstraints liefern mögliche Lesarten der unterspezifizierten Darstellung.

#### 5.3.3.2 Interessante Fragestellungen READY – LISA

Aufeinanderfolgende Sätze unterschiedlicher Sprecher weisen oft Eigenschaften der Kohärenz auf, wie sie für aufeinanderfolgende Sätze ein und desselben Sprechers erwartet werden. Auch Dialogturn-übergreifend bestehen Möglichkeiten der Ellipsenbildung und der Verwendung anaphorischer Bezüge. Die von LISA untersuchten Phänomene der Parallelität und des Kontrasts geben daher Anregungen für die möglichen Phänomene, die im READY-Kontext untersucht werden können:

- In Abschnitt 4.3.6.2 wurde auf die Bedeutung der Beantwortung von Entscheidungsfragen im Korpus hingewiesen. Insbesondere ist dabei für READY interessant, welchen Detailliertheitsgrad eine Antwort im Verhältnis zur Frage und zur Polarität[27] der Antwort aufweist, um daraus Inferenzen über den Zeitdruck des Benutzers ziehen zu können.

- Backchanneling-Signale im Korpus wiesen häufig die Form stereotyper Wiederholungen auf (Echo-Äußerungen, Abschnitt 3.4.3). READY möchte in diesem Zusammenhang erfahren, ob und in welchem Maß ein potentielles Backchanneling-Signal Information aus der vorangehenden Systemäußerung wiederholt.

- Obwohl keine Arbeiten zur Bedeutung von Ellipsen für das Maß der kognitiven Belastung und des Zeitdrucks bekannt sind, können aus den häufig auftretenden Phänomenen des Telegrammstils auch bereits ohne experimentelle Bestätigung Annahmen über den Zeitdruck eines Sprechers abgeleitet werden.

[25] HOCU ist lediglich eine Anpassung der HOU an die Erfordernisse der Kodierung von Steuerungsinformation.

[26] Die Variablen a und b stehen hier ihrerseits für Typen aus dieser Menge. Für eine Einführung in die Typtheorie siehe z.B. (Gamut, 1991, Kap.4).

[27] 'Ja' oder 'Nein'.

- Implikationen aus der Ellipsenbildung können für spätere Phasen des Projektes READY z.B. aufgrund eigener Experimente von Nutzen sein, wenn ein Zusammenhang zwischen Ellipsenbildung und Zeitdruck oder Belastung hergestellt wird, wie er sich bereits in den diskutierten Arbeiten von Levelt und Kelter (1982) andeutete.

- Prinzipiell kann CHOLI auch mit anderweitig unvollständiger Information[28] umgehen, da CHOLI auch Verarbeitungsmechanismen des Projektes CHORUS implementiert (Worm et al., 1997). Im Mittelpunkt stehen dabei Informationen über den Grad der Ungrammatikalität bei intern unvollständigen Äußerungen, indem Maße für die ausgelassene semantische Information und damit für die Bedeutung dieser Auslassungen zur Verfügung gestellt werden.

- In einer späteren Phase könnte ein System vom Typ CHOLIs ggf. Auskunft über den Zusammenhang von abgebrochenen Äußerungen mit nachfolgenden Korrekturen geben[29], indem Möglichkeiten der Vereinbarkeit der Semantik des abgebrochenen Satzes mit den nachfolgenden Sätzen geprüft werden.

#### 5.3.3.3 Eingrenzung des Phänomenbereiches

Als erste Stufe der Zusammenarbeit wurden die Detailliertheit bei der Beantwortung von Entscheidungsfragen und die Stereotypie von Backchanneling-Signalen für eine Kopplung der Systeme LISA und READY ausgewählt, da die Mechanismen für die in der obigen Aufstellung zuletzt genannten Phänomene in CHOLI derzeit noch nicht zur Verfügung stehen. Tabelle 5.1 gibt einige Beispieläußerungen aus dem READY-Korpus mit entsprechenden Varianten an, die im Rahmen dieser Zusammenarbeit behandelt wurden.

READY erkennt an einer geeigneten Sequenz von Dialogakten, ob es sich bei der vorliegenden Abfolge einer System- und einer Benutzeräußerung um ein Phänomen handelt, das von CHOLI analysiert werden kann. Für eine Dialogaktfolge[30] $D_S, D_P$ versucht READY genau dann eine Analyse von CHOLI zu erhalten, wenn eine der folgenden Bedingungen gilt:

- $D_S =$ \$question–yes-no
  $D_P =$ \$answer–yes-no

- $D_S \in$ Instruction
  $D_P \in$ Backchanneling

Ziel der Analyse dieser unterschiedlichen Äußerungen aus der Perspektive von READY ist es, Maße für die Detailliertheit[31] von Antworten auf Entscheidungsfragen (Beispiele 1-3 aus Tabelle 5.1) und die Stereotypie von Backchanneling-Signalen (Beispiel 4 und 5) zu erhalten.

[28] Z.B. mit intern oder extern unvollständigen Äußerungen.

[29] Im Laufe der vorliegenden Arbeit wurde bereits oft darauf hingewiesen, daß bislang automatische Unterscheidungsmöglichkeiten für Fehlansätze und Satzfragmente fehlen.

[30] Zu den verschiedenen Dialogakten und Dialogaktklassen siehe Tabelle 4.3.

[31] Im folgenden wird von *Explizitheit* als Spezialfall der *Detailliertheit* gesprochen, da letztere einen weiteren Bereich abdeckt.

Dafür ist ein Maß mit einer Basiseinheit zu definieren und zu beschreiben, wie den unterschiedlichen Analyseergebnissen diese Maße zugewiesen werden. Grundlage für die Bewertung sind die Strukturen der Analyse, daher werden im folgenden Abschnitt zunächst Ein- und Ausgabe bei der Kopplung von READY und CHOLI vorgestellt.

| | | | | |
|---|---|---|---|---|
| (1) | $\mathcal{S}$: | Haben Sie den Keilriemen gefunden? | $\mathcal{P}$: | Ja.<br><br>Ja. Ich habe ihn gefunden.<br>Ja. Ich habe den Keilriemen gefunden. |
| (2) | $\mathcal{S}$: | Ist der Keilriemen locker? | $\mathcal{P}$: | Ja.<br>Ja. Er ist locker.<br>Ja. Der Keilriemen ist locker.<br>Nein.<br>Nein. Er ist fest.<br>Nein. Der Keilriemen ist fest. |
| (3) | $\mathcal{S}$: | Zischt's noch? | $\mathcal{P}$: | Ja. zischt noch. |
| (4) | $\mathcal{S}$: | Der Luftfilter ist von Dir aus gesehen links. | $\mathcal{P}$: | Der Luftfilter ist von mir aus gesehen links.<br>Der Luftfilter.<br>Links. |
| (5) | $\mathcal{S}$: | Die Radmuttern mal nachziehen. | $\mathcal{P}$: | Nachziehen.<br>Radmuttern nachziehen. |

Tabelle 5.1: Beispielsätze aus der Zusammenarbeit READY–LISA

#### 5.3.3.4 Formalisierung der Beispieläußerungen und CHOLI-Resultate

Wie oben bereits angegeben, wird zur Auflösung unterspezifizierter Darstellungen in CHOLI Kontextunifikation eingesetzt. Dafür werden Kontextconstraints in Form von Gleichungen über lokalen Strukturen[32] aufgestellt. In den einzelnen Lösungsschritten werden den in den Gleichungen enthaltenen Variablen Werte zugeordnet, so daß die Lösungen aller Gleichungen konsistent sind. Die Darstellung der Frage im ersten Beispiel von Tabelle 5.1 in Logik höherer Stufe, die von CHOLI generiert wird, ist in 5.5 dargestellt, die der zweiten Antwortalternative dazu in 5.6. Die Darstellung gibt Ausgabedaten von CHOLI in der Form wieder, wie sie aus der CHOLI-internen OZ-Darstellung[33] in LISP-Code transformiert wurden. Bei dieser Darstel-

[32]Bei der Kontextunifikation handelt es sich dabei konkret um Bäume der Objektsprache, die ggf. Variablen enthalten.

[33]CHOLI ist in der Constraintsprache OZ (Smolka, 1995) implementiert.

lung sind gegenüber der prädikatenlogischen Notation die Quantoren $\forall$ durch `\forall` und $\exists$ durch `\exists` ersetzt worden. `\and` steht für $\wedge$ , `\equiv` für $\leftrightarrow$ und `\neg` für $\neg$. Da die Grammatik zur Äußerungsanalyse in CHOLI eine Grammatik des Englischen ist, müssen alle Beispiele vor der Übergabe an CHOLI in einem Übersetzungsschritt in eine geeignete Darstellung des Englischen überführt werden.[34] Die Prädikate der Objektsprache von CHOLI können prinzipiell beliebig kodiert werden, sind hier jedoch ebenfalls an das Englische angelehnt. Zur Vereinfachung sind die Übertragungen in prädikatenlogische Notation noch einmal angegeben.

(5.5) $\mathcal{S}$: Haben Sie den Keilriemen gefunden?
$\exists z \forall u(\text{fan\_belt}(u) \leftrightarrow z = u) \wedge \text{pos}(\text{find}(customer, z))$

```
\exists z(\forall u(fan_belt(u)\equiv z = u)
                  \and pos(find(customer)(z)))
```

(5.6) $\mathcal{P}$: Ja. Ich habe ihn gefunden.
$\exists \text{z} \forall \text{u}(C14(\text{u}) \leftrightarrow \text{z} = \text{u}) \wedge \text{pos}(\text{find}(customer, z))$

```
\exists z(\forall u(C14(u)\equiv z = u)
                    \and pos(find(customer)(z)))
```

Neben den Variablen $z$ und $u$, die durch Quantoren gebunden werden und insbesondere zur Semantik des bestimmten Artikels 'der'[35] bzw. der definiten Anapher '*ihn*' innerhalb der logischen Struktur beitragen, enthält 5.6 eine *Diskursvariable* $C14$, die für den Teil der Semantik von '*ihn*' steht, der über die Eigenschaft der Definitheit hinausgeht. $customer$ ist der Einfachheit halber ebenfalls als Variable dargestellt, steht jedoch nur als Kurzform für einen definiten Ausdruck im Sinne von Russell (siehe Levinson, 1990)[Kap. 4], der den einzigen und eindeutig referenzierbaren Anrufer in der Domäne bezeichnet. Die komplette Semantik des definiten Ausdrucks würde die Lesbarkeit der Formel nur unnötig erschweren.

Aus der Semantik der beiden Äußerungen bildet CHOLI passende Constraintgleichungen, die in 5.7 angegeben sind. Die *Kontextvariable* $C0$ dient zur Festlegung der Parallelität: Die Lösung für $C14$ muß, wenn sie in die erste Gleichung eingesetzt wird, mit der Semantik von $C0$ aus der zweiten Gleichung unifizierbar sein. Die Parallelität ist hier auf sehr einfacher Ebene hergestellt, da keine weiteren parallelen Elemente (wie etwa bei Gardent, 1997) vorliegen.

(5.7)
```
C0 = \exists z(\forall u(fan_belt(u)\equiv z = u)
     \and pos(find(customer)(z))) \\
```
$$C0 = \exists z \forall u(\text{fan\_belt}(u) \leftrightarrow z = u) \wedge \text{pos}(\text{find}(customer, z))$$

```
C0 = \exists z(\forall u(C14(u)\equiv z = u)
     \and pos(find(customer)(z)))
```

$$C0 = \exists z \forall u(C14(u) \leftrightarrow z = u) \wedge \text{pos}(\text{find}(customer, z))$$

[34]Das ist nicht immer unter exakter Beibehaltung der Phänomene möglich und wird mit der Entwicklung einer deutschen Grammatik für CHOLI zu beheben sein.

[35]Zur Semantik von definiten Ausdrücken siehe z.B. Gamut (1991) oder Levinson (1990, Kap.4).

CHOLI findet eine Bindung für die Diskursvariable(n)[36] und übergibt die Lösung an READY als Struktur der Form von 5.8. Dabei wird mit dem Slot :cvar die während der Kontextunifikation eingesetzte Kontextvariable beschrieben: Angegeben sind ihr Name (`:name`) und ihr semantischer Typ (`:type`). `:dvars` liefert Gleiches für die Diskursvariablen. READY benötigt diese Beschreibungen u.a. für die spätere Bewertung des Unifikationsergebnisses. Der Slot `:subst` gibt die Substitutionen für die beteiligten Variablen an, die durch den Constraintlösungsprozeß gefunden werden konnten. Für die Semantikkonstruktion wird in CHOLI ein getypter Lambda-Kalkül eingesetzt, daher ist auch die Lösung in dieser Form angegeben. Durch Anwendung der Lösung auf die zweite Gleichung von 5.7 und anschließende $\beta$-Reduktion erhält man die Semantik der Ausgangsäußerung. `\lambda` steht hierbei für den Operator $\lambda$.[37] Die Semantik (`:sem`) der Diskursvariable $C14$ entspricht der einfachen lexikalischen Semantik des logischen Prädikates 'Keilriemen'.

Die Interaktion von READY mit dem System CHOLI ist bis zur Übergabe dieser Strukturen noch einmal in Abbildung 5.6 dargestellt.

(5.8)
```
;Substitutions
:cvar  (:name C0
         :type t)
:dvars ((:name C14
          :type (t e)))
:subst ((:name C14
          :sem (\lambda B4.fan_belt(B4)))
         (:name C0
          :sem (\exists z(\forall u(fan_belt(u)\equiv z = u)\and
                                    pos(find(customer)(z)))))))
```

#### 5.3.3.5 Analyseschritte auf den CHOLI-Strukturen in READY und Ergebnisse

**Bewertung des semantischen Typs** Die Bewertung der Unifikationsergebnisse von CHOLI soll für READY die Basis für ein Maß liefern, mit dem die Explizitheit einer Benutzeräußerung in Abhängigkeit von der Systemfrage oder Systeminstruktion eingeschätzt werden kann. Von Bedeutung ist dabei vor allem der semantische Typ der vom Benutzer *nicht* sprachlich realisierten Konstituenten: Dieser Annahme liegt die Idee zugrunde, daß der Benutzer bei maximaler Explizitheit die komplette Proposition der Systemäußerung erneut verbalisieren würde.

Das Maß der Explizitheit wird entsprechend so definiert, daß sein höchster Wert auf 1 und der niedrigste auf 0 abgebildet wird.[38] Vereinfachend werden als relevant für die Beurteilung der Explizitheit die folgenden beiden Merkmale angenommen:

- Wurden relativ wichtige Diskursbestandteile (z.B. ganze Propositionen, Referenten und Relationen) oder weniger wichtige (wie z.B. Modifikatoren und Präpositionen) weggelassen? Diese Eigenschaft ist am semantischen Typ der Diskursvariable(n) ablesbar.

[36] Mehrere Diskursvariablen sind möglich. Dieser Fall kommt jedoch bei der Interaktion mit READY derzeit nicht vor.

[37] Für eine detailliertere Diskussion des Lambda-Kalküls siehe ebenfalls Gamut (1991).

[38] Das kommt auch Aspekten der Implementation entgegen. Siehe Anhang B.1.

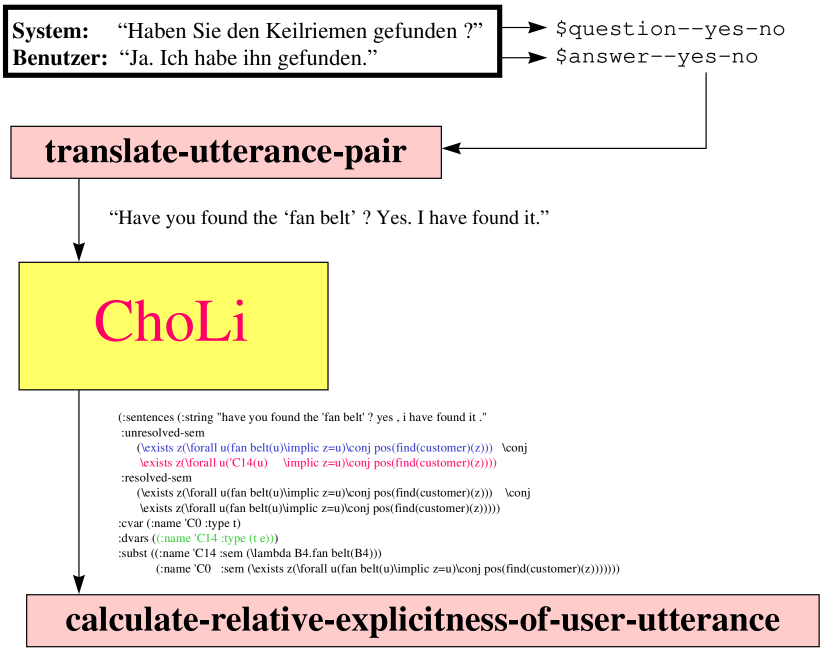


Abbildung 5.6: Eingabe und Ausgabe READY-CHOLI

- Wie komplex war die ausgelassene Konstituente? Diese Information gibt die Struktur der Diskursvariablen wieder.

Aus diesen Beobachtungen ergibt sich die Berechnung der Explizitheit wie folgt: Drei verschiedene semantische Typen können in dem hier verwendeten Szenario für elidierte oder anaphorische Ausdrücke auftreten. Ihnen wird durch die READY Funktion `evaluate-type` der in Tabelle 5.2 angegebene numerische Wert zugeordnet.[39] Dabei wird darauf geachtet, daß der Wert eines semantischen Typs umso höher ist, je unwichtiger er im Kontext eingestuft wird. Die Motivation dafür ist, daß ein Sprecher als Reaktion auf eine Systemäußerung getrost Dinge wiederholen darf. Eine Wiederholung läßt jedoch umso weniger Zeitdruck vermuten, je unwichtigere Fakten der Sprecher erneut versprachlicht. Referenten und Propositionen sind dabei im Diskurskontext meist besonders prominent und führen zu hoher Redundanz.[40]

[39] LISA verwendet für die Notation der Typen eine im Vergleich zu Gamut (1991) umgekehrte Schreibweise: Für eine Struktur A vom Typ $(b\ a)$ gilt: Wendet man A auf eine Struktur C vom Typ $a$ an, so besitzt die Resultatsstruktur B den Typ $b$. Bei Gamut (1991) steht dafür $< a, b >$.

[40] Wiederholt sei hier darauf hingewiesen, daß dieser Ansatz weit davon entfernt ist, Anspruch auf Endgültigkeit zu erheben. Er dient lediglich dazu, eine erste Annäherung und Verwendungsfähigkeit von Information aus Modulen der semantischen Verarbeitung bei der Benutzermodellierung unter Beweis zu stellen.

| Typ | Semantik zugeh. Ausdrücke | Beispiel | Wert |
|---|---|---|---|
| $t$ | Proposition | 'Ich habe den Keilriemen gefunden.' | 0 |
| $(t\ e)$ | Prädikat erster Ordnung | 'Keilriemen' | 0.5 |
| $((t\ e)\ (t\ e))$ | Prädikatsmodifikator | 'gefunden' | 0.8 |

Tabelle 5.2: Bewertung des semantischen Typs in READY

**Bewertung der Komplexität** Für die Bewertung des elidierten oder anaphorischen Ausdrucks ist es nicht nur notwendig, den semantischen Typ des Ausdrucks zu kennen, sondern auch zumindest grobes Wissen über seine Struktur zu besitzen. Zum Beispiel sind die Ausdrücke 'der Keilriemen' und 'der kleine grüne Keilriemen, der sich ganz hinten links befindet' nach vollständiger Reduktion vom Typ $(t\ e)$; der zweite Ausdruck ist jedoch aus wesentlich mehr Prädikaten und Modifikatoren aufgebaut.

Die Analyse der CHOLI-Strukturen macht dabei weitere vereinfachende Annahmen, wozu u.a. gehört, daß derzeit kein Strukturvergleich ausgeführt wird. Die Menge der Prädikats- und Modifikatorenkonstanten in der Diskursvariablen zeigt an, wie stark die ursprüngliche Äußerung vom Sprecher verkürzt wurde. Dieser Wert kann als Komplexitätsmaß eingesetzt werden. Er ist jedoch relativ zur gesamten, in der Äußerung enthaltenen Information zu beurteilen, so daß eine komplexe Ellipse in einer knappen Äußerung mehr Gewicht besitzt, als in einer sehr ausführlichen.

Daher wird stattdessen das Verhältnis der Anzahl von Konstanten in der Systemsemantik $P_S$ und der Anzahl von Prädikaten in der Benutzersemantik $P_P$ herangezogen.[41] Für komplexere Fälle oder Fälle mit mehreren Diskursvariablen ist diese Vorgehensweise möglicherweise nicht mehr adäquat, liefert jedoch für die vorliegenden Daten ein ausreichend differenzierbares Maß. Für einen Strukturvergleich müssen passende Komplexitätsmaße auf Bäumen mit sogenannten Löchern definiert werden. Ein breiterer Korpus mit elliptischen und anaphorischen Phänomenen als der READY vorliegende ist dafür jedoch notwendig.

**Gesamtmaß und Verwendung in READY** Die Evaluationsfunktion `rel-explicitness` berechnet aus der Formel 5.9 die relative Explizitheit $E$ der Benutzerantwort. $P_S$ und $P_P$ sind im vorangegangenen Abschnitt erklärt, $V_{sem-typ}$ steht für die Bewertung des semantischen Typs der Ellipse oder Anapher. Tabelle 5.3 gibt einige Beispiele der Berechnung.

Da die Bayes'schen Netze des READY-Systems (siehe Kapitel 6) möglichst keine Knoten enthalten sollen, die nur bei speziellen Äußerungstypen instantiiert werden, geht der Gesamtwert für die Explizitheit von Antworten in die Berechnung des Grades an Anaphern und Ellipsen $A$ nach Formel 5.10 ein.

[41]Voraussetzung dafür ist, daß der Benutzer die Systemproposition nicht durch zusätzliches sprachliches Material modifiziert. Derartige Fälle sind jedoch auch für CHOLI je nach Art der Modifikation derzeit nicht immer auflösbar.

$$E = \frac{P_S + 1}{P_P + 1} * V_{sem-typ} \tag{5.9}$$

$$A = 1 - E \tag{5.10}$$

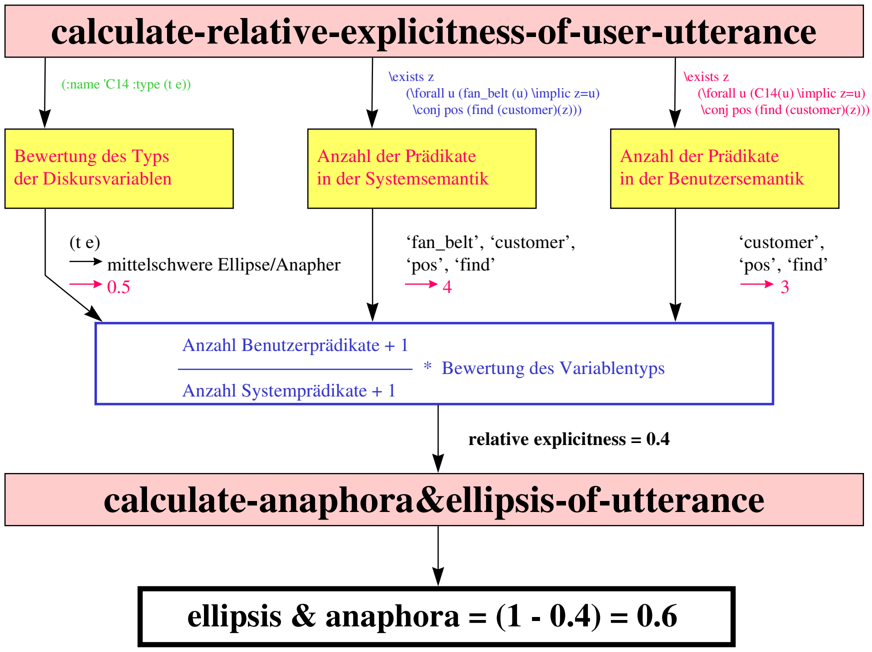


Abbildung 5.7: Analyseschritte auf der CHOLI-Ausgabestruktur

| Analyseschritt | 'Haben Sie den Keilriemen gefunden?' | |
|---|---|---|
| | 'Ja.' | 'Ja, ich habe ihn gefunden.' |
| CHOLI-Rückgabewert | | |
| Systemsemantik | $\exists z \forall u(\text{fan\_belt}(u) \leftrightarrow z = u) \wedge \text{pos}(\text{find}(customer, z))$ | |
| Benutzersemantik | $C1$ | $\exists z \forall u(C24(u) \leftrightarrow z = u) \wedge \text{pos}(\text{find}(customer, z))$ |
| Diskursvariable | C1 | C24 |
| Typ d. Diskursvariable | t | (t e) |
| Bewertung des sem. Typs | 0 | 0.5 |
| Präd. d. Systemäußerung $P_S$ | 4 | 4 |
| Präd. d. Benutzeräußerung $P_P$ | 1 | 3 |
| Verhältnis $P_P : P_S$ | 0.4 | 0.8 |
| Rel. Explizitheit nach Formel 5.9 | 0 | 0.4 |

Tabelle 5.3: Beispiele für die Berechnung der Explizitheit von Benutzerantworten

#### 5.3.3.6 Fazit und Ausblick

Mit der Kopplung der Systeme READY und CHOLI wurde gezeigt, daß auch mit formalen Mitteln die Auswertung semantischer Eigenschaften von Äußerungen in einer für die Benutzermodellierung sinnvollen Weise durchgeführt werden kann. Dabei wurde deutlich, daß nicht nur die endgültige Darstellung semantischer Information, sondern auch Zwischendarstellungen bzw. die Ausgaben von semantischen Analyseprozessen selbst zu Erkenntnissen für die Benutzermodellierung herangezogen werden können.

Grundlage für die Auswahl der Äußerungen war vor allem ein Kompromiß zwischen tatsächlich im Korpus der READY-Feldstudie vorkommenden Phänomenen mit einem experimentellen Hintergrund und den aktuellen Möglichkeiten der semantischen Verarbeitung in CHOLI. Die hier vorgestellten Mechanismen sollten daher als Anfang gesehen werden für eine Erweiterung der Auswertung semantischer Strukturen im Hinblick auf Verarbeitungsbeschränkungen.

Bei den hier angegebenen Äußerungen war es fast immer möglich, eine englische Übersetzung anzugeben und dabei die interessanten strukturellen Charakteristika zu erhalten. Die Zusammenarbeit sollte um neue Phänomenklassen erweitert werden wie z.B. weitere der oben angegebenen oder Phänomene struktureller syntaktischer Probleme, wie sie von Worm et al. (1997) angegeben werden. Dafür wäre eine robuste Grammatik des Deutschen für die Analyse in CHOLI jedoch wünschenswert. Mit einer Erweiterung der Fähigkeiten von CHOLI sollten experimentelle Studien im READY-Szenario einhergehen, die den Zusammenhang von Phänomenen der Ellipsis mit Verarbeitungsbeschränkungen untersuchen und damit die Verwendung derartiger Indikatoren auf eine breitere empirische Grundlage stellen.

### 5.3.4 Andere syntaktische und semantische Parameter

Die Berechnung der meisten Sprechparameter kann direkt auf der syntaktischen Eingabestruktur erfolgen, die die Markierungen aus der in Kapitel 4.2 vorgestellten Annotation enthält. Für einige der grundlegenden Parameter wie die Silbenzahl und spezifische Parameter mit vom natürlichen Vorgehen abweichender Berechnung soll in den folgenden Abschnitten kurz auf die Berechnungsform eingegangen werden.

#### 5.3.4.1 Berechnung von Silbenzahl und Sprechdauer

In READY steht für die Äußerungsanalyse kein Lexikon zur Verfügung. Aus diesem Grund muß ein alternativer Algorithmus für die Ermittlung der Silbenzahl herangezogen werden:

Die Struktur einer Silbe besteht im allgemeinen aus drei Teilen (Onset, Nukleus und Coda). Der Nukleus wird meist durch einen Vokal realisiert. Spezielle phonologische Regeln, die zu Silbenmodifikationen in konkreten Kontexten führen, sollen dabei der Einfachheit halber vernachlässigt werden. Im Deutschen können an der Stelle des Nukleus auch Diphthonge oder verlängerte Vokale (beide im Schriftbild durch zwei Buchstaben vertreten) eingesetzt werden. Daraus ergibt sich eine einfache Heuristik zur Ermittlung der Silbenzahl, die auf der Anzahl der Vokale für jedes Wort im Eingabstring aufbaut Die Methoden `number-of-syllables` und `count-syllables-of-word` ermitteln diese Parameter.

Die Sprechdauer von Äußerungen muß wegen der graphemischen Eingabe explizit in der Eingabekomponente angegeben werden. Eine Default-Einstellung mit einer Default-Artikulationsgeschwindigkeit wird vorgegeben und kann verwendet werden.

#### 5.3.4.2 Pausen

Die Erhebung der kognitiven Pausen geht auf den Algorithmus zur Pausenvorhersage von Gee und Grosjean (1983), die Aussagen von Butterworth (1980a), Roßnagel (1995a) etc. zur Bedeutung intraphrasaler Pausen sowie Erkenntnisse über gefüllte Pausen zurück. Der Algorithmus von Gee und Grosjean (1983) ist in 5.11 dargestellt, wird jedoch aus verschiedenen, unten erläuterten Gründen derzeit in READY nicht eingesetzt. Die Vorverabeitung in READY ist in 5.12 zu finden.

**Pausenvorhersage nach Gee und Grosjean (1983)**

Ziel des Algorithmus' von Gee und Grosjean (1983) ist es, aus prosodischen Informationen von Intonationseinheiten einen Intonationsbaum auf der Wortebene aufzubauen und allen Wortgrenzen eine relative Pausendauer im Satz zuzuordnen. Da bei Gee und Grosjean (1983) Binärverzweigung angenommen wird, entspricht ein Knoten des Intonationsbaumes einer Wortgrenze.

Der Algorithmus wurde für das Englische entworfen. Ob eine Evaluation auf deutsch- oder anderssprachigen Daten existiert, ist nicht bekannt. Zumindest von der Konstituentenstruktur her sind Deutsch und Englisch jedoch soweit verwandt, daß eine Übertragbarkeit der Analyse angenommen werden kann.[42]

Der Originalalgorithmus basiert auf dem Konzept prosodischer Einheiten der Artikulation, sogenannten $\Phi$-Phrasen, die sich aus Betonungsmustern von Worten und deren Zusammenfassung zu Artikulationsfolgen bilden lassen. Die Phrasenbildung orientiert sich neben diesen prosodischen Parametern vor allem an der syntaktischen Struktur des jeweiligen Satzes.

$\Phi$-Phrasen lassen sich zu größeren prosodischen Einheiten ($I$-Phrasen)[43] zusammenfassen. Der Algorithmus besitzt einen iterativen Kern, der innerhalb der höchsten Projektion S einer syntaktischen Struktur bottom-up komplexe prosodische Einheiten formt und die Wortgrenzen innerhalb dieser Struktur mit Vorhersagewerten für die erwarteten Pausen versieht. Die Darstellung des Algorithmus in 5.11 ist gegenüber Gee und Grosjean (1983) an eine sequentiell orientierte Verarbeitung angepaßt.

(5.11)
1. **Vorverarbeitung:** Kennzeichne die Grenze zwischen Funktionsworten[44] und nachfolgenden Worten mit dem Marker '#'. Markiere die syntaktischen Köpfe.
2. **Iterationsbeginn:** Wähle den initialen S-Knoten[45] der Struktur als Knoten N aus.
3. **Segmentierung:** Durchlaufe die Blattknoten von N von links nach rechts und fasse sie (rechtsverzweigend) zu $\Phi$-Phrasen zusammen: Sobald ein phrasaler Kopf gefunden wird, wird dieser mit dem vorangehenden Material bis zur letzten gefundenen $\Phi$-Phrase zu einer $\Phi$-Phrase zusammengefaßt.

[42] Einige notwendige Modifikationen des Algorithmus sind bereits im Aufsatz von Gee und Grosjean (1983) aufgezählt, die jedoch noch keine Berücksichtigung gefunden haben. Eine Verbesserung des Algorithmus ist hier nicht angestrebt.

[43] I steht für *intonational*.

[44] Funktionsworte wie Präpositionen sind normalerweise syntaktische Köpfe ihrer Phrasen, gleichzeitig jedoch prosodische Einheiten mit geringer Betonung. Sie werden daher von Gee und Grosjean (1983) als klitische Einheiten betrachtet, die sich mit dem nachfolgenden Wort verbinden, so daß dazwischen keine Pause entstehen sollte.

[45] Der initiale S-Knoten kann der S-Knoten des gesamten Satzes, jedoch auch die höchste Projektion eines eingebetteten Satzes sein.

4. **Markierung der $\Phi$-Phrasen:** Für jede $\Phi$-Phrase: Markiere alle syntaktischen Knoten K, die ein '#' dominieren, mit dem Wert 0. Sofern K einen Knoten K' dominiert, der seinerseits ein '#' dominiert, markiere ihn mit dem Wert 1. Markiere alle anderen syntaktischen Knoten M in der $\Phi$-Phrase mit dem Wert 'Anzahl der von M dominierten, verzweigenden syntaktischen Knoten + 1'

5. **Zusammenfassung von $\Phi$-Phrasen zu $I$-Phrasen:**

   (a) **Syntaktische Konstituenten:** Alle direkt aufeinanderfolgenden $\Phi$-Phrasen, die zur gleichen syntaktischen Konstituente (außer der VP) gehören, werden zu einer $I$-Phrase zusammengefaßt. Die Zusammensetzung erfolgt rechtsverzweigend.

   (b) **Verben:** Fasse die $\Phi$-Phrasen mit verbalem Kopf entweder mit der nachfolgenden oder mit der vorangehenden $\Phi$- bzw. $I$-Phrase zusammen, wobei als Entscheidungskriterium die geringere syntaktische Komplexität[46] der entstehenden Phrase dient.

   (c) **Verbleibende Elemente:** Bündele alle aufeinanderfolgenden $\Phi$- und $I$-Phrasen, die nicht von 5a und 5b erfaßt wurden, zu jeweils einer $I$-Phrase.

6. **Markierung der I-Phrasen:** Markiere alle $I$-Knoten K mit dem Wert 'Anzahl aller von K dominierten $\Phi$- und $I$-Knoten + 1'.[47]

7. **Iteration:** Wenn kein S-Knoten mehr existiert, der noch nicht bearbeitet wurde, gehe zu 8. Sonst weiter mit Schritt 3.

8. **Zusammenfassung von I-Phrasen aus S-Knoten zu komplexeren I-Phrasen:** Fasse alle I-Phrasen aus aufeinanderfolgenden S-Knoten zu einer $I$-Phrase zusammen und markiere sie wie in Schritt 6.

9. **Feinkorrektur 1:** Erhöhe den Wert aller Grenzen vor oder hinter einem komplexen Wort (d.h. ein Wort, das aus mehr als einer betonten Silbe besteht) um 1, sofern die rechte und linke Konstituente zur gleichen $\Phi$-Phrase gehört.

10. **Feinkorrektur 2:** Erhöhe den Wert von '#' -Grenzen um 1, wenn das nachfolgende Wort satzfinal ist und die Satzbetonung trägt.

### Pausenvorhersage in READY

Im Projekt READY wird die syntaktische Analyse von einem partiellen Parser (Chunker) ausgeführt, der Nominal-, Adverbial- und Präpositionalphrasen sowie Konjunktionen, jedoch keine Verbalphrasen erkennen kann. Solange keine vollständige syntaktische Struktur spezifiziert wird, ist die Anwendbarkeit der Pausenvorhersage eingeschränkt auf die Pausen im Inneren und an den Grenzen dieser Phrasen. In diesen Fällen ist eine detaillierte Vorhersage nicht sinnvoll und es werden pauschale Gewichtungen des der Pausen nach dem Grad ihrer Realisierung der kognitiven Verzögerungsfunktion vorgenommen.

Für von Hand spezifizierte syntaktische Strukturen wurde der Algorithmus testweise implementiert, wobei auch spontansprachliche Phänomene soweit einbezogen wurden, daß für die unvollständigen Teilstrukturen der unterschiedlichen Phasen von Selbstkorrekturen und Wiederholungen Teilvorhersagen getroffen werden können. Es ist jedoch offen, wie diese Vorhersagen in bezug auf die Lesevorhersagen von Gee und Grosjean (1983) zu bewerten sind bzw. welche Vorhersage für Unterbrechungsstellen, Pausen in der Korrekturphase und für Satzgrenzen getroffen werden sollen.

Das tatsächliche Pausenmuster wurde gegenüber der Vorhersage durch eine Verteilung der tatsächlichen Gesamtpausendauer (stiller Pausen) auf die vorliegenden Einzelpausen und Bildung der Differenz bewertet. Im allgemeinen ist dieses Vorgehen jedoch allein deshalb nicht anwendbar, da die Gesamtpausendauer bereits vom kognitiven Anteil an allen Pausen abhängt. Weiterführende Überlegungen hierzu bestehen derzeit nicht.

[46] Geringere syntaktische Komplexität bedeutet bei Gee und Grosjean (1983): weniger Verzweigungen.

[47] Wird hier bottom-up vorgegangen, kann der Wert auch aus den Werten der direkt dominierten Knoten berechnet werden.

Der Algorithmus von Gee und Grosjean (1983) wird daher derzeit in READY nicht eingesetzt und durch eine vereinfachte Variante ersetzt, die sich an den Möglichkeiten der automatischen syntaktischen Analyse mit dem Chunker (Abschnitt 5.3.2) orientiert:

(5.12)
- Gefüllte Pausen werden vereinfachend als rein kognitiv motiviert angesehen und gehen mit ihrer Gesamtdauer in die 'Effektive Dauer kognitiver Pausen' ein.
- Für stille Pausen können zwei unterschiedliche Bewertungen vorgenommen werden (mit Hilfe des Schalters `*use-pause-time*`):
  - Die stillen Pausen gehen ebenfalls mit ihrer Gesamtdauer in die Berechnung der effektiven Dauer kognitiver Pausen ein.[48]
  - Lediglich der Unterschied zwischen intraphrasalen, extraphrasalen und Pausen an Satzgrenzen wird berücksichtigt, indem die extraphrasalen, satzinternen Pausen mit einem Faktor `*weight-of-extraphrasal-pauses*` und die stillen Pausen zwischen Sätzen mit einem Faktor `*weight-of-clause-boundary-pauses*`[49] multipliziert werden und ihre Dauer so gewichtet in die Gesamtsumme eingeht.

#### 5.3.4.3 Selbstkorrekturen

Die Anzahl der Selbstkorrekturen wird durch die im Kapitel 4.2 vorgestellte Annotation des Eingabestrings zur Verfügung gestellt. Eine gesonderte Analyse der Qualität wird auf der Basis einer Wohlgeformtheits-Kennzeichnung der Korrektur[50] und dem Erfolg bei der Suche nach einem Korrekturterm hinter der Unterbrechungsstelle durchgeführt. Dieser Wert ergibt den Parameter 'Durchschnittliche Qualität von Selbstkorrekturen'. Wie bereits im Kapitel 3 angedeutet wurde, sind unterschiedliche Selbstkorrekturen im Zusammenhang mit Belastung und Zeitdruck jedoch noch relativ schwach untersucht und lassen ein erhebliches indikatorisches Potential erwarten. Daher ist die hier gewählte Analyse in hohem Maße als vorläufig anzusehen.

#### 5.3.4.4 Semantische Parameter

Den unterschiedlichen Eingabeäußerungen sind Einschätzungen der semantischen Eigenschaften auf einer Skala von 0 bis 1 zugeordnet. Automatische Analysemethoden für die einzelnen semantischen Merkmale existieren bis auf die im folgenden Abschnitt dargestellte Bewertung z.Zt. nicht. Sie müssen daher im Einzelfall noch nach informellen Kriterien beurteilt werden.

---

[48] In diesem Fall muß beachtet werden, daß eine geänderte Instantiierung der Bayes'schen Netze notwendig ist, damit nicht durch einen unangemessen hohen Pausenindex Vorhersageresultate beeinträchtigt werden.

[49] Während der Testphase des Systems ist *weight-of-extraphrasal-pauses* auf 0.8 und *weight-of-clause-boundary-pauses* auf 0.5 eingestellt.

[50] Eine automatische Analyse ist durch die unvollständige syntaktische Analyse nicht möglich.

### 5.3.5 Auswahl von Äußerungseigenschaften: Menüschnittstelle

Für die Interaktion mit dem Prototypen in READY wurde im Projekt eine Menüschnittstelle[51] realisiert. In deren Gestaltung sind zahlreiche Anregungen zur Systemsteuerung eingeflossen, die mit der Parametrisierung einzelner sprachlicher Symptome von Äußerungen verbunden sind. Insbesondere betrifft dies eine Auswahl aus verschiedenen möglichen Benutzeräußerungen, die Einstellung verschiedener spezifischer Werte für die unterschiedlichen Sprechparameter durch die Veränderung des Äußerungsstrings oder direkte Manipulation eines Parameterwertes. So ist es z.B. möglich, Wiederholungen oder Pausen in die Äußerung einzufügen oder den Wert der aktuellen Artikulationsgeschwindigkeit zu modifizieren, um verschiedene Anpassungen der Systemeinschätzung für die kognitive Belastung und den Zeitdruck des Benutzers zu testen.

Abbildung 5.8 zeigt einen Schnappschuß der Menüschnittstelle unmittelbar vor einer Äußerungsauswahl. Im oberen Teil sind die zur Auswahl stehenden Äußerungen aufgelistet. Pausen, Wiederholungen, Korrekturen etc. können an den Menü-Schaltknöpfen ausgewählt und in die Äußerung eingefügt werden. Für die aktuell ausgewählte Äußerung und ihre Form werden darunter die aktualisierten Werte der Äußerungsparameter angezeigt, mit denen das Netz der Benutzermodellierungskomponente nach Betätigen der *Abschicken*-Schaltfläche in der aktuellen Zeitscheibe instantiiert werden würde. Veränderliche Parameter sind am unteren Rand des Fensters zu finden und können ebenfalls per Menüauswahl modifiziert werden.

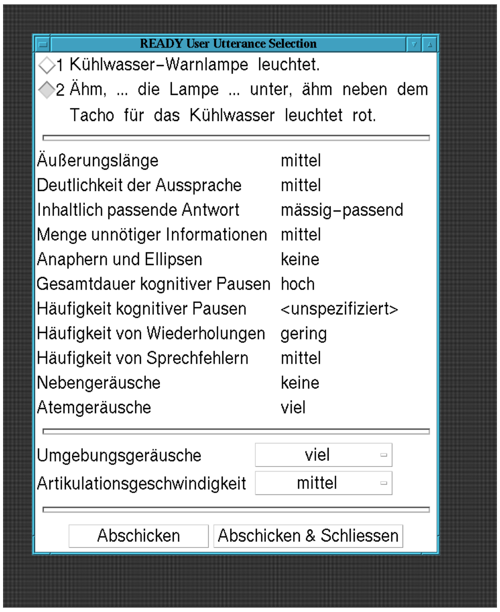


Abbildung 5.8: Bildschirmabzug einer READY-Menüauswahl

In Abbildung 5.9 ist ein Schnappschuß der Dialoggeschichte von READY zu sehen, die die angewählten Benutzeräußerungen mit ihren wichtigsten Merkmalen protokolliert. Die Anpassungen der Systemeinschätzung über die Variablenwerte der Benutzermodellierungskomponen-

[51] Die Menüschnittstelle wurde von Uwe Kern implementiert.

te kann am rechten Rand verfolgt werden: Die internen Variablen für spezifische Ressourcen, die in READY untersucht werden, erhalten einen Wert zwischen 0 und 1000 zugeordnet. Dabei steht 0 für den niedrigsten und 1000 für den höchsten möglichen Wert. Die Dialoggeschichte zeigt ebenfalls die vom System aufgrund des Benutzermodells ausgewählte Antwort auf die Benutzeräußerung bzw. die vorangehende Frage oder Instruktion (siehe Weis, 1999).

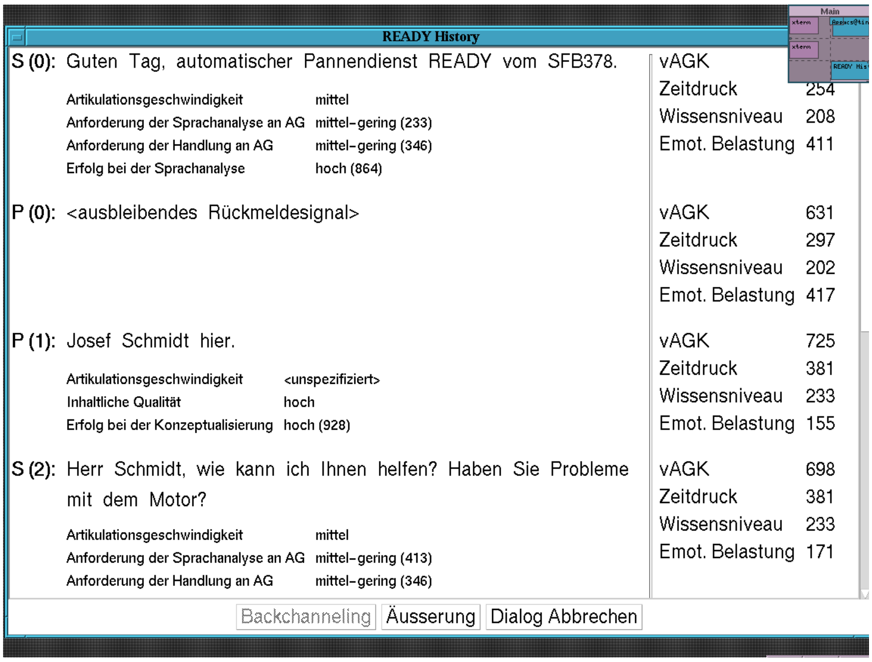


Abbildung 5.9: Bildschirmabzug der READY-Dialoggeschichte mit den Einschätzungen von verfügbarer Arbeitsgedächtniskapazität (vAGK) und Zeitdruck

# Kapitel 6

# Interpretation von Symptomen mit Bayes'schen Netzen

## 6.1 Bayes'sche Netze und Benutzermodellierung

### 6.1.1 Motivation

Bayes'sche Netze werden in den unterschiedlichsten Kontexten eingesetzt, in denen kausale Zusammenhänge zu modellieren sind. Das gilt besonders für Kontexte, in denen Inferenzen unter Unsicherheit erfolgen oder Zusammenhänge durch unterschiedliche Kontextfaktoren beeinflußt werden können. Durch ihren wahrscheinlichkeitsbasierten Ansatz stellen Bayes'sche Netze in diesen Domänen einen geeigneten Mechanismus bereit, der einen polarisierten Wahrheitsbegriff, wie er in der zweiwertigen Logik verwendet wird, vermeidet. Alternative Mechanismen werden in Jameson (1996) beschrieben. Weyrath (1998) begründet die Auswahl der Bayes'schen Netze für das System READY unter diesen verschiedenen Alternativen.

### 6.1.2 Kurzeinführung: Bayes'sche Netze

Bayes'sche Netze sind in zahlreichen Beiträgen ausführlich beschrieben worden. Daher sollen hier nur für das Verständnis wesentliche Aspekte kurz dargestellt und der interessierte Leser auf die entsprechende Literatur verwiesen werden.

#### 6.1.2.1 Relevante Literatur

Eine kurze und verständliche Einführung in den Formalismus der Bayes'schen Netze gibt Charniak (1991). Detailliertere Darstellungen finden sich in Jensen (1996) und Russel und Norvig (1995). Die Ausführungen in Netica User's Guide (1996) sind am praktischen Einsatz orientiert und können in einer Benutzerumgebung getestet werden, die als Testversion im WWW verfügbar ist[1]. Ein sehr gründliches, für den Überblick jedoch weniger geeignetes Standardwerk

[1] Netica wird im System READY für die Verwaltung der Bayes'schen Netze eingesetzt.

ist Pearl (1988). Neapolitan (1990) beschreibt vor allem Lösungsalgorithmen für Bayes'sche Netze.

Verschiedene praktische Anwendungen diskutiert Jameson (1996). Mit konkreten Implementationen Bayes'scher Netze für die Modellierung von Benutzerzielen und -plänen bzw. für Zielsetzungen wie Spracherkennung und Entscheidungsanalyse in Expertensystemen befassen sich Henrion, Breese und Horvitz (1991), Spiegelhalter, Dawid, Lauritzen und Cowell (1993), Heckerman und Horvitz (1998), Horvitz, Breese, Heckerman, Hovel und Rommelse (1998), Zweig und Russell (1998). Einige Anwendungen präsentieren auch Charniak (1991) und Russel und Norvig (1995).

Schäfer und Weyrath (1996) stellen die in READY verwendeten Dynamischen Bayes'schen Netze vor; Schäfer et al. (1997), Schäfer (1998), Weyrath (1998), und Jameson, Schäfer, Weis, Berthold und Weyrath (in Druck), Berthold und Jameson (1999) gehen auf spezielle Zusammenhänge und ihre Motivation in den Bayes'schen Netzen des Dialogsystems READY ein.

#### 6.1.2.2 Wichtige Aspekte und Begriffe Bayes'scher Netze

**Struktur und Semantik** Bayes'sche Netze stellen kausale Modelle von Ursachen und Folgen für Ereignisse dar. Die Knoten im Netz stehen für geeignete Zusammenfassungen von komplementären Elementarereignissen (atomaren Ereignissen), die z.B. den Zustand eines Objektes oder einer Relation in der Welt beschreiben. Eine solche Zusammenfassung bildet (unter bestimmten Annahmen) ein *vollständiges System von Ereignissen*, das durch eine *Zufallsvariable* bezeichnet werden kann. Die Elementarereignisse des Ereignissystems bilden die *Hypothesen* der Knoten.

Zwischen einzelnen Ereignissystemen können Beziehungen der bedingten Abhängigkeit bestehen, d.h. die Veränderung einer Variablen zieht eine Veränderung einer anderen Variablen nach sich. Knoten, deren Variablen miteinander in einer solchen wahrscheinlichkeitstheoretischen Abhängigkeit stehen, werden durch gerichtete Kanten im Netz verbunden. Eine Kante zwischen zwei Knoten verdeutlicht, daß das Verhältnis der zugehörigen Variablen durch *bedingte Wahrscheinlichkeiten* beschrieben werden kann, die auch die Stärke des Zusammenhangs angeben. Die Stärke der Bayes'schen Netze liegt vor allem darin, daß durch die Lokalität bei der Definition dieser quantitativen Zusammenhänge globale Verbindungen spezifiziert und konsistent aufgebaut werden können.

Ein Bayes'sches Netz definiert eine Wahrscheinlichkeitsverteilung auf den atomaren Ereignissen all seiner Variablen, und beschreibt damit die Wahrscheinlichkeiten, die ein Agent den unterschiedlichen Ereignissen in der modellierten Domäne zuordnet. Durch Beobachtungen, sogenannte *Evidenzen*, können Wahrscheinlichkeiten für das Zutreffen einzelner Elementarereignisse in der Welt direkt erfaßt und den zugehörigen Hypothesen im Netz zugeordnet werden.

Die Wahrscheinlichkeitsverteilung des Netzes muß zu jedem Zeitpunkt gewissen Konsistenzbedingungen genügen. Evidenzen, die zur *Instantiierung* ihrer zugehörigen Knoten führen, verursachen eine lokale Veränderung der Wahrscheinlichkeitsverteilung des Gesamtnetzes und verletzen dadurch dessen Konsistenzbedingungen. Daher werden beim *Lösen* eines Bayes'schen Netzes die Einschätzungen für die Hypothesen aller nichtinstantiierten Variablen so angepaßt, daß die Konsistenzbedingungen wieder erfüllt sind. Grundlage für die Spezifikation der Zusammenhänge und die Aufrechterhaltung der Konsistenz ist die namensgebende Bayes'sche Regel 6.1. Sie erlaubt die Berechnung einer bedingten Wahrscheinlichkeit aus einer Anfangswahrscheinlichkeit und vorgegebenen Kontextbedingungen.[2]

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} \tag{6.1}$$

**Verwendung** Trotz Ausrichtung der Beziehungen im Bayes'schen Netz erlaubt die Anwendung der Bayes'schen Regel zwei unterschiedliche Sichtweisen auf die lokalen Zusammenhänge:

- Die Wahrscheinlichkeitsverteilung für die Hypothesen eines Knotens liefert über die bedingten Wahrscheinlichkeiten seiner Kinder *Vorhersagen* über deren Wahrscheinlichkeitsverteilungen (Abwärtspropagierung). In dieser Sicht werden Bayes'sche Netze für *prädiktive Inferenz* herangezogen.
- Bei einer Instantiierung von Kindknoten muß die Wahrscheinlichkeitsverteilung der Elternknoten angepaßt werden (Aufwärtspropagierung). Diese Sicht versucht die Ursachen für das Eintreten eines atomaren Ereignisses zu erklären und wird daher als *diagnostische Inferenz* bezeichnet.

Die Wahrscheinlichkeitsverteilung auf den Hypothesen einer Variablen wird in der englischsprachigen Literatur *Belief* genannt; im folgenden wird der deutsche Begriff *Einschätzung* dafür verwendet. Für Variablen mit numerischen Hypothesen kann die Wahrscheinlichkeitsverteilung durch eine gewichtete Summe beschrieben werden, die *Erwartungswert* genannt wird (siehe Anhang B.3).

**CPT** Die bedingten Wahrscheinlichkeiten in einem Bayes'schen Netz beschreiben, wie stark die Veränderung der Einschätzung einer Ursache (Elternknoten) die Veränderung der Einschätzung ihrer möglichen Folgen oder Symptome (Kindknoten) beeinflußt (prädiktiv) bzw. wie stark sich die Wahrscheinlichkeitseinschätzung für Erklärungen einer bestimmten Evidenz durch ihre Eltern anpassen muß (diagnostisch). Da für eine einzelne Variable wegen der Komplementarität ihrer Elementarereignisse die Veränderung des Wertes für eine Hypothese zwangsweise die Veränderung der Werte für alle anderen Hypothesen nach sich zieht, werden die bedingten Wahrscheinlichkeiten ebenfalls als Wahrscheinlichkeitsverteilungen angegeben. Zur Darstellung dienen Tabellen, die *conditional probability table* genannt werden; abgekürzt[3] *CPT*.

In den Beschreibungen der unterschiedlichen Knoten des Sprachproduktionsmodells in READY im Anhang B sind Angaben zur Berechnung der jeweiligen CPT enthalten. Eine kurze Einführung in die Darstellung und Berechnung von CPTs gibt dort der Abschnitt B.3.

**Notation** Zur Verkürzung der Darstellung kommen im folgenden einige Abkürzungen zur Anwendung:

[2]Für eine detailliertere Diskussion siehe z.B. Pearl (1988), Kap. 2.

[3]Diese Kurzbezeichnung wird auch im folgenden verwendet.

(6.2) $P(K = k) =_{def} P(K)$
als Wahrscheinlichkeitsfunktion, die eine Wahrscheinlichkeitsverteilung auf den atomaren Ereignissen der Variable K definiert,

(6.3) $P(K = k \mid A = a, B = b) =_{def} P(K \mid A, B)$
für (bedingte) Wahrscheinlichkeitsfunktionen mit dem Definitionsbereich $< K, A, B >$,

(6.4) $P(K = k \mid A = a_i, B = b_j) =_{def} P(K \mid A = a_i, B = b_j)$
für Instanzen der bedingten Wahrscheinlichkeitsfunktion $P(K \mid A, B)$ an der Stelle $< a_i, b_j >$, d.h. eine Wahrscheinlichkeitsfunktion mit dem Definitionsbereich $K$

### 6.1.3 Modellierung sprachlicher Symptome im Dialogsystem READY

In der ersten Phase dieser Arbeit wurden die relevanten Symptome für Belastung und Zeitdruck aus der Literatur (Kapitel 3) identifiziert, in Zusammenhang mit einem Modell der Sprachproduktion gebracht und unter Beachtung einer ersten Grobauswertung der Dialoge aus der in Kapitel 4.3 dargestellten Feldstudie daraus ein qualitatives Modell abgeleitet und implementiert. Während der Implementation stellte sich vor allem immer wieder eines der zentralen Probleme beim Entwurf Bayes'scher Netze ein: Wie können konkrete quantitative Werte für die Konstruktion von CPTs abgeleitet werden (siehe auch Pearl, 1988; Jensen, 1996)? Nach Abschluß der Implementation wurde daher eine Reanalyse des Korpus (4.3) als notwendig erachtet und eine detailliertere Auswertung vorgenommen. Zusammen mit konkreten Meßwerten aus den Arbeiten aus Kapitel 3 sind damit ausreichende empirische Daten für die Anpassung der *quantitativen* Seite der implementierten Netze verfügbar geworden.

Es stellte sich heraus, daß die Architektur des stark am Sprachproduktionsmodell orientierten Ausgangsnetzes für die Spezifizierung der quantitativen Zusammenhänge nicht ohne weiteres geeignet ist. Daher wurden mit den quantitativen Veränderungen auch einige qualitative Aspekte der implementierten Netze erneut zur Disposition gestellt und Anpassungen vorgenommen, die sich aus dem neuen Stand der theoretischen und praktischen Erkenntnisse ergeben (siehe Berthold & Jameson, 1999). Zahlreiche Aspekte der Erstimplementation bleiben von den Veränderungen jedoch unberührt.

Abschnitt 6.3 skizziert die Ergebnisse der neuen Überlegungen und Verbesserungen des quantitativen Entwurfs der implementierten Netze. Die Überprüfung und Verfeinerung dieser Anpassungen erfolgte noch, während diese Arbeit bereits verfaßt wurde. Sie sind daher Bestandteil der aktuellen Forschungsaufgaben im Projekt READY. Eine Übernahme der Veränderungen in den implementierten Prototypen bedarf durch die enge Verbindung der Benutzermodellierung zum Dialogmanagement großer Sorgfalt und ist daher zum jetzigen Zeitpunkt nicht abgeschlossen. Die Daten dieses Modells finden sich im Anhang B.5.

An charakteristischen Beispielen soll im folgenden gezeigt werden,

- wie die identifizierten *qualitativen* Zusammenhänge aus empirischer Forschung und Sprachproduktionsmodell in ein Bayes'sches Netz des READY-Systems umgesetzt werden konnten (Abschnitt 6.2) und

- wie aus Referenzwerten der Feldstudie und Werten der empirischen Forschung quantitative Zusammenhänge für die zugrundeliegenden bedingten Wahrscheinlichkeiten angenähert werden können (Abschnitt 6.3).

In beiden Abschnitten wird zunächst ein Beispiel für eine einfache und direkte Modellierung dargestellt. Anschließend werden komplexere Zusammenhänge und Möglichkeiten zur Modellierung spezieller Formen der Abhängigkeit zwischen Parametern in Bayes'schen Netzen beschrieben.[4] Die dargestellten Modellierungskonzeptionen stehen stellvertretend für weitere, ähnlich zu behandelnde Symptome, deren Daten aus Platz- und Verständlichkeitsgründen jedoch lediglich im Anhang B.5 aufgelistet sind.

Da das System READY ein komplexes Softwaresystem darstellt, in dem neben der Modellierung sprachlicher Symptome weitere Mechanismen zur Erkennung von kognitiver Belastung und Zeitdruck und vor allem zur Anpassung des Systemverhaltens daran implementiert sind, werden spezielle Eigenschaften des Systems erst dort erläutert, wo sie für das Verständnis notwendig sind. Als weiterführende Literatur zum System READY und seiner Entwicklung können Wahlster, Jameson, Ndiaye, Schäfer und Weis (1995a), Schäfer und Weyrath (1996), Schäfer et al. (1997) und Jameson et al. (in Druck) sowie Jameson (1998) herangezogen werden.

## 6.2 Ein qualitatives Modell für die Zusammenhänge von Belastung, Zeitdruck und Sprachproduktion

### 6.2.1 Allgemeine Aspekte der Netzarchitektur

#### 6.2.1.1 Allokationsmodell und Zeitdruck

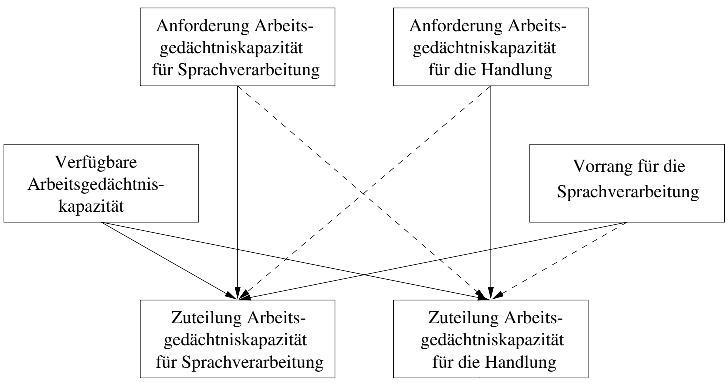


Abbildung 6.1: Das READY-Allokationsmodell für verfügbare Arbeitsgedächtniskapazität (Schäfer, 1998)

Abbildung 6.1 zeigt das Allokationsmodell der Benutzermodellierungskomponente in READY. Die Zuteilung von kognitiven Ressourcen wird nach diesem Modell für eine *Primäraufgabe*

[4]Die Reihenfolge der Darstellung ist daher hier nicht an der Zuordnung einzelner Symptome zu linguistischen Ebenen orientiert.

und eine *Nebenaufgabe* vorgenommen. Die drei bestimmenden Parameter dafür sind

- die Anforderungen der zugehörigen kognitiven Prozesse an kognitive Kapazität,
- die Präferenz des Benutzers bezüglich Zuteilung von Ressourcen an eine der beiden Aufgaben[5] und
- die verfügbare kognitive Kapazität (Arbeitsgedächtniskapazität = vAGK) des Benutzers.

Das Modell ist in Kapitel 4 von Schäfer (1998) ausführlich beschrieben. Die Einschätzung der kognitiven Belastung des Benutzers ergibt sich aus der Wahrscheinlichkeitsverteilung für den Knoten VERFÜGBARE ARBEITSGEDÄCHTNISKAPAZITÄT. Der Einfluß des Zeitdrucks wird in READY ebenso durch eine Variable erfaßt und seine aktuelle Einschätzung ebenfalls durch deren Wahrscheinlichkeitsverteilung beschrieben.

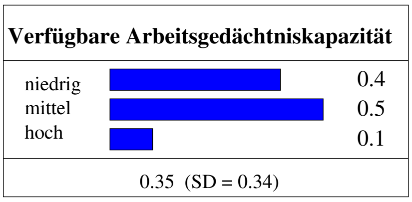


Abbildung 6.2: Beispiel für die Einschätzung der verfügbaren Arbeitsgedächtniskapazität

Abbildung 6.2 zeigt die Konfiguration des Knotens VERFÜGBARE ARBEITSGEDÄCHTNISKAPAZITÄT für die entsprechende diskretisierte Variable. Dabei liegt eine Wahrscheinlichkeitsverteilung vor, die die Kapazität eher als niedrig beschreibt. Derartige Einschätzungen für die Benutzerressourcen *verfügbare Arbeitsgedächtniskapazität* und *verfügbare Zeit* stellen die Basisparameter für das ressourcenadaptive Dialogverhalten des Systems dar (siehe Weis, 1999).

#### 6.2.1.2 Differenzierte Abhängigkeiten: Erfolg von Einzelphasen der Sprachproduktion

Kognitive Belastung und Zeitdruck sind im gewählten Sprachproduktionsmodell (Levelt, 1989) und unter der Sicht auf multiple kognitive Ressourcen (Wickens, 1992; Neumann, 1992)[6] keine direkten Ursachen für die untersuchten Phänomene. Vielmehr werden durch kognitive Belastung und Zeitdruck bestimmte kognitive Prozesse in ihrer Verarbeitungsfähigkeit beeinträchtigt und die *Beeinträchtigung* dieser Prozesse führt zum Entstehen der herausgearbeiteten Symptome für Belastung und Zeitdruck. Die Phasen der Sprachproduktion sind von dieser Beeinträchtigung in unterschiedlichem Maße betroffen. Entsprechend verschieden ist auch der diagnostische Wert der Symptome, die den verschiedenen Phasen der Sprachproduktion zuzuordnen sind.

Für den Entwurf der Bayes'schen Netze ergibt sich daraus, daß nicht alle Symptome als unabhängig[7] voneinander betrachtet werden können unter der Vorbedingung einer bestimmten

[5]Diese Präferenz kann durch zahlreiche Faktoren bestimmt werden, die in READY in einer Variable zusammengefaßt werden.

[6]Siehe auch Kapitel 2.

[7]Bedingte Unabhängigkeit (*conditional independence*) ist eines der Schlüsselkonzepte, um in Bayes'schen Netzen einzuschätzen, welche Variablen unter den aktuellen Evidenzen Auswirkungen aufeinander haben.

Beeinträchtigung der zugehörigen Prozesse. Insbesondere ist das der Fall für Gruppen von Symptomen, die auf die Beeinträchtigung der gleichen Sprachproduktionsphase zurückzuführen sind. Wenn z.B. die Phase der Konzeptualisierung stark gestört wird, erhöht sich die Wahrscheinlichkeit für das Auftreten *aller* Symptome aus dieser Phase. Um diese Abhängigkeit zu berücksichtigen, wird die Einführung sogenannter *Zwischenvariablen*[8] im Netz notwendig (siehe Jensen, 1996). *Zwischenvariablen* moderieren den Einfluß von *Informationsvariablen*, den Symptomen, auf die *nicht beobachtbaren Variablen* KOGNITIVE BELASTUNG (vAGK) und ZEITDRUCK.

Die Differenzierung der verschiedenen Sprachproduktionsphasen in bezug auf den Einfluß kognitiver Belastung und Zeitdruck führt daher zu einer Anzahl von Knoten, die zu einer Abstraktionsschicht zusammengefaßt werden können. Die Knoten dieser Schicht beschreiben die Beeinträchtigung kognitiver Prozesse durch Belastung und Zeitdruck, ohne Vorhersagen über konkrete Symptome zu treffen und abstrahieren damit z.B. von individuell unterschiedlichen Auswirkungen. Die Vorhersage über alle zu einer Teilphase gehörigen Symptome hängt damit direkt nur von der Einschätzung des Systems über die Leistungsfähigkeit in der betreffenden Teilphase ab (prädiktive Sicht). Umgekehrt werden die Systemeinschätzungen über die kognitive Belastung und den Zeitdruck des Benutzers nicht mehr durch Aggregierung des Einflusses unabhängig voneinander verursachter Einzelsymptome übermäßig stark angepaßt (diagnostische Sicht). Ein zu starker Einfluß von einzelnen, jedoch unter gegebener Beeinträchtigung einer Sprachproduktionsphase korrellierten Symptomen auf die Anpassung der Systemeinschätzungen wird dadurch vermieden.

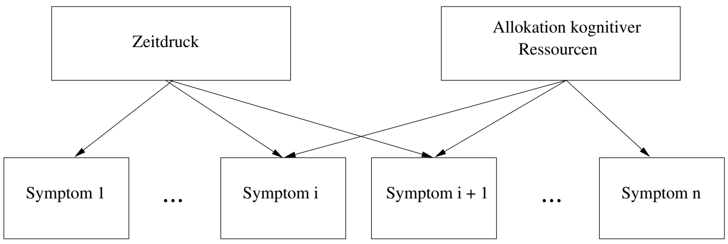


Abbildung 6.3: Einfache Modellierung des Zusammenhanges von Symptomen mit Belastung und Zeitdruck

Resultat einer Modellierung nach Abbildung 6.3 wäre z.B. eine zu starke Anpassung der Einschätzung für die ALLOKATION KOGNITIVER RESSOURCEN und den ZEITDRUCK unter scheinbar unabhängigen Evidenzen für die einzelnen Informationsvariablen.

(6.5) Zum Beispiel sind die Erhöhung des Pausenanteils und die Verringerung der Artikulationsgeschwindigkeit abhängig voneinander *unter der Vorbedingung verminderter Flüssigkeit der Sprachproduktion*, da sie zwei Effekte der gleichen Ursache darstellen. Die Wahrscheinlichkeit für das Auftreten von Fragmenten und Fehlansätzen steht dagegen nicht in direktem Zusammenhang mit der Häufigkeit von kognitiven Pausen, sondern ist lediglich davon abhängig, ob neben einer Verringerung der Geschwindigkeit der Sprachproduktion die Konzeptualisierungsphase beeinträchtigt wird.

Wie bereits mehrfach angesprochen wurde, verändert sich die Sprachproduktion unter Belastung und Zeitdruck sowohl in ihrer Qualität als auch in ihrer Geschwindigkeit. Die folgenden

[8]Die hier verwendete Terminologie geht auf Jensen (1996) zurück. Zwischenvariablen heißen dort *mediating variables*.

Zwischenvariablen werden den Sprachproduktionsphasen und ihrer jeweiligen Klasse zugeordnet (vgl. Abb 6.4):

ERFOLG BEI DER KONZEPTUALISIERUNG (Qualität) beschreibt die Einschätzung, inwiefern der Benutzer in der Lage ist, eine gute Konzeptualisierungsarbeit für die Sprachproduktion auszuführen, d.h. passende Information zu selektieren, zu linearisieren und die korrekten Voreinstellungen für die Formulierungsphase vorzunehmen. Erfolg ist ein relatives Maß, das von den Anforderungen der Dialogsituation und den Zielen des Sprechers beeinflußt wird.

ERFOLG BEI FORMULIERUNG UND ARTIKULATION (Qualität) beschreibt die qualitative Funktion der beiden automatisierten Teilphasen der Sprachproduktion. Der Erfolg der Formulierung und Artikulation ist vor allem als Nähe der Sprecheräußerungen zur idealen Sprachproduktion (Clark & Clark, 1977) zu verstehen.

AUSMASS DES MONITORINGS (Qualität) beschreibt die Einschätzung, in welchen Maße der Sprecher während der Sprachproduktion seine Äußerungen auf linguistische Wohlgeformtheit und Angemessenheit im Kontext überprüft.

FLÜSSIGKEIT DER SPRACHPRODUKTION (Geschwindigkeit) beschreibt die Einschätzung, inwiefern die inkrementelle Verarbeitung innerhalb des Sprachproduktionsmodells und ihr zeitlicher Verlauf reibungslos funktionieren. Die Qualität der inkrementellen Verarbeitung zeigt u.a. an, ob allen Teilphasen des Sprachproduktionsmodells stets ausreichend Eingabeinkremente zur Verfügung stehen, auf denen die Prozesse der Teilphasen ihre Verarbeitung autonom durchführen können. Schlechte inkrementelle Verarbeitung führt zur Entstehung von Verzögerungen.

Die Phasen der Formulierung und der Artikulation können in einem Knoten zusammengefaßt werden, weil die Prozesse beider Teilphasen als weitgehend automatisiert gelten (siehe u.a. Levelt, 1989). Die Vereinfachung beruht außerdem auf der Beobachtung, daß Symptome aus beiden Phasen sich als wenig relevant für die Einschätzung von Zeitdruck und Belastung erwiesen haben (siehe Kapitel 3). Die CPTs der Erfolgsknoten müssen daher berücksichtigen, daß gemäß den Erkenntnissen aus den Kapiteln 3 und 4 der Einfluß von Belastung und Zeitdruck auf den ERFOLG BEI DER KONZEPTUALISIERUNG, das AUSMASS DES MONITORINGS und die FLÜSSIGKEIT DER SPRACHPRODUKTION recht stark ist, während er für den ERFOLG BEI FORMULIERUNG UND ARTIKULATION nur geringe Bedeutung hat.

In der Grafik 6.4 sind Zwischenvariablen als fett umrandete Knoten, Informationsvariablen als gestrichelte Knoten und nicht beobachtbare Variablen (*not observable*) als normal umrandete Knoten dargestellt.[9] Der Knoten ZUTEILUNG KOGNITIVER RESSOURCEN steht stellvertretend für das Allokationsmodell aus Abbildung 6.1.

Auch für den Zeitdruck können unterschiedlich starke Auswirkungen auf die einzelnen Phasen der Sprachproduktion angenommen werden. Dieser Einfluß konnte aufgrund der bisherigen empirischen, experimentellen und theoretischen Forschung, die in Kapitel 3 dargestellt wurde,

[9] Die Terminologie der Knotenbezeichnungen stammt von Jensen (1996).

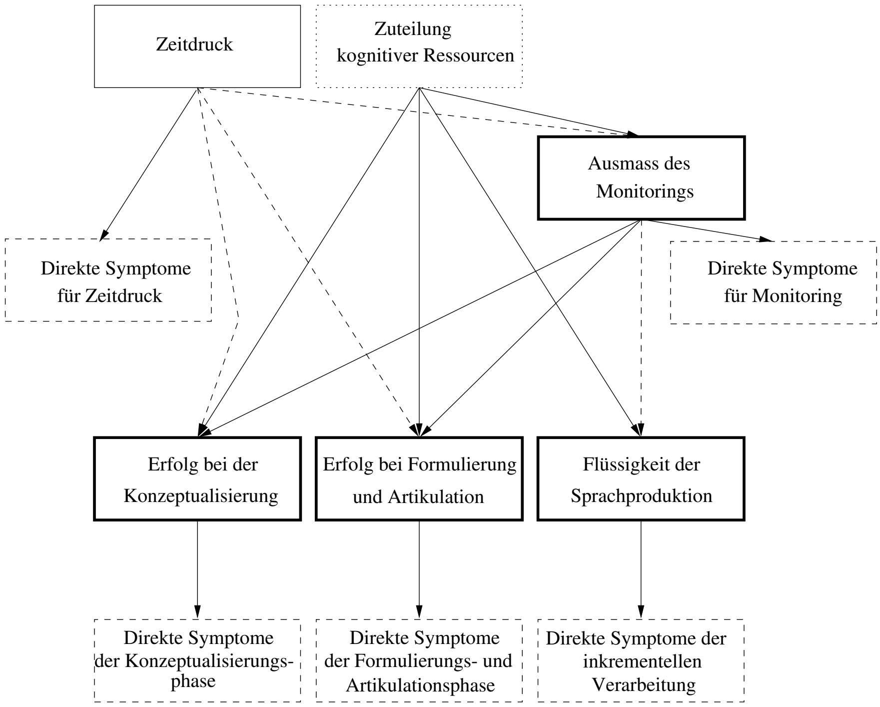


Abbildung 6.4: Abstraktion durch Einführung der Erfolgsknoten

jedoch weitaus weniger differenziert werden als für kognitive Belastung. Einige der sprachlichen Symptome für Zeitdruck werden daher als direkte Symptome des Zeitdrucks behandelt.

Eine Verbindung zwischen dem ZEITDRUCK und der FLÜSSIGKEIT DER SPRACHPRODUKTION fehlt im vorliegenden Modell 6.4, da keinerlei experimentelle Evidenz dafür vorliegt, daß die Geschwindigkeit der inkrementellen Verarbeitung durch Zeitdruck beschleunigt wird. In Abschnitt 6.2.3.1 wird darauf eingegangen, wie einige zusätzliche Interpolationen möglich sind, die die experimentell noch nicht untersuchten Zusammenhänge betreffen.

### 6.2.2 Beschreibung von CPTs

**Voraussetzung: Diskretisierung** Bei den meisten Sprechparametern handelt es sich um kontinuierliche Variablen, d.h. Variablen, die innerhalb eines bestimmten Intervalls beliebige reelle Werte annehmen können. Für die Verwendung in Bayes'schen Netzen ist eine Diskretisierung dieser kontinuierlichen Variablen auf Teilintervalle notwendig, die jeweils durch eine Variablenhypothese repräsentiert werden (siehe Anhang B.3).

**Funktionen statt Matrizen** Die Größe der CPT (Anzahl der Zellen) für einen Knoten wächst exponentiell mit der Zahl seiner Elternknoten und der maximalen Hypothesenzahl.[10] Es ist kaum möglich, konkrete Einträge für alle Zellen ohne eine umfangreiche Falldatenbasis oder anderweitig empirisch erhobene Werte anzugeben. Meist hängen die Werte der Wahrscheinlichkeitsverteilungen für die Hypothesen eines Knotens K unter jeweils einer bestimmten Elternkonfiguration jedoch auf eine spezifische Weise zusammen, die durch eine *Verteilungsfunktion* beschrieben werden kann. Um die Wahrscheinlichkeiten für *eine* Hypothese unter *verschiedenen* Elternkonfigurationen zu beschreiben, können ebenfalls Funktionen angewendet werden.[11] Durch Beschränkungen in Form von Constraintssystemen lassen sie sich so spezifisch wie nötig und so differenziert wie möglich definieren. Ein genereller Constraint für die CPT eines Knotens H mit k Hypothesen $h_j$ und einem Elternknoten A ist 6.6.

$$\sum_{j=1}^{k} P(H = h_j \mid A = a_i) = 1 \tag{6.6}$$

Spezifische Funktionsparameter für die Verteilungsfunktionen können den quantitativen Einfluß der Elternkonfiguration beschreiben. Dieser Einfluß der Elternhypothesen wird in einer Einflußfunktion zusammengefaßt, die einen Erwartungswert[12] für den Kindknoten zur Verfügung stellt. Durch die Instantiierung der Verteilungsfunktionen mit diesem Erwartungswert werden dann die konkreten Wahrscheinlichkeiten für die einzelnen Hypothesen des Kindes berechnet.

#### 6.2.2.1 Vorberechnung des Einflusses von Elternknoten

**Voraussetzung: Gleiche Wertebereiche** Der Erwartungswert eines Kindknotens in der Maßeinheit dieses Knotens unter einer bestimmten Elternkonfiguration kann durch eine Rechenvorschrift auf den Maßeinheiten der Elternhypothesen berechnet werden. Die Variablen der Erfolgsknoten und des Allokationsmodells sowie des Knotens ZEITDRUCK besitzen jedoch z.B. keine konkrete Maßeinheit, außerdem sind oft verschiedene Einheiten miteinander zu verrechnen. Zur Vereinfachung der Berechnung werden daher alle Hypothesen in READY auf Werte aus dem Normintervall [0;1] abgebildet[13]. Für einen Knoten lassen sich alle Hypothesen dann auf einer Skala von 0 bis 1 anordnen, wobei 0 meist den niedrigsten Wert für das Attribut darstellt, das durch die Knotenbezeichnung ausgedrückt wird und 1 den höchsten Wert (z.B.: ERFOLG DER KONZEPTUALISIERUNG = niedrig/hoch).

**Einflußfunktionen** Der Einfluß der Eltern auf einen Kindknoten H kann dann durch eine Funktion $I_H$ definiert werden. $I_H$ hat die Arität $j$, wobei $j$ die Anzahl der Elternknoten ist. Läßt sich z.B. die Wahrscheinlichkeitsverteilung über den Hypothesen eines Kindknotens angeben

[10]Für einen Knoten mit 5 Hypothesen und 3 Eltern mit ebenfalls je 5 Hypothesen besitzt eine CPT beispielsweise 625 Zellen.

[11]Funktionen sind auch bei der Anpassung der Implementation komfortabler als explizite CPTs, z.B. beim Test verschiedener Diskretisierungen und Diskretisierungsgranularitäten.

[12]Dieser Erwartungswert entspricht in einigen Fällen dem wahrscheinlichkeitstheoretischen Erwartungswert, jedoch wird der Begriff hier loser verwendet.

[13]Für Details dieser Abbildung siehe Anhang B.1.

als Normalverteilung[14] mit dem Mittelwert aus den Hypothesen zweier Elternknoten A und B, kann die Wahrscheinlichkeitsverteilung über den Hypothesen eines Knotens H unter einer bestimmten Konfiguration seiner Eltern durch die Formeln 6.7 und 6.8 berechnet werden.

$$I_H(a,b) = \frac{a+b}{2} \tag{6.7}$$

$$P(H = h_k \mid A = a_i, B = b_j) = \mathcal{N}_{I_H(a_i,b_j),\sigma}(h_k) \tag{6.8}$$

#### 6.2.2.2 Verteilungsfunktionen

**Normalverteilung** Normalverteilungsfunktionen sind auch unter dem Namen *Gauß-Funktionen* bekannt.[15] Sie erzeugen Graphen mit der Form der Gaußschen Glockenkurve (Abbildung 6.5) und ergeben sich oft als Beschreibung der Verteilung von Werten bestimmter Eigenschaften in einer Population. Z.B. ist die Körpergröße von Menschen eine auf die Bevölkerung eines Staates weitgehend normalverteilte Eigenschaft. Auch die Verteilung für die Ausprägung bestimmter Sprechparameter, wie z.B. der Artikulationsgeschwindigkeit über einen Dialog hinweg kann als Normalverteilung beschrieben werden, wie am Vergleich von Abbildung 6.5 und Abbildung 4.5 zu erkennen ist.

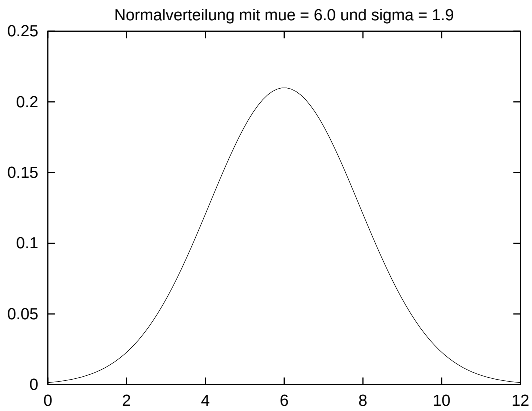


Abbildung 6.5: Darstellung einer typischen Normalverteilungskurve

Neben dem Argument x werden Normalverteilungen mit zwei weiteren Parametern instantiiert: $\mu$ gibt den Erwartungswert der Verteilung an, $\sigma$ ihre Streuung. Der Streuungsparameter $\sigma$ kann als Maß für die Unsicherheit gesehen werden, mit der der Erwartungswert $\mu$ unter der vorgegebenen Elternkonfiguration eintreten wird. (Der Erwartungswert $\mu$ kann z.B. mithilfe der zum Knoten gehörigen Einflußfunktion $I_H$ berechnet werden.) Im folgenden wird als Kurzdarstellung die Form $\mathcal{N}_{\mu,\sigma}(x)$ verwendet.

Normalverteilungsfunktionen sind jedoch nicht für diskrete, sondern für kontinuierliche Zufallsgrößen definiert. Da für alle Wahrscheinlichkeitsverteilungen die Bedingung 6.6 gelten

[14] Zum Begriff 'Normalverteilung' siehe Abschnitt 6.2.2.2.

[15] Eine Diskussion des Wertes von Normalverteilungsfunktionen in der empirischen psychologischen Forschung gibt z.B. Robson (1973).

muß, sind die Funktionswerte von Normalverteilungen nach ihrer *Diskretisierung* einer *Normalisierung* zu unterziehen. Ein Beispiel wird in Abschnitt 6.2.3 angegeben.

**Zweipunktverteilung** Für *binäre* Knoten H mit zwei Hypothesen $h_1$ und $h_2$ lassen sich die bedingten Wahrscheinlichkeiten vollständig durch die Angabe des Wertes für eine der beiden Hypothesen beschreiben, da für eine Elternkonfiguration von Knoten $A_1, ...A_k$ Gleichung 6.9 als Spezialfall von Gleichung 6.6 gilt. Da die hier verwendeten Werte für $h_i$ der Menge $\{0{,}1\}$ entsprechen, wird die Zweipunktverteilung auch als Null-Eins-Verteilung bezeichnet.

$$P(H = h_2 \mid A_1 = a_1, ..., A_k = a_k) = 1 - P(H = h_1 \mid A_1 = a_1, ..., A_k = a_k) \tag{6.9}$$

**Weitere Funktionen** Prinzipiell sind beliebige Funktionen in x (und weiteren Parametern) als Verteilungsfunktionen denkbar, wenn durch eine Normierung Constraint 6.6 sichergestellt wird. Verschiedene Funktionen sind jedoch wie die Normalverteilung durch ihren empirischen Wert dafür besonders geeignet. Zu diesen Funktionen zählen u.a. die Poisson-Verteilung, die hypergeometrische Verteilung, die Polya-Verteilung und die Binomialverteilung. Diese Funktionen kommen im vorgestellten Modell derzeit nicht zum Einsatz, weshalb sie hier aus Platzgründen nicht näher vorgestellt werden sollen. Für nicht normalverteilte Zufallsgrößen wie z.B. die Einsatzlatenz (siehe Abb. 4.13) werden sie jedoch bei der weiteren Spezifizierung des detaillierten *quantitativen* Modells (siehe Abschnitt 6.3) von Bedeutung sein.

## 6.2.3 Eindeutige Zusammenhänge zwischen Symptom und Ursache

### 6.2.3.1 Qualitative Struktur

**Ursachen, Folgen und die resultierende Netzarchitektur**

In den Kapiteln 3, 4 und 5 wurde darauf hingewiesen, daß es mit formalen Mitteln schwierig ist, Fehlansätze und Fragmente eindeutig voneinander zu trennen, wenn die Äußerung nach dem Fragment nicht endet. Bei der Auswertung der Feldstudie wurden die Werte für Fragmente und Fehlansätze gemeinsam dargestellt. Die beiden unterschiedlichen Phänomene werden auch für die Modellierung zusammengefaßt.

Fehlansätze wurden in Kapitel 3 mit Problemen der Linearisierung innerhalb der Teilphase der Konzeptualisierung oder wechselnder Kontextinformation während des Sprechens erklärt. Auch Satzfragmente lassen sich auf Probleme der Informationsauswahl in der Phase der Konzeptualisierung zurückführen.

READY's Zeitscheibenmodell[16] führt die vereinfachende Annahme ein, daß sich der Dialogkontext innerhalb einer Zeitscheibe nicht verändert. GEÄNDERTER KONTEXT als Ursache für Fehlansätze entfällt daher. Probleme der Konzeptualisierung verbleiben als einzige Erklärung für Fehlansätze und Fragmente im vorgegebenen Sprachproduktionsmodell. Zusammen mit den Annahmen aus dem Abschnitt 6.2.1.2 führt dieser direkte Zusammenhang zu dem einfachen Bayes'schen Teilnetz aus 2 Knoten, wie es in Abbildung 6.6 angegeben ist.

[16] Siehe Abschnitt 6.2.6.2.

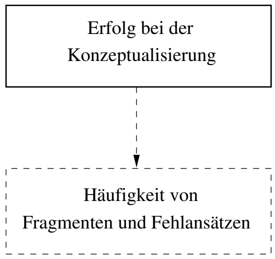


Abbildung 6.6: Zusammenhang eines Symptoms mit einer einzelnen Ursache

Wodurch die Probleme bei der Konzeptualisierung entstanden sind, ist hier irrelevant, da dieser Zusammenhang durch die CPT des Knotens ERFOLG BEI DER KONZEPTUALISIERUNG erfaßt wird. Da der Erfolg der Konzeptualisierung von mehreren Ursachen abhängt, ist sichergestellt, daß nicht jeder Fehlansatz durch kognitive Belastung erklärt wird. Wichtig ist aber, daß folgendes gilt: Solange Erfolgsknoten im Bayes'schen Netz nicht direkt instantiiert werden, ist der Vorhersagewert eines Symptoms immer dann vom Grad der kognitiven Belastung bzw. des Zeitdrucks abhängig, wenn ein gerichteter Pfad vom entsprechenden Knoten zum Symptom existiert. Das wird jedoch immer der Fall sein, da das System nie direkte Evidenz für den Erfolg der Konzeptualisierungsphase erlangen kann.

**Implizite Annahmen der Modellierung**

Für den Zusammenhang von Fehlansätzen und Fragmenten mit Zeitdruck liegen keine Untersuchungen vor. Für einige andere Symptome, die auf Probleme in der Konzeptualisierungsphase zurückzuführen sind, konnte ein Effekt von Zeitdruck jedoch festgestellt werden: Zeitdruck mindert den ERFOLG DER KONZEPTUALISIERUNG. Eine entsprechende Kante zwischen diesen Knoten wurde daher in der Netzarchitektur von READY eingeführt. Durch die Zwischenvariablen und die Klassifizierung von Symptomen nach Zugehörigkeit zu den unterschiedlichen Sprachproduktionsphasen ergeben sich daher Vorhersagen über *mögliche* Versuchsergebnisse für die noch nicht untersuchten Zusammenhänge. Dieses Vorgehen entspricht der Hypothesenbildung im Vorfeld experimenteller Studien, die sich ausschließlich auf bisherige Arbeiten und Schlußfolgerungen aus daraus gebildeten Theorien stützt.[17]

Da die so entstandenen Verbindungen hypothetisch sind, können sie sich auch experimentell als unzutreffend herausstellen. Läßt sich eine Wirkung von Belastung oder Zeitdruck auf einen Parameter nicht experimentell nachweisen, kann dafür z.B. ein unterschiedlicher Einfluß der beiden Größen auf die einzelnen Subprozesse verantwortlich sein. Eine differenziertere Netzarchitektur kann sich an diese Erkenntnisse anpassen: So lassen sich die Prozesse der Konzeptualisierung z.B. einerseits in Prozesse der *Makroplanung* und *Mikroplanung* unterteilen, andererseits bestehen beide aus Teilphasen der *Informationsselektion* und der *Informationslinearisierung*. Durch die Neuordnung der einzelnen Symptome nach *Subprozessen* können die

[17]Die Bestätigung der jeweiligen Hypothesen ist nicht Inhalt des Projektes READY. Ausgewählte Schlußfolgerungen werden im Rahmen des Projekts READY jedoch derzeit experimentell überprüft.

Vorhersagen über die Wirkung der beiden Einflußgrößen präzisiert werden. Abbildung 6.7 verdeutlicht diese Erweiterbarkeit des gewählten Ansatzes und seine Adaptierbarkeit an neue Forschungsergebnisse.

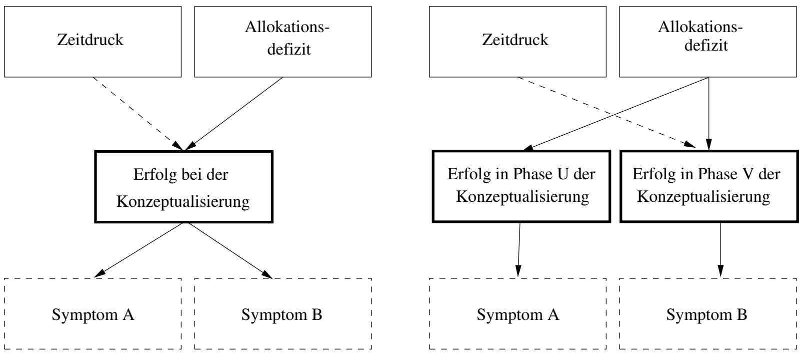


Abbildung 6.7: Vorhersagen und ihre hypothetische Präzisierung aufgrund neuer experimenteller Ergebnisse

(6.10) Beispiel zu Abbildung 6.7: Für die Symptome A und B sei der Einfluß kognitiver Belastung experimentell bestätigt, der Einfluß von Zeitdruck jedoch nur für Symptom B. Die aktuelle Modellierung (links) überträgt diese Annahmen auf den unbestätigten Zusammenhang von Symptom A mit Zeitdruck. In späteren Experimenten stellt sich heraus, daß A und B unterschiedlichen Phasen U und V der Konzeptualisierung zuzuordnen sind, wobei Zeitdruck keinen Effekt auf Phase U zeigt. Daraufhin kann die Architektur des Netzes so angepaßt werden, daß diesen neuen Ergebnissen Rechnung getragen wird (rechtes Modell).

Zum gegenwärtigen Zeitpunkt ist die empirische Basis jedoch für Differenzierungen dieser Art zu schwach. Die Modellierung in READY soll in ihrer aktuellen Form vor allem an die Ergebnisse der empirischen Forschung angepaßt sein, ohne zu viele Interpolationen vorzunehmen, andererseits jedoch so viele Resultate wie möglich berücksichtigen.

#### 6.2.3.2 Quantitative Struktur

Die Anzahl von Fragmenten oder Fehlansätzen wurde in den meisten Arbeiten aus den Tabellen 3.17 und 3.16 an der Gesamtäußerungsmenge relativiert. Aus diesem Grund wird die HÄUFIGKEIT VON FRAGMENTEN UND FEHLANSÄTZEN (F) in der aktuellen Implementation mit einer Granularität von 3 erhoben. Für die unterschiedlichen Konfigurationen des Knotens ERFOLG DER KONZEPTUALISIERUNG (K) wird die Wahrscheinlichkeit für das Auftreten einer bestimmten relativen Anzahl von Fehlansätzen und Fragmenten in Abhängigkeit von dessen Hypothesen $k_i$ als Normalverteilung um den Erwartungswert $I_F(k_i)$ mit einer Standardabweichung von 0.5 definiert (Gleichung 6.11).

$$P(F = f_i \mid K = k_i) = \mathcal{N}_{I_F(k_i),0.5}(f_i) \tag{6.11}$$
$$I_F(k) = 1 - k \tag{6.12}$$

Gleichung 6.12 beschreibt den Einfluß des Konzeptualisierungserfolges auf die Erwartung des Symptoms: Mit abnehmendem Konzeptualisierungserfolg steigt die erwartete Häufigkeit für einen Fehlansatz oder Satzfragment. Es wird angenommen, daß die Entwicklung linear-komplementär ist. Grafisch lassen sich diese Zusammenhänge wie in Abbildung 6.8 (Einflüsse der Funktionen) und 6.9 (nach Diskretisierung und Normalisierung) darstellen.

Die Abbildungen zeigen die Verteilung der bedingten Wahrscheinlichkeiten $P(F = f_i \mid K = k_j)$ für die Hypothesen K = {hoch, mittel, niedrig}, d.h. die bedingten Wahrscheinlichkeiten für das Auftreten von wenig (0), einigen (0.5) und vielen (1) FRAGMENTEN UND FEHLANSÄTZEN unter der Vorbedingung eines bestimmten ERFOLGES DER KONZEPTUALISIERUNG. Die Wahrscheinlichkeit für viele Fragmente und Fehlansätze (F = 1) nimmt mit abnehmendem ERFOLG DER KONZEPTUALISIERUNG zu.

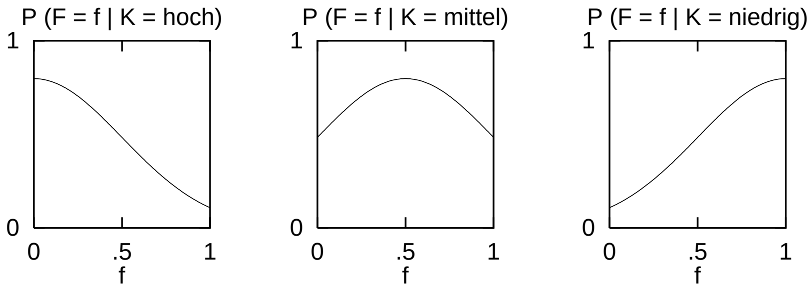


Abbildung 6.8: Definition der CPT des Knotens HÄUFIGKEIT VON FRAGMENTEN UND FEHLANSÄTZEN durch verschiedene Verteilungsfunktionen

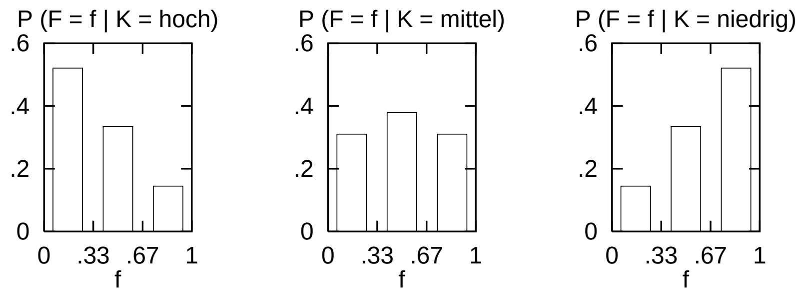


Abbildung 6.9: CPT des Knotens HÄUFIGKEIT VON FRAGMENTEN UND FEHLANSÄTZEN nach Diskretisierung und Normalisierung

**Verwandte Symptome** Symptome, die im vorliegenden Sprachproduktionsmodell und unter den für READY gültigen Annahmen durch eine einzelne Ursache erklärt werden sind die HÄUFIGKEIT VON LEXIKALISCHEN, SYNTAKTISCHEN UND PHONOLOGISCHEN SPRECHFEHLERN, MENGE ZUSÄTZLICHER INFORMATIONEN, EXPLIZITHEIT (ELLIPSEN UND ANAPHERN), und die INHALTLICH PASSENDE ANTWORT.

### 6.2.4 Zusammenhänge eines Symptoms mit mehreren Ursachen

#### 6.2.4.1 Qualitative Struktur

Für das Auftreten von Wiederholungen wurden in der Literatur zwei verschiedene Hypothesen favorisiert: Einerseits wurden sie als Verzögerungsphänomen eingestuft, andererseits erklärte Levelt (1983) sie mit einem Fehlalarm des Monitoring-Prozesses. Zusammen mit den Zusammenhängen, die für die Flüssigkeit der Sprachproduktion und die Funktionsweise des Monitoring-Prozesses mit Zeitdruck und kognitiver Belastung herausgearbeitet wurden, ergibt sich das in Abbildung 6.10 dargestellte Teilnetz.

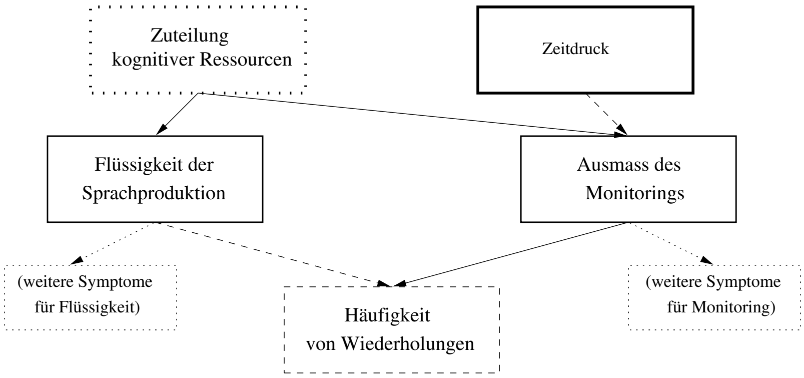


Abbildung 6.10: Mehrfache Ursachen für das Entstehen eines Symptoms

#### 6.2.4.2 Quantitative Struktur

Bei der Ableitung der bedingten Wahrscheinlichkeiten müssen nicht nur mehrfache Ursachen für eine Veränderung des Parameters HÄUFIGKEIT VON WIEDERHOLUNGEN berücksichtigt werden, sondern auch die *gegensätzlichen* Einflüsse der beiden Ursachen auf den Parameter: Eine Verbesserung des AUSMASSES AN MONITORING führt zu einer Erhöhung der HÄUFIGKEIT VON WIEDERHOLUNGEN, da die Wahrscheinlichkeit für einen falschen Alarm bei höherer Aktivität des Monitorings steigt. Andererseits führt eine verbesserte FLÜSSIGKEIT DER SPRACHPRODUKTION zu einer verringerten Notwendigkeit, Verzögerungsphänomene einzusetzen und daher zu weniger Wiederholungen. Die unterschiedlichen Einflüsse werden in der Einflußfunktion 6.13 zusammengefaßt.

$$I_W(f, m) = 0.4 * (1 - f) + 0.6 * m \tag{6.13}$$

Diese Formel sagt u.a. aus, daß der Erwartungswert für Wiederholungen maximal wird, wenn die FLÜSSIGKEIT minimal und das AUSMASS DES MONITORINGS maximal ist. Die unterschiedlichen Einflüsse der beiden Elternknoten gehen durch eine gewichtete Durchschnittsbildung auf den jeweiligen Werten der Elternknoten in die Berechnung ein: Dem höheren Erklärungswert von Levelts Hypothese wird Rechnung getragen, indem der Wert des Monitorings ($m$) mit einem stärkeren Gewicht versehen wird. Der Einfluß des Knotens FLÜSSIGKEIT DER SPRACHPRODUKTION ist komplementär linear: Je höher die Flüssigkeit ($f$), desto weniger Wiederholungen treten auf.

Für die Wahrscheinlichkeitsverteilungen unter den verschiedenen Elternkonfigurationen werden auch hier Normalverteilungen eingesetzt, die allerdings mit einem $\sigma$ von 0.4 instantiiert werden. Grund dafür ist, daß die *Sicherheit*, mit der Wiederholungen unter einer gegebenen Elternkonfiguration auftreten werden, als höher angesehen wird, als die für Fehlansätze oder Fragmente.[18] Diese Auffassung erscheint unmotiviert: Ohne empirische Daten kann jedoch weder das Gegenteil festgestellt noch eine Ähnlichkeit behauptet werden. Die Notwendigkeit einer empirischen Basis für ein Dialogsystem vom Typ READYs wird damit erneut unter Beweis gestellt. Abschnitt 6.3 wird zeigen, woher die fehlenden Anhaltspunkte beim Entwurf von CPTs kommen können.

Die vollständige CPT für den Knoten HÄUFIGKEIT VON WIEDERHOLUNGEN bei Elternknoten mit je 3 Hypothesen zeigt Formel 6.14.

$$P(\text{W} = (\text{niedrig, mittel, hoch}) \mid \text{F} = f, \text{M} = m) = \begin{cases} (0.35 \quad 0.41 \quad 0.24) & : \quad f = \text{niedrig}, m = \text{niedrig} \\ (0.19 \quad 0.40 \quad 0.41) & : \quad f = \text{niedrig}, m = \text{mittel} \\ (0.08 \quad 0.32 \quad 0.60) & : \quad f = \text{niedrig}, m = \text{hoch} \\ (0.48 \quad 0.37 \quad 0.15) & : \quad f = \text{mittel}, m = \text{niedrig} \\ (0.29 \quad 0.41 \quad 0.30) & : \quad f = \text{mittel}, m = \text{mittel} \\ (0.15 \quad 0.37 \quad 0.48) & : \quad f = \text{mittel}, m = \text{hoch} \\ (0.60 \quad 0.32 \quad 0.08) & : \quad f = \text{hoch}, m = \text{niedrig} \\ (0.41 \quad 0.40 \quad 0.19) & : \quad f = \text{hoch}, m = \text{mittel} \\ (0.24 \quad 0.41 \quad 0.35) & : \quad f = \text{hoch}, m = \text{hoch} \end{cases} \tag{6.14}$$

## 6.2.5 Symptome mit mehreren Dimensionen

### 6.2.5.1 Qualitative Struktur

Die umfangreiche Literatur zur Pausenforschung lieferte vor allem die Schlußfolgerung, daß sowohl eine Erhöhung der Anzahl als auch eine Verlängerung der durchschnittlichen Dauer von Pausen Indikatoren für kognitive Belastung sein können. Da individuelle Unterschiede verschiedene Strategien eines Sprechers zur Anpassung seines Sprechverhaltens an kognitive Belastung und Zeitdruck nach sich ziehen können, müssen die beiden Dimensionen des Parameters 'Pausen' auch in den Bayes'schen Netzen getrennt modelliert werden. Die linke Seite von Abbildung 6.11 zeigt ein mögliches Teilnetz, das diese Aspekte berücksichtigt.

Aus den Zusammenhängen in Bayes'schen Netzen ergibt sich jedoch ein Problem: Die beiden Dimensionen des Parameters 'Pausen' sind nicht unabhängig voneinander. Die Modellierung links in Abbildung 6.11 sagt das jedoch aus. Ein Sprecher kann zwar die Dauer seiner Pausen gegenüber Situationen durchschnittlicher Belastung verlängern, ohne zusätzliche Pausen zu setzen. Durch eine erhöhte Anzahl an Pausen nimmt jedoch (i.a.) deren Gesamtdauer zu. Insgesamt verändert sich in jedem der beiden Fälle das Verhältnis der Pausendauer zur Gesamtsprechdauer. Die Pausendaten eines Sprechers können somit zusammengefaßt werden, so daß nur ein Knoten im Netz existieren muß. Daraus ergibt sich ein ähnliches Teilmodell wie in Abschnitt 6.2.3.

[18] Eine Normalverteilung mit einem kleineren $\sigma$ ergibt eine geringere Streuung um den Mittelwert.

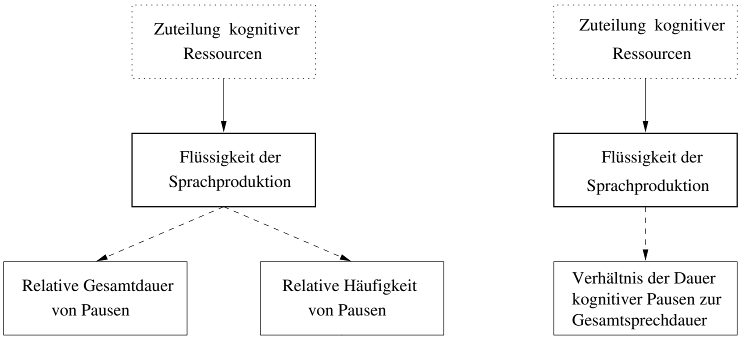


Abbildung 6.11: Varianten der Modellierung zweier unterschiedlicher Dimensionen ein und desselben Parameters: problematisch (links) und unproblematisch (rechts)

In Kapitel 3 zeigte sich jedoch auch die Notwendigkeit, vor allem für stille Pausen zu unterscheiden, ob sie kognitive oder kommunikative Ursachen haben. Gesamtpausendauer und Gesamtpausenzahl, die als Symptome kognitiver Belastung dienen, sollten Pausen aufgrund von kommunikativen Zielen des Sprechers nicht einschließen.

Aus diesem Grund wird in READY eine Vorverarbeitung für alle Einzelpausen durchgeführt, die den 'kognitiven Anteil' in den Pausendaten *vor* Instantiierung des Netzes isolieren soll. Dieser Vorverarbeitungsschritt ist in Abschnitt 5.3.4.2 dargestellt.

Ergebnis ist ein Vergleich der aktuellen Pausendauer für eine stille Pause mit der tatsächlichen Pausendauer an ihrer syntaktischen Position. Die Vorhersagewerte werden von den Meßwerten abgezogen, ggf. negative Werte zu 0 korrigiert.[19] Die Summe der resultierenden Einzelwerte ergibt die *Effektive Dauer stiller kognitiver Pausen* $E_s$. Für gefüllte Pausen liefert die Vorverarbeitung in der aktuellen Form keine Anpassung. Die *Effektive Dauer gefüllter Pausen* $E_f$ ist mit der Gesamtdauer gefüllter Pausen identisch. Die Summe $E_s + E_f$ ergibt die *Effektive Dauer kognitiver Pausen* $E_k$, welche ins Verhältnis zur Gesamtsprechdauer gesetzt wird. Dieses Verhältnis erzeugt den Wert, mit dem der Knoten VERHÄLTNIS DER DAUER KOGNITIVER PAUSEN ZUR GESAMTSPRECHDAUER im rechten Modell von Abbildung 6.11 bzw. im gesamten Teilnetz von Abbildung 6.14 instantiiert wird.

In Abschnitt 5.3.4.2 wurde angegeben, daß in READY zwei verschiedene Mechanismen zur Verfügung stehen, um die Pausenvorhersage für stille Pausen zu simulieren und damit deren kognitive Relevanz zu berechnen. Durch die Abbildung aller Werte auf das Einheitsintervall (Anhang B.1) braucht bei geänderter Vorverarbeitung jedoch keine CPT geändert zu werden: Lediglich die Rahmenwerte für die Intervallgrenzen im Abbildungsschritt werden angepaßt, wodurch trotz geänderter Vorhersage für ein und dieselbe Äußerung die gleiche Hypothese für das Pausenverhältnis instantiiert werden kann.[20] Zu beachten ist allerdings der damit verbun-

[19]Eine Pause, die kürzer ist als der vorhergesagte Wert aufgrund der prosodischen-syntaktischen Struktur, hat für READY keine Bedeutung.

[20]D.h., daß der höhere Wert bei `*use-pause-time* = 'all` auf ein niedrigeres Intervall abgebildet wird, da er weniger aussagekräftig für die kognitive Belastung des Sprechers ist.

dene Informationsverlust: Äußerungen mit einer unterschiedlichen kognitiven Relevanz ihrer Einzelpausen (d.h. einem unterschiedlichen $E_k$) instantiieren bei gleichem Verhältnis der Gesamtdauer von Pausen zur Äußerungsdauer die gleiche Hypothese, während eine detailliertere Vorverarbeitung zu einer präziseren Instantiierung führen könnte.

#### 6.2.5.2 Verwandte Symptome

Eine Aufspaltung verschiedener Dimensionen ein und desselben Parameters ergibt sich auch für einige andere Parameter: Bei Selbstkorrekturen stehen *Häufigkeit*, *Wohlgeformtheit* und *Anwesenheit eines Korrekturterms* zur Auswahl.

Die *durchschnittliche Wohlgeformtheit* aller Selbstkorrekturen einer Äußerung, sowie die *durchschnittliche Häufigkeit von Korrekturtermen* geben Auskunft über die Funktionsweise des Monitoring-Prozesses. Das Maß DURCHSCHNITTLICHE QUALITÄT VON SELBSTKORREKTUREN (siehe den entsprechenden Knoten in Abbildung 6.14) faßt diese beiden Aspekte zusammen. Beide Qualitätsmaße werden derzeit binär für jede Korrektur in der Äußerung vergeben und aus ihnen der Durchschnitt gebildet.

Die Bedeutung der *relativen Häufigkeit von Selbstkorrekturen* für die Einschätzung der kognitiven Belastung ist durch die widersprüchlichen experimentellen Ergebnisse (siehe Tabelle 3.12) derzeit nicht ohne weiteres verwendbar. Eine entsprechende Variable im Netz von Abbildung 6.14 wird jedoch trotz allem eingeführt, da Selbstkorrekturen eines der wichtigsten spontansprachlichen Phänomene darstellen und der Zusammenhang zwischen der Häufigkeit von Korrekturen und der Funktionsweise des Monitoringprozesses nicht in Zweifel zu ziehen ist. In Abschnitt 6.2.3.1 wurde ein möglicher Weg gezeigt, wie durch Differenzierung des Effektes von Zeitdruck und Belastung auf Teilaspekte der Phasen der Sprachproduktion möglicherweise die unterschiedlichen Experimentaleffekte erklärt und damit der Parameter in der Zukunft für die Modellierung besser berücksichtigt werden kann.

Die Daten der beiden Knoten im Netz sowie die Angaben zur Bewertung von Einzelkorrekturen sind im Anhang B.5 zu finden.

Das *Verhältnis von Korrekturen zu (erkennbaren) Sprechfehlern* könnte eine weitere Dimension des Parameters Selbstkorrekturen darstellen, die ebenfalls für die Beurteilung des Monitorings einsetzbar ist. Allerdings ist dieser Wert von der Erkennbarkeit von Fehlern abhängig und kann daher nur Korrekturen und Fehler der *Error*-Kategorie betreffen. Da die zugehörigen Fehler außerordentlich selten sind, wird auf die Verwendung dieses Parameters aktuell verzichtet.

### 6.2.6 Symptome mit individuellen Basisparametern

Individuelle Unterschiede spielen bei der Beurteilung bestimmter Sprechparameter eine wesentliche Rolle (siehe Tabelle 3.22). Als prominenteste Vertreter für diese Parameter wurden die Artikulationsgeschwindigkeit und die Verwendung gefüllter Pausen identifiziert. Im folgenden wird am Beispiel der Artikulationsgeschwindigkeit demonstriert, wie individuelle Basisparameter in die Modellierung mit dynamischen Bayes'schen Netzen einbezogen werden können.

#### 6.2.6.1 Qualitative Struktur

Individuelle Basiswerte stellen eine der Ursachen für die aktuelle Ausprägung des zugehörigen Sprechparameters dar und sind daher im Bayes'schen Netz als Elternknoten modellierbar. Ausschlaggebend für die Anpassung der Einschätzung der kognitiven Belastung und des Zeitdrucks (indirekt über die Flüssigkeit der Sprachproduktion) ist nicht unmittelbar die Ausprägung des beobachteten Parameters, sondern die *Abweichung* von seinem Basiswert. Daraus ergibt sich für die Artikulationsgeschwindigkeit das in Abbildung 6.12 dargestellte kausale Modell. Die gestrichelte Linie für die Verbindungen zwischen ZEITDRUCK und ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT, sowie letzterer und BEOBACHTETER ARTIKULATIONSGESCHWINDIGKEIT deutet darauf hin, daß der Zusammenhang keine eindeutige Richtung besitzt wie bisher. Dafür ist die Semantik des Knotens für die Abweichung verantwortlich (siehe 6.2.6.5).

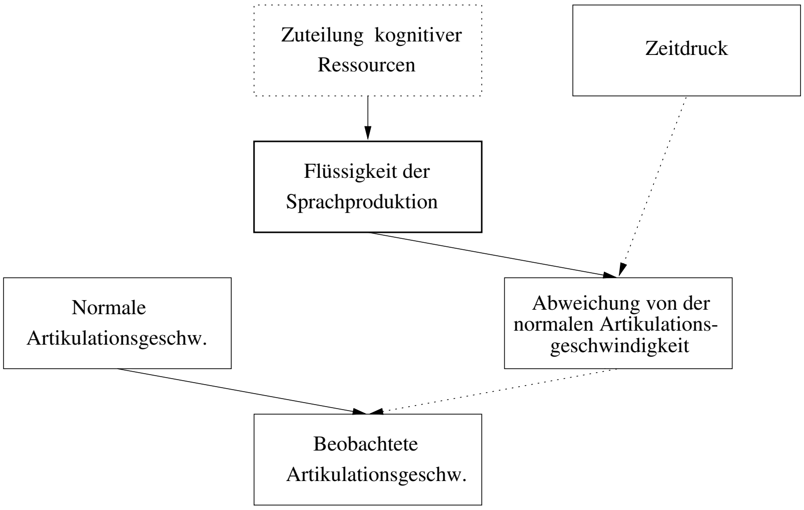


Abbildung 6.12: Modellierung individueller Basiswerte am Beispiel der Artikulationsgeschwindigkeit

Woher kann ein Dialogsystem jedoch Wissen über die Basisartikulationsgeschwindigkeit eines Benutzers erlangen? Der Knoten NORMALE ARTIKULATIONSGESCHWINDIGKEIT kann nie instantiiert werden, da für ihn keine direkte Evidenz zu erlangen ist. Als Alternative steht einem Dialogsystem daher lediglich offen, daß es versucht, diesen Wert aus den Daten der vergangenen Dialogturns zu *lernen*. Um zu klären, wie mit den Bayes'schen Netzen in READY der Lernvorgang modelliert werden kann, ist ein Blick auf das Zeitscheibenmodell im System und die verschiedenen Kategorien von Knoten in den sich daraus ergebenden temporalen Bayes'schen Netzen des Systems notwendig.

#### 6.2.6.2 Begriffliche Details: Temporale Bayes'sche Netze und ihre Knotenkategorien

Bayes'sche Netze können für die Modellierung zeitlich veränderlicher Information eingesetzt werden. Hauptgedanke dabei ist, daß die Einschätzung bestimmter Variablen $V$ zu einem Zeitpunkt $t_k$ auf zwei Einflüssen beruhen kann:

- die direkte und indirekte Evidenz für $V$, die zum Zeitpunkt $t_k$ erhoben werden kann.
- der Wert von $V$ zum Zeitpunkt $t_{k-1}$.

Damit unterschiedliche Werte in Abhängigkeit von Zeitpunkten repräsentiert werden können, ist eine Diskretisierung des Zeitintervalls notwendig, in dem verschiedene Evidenzen erhoben und Variablenwerte ausgelesen werden. Im System READY wird ein Zeitscheibenmodell eingesetzt, in dem derzeit eine Zeitscheibe einem Dialogturn entspricht (Schäfer & Weyrath, 1996; Schäfer, 1998).[21] Die Symptomknoten des Bayes'schen Netzes für die jeweilige Zeitscheibe werden mit den Evidenzen aus dieser Zeitscheibe, d.h. aus dem zugehörigen Dialogturn, instantiiert. Das Lösen des Netzes führt dann zu einer Wahrscheinlichkeitsverteilung, die die Systemeinschätzung der Variablen des Netzes unmittelbar *nach* der Interpretation der erlangten Evidenz wiedergibt, d.h. nach Ablauf des entspechenden Dialogturns. Drei Variablentypen sind dabei zu unterscheiden:

**Dynamische Variablen** dienen der Modellierung von Merkmalen, deren Wert sich im zeitlichen Verlauf wesentlich ändern kann. In jeder Zeitscheibe wird dafür eine neue Inkarnation eines *dynamischen Knotens* eingeführt. Regelhaftigkeiten der Veränderung ihres Wertes werden durch die Abhängigkeit von der Inkarnation der jeweiligen Variablen in der vorangegangenen Zeitscheibe erfaßt. Dynamische Knoten in Zeitscheibe $t_k$ (für $k \neq 0$) haben damit stets mindestens einen Vorgänger, der die Einschätzung der Variablen zum vorangegangenen Betrachtungszeitpunkt, also in Zeitscheibe $t_{k-1}$ angibt. Dynamische Variablen in READY sind zum Beispiel die beiden zentralen Benutzereigenschaften VERFÜGBARE ARBEITSGEDÄCHTNISKAPAZITÄT und ZEITDRUCK.[22]

**Temporäre Variablen** einer Zeitscheibe sind Variablen, die umittelbar nur mit Variablen aus dieser Zeitscheibe in Verbindung stehen. Zwischen Inkarnationen derartiger Variablen in verschiedenen Zeitscheiben besteht keine direkte Abhängigkeit. Alle sprachlichen Symptome, die für eine einzelne Äußerung erhoben werden, sind durch *temporäre Knoten* in READY vertreten, da sie lediglich für die aktuelle Äußerung Gültigkeit haben.

**Statische Variablen** sind Variablen, deren tatsächlicher Wert sich während des Dialoges nicht verändert, für die das System jedoch über eine Einschätzung verfügen muß.[23] Sie haben im Gegensatz zu temporären und dynamischen Variablen *nur eine einzige Knoteninstanz* im gesamten dynamischen Netz. Die Einschätzung der Wahrscheinlichkeitsverteilung dieses Knotens in Zeitscheibe $k$ basiert auf *allen* für die Variable relevanten Symptomen in *allen* bisher erzeugten Zeitscheiben (d.h. zu allen Zeitpunkten $t_i$ mit $0 \leq i \leq k$). Zu den statischen Variablen zählen in READY Benutzereigenschaften wie das WISSENSNIVEAU und Basissprechparameter wie die NORMALE ARTIKULATIONSGESCHWINDIGKEIT, da von ihnen angenommen wird, daß sie sich während des Dialogverlaufs nicht ändern.

[21] Diese Diskretisierung ist der Grund dafür, daß alle Sprechparameter für Äußerungen und nicht für Sätze erhoben werden.

[22] Hinzu kommen Variablen, deren Wert in einer Zeitscheibe nur als Anpassung des Wertes der vorangegangenen Zeitscheibe gesehen werden kann. Dazu zählen EMOTIONALE BELASTUNG und SITUATIVE ABLENKUNG.

[23] Im Laufe des Dialoges ändert sich die *Einschätzung* des Wertes durch das System, jedoch nicht der tatsächliche Wert selbst!

#### 6.2.6.3 Basisparameter NORMALE ARTIKULATIONSGESCHWINDIGKEIT

Die NORMALE ARTIKULATIONSGESCHWINDIGKEIT (N) eines Sprechers kann aus den Sprachdaten eines Sprechers $P$ in einem Dialog wie folgt abgeleitet werden: In jeder Zeitscheibe $t_k$ äußert sich $P$ mit einer bestimmten Artikulationsgeschwindigkeit. In einigen Fällen kann die Geschwindigkeit (indirekt) von außerordentlicher Belastung oder Zeitdruck beeinflußt werden. Für jede Zeitscheibe $t_k$ existiert ein Knoten $B_k$ im Netz, der mit dem beobachteten Wert der Artikulationsgeschwindigkeit zum Zeitpunkt $t_k$ instantiiert wird.[24] Aufgrund dieser $k$ Evidenzen verfügt der Knoten NORMALE ARTIKULATIONSGESCHWINDIGKEIT zum Zeitpunkt $t_k$ über eine Wahrscheinlichkeitsverteilung, die die a-posteriori-Wahrscheinlichkeiten für N unter $k$ Evidenzen beschreibt, d.h.

$$P(N \mid B_0 = b_0, B_1 = b_1, ...B_k = b_k) \tag{6.15}$$

**A-Priori-Annahmen** Die a-posteriori-Wahrscheinlichkeiten für N können in Zeitscheibe $t_0$ nur berechnet werden, wenn das System über eine initiale Einschätzung, d.h. a-priori-Wahrscheinlichkeiten, verfügt. Diese Einschätzung gibt an, mit welcher Wahrscheinlichkeit ein Sprecher einer der unterschiedlichen Kategorien *schneller*, *normaler* und *langsamer* Sprecher zugeordnet werden kann, solange das System noch nichts anderes über ihn weiß.

Die Verteilung der Artikulationsgeschwindigkeiten wird für unterschiedliche Sprecher als normalverteilt angenommen. Mittelwert und Standardabweichung für diese Normalverteilung werden aus den Daten der Feldstudie entnommen: Die Anrufer wiesen eine durchschnittliche Artikulationsgeschwindigkeit von ca. 6.0 Silben/s auf bei einer Standardabweichung von 1.9. Daraus ergibt sich die Verteilung 6.16. Nach Diskretisierung und Abbildung der Artikulationsgeschwindigkeit auf das Intervall [2,8] (Silben/s) ist das gleichbedeutend mit Verteilung[25] 6.17. Zu beachten ist, daß die Werte der einzelnen Hypothesen $n_i$ nicht von Einflüssen eines Elternknotens abhängen, da N ein Wurzelknoten ist. Daher wird keine Einflußfunktion angegeben.

$$P(N = n_i) = \mathcal{N}_{6,1.9}(n_i) \tag{6.16}$$

$$P(\mathrm{N} = n) = \begin{cases} 0.16 & : \quad n = niedrig \quad ([2;4]] \\ 0.42 & : \quad n = mittel \quad ([4;6]) \\ 0.42 & : \quad n = hoch \quad ([6;8]) \end{cases} \tag{6.17}$$

[24] Im echten READY-System wird $B_k$ nur dann instantiiert, wenn die Silbenzahl der zugehörigen Äußerung $\geq 4$ ist. Das führt zu weniger Evidenzen für den Basisparameter, da nicht instantiierte Blattknoten keinen Einfluß auf das restliche Netz haben.

[25] Bei der Implementierung sind auch die beiden Normalverteilungsparameter abhängig von der Abbildung der Extremwerte auf die Intervallwerte 0 und 1: Bei der angegebenen Abbildung von 2.0 Silben/s auf 0 und 8.0 Silben/s auf 1 wird das Intervall auf ein Sechstel seiner ursprünglichen Länge zusammengezogen. Dadurch ergeben sich $\sigma$ = 1.9/6 = 0.317 und $\mu$ = 0.67.

#### 6.2.6.4 Informationsvariable: BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT

In jedem Dialogturn des Benutzers wird von READY eine bestimmte Artikulationsgeschwindigkeit beobachtet. Diese kann nicht selbst Einfluß auf die Einschätzung der FLÜSSIGKEIT DER SPRACHPRODUKTION haben, da sie vom Sprecher abhängig ist. Direkte Folge einer verminderten Flüssigkeit ist lediglich die ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT.

Die BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT (B) eines Sprechers ist nach dem in Abbildung 6.12 angegebenen Modell abhängig von der NORMALEN ARTIKULATIONSGESCHWINDIGKEIT (N) und der ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT (A). Sie berechnet sich als Summe aus diesen beiden Einflüssen, wobei die Grenzen des Modellierungsintervalls $u$ und $o$ zu berücksichtigen sind. Formel 6.18 spezifiziert eine entsprechende Behandlungsroutine.

$$\begin{aligned} I_B(n,a) \;=\; & \texttt{if } (n+a>o) \texttt{ then } o \\ & \quad \texttt{elseif } (n+a<u) \texttt{ then } u \\ & \quad \texttt{else } n+a \\ & \quad \texttt{fi} \end{aligned} \tag{6.18}$$

Wenn Sprecher keine Abweichungen von ihrem Basiswert zeigen, muß die gleiche Verteilung wie für den Basiswert entstehen. Die Wahrscheinlichkeitsverteilung für die Konfiguration der Elternknoten aus der Hypothese '*keine* ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT' und einer beliebigen Hypothese für NORMALE ARTIKULATIONSGESCHWINDIGKEIT ist damit vorgegeben:

$$P(B=n \mid N=n, A=0) = P(N=n) \tag{6.19}$$

Auch die Wahrscheinlichkeiten für die Hypothesen der BEOBACHTETEN ARTIKULATIONSGESCHWINDIGKEIT unter einer bestimmten Abweichung lassen sich einfach berechnen: Sie entsprechen einer Verschiebung der Normalverteilung für die normale Artikulationsgeschwindigkeit um den Offset der Abweichung (6.20). Abbildung 6.13 zeigt den Effekt für eine Abweichung von $-0.9$ Silben/s.

$$P(B=b_i \mid N=n, A=a) = \mathcal{N}_{n+a,1.9}(b_i) \tag{6.20}$$

#### 6.2.6.5 Unabhängigkeit von individuellen Basiswerten: ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT

**Qualitativer Zusammenhang**

Unabhängig von individuellen Basiswerten hat nun lediglich die ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT (A) einen Einfluß auf die Einschätzung der FLÜSSIGKEIT. Die Hypothesen von A weisen aber eine Besonderheit auf: Sie werden nicht auf Werte einer von 0 oder einem anderen Minimalwert aus monoton ansteigenden Funktion abgebildet. Um eine Abweichung in beide Richtungen interpretieren zu können, wird der mittleren Hypothese dieses

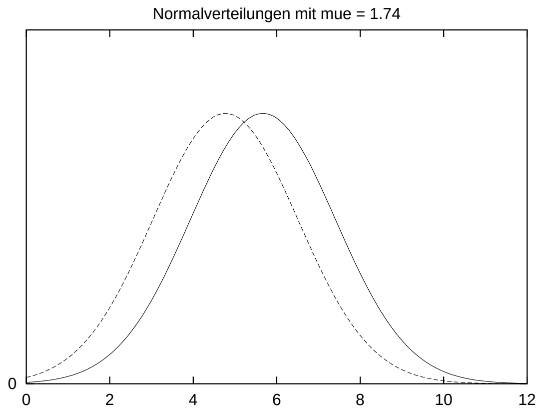


Abbildung 6.13: Effekt der ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT auf die Vorhersagen für die BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT

Knotens die Bedeutung '*keine* Abweichung', d.h. *Abweichung = 0* zugeordnet, den darunterliegenden eine *negative* Abweichung ('−', d.h. eine Verlangsamung) und den darüberliegenden eine *positive* Abweichung ('+', d.h. eine Beschleunigung).

In Abschnitt 6.2.3.1 wurde auf die fehlenden Differenzierungsmöglichkeiten für die Einflüsse von Zeitdruck auf unterschiedliche Phasen der Sprachproduktion hingewiesen und ein Beispiel gegeben, wie diese Einflüsse aus bekannten Daten interpoliert werden können.

Um einen Zusammenhang der FLÜSSIGKEIT DER SPRACHPRODUKTION mit ZEITDRUCK herzustellen, können jedoch die impliziten Annahmen der Modellierung, die in Abschnitt 6.2.3.1 erläutert wurden, nicht angewendet werden: Symptome für die Einschätzung der Flüssigkeit der Sprachproduktion sind *Pausen*, die *Artikulationsgeschwindigkeit* und die *Häufigkeit von Wiederholungen*. Für keinen der drei Parameter wurde ein überzeugender Nachweis für eine klare Tendenz unter Zeitdruck erbracht.[26] Über den Zusammenhang von Zeitdruck mit der Flüssigkeit der Sprachproduktion ist daher nichts Eindeutiges bekannt und auf eine entsprechende Kante im Modell wurde daher verzichtet.

In der quasi-experimentellen Studie von Weyrath (1998) zeigte sich jedoch, daß Versuchspersonen der Meinung sind, daß ein Einfluß von Zeitdruck auf *Sprechgeschwindigkeit* existiert. Eine Temposteigerung bei der Ausführung von Handlungen unter Zeitdruck wurde u.a. von Marx (1984) in ihren Vorüberlegungen identifiziert. Die Semantik des Knotens FLÜSSIGKEIT DER SPRACHPRODUKTION paßt jedoch nicht zu dieser Abhängigkeit: Temposteigerungen in der Artikulationsphase können, aber müssen nicht mit der Qualität der inkrementellen Verarbeitung zusammenhängen. Die Artikulationsphase stellt ohnehin einen Engpaß des Sprachproduktionssystems dar, da sie nicht unbegrenzt beschleunigt werden kann, jedoch Sprecher oft mehr sagen würden, wenn es motorisch möglich wäre. Statt einen zusätzlichen Erklärungsknoten für die Steigerung der Artikulationsgeschwindigkeit ins Netz aufzunehmen, wird daher vereinfachend

[26]Die einzelne Evidenz aus Marx (1984) für Wiederholungen kann als Hinweis verwendet werden, ist jedoch vor allem auch wegen der widersprüchlichen Ergebnisse ihrer beiden Bedingungen problematisch.

einer direkten Verbindung zwischen ZEITDRUCK (Z) und ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT (A) der Vorzug gegeben, solange keine differenzierteren Forschungen zur Wirkung von Zeitdruck verfügbar sind (siehe Abb. 6.14).

**Quantitativer Zusammenhang**
Wie stark die Artikulationsgeschwindigkeit von ihrem Basiswert abweichen wird, ist von der Flüssigkeit der Sprachproduktion und vom Zeitdruck abhängig. Für die Bedingung hoher Flüssigkeit unter geringem Zeitdruck ist die Wahrscheinlichkeit für eine normale Artikulationsgeschwindigkeit, d.h. eine fehlende Abweichung, recht hoch.

Dieser Wert wird mit 0.9 angenommen, was bedeutet, daß für 10 % aller Sprecher auch bei guter inkrementeller Verarbeitung und fehlendem Zeitdruck eine Abweichung von der normalen Artikulationsgeschwindigkeit zu erwarten ist. Damit sind gewissermaßen Randbedingungen definiert, die als 'Normalwerte' dienen können und einen Fixierungspunkt für die Entwicklung der CPT darstellen. Prinzipiell entspricht die verbleibende Unsicherheit von 0.1 möglichen Einflüssen, die im Netz nicht erfaßt sind.

Bei niedriger Flüssigkeit der Sprachproduktion, d.h. u.a. unter außerordentlich hoher Belastung, ist die Wahrscheinlichkeit für eine Verringerung der Artikulationsgeschwindigkeit höher, während bei zunehmendem Zeitdruck die Wahrscheinlichkeit für die Hypothese der Beschleunigung zunimmt. Diese gegensätzlichen Einflüsse werden über die Constraints 6.21 und 6.22 integriert.

$$I_A(f,z) = \begin{cases} 0.5 & : f = hoch, \quad z = niedrig \\ > 0.5 & : f = hoch, \quad z = hoch \\ < 0.5 & : f = niedrig, \quad z = niedrig \end{cases} \tag{6.21}$$

$$\begin{aligned} P(A = 0 \mid Z = \text{niedrig}) &= P(A = + \mid Z = \text{hoch}) \\ P(A = 0 \mid F = \text{hoch}) &= P(A = - \mid F = \text{niedrig}) \end{aligned} \tag{6.22}$$

In der aktuellen Implementation sind für den Knoten ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT lediglich drei Hypothesen vorgesehen. Diese Knappheit bei der Diskretisierung ist, wie auch für andere Variablen, auf Laufzeitprobleme zurückzuführen: Die Lösungsdauer für ein Netz hängt u.a. von der Gesamtzahl der Hypothesen im Netz ab. Im Laufe des Dialoges wächst das Bayes'sche Netz in READY stark an, was auch bei wenigen Hypothesen je Knoten schon zu Lösungszeiten von mehreren Sekunden in den ersten Zeitscheiben führt. Der weitere Entwurf der CPT wird nach dem Vorbild von Abschnitt 6.2.5 durchgeführt.

#### 6.2.6.6 Anpassungen der Einschätzung für NORMALE ARTIKULATIONSGESCHWINDIGKEIT

Die anfängliche Unsicherheit des Systems über den tatsächlichen Sprecher wird durch relativ starke Schwankungen der Einschätzung für die NORMALE ARTIKULATIONSGESCHWINDIGKEIT in den ersten Zeitscheiben deutlich. Nach einer gewissen Zahl von Evidenzen stabilisiert sich diese Einschätzung jedoch und kann von einzelnen lokalen Extremwerten für BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT in späteren Zeitscheiben nur noch geringfügig beeinflußt werden. Eine über mehrere Zeitscheiben hinweg ähnliche, von der bisherigen Einschätzung stark abweichende beobachtete Artikulationsgeschwindigkeit führt dagegen zu einer allmählichen Anpassung an diese neuen Evidenzen.

Man mag einwenden, daß in einer ganz konkreten Dialogsituation READY den Wert der normalen Artikulationsgeschwindigkeit nicht korrekt einschätzt: Für einen Sprecher, der *während des gesamten Dialoges* unter Zeitdruck oder kognitiver Belastung steht und bei dem die Artikulationsgeschwindigkeit *während der gesamten Zeit* ober- oder unterhalb ihres Basiswertes bleibt, nimmt READY letztendlich scheinbar diese Artikulationsgeschwindigkeit als Normalwert an. Eine solche Fehleinschätzung wäre eine logische Konsequenz aus fehlender weiterer Information über den Sprecher. READY's Inferenzen würden in dieser Ausnahmesituation der Schlußweise von Menschen unter den gleichen Bedingungen fehlender kontextueller Information ähneln: Wenn jemand seit Monaten nur Regen vor seinem Fenster beobachtet, wird er dies irgendwann zum Normalfall erklären. Erst eine Mehrzahl von Sonnentagen kann ihn dazu verleiten, seine Annahme (langsam) zu revidieren.

READY *verfügt* jedoch über die notwendige zusätzliche Information: Bei einem permanent belasteten Sprecher sind über den gesamten Dialog hinweg auch andere Belastungssymptome zu erwarten, die nicht von Basiswerten abhängen. Diese führen (diagnostisch) zu der Einschätzung, daß die Belastung in der Tat hoch ist. Dadurch werden jedoch auch die Vorhersagen für die ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT angepaßt und besitzen für eine geringe beobachtete Geschwindigkeit einen höheren Erklärungswert als der Basiswert. Für eine niedrige Artikulationsgeschwindigkeit wird die Einschätzung des Basiswertes dann stärker in Richtung auf einen höheren Wert angenommen: Die starke Wahrscheinlichkeit einer Abweichung nach unten erklärt die niedrige Artikulationsgeschwindigkeit zu einem erheblichen Teil.

#### 6.2.6.7 Fazit

Mit der gewählten Modellierung ist auf elegante Weise erreicht, daß der tatsächliche Wert der Artikulationsgeschwindigkeit eines Sprechers nicht mehr in direkten Zusammenhang mit seiner kognitiven Belastung gebracht wird, sondern nur noch ein Abweichungswert von seiner Basisgeschwindigkeit. Das gleiche Vorgehen kann für die Verteilung von gefüllten Pausen oder anderen, einen Basiswert voraussetzende Sprechparameter in ähnlicher Weise angewandt werden.

### 6.2.7 Symptome mit unterschiedlichem diagnostischen Wert und seltene Symptome

Für einzelne Symptome wurde festgestellt, daß sie im Korpus nur sehr selten auftraten und ihr Zusammenhang mit kognitiver Belastung durch die Zuordnung zur Phase der Formulierung nur als gering einzuschätzen ist. Wenn jedes einzelne dieser Symptome durch einen eigenen Knoten im Netz vertreten wäre, ergäben sich zahlreiche Knoten, die nur in wenigen Zeitscheiben instantiiert werden können.[27] Um ihrer geringen Bedeutung Rechnung zu tragen, die Symptome jedoch für die Analyse nicht vollkommen auszuschließen, wurden daher z.B. phonologische, flexionsmorphologische und lexikalische Sprechfehler zu einem Knoten HÄUFIGKEIT SPRECHFEHLER zusammengefaßt.

Dieser Knoten wird als binär, d.h. mit zwei Hypothesen modelliert. Das bedeutet, daß nicht differenziert wird, wieviele dieser Fehler auftraten, sondern nur, *ob* sie auftraten. Ursache für ihr Auftreten ist ein geringer *Erfolg der Formulierung und Artikulation*. Durch den flachen Verlauf der Wahrscheinlichkeitsverteilungen für einzelne Hypothesen in der CPT dieses Knotens (siehe Anhang B.5) ist wiederum indirekt sichergestellt, daß phonologische, flexionsmorphologische und lexikalische Sprechfehler als Symptome geringen Erfolges von Formulierung und Artikulation nur einen geringen Einfluß auf die Einschätzung von Belastung und Zeitdruck erreichen. Der Aufbau der CPT von HÄUFIGKEIT SPRECHFEHLER entspricht der Vorgehensweise in Abschnitt 6.2.3 und wird daher im Anhang angegeben.

Einen anders gelagerten Fall stellt der Knoten VERWENDUNG NICHT AUFLÖSBARER ORTSDEIXIS dar: Die Verwendung von nicht auflösbarer Ortsdeixis gehört einerseits ebenfalls zur Kategorie der seltenen Symptome. Das spiegelt sich in der Form der CPT des zugehörigen Knotens wider, die auch unter geringem Erfolg der Konzeptualisierung und des Monitorings nur eine geringe Wahrscheinlichkeit für das Auftreten ortsdeiktischer Ausdrücke vorhersagt (siehe Anhang B.5). Durch die Wahl eines binären Knotens ist jedoch der Effekt eines tatsächlich auftretenden ortsdeiktischen Ausdrucks recht stark, da er jeder möglichen Vorhersage für die Hypothese 'nein' widerspricht. Für die bedingten Wahrscheinlichkeiten der Ortsdeixis gilt:

$$0 < P(O = 1 \mid K = k, M = m) < 0.1 \tag{6.23}$$

Die Instantiierung der Hypothese 'ja' für diesen Knoten zieht daher (natürlich in Abhängigkeit von der restlichen Netzkonfiguration) eine relativ drastische Anpassung der Einschätzungen für die Verteilungen der Elternknoten nach sich. Die Ortsdeixis ist daher ein Symptom mit relativ hohem diagnostischem Wert. Differenziertere Maße können diese Sonderstellung in der Zukunft einschränken helfen, indem z.B. die Anzahl deiktischer Ausdrücke, wie bereits vorgeschlagen, an der Gesamtzahl der Referenten und Lokalisationsausdrücke relativiert wird.

### 6.2.8 Ausnutzung des Zeitscheibenmodells für kognitive Subprozesse

Neben der Modellierung dynamischer Variablen bietet sich mit dem Zeitscheibenmodell in READY die Möglichkeit, verschiedene Bestandteile einer Benutzeräußerung unterschiedlichen kognitiven Prozessen zuzuordnen: Im Abschnitt 3.2.6 wurde darauf hingewiesen, daß der Moment, in dem die Sprachanalyse abgeschlossen ist und der Benutzer mit einer Generierung einer eigenen Äußerung beginnt, nur dann festzustellen ist, wenn der Sprecher ein explizites Backchanneling-Signal gibt. In READY wird dieses Backchanneling-Signal auf eine Systemfrage oder -instruktion für eine erste Revision der Systemeinschätzung über die kognitive Belastung des Benutzers verwendet: Gibt der Sprecher ein Rückmeldesignal, zeigt er damit an, daß er zumindest die Frage verstanden hat. Mangelndes Sprachverständnis (Knoten ERFOLG BEI DER

[27] Alternativ werden die Knoten jeweils mit dem gleichen Wert instantiiert und haben daher immer den gleichen Effekt auf die Anpassung der Systemeinschätzung.

SPRACHANALYSE) wird damit als Erklärung für einen mangelnden Erfolg bei der Antwortgenerierung weitgehend ausgeschlossen.

Durch die Segmentierung des Dialoges in Zeitscheiben hat READY die Möglichkeit, seine Einschätzung des Benutzers sowohl aufgrund neuer Evidenzen als auch unter Berücksichtigung alter Inferenzen anzupassen, ohne diese zu verwerfen.

### 6.2.9 Gesamtes Sprachproduktionsnetz

Alle Knoten, die sich aus der Modellierung ergeben, sind in Anhang B.5 zusammen mit ihren wichtigsten Parametern dargestellt. Gemeinsam erzeugen sie das in Abbildung 6.14 dargestellte Bayes'sche Netz, das die Modellierung der sprachlichen Symptome der Benutzeräußerung in READY übernimmt.

In der Abbildung sind alle Variablen als Rechtecke dargestellt. Symptomvariablen sind an einer gestrichelten Umrandung erkennbar, die beiden modellierten internen Variablen an einer fetten Umrandung. Individuelle Basisparameter haben gepunktete Ränder. Alle anderen Knoten stellen Zwischenvariablen dar. Kausalzusammenhänge sind als durchgehende Pfeile gekennzeichnet, wenn die Wahrscheinlichkeit für das Auftreten einer bestimmten Hypothese des Kindknotens unter der Voraussetzung des Eintretens einer höher bewerteten Hypothese des Elternknotens gegenüber niedriger bewerteten Hypothesen des Elternknotens steigt. Andernfalls ist die Kante als gestrichelte Linie markiert. An einem Beispiel wird diese Aussage klarer: Unter stärkerem Monitoring steigt z.B. die Wahrscheinlichkeit für das Auftreten von Selbstkorrekturen an, während die Flüssigkeit der Sprachproduktion abnimmt.

## 6.3 Quantitative Verbesserungsmöglichkeiten des bestehenden Modells

Eines der Probleme der Modellierung aus Abschnitt 6.2 bestand darin, die verwendeten bedingten Wahrscheinlichkeiten für verschiedene Knoten zu motivieren. Im folgenden sollen Möglichkeiten für die Ableitung von CPTs aus empirischen Werten skizziert werden. Dieser Punkt ist Inhalt der aktuellen Forschungsaufgaben in READY (siehe z.B. Berthold & Jameson, 1999). Da alle CPTs im bisher beschriebenen Modell durch Constraints und Funktionen beschrieben wurden, müssen keine Zellen etwaiger Tabellen angepaßt werden, sondern es genügt für eine Anpassung einer CPT, ihre Constraints zu verändern.

### 6.3.1 Woher die Zahlen nehmen?

Sowohl Pearl (1988) als auch Charniak (1991), Russel und Norvig (1995) und Jensen (1996) gehen auf ein entscheidendes Problem beim Entwurf von Bayes'schen Netzen ein: Wo kommen all die Zahlen her? Russel und Norvig (1995) und Jensen (1996) geben dafür drei verschiedene Möglichkeiten an:

1. Häufigkeiten als Ergebnisse von Experimenten oder aus Falldatenbasen

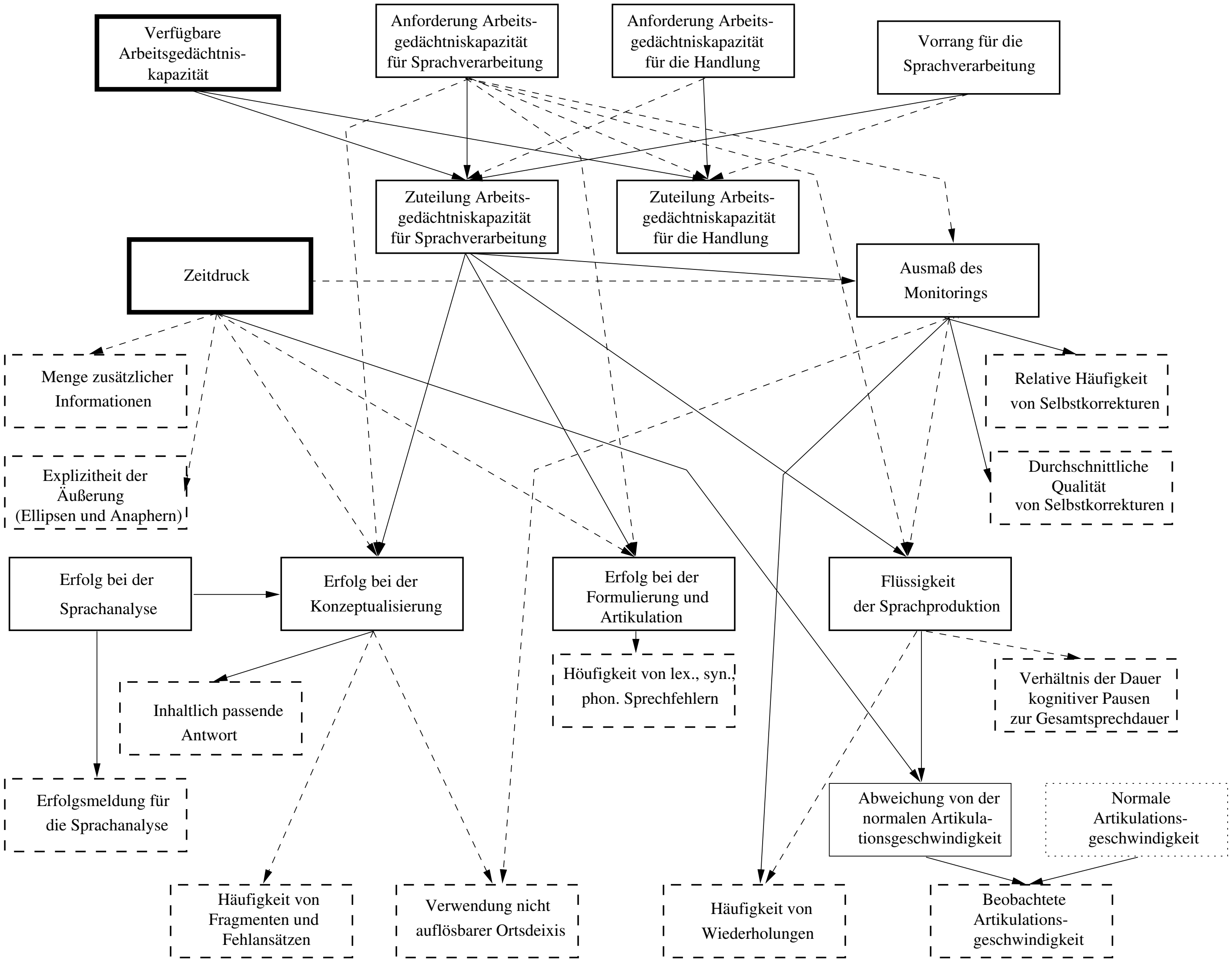


Abbildung 6.14: Das gesamte Bayes'sche Teilnetz zur Sprachproduktion in READY

2. Objektive Wahrscheinlichkeiten, die sich aus einer Theorie ergeben

3. Subjektive Einschätzungen von Experten aufgrund ihrer Erfahrungen

Eine Datenbasis zu grundlegenden Wahrscheinlichkeiten für das Auftreten unterschiedlicher Symptome in der Domäne stellt die Auswertung der Feldstudie (4.3) dar. Hinweise auf das Verhältnis der Werte in einer belastenden und einer nicht belastenden Situation können die konkreten Meßwerte der Arbeiten in Kapitel 3 liefern. Problematisch ist dabei allerdings die Integration der unterschiedlichen Ergebnisse, da jeder Experimentalaufbau mit einer anderen Form von Belastung oder Zeitdruck arbeitete und für die Feldstudie gar kein objektives Maß der Belastung oder des tatsächlich von den Versuchspersonen erlebten Zeitdrucks zur Verfügung steht.

Da nicht sämtliche Zusammenhänge domänenspezifisch neu überprüft werden können, müssen die Experimentalergebnisse aus Literatur und Feldstudie unter speziellen Annahmen in die vorliegende Arbeit integriert werden. Diese speziellen Annahmen sind:

1. Die Kontrollbedingungen der einzelnen Experimente stellen Bedingungen geringer kognitiver Belastung bzw. geringen Zeitdrucks dar.

2. Die Experimentalbedingungen, in denen *Zusatzaufgaben* (Doppelaufgaben) ausgeführt werden mußten, stellen die Bedingungen mit dem höchsten Beanspruchungsgrad unter allen Experimentalbedingungen dar.

3. In der Feldstudie wird durchschnittlich eine mittlere Belastung und mittlerer Zeitdruck angenommen. Da beide im Dialogverlauf variieren, wird ferner angenommen, daß die Durchschnittswerte der Parameter Werten unter mittlerer Belastung entsprechen.

Im folgenden wird an einem Beispiel gezeigt, wie die CPTs mithilfe dieser Annahmen präzisiert und domänenspezifisch angepaßt werden können. Voraussetzung dafür ist eine modifizierte Sichtweise vor allem auf die Schicht der Erfolgsknoten.

### 6.3.2 Qualitative Anpassungen

Das oben vorgestellte Modell (6.2) sagt aus, daß für einen Sprecher unter Belastung sowohl Einbußen in der Qualität seiner Sprachproduktion *als auch* eine Reduktion der Flüssigkeit und des Monitorings vorhergesagt werden. Diese Annahme ist jedoch diskussionsfähig:

Wenn ein Sprecher in einer bestimmten Situation nur über eine geringe verbleibende Arbeitsgedächtniskapazität verfügt, weil eine Nebenaufgabe, situative Ablenkung oder emotionale Belastung diese Kapazität stark einschränken, dann hat er prinzipiell zwei Strategien zur Verfügung, um die Effekte dieser Einschränkung aufzufangen:

- Durch eine Reduktion der Geschwindigkeit der Sprachproduktion wird die durchschnittliche Planungsleistung pro Zeiteinheit reduziert. Dadurch wird pro Zeiteinheit weniger Arbeitsgedächtniskapazität angefordert und das resultierende Defizit ist geringer. Ein Rückgang der Qualität der Sprachproduktion und des Monitorings werden dadurch vermindert.

- Hält ein Sprecher seine Generierungsgeschwindigkeit bei, sind Einbußen der Qualität, z.B. beim Erfolg der Konzeptualisierung, unvermeidlich. Der Sprecher muß mit einem höheren Defizit zwischen angeforderter und zugeteilter Kapazität generieren und erreicht dadurch nur eine verminderte Leistungsfähigkeit, wie es in den Grafiken des Kapitels 2 (Abb. 2.1 und 2.2) angedeutet wurde.

Dieser Gedanke legt nahe, daß die FLÜSSIGKEIT DER SPRACHPRODUKTION, ERFOLG BEI DER KONZEPTUALISIERUNG und AUSMASS DES MONITORINGS nicht unabhängig voneinander sind unter einer vorgegebenen kognitiven Belastung. Berthold und Jameson (1999) sehen den Effekt vor allem so, daß das POTENTIELLE ARBEITSGEDÄCHTNISDEFIZIT zusammen mit einer Anpassung der Generierungsrate der Sprachproduktion dazu führt, daß das TATSÄCHLICHE ARBEITSGEDÄCHTNISDEFIZIT verringert wird. Der Einfluß der Geschwindigkeit auf die Qualität und umgekehrt wird mit einer direkten kausalen Verbindung der beiden Variablen erfaßt (Abb. 6.15).

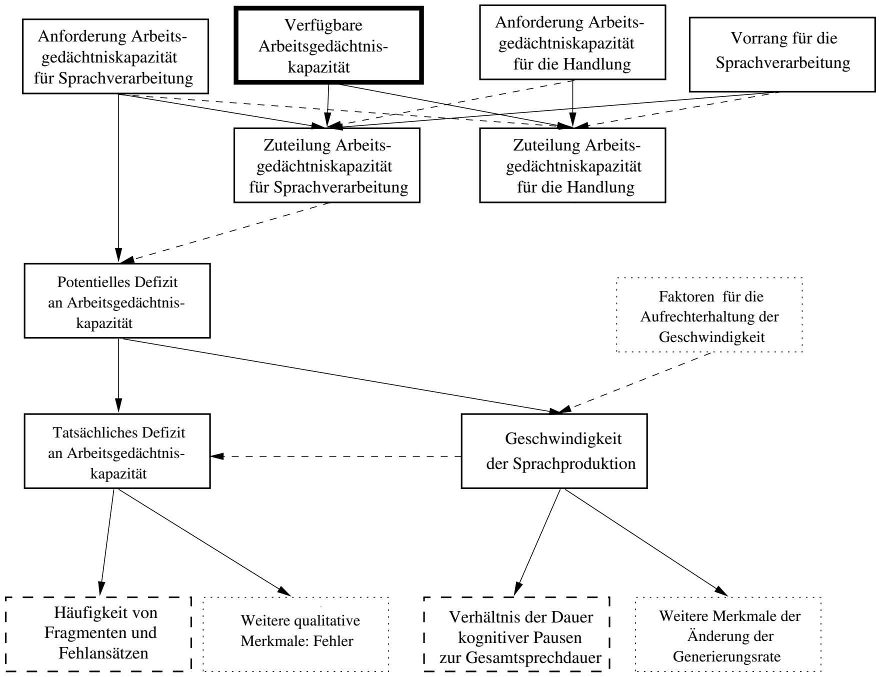


Abbildung 6.15: Veränderte qualitative Aspekte des Modells als Voraussetzung für eine stärker empirisch fundierte quantitative Modellierung

### 6.3.3 Quantitative Struktur: Von empirischen Werten zu CPTs

Am Beispiel der Fragmente und Fehlansätze soll skizziert werden, in welcher Form unter den in 6.3.1 angegebenen Annahmen genauere CPTs gewonnen werden können, als das im Modell aus Abschnitt 6.2 der Fall war.

**Rahmenwerte** Fehlansätze und Satzfragmente werden in den Dialogen der READY-Domäne nach den Daten der Feldstudie nur selten erwartet: 91 % aller Äußerungen enthielten weder ein

Fragment noch einen Fehlansatz.

Verschiedene Autoren haben darauf hingewiesen, daß für bestimmte sprachliche Phänomene die Möglichkeit, sie überhaupt zu beobachten, von der Gesamtäußerungsmenge abhängig ist (siehe z.B. Bromme & Wehner, 1987). Entsprechend wurden im vorgestellten Modell (6.2) und bei der Auswertung der Voruntersuchung in Abschnitt 4.3 die Meßwerte für die Anzahl von Fragmenten und Fehlansätzen einer Relativierung unterzogen. Für Entwurf, Überprüfung und Anpassung des Systemverhaltens machen jedoch zwei Punkte die Situation komplizierter als notwendig:

- Die Äußerungen in der READY-Domäne sind durchschnittlich von sehr kurzer Dauer. Eine Relativierung an der Gesamtäußerungsmenge liefert daher nur selten wirklich wesentliche Unterschiede.
- Die relativierten Werte sind für eine Kontrolle der Inferenzen von READY nur wenig aussagekräftig. Welche Äußerung muß man sich z.B. zu einem Meßwert von 0.05 für die Häufigkeit von Fragmenten vorstellen?

Zusätzlich basieren die Daten, die aus der Feldstudie erhoben werden konnten, auf einem verhältnismäßig kleinen Korpus und verursachen daher vor allem bei einer zu feinkörnigen Modellierung stark stichprobenabhängige Werte. Aus diesen Gründen soll die Modellierung für Symptome der Häufigkeit in READY wie folgt angepaßt werden: Die Beurteilung erfolgt nur noch binär, d.h. mit der Einschätzung 'Symptom vorhanden' (Hypothese 1) vs. 'Symptom nicht vorhanden' (Hypothese 0).[28] Ein zusätzlicher Abbildungsschritt auf das Einheitsintervall entfällt für diese Symptome damit ebenfalls. Tabelle 6.1 gibt die korrigierten Werte von Tabelle 4.13 unter diesen Annahmen wider.

| **Häufigkeit Fragmente** und Fehlansätze **- Verteilungen (korrigiert) -** | | |
|---|---|---|
| **Sprecher** | **0** | **1** |
| Mechaniker | 0.920 | 0.080 |
| Anrufer | 0.914 | 0.086 |
| gesamt | 0.917 | 0.083 |

Tabelle 6.1: Anpassung von Verteilungen aus der Feldstudie

**Beschränkungen der CPT** Für die beiden Hypothesen des Knotens HÄUFIGKEIT VON FRAGMENTEN UND FEHLANSÄTZEN lassen sich die bedingten Wahrscheinlichkeiten für die Erwartung eines Fragments oder eines Fehlansatzes aufgrund der Binärität relativ direkt ableiten:

- Es ergibt sich ein durchschnittliches Verhältnis von 3.3 für die Wahrscheinlichkeit von Fehlansätzen und Fragmenten unter Belastung gegenüber fehlender Belastung. Dieser Wert entsteht aus dem durchschnittlichen Verhältnis jeweils zweier vergleichbarer Werte aus den unterschiedlichen Arbeiten *mit Doppelaufgaben*, die in den Tabellen 3.16 und

[28] Für diesen Vorschlag zur Vereinfachung danke ich Anthony Jameson.

3.17 die überwiegende Tendenz eines Anstiegs der Menge von Fragmenten und Fehlansätzen unter Belastung nachgewiesen haben. In der Sichtweise auf das *tatsächliche Arbeitsgedächtnis-Defizit* läßt sich das als Constraint 6.24 über der CPT des Knotens ausdrücken.

- Die durchschnittliche Häufigkeit von Äußerungen ohne Fragmente und Fehlansätze aus der Feldstudie beträgt 0.92, was mit der Wahrscheinlichkeit dafür gleichgesetzt wird, daß kein Fragment oder Fehlansatz auftritt. Unter Annahme 3 aus 6.3.1 gilt daher Gleichung 6.25.

$$\frac{P(F=1 \mid D=hoch)}{P(F=1 \mid D=niedrig)} = 3.3 \tag{6.24}$$

$$P(F=1 \mid D=mittel) = 1 - 0.92 = 0.08 \tag{6.25}$$

$$\tag{6.26}$$

Zu klären ist weiterhin, in welcher Weise sich die Wahrscheinlichkeiten der Hypothesen für Fragmente und Fehlansätze bei einer Veränderung der Einschätzung ihres Elternknotens ändern. Entsprechende Hinweise liefern auch die vorliegenden empirischen Daten nicht, da die kognitive Belastung in den unterschiedlichen Experimenten nicht mittels einer Maßeinheit vergleichbar sind. Der Anstieg der Wahrscheinlichkeit für Fragmente und Fehlansätze erfolge daher wie bisher mit der Verringerung des Konzeptualisierungserfolges linear (Gleichung 6.27). Die Konstante $c$ moduliert den Anstieg dieser linearen Kurve.[29].

$$P(F=1 \mid D=d_j) = c * (d_j - d_i) * P(F=1 \mid D=d_i=mittel) \tag{6.27}$$

Ein Constraint für die Verteilung der Wahrscheinlichkeiten unter ein und derselben Elternkonfiguration wird wegen Gleichung 6.9 nicht benötigt. Die CPT für den Knoten HÄUFIGKEIT VON FRAGMENTEN UND FEHLANSÄTZEN mit den beiden Hypothesen $f_1$ und $f_2$ ist mit den aus empirischen Daten abgeleiteten Constraints 6.24 bis 6.27 vollständig definiert.

$$P(\mathrm{F}=(0,1) \mid \mathrm{K}=k) = \begin{cases} (0.96 \quad 0.04) & : \quad k \quad = \quad 1 \\ (0.92 \quad 0.08) & : \quad k \quad = \quad 0.5 \\ (0.88 \quad 0.12) & : \quad k \quad = \quad 0 \end{cases} \tag{6.28}$$

### 6.3.4 Fazit

Statt Beschreibungen für einzelne CPTs auf der Basis von Einschätzungen 'subjektiver Experten' abzuleiten, wie es im Modell 6.2 geschehen ist, können diese weitgehend durch empirische Daten und einige daraus gebildete Constraints beschrieben werden. Möglich ist z.B. die folgende Kombination von Constraints:

[29] Bei Constraints dieser Form ist nicht immer gewährleistet, daß die Werte für die einzelnen $f_i$ im Intervall [0;1] liegen. Eine nachträgliche Normalisierung kann daher notwendig sein.

- ein Mittelwert für durchschnittliche Belastungsbedingungen (vgl. Gleichung 6.25),
- ein Verhältnisconstraint für die Zusammenhänge von bedingten Wahrscheinlichkeiten *einer* Hypothese für *verschiedene* Elternkonfigurationen (vgl. Gleichung 6.24)
- eine Funktion, die die Entwicklung der Wahrscheinlichkeiten für diejenige Hypothese angibt, für die der Mittelwert verwendet wurde (vgl. Gleichung 6.27) sowie
- eine Beschränkung über der Wahrscheinlichkeitsverteilung für *verschiedene* Hypothesen unter der *gleichen* Elternkonfiguration (vgl. Gleichung 6.6)

Die Werte und Verhältnisse, welche aus den empirischen Studien und einer Feldstudie abgeleitet werden konnten, stellen keine perfekte Annäherung an die Realität dar und sollen diesen Anspruch auch nicht erheben. Sie geben jedoch eine bestmögliche Annäherung an Werte in realen Situationen an und liefern damit eine adaptierbare Basis für eine korrekte Modellierung.

# Kapitel 7

# Schlußfolgerungen und Ausblick

## 7.1 Schlußfolgerungen

Die vorliegende Arbeit leistet einen umfassenden Beitrag zur Identifikation und Verarbeitung sprachlicher Symptome für die Erkennung von Ressourcenbeschränkungen des Benutzers in Dialogsystemen. Durch die Aufarbeitung relevanter psychologischer Grundlagen wurde ein wichtiger Teil der Basis für die Modellierung zweier Ressourcen des Benutzers gelegt, die bei der Interaktion mit der Umwelt und damit heute mehr und mehr auch mit komplexen Softwaresystemen von besonderer Bedeutung sind. Diese beiden Ressourcen sind die verfügbare Arbeitsgedächtniskapazität und die verfügbare Zeit.

Eine umfangreiche Rückschau auf die sprachpsychologische experimentelle Forschung erlaubte es, wichtige Effekte von kognitiver Belastung und Zeitdruck auf sprachliche Äußerungen zu identifizieren. Durch die Diskussion dieser Effekte in einem Modell der Sprachproduktion wurden die einzelnen experimentellen Ergebnisse in einen einheitlichen theoretischen Rahmen eingeordnet. Damit konnten vor allem widersprüchliche Ergebnisse erkannt und unterschiedlichste Interpretationen der experimentellen Resultate konsistent dargestellt werden. Gleichzeitig zeigte sich, daß die Forschung zum Zusammenhang von Sprachproduktion, Belastung und Zeitdruck ein sehr komplexes Gebiet umfaßt und noch viele bislang wenig beachtete Fragestellungen für die Zukunft bereithält.

Zahlreiche mögliche Symptome wurden in bezug auf ihre Bedeutsamkeit für die Einschätzung kognitiver und außerkognitiver Benutzerressourcen bewertet und damit die wichtigsten Indikatoren für Belastung und Zeitdruck identifiziert. Durch die Auswertung eines Korpus konnten domänenabhängig quantitative Werte abgeleitet und gleichzeitig noch nicht untersuchte mögliche Symptome herausgearbeitet werden. Die Bedeutung einer domänenabhängigen Vorstudie für den Entwurf und die Parametrisierung einer Benutzermodellierungskomponente wurde dabei ebenfalls aufgezeigt.

Vorschläge zur Repräsentation der unterschiedlichen Symptome mündeten in eine Darstellung und Diskussion unterschiedlicher Möglichkeiten der automatischen Erkennung dieser Phänomene. Dabei wurde deutlich, daß nicht alle Methoden der linguistischen Analyse für die Erhebung sprachlicher Indikatoren von Ressourcenbeschränkungen und Zeitdruck geeignet sind.

Am Beispiel der Zusammenarbeit mit dem Projekt LISA wurde für einen existierenden semantischen Verarbeitungsformalismus gezeigt, wie einerseits allgemeine Eigenschaften semantischer Repräsentationen für die Erkennung von Ressourcenbeschränkungen ausgewertet werden können, wie andererseits jedoch auch aus der prozeduralen Verarbeitung spezifischer Systeme und auf der Basis spezieller Mechanismen der Computerlinguistik wichtige Hinweise für die Benutzermodellierung extrahiert werden können.

Mit der Umsetzung der Erkenntnisse aus der empirischen Forschung in einem Computermodell zur Modellierung sprachlicher Symptome von Belastung und Zeitdruck wurde zur Systemarchitektur eines Softwaresystems beigetragen, das in einem Auskunftsdialog die genannten Benutzereigenschaften aufgrund verschiedener Evidenzen erkennen soll. Die möglichen Evidenzen aus der Sprachproduktion wurden in einem Teil der Bayes'schen Netze im Prototypen des READY-Projektes implementiert und getestet. Es wurde gezeigt, daß der gewählte Ansatz auf qualitativen Zusammenhängen der psychologischen Forschung beruht und unter neuen Erkenntnissen leicht adaptierbar ist. Es ergaben sich dabei u.a. auch zahlreiche interessante Fragestellungen für die weitere Forschung.

Schließlich wurde gezeigt, wie individuelle Basisparameter, die eines der Hauptprobleme bei der Modellierung sprachlicher Eigenschaften eines Benutzers darstellen, behandelt werden können. Damit werden trotz starker individueller Unterschiede auch sprecherunabhängige Modellierungen möglich, weil das zugehörige System geeignete Lernverfahren bereits während der Interaktion mit dem Benutzer einsetzen kann und somit nicht auf etwaige Trainingssitzungen angewiesen ist. Diese Fähigkeit ist eine wichtige Voraussetzung für den universellen und mobilen Einsatz ressourcenadaptiver Dialogsysteme in verschiedenen Umgebungen.

## 7.2 Ausblick

Die vorliegende Arbeit hat mit der Integration zahlreicher Basisarbeiten eine der Türen zum Eintritt in ein relativ junges Forschungsgebiet geöffnet. Aus diesem Grund stellen die Erkenntnisse und Modellierungsansätze keine abschließenden Darstellungen oder Aussagen dar. Sowohl neue Erkenntnisse zur internen Organisation der Sprachproduktion, vor allem in bezug auf ihre verschiedenen Repräsentationsebenen und Verarbeitungsmechanismen, als auch zum weiteren Zusammenhang von Sprache, Denken und Handlungsplanung müssen dazu beitragen, diese Ansätze zu verfeinern und die zugrundeliegenden Aussagen zu überprüfen.

Die unterschiedliche Gewichtung von prosodischen und syntaktischen Aspekten gegenüber inhaltlichen Parametern der Sprachproduktion ist vor allem auf eine entsprechende Gewichtung in den experimentellen Vorarbeiten zurückzuführen. Es steht jedoch außer Frage, daß gerade semantische und pragmatische Aspekte der Sprachproduktion noch ein reiches Potential an Indikatoren für kognitive Belastung und Zeitdruck bergen, das es zu erschließen gilt. Zu den Erkenntnissen dieser Arbeit gehört dabei vor allem, daß der Weg dorthin über kontextabhängige Analysen führen muß.

Verzahnt mit der Erweiterung der empirischen Grundlagen können Verfahren entwickelt werden, die eine automatische Analyse der potentiellen Indikatoren für Belastung und Zeitdruck auch bereits in den experimentellen Studien liefern können und dadurch zu vergleichbaren Ergebnissen führen, die derzeit nur für solch 'einfache' und unstrittige Parameter, wie z.B.

die Artikulationsgeschwindigkeit vorliegen. Gleichzeitig stellt die Modellbildung in Systemen der künstlichen Intelligenz beiden Gebieten ein mächtiges Potential an Evaluationsmöglichkeiten zur Verfügung, das ausgenutzt werden sollte. Dazu gehört die Überprüfung von Inferenzen von Systemen durch Plausibilitäts- und Interaktionsstudien mit echten Benutzern. Diese gegenseitige Unterstützung von empirischer Forschung, computerlinguistischer Verarbeitung sowie Evaluation von Erkenntnissen und Überprüfung von Hypothesen mit Mechanismen der Künstlichen Intelligenz kann sehr fruchtbar sein und Gewinne für alle Seiten abwerfen.

Bei der weiteren empirischen Forschung und Modellbildung ist aufgrund der Komplexität des Gebietes eine Konzentration auf abgegrenzte Teilaspekte der Empirie, Erkennung und Modellierung notwendig. Durch lokale Erfolge, wie sie unter anderem in der Kooperation mit LISA entstanden, können Einzelerkenntnisse präzisiert werden und somit schrittweise zur Verbesserung der Benutzermodellierung insgesamt beitragen.

Wichtige Fragestellungen bei diesem Prozeß sind die folgenden:

- Wie können die Möglichkeiten zur Erkennung individueller Basisparameter erweitert werden? Dabei sind vor allem Lernergebnisse durch Korrelationen mit anderen Sprech- und Verhaltensparametern zu präzisieren und zu fundieren.

- Welche Situationsparameter beeinflussen das Benutzerverhalten und seine Anpassungsleistung an Belastung und Zeitdruck? Das erfordert eine klare Trennung zwischen verschiedenen situativen Parametern, die zu unterschiedlichen Formen der Belastung führen bzw. unterschiedlich starke Effekte bei und Anpassungsleistungen von Sprechern verursachen.

- Wie können Erkenntnisse zu kognitiven Subprozessen für die Verarbeitung zugänglich gemacht werden? Dieser Punkt umfaßt eine Verfeinerung der Granularität der Zeitscheiben in READY auf lokale Ebenen unterhalb des Äußerungsniveaus.

- Welche Möglichkeiten der Anytime-Analyse erlauben es einem Dialogsystem bereits zum frühestmöglichen Zeitpunkt Erkenntnisse über die Benutzerressourcen zu erlangen und damit seine eigenen Ressourcen gezielt für die Berücksichtigung möglicher Beschränkungen einzusetzen? Partielle Evidenzen aus einer inkrementellen Verarbeitung, wie sie im Laufe dieser Arbeit bereits verschiedentlich angedeutet wurden, können dieses Vorgehen unterstützen.

- Auf welche Weise können die Modelle von Ursachen und Symptomen sowie die zugehörigen Erkennungsmethoden auf andere Eingabemodalitäten als die Spontansprache verallgemeinert werden? Dazu gehören z.B. handschriftliche oder Tastatur-Eingaben von Sprache, Auswahleingaben aus Menü- oder anderweitig vorstrukturierten Präsentationen. Dieser Punkt wird vor allem in der zweiten Förderungsperiode des Projektes READY (Wahlster & Jameson, 1998a) im Zentrum des Interesses stehen.

- Können aus größeren, domänenspezifischen Korpora Parameter oder Strukturen Bayes'scher Netze gelernt werden? Auch dieser Punkt gehört zu den Zielen des Fortsetzungsprojektes zu READY.

- Welche Symptome erweisen sich als am aussagekräftigsten bzw. welche Symptome können in der Zukunft vernachlässigt werden? Dafür ist eine Evaluierung von Systemleistungen in experimentellen Studien mit den bis dahin entwickelten Prototypen notwendig, in denen die tatsächliche kognitive Belastung des Benutzers kontrolliert und damit mit den Systemeinschätzungen verglichen werden kann.

In vielen der genannten Gebiete bestehen bereits gutdurchdachte Ansätze, die auf neue Einsatzgebiete nur warten. Ich hoffe, mit dieser Arbeit auch für die Beschleunigung dieses Prozesses einen Beitrag geleistet zu haben.

# Anhang A

# Kurzbeschreibung der Dialoge

Im folgenden wird eine Kurzbeschreibung der Dialoge aus der Feldstudie gegeben. Die transliterierten und annotierten Dialoge können im Internet eingesehen werden (Berthold, 1998). Ziel der Darstellung ist es einerseits, dem Leser die Einordnung der im Text angegebenen Beispiele in den jeweiligen Dialogverlauf zu ermöglichen. Andererseits soll damit auch auf die Schwierigkeiten hingewiesen werden, die gleiche Instruktionen bei verschiedenen Anrufern bzw. unterschiedliche Instruktionen beim gleichen Anrufer verursachten. In Klammern hinter Aktionen oder Instruktionen sind die Nummern des den Abschnitt eröffnenden und schließenden Dialogturns angegeben. Dadurch ist es möglich, die Länge des entsprechenden Dialogabschnitts ohne genaue Kenntnis des Dialoges in etwa abzuschätzen.

**Dialog 1:** P berichtet S, daß es im Innenraum seines Fahrzeugs nach Benzin riecht (001). Es folgt eine kurze Verständigung über das Ziel des Anrufes (003-010). Die Motorhaube ist bereits geöffnet, daher läßt S P den Benzinfilter suchen (010-026), dann die Befestigung der Leitungen überprüfen (027-039). P kann beides nicht ohne weiteres ausführen. Da alle Schellen festgezogen sind, soll P den Luftfilter suchen (040). P meldet, es sei Benzin am Benzinfilter (041-046). Die Aufforderung, die Befestigung des Benzinfilters zu überprüfen, quittiert P mit der Antwort, er habe ihn wieder richtig dazwischengesteckt (046-051). Damit ist die Panne behoben. Danach folgt die Klärung technischer Probleme der Kommunikation (052-065).

**Dialog 2:** P berichtet von einem Klackern am linken Vorderreifen, das beim Bremsen und Gasgeben geringer wird (000-009). Die Radkappen sind bereits abgenommen, nach einigem Nachdenken und Nachfragen (009-017) läßt S P die Radschrauben überprüfen. Es ergibt sich, daß drei davon locker sind, P zieht sie sofort fest und bestätigt das (017-030). Der Mechaniker gibt einen Hinweis auf eine weitere mögliche Ursache, falls das Geräusch bestehen bleibt. Die Panne ist behoben. Das Zeitlimit von 4 Minuten wurde eingehalten.

**Dialog 3:** P berichtet, daß seine Scheibe schmutzig ist, sein Scheibenwischer nicht funktioniert und kein Wasser aus der Waschanlage kommt (001-009). S läßt P daraufhin den Sicherungskasten im Fahrerraum suchen. Nach mehreren vergeblichen Anläufen findet ihn P (010-023). P wird zur Überprüfung der Sicherungen aufgefordert, was er erst nach Rückfragen ausführen kann (029-040). Der Dialog wird vorzeitig abgebrochen, ohne daß die Panne behoben wurde.

**Dialog 4:** P berichtet von einem Ruckeln beim Fahren, mangelnder Beschleunigung und davon, daß das Fahrzeug ab und zu ausgeht (002-008). S läßt die Motorhaube öffnen, was P erst nach mehreren Rückfragen erfolgreich ausführen kann (009-026). S läßt daraufhin die Befestigung von Kabeln an einer 'orangenen Kappe' überprüfen (problemlos gefunden und ausgeführt, 027-035). P befestigt eines der Kabel sofort wieder und kommt schnell der Aufforderung zum Starten nach (036-040). Die Panne ist behoben. Das Zeitlimit von 4 Minuten wurde eingehalten.

**Dialog 5:** P berichtet von einer verschmutzten Scheibe, funktionierendem Scheibenwischer, jedoch mangelnder Wasserzufuhr (003-010). Die Motorhaube ist bereits geöffnet. S läßt P den Waschwasserbehälter suchen, was nicht sofort klappt (013-022). Anschließend soll das Kabel, das an dem Behälter befestigt ist, überprüft werden (problemlos, 023-029). Das Kabel ist fest. Daher läßt S P den Wasserstand im Behälter überprüfen (schnell erledigt, 030-031a). Der Behälter ist leer und S fordert P auf, Wasser nachzufüllen. Das braucht Zeit und Rückfragen (032-043). Danach soll P starten und die Scheibenwaschanlage ausprobieren (problemlos erledigt, 047-064). Die Panne ist behoben. Das Zeitlimit von drei Minuten wurde eingehalten.

**Dialog 6:** P berichtet von aus der Motorhaube aufsteigendem Wasserdampf und einem roten Lämpchen, das im Cockpit leuchtet (001-007). S läßt P die Motorhaube öffnen (keine Probleme, 008-009), einen „schwarzen Deckel"(Kühler) suchen (keine Probleme, 010-013) und überprüfen, ob dieser verschlossen ist (020-021). P stellt schnell fest, daß das der Fall ist, es jedoch beim Aufdrehen zischt (024-028a). P soll langsam weiter aufdrehen (029-038). Der Aufforderung von S, die Schläuche zu überprüfen, kommt P problemlos nach. Alle sind befestigt (039-046). Da es inzwischen aufhört, zu dampfen, läßt S P den Wasserstand im Kühler überprüfen (problemlos erledigt, 055-060). Der Wasserstand ist sehr niedrig. Daher soll P vorsichtig Wasser auffüllen. Das braucht etwas Zeit (063-095). Der Deckel ist wieder zu schließen (schnell erledigt, 108-109). Ein Startversuch (problemlos, 110-113) ergibt, daß die Panne behoben ist.

**Dialog 7:** P berichtet von schwächer werdendem Licht, ausgehendem Radio und liegenbleibendem Fahrzeug (001). S läßt P die Motorhaube öffnen (ohne Rückfragen, 004-011). P soll den Keilriemen suchen, den er sofort findet (012-013), und dessen Sitz überprüfen. P stellt schnell fest, daß er fest sitzt (014-021). Daher soll P die Stecker an der Lichtmaschine überprüfen, kann jedoch mit dem Begriff Lichtmaschine nichts anfangen (022-023). Nach erneuter Beschreibung durch S findet er sie schnell (024-025). Die Überprüfung dauert nicht lange (026-027). Da ein Stecker lose ist, soll er wieder befestigt werden (ab 030). Das Gespräch bricht wegen technischer Probleme nach 2 Minuten ab (was der Dauer des Zeitlimits entspricht). Die Panne wurde nicht behoben.

**Dialog 8:** P berichtet von einem blinkenden Lämpchen im Cockpit. S erfragt weitere Informationen und fordert zum Öffnen der Motorhaube auf (003-006). P weiß nicht, wie das geht, daher läßt S ihn einen Hebel im Fußraum suchen. Mehrere verschiedene Beschreibungen sind notwendig, bis P den Hebel findet (008-025). Vor dem Fahrzeug stehend, braucht P dann weitere Unterstützung beim Öffnen der Haube. Nach Verständigungsproblemen zeigt sich, daß P die Motorhaube geöffnet hat und bereits Beschreibungen zum Motorraum abgegeben hat, während S weiter versuchte, ihn beim Öffnen zu unterstützen (026-049). P nimmt selbständig Manipulationen am Fahrzeug vor, die S irritieren (049-054). Als S P daraufhin nach dem Ölmeßstab suchen läßt, läuft das Zeitlimit ab (058). Der Dialog wird fortgesetzt: P findet nach mehreren alternativen Beschreibungen von S und vielen Rückfragen den Ölmeßstab (054-172). Technische Probleme erschweren außerordentlich die Kommunikation. P soll den Stab herausziehen, abputzen und wieder einstecken. Das macht leichte Probleme (175-199). Der Ölstand ist niedrig, daher soll P Motoröl auffüllen. Währenddessen bricht das Gespräch ab (198-217). Die Panne wurde nicht behoben, das Zeitlimit überschritten.

# Anhang B

# Technische Details zur Dokumentation des Modells in Kapitel 6

## B.1 Einheitsintervall

Die Instantiierung der Bayes'schen Netze in der Benutzermodellierungskomponente von READY soll unabhängig von konkreten Maßeinheiten möglich sein. Dafür werden alle Meßwerte in das Einheitsintervall [0,1] abgebildet. Die Abbildung liefert Funktion B.1, wobei die Parameter `minimal` und `maximal` mit spezifischen Rahmenwerten und `value` mit dem aktuellen Wert des Symptoms instantiiert werden. Beispiele für Rahmenwerte sind in Beispiel B.2 für den Parameter FRAGMENTE UND FEHLANSÄTZE angegeben. In der Dokumentation der Knoten des Modells in B.5 werden diese Umrechnungswerte als Wert des Knoten-Attributs *Abbildung* angegeben. Die Rahmenwerte lassen sich aus der Feldstudie (Kap. 4.3) und der experimentellen Literatur ableiten. Sie werden für jedes Symptom so festgelegt, daß sich die Meßwerte für das Symptom bei möglichst nur einem geringen Teil aller Äußerungen, sowohl ohne als auch mit Belastung und/oder Zeitdruck, außerhalb dieser Werte ansiedeln.

(B.1)
```
(defun standard-rel-function (minimal maximal value)
          (if (> value minimal)
                    (/ (- (max value minimal) minimal)
                       (- maximal minimal))
               0))
```

(B.2)
```
(defparameter *min-frequency-of-fragments-and-false-starts* 0.050)
(defparameter *max-frequency-of-fragments-and-false-starts* 0.200)
```

## B.2 Diskretisierung

Da die Normalverteilung für *stetige* (*kontinuierliche*) Zufallsgrößen definiert ist, jedoch in READY durch die Intervallaufteilung eine Modellierung mit *diskreten* Größen erfolgt, müssen statt echter Funktionswerte die Flächen unter den Kurven innerhalb eines Teilintervalls als Funktionswerte approximiert werden. Dafür können im allgemeinen Mechanismen der Integralrechnung verwendet werden. Um die Normierung eines Zeilenvektors der zugehörigen CPT auf 1 zu erreichen, müssen die so berechneten Werte anschließend normiert werden.

Aus diesem Grund werden alle Meßwerte für Einzelsymptome nach der Umrechnung ins Einheitsintervall auf Intervallwerte abgebildet. Der dabei entstehende Informationsverlust ist vernachlässigbar, sofern eine ausreichend hohe Anzahl von Hypothesen zur Verfügung steht, die mit ihrer Verteilung empirisch gut genug abgesichert sind.

**Komplexität in READY**
Der Rechenaufwand zum Lösen Bayes'scher Netze ist u.a. von der Anzahl an Hypothesenkombinationen in allen lokalen Teilnetzen abhängig. Die Zahl dieser Kombinationen bei mehreren Elternknoten wächst exponentiell in der Anzahl der Elternknoten auf der Basis der Hypothesenmengen. Im READY-System wird daher die maximale Anzahl von Hypothesen für Knoten in Abhängigkeit von ihrer Stellung im Netz definiert. Für Knoten, die die einzuschätzenden Benutzereigenschaften modellieren (Allokationsmodell und Zeitdruck) beträgt dieser Wert 5, für alle anderen Knoten sind derzeit maximal 3 Hypothesen möglich. Die Granularitäten aus der Feldstudie liegen meist über diesem Wert. Daher müssen geeignete Teilintervalle zusammengefaßt werden.

## B.3 Explizite Form von CPTs

Die bedingten Wahrscheinlichkeiten für Knoten eines Bayes'schen Netzes bilden eine Matrix, wobei in der CPT eines Knotens $\kappa$ jede Zeile für eine einzelne Kombination von Hypothesen der Elternknoten steht und die Zellen der Zeile die Wahrscheinlichkeiten für die unterschiedlichen Hypothesen von $\kappa$ angeben. Eine CPT in Matrixform für das Netz in Abbildung B.1 hat z.B. das Aussehen von Tabelle B.1.

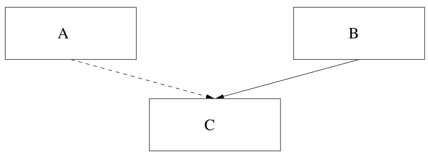


Abbildung B.1: Beispiel für ein einfaches Bayes'sches Netz mit der CPT von Tabelle B.1

| Eltern | | Kind | | | |
|---|---|---|---|---|---|
| A | B | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
| $a_1$ | $b_1$ | $p_{11}$ | $p_{12}$ | $p_{13}$ | $p_{14}$ |
| $a_1$ | $b_2$ | $p_{21}$ | $p_{22}$ | $p_{23}$ | $p_{24}$ |
| $a_1$ | $b_3$ | $p_{31}$ | $p_{32}$ | $p_{33}$ | $p_{34}$ |
| $a_2$ | $b_1$ | $p_{41}$ | $p_{42}$ | $p_{43}$ | $p_{44}$ |
| $a_2$ | $b_2$ | $p_{51}$ | $p_{52}$ | $p_{53}$ | $p_{54}$ |
| $a_2$ | $b_3$ | $p_{61}$ | $p_{62}$ | $p_{63}$ | $p_{64}$ |

Tabelle B.1: CPT in Matrixform für das Netz in Abb. B.1

Der Knoten C hat die Hypothesen $(c_1, c_2, c_3, c_4)$ sowie zwei Elternknoten A und B mit den Hypothesenmengen $(a_1, a_2, a_3)$ bzw. $(b_1, b_2)$. Betrachtet man die Zeilen einer CPT als Vektoren, so gibt jedes Vektorelement die Wahrscheinlichkeit für das Eintreffen der dem Vektorelement zugeordneten Hypothese an unter der Voraussetzung, daß die in der Zeile angegebene Hypothesenkombination für die Elternknoten eingetreten ist. Die Zeile der Matrix mit der Kombination $(a_2, b_1)$ bildet also einen vierstelligen Vektor $(p_{41}, p_{42}, p_{43}, p_{44})$, wobei die $p_{4i}$ für die folgenden bedingten Wahrscheinlichkeiten stehen:

$$p_{4i} = P(\mathsf{C} = c_i \mid \mathsf{A} = a_2, \mathsf{B} = b_1) \tag{B.3}$$

# B.4 Intervallwerte und Erwartungswert

## B.4.1 Ordnung von Hypothesen

Um die notwendigen numerischen Werte für die Berechnung von Erwartungswerten zu erhalten, werden die Hypothesen aller Knoten in READY auf diskrete Werte aus dem Einheitsintervall [0;1] abgebildet. Der numerische Wert $n_i$ (i $\geq$ 0) einer Hypothese $b_i$ eines Knotens B ergibt sich aus einer Ordnung $O$ der k Hypothesen von B, wobei $i$ die Position der Hypothese in der Ordnung $O$ sei (Formel B.4). Den Hypothesen einer Ordnung auf einer Menge der Mächtigkeit 3 werden durch die obige Rechenvorschrift z.B. die Werte 0, 0.5 und 1 zugeordnet. Diese Werte werden **Intervallwerte** genannt.

$$n_i = i * \frac{1}{k-1} \tag{B.4}$$

Als Ordungskriterium kommt beispielsweise die Relation $<$ auf den ordinalen Bezeichnungen numerischer Werte für die Häufigkeit eines Parameters in Frage. Für einen Knoten mit 3 Hypothesen ist eine mögliche Ordnung z.B. <*niedrig*, *mittel*, *hoch*>. Die Aufzählung der Hypothesen der unterschiedlichen Knoten des Sprachproduktionsmodells in Abschnitt B.5 erfolgt nach der jeweils zugrundegelegten Ordnung. Die numerischen Werte können damit direkt nach der oben angegebenen Vorschrift zugeordnet werden.

Tabelle B.2 gibt ein Beispiel, wie den unterschiedlichen Hypothesen des Knotens HÄUFIGKEIT VON FRAGMENTEN UND FEHLANSÄTZEN Intervallwerten zugeordnet werden.

| **Häufigkeit Fragmente und Fehlansätze - Intervalle -** | | |
|---|---|---|
| **Wert** | **Intervallwert** | **Bezeichnung** |
| $<$ 1/20 Silben | 0 | niedrig |
| 1/20 Silben bis 3/20 Silben | 0.5 | mittel |
| $>$ 3/20 Silben | 1 | hoch |

Tabelle B.2: Abbildung von Parameter-Intervallen auf diskrete Werte des Normintervalls [0;1]

### B.4.2 Erwartungswert

Der Erwartungswert $E_K$ eines Knotens K bietet eine Möglichkeit, die Einschätzung der Wahrscheinlichkeitsverteilung für K durch ein einzelnes Maß darzustellen. Er wird berechnet durch die Summe B.5.[1]

$$E_K = \sum_{i=1}^{k} p(b_i) * n_i \tag{B.5}$$

## B.5 Dokumentation der Knoten des Sprachproduktionsmodells

Dieser Abschnitt beschreibt die konkreten technischen Daten des Modells aus Abbildung 6.14 in Kapitel 6. Zugrundegelegt für einen Eintrag wird die folgende Struktur:

### Legende

| NAME DES KNOTENS ($\mathcal{K}$urzform) | Status |
|---|---|
| Eltern: | NAMEN DER ELTERNKNOTEN ($\mathcal{E}$), |
| Zustände: | Anzahl der Zustände (Bezeichnungen der Zustände) |
| CPT: | |
| Einfluß: | Einflußfunktion der Elternknoten |
| Verteilung: | Verteilungsfunktion für 1 Hypothesenkombination der Eltern |
| Beschreibung: | wichtige auf die Variable eingehende Abschnitte der Arbeit |
| Maßeinheit: | Maßeinheit (nur bei Symptomen) |
| Typ: | Z = zusammengesetzter Parameter, B = Basisparameter (nur für Symptome) |
| Slot/Methode: | `Slot in den 'Utterance-Objekten (bei Basisparametern) oder Berechnungsmethode (bei zusammengesetzten Parametern)` |
| Berechnung: | `Zur Berechnung verwendete Parametermethoden/-slots (nur bei zusammengesetzten Parametern)` |
| Abbildung: | Umrechnungsparameter für das Einheitsintervall |

Die unter '*Beschreibung*' angegebenen Teile der Arbeit umfassen nur die wichtigsten Unterabschnitte, in denen jeweilige Variable genauer behandelt wurde. Die Rahmenwerte für die Abbildung des Einheitsintervalls sind Kapitel 4 entnommen. Wegen der Einschränkungen bei der Implementation war es nicht immer möglich, die dort ebenfalls empfohlenen Intervallgranularitäten zu verwenden.

[1] Da es sich beim Erwartungswert um ein statistisches Maß der Verteilung handelt, kann für $E_K$ auch eine Standardabweichung angegeben werden.

| FLÜSSIGKEIT DER INKREMENTELLEN VERARBEITUNG BEI DER SPRACHPRODUKTION ($\mathcal{F}$) | Zwischenvar. |
|---|---|
| Eltern: | ANFORDERUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{A}$), ZUTEILUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{C}$), AUSMASS DES MONITORINGS ($\mathcal{M}$), |
| Zustände: | 3 (hoch,mittel,niedrig) |
| CPT: | |
| Einfluß: | $I_K(a, c, m) = \mathtt{min}(\mathtt{allok\text{-}defizit}(a, c) - \mathtt{fac} * (m - m_0)), 1)$<br>fac = Einfluß des Monitorings auf Erfolg, $m_0$ = Monitoring- Normalwert |
| Verteilung: | $P(F = f \mid A = a, C = c, M = m) = \mathcal{N}_{\mu,\sigma}(f)$<br>$\mu = I_F(a, c, m),\ \sigma = 0.1$ |
| Beschreibung: | 2.4.3; 3.2; 6.2.1.2 |

| ERFOLG BEI FORMULIERUNG UND ARTIKULATION ($\mathcal{F}$) | Zwischenvar. |
|---|---|
| Eltern: | ANFORDERUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{A}$), ZUTEILUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{C}$), ZEITDRUCK ($\mathcal{Z}$), AUSMASS DES MONITORINGS ($\mathcal{M}$), |
| Zustände: | 3 (hoch,mittel,niedrig) |
| CPT: | |
| Einfluß: | $I_{EA}(a, c, z, m) = \mathtt{min}(\mathtt{allok\text{-}defizit}(a, c) * (1 - \mathtt{fac} * z) + m), 1)$<br>fac = Einfluß des Zeitdrucks auf Erfolg |
| Verteilung: | $P(F = f \mid A = a, C = c, Z = z, M = m) = \mathcal{N}_{\mu,\sigma}(f)$<br>$\mu = I_{EA}(a, c, z, m),\ \sigma = 0.2$ |
| Beschreibung: | 2.4.3; 3.3; 6.2.1.2 |

| ERFOLG BEI DER KONZEPTUALISIERUNG ($\mathcal{K}$) | Zwischenvar. |
|---|---|
| Eltern: | ANFORDERUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{A}$), ZUTEILUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{C}$), ZEITDRUCK ($\mathcal{Z}$), AUSMASS DES MONITORINGS ($\mathcal{M}$), ERFOLG DER SPRACHANALYSE (T-1) ($\mathcal{E}$), |
| Zustände: | 3 (hoch, mittel, niedrig) |
| CPT: | |
| Einfluß: | $I_K(a, c, z, m, e) = \mathtt{min}((e * \mathtt{allok\text{-}defizit}(a, c) * (1 - \mathtt{fac} * z)) + m), 1)$<br>fac = Einfluß des Zeitdrucks auf Erfolg |
| Verteilung: | $P(K = k \mid A = a, C = c, Z = z, M = m, E_{t-1} = e) = \mathcal{N}_{\mu,\sigma}(d)$<br>$\mu = I_K(a, c, z, m, e),\ \sigma = 0.2$ |
| Beschreibung: | 2.4.3; 3.2; 3.5; 6.2.1.2 |

| AUSMASS DES MONITORINGS ($\mathcal{M}$) | Zwischenvar. |
|---|---|
| Eltern: | ANFORDERUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{A}$), ZUTEILUNG AGK FÜR DIE SPRACHPRODUKTION ($\mathcal{C}$), ZEITDRUCK ($\mathcal{Z}$), |
| Zustände: | 3 (hoch,normal,niedrig) |
| CPT: | |
| Einfluß: | $I_M(a, c, z) = \frac{\mathtt{allok\text{-}defizit}(a, c) + \mathtt{fac} * z}{1 + \mathtt{fac}}$<br>fac = Einfluß des Zeitdrucks auf Monitoring |

| | |
|---|---|
| Verteilung: | $P(M = m \mid A = a, C = c, Z = z) = \mathcal{N}_{\mu,\sigma}(m)$ <br> $\mu = I_F(a, c, m),\ \sigma = 0.1$ |
| Beschreibung: | 2.4.3; 3.3.4; 3.6; 6.2.1.2 |

| ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT ($\mathcal{A}$) | Zwischenvar. |
|---|---|
| Eltern: | FLÜSSIGKEIT DER INKREMENTELLEN VERARBEITUNG BEI DER SPRACHPRODUKTION ($\mathcal{F}$), ZEITDRUCK ($\mathcal{Z}$), |
| Zustände: | 3 (positiv, keine, negativ) |
| CPT: | |
| Einfluß: | z.Zt. explizite Matrix |
| Verteilung: | z.Zt. explizite Matrix |
| Beschreibung: | 3.2.1; 4.3.3.1; 6.2.6.5 |
| Maßeinheit: | Silben/s |

| BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT ($\mathcal{B}$) | Symptom |
|---|---|
| Eltern: | NORMALE ARTIKULATIONSGESCHWINDIGKEIT ($\mathcal{N}$), ABWEICHUNG VON DER NORMALEN ARTIKULATIONSGESCHWINDIGKEIT ($\mathcal{A}$), |
| Zustände: | 3 (niedrig, mittel, hoch) |
| CPT: | |
| Einfluß: | $I_B(n, a) = (min(n + a - 0.5), 1)$ |
| Verteilung: | $P(B = b \mid N = n, A = a) = \mathcal{N}_{\mu,\sigma}(b)$ <br> $\mu = I_B(n, a), \sigma = 0.1$ |
| Beschreibung: | 3.2.1; 4.3.3.1; 6.2.6.4 |
| Maßeinheit: | Silben/s |
| Typ: | Z |
| Slot/Methode: | `speed, (articulation-rate)` |
| Berechnung: | `number-of-syllables, duration, all-pauses` |
| Abbildung: | niedrig = 2 Silben/s, hoch = 8 Silben/s |

| NORMALE ARTIKULATIONSGESCHWINDIGKEIT ($\mathcal{N}$) | Basiswert |
|---|---|
| Eltern: | − |
| Zustände: | 3 (niedrig, mittel, hoch) |
| CPT: | |
| Einfluß: | −(a-priori) |
| Verteilung: | $P(N = n) = \mathcal{N}_{\mu,\sigma}(n)$ <br> $\mu = 0.67, \sigma = 0.317$ |
| Beschreibung: | 3.2.1; 4.3.3.1; 6.2.6.3 |
| Maßeinheit: | Silben/s |
| Typ: | Z |
| Berechnung: | `− (wird gelernt)` |
| Abbildung: | niedrig = 2 Silben/s, hoch = 8 Silben/s |

| DURCHSCHNITTLICHE QUALITÄT VON SELBSTKORREKTUREN ($\mathcal{Q}$) | Symptom |
|---|---|
| Eltern: | AUSMASS DES MONITORINGS ($\mathcal{M}$), |
| Zustände: | 3 (schlecht, mittel, gut) |
| CPT: | |
| Einfluß: | $I_Q(m) = m$ |
| Verteilung: | $P(Q = q \mid M = m) = \mathcal{N}_{\mu,\sigma}(m)$ <br> $\mu = I_Q(m),\ \sigma = 0.2$ |
| Beschreibung: | 3.3.4, 4.3.4.2, 5.3.4.3, 6.2.5 |
| Maßeinheit: | $\in$ [0;1] |
| Typ: | Z |
| Slot/Methode: | `quality-of-self-corrections` |
| Berechnung: | `number-of-wellformed-self-corrections,` <br> `number-of-editing-terms-in-self-corrections,` <br> `number-of-self-corrections` |
| Abbildung: | schlecht = 0, gut = 1 |

| EXPLIZITHEIT DER ÄUßERUNG (ELLIPSEN + ANAPHERN) ($\mathcal{E}$) | Symptom |
|---|---|
| Eltern: | ZEITDRUCK ($\mathcal{Z}$), |
| Zustände: | 3 (gering, mittel, hoch) |
| CPT: | |
| Einfluß: | $I_E(z) = z$ |
| Verteilung: | $P(E = e \mid Z = z) = \mathcal{N}_{\mu,\sigma}(z)$ <br> $\mu = I_E(z),\ \sigma = 0.5$ |
| Beschreibung: | 3.4.2; 4.3.5.2; 5.3.3 |
| Maßeinheit: | $\in$ [0;1] |
| Typ: | Z |
| Slot/Methode: | `anaphora-and-ellipsis` |
| Berechnung: | `sem-info, gesetzter Wert` |
| Abbildung: | gering = 0, hoch = 1 |

| RELATIVE HÄUFIGKEIT VON FEHLANSÄTZEN ($\mathcal{F}$) | Symptom |
|---|---|
| Eltern: | ERFOLG BEI DER KONZEPTUALISIERUNG ($\mathcal{K}$), |
| Zustände: | 3 (wenig, mäßig viele, viele) |
| CPT: | |
| Einfluß: | $I_F(k) = 1 - k$ |
| Verteilung: | $P(F = f \mid K = k) = \mathcal{N}_{\mu,\sigma}(k)$ <br> $\mu = I_F(k),\ \sigma = 0.5$ |
| Beschreibung: | 3.3.6; 3.4.1; 4.3.5.1; 5.2.2; 6.2.3 |
| Maßeinheit: | Anzahl/Silbe |
| Typ: | Z |
| Slot/Methode: | `frequency-of-fragments-and-false-starts` |
| Berechnung: | `number-of-syllables, number-of-fragments,` <br> `number-of-false-starts` |
| Abbildung: | wenig = 0.05, viele = 0.2 |

| HÄUFIGKEIT VON SELBSTKORREKTUREN ($\mathcal{V}$) | Symptom |
|---|---|
| Eltern: | AUSMASS DES MONITORINGS ($\mathcal{M}$), |
| Zustände: | 3 (wenig, mäßig viele, viele) |
| CPT: | |
| Einfluß: | $I_V(m) = m$ |
| Verteilung: | $P(V = v \mid M = m) = \mathcal{N}_{\mu,\sigma}(v)$ |
| | $\mu = I_V(m),\ \sigma = 0.2$ |
| Beschreibung: | 3.3.4, 4.3.4.2 |
| Maßeinheit: | Anzahl/Silbe |
| Typ: | Z |
| Slot/Methode: | `frequency-of-self-corrections` |
| Berechnung: | `number-of-self-corrections, number-of-syllables` |
| Abbildung: | wenig = 0.05, viele = 0.15 |

| PHONOLOGISCHE, SYNTAKTISCHE UND LEXIKALISCHE SPRECHFEHLER? ($\mathcal{S}$) | Symptom |
|---|---|
| Eltern: | ERFOLG BEI FORMULIERUNG UND ARTIKULATION ($\mathcal{F}$), |
| Zustände: | 2 (ja, nein) |
| CPT: | |
| Einfluß: | $I_S(f) = 1 - f$ |
| Verteilung: | $P(S = s \mid F = f) = \mathcal{N}_{\mu,\sigma}(s)$ |
| | $\mu = I_S(f),\ \sigma = 0.5$ |
| Beschreibung: | 3.2.8; 3.3.2; 3.5.2; 3.3.4; 4.3.7; 6.2.7 |
| Maßeinheit: | $\in \{0{,}1\}$ |
| Typ: | Z |
| Slot/Methode: | `frequency-of-errors` |
| Berechnung: | `number-of-syllables, number-of-errors` |
| Abbildung: | nein = 0, ja = 1 |

| RELATIVE HÄUFIGKEIT VON WIEDERHOLUNGEN ($\mathcal{W}$) | Symptom |
|---|---|
| Eltern: | FLÜSSIGKEIT DER INKREMENTELLEN VERARBEITUNG BEI DER SPRACHPRODUKTION ($\mathcal{F}$), AUSMASS DES MONITORINGS ($\mathcal{M}$), |
| Zustände: | 3 (wenig, mäßig viele, viele) |
| CPT: | |
| Einfluß: | $I_W(f, m) = 0.4 * f + 0.6 * m$ |
| Verteilung: | $P(W = w \mid F = f, M = m) = \mathcal{N}_{\mu,\sigma}(w)$ |
| | $\mu = I_W(f, m),\ \sigma = 0.4$ |
| Beschreibung: | 3.3.4; 3.3.5; 4.3.4.3; 5.2.2; 6.2.4 |
| Maßeinheit: | Anzahl/Silbe |
| Typ: | Z |
| Slot/Methode: | `frequency-of-repetitions` |
| Berechnung: | `number-of-syllables, number-of-repetitions` |
| Abbildung: | wenig = 0.05 , viele = 0.2 |

| INHALTLICH PASSENDE ANTWORT ($\mathcal{I}$) | Symptom |
|---|---|
| Eltern: | ERFOLG BEI DER KONZEPTUALISIERUNG ($\mathcal{E}$), |
| Zustände: | 3 (passend, mäßig passend, unpassend) |
| CPT: | |
| Einfluß: | $I_I(k) = 0.25 + 0.75 * k$ |
| Verteilung: | $P(I = i \mid K = k) = \mathcal{N}_{\mu,\sigma}(i)$ <br> $\mu = I_I(k),\ \sigma = 0.2$ |
| Beschreibung: | 3.5.7; 4.3.6.3 |
| Maßeinheit: | $\in$ [0;1] |
| Typ: | Z |
| Slot/Methode: | `quality-of-content` |
| Berechnung: | `- (wird gesetzt)` |
| Abbildung: | unpassend = 0, passend = 1 |

| VERHÄLTNIS DER DAUER KOGNITIVER PAUSEN ZUR GESAMTSPRECHDAUER ($\mathcal{K}$) | Symptom |
|---|---|
| Eltern: | FLÜSSIGKEIT DER INKREMENTELLEN VERARBEITUNG BEI DER SPRACHPRODUKTION ($\mathcal{F}$), |
| Zustände: | 3 (hoch, mittel, niedrig) |
| CPT: | |
| Einfluß: | $I_K(f) = 1 - f$ |
| Verteilung: | $P(K = k \mid F = f) = \mathcal{N}_{\mu,\sigma}(k)$ <br> $\mu = I_K(f), \sigma = 0.5$ |
| Beschreibung: | 3.2.2; 4.3.3.2; 5.2.1; 5.3.4.2; 6.2.5 |
| Maßeinheit: | $\in$ [0;1] |
| Typ: | Z |
| Slot/Methode: | `cognitive-pauses` |
| Berechnung: | `duration, syn-structure` |
| Abbildung: | niedrig = 0 %, hoch = 100 % |

| VERWENDUNG VON NICHTAUFLÖSBARER ORTSDEIXIS ($\mathcal{O}$) | Symptom |
|---|---|
| Eltern: | ERFOLG BEI DER KONZEPTUALISIERUNG ($\mathcal{K}$), AUSMASS DES MONITORINGS ($\mathcal{M}$), |
| Zustände: | 2 (ja, nein) |
| CPT: | |
| Einfluß: | $I_O(k,m) = 0.90 - 0.05 * \frac{k+m}{2}$ |
| Verteilung: | $P(O = o \mid K = k, M = m) = \begin{cases} I_O(f,m) & : \quad o = 0 \\ 1 - I_O(f,m) & : \quad o = 1 \end{cases}$ |
| Beschreibung: | 3.6; 4.3.7; 6.2.7 |
| Maßeinheit: | $\in \{0,1\}$ |
| Typ: | B |
| Slot/Methode: | `improper-deixis` |
| Berechnung: | `– (wird gesetzt)` |
| Abbildung: | nein = 0, ja = 1 |

| MENGE ZUSÄTZLICHER INFORMATIONEN ($\mathcal{U}$) | Symptom |
|---|---|
| Eltern: | ZEITDRUCK ($\mathcal{Z}$), |
| Zustände: | 3 (wenig, mäßig viele, viele) |
| CPT: | |
| Einfluß: | $I_U(z) = 1 - z$ |
| Verteilung: | $P(U = u \mid Z = z) = \mathcal{N}_{\mu,\sigma}(z)$ |
| | $\mu = I_U(z),\ \sigma = 0.5$ |
| Beschreibung: | 3.5.3; 4.3.6.1 |
| Maßeinheit: | $\in$ [0;1] |
| Typ: | Z |
| Slot/Methode: | `superfluous-information` |
| Berechnung: | `– (wird gesetzt)` |
| Abbildung: | wenig = 0, viele = 1 |

# Anhang C

# Verwendete Tagsets

| POS- Tag | Beschreibung | Beispiel |
|---|---|---|
| ADJA ADJD | Adjektive | [das] große [Haus]; [er fährt] schnell |
| ADV | Adverb | schon,bald,doch |
| APPR APPRART APPO APZR | Prä-,Zirkum-,Postposition | in [der Stadt] |
| ART | Artikel | der, die, das |
| CARD | Kardinalzahl | zwei, 1994 |
| FM | Fremdsprachl. Material | A big fish |
| ITJ | Interjektion | (in READY nicht verwendet !) |
| KOUI KOUS KON KOKOM | Konjunktionen | anstatt, weil, daß |
| NN NE | Nomen | Tisch, Hans |
| PDS PDAT | Demonstrativpronomen | dieser; jener |
| PIS PIAT PIDAT | Indefinitpronomen | keiner; irgendein; beide |
| PPER | Personalpronomen | ich, du |
| PPOSS PPOSAT | Possessivpronomen | meins; mein |
| PRELS PRELAT | Relativpronomen | dessen, welcher |
| PRF | reflexives Personalpronomen | sich, einander |
| PWS PWAT PWAV | Interrogativ-, Relativpronomen | wer; welche; warum |
| PAV | Pronominaladverb (dafür, dabei) | dafür, dabei |
| PTKZU PTKNEG PTKVZ PTKANT PTKA | Partikeln | nicht, zu, ja, |
| TRUNC | Kompositions-Erstglied | An- [und Abreise] |
| VVFIN VVIMP VVINF VVIZU VVPP VAFIN VAIMP VAINF VAPP VMFIN VMINF VMPP | Verben | gehst; komm; ankommen; gewesen |
| XY | Nichtwort oder mit Sonderzeichen | D2XW3 |
| $. $, $( | Satzzeichen | . ? ! , ( ) - |

Tabelle C.1: Das Stuttgart- Tübingen- Tagset nach Schiller et al. (1995)

| Kategorie | Beschreibung | Beispiel |
|---|---|---|
| SB | Subjekt | Ich [finde den Hebel nicht.] |
| MO | Modifikator | Mit dem Schraubenzieher [mal festziehen.] |
| OA | Akkusativ-Objekt | [Ich finde] den Hebel [nicht.] |
| OC | Klausales Objekt | [Mal probieren,] ob es funktioniert. |
| OP | Präpositionales Objekt | [Suchen Sie] nach einem Stab. |

Tabelle C.2: Verwendete grammatische Kategorien

# Literatur


Anderson, J. R. (1985). *Cognitive psychology and its implications.* New York: W. H. Freeman.

Arnold, S. R. (1997). Wo das Perlhuhn trudelt. *Die ZEIT vom 11.7.1997*, S. 33-34.

Baber, C. & Mellor, B. (1996). The effects of workload on speaking: Implications for the design of speech recognition systems. *Contemporary Ergonomics*, 513-517.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*, 353-356.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D., Gathercole, S. & Constanza, P. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158-173.

Batliner, A., Burger, S. & Kießling, A. (1994). *Außergrammatische Phänomene in der Spontansprache: Gegenstandsbereich, Beschreibung, Merkmalinventar* (Verbmobil Report Nr. 57). München.

Batliner, A., Kießling, A., Burger, S. & Nöth, E. (1995). *Filled pauses in spontaneous speech* (Verbmobil Report Nr. 88). München.

Bear, J., Dowding, J. & Shriberg, E. (1992). Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proceedings of the Twenth-Ninth Annual Meeting of the Association for Computational Linguistics.* Newark, DE: 56-63.

Beattie, G. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation.* Milton Keynes: Open University Press.

Beattie, G. (1984). Are there cognitive rhythms in speech? A reply to Power (1983). *Language and Speech*, *27*, 193-195.

Berthold, A. (1997). „Ich – ähm – denke, daß ...": Zu sprachlichen Indikatoren für Ressourcenbeschränkungen und ihrer Verarbeitung in natürlichsprachlichen Dialogsystemen. In R. Schäfer & M. Bauer (Hrsg.), *ABIS-97, Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen* (S. 139-147). Saarbrücken: SFB 378, Universität des Saarlandes.

Berthold, A. (1998). *Transliteration der Dialoge aus der* READY-*Voruntersuchung.* http://www.coli.uni-sb.de/~berthold/ready/Dialoge.

Berthold, A. & Jameson, A. (1999). Interpreting symptoms of cognitive load in speech input. *Zur Veröffentlichung eingereicht.*

Beun, R. J. (1985). The function of repetitions in information dialogues. *IPO Annual Progress Report*, *20*, 91-98.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148-158.

Bourdin, B. & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, *29*, 591-620.

Breznitz, Z. (1989). Vocalization and pauses in fast-paced reading. *Journal of General Psychology*, *117*, 153-159.

Bromme, R. & Wehner, T. (1987). Zum Zusammenhang von Sprechgeschwindigkeit und Sprechfehlern mit der Aufgabenschwierigkeit beim lauten Denken. *Zeitschrift für experimentelle und angewandte Psychologie*, *34*, 1-16.

Bußmann, H. (1990). *Lexikon der Sprachwissenschaft.* Stuttgart: Kröner.

Butcher, A. (1980). Pause and syntactic structure. In H. W. Dechert & M. Raupach (Hrsg.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (S. 85-90). The Hague: Mouton.

Butterworth, B. (1980a). Evidence from pauses in speech. In B. Butterworth (Hrsg.), *Language production* (Bd. 1, S. 155-176). New York, London: Academic Press.

Butterworth, B. (1980b). *Language production* (Bd. 1). New York, London: Academic Press.

Carletta, J., Caley, R. & Isard, S. (1995). Simulating time-constrained language production. *Language and Cognitive Processes*, *10*, 357-361.

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine*, *12*(4), 50-63.

Clark, H. H. & Clark, E. V. (1977). *Psychology and language.* New York: Harcourt Brace Jovanovich.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39.

Deese, J. (1980). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Hrsg.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (S. 69-84). The Hague: Mouton.

Dell, G. S. & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the 'model of the listener'. In D. Napoli & J. Kegel (Hrsg.), *Bridges between psychology and linguistics: A swarthmore festschrift for Lila Gleitman.* Hove, UK: Erlbaum.

Duez, D. (1982). Silent and non-silent pauses in three different speech styles. *Language and Speech*, *25*, 11-28.

Finkler, W. (1997). *Automatische Selbstkorrektur bei der inkrementellen Generierung gesprochener Sprache unter Realzeitbedingungen.* Dissertation, Universität des Saarlandes, Saarbrücken.

Gamut, L. (1991). *Logic, language and meaning* (Bd. 2: Intensional Logic and Logical Grammar). Chicago: Chicago University Press.

Gardent, C. (1997). Parallelism, HOU and deaccenting. *In Vorbereitung*.

Gardent, C. & Kohlhase, M. (1996). Higher order coloured unification and natural language semantics. In *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics.* Santa Cruz, USA.

Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language.* Hove, UK: Erlbaum.

Gee, J. P. & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411-458.

Gerrig, R. J. & Littman, M. L. (1990). Disambiguation by community membership. *Memory and Cognition*, *18*, 331-338.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown.* Dissertation, Carnegie Mellon University.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press.

Grabowski, J. (1991). *Der propositionale Ansatz der Textverständlichkeit: Kohärenz, Interessantheit und Behalten.* Münster: Aschendorff.

Greene, J. O. (1984). Speech preparation processes and verbal fluency. *Human Communication Research*, *11*, 61-84.

Greene, J. O., Lindsey, A. E. & Hawn, J. J. (1990). Social goals and speech production: Effects of multiple goals on pausal phenomena. *Journal of Language and Social Psychology*, *9*, 119-134.

Grice, H. P. (1975). Logic and conversation. In P. Coles & J. L. Morgan (Hrsg.), *Syntax and semantics* (Bd. 3, S. 41-58). New York: Academic Press.

Grosjean, F. & Deschamps, A. (1973). Analyse des variables temporelles du français spontané. *Phonetica*, *28*, 191-226.

Guggisberg, R., Laederach, K. & Adler, R. (1981). Formal speech stylistics and type A behaviour in 38 subjects during non-stress interviews. *Psychotherapy and Psychosomatics*, *36*, 86-91.

Heckerman, D. & Horvitz, E. (1998). Inferring informational goals from free-text queries: A Bayesian approach. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (S. 230-237). Madison, WI: Morgan Kaufman.

Heeman, P. & Allen, J. (1994). Detecting and correcting speech repairs. In *Proceedings of the Thirty-Second Annual Meeting of the Association for Computational Linguistics* (S. 295-302). Las Cruces, NM.

Heeman, P. A. & Allen, J. F. (1997). Intonational boundaries, speech repairs and discourse markers: Modeling spoken dialog. In *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics.* Madrid, Spanien.

Henderson, A., Goldman-Eisler, F. & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, *9*, 207-216.

Henrion, M., Breese, J. S. & Horvitz, E. J. (1991). Decision analysis and expert systems. *AI Magazine*, *12*(4), 64-91.

Hermann, T. & Grabowski, J. (1994). *Sprechen: Psychologie der Sprachpoduktion.* Heidelberg: Spektrum.

Hieke, A. E., Kowal, S. & O'Connell, D. (1983). The trouble with articulatory pauses. *Language and Speech*, *26*, 203-214.

Hindle, D. (1983). Deterministic parsing of syntactic non-fluencies. In *Proceedings of the Twenty-First Annual Meeting of the Association for Computational Linguistics* (S. 123-128). Cambridge, MA.

Horton, W. S. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91-117.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D. & Rommelse, K. (1998). The Lumière project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (S. 256-265). Madison, WI.

Jameson, A. (1996). Numerical uncertainty management in user and student modeling: An overview of systems and issues. *User Modeling and User-Adapted Interaction*, *5*, 193-251. (Verfügbar über http://www.cs.uni-sb.de/users/jameson/)

Jameson, A. (1997). Wie gehen wir mit dem Arbeitsgedächtnis unserer Dialogpartner um? Eine Integration von Ergebnissen aus vier Forschungsrichtungen. In H. Mandl (Hrsg.), *Bericht über den 40. Kongreß der Deutschen Gesellschaft für Psychologie in München 1996* (S. 258-263). Göttingen: Hogrefe.

Jameson, A. (1998). Adapting to the user's time and working memory limitations: New directions of research. In U. J. Timm & M. Rössel (Hrsg.), *ABIS-98, Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen.* Erlangen: FORWISS.

Jameson, A. & Buchholz, K. (1998). Einleitung zum Themenheft „Ressourcenadaptive kognitive Prozesse" . *Kognitionswissenschaft*, *7*(3).

Jameson, A., Paris, C. & Tasso, C. (Hrsg.). (1997). *User modeling: Proceedings of the Sixth International Conference, UM97.* Wien, New York: Springer Wien New York. (Elektronisch verfügbar über http://www.cs.uni-sb.de/UM97/ and http://um.org/um_97/proceedings-overview.html)

Jameson, A., Schäfer, R., Weis, T., Berthold, A. & Weyrath, T. (in Druck). Making systems sensitive to the user's time and working memory constraints. In M. T. Maybury (Hrsg.), *IUI '99: International Conference on Intelligent User Interfaces.* New York: ACM.

Jensen, F. V. (1996). *An introduction to Bayesian networks.* New York: Springer.

Jou, J. & Harris, R. J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society*, *30*, 301-304.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kamp, H. & Reyle, U. (1993). *From discourse to logic.* Dordrecht: Kluwer.

Kelly, K. R. & Stone, G. L. (1982). Effects of time limits on the interview behaviour of type A and B persons within a brief counseling interview. *Journal of Counseling Psychology*, *29*, 454-459.

Klesen, M. (1997). *Statistische Klassifikation von Dialogakten.* Diplomarbeit, Fachbereich Informatik, Universität des Saarlandes, Saarbrücken.

Klos, T. & Ellgring, H. (1987). Manuelle vs. elektronische Analyse von Sprechpausen. *Zeitschrift für experimentelle und angewandte Psychologie*, *34*, 64-71.

Kobsa, A. & Wahlster, W. (Hrsg.). (1989). *User models in dialog systems.* Berlin: Springer.

Kohler, K., Lex, G., Pätzold, M., Scheffers, M., Simpson, A. & Thon, W. (1994). *Handbuch zur Datenaufnahme und Transliteration in TP 14 von Verbmobil 3.0* (Verbmobil Technisches Dokument Nr. 11). Kiel.

Kowal, S. (1991). *Über die zeitliche Organisation des Sprechens in der Öffentlichkeit.* Bern: Huber.

Kowal, S. & O'Connell, D. (1987). Some temporal aspects of stories told while or after watching a film. *Bulletin of the Psychonomic Society*, *25*, 364-366.

Kowal, S., Wiese, R. & O'Connell, D. (1983). The use of time in storytelling. *Language and Speech*, *26*, 377-392.

Lazarus-Mainka, G. & Arnold, M. (1987). Implizite Strategien bei Doppeltätigkeit: Sprechen – Zuhören und Sortieren. *Zeitschrift für experimentelle und angewandte Psychologie*, *34*, 286-300.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge: ACL/MIT.

Levelt, W. J. M. & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*, 78-106.

Levinson, S. C. (1990). *Pragmatik.* Tübingen: Niemeyer.

Lewis, R. (1993). Specifying architectures for language processing: process, control, and memory in parsing and interpretation. In M. Crocker, M.Pickering & C. Clifton (Hrsg.), *Architectures and mechanisms for language processing.* Cambridge: Cambridge University Press.

Malloch, M., Shapiro, J., Hitch, G., Culpin, V. & Towse, J. (1995). Temporal effects in immediate verbal memory: A combined experimental/modelling approach. *Language and Cognitive Processes*, *10*, 401-405.

Marx, E. (1984). *Über die Wirkung von Zeitdruck auf Sprachproduktionsprozesse.* Dissertation, Universität Münster.

McLeod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *Quarterly Journal of Experimental Psychology*, *29*, 651-667.

Miller. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Mitchell, H. L., Hoit, J. D. & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. *Journal of Speech and Hearing Research*, *39*, 93-104.

Montague, R. (1974). The proper treatment of quantification in ordinary English. In R. H. Thomason (Hrsg.), *Formal philosophy: Selected papers of Richard Montague.* New Haven: Yale University Press.

Nakatani, C. & Hirschberg, J. (1993). A speech-first model for repair detection and correction. In *Proceedings of the Thirty-First Annual Meeting of the Association for Computational Linguistics.* Columbus, OH.

Naumann, S. & Langer, H. (1994). *Parsing. Eine Einführung in die maschinelle Analyse natürlicher Sprache.* Stuttgart: Teubner.

Neapolitan, R. E. (1990). *Probabilistic reasoning in expert systems: Theory and algorithms.* New York: Wiley.

Nederhof, M.-J., Bouma, G., Koeling, R. & van Noord, G. (1997). Grammatical analysis in the OVIS spoken-dialogue system. In *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics.* Madrid, Spanien.

*Netica: Application for belief networks and influence diagrams. User's guide. 1.03.* (1996). Vancouver, Canada. (Veröffentlicht von Norsys Software Corp., Vancouver, BC, Canada, http://www.norsys.com/)

Neumann, O. (1992). Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen. *Psychologische Rundschau*, *43*, 83-101.

Niederberger, U., Engemann, A. & Radtke, M. (1987). Umfang der Informationsverarbeitung bei Entscheidungen: Der Einfluß von Gedächtnisbelastung und Handlungsorientierung. *Zeitschrift für experimentelle und angewandte Psychologie*, *34*, 80-100.

Niehren, J., Pinkal, M. & Ruhrberg, P. (1997a). On equality up-to constraints over finite trees ,context unification and one-step rewriting. In *Proceedings of the International Conference on Automated Deduction* (Bd. 1249, S. 34-48). Townsville, Australia: Springer-Verlag. (Erhältlich unter http://www.ps.uni-sb.de/Papers/abstracts/fullContext.html)

Niehren, J., Pinkal, M. & Ruhrberg, P. (1997b). A uniform approach to underspecification and parallelism. In *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics.* Madrid, Spanien.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*, 19-35.

Oviatt, S., Levov, G.-A., MacEachern, M. & Kuhn, K. (1996). Modeling hyperarticulate speech during human-computer error resolution. In *Proceedings of the International Conference on Spoken Language Processing* (Bd. 2, S. 801-804). Philadelphia, DE: A. I. duPont and University of Delaware.

Oviatt, S. & VanGent, R. (1996). Error resolution during multimodal human-computer-interaction. In *Proceedings of the International Conference on Spoken Language Processing* (Bd. 1, S. 204-207). Philadelphia, DE: A. I. duPont and University of Delaware.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Pechmann, T. (1991). Besprechung: Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge, Mass.: MIT Press. *Sprache und Kognition*, *10*, 171-177.

Pinkal, M. (1985). Kontextabhängigkeit, Vagheit, Mehrdeutigkeit. In C. Schwarze & D. Wunderlich (Hrsg.), *Handbuch der Lexikologie* (S. 27-63). Königstein/Taunus: Athenäum.

Pinkal, M. (1993). *Semantikformalismen für die sprachverarbeitung* (CLAUS Report Nr. 26). Universität des Saarlandes.

Pinkal, M. (1996). *Vagueness, ambiguity, and underspecification* (CLAUS Report Nr. 82). Universität des Saarlandes.

Pinkal, M. & Siekmann, J. (1998). LISA. Ressourcenadaptive linguistische Inferenz. In W. H. Tack (Hrsg.), *Ressourcenadaptive kognitive Prozesse: Arbeits- und Ergebnisbericht Januar 1996 – Januar 1998.* Saarbrücken: SFB 378, Universität des Saarlandes.

Postma, A., Kolk, H. & Povel, D.-J. (1990). On the relation among speech errors, disfluencies and self-repairs. *Language and Speech*, *33*, 19-29.

Power, M. J. (1983). Are there cognitive rhythms in speech? *Language and Speech*, *26*, 253-261.

Power, M. J. (1984). Are there cognitive rhythms in speech? A reply to Beattie (1984). *Language and Speech*, *27*, 197-201.

Power, M. J. (1985). Sentence production and working memory. *The Quarterly Journal of Experimental Psychology*, *37A*, 367-385.

Power, M. J. (1986). A technique for measuring processing load during speech production. *Journal of Psycholinguistic Research*, *15*, 371-382.

Robson, C. (1973). *Experiment, design and statistics in psychology.* Harmondsworth, UK: Penguin.

Rosé, C. P. & Waibel, A. H. (1996). Recovering from parser failures: A hybrid statistical and symbolic approach. In J. H. Klavans & P. Resnik (Hrsg.), *The balancing act: Combining symbolic and statistical approaches to language* (S. 157-179). Cambridge, MA: MIT Press.

Roßnagel, C. (1995a). Übung und Hörerorientierung beim monologischen Instruieren: Zur Differenzierung einer Grundannahme. *Sprache und Kognition*, *14*, 16-26.

Roßnagel, C. (1995b). Kognitive Belastung und Hörerorientierung beim monologischen Instruieren. *Zeitschrift für experimentelle und angewandte Psychologie*, *42*, 94-110.

Ruland, T., Rupp, C. J., Spilker, J., Weber, H. & Worm, K. L. (1998). *Making the most of multiplicity: A multi-parser multi-strategy architecture for the robust processing of spoken language* (Verbmobil Report Nr. 230). Erlangen.

Rummer, R. (1996a). *Kognitive Beanspruchung beim Sprechen.* Weinheim: Beltz.

Rummer, R. (1996b). Sprechplanung und Arbeitsgedächtnis. In R. H. Kluwe & M. May (Hrsg.), *Proceedings der 2. Fachtagung der Gesellschaft für Kognitionswissenschaft* (S. 145-146). Hamburg: Universität der Bundeswehr.

Rummer, R. (1997). Verarbeitungsbeschränkungen bei der Sprachproduktion: Ein experimenteller Ansatz zur Erforschung sprachlicher Makroplanungsprozesse. In R. H. Kluwe (Hrsg.), *Strukturen und Prozesse intelligenter Systeme* (S. 41-63). Wiesbaden: Deutscher Universitäts-Verlag.

Rummer, R. & Grabowski, J. (1996). Control processes in language production: Theoretical, methodological and empirical aspects. *Unveröffentlichtes Manuskript*.

Rummer, R., Grabowski, J. & Vorwerg, C. (1995). Kontrollprozesse beim Sprechen: Flexibilität und Determination der ereignisbezogenen Äußerungsplanung. *Zeitschrift für Psychologie*, *203*, 25-51.

Russel, S. & Norvig, P. (1995). *Artificial intelligence: A modern approach.* Englewood Cliffs, NJ: Prentice Hall.

Schäfer, R. (1998). *Benutzermodellierung mit dynamischen Bayes'schen Netzen als Grundlage adaptiver Dialogsysteme.* Dissertation, Universität des Saarlandes, Saarbrücken.

Schäfer, R., Weis, T., Weyrath, T. & Jameson, A. (1997). Wie können Ressourcenbeschränkungen eines Dialogpartners erkannt und berücksichtigt werden?. *Kognitionswissenschaft*, *6*, 151-164.

Schäfer, R. & Weyrath, T. (1996). Einschätzung von verfügbarer Arbeitsgedächtniskapazität mit temporalen Bayesschen Netzen. In H.-G. Lindner (Hrsg.), *ABIS-96, Adaptivität und Benutzermodellierung in interaktiven Systemen: Workshop-Beiträge* (S. 5.1-5.9). Dortmund: VEW AG.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society London, Series B*, *298*, 199-209.

Shiffrin, R. M. & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, *101*, 357-361.

Skut, W. & Brants, T. (1998a). Chunk tagger: Statistical recognition of noun phrases. In *Proceedings of the ESSLLI. Workshop on automated acquisition of syntax and parsing.* Saarbrücken.

Skut, W. & Brants, T. (1998b). A maximum-entropy partial parser for unrestricted text. In *Proceedings of the sixth workshop on very large corpora.* Montreal, CA.

Smolka, G. (1995). The OZ programming model. In J. van Leeuwen (Hrsg.), *Computer science today* (Bd. 1000, S. 324-343). New York: Springer.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L. & Cowell, R. G. (1993). Bayesian analysis in expert systems. *Statistical Science*, *8*, 219-283.

Stemberger, J. P. (1982). Syntactic errors in speech. *Journal of Psycholinguistic Research*, *11*, 313-345.

Strube, G. & Pentenrieder, I. (1987). Induzierte Verarbeitungsstrategien bei Mehrfachtätigkeit. *Zeitschrift für experimentelle und angewandte Psychologie*, *34*, 331-339.

Traum, D. R. & Allen, J. F. (1994). Discourse obligations in dialogue processing. In *Proceedings of the Thirty-Second Annual Meeting of the Association for Computational Linguistics* (S. 1-8). Las Cruces, NM.

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S. & Skut, W. (1998). Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft*, *7*(3), 129-133.

Wahlster, W. & Jameson, A. (1998a). READY: Benutzerorientierte Ressourcenadaptation. In W. H. Tack (Hrsg.), *Ressourcenadaptive kognitive Prozesse: Fortsetzungsantrag, Januar 1999 – Dezember 2001.* Saarbrücken: SFB 378, Universität des Saarlandes.

Wahlster, W. & Jameson, A. (1998b). READY: Ressourcenadaptive Dialogführung. In W. H. Tack (Hrsg.), *Ressourcenadaptive kognitive Prozesse: Arbeits- und Ergebnisbericht, Januar 1996 – Januar 1998.* Saarbrücken: SFB 378, Universität des Saarlandes.

Wahlster, W., Jameson, A., Ndiaye, A., Schäfer, R. & Weis, T. (1995a). READY: Ressourcenadaptive Dialogführung. In W. H. Tack (Hrsg.), *Ressourcenadaptive kognitive Prozesse: Finanzierungsantrag 1996 – 1997 – 1998.* Saarbrücken: Universität des Saarlandes.

Wahlster, W., Jameson, A., Ndiaye, A., Schäfer, R. & Weis, T. (1995b). Ressourcenadaptive Dialogführung: Ein interdisziplinärer Forschungsansatz. *Künstliche Intelligenz*, *9*(6), 17-21.

Wegner, D. M. & Erber, R. (1992). The hyperaccessibility of suppressed thoughts. *Journal of Personality and Social Psychology*, *63*, 903-912.

Weis, T. (1999). *Entscheidungstheoretisches ressourcenadaptives Dialogmanagement* . Dissertation, Universität des Saarlandes. (In Vorbereitung)

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, *9*, 36-45.

Weyrath, T. (1998). *Erkennung von Arbeitsgedächtnisbelastung und Zeitdruck in Dialogen – Empirie und Modellierung mit Bayesschen Netzen.* Diplomarbeit, Fachbereich Informatik, Universität des Saarlandes.

Wickens, C. D. (1992). *Engineering psychology and human performance.* New York: HarperCollins.

Wiese, R. (1983). *Psycholinguistische Aspekte der Sprachproduktion.* Hamburg: Buske.

Worm, K. L. (1998). *A model for robust processing of spontaneous speech by integrating viable fragments* (Verbmobil Report Nr. 223). Saarbrücken.

Worm, K. L., Egg, M. & Winter, V. (1997). *Robustheit in CHORUS.* (Interner Bericht, Sonderforschungsbereich 378, Computerlinguistik, Universität des Saarlandes)

Ziesche, S. & Klein, A. (1997). Resource-constrained generation with the principle of adequacy drop. In *Proceedings of the Sixth European Workshop on Natural Language Generation.* Duisburg.

Zweig, G. & Russell, S. (1998). Speech recognition with dynamic bayesian networks. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence.* Madison, WI: AAAI Press.

# Verwendete Abkürzungen

In Tabellen und Grafiken wird aus Platzgründen für Arbeitsgedächtnisbelastung bzw. kognitive Belastung die Kurzform AGB und für Zeitdruck die Kurzform ZD verwendet.

$S$ und $P$ stehen als Kürzel für zwei Dialogteilnehmer, wobei in speziellen Kontexten $S$ für 'System' und $P$ für 'Partner' (im Kontext von READY der Anrufer) verwendet werden.

GPD steht für *Gesamtpausendauer*, GSD für *Gesamtsprechdauer*.