

Symptome von Zeitdruck und kognitiver Belastung in gesprochener Sprache: eine experimentelle Untersuchung

Diplomarbeit im Fachbereich Computerlinguistik

von
Christian Alfons Müller

angefertigt im Rahmen des
Sonderforschungsbereichs 378 „Ressourcenadaptive kognitive Prozesse“ (Projekt
READY)
an der Universität des Saarlandes

Saarbrücken 15. März 2001

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen verwendet habe.

Christian Müller, Saarbrücken, den 15. März 2001

Danksagung

Ich möchte mich bei allen Mitarbeitern des Projektes READY des SFB 378 für ihre Unterstützung bedanken. Bei Sylvia Bach für ihren nervenstarken Einsatz bei der Erhebung der Daten; bei Barbara Großmann-Hutter, die für die Entwicklung und Anpassung des Experimentalsystems bisweilen in die klassische Rolle des Sisyphus geschlüpft ist; bei Tore Knabe und Frank Wittig für ihre Arbeit bei der Analyse der Daten. Insbesondere bedanke ich mich bei Anthony Jameson, der bei der Betreuung der Arbeit seine Zeit für unzählige individuelle Besprechungen geopfert hat. Dank gilt auch denjenigen, die sich im Vorfeld des Experiments als Probe-Versuchsperson zur Verfügung gestellt haben. Prof. Manfred Pinkal danke ich dafür, dass er mir die Freiheit eingeräumt hat, für meine Diplomarbeit auf diesem interdisziplinären Gebiet zu forschen. Unter anderem für Ideen und Anregungen bedanke ich mich bei Simone Könen.

Inhaltsverzeichnis

1	Einleitung	11
1.1	Zugrundeliegende Motivation	11
1.2	Fragestellungen dieser Arbeit	11
1.3	Überblick über die Arbeit	12
2	Theoretische Grundlagen	14
2.1	Bisherige Studien	14
2.1.1	Eine vergleichende Literaturstudie von Berthold (1998)	14
2.1.2	Vergleich der Übersichtsstudien	22
2.2	Der Einfluß der Äußerungskomplexität auf die Symptome	23
3	Eigene empirische Studien	24
3.1	Einleitung	24
3.1.1	Motivation	24
3.1.2	Zusammenfassung der Fragestellungen	27
3.2	Eine Feldstudie am Frankfurter Flughafen	28
3.3	Ein Experiment zur Untersuchung von Sprache unter Belastung	29
3.3.1	Vorüberlegungen	29
3.3.2	Methode	37
3.3.3	Einteilung und Definition der Variablen	47
3.3.4	Spezifische Vorhersagen	52
3.4	Ergebnisse	54
3.4.1	Auswirkungen der Äußerungskomplexität	54
3.4.2	Analyseverfahren	54
3.4.3	Zeitsymptome	55
3.4.4	Weitere Analysen der einzelnen Faktoren	57
3.4.5	Qualitätssymptome	61
3.4.6	Längensymptome	62
3.4.7	Die Artikulationsgeschwindigkeit	65
3.4.8	Inhaltliche Qualität	65
3.4.9	Belastung durch die Aufgabenkomplexität	65
3.4.10	Zusammenfassung der Ergebnisse	65
3.5	Belastung durch die Nebenaufgabe	68
3.6	Bewertung der Ergebnisse	69

4	Modellierung	71
4.1	Methodische Analysen	71
4.2	Erkennung von Ressourcenbeschränkungen	72
4.2.1	Bayssche Netze	72
4.2.2	Die Struktur des Baysschen Netzes	73
4.2.3	Lernen eines Bayssches Netzes	73
4.2.4	Evaluation	75
5	Bewertung und Ausblick	79
5.1	Bewertung	79
5.2	Ausblick	80
	Literaturverzeichnis	81

Abbildungsverzeichnis

3.1	Bayssches Netz zur Interpretation von Symptomen	26
3.2	Inhalte der am Frankfurter Flughafen aufgezeichneten Anfragen	28
3.3	Experimentelle Stimuli bei Eklund (1999)	34
3.4	Stimuli bei (Eklund, 1999)	35
3.5	Experimentelle Stimuli	38
3.6	Organisation des Experimentes	39
3.7	Experimentalsystem ohne Navigationsaufgabe	40
3.8	Oberfläche des Experimentalsystems mit Navigationsaufgabe	41
3.9	Abfrage der Zielnummer	41
3.10	Aufzeichnung der Sprache	50
3.11	Verarbeitung der digitalisierten Äußerungen	51
3.12	Durchschnittliche Werte für gefüllte Pausen	57
3.13	Durchschnittliche Werte für stille Pausen	58
3.14	Durchschnittliche Werte für die Einsatzlatenz	59
3.15	Durchschnittliche Werte für die Wiederholungen	60
3.16	Durchschnittliche Werte für Begrüßungen und Anreden	60
3.17	Durchschnittliche Werte für Silbendehnungen und Zögerungen	61
3.18	Durchschnittliche Werte für die Fehlansätze	63
3.19	Durchschnittliche Werte für die Satzfragmente	63
3.20	Durchschnittliche Werte für inh. Selbstkorrekturen	64
3.21	Durchschnittliche Werte für die Längensymptome	66
3.22	Durchschnittliche Werte für die Artikulationsgeschwindigkeit	67
3.23	Bewertung der inhaltlichen Qualität	67
4.1	Struktur des Baysschen Netzes	74
4.2	Genauigkeit des Baysschen Netzes für Zeitdruck	76
4.3	Genauigkeit des Baysschen Netzes für die Nebenaufgabe	77

Tabellenverzeichnis

2.1	Symptome kognitiver Belastung nach	15
2.2	Ressourcenabhängigkeit verschiedener Teilprozesse der Sprachproduktion	22
3.1	Einteilung von Nebenaufgaben nach Rummer (1996)	31
3.2	Design des Flughafenexperiments	37
3.3	Reihenfolgevarianten nach dem Rotationsprinzip	39
3.4	Bedeutung der Variablen und Aufteilung in die drei Hauptkategorien . . .	49
3.5	Faktoranalyse aller Zeitsymptome	56
3.6	Faktoranalyse aller Qualitätssymptome	61
3.7	Faktoranalyse der Längensymptome	64
3.8	Korrelation der Aufgabenkomplexität mit den einzelnen Symptomen . . .	68

Kapitel 1

Einleitung

1.1 Zugrundeliegende Motivation

Großflughäfen wie der Frankfurter Flughafen sind hochkomplexe künstliche Räume, die den Menschen, der sich darin zurecht finden soll, nicht selten überfordern. Mit Expansion des internationalen Flugverkehrs, ist es zu erwarten, dass dieses Problem in der Zukunft noch gravierender werden wird. Um den Reisenden auf seinem Weg durch einen solchen Großflughafen zu unterstützen, wäre ein elektronisches Assistenzsystem wünschenswert. Ein solches tragbares System könnte mithilfe einer Wissensbasis über Strukturen und Abläufe am Großflughafen, dem Benutzer relevante Informationen wie Wegbeschreibungen oder die Abflugzeit seines Fluges zur Verfügung stellen. Es könnte ihm auch Möglichkeiten aufzeigen, wie er die Wartezeit bis zum Abflug auf angenehme Art überbrücken kann. Eine besondere Eigenschaft des Assistenten sollte jedoch sein, dass er den mentalen Zustand berücksichtigt, in dem sich der Benutzer gerade befindet. Falls es sich nämlich bei dem Reisenden nicht gerade um einen Vielflieger handelt, wird diese Person evtl. etwas Aufregung verspüren, vielleicht sogar „Reisefieber“. Sie wird vielleicht bereits eine lange Anfahrt hinter sich haben. Am Flughafen wird sie mit Informationen überschüttet: Hinweistafeln, Lautsprecherdurchsagen, u.Ä. gilt es zu beachten. Eventuell befindet sich die Person unter starkem Zeitdruck, weil bis zum Abflug der Maschine nicht mehr viel Zeit bleibt. Aber auch wenn der Abflug nicht unmittelbar bevorsteht, wird sie einen gewissen Zeitdruck empfinden, denn es gilt noch einige Dinge zu erledigen, bevor sie entspannt in der Wartehalle Platz nehmen, oder das Unterhaltungsangebot wahrnehmen kann.

1.2 Fragestellungen dieser Arbeit

Die Entwicklung eines solchen Assistenzsystems stellt eines von zwei Forschungsszenarien in der aktuellen Förderungsperiode des Sonderforschungsbereichs 378 „Ressourcenadaptive Kognitive Prozesse“ dar, in dessen Rahmen Projekte aus unterschiedlichen Disziplinen forschen. Diese Disziplinen sind die Computerlinguistik, die Informatik, die Psychologie, und die Philosophie. Die interdisziplinäre Zusammenarbeit ist eine wichtige Eigenschaft des Sonderforschungsbereichs, wozu die vorliegende Arbeit einen Beitrag zu leisten versucht.

Es soll untersucht werden, inwieweit die Eigenschaften der Sprache des Reisenden Rückschlüsse über seinen mentalen Zustand zulassen, insbesondere ob er unter Zeitdruck steht, oder einer erhöhten kognitiven Belastung unterliegt. Geht man nämlich von der Annahme aus, dass die gesprochene Sprache zumindest eine von mehreren Eingabemodalitäten des Assistenten sein wird, ist es sinnvoll, diese als Informationsquelle zur Modellierung des Benutzers zugrunde zu legen.

Was diesen Aspekt angeht, nämlich die Untersuchung von Symptomen kognitiver Belastung und Zeitdruck in der Eingabesprache des Benutzers, baut diese Arbeit zu einem Großteil auf Ergebnissen von Berthold (1998) auf. Dort wurde eine umfangreiche Literaturstudie durchgeführt und Ergebnisse aus einer Vielzahl von Experimenten zusammengetragen. Obwohl diese Ergebnisse nicht immer eindeutig waren, konnte so eine Menge sprachlicher Phänomene zusammengestellt werden, die in unserem Sinne als Symptome von Zeitdruck und Belastung betrachtet werden können. In Abschnitt 2.1 werden diese Phänomene im Einzelnen diskutiert. Die bei Berthold (1998) aufgeführten Studien beschäftigten sich jedoch vor allem mit der Sprache im Zusammenhang mit kognitiver Belastung. Die Auswirkungen von Zeitdruck sind bis dato dagegen weniger gut erforscht. Die bisherigen Studien untersuchten außerdem jeweils nur einzelne der relevanten Phänomene und wurden im Hinblick auf unterschiedliche Fragestellungen durchgeführt. In der vorliegenden Arbeit wird ein Experiment beschrieben, bei dem alle von Berthold (1998) herausgestellten Symptome zusammen mit direktem Bezug auf das Flughafenzenario untersucht wurden. Das Experiment wird im folgenden als *Flughafenexperiment* bezeichnet. Die Ergebnisse wurden innerhalb des READY-Projektes als Rohdaten für den Aufbau von Modellen verwendet, anhand derer der Zeitdruck und die kognitive Belastung des Benutzers eingeschätzt werden soll.

1.3 Überblick über die Arbeit

In Kapitel 2 werden die theoretischen Grundlagen dieser Arbeit erläutert und ein Überblick über den Stand der Forschung gegeben. In Abschnitt 2.1 werden die Ergebnisse zweier Literaturstudien gegenübergestellt, die sich mit den Auswirkungen von kognitiver Belastung und Zeitdruck auf die Sprache beschäftigen. In Abschnitt 2.2 werden die Ergebnisse einer Studie von Oviatt (1997) berichtet, die den Einfluß der Äußerungskomplexität auf das Vorkommen von sprachlichen Oberflächenphänomenen untersuchte Oviatt (1997).

In Kapitel 3 werden eigene empirische Untersuchungen beschrieben. Dazu werden in Abschnitt 3.1 zunächst einige Vorüberlegungen angestellt, z.B. was die Motivation angeht, solche Untersuchungen durchzuführen. In Abschnitt 3.2 wird dann eine Feldstudie am Frankfurter Flughafen vorgestellt, die vom Verfasser zur Vorbereitung eines Doppelaufgabenexperimentes durchgeführt wurde. Das Flughafenexperiment an sich wird in Abschnitt 3.3 beschrieben. Dabei werden zunächst in Abschnitt 3.3.1 wichtige Designentscheidungen diskutiert. Die angewandte Methode wird dann in Abschnitt 3.3.2 beschrieben. In Abschnitt 3.4 werden schließlich die Ergebnisse vorgestellt und in Abschnitt 3.6 bewertet.

Kapitel 4 beschäftigt sich mit der Verwendung der gewonnenen Daten zum Zweck der Modellierung von kognitiver Belastung und Zeitdruck eines Benutzers. In Abschnitt 4.1 werden verschiedene methodische Analysen vorgestellt, die im Rahmen des Projektes

READY auf Basis der gewonnenen Daten durchgeführt wurden. In Abschnitt 4.2 wird ein Modell zur Erkennung von Belastung und Zeitdruck eines Benutzers vorgestellt und evaluiert.

Kapitel 5 stellt eine Bewertung der hier beschriebenen Arbeit dar und gibt einen Ausblick darauf, wie die Ergebnisse in zukünftigen Untersuchungen ausgenutzt werden können.

Kapitel 2

Theoretische Grundlagen

Um einen Katalog von Symptomen von Zeitdruck und kognitiver Belastung zu Erstellen, die im Rahmen dieser Arbeit untersucht werden sollen, werden in diesem Kapitel im Abschnitt 2.1 zunächst die Ergebnisse der umfangreichen Literaturstudie von Berthold (1998) zusammengefasst und anschließend mit den Ergebnissen einer weiteren Überblickstudie von Oberauer und Hockl (2000) verglichen. Letztere strukturieren ihre Untersuchung nach Phasen der Sprachproduktion, was eine interessante Betrachtungsweise auf die relevanten sprachlichen Phänomene darstellt. In Abschnitt 2.2 wird der Einfluß der Äußerungskomplexität auf das Vorkommen von Belastungssymptomen untersucht. Dabei wird auf die Ergebnisse einer Untersuchung von Oviatt (1997) eingegangen, die festgestellt hat, dass die Häufigkeit von *Disfluencies*, wie die Symptome von Zeitdruck und kognitiver Belastung bei ihr genannt werden, mit zunehmender Äußerungskomplexität nicht nur linear, sondern exponentiell ansteigt. Diese Feststellung hat einen unmittelbaren Einfluß auf die Interpretation der Ergebnisse des Flughafenexperiments.

2.1 Bisherige Studien

2.1.1 Eine vergleichende Literaturstudie von Berthold (1998)

Berthold (1998) führte eine umfangreiche Literaturstudie durch, bei der Ergebnisse zusammengetragen wurden, die Symptome von kognitiver Belastung in gesprochener Sprache betreffen (s. auch Berthold & Jameson, 1999). In Tabelle 2.1 werden die Ergebnisse zusammenfassend dargestellt. Berthold und Jameson (1999) führen eine Klassifikation ein, bei der die betrachteten Phänomene grob unterteilt werden in Symptome, die die Sprachqualität betreffen, und Symptome, die sich auf zeitliche Faktoren der Äußerung auswirken. Diese Klassifikation spiegelt sich auch in Tabelle 2.1 wider. Auf der linken Seite werden die *Qualitätssymptome* aufgeführt, auf der rechten Seite die *Geschwindigkeitssymptome*. Die Klasse der Qualitätssymptome umfasst Phänomene, die die formale Qualität der Äußerung betreffen. Dabei wird davon ausgegangen, dass der Sprecher die Äußerung mit hoher formaler Qualität geplant hatte, und das Vorkommen dieser Symptome auf eine Ressourcenbeschränkung schließen lässt. Als Geschwindigkeitssymptome werden diejenigen Phänomene bezeichnet, die sich auf die Länge der Äußerung auswirken. Ihr Vorkommen wird interpretiert als der Versuch des Sprechers, durch Einstreuung redundanten Materials eine Entlastung der Ressourcen zu erwirken. Berthold (1998) zählt die Äußerungsgeschwindigkeit ebenfalls zu dieser Klasse, da eine Reduzierung derselben

den gleichen Effekt hat, wie die Verzögerung durch Pausen usw. Jedes dieser Phänomene ist mit einem Tendenzwert versehen, der angibt, ob ein Anstieg oder ein Abfallen des entsprechenden Wertes unter kognitiver Belastung zu erwarten ist. Die *Übereinstimmung* betrifft den Konsens in den analysierten Studien. Was das Symptom *Satzfragmente* angeht, wurde z.B. in vier von fünf gesichteten Studien ein Ansteigen unter kognitiver Belastung festgestellt.

Tabelle 2.1: Übersicht einiger wichtiger Symptome kognitiver Belastung in der Sprache nach Berthold (1998) (s. auch Berthold & Jameson, 1999)

Qualitätssymptome			Geschwindigkeitssymptome		
Symptom	Tendenz	Übereinstimmung	Symptom	Tendenz	Übereinstimmung
Satzfragmente (Anzahl)	+	4/5	Artikulationsgeschwindigkeit	-	7/7
Fehlansätze (Anzahl)	+	2/4	Sprechgeschwindigkeit	-	7/7
Syntaxfehler (Anzahl)	+	1/1	Einsatzlatenz (Dauer)	+	9/11
Selbstkorrekturen	+,-,0	2,1,4	Stille Pausen (Anzahl)	+	4/5
			Stille Pausen (Dauer)	+	8/10
			Gefüllte Pausen (Anzahl)	+	4/6
			Gefüllte Pausen (Dauer)	+	1/2
			Wiederholungen (Anzahl)	+	5/6

Im Folgenden werden die Definitionen der Symptome aufgeführt, wie sie bei Berthold (1998) verwendet werden. Diese Klassifikation diene als Grundlage für den Begriff der „Symptome von Zeitdruck und kognitiver Belastung“ wie er in dieser Arbeit verwendet wird. Bei der Analyse der experimentellen Daten wurde er jedoch um einige Phänomene erweitert. In Abschnitt 3.3 wird darauf näher eingegangen. Die Beispiele für die unten aufgeführten Definitionen stammen aus dem Korpus des Flughafenexperimentes.

Fehlansätze und Satzfragmente Ein Satzfragment definiert als unvollständige syntaktische Struktur *i*, für die es eine syntaktische Fortführung *c* gibt, so dass *ic* einen wohlgeformten Satz ergibt. Der Sprecher bricht entweder ab (s. Beispiel 2.1) oder beginnt einen neuen Satz (s. Beispiel 2.2). Im letzteren Fall handelt es sich um einen Fehlansatz. In vier von fünf von Berthold (1998) analysierten Studien konnte ein Ansteigen der Häufigkeit dieser Phänomene unter kognitiver Belastung festgestellt werden.

- (2.1) <22/200300/G+/5/001/3.10>
 <EL:2.31> Kann ich die <-/> Roll <-/>

(2.2) <02/100300/G+/3/001/3.12>

<FS/> Wie schwer ist </FS> Wo kann ich mein Gepäck wiegen?

Syntaxfehler Ein Syntaxfehler kann verschiedene Ausprägungen haben. In diesem Symptom fasst Berthold (1998) verschiedene Erscheinungen zusammen, wie fehlende Argumente oder Kongruenzfehler (s. Beispiel 2.3). Zumindest einer der von Berthold (1998) gesichteten Studien untersuchte den Zusammenhang zwischen kognitiver Belastung und Syntaxfehlern. Wie aus Tabelle 2.1 hervorgeht, wurde dort ein Ansteigen der Häufigkeit festgestellt.

(2.3) <09/020600/Q1/4/001/12.47>

Ich möchte gerne eine </> Last-Minute-Ticket nach Los Angeles buchen. Gibt es Last-Minute-Tickets?

Selbstkorrekturen Bei einer Selbstkorrektur werden Teile der bereits artikulierten Äußerung vom Sprecher als fehlerhaft bewertet und korrigiert. Selbstkorrekturen lassen sich nach Berthold (1998) grob in Fehler-Korrekturen und Korrekturen der Angemessenheit unterteilen. Bei Fehler-Korrekturen wird die konzeptuelle Struktur nicht geändert. Es erfolgt eine Korrektur von lexikalischen, syntaktischen und phonetischen Fehlern (s. Beispiel 2.4). Bei der Korrektur der Angemessenheit dagegen erfolgt eine Revision des konzeptuellen Inhaltes im Hinblick auf die Angemessenheit der Äußerung bezüglich Hörer, Sprecherzielen und Dialogsituation (s. Beispiel 2.5). Diese Unterteilung wird in der vorliegenden Arbeit ebenfalls gemacht. Allerdings sprechen wir in diesem Zusammenhang von syntaktischen und inhaltlichen Selbstkorrekturen.

(2.4) <05/200300/G+/3/002/2.65>

<EL:0.84> Wo finde ich ich denn <+s/> der </+s> den Schalter der Lufthansa <S>?

(2.5) <09/020600/G+/3/001/3.34>

<EL:1.69> Wo <+i/> kann </+i> krieg' ich einen Mietwagen <S>?

Die in Tabelle 2.1 aufgeführten Ergebnisse bezüglich der Selbstkorrekturen sind widersprüchlich. Zwei der insgesamt sieben relevanten Studien berichteten von einem Anstieg der Selbstkorrekturen unter Belastung, während vier Studien keine Veränderung feststellen konnten. In einer der Quellen wurde sogar von einem Abfallen der Häufigkeit berichtet. Bei genauerer Betrachtung der Eigenschaften von Selbstkorrekturen ist dieser Widerspruch nicht überraschend. Wir kommen in Abschnitt 3.3.4 darauf zurück.

Artikulations- und Sprechgeschwindigkeit Betrachten wir nun die rechte Seite von Tabelle 2.1, wo die Geschwindigkeitssymptome aufgeführt sind. Die Artikulationsrate definiert sich als die Anzahl der geäußerten Silben geteilt durch die Gesamtdauer der Silben. Gleiches gilt für die Sprechgeschwindigkeit. Berthold (1998) unterscheidet die beiden Maße dahingehend, dass bei der Sprechgeschwindigkeit sämtliche Pausen in die Gesamtdauer eingerechnet werden, während bei der Artikulationsgeschwindigkeit die reine Artikulationsdauer zugrundegelegt wird.

Das Absinken von Artikulations- und Sprechgeschwindigkeit unter Belastung ist ein stabiler Befund in der Literatur (s. Tabelle 2.1). Alle hier zitierten Studien erreichten die Belastung, in dem sie die Schwierigkeit der Sprachaufgabe erhöhten. Berthold und Jameson (1999) erwarten, dass die Verlangsamung durch Einführung einer zweiten Aufgabe noch deutlicher wird.

Stille und gefüllte Pausen Sprechpausen gehören nach Berthold und Jameson (1999) ebenfalls zur Klasse der Geschwindigkeitssymptome. Stille Pausen stellen eine Unterbrechung des Sprachflusses für einen bestimmten Zeitraum dar. In der Literatur werden häufig 200 ms als Untergrenze für eine Pause verwendet. Nach Berthold (1998) sollten die hier interessanten kognitiven Pausen von kommunikativen Pausen unterschieden werden. Kognitive Pausen stellen eine Verzögerung der Sprachproduktion zum Zwecke der Planung dar. Kommunikative Pausen dienen der Markierung von Passagen einer Äußerung, um die Aufmerksamkeit des Hörers zu fokussieren. Letztere können nicht als Belastungssymptome angesehen werden, da sie vom Sprecher bewusst eingesetzt werden. Im Gegensatz dazu gehören gefüllte Pausen ausschließlich der Klasse der kognitiven Pausen an. Es handelt sich um Unterbrechungen des Sprachflusses, die vom Sprecher durch die Äußerung eines Fülllautes überbrückt werden (s. Beispiel 2.6). Nach Berthold (1998) haben gefüllte Pausen eine größere Relevanz als Symptome kognitiver Belastung als stille Pausen. Mit mehr oder weniger großer Übereinstimmung stellten die von Berthold (1998) analysierten Studien ein Ansteigen sowohl von Anzahl, als auch von Dauer der stillen und gefüllten Pausen unter Belastung fest.

(2.6) <27/110500/Q+/3/005/10.26>

<EL:6.51> Wo kann ich <aeh:0.49> hier <aeh:0.31> Wasser zum Trinken holen
<S>?

Wiederholungen Für den Begriff der Wiederholung schlägt Berthold (1998) folgende Definition vor:

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.

Die Äußerung in 2.7 stellt ein Beispiel für eine Wiederholung aus dem Korpus des Flughafenexperimentes dar.

(2.7) <29/230500/G+/4/004/3.04>

<EL:1.04> Wo ist <=/> Wo ist </=> die nächste Apotheke <S>?

Wiederholungen sind verwandt mit gefüllten Pausen, da bei beiden Phänomenen eine Verzögerung im Sprachfluß entsteht, die durch Artikulation von nicht informativen Materials überbrückt werden soll. Dementsprechend wurden die Wiederholungen bei Berthold und Jameson (1999) auch als Geschwindigkeitssymptome klassifiziert. In der Mehrzahl der zugrundeliegenden Studien wurde ein Anstieg unter kognitiver Belastung festgestellt.

2.1.1.1 Eine vergleichende Studie von Oberauer und Hockl (2000)

Oberauer und Hockl (2000) führten ebenfalls eine vergleichende Literaturstudie zu diesem Thema durch. Sie interessierten sich für die Frage, welche Prozesse der Sprachproduktion in welchem Maße die Kapazität des Arbeitsgedächtnisses beanspruchen und wie sich Zeitdruck auf die Sprachproduktion auswirkt. Die bei den beteiligten Prozessen zugrundeliegende Arbeitsgedächtnisstruktur ist im Rahmen dieser Arbeit nicht von zentraler Bedeutung. Die verschiedenen Sprachproduktionsmodelle machen jedoch unterschiedliche Vorhersagen über zu erwartende Symptome, da Prozesse je nach Modell als ressourcenverbrauchend angesehen werden oder nicht. Dabei wird grundsätzlich davon ausgegangen, dass ressourcenverbrauchende Teilprozesse ihrerseits anfällig sind gegen Ressourcenbeschränkungen. Es ist deshalb anzunehmen, dass diese Prozesse zu der Entstehung von Symptomen an der Oberfläche beitragen. Prozesse, die ressourcenunabhängig sind, werden dagegen keine entsprechenden Symptome verursachen.

Die meisten Modelle unterscheiden zwischen automatischen und kontrollierten Teilprozessen der Sprachproduktion. Kontrollierte Prozesse verbrauchen Ressourcen und sind deshalb auch anfällig gegenüber Einschränkungen derselben. Automatische Prozesse dagegen verbrauchen keine Ressourcen. Diese sollten daher auch nicht durch kognitive Belastung beeinflusst werden (s. z.B. Levelt, 1989).

Bock und Levelt (1994) beschreiben folgende Teilprozesse der Sprachproduktion:

1. Aufbau einer nichtsprachliche Konzeption
2. Herstellen der zeitlichen Ordnung zur Erzeugung eines kohärenten Diskurses
3. Planen einzelner Sätze (syntaktische Struktur, Wortauswahl)
4. Festlegen der Artikulation
5. Überprüfen der Angemessenheit und ggf. Korrektur

Nach Levelt (1989) nehmen vor allem Prozesse zur Konzeptualisierung und Überwachung Ressourcen in Anspruch. Bei der Konzeptualisierung müssen das Situationswissen und der bisherige Diskursablauf im Arbeitsgedächtnis gehalten werden. Bei der Überwachung erfolgt ein Parsing der Äußerung. Die entstandene Repräsentation verbraucht für den Zeitraum der Fehlerkorrektur ebenfalls Arbeitsgedächtniskapazität. Beim Überwachungsprozess spielt außerdem das Hörermodell eine Rolle. Die Äußerung muß auf ihre Angemessenheit in Bezug auf den Hörer hin überprüft werden. Zu den automatischen Prozessen gehören nach Bock (1982) dagegen syntaktische und phonologische Planung sowie der Zugriff auf das Lexikon.

Von einem differenzierteren Modell der Sprachproduktion gehen Herrmann und Grabowski (1994) aus. Eine *zentrale Exekutive* bearbeitet die konzeptuelle Repräsentation und die Überwachung. Untergeordnete Enkodiermechanismen programmieren die genaue Phonemabfolge. Die zentrale Exekutive ist die ressourcenverbrauchende Komponente. Die Subsysteme arbeiten weitestgehend automatisch. Die zentrale Exekutive kann jedoch praktisch jede Aufgabe an sich ziehen, wenn die Hilfssysteme scheitern. In diesem Fall verbrauchen diese Prozesse ebenfalls Ressourcen. In welchem Maße das geschieht, hängt unter anderem von der Kommunikationssituation ab (s. Rummer, 1996).

Oberauer und Hockl (2000) vergleichen eine Reihe von Studien auf deren Aussagen über den Ressourcenverbrauch der einzelnen Teilprozesse hin. Die Aussagen, die sie auf Basis der Ergebnisse dieses Vergleichs machen, werden im Folgenden aufgeführt.

Konzeptualisierung Bei der Konzeptualisierung wird der unmittelbare Gegenstand des Sprechens im Arbeitsgedächtnis gehalten. Darüberhinaus müssen Informationen aus dem Langzeitgedächtnis abgerufen werden. Einige Quellen wie z.B. Craik, Govoni, Naveh-Benjamin und Anderson (1996) gehen davon aus, dass das kognitive System in einen *Abrufmodus* schaltet, der dem Langzeitgedächtnisabruf gegenüber anderen Prozess eine höhere Priorität zuweist. Aufgrund der bevorzugten Zuteilung von Arbeitsgedächtniskapazität sollte dieser Prozess also durch eine zusätzliche Belastung, wie etwa durch eine Sekundäraufgabe, nicht beeinflusst werden. Dagegen ist zu erwarten, dass die Leistungen in der Sekundäraufgabe sinken.

Diskursplanung Die Koordination von Handlungszielen ist eine Kernfunktion des Arbeitsgedächtnisses. Bei längeren Diskursen kann jedoch nicht der gesamte Verlauf im Arbeitsgedächtnis gehalten werden. In dem Fall ist eine Interaktion mit dem Langzeitgedächtnis zu erwarten. Dennoch fanden die Autoren wenig Evidenz, dass die Diskursplanung und das Behalten des Diskursprotokolls von der Arbeitsgedächtniskapazität abhängt. Im Rahmen der vorliegenden Arbeit ist die Diskursplanung weniger interessant, weil die Versuchspersonen nur ein bis zwei Sätze produzierten (s. Abschnitt 3.3).

Planung einzelner Sätze Die Bildung einer syntaktischen Struktur wird dann durch eingeschränkte kognitive Ressourcen beeinträchtigt, wenn Teile zur Weiterverarbeitung im Arbeitsgedächtnis zwischengespeichert werden müssen, wie etwa die Speicherung unvollständiger Teilsätze aufgrund intervenierender Satzteile. Aus den Ergebnissen verschiedener Studien schließen Oberauer und Hockl (2000), dass die syntaktische Planung relativ stark von kognitiven Ressourcen abhängig ist. Sie berufen sich unter anderem auf Arbeiten aus der Altersforschung, die einen Rückgang der syntaktischen Komplexität bei Äußerungen feststellten, die von älteren Menschen produziert werden, was auf eine Korrelation mit der Arbeitsgedächtniskapazität schließen lässt. Auch Alzheimer-Patienten bildeten mit zunehmender Stärke der Krankheit weniger komplexe Konstruktionen.

Sprechpausen als Indikatoren für Planungsphasen während der Artikulation, sind nach Oberauer und Hockl (2000) nur bedingt von einer eingeschränkten kognitiven Ressource abhängig. Die von ihnen gesichteten Studien ergaben unterschiedliche Befunde. Während Barch und Berenbaum (1994), Rummer (1996) und Jou und Harris (1992) einen Anstieg der Sprechpausen bei kognitiver Belastung berichteten, stellte Roßnagel (1995) fest, dass Versuchspersonen unter höherer kognitiver Belastung weniger Sprachpausen innerhalb der Phrasen, dafür mehr extraphrasale Pausen machten. Die unspezifischen Aussagen über Sprechpausen sind nach Meinung von Oberauer und Hockl (2000) z.T. darin begründet, dass diese sehr unterschiedliche Ursachen haben können: die Suche des Sprechers nach dem richtigen Wort, die Auswahl der richtigen Verbalisierung eines gegebenen Bedeutungsinhaltes, oder auch die Verfolgung kommunikativer Ziele wie die Übergabe des Rederechts.

Auswahl von Wörtern Die meisten Theorien der Sprachproduktion beschreiben diesen Vorgang als Aktivierung von Knoten innerhalb eines Netzwerkes (s. z.B. Levelt, 1989; McClelland & Elman, 1986). Betrachtet man das Arbeitsgedächtnis als insgesamt zur Verfügung stehende Aktivationsenergie, so kann diese auch durch andere Prozesse verbraucht werden. Eine adäquate Aktivierung lexikalischer Knoten kann somit unter Umständen beeinträchtigt werden (s. Just & Carpenter, 1992). In anderen von den Autoren untersuchten Quellen wird eine alternative Erklärung angeboten: das Arbeitsgedächtnis wird benötigt zur Unterdrückung irrelevanter Informationen. Unter Belastung treten vermehrt Fehler bei der Wortauswahl oder Verzögerungen auf (s. Engle, Conway, Tuholski & Shisler, 1995). In anderen Studien (s. z.B. Engle et al., 1995) konnte dagegen kein Niveauabfall beim lexikalischen Zugriff von Versuchspersonen geringer Arbeitsgedächtniskapazität festgestellt werden. Insgesamt ziehen Oberauer und Hockl (2000) den Schluß, dass es für eine Beschränkung der Auswahl von Wörtern durch kognitive Ressourcen „keine überzeugende empirische Unterstützung“ gibt.

Phonologische Kodierung Die Abhängigkeit der phonologischen Kodierung von der kognitiven Ressource wird häufig anhand von Versuchen mit Stotterern untersucht. Bosshardt (1999) berichtete von einem kurzzeitigen Absinken der Stotterrate, wenn eine Sekundäraufgabe präsentiert wird, und erklärt dies mit einer temporären Aktivierung der zentralen Exekutive zugunsten der Sprechflüssigkeit. Während der Bearbeitung der Sekundäraufgabe stieg die Stotterrate jedoch stark an. Die Leistungen in der Nebenaufgabe gingen gleichzeitig zurück. Dies ist ein Zeichen dafür, dass beide Prozesse auf die gleiche Ressource zugriffen, nämlich die zentrale Exekutive. Nach Starkweather (1995) ist ein Anstieg der Stotterrate auch unter Zeitdruck zu erwarten. Als weiterer Indikator für die Belastung können nach Oberauer und Hockl (2000) Versprecher angesehen werden. Sie berufen sich auf eine Studie von Daneman und Green (1986) sowie Bredenkamp und Dillger (1998), die bei ProbandInnen niedrigerer Arbeitsgedächtniskapazität deutlich mehr sogenannte *Spoonerismen* hervorrufen konnten, als bei Versuchspersonen, die über eine größere Arbeitsgedächtnisleistung verfügten. Spoonerismen gehören zu den bekanntesten Sprechfehlern. Dabei werden die Anfangsphoneme zweier oder mehrerer Wörter bei der Artikulation vertauscht. Der Priester William Archibald Spooner, nach dem die Spoonerismen benannt sind, produzierte bekannte Beispiele wie „You have hist my mystery lectures“ (s. Eysenck & Keene, 1990). Solche Sprechfehler werden oft in Experimenten untersucht, bei denen Aspekte der menschlichen Sprachproduktion erforscht werden sollen. Die Relevanz der Befunde solcher Sprechfehler bezüglich allgemeiner Fragen der Sprachproduktion sind jedoch umstritten (s. Harley, 1995). Oberauer und Hockl (2000) gehen davon aus, dass dieser Teilprozeß eher nicht von einer eingeschränkten kognitiven Ressource abhängt.

Überwachung und Selbstkorrektur Korrekturen geplanter Äußerungen, die noch vor Beginn der Artikulation vorgenommen werden, lassen sich empirisch kaum nachweisen. Anders ist es mit Korrekturen nach der Artikulation. Jou und Harris (1992) fanden unter Doppelaufgabenbelastung mehr Sprechfehler, jedoch nicht mehr Selbstkorrekturen. Die Autoren schlossen daraus eine hohe Anforderung der Korrektur an die Aufmerksamkeitsressource. Auch Roßnagel (1995) fand unter geringerer Belastung mehr abgebrochene Sätze, was er als Zeichen einer enger kontrollierten Sprachproduktion deutet. Die Kor-

rektur als beobachtbares Zeichen der Überwachung birgt ein grundsätzliches Problem: es gibt begründete Annahmen sowohl für das vermehrte als auch das verminderte Auftreten von Korrekturen unter kognitiver Belastung. Ein vermehrtes Auftreten könnte man dann annehmen, wenn durch die Belastung mehr Fehler gemacht und entdeckt werden, die dann korrigiert werden. Es ist jedoch auch möglich, dass die Überwachung so stark eingeschränkt wird, dass weniger Fehler entdeckt und korrigiert werden können. Aus diesem Grund ist auch nach Oberauer und Hockl (2000) in dieser Frage „keine Aussage möglich“. In Abschnitt 3.3.4 wird auf dieses Problem nochmals eingegangen.

Partnerorientierung Nach Oberauer und Hockl (2000) werden die Informationen über den Gesprächspartner in einem Hörermodell ständig im Arbeitsgedächtnis verfügbar gehalten. Herrmann und Grabowski (1994) und Roßnagel (1995) ließen die Versuchspersonen den Aufbau einer Maschine einem Studenten bzw. einem Kind erklären. In der unbelasteten Bedingung verwendeten die Sprecher gegenüber dem Kind weniger Fachbegriffe und gaben genauere Anweisungen. Unter Belastung verschwand der Unterschied. Die Hörerorientierung erfolgt demnach nicht ohne Kosten und wird bei Belastung vermindert.

Roßnagel (1996) führte weitere Untersuchungen bezüglich der Partnerorientierung durch. Er lies seine Versuchspersonen einem anwesenden Hörer den Aufbau einer Maschine erklären, den sie zuvor selbst erlernt hatten. Neben der Belastung manipulierte Roßnagel (1996) die Motivation der Versuchspersonen: eine Gruppe erhielt eine Belohnung, wenn der Hörer aufgrund ihrer Instruktion ein Modell der Maschine anschließend auf Anhieb zusammenbauen konnte. Die Partnerorientierung der belasteten Gruppe ohne Belohnung war geringer als die der entsprechenden unbelasteten Gruppe. Die Gruppe mit Belohnung zeigte insgesamt mehr Partnerorientierung. Außerdem waren belastete und unbelastete Gruppen nahezu gleich. Nach Roßnagel (1996) legen diese Ergebnisse nahe, dass unter kognitiver Belastung die Partnerorientierung als zusätzliche Beanspruchung vermieden wird. Gleichzeitig weist er auf die Bedeutung motivationaler Faktoren bei der Modellierung der kognitiven Prozesse beim Sprechen hin.

Horton und Keysar (1996) führten ebenfalls Untersuchungen zur Partnerbezogenheit durch. Nach ihrem *Monitoring and Adjustment* Modell gibt es unter den Teilprozessen der Sprachproduktion einen separaten Kontrollmechanismus, der für die Anpassung der Äußerung an den Partner zuständig ist. In einem Experiment untersuchten sie die Auswirkungen von Zeitdruck auf die Partnerbezogenheit der Äußerungen ihrer Versuchspersonen. Dabei gingen sie davon aus, dass Überwachungsprozesse besonders sensitiv gegenüber Zeitrestriktionen sind. Sie fanden heraus, dass die Partnerbezogenheit unter Zeitdruck deutlich zurückgeht, was sie als Evidenz für das Monitoring and Adjustment Modell interpretieren.

In Tabelle 2.2 werden die Ergebnisse von Oberauer und Hockl (2000) zusammenfassend dargestellt. „Vergleichsweise eindeutige Evidenz“ für die Abhängigkeit von einer begrenzten Ressource konnten sie für das Erstellen des konzeptuellen Inhaltes, die Syntaxplanung und die Anpassung der Sprache an den Partner feststellen. Eine wichtige Rolle spielt nach ihrer Ansicht dabei das Arbeitsgedächtnis.

Tabelle 2.2: Abhängigkeit verschiedener Teilprozesse der Sprachproduktion von kognitiven Ressourcen nach Oberauer und Hockl (2000)

Teilprozeß	Aussage aufgrund bisheriger empirischer Ergebnisse
Konzeptualisierung	abhängig
Diskursplanung	keine Aussage möglich
Syntaxplanung	abhängig
Wortauswahl	nicht abhängig
Artikulation	mit Einschränkung nicht abhängig
Überwachung	keine Aussage möglich
Partnerorientierung	abhängig

2.1.2 Vergleich der Übersichtsstudien

Die Phase der Konzeptualisierung ist nach Oberauer und Hockl (2000) ressourcenabhängig. Das bedeutet, dass eine Ressourcenbeschränkung sich an der Oberfläche durch folgende Symptome bemerkbar machen sollte: Die Einsatzlatenz als initiale Planungsphase, sollte länger werden. Sprechpausen, insofern sie kognitive Pausen sind, sollten zunehmen bzw. länger werden. Außerdem sollte ein Zuwachs von Symptomen zu erkennen sein, die als Versuch der Gewinnung zusätzlicher Planungszeit interpretiert werden können, wie z.B. Wiederholungen. Diese Überlegungen sind konsistent mit den Ergebnissen von Berthold (1998), die in Tabelle 2.1 dargestellt sind.

Wenn Satzfragmente und Fehlansätze als Folge fehlerhafter Konzeptualisierung angesehen werden können, ist ihr Ansteigen unter Belastung ebenfalls mit den Ergebnissen von Oberauer und Hockl (2000) konsistent. Bricht ein Sprecher einen Satz aufgrund einer fehlerhaften Konzeptualisierung ab, sind dazu jedoch auch Überwachungsprozesse nötig. Oberauer und Hockl (2000) treffen keine konkrete Aussage über die Auswirkungen einer Ressourcenbeschränkung auf Überwachungsprozesse. In noch stärkerem Maße betrifft dies die Selbstkorrekturen, was sich auch bei Berthold (1998) in der widersprüchlichen Befundlage bezüglich dieses Symptoms widerspiegelt. In Abschnitt 3.3.4 wird auf diese Problematik genauer eingegangen.

Oberauer und Hockl (2000) erachten die Planung der syntaktischen Struktur als ressourcenabhängig. Dies ist konsistent mit den Ergebnissen von Berthold (1998), obwohl dort lediglich eine einzelne Studie ein Ansteigen von Syntaxfehlern unter Belastung feststellte (s. Tabelle 2.1).

Das Ergebnis aller sieben relevanter Studien, die Berthold (1998) analysierte, war ein Absinken von Artikulationsgeschwindigkeit bzw. Sprechgeschwindigkeit bei einer Ressourcenbeschränkung. Dies ist nicht konsistent mit der Annahme von Oberauer und Hockl (2000), dass der Teilprozeß der Artikulation nicht ressourcenabhängig ist. Unter Umständen ist das darauf zurückzuführen, dass die Artikulationsgeschwindigkeit nicht direkt mit besagtem Teilprozeß in Beziehung steht. Vielmehr legt dies den Schluß nahe, dass der Rückgang der Artikulationsgeschwindigkeit als Versuch des Sprechers zu inter-

pretieren ist, Planungszeit zu gewinnen.

2.2 Der Einfluß der Äußerungskomplexität auf das Vorkommen von Belastungssymptomen

Oviatt (1997) untersuchte in einer Reihe von Experimenten die Eigenschaften multimodaler Eingaben in verschiedenen Situationen. Sie ließ ihre Versuchspersonen mit elektronischen Landkarten interagieren, wobei die Eingabemodalität zwischen Stifteingabe, Spracheingabe und multimodaler Eingabe (sowohl Stift als auch Sprache) wechselte. Oviatt (1997) war primär an der Frage interessiert, welcher Eingabemodus in welcher Situation von den Versuchspersonen bevorzugt wird. Sie stellte fest, dass eine allgemeine Präferenz zu multimodalen Schnittstellen besteht.

Ein Ziel von Oviatt (1997) war die Reduzierung von Eingabefehlern durch die Wahl einer Kombination von Eingabemodi, die der Aufgabe angemessen ist. Was die Spracheingabe betrifft, stellte sie unter anderem fest, dass es eine starke Beziehung zwischen der Äußerungskomplexität und der Anzahl von *Disfluencies* gibt. Als *Disfluencies* bezeichnet sie die Menge der sprachlichen Phänomene, die hier Belastungssymptome genannt werden. Sie untersuchte das Vorkommen von Selbstkorrekturen, Fehlsätzen, syntaktischen Fehlern, Wiederholungen, gefüllten Pausen, und Wortfragmenten. Als Maß für die Äußerungskomplexität legte sie die Anzahl der Wörter zugrunde. Sie stellte fest, dass es eine starke Korrelation zwischen Äußerungskomplexität und dem Vorkommen von *Disfluencies* gibt ($F = 4.0, p < 0.05$).

Unter Berücksichtigung dieses Ergebnisses ist zu erwarten, dass die Komplexität der Äußerung einen direkten Einfluß auf das Vorkommen von Symptomen hat. Eine Erhöhung von Zeitdruck und/oder kognitiver Belastung könnte dazu führen, dass der Sprecher Äußerungen von geringerer Komplexität erzeugt, um die Mehrbelastung zu kompensieren. Das würde wiederum bedeuten, dass an der Oberfläche eine geringere Anzahl von Symptomen zu messen ist. Diese Komplikation ist bei der Interpretation von Ergebnissen von Studien zu beachten, die sich mit der Beziehung von Sprachproduktion und Ressourcenbeschränkungen beschäftigen. Das gilt auch für die Ergebnisse des Flughafenexperimentes. In Abschnitt 3.4.1 wird darauf näher eingegangen.

Kapitel 3

Eigene empirische Studien

3.1 Einleitung

In Kapitel 2 wurden Ergebnisse aus früheren Studien aufgeführt, die sich mit den Auswirkungen von Zeitdruck und/oder kognitiver Belastung auf die Sprachproduktion beschäftigen. Es erfolgte in erster Linie ein Vergleich zweier Literaturstudien von Berthold (1998) und Oberauer und Hockl (2000), die aus unterschiedlichen Blickwinkeln jeweils eine zusammenfassende Darstellung entsprechender Studien präsentierten. In diesem Kapitel sollen eigene empirische Untersuchungen des Verfassers beschrieben werden. In Abschnitt 3.2 wird eine Feldstudie am Frankfurter Flughafen präsentiert, die als Vorstudie für das Flughafenexperiment diente. Das Experiment selbst ist Thema des restlichen Teils dieses Kapitels. Zunächst wird jedoch darauf eingegangen, warum es angesichts des zuvor beschriebenen Stand der Forschung notwendig erschien, ein solches Experiment durchzuführen.

3.1.1 Motivation

Menge der untersuchten Phänomene Die Quellen, die Berthold (1998) und Oberauer und Hockl (2000) als Grundlage für ihre Übersichtsstudien dienten, untersuchten jeweils nur einen Teil der relevanten Phänomene. Zum Teil könnte das daran liegen, dass die ursprünglichen Fragestellungen speziell auf einen bestimmten Aspekt der Sprachproduktion beschränkt waren, und aus diesem Grund keine Notwendigkeit bestand, die Untersuchungen auf andere Phänomene auszuweiten. Roßnagel (1996) z.B. war bei seiner Studie lediglich an denjenigen Aspekten der Sprachproduktion interessiert, die einen direkten Bezug zur Partnerorientierung hatte. Es war deshalb unklar, welche Auswirkungen es auf die oben beschriebenen Ergebnisse bezüglich der Sprachproduktion unter Belastung bzw. Zeitdruck haben würde, wenn alle relevanten Phänomene in einer einzigen Studie untersucht würden.

Übertragbarkeit der Ergebnisse auf das Flughafenszenario Darüberhinaus stellt sich die Frage, ob die Ergebnisse aus den bisherigen Studien auf die Situation übertragbar sind, die durch das Flughafenszenario gegeben ist. Zum einen betrifft das die Sprachproduktion selbst: sie erfolgt innerhalb einer bestimmten Kommunikationssituation (Mensch-Maschine) und ist inhaltlich auf die eingeschränkte Domäne „Flughafen“ beschränkt. Die Art der Belastung spielt ebenfalls eine Rolle. Wie bereits in Abschnitt 1.1

erwähnt, unterliegen Reisende am Flughafen ganz bestimmten Ressourcenbeschränkungen. Es ist denkbar, dass sich diese Art der Belastung anders auf die Sprache auswirkt, als es bei den bisherigen Untersuchungen der Fall war.

Zeitdruck als Ressourcenbeschränkung Obwohl Berthold (1998) von einigen Quellen berichtet, die die Auswirkungen von Zeitdruck auf die Sprachproduktion untersuchen, stellt dieser einen weniger gut erforschten Aspekt von Ressourcenbeschränkung dar. Die überwiegende Mehrheit der oben aufgeführten Studien untersuchen lediglich die Auswirkungen von Arbeitsgedächtnisbelastung. Zeitdruck ist jedoch gerade im Hinblick auf das gegebene Flughafenszenario eine wichtige Ressourcenbeschränkung (s. Abschnitt 1.1)

Die Existenz eines Speed-Accuracy-Tradeoffs in der Sprachproduktion Wie bereits aus Tabelle 2.1 hervorgeht, unterscheidet Berthold (1998) zwischen Qualitätssymptomen und Geschwindigkeitssymptomen. Die Ursache für Qualitätssymptome sieht Berthold (1998) darin, dass der Sprecher durch die gegebene Ressourcenbeschränkung nicht genügend Kapazität zur Verfügung hat, um die Qualität der Sprache aufrecht zu erhalten. Die Geschwindigkeitssymptome dagegen sind eine Folge einer Verlangsamung der Sprachproduktion zur Kompensation einer Ressourcenbeschränkung. Berthold und Jameson (1999) schlagen vor, dass es zwischen Qualität und Geschwindigkeit einen gegenseitigen negativen Einfluß gibt (*Speed-Accuracy-Tradeoff*). Versucht der Sprecher, die Qualität der Sprache bei einer Ressourcenbeschränkung aufrecht zu erhalten, kann dies zu einer Verlangsamung im Sinne der oben genannten Geschwindigkeitssymptome führen. Eine Aufrechterhaltung der Geschwindigkeit kann umgekehrt dazu führen, dass die Qualität der Sprache abnimmt, was sich in einer zunehmenden Häufigkeit der entsprechenden Symptome zeigen sollte. Berthold und Jameson (1999) gehen dabei davon aus, dass dies zu einem gewissen Grad von der Strategie des Sprechers abhängig ist. Je nach Situation kann mehr Wert auf die Qualität bzw. die Geschwindigkeit gelegt werden.

Abbildung 3.1 stellt die Formulierung dieses Speed-Accuracy-Tradeoffs in Form eines vereinfachten Baysschen Netzes dar. In einem solchen Netz werden kausale Zusammenhänge zwischen verschiedenen Variablen zum Ausdruck gebracht¹. Ein Rechteck mit durchgezogener Linie entspricht einer einzelnen Variable. Rechtecke mit unterbrochenen Linien stellen Gruppen von Variablen dar, die eine ähnliche Rolle innerhalb des Netzes spielen. Die durchgezogenen und unterbrochenen Kanten bedeuten einen positiven bzw. negativen kausalen Einfluß.

Das Netz ist wie folgt zu interpretieren: angenommen, ein Sprecher unterliegt einer bestimmten POTENTIELLEN AG-BELASTUNG, die er nicht mühelos bewältigen kann. Diese Überlastung kann der Sprecher vermindern, indem er die Geschwindigkeit der Sprachproduktion reduziert, z.B. durch Einfügen von zusätzlichen Planungspausen. In welchem Maß er das tut kann von den Gegebenheiten der Aufgabe abhängen, sowie von dem Zeitdruck, dem der Sprecher unterliegt und seinen Präferenzen. Die Geschwindigkeitsverminderung kann sich z.B. in einer Reduzierung der ARTIKULATIONSGESCHWINDIGKEIT äußern oder in einer Reihe ANDERER GESCHWINDIGKEITSVERRINGERNDER SYMPTOME, die in Tabelle 2.1 als Geschwindigkeitssymptome bezeichnet werden. Aufgrund der Verlangsamung kann die TATSÄCHLICHE AG BELASTUNG reduziert werden. Diese entspricht nach Berthold und Jameson (1999)

¹In Abschnitt 3.3.2 wird näher auf die Bayssche Netze als Inferenzmechanismus eingegangen

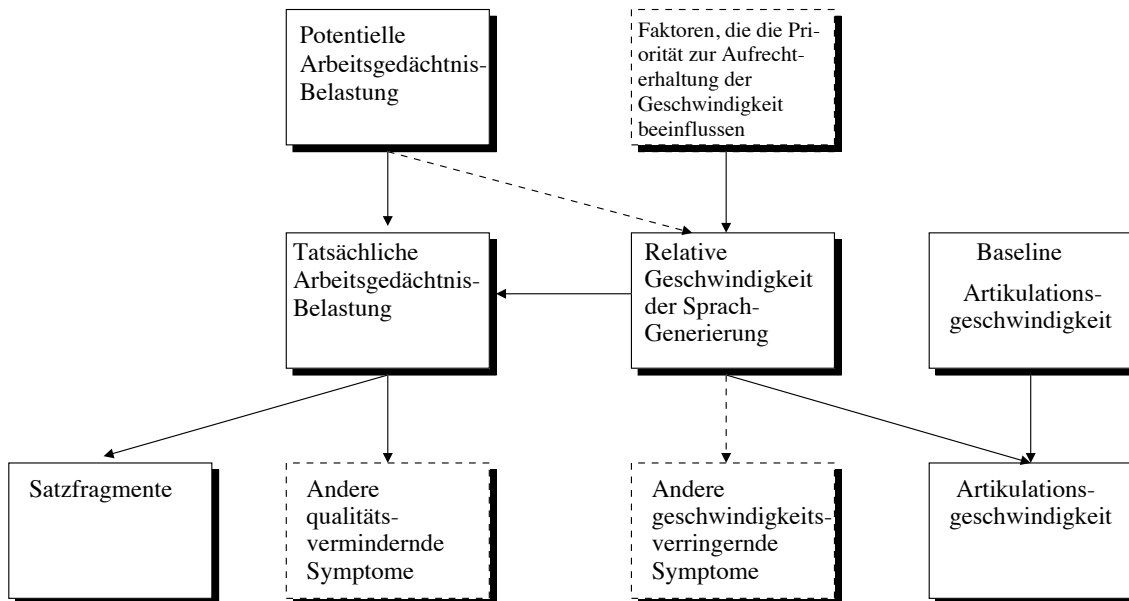


Abbildung 3.1: Vereinfachtes Bayssches Netz zur Interpretation von Symptomen kognitiver Belastung nach Berthold und Jameson (1999)

der Menge der kognitiven Arbeit, die in einem bestimmten Zeitraum durchzuführen ist. Auf der anderen Seite sprechen auch Gründe dafür, dass der Sprecher eine solche Verlangsamung vermeiden sollte. In so einem Fall ist es wahrscheinlich, dass die TATSÄCHLICHE AG BELASTUNG sich in einem Qualitätsrückgang bei der Sprachproduktion niederschlägt, z.B. Form eines verstärkten Auftretens von SATZFRAGMENTEN ODER ANDEREN QUALITÄTSVERMINDERNDEN SYMPTOMEN (Qualitätssymptome aus Tabelle 2.1).

Der Begriff Speed-Accuracy-Tradeoff bezeichnet das in der kognitiven Psychologie bekannte Phänomen, dass die Genauigkeit zurückgeht, wenn die Geschwindigkeit einer Entscheidung oder Handlung steigt (s. z.B. Townsend & Ashby, 1983). Die numerischen Relationen zwischen Genauigkeit und Geschwindigkeit sind in den meisten Fällen nicht genau bekannt und basieren auf empirischen Untersuchungen.

Eine Reihe von Untersuchungen beschäftigen sich mit dem Thema der Speed-Accuracy-Tradeoffs, ohne jedoch einen direkten Bezug auf die Sprachproduktion zu nehmen. Kray und Lindenberger (2000) untersuchen die Leistungen von Personen verschiedenen Alters beim Gehen über einen vorgegebenen Parcours. Personengruppen im Alter von 20-30 Jahren, 40-50 Jahren, und 60-70 Jahren wurden instruiert, so schnell wie möglich und so exakt wie möglich schmalen Pfaden zu folgen. Die Geschwindigkeit und Genauigkeit des Gehens wurde mit und ohne zusätzliche Belastung ermittelt.

Die Personen der oberen Altersgruppen verminderten ihre Geschwindigkeit beim Gehen unter Belastung stärker als die Personen zwischen 20 und 30 Jahren. Am stärksten war die Geschwindigkeitsverminderung bei der mittleren Personengruppe. Die Genauigkeit des Gehens (gemessen in der Abweichung vom vorgegebenen Parcours) sank jedoch lediglich bei der ältesten Gruppe ab. Hier liegt nach Kray und Lindenberger (2000) möglicherweise ein Speed-Accuracy-Tradeoff vor: während die Gruppe mittleren Alters

die Auslastung des Arbeitsgedächtnisses durch Verlangsamung weitestgehend ausgleichen konnte, ist dies bei den Personen von 60-70 Jahren nicht mehr in ausreichendem Maße möglich, was sich in der Genauigkeit des Gehens auswirkt.

In der Belastungsbedingung memorierten die Versuchspersonen Wortlisten. Kray und Lindenberger (2000) stellten fest, dass sie mit zunehmendem Alter für das Gehen mehr Aufmerksamkeit benötigten, was sich vor allem unter Belastung bemerkbar machte. Dies versuchten sie durch Verlangsamung auszugleichen.

Was die Sprachproduktion angeht, gibt es meines Wissens derzeit keine Untersuchungen, die sich mit der Relation zwischen Äußerungsqualität und -geschwindigkeit befassen. Lediglich bei spezifischen Teilprozessen, die eine Entscheidungskomponente beinhalten (z.B. Lexikonzugriff), wurde ein Speed-Accuracy-Tradeoff postuliert (s. Townsend & Ashby, 1983). Um Evidenz für die Existenz eines solchen Tradeoffs im Sinne von Berthold und Jameson (1999) zu erhalten, könnte ein Experiment sinnvoll sein, bei dem die Priorität zur Aufrechterhaltung der Geschwindigkeit direkt manipuliert wird. Die genaue Evaluation der Struktur, wie sie in Abbildung 3.1 dargestellt wird, kann im Rahmen dieser Arbeit jedoch nicht erfolgen, weil verborgene Variablen enthalten sind. Es wird überprüft, ob sich Geschwindigkeits- und Qualitätssymptome wie zu erwarten verhalten.

3.1.2 Zusammenfassung der Fragestellungen

Zusammenfassend können die Fragestellungen des Flughafenexperimentes wie folgt dargelegt werden:

1. Wie sind die Auswirkungen von Zeitdruck auf die sprachlichen Phänomene, die bislang zumeist nur im Zusammenhang mit Arbeitsgedächtnisbelastung untersucht wurden?
2. Können die Ergebnisse aus den bisherigen Studien repliziert werden, wenn die Symptome alle gleichzeitig in einem Experiment untersucht werden?
3. Können die Ergebnisse aus den bisherigen Studien repliziert werden, die Sprachaufgabe und die Art der Belastung möglichst realistisch im Bezug auf das Flughafenszenario gestaltet werden?
4. Inwieweit existiert das von Berthold und Jameson (1999) postulierte Speed-Accuracy-Tradeoff in der Sprachproduktion?

Um eine Antwort auf diese Fragen zu finden, wurde ein Experiment entworfen, mit dem die Sprachproduktion im Kontext des Flughafenszenarios untersucht wurde. Die Versuchspersonen schlüpfen in die Rolle von Reisenden an einem Großflughafen, die Informationen zu bestimmten Problemen benötigen, und deshalb entsprechende Anfragen stellen. Neben der kognitiven Belastung wurden diejenigen Faktoren direkt manipuliert, die die Wahl der Strategie im Sinne von Abbildung 3.1 bestimmen: die Anforderungen an die Geschwindigkeit und die Qualität der Äußerungen.

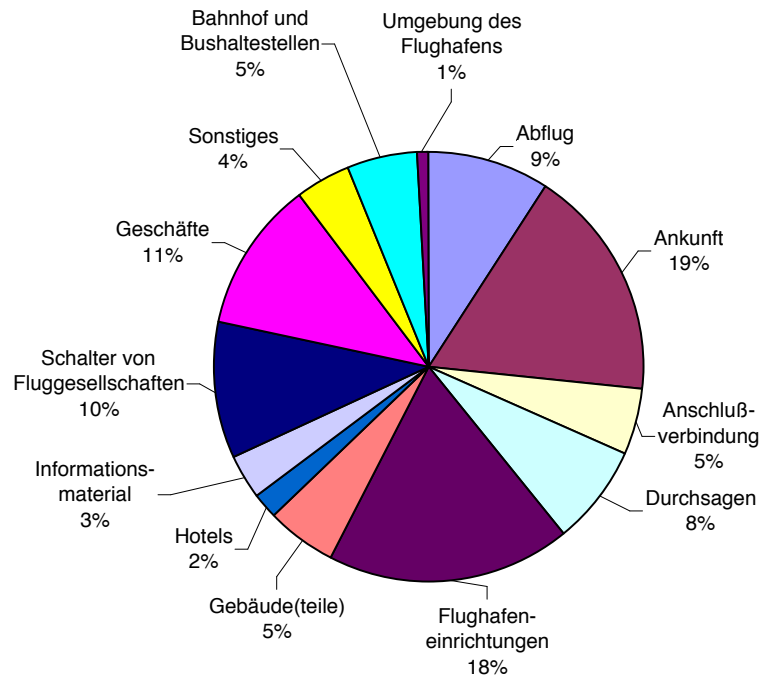


Abbildung 3.2: Inhalte der am Frankfurter Flughafen aufgezeichneten Anfragen

3.2 Eine Feldstudie am Frankfurter Flughafen

Um einen Eindruck zu gewinnen, welche Art von Anfragen an einem konventionellen Informationsschalter am Flughafen anfallen, wurde vom Verfasser im Vorfeld des Experimentes eine Feldstudie am Flughafen Frankfurt (Main) durchgeführt. Einen Tag lang wurden an verschiedenen Schaltern insgesamt 250 Anfragen notiert. Außerdem haben Befragungen des Personals betreffend ihrer Erfahrungen stattgefunden. Die Ergebnisse dieser Feldstudie werden in diesem Abschnitt aufgeführt.

Im Allgemeinen konnte beobachtet werden, dass immer wieder die gleichen Anfragen gestellt werden. Am Häufigsten beziehen sie sich auf Objekte oder Einrichtungen in der unmittelbaren Nähe des Schalters. Abbildung 3.2 stellt einen Überblick über die Ergebnisse dar. Eine grobe Untergliederung lässt sich, ganz entsprechend der Struktur des Flughafens, nach Ankunfts- und Abflugsbereich vornehmen. Im Abflugsbereich fragen Reisende nach Schaltern von Fluggesellschaften oder Geschäften. Darüberhinaus stellen sie Fragen, die direkt ihren Abflug betreffen (Uhrzeit, Gatenummer usw.).

Im Ankunfts-bereich dagegen sind Hotels und Anschlußverbindungen Gegenstand der Anfragen. Darüberhinaus konnte festgestellt werden, dass die Informationsschalter neben den Reisenden dort häufig von einer zweiten Personengruppe besucht werden, nämlich den „Abholenden“. Damit sind diejenigen Personen gemeint, die einen Reisenden am Flughafen in Empfang nehmen wollen (oder sollen). Sie interessieren sich vor allen Dingen dafür, wann ein bestimmtes Flugzeug landet, oder welchen Weg die Passagiere nehmen, wenn sie den geschlossenen Bereich verlassen. In Abbildung 3.2 entsprechen diese Art von Anfragen der Kategorie „Ankunft“. Oft bitten sie das Personal am Informati-

onsschalter, einen Passagier über die Lautsprecheranlage ausrufen zu lassen. Für diese Personengruppe wäre eine Unterstützung durch ein Assistenzsystem ebenfalls denkbar, z.B. durch Wegbeschreibungen und Information über relevante Ankunftszeiten. Wenn man davon ausgeht, dass sowohl Reisende als auch Abholende über ein Assistenzsystem verfügen, wäre auch die Zusammenführung von Paaren beider Gruppen vorstellbar.

Im Rahmen des Flughafenexperimentes sind jedoch lediglich die Anfragen der Reisenden relevant. Aus den notierten Beispielen wurde ein Katalog von realen Problemsituationen erstellt, der als Grundlage für den Entwurf der experimentellen Stimuli diente (s. Abschnitt 3.3.2.3)

3.3 Ein Experiment zur Untersuchung von Sprache unter kognitiver Belastung und Zeitdruck

3.3.1 Vorüberlegungen

Bevor die Methode erläutert wird, die in diesem Experiment angewandt wurde, werden im folgenden Abschnitt einige Vorüberlegungen besprochen, die zentrale Designentscheidungen beeinflusst haben. Zunächst werden in Abschnitt 3.3.1.1 einige theoretische Überlegungen zum Doppelaufgabenparadigma gemacht, wobei insbesondere die Frage der Auswahl einer geeigneten Nebenaufgabe von Interesse ist. Anschließend wird in Abschnitt 3.3.1.2 die Auswahl der Kommunikationssituation diskutiert. Abschnitt 3.3.1.3 beschäftigt sich mit alternativen Möglichkeiten des Herstellens von Zeitdruck. In Abschnitt 3.3.2.9 wird schließlich auf die besondere Rolle der Anweisung in diesem Experiment eingegangen.

3.3.1.1 Kriterien bei der Auswahl einer geeigneten Nebenaufgabe in Doppelaufgabenexperimenten

Das Doppelaufgabenparadigma ist eine weit verbreitete experimentalpsychologische Forschungsmethode (s. Gopher, 1994). Der Grundgedanke dieses Paradigmas ist, dass Menschen aufgrund einer kapazitativen Begrenzung des kognitiven Systems nicht dazu in der Lage sind, unbegrenzt viele informationsverarbeitende Operationen gleichzeitig auszuführen. In dem Moment, in dem eine energetische Ressource ausgeschöpft ist, kommt es zu Interferenzen (s. Just & Carpenter, 1992). Im Rahmen von Doppelaufgabenexperimenten machen sich diese Interferenzen meist durch Rückgang der Verarbeitungsgeschwindigkeit und/oder eine Erhöhung der Anzahl von Handlungsfehlern bemerkbar. Diese Methode bringt eine Besonderheit mit sich: der Primär- und Sekundäraufgabe kommt sowohl die Funktion einer abhängigen als auch einer unabhängigen Variablen zu. Die Leistungen der Versuchsperson in jeder der Aufgabe zeigt die Belastung durch die jeweils andere Aufgabe an.

Ein entscheidendes Kriterium für den Erfolg eines Doppelaufgabenexperimentes ist die Wahl einer geeigneten Nebenaufgabe. Der folgende Abschnitt wird sich deshalb mit der Frage auseinandersetzen, welche Faktoren bei der Auswahl eine Rolle spielen.

Klassifikation von Nebenaufgaben Prinzipiell kann jede Aufgabe mit einer anderen Aufgabe kombiniert werden. Nach Wieland-Eckelmann (1992) sind zur besseren Kon-

trolle jedoch standardisierte Nebenaufgaben wünschenswert. Gerade im Zusammenhang mit einer Sprachproduktionsaufgabe, die an sich nur schwer zu reglementieren ist, sollte die Sekundäraufgabe so standardisiert wie möglich sein.

Rummer (1996) schlägt folgende Unterscheidungskriterien von Nebenaufgaben vor:

1. zentrale oder periphere Belastung
2. kontinuierliche oder diskontinuierliche Belastung
3. lokale oder globale Messung der Leistung

Unter zentral belastend verstehen wir eine Belastung der zentralen Ressource (s. Abschnitt 2.1.1.1). Peripher ist eine Belastung dagegen dann, wenn sie sich auf spezifisches Subsystem bezieht, wie z.B. die *phonologische Schleife*, wie sie von Baddeley (1986) postuliert wird. Diese ist Teil des Baddeley'schen Arbeitsgedächtnismodells und hat spezifische Aufgaben, die z.B. beim Memorieren von Ziffernfolgen eine Rolle spielen. In einer Reihe von Experimenten im Rahmen der Arbeitsgedächtnisforschung wurde die phonologische Schleife durch geeignete Aufgaben gezielt gestört. Eine solche Belastung bezeichnen wir als peripher.

Zentrale Belastung wird nach Baddeley (1993) vor allem durch Entscheidungsaufgaben verursacht. Es sollte daher bei der Gestaltung von Nebenaufgaben darauf geachtet werden, dass eine solche Entscheidungskomponente enthalten ist. Dagegen sollte die Aufgabe möglichst wenig strukturelle Ähnlichkeiten mit der Hauptaufgabe haben, um sicher gehen zu können, dass die vorgefundenen Effekte nicht auf strukturelle Interferenzen zurückgehen.

Beansprucht die Nebenaufgabe das kognitive System gleichmäßig über den gesamten Bearbeitungszeitraum, spricht man von einer kontinuierlich belastenden Aufgabe. Diskontinuierlich belastende Aufgaben weisen dagegen punktuelle Maxima auf. In der unten aufgeführten Übersicht werden Beispiele aus beiden Kategorien genannt.

Eine Besonderheit des Doppelaufgabenparadigmas stellt die Tatsache dar, dass anhand der Leistungen in der einen Aufgabe, die Belastung der anderen Aufgabe gemessen werden kann. Können die Leistungen in der Nebenaufgabe ständig gemessen werden, ist eine Synchronisation mit der Hauptaufgabe möglich (lokale Messungen). Ist die Nebenaufgabe jedoch so geartet, dass eine permanente Leistungsmessung nicht erfolgen kann, kann keine Aussage über den Belastungsverlauf der Hauptaufgabe getroffen werden (globale Messungen).

Nach diesen Kriterien teilt Rummer (1996) Nebenaufgaben in acht verschiedene Typen ein (s. Tabelle 3.1).

Die meisten Doppelaufgaben sind jedoch keine reinen Ausprägungen dieser Typen. Diskontinuierliche Aufgaben können z.B. durch Einführung einer Gedächtnisanforderung kontinuierlicher gestaltet werden.

Ein Beispiel für eine *Typ-I-Aufgabe* ist die *Random Generation* Aufgabe, bei der die Versuchspersonen eine bewusst zufällige Zahlenfolge produzieren soll (s. Baddeley, 1986). Als Maß für die Leistung wird der Grad der Wiederholung bestimmter Zahlenfolgen zugrunde gelegt. In der Regel werden die Zahlen artikuliert, was die Kombination mit einer weiteren Sprachproduktionsaufgabe ausschließt.

Die *Kettenadditionsaufgabe* ist ein Beispiel für eine *Typ-II-Aufgabe*. Hierbei handelt es sich um eine einfache Additionsaufgabe, bei der eine Zahl auf die Summe der bereits

Tabelle 3.1: Einteilung von Nebenaufgaben nach Rummer (1996)

	globale Messungen		lokale Messungen	
	kontinuierliche Belastung	diskontinuierliche Belastung	kontinuierliche Belastung	diskontinuierliche Belastung
zentral belastend	Typ I	Typ II	(Typ V)	Typ VI
peripher belastend	Typ III	Typ IV	Typ VII	Typ VIII

präsentierten Zahlen hinzu addiert werden soll. Das globale Maß ist die Summe aller Zahlen, die die Versuchsperson am Ende der Aufgabe angibt. Kontinuierlich belastend ist die Aufgabe besonders deshalb, weil es notwendig ist, das Zwischenergebnis im Arbeitsgedächtnis zu halten.

Aufgaben, die das Speichersystem systematisch durch das Memorieren von Ziffern belasten, sind Beispiele für *Typ-III-Aufgaben*. Die Versuchsperson merkt sich während der Ausführung der Primäraufgabe eine Anzahl von Ziffern und gibt sie nach Beendigung wieder. Die Erinnerungsleistung stellt das globale Maß für die Sekundäraufgabe dar.

Aufgaben des *Typs V* wären für die Kombination mit einer Sprachproduktionsaufgabe ideal. Dieser Typ kommt jedoch prinzipiell nicht vor. Rummer (1996) führt dies darauf zurück, dass lokale Entscheidungen für eine zentrale Belastung verantwortlich sind, die naturgemäß nicht kontinuierlich sein können.

Ein Beispiel für *Typ-VI-Aufgaben* stellt eine Serie von in unregelmäßigen Abständen dargebotenen Entscheidungsaufgaben dar, bei denen die Reaktionszeit gemessen wird (lokales Maß).

Tracking-Aufgaben sind als peripher und kontinuierlich belastend zu klassifizieren und erlauben lokale Messungen. Diese Art von Aufgaben ist demnach ein Beispiel für *Typ-VII-Aufgaben*. Die Versuchsperson muß dabei ein sich bewegendes Ziel manuell verfolgen. Dies erfolgt häufig mittels der Maussteuerung eines Rechners. Die Abweichungen der „Spur“ von der Vorgabe dient als lokales Maß.

Eine wesentliche Voraussetzung für die Effizienz von Doppelaufgabenexperimenten ist, dass die Verarbeitungskapazität einer Versuchsperson durch die Kombination der beiden Aufgaben zumindest erschöpft ist, da es ansonsten nicht zu Interferenzen kommt.

Für das Flughafenexperiment wäre gemäß dieser Übersicht eine Tracking-Aufgabe am besten geeignet. Durch die kontinuierliche Belastung sind am ehesten interpretierbare Leistungsrückgänge in der Sprachproduktion zu erwarten. Punktuelle Belastungen dagegen könnten dazu führen, dass diese auf die Hauptaufgabe keinen Einfluß haben, da sie innerhalb einer Phase auftreten, in der nicht artikuliert wird. Außerdem bergen sie die Gefahr, dass die Sprachproduktion völlig zusammenbricht, was sich negativ auf die Interpretierbarkeit auswirken würde. Ebenso wichtig ist die zentrale Belastung, da in diesem Experiment die Auswirkungen von Belastung auf die gesamte Sprachproduktion untersucht werden soll, und nicht etwa spezifische Teilprozesse. Lokale Messungen spielen insofern eine Rolle, als sichergestellt werden muß, dass die Sprecher zur Aufrechterhaltung der

Leistungen in der Sprachproduktionsaufgabe nicht die Nebenaufgabe vernachlässigen. Bei globalen Messungen ist dies schwierig, da keine Überwachung über den gesamten Bearbeitungszeitraum der Hauptaufgabe erfolgen kann.

Anforderungen an die Nebenaufgabe von Seiten des Szenarios Beim Entwurf des Flughafenexperiments wurden jedoch neben diesen experimentalpsychologischen Gesichtspunkten noch weitere Anforderungen an die Nebenaufgabe gestellt. Es wurde der Anspruch erhoben, eine Situation herzustellen möglichst realistisch ist². Neben einer entsprechenden Gestaltung der Hauptaufgabe (s. Abschnitt 3.3.2.3) sollte dies auch für die Nebenaufgabe gelten.

Aus diesem Grund wurde darauf verzichtet, eine Standard-Nebenaufgabe wie z.B. die Trackingaufgabe zu wählen. Vielmehr wurde eine Nebenaufgabe konstruiert, die eine Art von Belastung erzeugt, welche der tatsächlichen Belastung eines Reisenden am Flughafen möglichst nahen kommen sollte. Das Arbeitsgedächtnis eines Reisenden an einem Großflughafen wird belastet durch eine unüberschaubare Anzahl von Zeichen (s. Schmaucks, 1999) sowie durch die Daten, die der Reisende sich merken muß (Gate-Nummer, Flugnummer etc). Außerdem muß die Person vor der Abreise eine Reihe von Dingen erledigen. Zusammen mit der ungewohnten Situation führt dies dazu, dass sie unter erhöhter kognitiver Belastung steht. Diese beiden Aspekte, also Memorieren von Ziffernkombinationen und das Erledigen einer Aufgabe, sollte demnach durch die Nebenaufgabe nachempfunden werden.

Letzteres wurde durch eine Aufgabe realisiert, die im Folgenden „Navigationsaufgabe“ genannt und in Abschnitt 3.3.2.5 genauer beschrieben wird. Bei der Navigationsaufgabe steuert die Versuchsperson eine „Spielfigur“ (piktorielles Symbol) über einen Lageplan, der den Flughafen repräsentiert. Nacheinander leuchten „Stationen“ auf, d.h. Stellen auf dem Lageplan, zu denen die Spielfigur bewegt werden muß. Die Stationen repräsentieren Einrichtungen des Flughafens. Die Bewegung der Spielfigur erfolgt über die Pfeiltasten. Durch die Notwendigkeit, die Richtung der Bewegung zu bestimmen, sowie ständig auf Hindernisse zu achten, wird eine zentrale Belastung im Sinne der oben aufgeführten Klassifikation erreicht. Die Leistungen können dabei in geeigneter Granularität lokal gemessen werden. Das Maß sind dabei richtig bzw. falsch angelaufene Stationen sowie Kollisionen mit den Hindernissen (s. Abschnitt 3.3.2.5). Die nacheinander aufleuchtenden Stationen müssen darüberhinaus im Gedächtnis gehalten werden, was das Arbeitsgedächtnis zusätzlich belastet.

Um eine zusätzliche, kontinuierliche Belastung zu erreichen, wurde, wie von Rummer (1996) empfohlen, eine Memorier-Aufgabe hinzugefügt. Zu Anfang der Bearbeitung der Hauptaufgabe³ soll sich die Versuchsperson eine Kombination aus alphanumerischen Zeichen merken, die eine Gate-Nummer repräsentiert. Nach Beenden der Hauptaufgabe wird diese wiedergegeben (globales Maß).

Insgesamt stellt die hier aus den Gesichtspunkt der erzeugten Belastung beschriebene Nebenaufgabe einen Kompromiß zwischen den aus der Literatur bekannten Anforderungen und dem Anspruch der Realitätsnähe dar. Unter Beachtung der Voraussetzungen zum Erreichen einer geeigneten Art der Arbeitsgedächtnisbelastung, wurde mit der Navigati-

²Dabei sind die Einschränkungen zu berücksichtigen, die durch die Laborbedingungen gegeben sind

³Genaugenommen erstreckt sich die Memorier-Aufgabe über den Bearbeitungszeitraum von mehreren Iterationen der Hauptaufgabe. Siehe dazu Abschnitt 3.3.2.5

onsaufgabe eine im Hinblick auf das Flughafenszenario möglichst realistische Aufgabe entwickelt.

3.3.1.2 Die Kommunikationssituation als entscheidender Faktor bei der Untersuchung von Sprache

Nachdem im vorangehenden Abschnitt die Wahl der Nebenaufgabe diskutiert wurde, widmet sich dieser Abschnitt einem wichtigen Aspekt der Hauptaufgabe, nämlich der Kommunikationssituation. Da es sich bei der Hauptaufgabe um eine Sprachproduktionsaufgabe handelt, die unter experimentellen Bedingungen stattfinden soll, kommt der Kommunikationssituation eine entscheidende Rolle zu (s. Rummer, 1996). Es wird zunächst von einem Experiment beim schwedischen *Telia* Forschungsinstitut berichtet, bei dem unter ähnlichen Rahmenbedingungen die Auswirkungen verschiedener Kommunikationssituation auf die Sprache der Versuchspersonen untersucht wurde (s. Eklund, 1999). Anschließend wird diskutiert, welche Faktoren außerdem bei der Wahl der Kommunikationssituation eine Rolle gespielt haben, und wie diese mit den Ergebnissen von Eklund (1999) in Beziehung stehen.

Vergleich verschiedener Kommunikationssituationen Im Rahmen des *Spoken Language Translator* (SLT) Projektes beim schwedischen *Telia* Forschungsinstitut führte Eklund (1999) eine vergleichende Analyse von *Disfluencies*⁴ in vier Reisedialog-Korpora durch:

1. Mensch-„Maschine“-Mensch Dialoge aus sogenannten „Wizard of Oz“ Sitzungen, bei denen der Sprecher instruiert wird, wie zu einer Maschine zu reden. Die Verarbeitung erfolgte jedoch in Wirklichkeit durch einen Menschen (WOZ-1).
2. Mensch-„Maschine“ Dialoge. Ebenfalls nach dem „Wizard of Oz“ Prinzip (WOZ-2).
3. Mensch-Mensch Dialoge.
4. Mensch-Maschine Dialoge. Bei dieser Versuchsreihe wurde die Verarbeitung zumindest teilweise maschinell vorgenommen. Ein menschlicher Assistent wurde zur Verbesserung der Erkennensleistung eingesetzt.

Sowohl was die Domäne als auch was die verwendeten Aufgaben betrifft, waren die Experimente von Eklund (1999) mit unserem vergleichbar. Beim ersten Experiment (WOZ-1) wurden Stimuli eingesetzt, die eine Mischung von geschriebener und visueller Information darstellten (s. Abbildung 3.3).

Zusammen mit dem Bild wurde eine schriftliche Instruktion präsentiert:

Nach einer Woche in NY erfahren Sie, dass Ihr Lieblingsmusiker am 10. Mai ein Konzert in Boston gibt. Sie möchten am 10. Mai nach 13:00 Uhr abreisen. Sie möchten die Abflugzeiten wissen und ob es Zwischenlandungen gibt. Außerdem möchten Sie so billig reisen wie möglich. Rufen Sie an und buchen Sie !

⁴Wie bereits bei Oviatt (1997) wird hier der Begriff *Disfluencies* für die Symptome von Zeitdruck und kognitiver Belastung. Konkret wurden untersucht: gefüllte Pausen, stille Pausen, Wortfragmente, explizite Korrektursignale und Silbendehnungen.

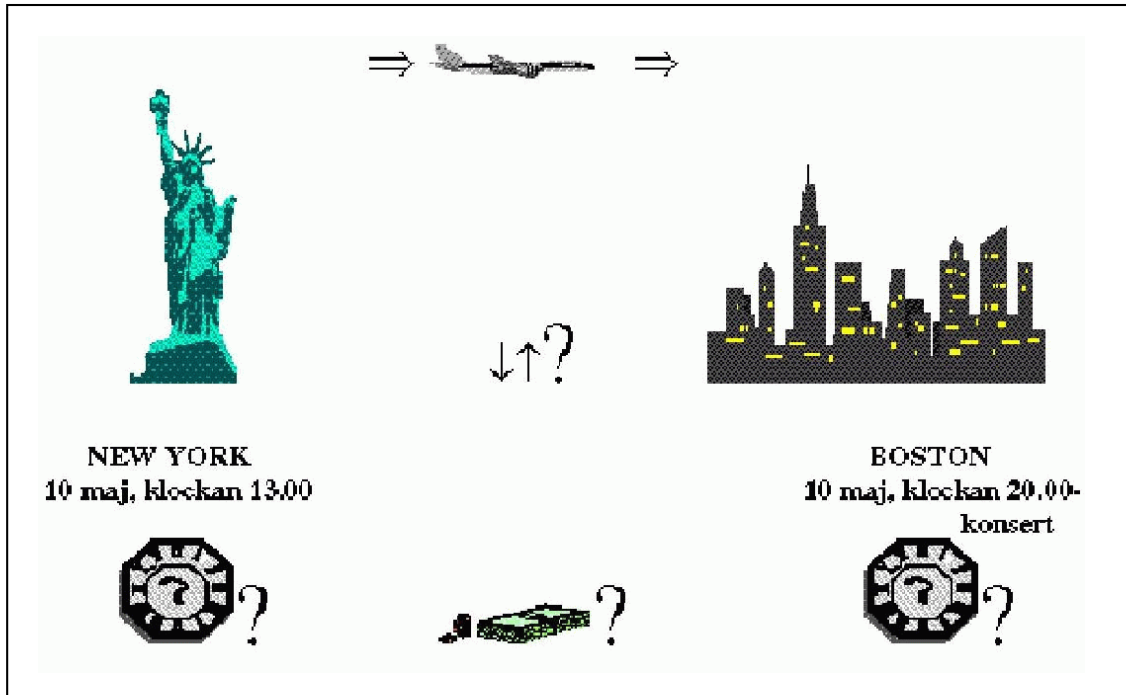


Abbildung 3.3: Experimentelle Stimuli bei Eklund (1999)

Die Versuchspersonen durften vor dem Anruf Notizen machen. Nach dem „Wizard of Oz“ Prinzip redeten sie via Telefon zu einem simulierten Auskunftssystem.

In einem zweiten Experiment wurden die Daten fast ausschließlich in piktorieller Form dargeboten. Diese Stimuli entsprachen annähernd denjenigen, die im Flughafenexperiment verwendet wurden. In Abbildung 3.4 wird auf der linken Seite ein Bild aus unserer Stimulusmenge dargestellt. Es entspricht einer Äußerung wie sie in Beispiel 3.1 dargestellt wird. Sie stammt aus dem Ergebnis-Korpus des Flughafenexperimentes. Das Bild rechts stammt aus Eklund (1999). Es könnte in etwa der Äußerung in Beispiel 3.2 entsprechen⁵.

(3.1) <04/170300/G-/2/004/3.55>

<EL:0.93> Was kostet ein Bahnticket von hier nach Los Angeles?

(3.2) Ist es billiger mit der Bahn zu reisen oder mit dem Flugzeug?

Bei diesem Experiment kommunizierten die Versuchspersonen mit einer simulierten Datenbank-Anwendung (WOZ-2). Der Mensch-Mensch Korpus wurde mit den gleichen Aufgaben erzeugt. Die Sprecher hatten dabei einen menschlichen Dialogpartner. In der Mensch-Maschine Situation wurden die Beiträge des Partners weitestgehend von einem Spracherkennungs- bzw. -produktionssystem übernommen. Die Erkennung wurde jedoch von einem menschlichen Assistenten unterstützt.

⁵Hier handelt es sich um eine Interpretation des Bildes, die nicht explizit bei Eklund (1999) genannt wird.

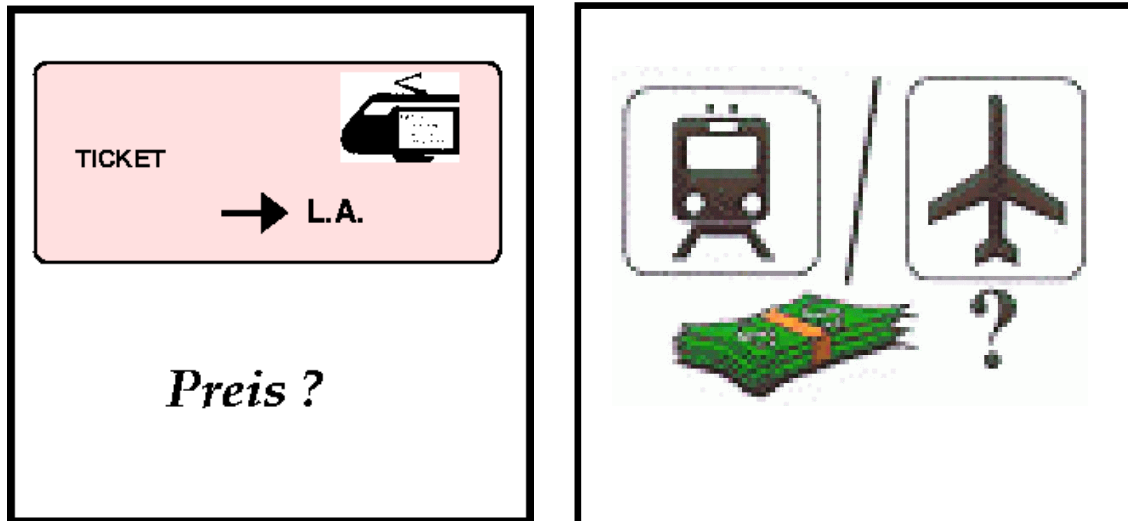


Abbildung 3.4: Links: „Was kostet ein Bahnticket nach Los Angeles?“ im Flughafenexperiment. Rechts: „Wieviel kostet ein Bahnticket im Vergleich zum Flug?“ nach Eklund (1999)

Eklund (1999) interessierte sich in erster Linie für gefüllte Pausen, stille Pausen, Wortfragmente, explizite Korrektursignale und Silbendehnungen. Die betrachteten Phänomene kamen in der Mensch-Maschine Bedingung am häufigsten vor. Eklund (1999) führt dies auf die Qualität des Feedbacks zurück, das im Fall der Mensch-Maschine Situation von einem Synthesizer übernommen wurde. Eine Ausnahme dieses Musters bildeten die Wortfragmente, die in der Mensch-Mensch-Bedingung häufiger auftraten, als in den anderen Bedingungen. Nach Eklund (1999) ist der Grund dafür möglicherweise in der Definition zu sehen: als Wortfragmente im Sinne von *Disfluencies* wurden auch diejenigen gezählt, die durch Satzabbrüche durch den Dialogpartner erfolgten. Eklund (1999) nimmt an, dass diese ausschließlich in der Mensch-Mensch Bedingung auftraten.

Auswahl einer geeigneten Kommunikationssituation Das Flughafenszenario sieht vor, dass der Reisende mit einem elektronischen Assistenten kommuniziert, um von diesem Informationen zu erhalten. Um eine anwendungsnahe Kommunikationssituation zu schaffen, sollte die Wahl auf eine Mensch-Maschine Situation fallen. Dies hätte jedoch einige Schwierigkeiten mit sich gebracht. Zum einen wäre die Entwicklung einer tatsächlichen Spracherkennung und -produktion mit erheblichem Aufwand verbunden gewesen. Auch bei Eklund (1999) konnte dies nur teilweise realisiert werden. Zum anderen war zu erwarten, dass die Reaktionen der Versuchspersonen auf eine solche Situation unterschiedlich ausfallen würden. Es ist anzunehmen, dass ein Teil der Versuchspersonen wie zu einem menschlichen Dialogpartner gesprochen hätten. Andere hätten jedoch vermutlich ihre Sprache angepasst, z.B. durch diskrete Aussprache der Wörter, oder durch Veränderung der Sprechgeschwindigkeit. Um diese Störvariable auszuschließen, entschieden wir uns für eine (simulierte) Mensch-Mensch Kommunikation. Nach den Ergebnissen von Eklund (1999) sollte dies dazu führen, dass die Anzahl der beob-

achtbaren Phänomene weniger hoch ausfällt. Es ist demnach zu erwarten, dass in einer realen Situation die Häufigkeit der Symptome größer ist. Wie die Kommunikationssituation hergestellt wurde, ist Abschnitt 3.3.2.9) zu entnehmen.

3.3.1.3 Untersuchung der Auswirkungen von Zeitdruck

Neben der kognitiven Belastung sollten in der hier beschriebenen Studie auch die Auswirkungen von Zeitdruck auf die Sprache untersucht werden. Dazu waren einige Vorüberlegungen notwendig, was das Erzeugen von Zeitdruck unter experimentellen Bedingungen betrifft. In diesem Abschnitt werden einige alternative Möglichkeiten aufgeführt, wie Zeitdruck in einem Experiment wie dem gegebenen erzeugt werden kann, und wie sich diese Methoden auf die Sprachproduktion auswirken können.

Eine Alternative ist die Erzeugung von Zeitdruck nach dem *Sanduhr-Prinzip*. Dabei wird eine Zeitspanne vorgegeben, in der eine Aufgabe abgeschlossen sein soll. Die noch zur Verfügung stehende Zeit kann dabei visuell oder akustisch dargestellt werden. Eine akustische Präsentation kam aufgrund der Interferenzen mit der Sprachproduktionsaufgabe von vornherein nicht in Frage. Die visuelle Präsentation wurde ebenfalls als problematisch angesehen, weil sie die Aufmerksamkeit des Sprechers erfordert hätte, und somit wie eine zusätzliche Aufgabe zu behandeln gewesen wäre, deren Interferenz mit Primär- und Sekundäraufgabe nicht vorhersagbar ist. Das Sanduhr-Prinzip bringt jedoch noch weitere Probleme mit sich. Erstens muß damit gerechnet werden, dass einige Aufgaben nicht rechtzeitig beendet werden können, und somit für die Auswertung verloren gehen. Zweitens ist der Zeitdruck nicht gleichmäßig über den Bearbeitungszeitraum verteilt, sondern wird mit Ablauf der Zeitspanne stärker.

Das *Stoppuhr-Prinzip* überwindet den zuletzt genannten Nachteil. Dabei wird ein nach oben offenes Zeitintervall vorgegeben. Zusammen mit der Anweisung, möglichst schnell fertig zu werden, entsteht der Zeitdruck durch die Wahrnehmung der fortlaufenden Zeit. Da keine feste Zeitspanne vorgegeben ist, kommt es weder zum Verlust von verwertbaren Beobachtungen, noch zum Ansteigen des Zeitdrucks. Die Abbildung der Zeit auf die Leistung der Versuchsperson ist für diese jedoch nicht direkt nachvollziehbar. Sie kann also nicht einschätzen, ob sie schnell oder langsam ist. Gibt man eine Schranke an, vor der die Aufgabe bearbeitet sein sollte, entsteht ein ähnlich ungleichmäßiger Zeitdruck wie beim Sanduhr-Prinzip. Außerdem bleibt die Ablenkung durch die notwendige Beobachtung der Uhr bestehen.

Im Allgemeinen gibt es Evidenzen, die gegen die Erzeugung eines restriktiven Zeitdrucks in einer experimentellen Untersuchung sprechen. Der Grund dafür sind nicht vorhersagbare Effekte, die auf die unterschiedliche Reaktion der Versuchspersonen zurückzuführen sind (s. Rummer, 1996). Eine Alternative ist die Erzeugung eines *weichen* Zeitdrucks. Das Attribut weich ist in diesem Zusammenhang als weniger restriktiv zu verstehen. Dies kann durch eine geeignete Instruktion der Versuchspersonen erreicht werden. Die Anweisung sollte so formuliert sein, dass die Versuchspersonen motiviert sind, in der Zeitdruckbedingung die gestellte Aufgabe in geringerer Zeit durchzuführen, als in der Kontrollbedingung. Diese Vorgehensweise wurde bei dem Design des Flughafenexperimentes gewählt. In Abschnitt 3.3.2.9 wird die Formulierung der entsprechenden Anweisung diskutiert.

3.3.2 Methode

3.3.2.1 Versuchspersonen

Zweiunddreißig Studenten der Universität des Saarlandes nahmen als Versuchspersonen an dem Experiment teil. Es handelte sich durchweg um deutsche Muttersprachler, die in verschiedenen Studiengängen eingeschrieben waren. Die Versuchspersonen wurden für ihre Teilnahme bezahlt. Um eine zusätzliche Motivation zu erreichen, wurde unter den fünf besten Versuchspersonen ein Geldpreis in Höhe der doppelten Aufwandsentschädigung verlost. Die Verteilung der Geschlechter betrug zwölf weibliche gegenüber zwanzig männlichen Teilnehmern.

Die Dauer einer Sitzung schwankte zwischen 60 und 75 Minuten. Die Unterschiede sind vor allem auf selbstgesteuerte Pausenlängen zwischen den einzelnen Aufgaben und auf Rückfragen zu den Anweisungen zurückzuführen.

3.3.2.2 Design

Die experimentellen Bedingungen wurden nach einem zwei mal zwei Design innerhalb der Versuchsperson manipuliert (s. Tabelle 3.2)

Tabelle 3.2: Design des Flughafenexperiments

	Qualitätsbedingung	Geschwindigkeitsbedingung
ohne Nebenaufgabe	Q–	G–
mit Nebenaufgabe	Q+	G+

In den Bedingungen mit Q+ und Q– wurden die Versuchspersonen angewiesen, auf die Qualität ihrer Äußerungen zu achten. In den Bedingungen G+ und G– dagegen sollten sie die Sprachproduktionsaufgabe möglichst schnell erledigen. Wie diese Bedingungen hergestellt wurden, geht aus Abschnitt 3.3.2.9 hervor. Die Eigenschaften der Nebenaufgabe werden in Abschnitt 3.3.2.5 beschrieben.

3.3.2.3 Stimuli

Die Anfragen, die bei der Feldstudie am Frankfurter Flughafen gesammelt wurden, dienten als Grundlage für die Entwicklung der experimentellen Stimuli (s. Abschnitt 3.2). Die Anzahl unterschiedlicher Anfragen war jedoch zu gering. Der Katalog wurde deshalb um fiktive, jedoch ebenfalls realistische Problemsituationen erweitert, so dass insgesamt 100 verschiedene „Bedürfnisse“ zur Verfügung standen.

Bilder Die Reisenden, die sich am Flughafen an einen Informationsschalter wandten und eine Anfrage artikulierten, taten dies aufgrund einer realen Bedürfnissituation, was unter experimentellen Bedingungen nicht der Fall ist. Es bestand deshalb die Problematik, aus dem Anfragenkatalog Stimuli zu erzeugen, die die Versuchspersonen dazu veranlassen sollten, eine Anfrage zu stellen. Dabei durfte die sprachliche Form nicht vorgegeben werden, da es sich ansonsten nicht um eine Sprachproduktionsaufgabe gehandelt hätte.

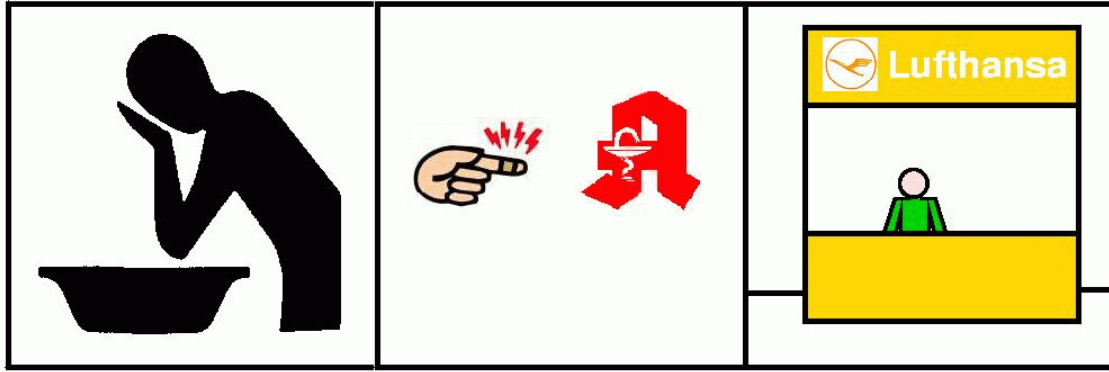


Abbildung 3.5: Experimentelle Stimuli

Wir wählten aus diesem Grund eine grafische Repräsentation. Unter Berücksichtigung der Arbeiten von Schmaucks (1999) über semiotische Aspekte der Fortbewegung an Großflughäfen, wurden 100 Bilder entworfen, die jeweils eine Problemsituation darstellten.

In Abbildung 3.5 werden drei Beispiele dargestellt. Von links nach rechts gesehen, entsprechen sie den Anfragen aus Beispiel 3.3 bis 3.5.

- (3.3) <30/020600/G-/5/002/3.85>
<EL:0.92> Ich möchte mich frisch machen. Wo finde ich eine Gelegenheit dazu?
- (3.4) <32/300500/Q-/4/004/4.98>
<EL:0.93> Ich habe mich verletzt. <IS:1.12> Wo ist die nächste Apotheke, wo ich Pflaster kaufen kann?
- (3.5) <10/140300/Q+/3/003/3.75>
<EL:1.17> Können Sie mir sagen wo der Schalter der Lufthansa ist ?

Bildergeschichte Die Gestaltung der Bilder basierte zum Teil auf Piktogrammen, die an verschiedenen Flughäfen dieser Welt zur Orientierung der Reisenden eingesetzt werden. Diese werden nach Gesichtspunkten der Interpretierbarkeit entwickelt werden (s. Modley, 1974). Vor Beginn des Experimentes wurden die Versuchspersonen mit den Bildern vertraut gemacht. Sie wurden im Rahmen einer Bildergeschichte präsentiert, um eine sprachliche Vorgabe der Äußerungen als Fragen zu vermeiden. Die Geschichte handelte von der fiktiven Familie „Kümmerling“, die von Frankfurt nach San Francisco reiste. Ihre Erlebnisse aus der Reise entsprachen der Bedeutung der Bilder. Zu jedem Bild wurde ein entsprechender Abschnitt auf der Geschichte über die Lautsprecher ausgegeben. Die Versuchspersonen konnten durch Drücken einer Taste selbst bestimmen, wann das nächste Bild erscheinen sollte. Da es sich bei den Texten um Problembeschreibungen handelte, die in einer deklarativen Form präsentiert wurden, wurde die sprachliche Form der späteren Äußerung nicht vorgegeben, da diese von der Versuchsperson als Frage formuliert werden sollten. Die Versuchsperson wurden vor dem Anhören der Geschichte nicht über ihre spätere Aufgabe informiert.

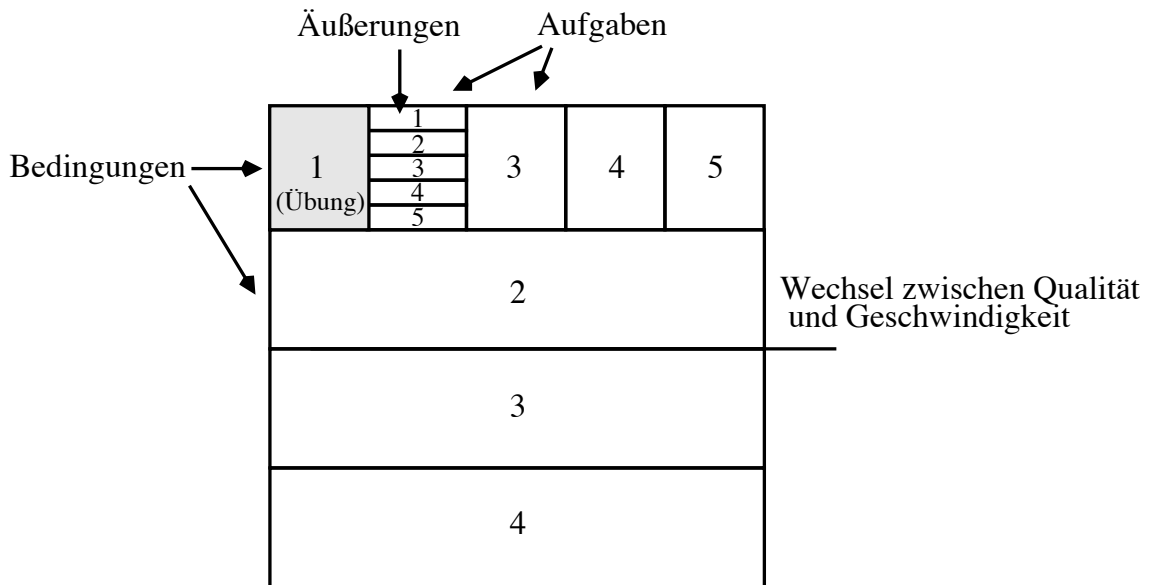


Abbildung 3.6: Organisation des Experimentes in Bedingungen, Aufgaben und Äußerungen

3.3.2.4 Reihenfolge und Aufbau der Bedingungen

Um Lerneffekte zu vermeiden, wurden die Bedingungen in rotierender Reihenfolge abgearbeitet. In Tabelle 3.3 werden die unterschiedlichen Varianten dargestellt. Die Bezeichnung der Bedingungen kann aus Tabelle 3.2 entnommen werden.

Tabelle 3.3: Reihenfolgevarianten nach dem Rotationsprinzip

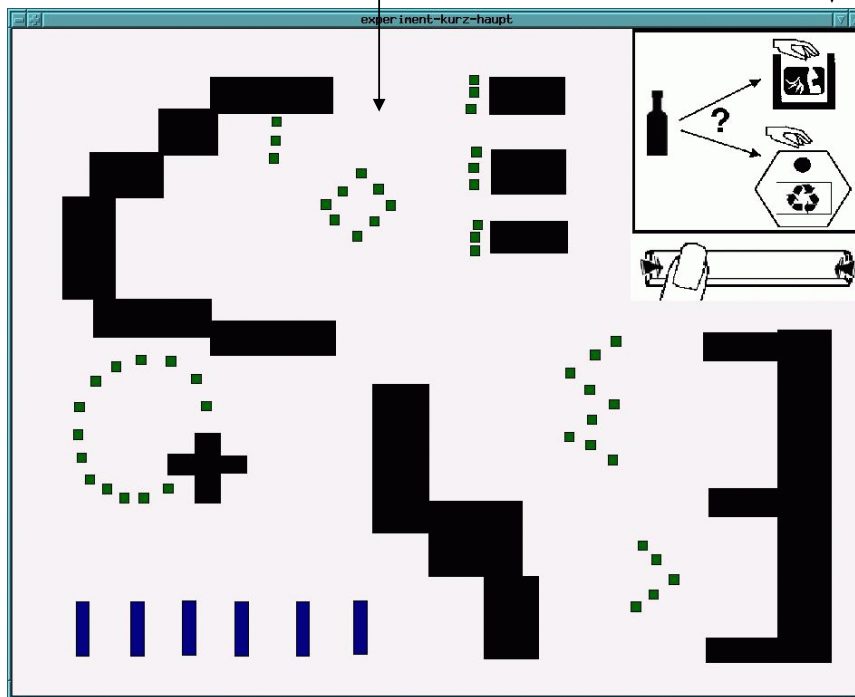
	Bed. 1	Bed. 2	Bed. 3	Bed. 4
Variante 1	Q-	Q+	G-	G+
Variante 2	Q+	Q-	G+	G-
Variante 3	G-	G+	Q-	Q+
Variante 4	G+	G-	Q+	Q-

Eine Alternation von Qualität und Geschwindigkeit wurde nicht vorgenommen, weil die Umgewöhnung an die jeweils andere Bedingung relativ schwierig ist. Während nämlich die Versuchspersonen beim Wechsel zwischen belasteter und unbelasteter Bedingung eine direkte Orientierungshilfe durch die anders gestaltete Oberfläche hatten (s. Abbildung 3.7 und 3.8), war dies bei dem Wechsel zwischen Qualitäts- und Geschwindigkeitsbetonung nicht der Fall. Alternierende Bedingungen hätten unter Umständen dazu geführt, dass den Versuchspersonen die aktuelle Anforderung weniger bewusst gewesen wäre (s. Abschnitt 3.3.2.9).

Abbildung 3.6 stellt die Organisation des Experimentes in Bedingungen, Aufgaben und Äußerungen dar. Eine Bedingung bestand aus fünf Aufgaben, die wiederum in fünf

Darstellung des Hintergrundes aus der Navigationsaufgabe mit weniger Details

Präsentation des Stimulus



Erinnerung an das Drücken der Leertaste nach dem Sprechen

Abbildung 3.7: Oberfläche des Experimentalsystems in den Bedingungen ohne Navigationsaufgabe

Äußerungen unterteilt waren. Das Bildmaterial wurde so zusammengestellt, dass innerhalb einer Aufgabe die typischen Ablaufschemata an einem Flughafen nicht verletzt wurden. Bilder, die etwas mit Erledigungen im Flughafen zu tun haben, wurden z.B. vor Bildern präsentiert, die zu der Abflugphase gehören. Die erste Aufgabe einer Bedingung war eine Übungsaufgabe, die auch als solche deklariert und später bei der Auswertung nicht berücksichtigt wurde. In einer Bedingung gab es demnach vier analysierbare Aufgaben, also insgesamt 20 Äußerungen.

3.3.2.5 Sekundäraufgabe

In Abbildung 3.8 wird die Oberfläche der Sekundäraufgabe dargestellt. Sie repräsentiert einen „virtuellen Flughafen“. Die schwarzen Flächen stellen Gebäudeteile mit verschiedenen Einrichtung (gelbe Quadrate) dar. Die Spielfigur wird durch das rote Quadrat repräsentiert. Sie markiert die aktuelle Position der Versuchspersonen im virtuellen Flughafen. Die kleinen dunklen Quadrate stellen Hindernisse dar, wie z.B. andere Personen oder statische Hindernisse.

Die Aufgabe der Versuchsperson war es, die Figur durch Drücken der Pfeiltasten durch den „Flughafen“ zu bewegen. Dabei sollte sie verschiedene Einrichtungen besuchen, die

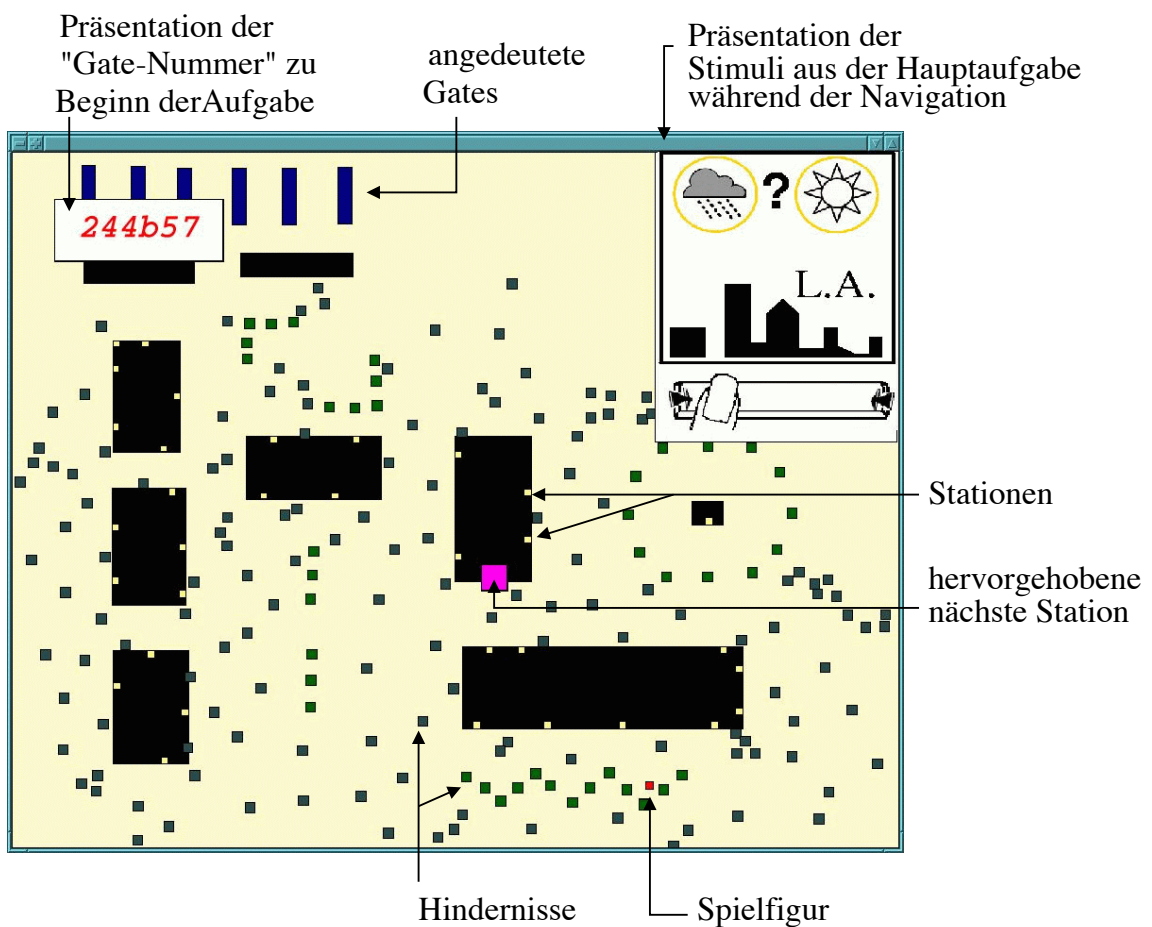


Abbildung 3.8: Oberfläche des Experimentalsystems mit Navigationsaufgabe

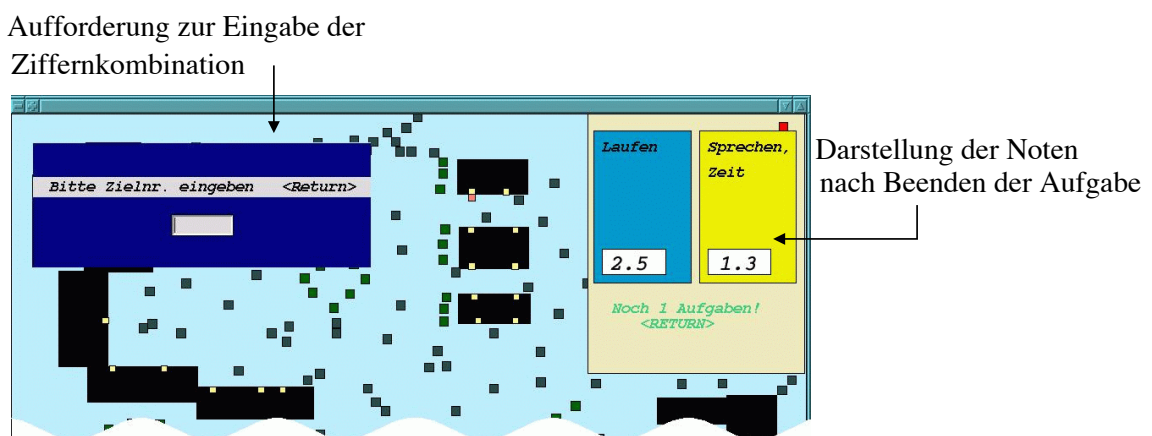


Abbildung 3.9: Abfrage der Zielnummer und anschließende Präsentation der Online-Bewertungen

im Folgenden *Stationen* genannt werden. Die Stationen wurden nacheinander markiert, d.h. wenn die Versuchsperson eine Station erreicht hatte, wurde diejenige hervorgehoben, die als nächstes angelaufen werden sollte. Auf dem Weg dorthin wurde zu einem nicht vorhersagbaren Zeitpunkt in der rechten oberen Ecke ein Stimulus (Bild) der Hauptaufgabe dargestellt. Die Navigation durch einen solchen Plan, wie er in Abbildung 3.8 dargestellt wird, entsprach der Einheit „Aufgabe“ (s. Abbildung 3.6). Dementsprechend wurden pro Plan fünf Bilder präsentiert.

Um zu erreichen, dass die Navigationsaufgabe kontinuierlich belastend ist (s. Abschnitt 3.3.1.1), musste sichergestellt werden, dass die Versuchsperson nicht stehenblieb, und die Aufgabe zu jedem Zeitpunkt aufmerksam bearbeiten musste.

Eine Reihe von Vorversuchen diente unter anderem dazu, die Eigenschaften der Nebenaufgabe zu verbessern. Aufgrund der dort ermittelten Daten wurde die Anzahl der Hindernisse auf das Maß erhöht, wie es in Abbildung 3.8 dargestellt ist. Dadurch wurde gewährleistet, dass die Versuchsperson die Navigation nicht durch einfaches Drücken der Pfeiltasten in eine gleichbleibende Richtung bearbeiten konnte. Jeder Kontakt mit einem Hindernis wurde akustisch signalisiert und führte zu Punktabzug in der Sekundäraufgabe, was den Versuchspersonen in der Anweisung deutlich gemacht wurde (s. Abschnitt 3.3.2.9). Für jeden Schritt in eine bestimmte Richtung musste die entsprechende Pfeiltaste einmal gedrückt werden. Ein Überwinden einer Strecke durch Gedrückthalten einer Taste war also nicht möglich. Blieb die Versuchsperson stehen, wurde dies ebenfalls durch einen deutlichen Punktabzug geahndet. In der Übungsphase zu Anfang wurde auf diesen Aspekt außerdem besonders geachtet (s. Abschnitt 3.3.2.6).

Zu Beginn der Nebenaufgabe wurde eine sechsstellige alphanumerische Ziffernkombination präsentiert, die eine Gate-Nummer darstellte. Die Versuchspersonen sollte sie sich merken und am Ende der Aufgabe wiedergegeben. Dies erfolgte mittels Eingabe der Kombination in ein Textfeld. Die Ziffern mussten in der richtigen Reihenfolge eingegeben werden. Jede richtig eingegebene Ziffer wurde gewertet.

3.3.2.6 Übungsphase

Nach der Einführung der Bilder mithilfe der Bildergeschichte, wurden Haupt- und Nebenaufgabe geübt. Zunächst übten die Versuchspersonen anhand von fünf Bildern das Stellen von Anfragen. Durch direkte Rückmeldung von Seiten der Versuchsleiterin konnten Mißverständnisse direkt beseitigt werden. Danach wurde die Sekundäraufgabe geübt, erst separat und dann in Kombination mit der Primäraufgabe. In dieser Phase wurde besonders darauf geachtet, dass die Versuchspersonen während der Artikulation die Bearbeitung der Navigationsaufgabe nicht unterbrach.

3.3.2.7 Hauptphase

Anschließend wurde der Hauptteil des Experiments gestartet. Entsprechend den Bedingungen bestand dieser aus vier Blöcken. Zu Beginn eines jeden Blockes wurde jeweils eine entsprechende Anweisung gegeben.

Die folgenden Schemata stellen den Ablauf der Aufgaben in einer belasteten bzw. unbelasteten Bedingung dar. Der Ablauf einer Aufgabe in einer unbelasteten Bedingung (ohne Sekundäraufgabe) war der Folgende:

1. Präsentation des ersten Bildes.

2. Artikulation einer entsprechenden Anfrage durch die Versuchsperson.
3. Betätigung der Leertaste durch die Versuchsperson als Zeichen, dass die Äußerung beendet ist.
4. Pause von drei Sekunden.
5. Von der Versuchsperson bestimmte zusätzliche Pause, die durch das Drücken der Leertaste beendet wurde.
6. Präsentation des nächsten Bildes.

Der Ablauf der Kombination der Aufgaben in der Belastungsbedingung war folgendermaßen:

1. Anzeige der Zielbezeichnung, die memoriert werden sollte.
2. Hervorheben der aktuellen Position der zu bewegenden Figur.
3. Aufleuchten des aktuellen Ziels.
4. Start der Navigation von Seiten der Versuchsperson, um die aktuelle Station zu erreichen.
5. Präsentation des Stimulus der Primäraufgabe.
6. Artikulation einer entsprechenden Anfrage durch die Versuchsperson.
7. Betätigung der Leertaste als Zeichen, dass die Äußerung abgeschlossen ist.
8. Erreichen der aktuellen Station. Dies konnte auch die falsche Station sein, wenn die Versuchsperson sich irrte.
9. Aufleuchten der nächsten Station.
10. Nach fünf Stationen: Aufforderung zur Wiedergabe der memorierten Zielnummer.

3.3.2.8 Rückmeldung der Leistungen für die Versuchspersonen

Zur Orientierung und zusätzlichen Motivation der Versuchspersonen, wurden deren Leistungen in Form von *Online Bewertungen* nach Ablauf einer Aufgabe angezeigt. Es wurden zwei verschiedene Noten vergeben: eine Note für die Leistungen in der Hauptaufgabe und (in den Belastungsbedingungen) eine Note für die Leistungen in der Nebenaufgabe.

Die Noten für die Sprachproduktion basierten in der Geschwindigkeitsbedingung auf der verbrauchten Zeit zwischen Präsentation des Bildes und Beenden der Äußerung. In der Qualitätsbedingung basierte sie auf der subjektiven Einschätzung der Versuchsleiterin. Diese wurde über die Maustasten eingegeben.

Die Note für die Nebenaufgabe setzte sich aus verschiedenen Faktoren zusammen. Es gab Punkte für korrekt angelaufene Stationen und richtig eingegebene Ziffern der Zielnummer. Punktabzug gab es für Kollision mit einem Hindernis und Stehenbleiben.

3.3.2.9 Anweisung

Obwohl dies aus Literaturanalysen bereits bekannt war (s. Rummer, 1996), hat es sich spätestens bei der Durchführung der Vorversuche herausgestellt, dass die Anweisung einen wichtigen Einfluß auf Interpretierbarkeit der Beobachtungen hat. Das hat dazu geführt, dass neben der Anpassung des Experimentalsystems, vor allen Dingen die Anweisung intensiv überarbeitet wurde.

Ein Grund für die Bedeutung der Anweisung war z.B. die Tatsache, dass die Unterscheidung zwischen Qualitäts- und Geschwindigkeitsbedingung ausschließlich auf ihr beruhte. Das ist eine Folge der Vorüberlegungen bezüglich der Kommunikationssituation und der Herstellung von Zeitdruck. Desweiteren galt es die relativ komplexen Abläufe von Haupt- und Nebenaufgabe zu erläutern. Nachfolgend werden die relevanten Teile der Anweisung dargestellt. Passagen in eckigen Klammern stellen Kommentare dar, die Zeichenfolge [...] bedeutet eine Auslassung. Die Abschnitte, die fett gedruckt und mit einer hochgestellten Zahl versehen wurden, werden im darauffolgenden Abschnitt diskutiert.

[zur Bildergeschichte]

[...*Familie Kümmerling macht eine Flugreise*] Während sie versuchen, sich in dem riesigen Flughafenkomplex zurecht zu finden, werden die Kümmerlings immer wieder mit kleineren und größeren Problemen konfrontiert. Die Bilder, die Du siehst, während Du die Geschichte hörst, stellen das jeweilige Problem dar. [...]

[Zu Beginn der Übungsphase]

[...] Im Unterschied zu den Kümmerlings profitierst Du von einem neuen Service: **Du hast eine Art Sprechfunkgerät dabei, über das Du mit einer Informationszentrale in Verbindung stehst¹**. Wann immer Du auf ein Problem stößt, kannst Du Dich an das freundliche Personal wenden, das über einen Computermonitor Deine genaue Position innerhalb des Flughafens feststellen kann.

Damit das Servicepersonal Deine Anfrage beantworten kann, musst Du folgendes beachten:

- leite die Frage ein, indem Du die **Situation beschreibst²**, in der Du Dich befindest.
- stelle dann Deine Frage.

Bsp: „Ich habe noch etwas Zeit und möchte die Stadt besichtigen. Können Sie mir sagen wie lange ich bis da hin brauche?“

[...] Auf Deinem Weg durch einen virtuellen Flughafen, wirst Du mit Problemen konfrontiert, die durch die Bilder dargestellt werden, wie Du sie aus der Geschichte mit den Kümmerlings kennst. Das Mikrophon vor Dir ersetzt das Sprechfunkgerät und Deine Anfragen werden aufgezeichnet.

[Übungsphase 1: nur Artikulieren]

[...] Du wirst mit einigen Problemen konfrontiert, die durch Bilder dargestellt werden. [...] **Wenn Du das Bild siehst, beschreibe das Problem, das Dir dabei als erstes einfällt. Es kommt nicht darauf an, dass es sich um die gleiche Situation handelt, in der auch die Kümmerlings waren.³** [...]

[Übungsphase 2: nur Navigieren]

[...] [während die Versuchspersonen diesen Teil der Anweisung hörten, wurde der Ablauf der Nebenaufgabe am Bildschirm automatisch abgespielt. Die Versuchsführerin deutete auf die relevanten Objekte. s. Abbildung 3.8.] Die Ziele auf unserem Lauf

durch den virtuellen Flughafen kannst Du Dir entsprechend als einen Block von Gates bzw. Gepäckbändern vorstellen. Bevor Du jedoch dort hingehen kannst, musst Du einige Dinge erledigen. Das heisst, Du musst einige Stationen anlaufen.

Du siehst jetzt den virtuellen Flughafen. Das hier [Zeigegeste auf die schwarzen Flächen] sind Gebäudeteile, in denen sich z.B. Check-In Schalter und verschiedene Läden usw. befinden. [...] Jetzt wird Dir die Nummer des Ziels angezeigt, die Du Dir merken musst. Du musst sie später wieder eingeben. [...] Diese [Zeigegeste] kleinen gelben Quadrate sind die Stationen, von denen Du einige anlaufen musst. Die Station, die Du als nächstes anlaufen sollst, leuchtet pink auf [Zeigegeste]. Merke Dir, welche aufgeleuchtet hat.

Nun erscheint die dunkelrote Spielfigur auf der Bildfläche [Zeigegeste]. Die Figur kannst Du mit den Pfeiltasten bewegen. Für jeden Schritt musst Du die entsprechende Pfeiltaste einmal drücken. Das hier sind andere Personen, gegen die Du nicht laufen solltest. Tust Du es doch, beschweren Sie sich und Dir werden Punkte abgezogen und es kostet Dich Zeit. Gegen die Mauern darfst Du auch nicht laufen. [...]

Immer wenn Du eine Station erreicht hast, wird Dir die nächste angezeigt. [...] laufe zur nächsten Station. Wenn sich der Plan blau färbt, hast Du alle Stationen angelaufen, und Du kannst Dich zum Ziel begeben. Dazu musst Du zunächst die richtige Nummer eingeben. Wir bewerten jede richtig eingegebene Stelle. [...] Es kommt [...] darauf an, dass Du die richtigen Stationen erreichst und am Schluss die richtigen Zielnummern eingibst. Für diese Leistungen werden Noten vergeben. [...]

[Übungsphase 3: Kombination von Navigieren und Artikulieren]

[...] Wir üben jetzt die Bedienung des Auskunft-Systems auf dem Weg durch unseren virtuellen Flughafen. Während Du dich auf dem Weg zu einer Station befindest, wird in der rechten oberen Ecke [Zeigegeste] jeweils ein Bild angezeigt, das Dein Problem darstellt.

Wenn ein Bild aufleuchtet, stelle so schnell wie möglich eine passende Anfrage. **Gehe von dem Problem aus, das Dir beim Anblick des Bildes als erstes in den Sinn kommt³**. Denke daran, dass Du zuerst beschreiben sollst, in welcher Situation Du Dich befindest. [...] Vergesse dabei das Laufen nicht. **Wenn Du stehenbleibst, unterbricht die Verbindung zum Hilfesystem und das Bild verschwindet. [...] für jede nicht vollständig gestellte Anfrage, werden Dir Punkte abgezogen⁴**. Sowohl das Laufen, als auch das Stellen von Anfragen werden nun bewertet. [...] Am Ende der Aufgabe werden Dir Deine Leistungen in dieser Aufgabe angezeigt. [...]

[Übergang von der Übungsphase zur Hauptphase]

[...] Du sollst nun vier verschiedene Blöcke von Aufgaben bearbeiten. In zwei der Blöcke musst Du nur Anfragen stellen. In den anderen beiden Blöcken sollst Du sowohl Anfragen stellen als auch laufen. [...] Jeder Block besteht aus insgesamt 5 Aufgaben, von denen die erste Aufgabe eine Übungsaufgabe ist, die nicht in die Endbewertung eingeht.

Unter den besten 5 Versuchspersonen werden wir einen Geldpreis in Höhe von 50 DM verlosen⁵. Um eine gute Gesamtnote zu erhalten, wirst Du beide Aspekte des Experiments (das Sprechen und das Laufen) gut bearbeiten müssen. Zur Kontrolle für Dich werden während des Experimentes Noten angezeigt. *Wenn Du einen Aspekt des Experimentes vernachlässigst, werde ich [die Versuchsleiterin] Dich darauf aufmerksam machen.⁶* [...]

[Hauptphase Geschwindigkeit ohne Nebenaufgabe]

Du musst jetzt Anfragen stellen ohne zu Laufen. **In der Servicezentrale hat Frau Altmeier Dienst. Sie hat sehr viel Erfahrung im Bearbeiten von Anfragen. Leider ist sie auch sehr gefragt und hat deshalb wenig Zeit. Versuche also, mit Deiner Anfrage so bald wie möglich fertig zu werden und sofort auf die Leertaste zu drücken⁷.** Die Zeit, die Du für Deine Anfrage benötigst, wird gemessen und bildet die Grundlage für Deine Sprachnote.

[Hauptphase Geschwindigkeit mit Nebenaufgabe]

Du musst jetzt Anfragen stellen und dabei Laufen. Beim Laufen ist es wichtig, dass Du nicht stehenbleibst. Frau Altmeier [...*sinngemäß wie 7*] Das Laufen und das Stellen von Anfragen sind gleich wichtig, und **Deine Leistungen in diesen beiden Aufgaben gehen in die Bewertung ein. Du sollst Dich also nicht nur auf eine Aufgabe konzentrieren, sondern beide Aufgaben mit gleicher Aufmerksamkeit bearbeiten.**⁸

[Hauptphase Qualität ohne Nebenaufgabe]

Du musst jetzt Anfragen stellen ohne zu Laufen. **In der Servicezentrale hat Frau Neumann Dienst, die dort erst seit kurzem arbeitet. Komme ihr entgegen, indem Du Deine Anfragen eindeutig und in gutem Deutsch formulierst. Das ist jetzt auch die Grundlage für Deine Sprachnote.**⁹ [...]

[Hauptphase Qualität mit Nebenaufgabe]

Du musst jetzt Anfragen stellen und dabei Laufen. Beim Laufen ist es wichtig, dass Du nicht stehenbleibst. Frau Neumann [...*sinngemäß wie 9*]

Das Laufen und das Stellen von Anfragen sind gleich wichtig [...*wie 8*]

Im folgenden Abschnitt wird auf diejenigen Aspekte der Anweisung näher eingegangen, die mit hochgestellten Ziffern markiert sind.

Zu 1 Wie bereits in Abschnitt 3.3.1.2 diskutiert, wurde eine Mensch-Mensch-Kommunikationssituation gewählt. Auf diese Weise wurde sie hergestellt.

Zu 2 Die Situationsbeschreibung wurde deshalb eingeführt, weil in den Vorversuchen die Äußerungen durchweg sehr kurz waren. Die zu geringe Anzahl von Symptomen führte zu Bodeneffekten. Außerdem sollten stereotype Äußerungen vermieden werden. Ohne die Anweisung, die Situationen zu beschreiben, konnte zu fast jedem Bild eine Äußerung wie „Wo kann ich XY finden?“ produziert werden. In Abschnitt 2.2 wird auf die Bedeutung der Äußerungslänge bzw. -komplexität eingegangen.

Zu 3 Um zu vermeiden, dass die Versuchspersonen sich an die Bildergeschichte zu erinnern versuchen, damit sie eine exakte Übereinstimmung erreichen, wurde deutlich gemacht, dass sie selbständig formulieren sollen.

Zu 4 Diese besonders deutliche Betonung des Nicht-Stehenbleibens wurde ebenfalls aufgrund der Erfahrungen aus den Vorversuchen eingeführt. So konnten die Unterbrechungen der Nebenaufgabe reduziert werden.

Zu 5 Der Geldpreis als leistungsabhängige Belohnung wurde vor allem deshalb eingeführt, weil viele Aspekte des Experiments davon abhängig waren, dass die Versuchspersonen sich an die Anweisung hielten. Damit sich jeder eine realistische Chance ausrechnen konnte, wurde der Preis unter den besten fünf Versuchspersonen verlost.

Zu 6 Nach jedem Übungsblock hielt die Versuchsleiterin Rücksprache mit der Versuchsperson, um sie auf eine eventuelle Vernachlässigung der Haupt- bzw. Nebenaufgabe aufmerksam zu machen. Nach der Präsentation der Anweisungsteile wurde darüberhinaus das Verständnis durch gezielte Rückfragen überprüft.

Zu 7 Die Einführung der fiktiven Personen Frau Altmeier und Frau Neumann betrifft in erster Linie die Unterscheidung von Qualitäts- und Geschwindigkeitsbedingung. Frau Altmeier wurde als eine Person mit viel Erfahrung und wenig Zeit beschrieben. Die Anpassung der Sprache an den Dialogpartner ist für den Sprecher eine bekannte Situation. Die Beschreibung der Eigenschaften von Frau Altmeier diente somit dazu, Zeitdruck zu erzeugen (s. Abschnitt 3.3.1.3). Die explizite Nennung von Namen diente dazu, dass sich die Versuchspersonen das Muster von Eigenschaften besser merken konnten. Einige Versuchspersonen sprachen ihren „Dialogpartner“ in ihren Äußerungen mit dem Namen an. Auf diesen Aspekt wird in Abschnitt 3.4 näher eingegangen.

Zu 8 Das Doppelaufgabenparadigma sieht vor, dass beide Aufgaben mit gleich hoher Aufmerksamkeit bearbeitet werden, damit gezielt eine Überlastung hergestellt werden kann. (s. Abschnitt 3.3.1.1). Deshalb wurde auch dieser Aspekt in der Anweisung besonders hervorgehoben. Bei einer großen Diskrepanz zwischen den Noten für die Haupt- bzw. Nebenaufgabe in der Übungsphase, wurden die Versuchspersonen darüberhinaus von der Versuchsleiterin ermahnt.

Zu 9 Die Eigenschaften von Frau Neumann als Partnerin dienten dazu, die Betonung auf Qualität zu erreichen, ohne konkret die Art der Anpassung nennen zu müssen. Statt z.B. zu sagen „Achte darauf, möglichst keine Wortfragmente zu erzeugen“, wurde Frau Neumann als Person dargestellt, die unerfahren ist, und deshalb durch Äußerungen guter Qualität unterstützt werden sollte.

3.3.3 Einteilung und Definition der Variablen

Eine grobe Einteilung der abhängigen Variablen betrifft die drei Kategorien Zeitsymptome, Qualitätssymptome und Längensymptome. Die Kategorien Zeitsymptome und Qualitätssymptome entsprechen denjenigen, die von Berthold (1998) vorgeschlagen wurden (s. Tabelle 2.1). Die dritte Kategorie Längensymptome subsumiert alle Variablen, die direkt mit der Länge einer Äußerung in Beziehung stehen: Gesamtzeit, Anzahl der Wörter, Anzahl der Silben und Anzahl der Sätze. Da das Vorkommen vieler Symptome stark von der Äußerungslänge abhängt, erfolgte für die Interpretation eine entsprechende Normierung. Dazu kamen zwei Variablen aus der Kategorie der Längensymptome in Frage: die Anzahl der Silben und die Anzahl der Wörtern. Wir entschieden uns für eine Normierung

auf 100 Wörter, wie sie bei Oviatt (1997) vorgenommen wird ⁶. Bei anderen Variablen erfolgte z.T. eine andere Normierung. In Tabelle 3.4 wird eine Übersicht dargestellt.

3.3.3.1 Apparatur

In Abbildung 3.10 wird die Apparatur dargestellt, die beim Experiment (online) und bei der Auswertung der Daten (offline) verwendet wurde. Das Experimentalsystem lief auf einem Intel™ PENTIUM™ PC mit 200 MHz unter dem Betriebssystem Linux. Über die integrierte Soundkarte wurde die Geschichte auf Lautsprecher ausgegeben, die zuvor per Mikrophon aufgezeichnet und digitalisiert wurde. Die Anweisung wurde auf eine CD aufgenommen und während des Experimentes mithilfe eines CD-Players abgespielt. Die Reihenfolge der Anweisungsteile konnte so je nach Abfolge der Bedingungen gewählt werden. Die Signale des Experimentalsystems wurden über den internen Lautsprecherausgang ausgegeben. Damit sie als akustisches Signal auf das Tonband aufgenommen werden konnten, mussten sie zuvor über einen kleinen Lautsprecher ausgegeben und mit einem Mikrophon wieder aufgenommen werden. Dazu wurde eine Vorrichtung hergestellt, die „Black Box“, in deren Innern Lautsprecher und Mikrophon untergebracht wurden. Sie wurde isoliert, so dass die Signale nicht nach außen drangen und somit für die Versuchspersonen nicht hörbar waren. Die Sprache der Versuchspersonen wurde über ein Tischmikrophon aufgezeichnet. Signale und Sprache belegten jeweils einen Kanal des Stereoeingangs des Tapedecks. Sie wurden getrennt angesteuert, um in geeigneter Intensität auf das Band aufgezeichnet werden zu können.

Nach Ablauf des Experimentes wurden die Bänder mit dem gleichen Tapedeck abgespielt, verstärkt und über die Line-in Schnittstelle eines Intel™ PENTIUM™ kompatiblen PCs mit 400 MHz digitalisiert. Die entstandenen Wavedateien wurden mithilfe eines Wave-Editors weiterverarbeitet und anschließend im gleichen Format auf CD archiviert.

Abbildung 3.11 verdeutlicht, wie die weitere Verarbeitung vorgenommen wurde. Die separaten Spuren werden übereinander liegend dargestellt. Die obere (linke) Spur enthält die Signale des Experimentalsystems. Das erste Signal repräsentiert den Zeitpunkt, an dem der Stimulus der Hauptaufgabe präsentiert wurde. Die Zeit zwischen diesem Signal und dem Beginn der Äußerung bildet die Einsatzlatenz, die Zeit bis zum Ende der Äußerung die Gesamtdauer. Diese beiden Werte wurden als abhängige Variablen verwendet (s. Tabelle 3.4). Das zweite Signal auf der linken Spur repräsentiert das Beenden der Stimulus-Präsentation. Dies erfolgte entweder, wenn die Versuchspersonen die Leertaste drückte, oder wenn die Präsentation abgebrochen wurden, z.B. wegen Stehenbleibens der Spielfigur in der Nebenaufgabe (s. Abschnitt 3.3.2.5). Durch die Trennung der Signale auf zwei Spuren wurde gewährleistet, dass die Sprache ohne störende Signale von Experimentalsystem abgespielt werden konnte.

3.3.3.2 Bewertung der inhaltlichen Qualität

Um neben den Symptomen, die an der Oberfläche der Äußerungen zu beobachten sind, auch ein objektives Maß der inhaltlichen Qualität zu bekommen, wurden entsprechende Bewertungen eingeholt. Vier Muttersprachlern wurden die transkribierten Äußerungen zur Bewertung vorgelegt. Sie wurden für diese Aufgabe bezahlt. Für jedes Bild aus dem

⁶Es bestätigte sich ohnehin bei den Datenanalysen, dass die beiden Maße Anzahl der Silben und Anzahl der Wörter stark korrelieren (s. Abschnitt 3.4.6)

Tabelle 3.4: Bedeutung der Variablen und Aufteilung in die drei Hauptkategorien

Zeitsymptome	
Anzahl der gefüllten Pausen	normiert auf 100 Wörter
Dauer der gefüllten Pausen	“ “
Anzahl der stillen Pausen	“ “
Dauer der stillen Pausen	“ “
Einsatzlatenz	Intervall zwischen Beginn der Präsentation des Bildes und Einsetzen der Artikulation in ms
Zögerungen	Stille Pausen unter 200 ms, normiert auf 100 Wörter
Silbendehnungen	normiert auf 100 Wörter
Wiederholungen	“ “
Begrüßungen	Vorkommen einer Begrüßungsphrase
Anreden	Vorkommen einer direkten Anrede
Artikulationsgeschwindigkeit	
Artikulationsgeschwindigkeit	Anzahl der Silben pro Sekunde
Qualitätssymptome	
Fehlansätze	normiert auf Anzahl der Sätze
Satzabbrüche	“ “
inhaltliche Selbstkorrekturen	normiert auf 100 Wörter
syntaktische Selbstkorrekturen	
Wortabbrüche	
Syntaktische Fehler	
Längensymptome	
Gesamtdauer	verbrauchte Zeit für die gesamte Äußerung in ms (s. Abbildung 3.11)
Anzahl der Silben	
Anzahl der Wörter	
Anzahl der Sätze	

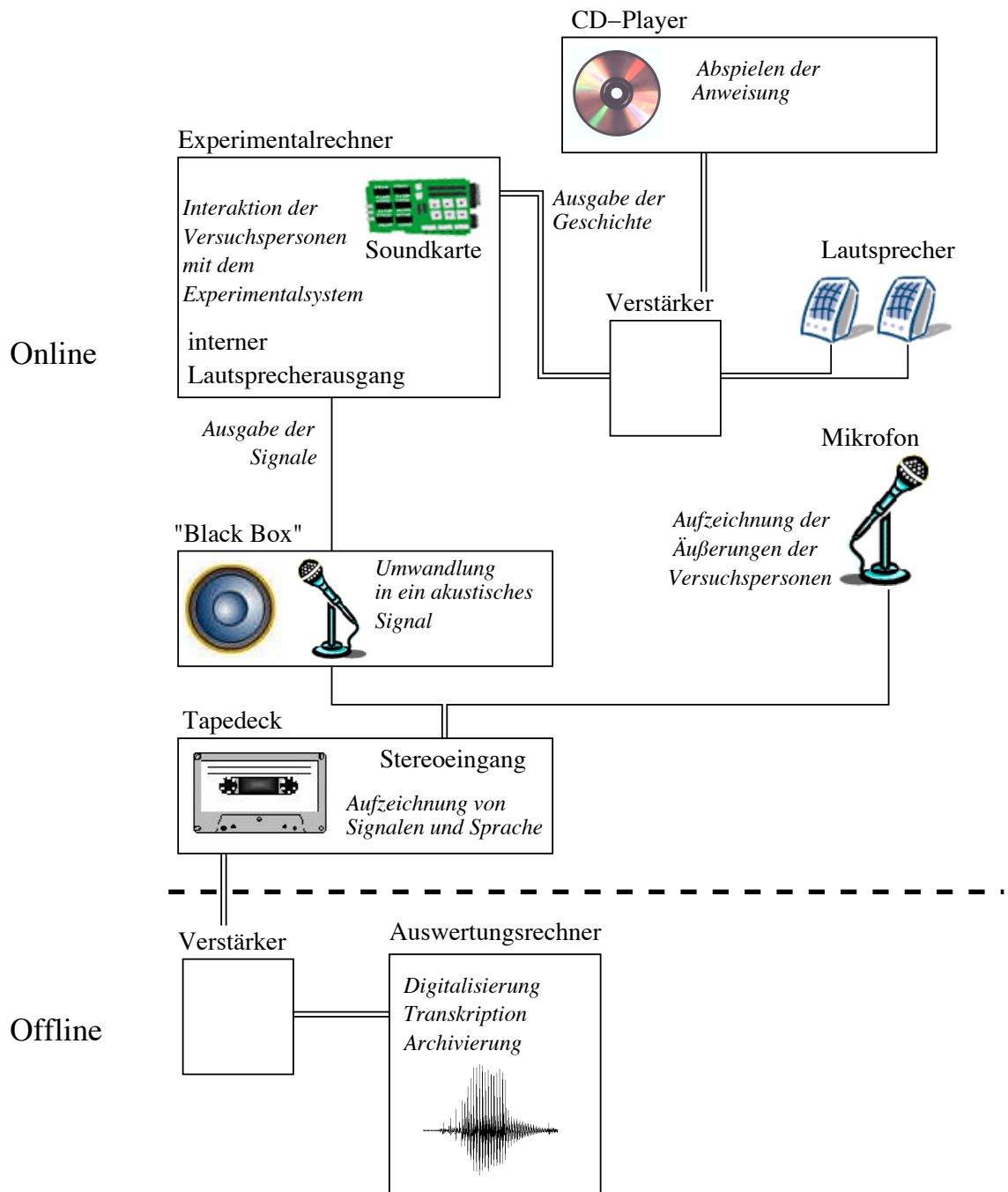


Abbildung 3.10: Aufzeichnung von Sprache und Synchronisation mit Signalen vom Experimentalsystem

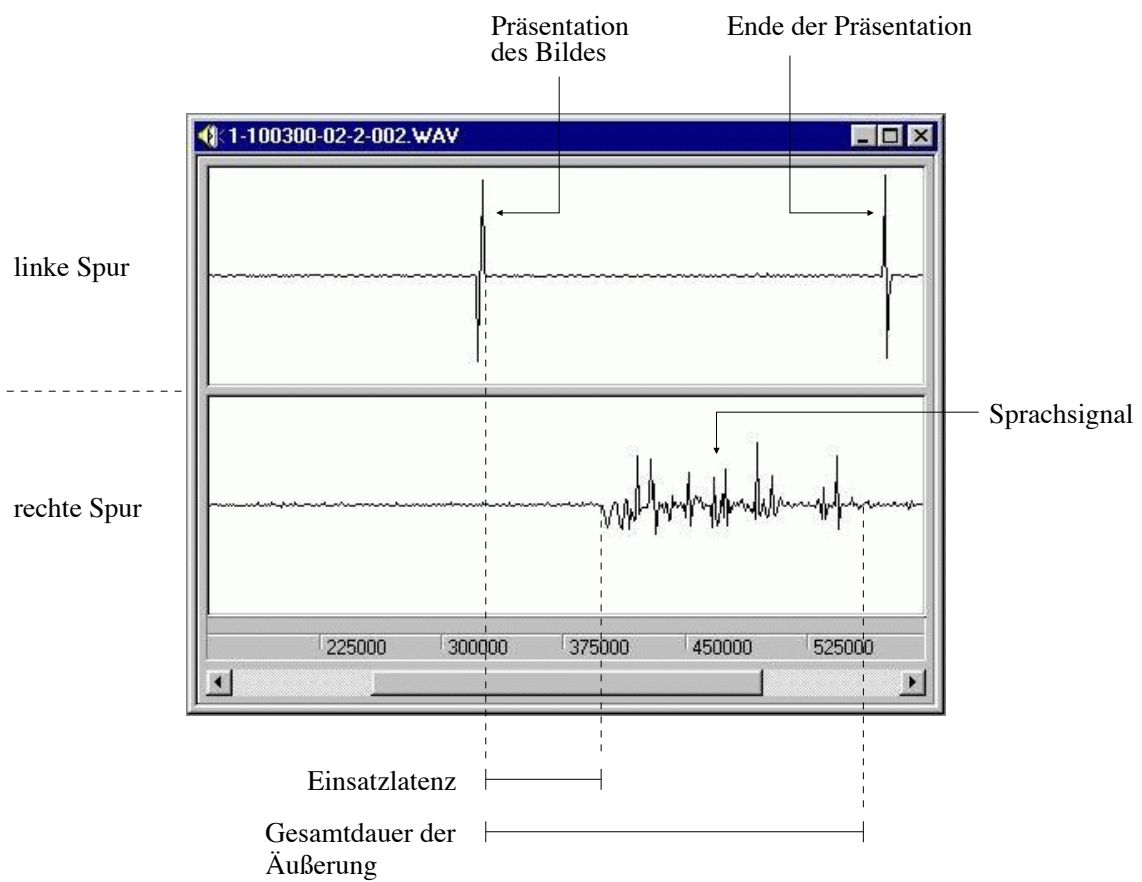


Abbildung 3.11: Verarbeitung der digitalisierten Äußerungen

Hauptteil des Experimentes lagen 32 Äußerungen der Versuchspersonen vor. Gemäß dem Rotationsprinzip, das beim Experiment angewendet wurde, waren diese gleichmäßig auf die vier Bedingungen verteilt. Die Bewerter wurden gebeten, für jedes der insgesamt 80 Bilder die Reihenfolge der inhaltlichen Qualität der entsprechenden Äußerungen zu bestimmen.

3.3.3.3 Bewertung der Bildkomplexität

Um ein objektives Maß für die Komplexität der Sprachaufgabe zu erhalten, wurden die verwendeten Bilder entsprechend kategorisiert. Dieselben Personen, die die Bewertung der inhaltlichen Qualität vorgenommen haben, ordneten die Bilder in eine der folgenden fünf Kategorien ein:

1. Sehr hohe Komplexität
2. Eher hohe Komplexität
3. Mittlere Komplexität
4. Eher niedrige Komplexität
5. Sehr niedrige Komplexität

3.3.4 Spezifische Vorhersagen

In Abschnitt 2.1 wurden einige theoretische Überlegungen über die Eigenschaften der Sprache unter Belastung (und Zeitdruck) gemacht. Diese basieren auf den Ergebnissen verschiedener Literaturstudien. Im folgenden Abschnitt werden die qualitativen Hypothesen beschrieben, die daraus abgeleitet werden können.

Die formale Qualität sollte insgesamt unter Belastung leiden. Unklar ist der Effekt von Zeitdruck: Einerseits sollte die Qualitätsaufforderung zu besserer formaler Qualität führen, andererseits war die Anweisung vor allem auf inhaltliche Qualität gerichtet. Diese kann teilweise auf Kosten von formaler Qualität erkaufte werden (z.B. wenn die Versuchsperson einen Satz abbricht, um eine bessere Formulierung zu finden). Was die Satzfragmente und Fehlsätzen angeht, erwarten wir ein Ansteigen unter Belastung in beiden Bedingungen. Die Häufigkeit sollte in der Qualitätsbedingung insgesamt höher sein, wegen der zu erwartenden komplexeren Äußerungen (s. Abschnitt 2.2).

Wir erwarten kein starkes Ansteigen der Syntax-Fehler unter Belastung. Die empirische Basis für eine solche Annahme ist zu schwach. Die Angaben bei Berthold (1998) nennen eine einzelne Studie.

Bezüglich der Artikulationsgeschwindigkeit erwarten wir gemäß Berthold (1998) ein Absinken unter Belastung, wobei sie in der Geschwindigkeitsbedingung insgesamt höher sein sollte, als in der Qualitätsbedingung. Für die restlichen Phänomene, die zu der Klasse der Geschwindigkeitssymptome gehören, erwarten wir ein Ansteigen unter Belastung.

Die Belastung sollte einen negativen Einfluß auf die Partnerorientierung haben. Dies könnte sich auf die Unterscheidung zwischen Qualitätsbedingung und Geschwindigkeitsbedingung insgesamt auswirken, da diese fast ausschließlich auf der Anpassung an den Dialogpartner basiert.

Bezüglich der inhaltlichen Qualität erwarten wird einen Haupteffekt von Qualität vs. Geschwindigkeit, d.h. eine durchschnittlich bessere Bewertung der Äußerungen aus der Qualitätsbedingung. Außerdem erwarten wird einen Haupteffekt von Belastung, d.h. eine bessere Bewertung der Äußerungen aus der unbelasteten Bedingung.

Was die Selbstkorrekturen angeht, können wir aufgrund der widersprüchlichen Befundlage keine Vorhersagen bezüglich des Haupteffektes der Belastung machen. Folgende Überlegung könnte dazu beitragen, eine Erklärung dafür zu finden, warum über dieses Phänomen in der Literatur derartig gegensätzliche Ergebnisse berichtet werden, und warum keine konkreten Vorhersagen gemacht werden können.

Nehmen wir einmal an, wir könnten beobachten, wann

1. kein inhaltlicher Fehler gemacht wurde
2. ein Fehler korrigiert wurde
3. ein unkorrigierter Fehler vorliegt

Wir nennen dies *vollständige Beobachtung*. Dann könnten unsere Erwartungen bezüglich der inhaltlichen Selbstkorrekturen wie folgt beschrieben werden.

	ohne Belastung	mit Belastung
$p(\text{kein Fehler})$	a	d
$p(\text{korrigierter Fehler})$	b	e
$p(\text{unkorrigierter Fehler})$	c	f

Die Buchstaben in der Tabelle repräsentieren Wahrscheinlichkeiten. Der Buchstabe a z.B. stellt $p(\text{kein Fehler}|\text{keine Belastung})$ dar, also die Wahrscheinlichkeit, dass kein inhaltlicher Fehler in der Äußerung gemacht wird unter der Voraussetzung, dass keine Arbeitsgedächtnisbelastung vorlag. Die allgemeinen Beziehungen zwischen den Wahrscheinlichkeiten (bei vollständiger Beobachtung) sind vermutlich wie folgt: $a > d, c < f$. Über die Beziehung zwischen b und e lassen sich jedoch keine genauen Aussagen machen. Es könnte sein, dass unter Belastung mehr Selbstkorrekturen zu beobachten sind ($b < e$), dann nämlich, wenn insgesamt mehr Fehler gemacht wurden ($e + f > b + c$). Auf der anderen Seite könnte es jedoch auch sein, dass die Reduktion der Korrekturquote überwiegt ($e/e + f \leq b/b + c$), was dazu führt, dass unter Belastung weniger Selbstkorrekturen zu beobachten sind. In Wirklichkeit sind die Zustände *kein Fehler* und *unkorrigierter Fehler* nicht zu unterscheiden, da die eigentliche Äußerungsplanung des Sprechers unter normalen Umständen nicht eingesehen werden kann. Vorhersagen in Bezug auf das an der Oberfläche zu beobachtenden Korrekturphänomens sind jedoch problematisch. Auch wenn in einem Experiment ein eindeutiges Ergebnis herauskommt, wird dieses schwer zu verallgemeinern sein. Es handelt sich um zwei entgegengesetzte Tendenzen: Fehler-rate und Korrekturrate. Allein die Tatsache, dass in diesem speziellen Fall eine Tendenz überwiegt, heisst nicht, dass sie das immer tun wird.

3.4 Ergebnisse

3.4.1 Auswirkungen der Äußerungskomplexität

In Abschnitt 2.2 wurde von den Ergebnissen einer Untersuchung von Oviatt (1997) berichtet, die einen exponentiellen Anstieg von *Disfluencies* in Abhängigkeit der Äußerungskomplexität beobachten konnte. Als Maß für die Komplexität der Äußerung wurde bei Oviatt (1997) die Anzahl der Wörter zugrunde gelegt, die jedoch im Experiment nicht kontrolliert manipuliert wurde. Um den *Oviatt-Effekt* zu berücksichtigen, wurde der Einfluß der Äußerungskomplexität auf die Ergebnisse dieses Experimentes in einer Reihe von Analysen untersucht. Zunächst wurden die gleichen Symptome zugrunde gelegt, wie Oviatt (1997) sie verwendet hat, nämlich Selbstkorrekturen⁷, syntaktische Fehler, Fehlansätze, Wiederholungen, Anzahl der gefüllten Pausen, Wortfragmente und Satzfragmente. Das Vorkommen dieser Symptome pro 100 Wörter wurde berechnet, um den einfachen Effekt der Äußerungslänge zu neutralisieren. Anschließend wurde die Korrelation mit der Anzahl der Wörter ermittelt. Um eine Verzerrung des Ergebnisses durch Ausnahmefälle zu vermeiden, wurden nur Sätze berücksichtigt, die mindestens 4 und höchstens 25 Wörter umfassten, und von deren Typ (Anzahl der Wörter) es mindestens neun weitere Äußerungen gab. Der Oviatt-Effekt konnte auch mit den Daten des Flughafenexperimentes nachgewiesen werden. Die Korrelation ist signifikant ($p < 0,05$), jedoch mit 0,523 deutlich geringer als in der Originalstudie (s. Abschnitt 2.2). Dies ist vermutlich darauf zurückzuführen, dass die Äußerungen in diesem Experiment zum Großteil aus zwei Sätzen bestanden.

Aus diesem Grund wurde eine zweite Analyse vorgenommen, bei der als Maß für die Äußerungskomplexität die Einschätzung der Komplexität der Bilder zugrunde gelegt wurde (s. Abschnitt 3.3.3.3). Dabei wurde angenommen, dass die Komplexität der Sprachaufgabe mit der Komplexität der resultierenden Äußerung korreliert. Die Korrelation betrug in diesem Fall (0,756 bei einer Signifikanz von $p < 0,01$).

Diese Ergebnisse bestätigen die Resultate von Oviatt (1997), wonach die Äußerungskomplexität ein entscheidender Faktor für die Häufigkeit von Symptomen ist. Bei der Interpretation der Hauptergebnisse des Experimentes, die in den folgenden Abschnitten vorgestellt werden, sollte dies berücksichtigt werden.

3.4.2 Analyseverfahren

In Abschnitt 3.3.3 wurde die grobe Einteilung der abhängigen Variablen beschrieben. Wie man in Tabelle 3.4 sieht, umfasst auch jede der Hauptgruppen einige Variablen, die vermutlich stark miteinander zusammenhängen. Bevor wir zur Analyse einzelner Effekte übergehen, sollten wir eine Übersicht über die Beziehungen unter den Variablen innerhalb jeder Gruppe erhalten.

3.4.2.1 Faktoranalysen

Hierzu wurde innerhalb jeder Kategorie von Variablen eine Faktoranalyse ausgeführt (Hauptachsenfaktorisierung mit Rotation der Achsen nach der Methode Oblimin mit Kaiser-Normalisierung). Im wesentlichen identifiziert diese Analysemethode Gruppen

⁷Oviatt unterschied nicht zwischen syntaktischen und inhaltlichen Selbstkorrekturen

von Variablen, bei denen die Variablen innerhalb jeder Gruppe relativ stark (über Äußerungen) zusammenhängen. Im Gegensatz zu Faktoranalysen ohne Rotation wird hierbei *nicht* verlangt, dass die Gruppen orthogonal (unkorreliert) sind. Zwei Gruppen (z.B. stille Pausen und gefüllte Pausen) können also durchaus noch miteinander korreliert sein. Eine Analyse mit Rotation wurde gewählt, weil die dadurch entstehenden Gruppen in der Regel interpretierbarer sind als das ohne Rotation möglich ist.

In die Faktoranalysen gingen die Daten für einzelne Äußerungen ein (und nicht etwa Durchschnitte für Versuchspersonen oder Stimulusbilder).

3.4.2.2 Multivariate Varianzanalysen

Wenn in einem Experiment mehrere abhängige Variablen erfasst werden, sollte grundsätzlich eine multivariate Varianzanalyse (MANOVA) durchgeführt werden, bevor Effekte auf einzelne Variablen untersucht werden. Sonst können schon wegen der Vielzahl der Variablen einzelne signifikante Effekte auftreten, deren tatsächliche Bedeutung schwer einzuschätzen ist. Eine MANOVA prüft, ob die unabhängigen Variablen überhaupt signifikante Effekte auf die Gesamtheit der abhängigen Variablen hatten.

Theoretisch hätte man für sämtliche abhängigen Variablen des Experiments eine einzige MANOVA durchführen können. Dieses Verfahren wäre bei unserer Anzahl von Variablen jedoch praktisch nicht möglich gewesen. Stattdessen wurde für jede der drei Hauptkategorien von Variablen eine MANOVA durchgeführt - und ggf. auch für einzelne Gruppen von Variablen, die durch die Faktoranalyse identifiziert wurden.

Sowohl die MANOVAs als auch die univariaten Varianzanalysen (ANOVAs) berücksichtigten die Tatsache, dass von jeder Versuchsperson 80 Beobachtungen erhalten wurden.

In den folgenden Abschnitten werden die Ergebnisse der Analyse der Daten dargestellt. Für jede der drei Hauptkategorien werden zunächst die Ergebnisse der Faktoranalysen besprochen. Anschließend werden MANOVA und ANOVA präsentiert. Wo signifikante Effekte zu beobachten sind, werden die Tendenzen der Einzelvariablen anhand eines Diagramms dargestellt und diskutiert. Die Artikulationsgeschwindigkeit wird separat besprochen. Obwohl sie für die Dauer der Äußerung relevant ist, wurde sie nicht zu den Zeitsymptomen gezählt. Der Grund dafür ist, dass die Artikulationsgeschwindigkeit eine andere Einheit hat (pro Sekunde statt pro Äußerung).

3.4.3 Zeitsymptome

In Tabelle 3.5 werden die Ergebnisse der Faktoranalyse aller Zeitsymptome dargestellt. Auf Faktor eins laden Anzahl und Dauer der gefüllten Pausen. Auf Faktor zwei laden Anzahl und Dauer der stillen Pausen und - obwohl weniger deutlich - auch Einsatzlatenz und Wiederholungen. Der Grund für die geringere Korrelation der Einsatzlatenz als „Stille Pause am Anfang der Äußerung“ mit den stillen Pausen im Innern ist eventuell die experimentelle Situation: ein Großteil der Einsatzlatenz muß interpretiert werden als Zeit, die die Versuchsperson braucht, um das dargestellte Bild zu verstehen. Das führt dazu, dass die Einsatzlatenzen im Bereich von ein bis zwei Sekunden liegen, während die stillen Pausen in der Regeln wesentlich kürzer sind (200 bis 800 ms, selten mehr). Die Korrelation des geringeren Planungsanteils der Einsatzlatenz mit den stillen Pausen ist dementsprechend niedrig. Die Wiederholungen sind ein eher seltenes Phänomen, was

Tabelle 3.5: Faktoranalyse aller Zeitsymptome (Hauptachsenfaktorierung mit Rotation der Achsen nach der Methode Oblimin mit Kaiser-Normalisierung). Die Werte, die mit Null beziffert sind, stellen in Wirklichkeit eine verschwindend kleine Zahl dar.

	Faktor 1	Faktor 2	Faktor 3	Faktor 4
Gefüllte Pausen (Anzahl)	0,917	0	0	0
Gefüllte Pausen (Dauer)	0,961	0	0	0
Stille Pausen (Anzahl)	0	0,821	0	0
Stille Pausen (Dauer)	0	0,910	0	0
Einsatzlatenz	0	0,123	-0,108	0
Wiederholungen	0	0,157	0	0
Silbendehnungen	0	0	0	0,125
Zögerungen	0	0	0	0,547
Begrüßungen	0	0	0,693	0
Anreden	0	0	0,595	0

die geringe Korrelation dieser Variable erklärt. Faktor eins und Faktor zwei sind für die Interpretation besonders interessant, weil die wichtigsten Symptome vertreten sind. Wichtig sind sie deshalb, weil sie relativ häufig vorkommen (im Fall von Pausen) und in den bisherigen Studien bereits untersucht wurden (s. Berthold & Jameson, 1999).

Anreden und Begrüßungen laden auf Faktor 3 dar. Für die Interpretation sind sie jedoch nicht zentral, da sie lediglich bei drei der Versuchspersonen vorkamen. Unter Begrüßungen verstehen wir das Vorkommen einer Begrüßungsformel, unter Anreden die Nennung eines der Namen der beiden Kommunikationspartnerinnen, die in der Anweisung zum Experiment genannt wurden (s. 3.3.2.9). In Beispiel 3.6 sind beide Phänomene enthalten. Sie treten häufig gemeinsam auf, was die Zugehörigkeit zum selben Faktor erklärt.

(3.6) <31/230500/Q+/2/002/4.40>

<EL:0.71> Hallo <G> Frau Neumann <N> <iaeh:0.15> Ich habe etwas verloren. Wo ist denn die Fundstelle <S> <S>?

Auf diesen Faktor lädt, mit negativer Korrelation zu Anreden und Begrüßungen, auch die Einsatzlatenz. Wenn Anreden und Begrüßungen vorkommen, ist also die Einsatzlatenz geringer. Wir können erstere demnach als Füllphrasen interpretieren, die vom Sprecher eingesetzt werden, um die Einsatzlatenz zu verringern. Es ist nicht ganz klar, ob sie sich nur auf die Einsatzlatenz im engeren Sinne auswirken oder ob sie bereits während der Phase der Interpretation des Bildes geäußert werden. Wie schon bei den stillen Pausen in Faktor zwei, ist die Korrelation gering.

Auf Faktor vier laden die Silbendehnungen und Zögerungen. Auch hier handelt es sich um Variablen, die selten vorkommen, mit entsprechend geringer Korrelation. Dieser Faktor spielt deshalb wie Faktor drei bei der weiteren Analyse eine untergeordnete Rolle.

Multivariate Analyse aller Zeitsymptome Eine multivariate Analyse wurde durchgeführt mit den Zeitsymptomen als abhängige Variablen. Die unabhängigen Variablen

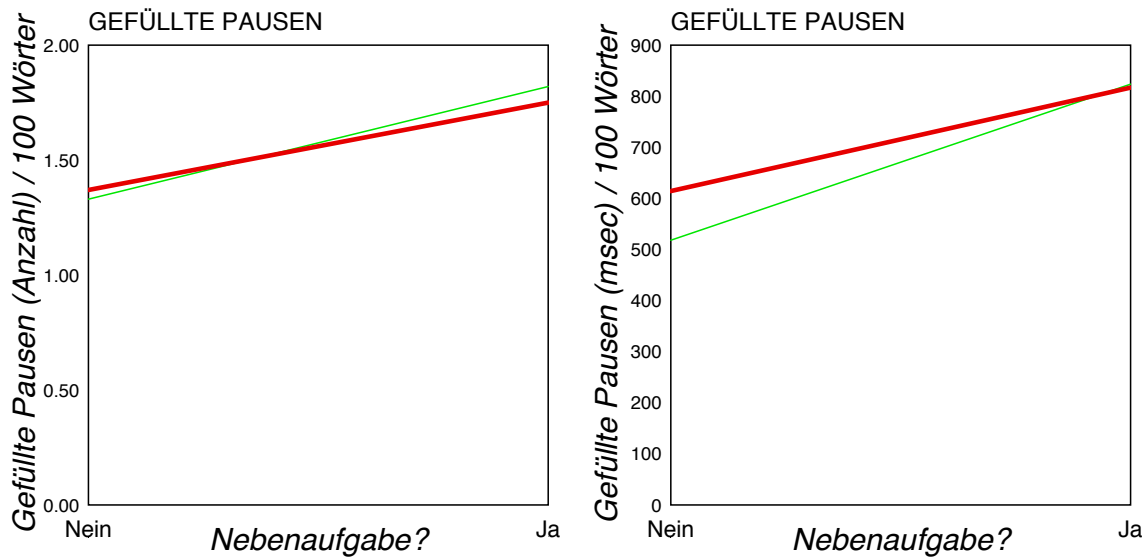


Abbildung 3.12: Durchschnittliche Werte für Anzahl und Dauer der gefüllten Pausen in den vier Bedingungen

waren

- Belastung vs. keine Belastung
- Qualität vs. Geschwindigkeit
- Qualität/Geschwindigkeit ohne Belastung vs. Qualität/Geschwindigkeit mit Belastung

Die MANOVA sollte zeigen, ob überhaupt Effekt der abhängigen Variablen auf die Zeitsymptome vorliegen. Ist dies nicht der Fall, ist jede weitere Interpretation nicht gerechtfertigt. Die beiden Haupteffekte sind signifikant mit $F(10, 22) = 11.327, p < 0.01$ für Qualität vs. Geschwindigkeit bzw. $F(10, 22) = 4.191, p < 0.01$ für Belastung vs. keine Belastung. Die Interaktion ist ebenfalls signifikant mit $F(10, 22) = 2.929, p < 0,05$. Die Interpretation der univariaten Analysen ist deshalb zulässig.

3.4.4 Weitere Analysen der einzelnen Faktoren

Faktor eins: gefüllte Pausen Mit dem ersten Faktor (gefüllte Pausen) wurde keine MANOVA ausgeführt, da die Variablen dieses Faktors sehr interkorreliert sind. Die univariate Analyse ergab einen signifikanten Haupteffekt von Belastung auf die gefüllten Pausen mit $F(1, 31) = 4,435, p < 0,05$ für die Anzahl und $F(1, 31) = 5,924, p < 0,05$ für die Dauer. Der Haupteffekt von Qualität vs. Geschwindigkeit ist nicht signifikant. Ebensowenig die Interaktion. Aus Abbildung 3.12 sind die Tendenzen ersichtlich.

Die Sprecher machten unter Belastung mehr gefüllte Pausen als in den unbelasteten Bedingungen, was unter Berücksichtigung der Ergebnisse früherer Studien zu erwarten war (s. Abschnitt 2.1).

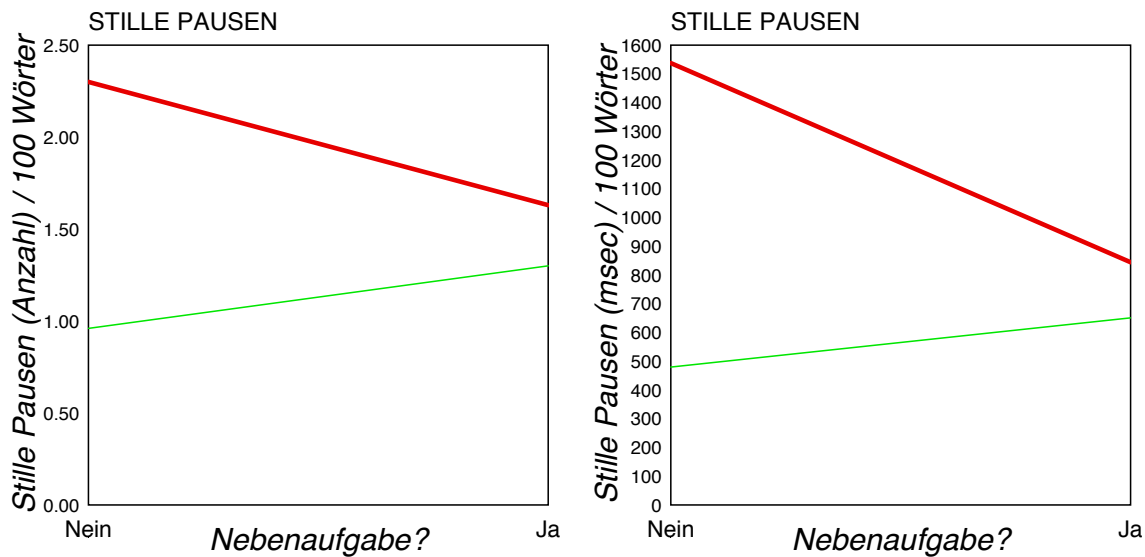


Abbildung 3.13: Durchschnittliche Werte für Anzahl und Dauer der stillen Pausen in den vier Bedingungen

Faktor zwei: stille Pausen und Wiederholungen Für den zweiten Faktor, stille Pausen und Wiederholungen, wurde eine multivariate Analyse vorgenommen. Der Haupteffekt von Qualität vs. Geschwindigkeit ist signifikant mit $F(4, 28) = 19,6, p < 0,01$. Der Haupteffekt von Belastung ist marginal signifikant mit $F(4, 28) = 2,299, p < 0,08$. Die Interaktion ist signifikant mit $F(4, 28) = 4,921, p < 0,05$. Abgesehen von möglichen Haupteffekten von Belastung ist die weitere Analyse dieses Faktors also gerechtfertigt.

Die ANOVA des Faktors der stillen Pausen und Wiederholungen erbrachte folgende Ergebnisse: Der Haupteffekt von Qualität vs. Geschwindigkeit auf die stillen Pausen ist signifikant mit $F(1, 31) = 20,844, p < 0,01$ für die Anzahl und $F(1, 31) = 15,828, p < 0,01$ für die Dauer. In der Qualitätsbedingung ist ein höheres Maß an Planung erforderlich, was sich auf die Häufigkeit und Länge der stillen Pausen auswirkt. Was die Einsatzlatenz betrifft, war ebenfalls ein signifikanter Effekt nachweisbar mit $F(1, 31) = 64,291, p < 0,01$. Auch dieser Effekt ist auf einen erhöhten Planungsaufwand bei der Produktion von Äußerungen höherer Qualität zurückzuführen. In Bezug auf die Wiederholungen war kein signifikanter Haupteffekt von Qualität vs. Geschwindigkeit feststellbar. Der Haupteffekt von Belastung sollte aufgrund der Ergebnisse der multivariaten Analyse nicht weiter untersucht werden. Die Interaktion ist im Falle der stillen Pausen signifikant mit $F(1, 31) = 15,032, p < 0,01$ für die Anzahl und $F(1, 31) = 11,010, p < 0,01$ für die Dauer. Was die Einsatzlatenz betrifft, war ebenfalls eine signifikante Interaktion feststellbar mit $F(1, 31) = 5,412, p < 0,05$. Nicht signifikant ist die Interaktion in Bezug auf die Wiederholungen.

Abbildung 3.13 zeigt, warum ein positiver Haupteffekt von Belastung, der intuitiv plausibel gewesen wäre, nicht signifikant auftritt: die Effekte von Belastung sind für die Qualitätsbedingung und die Geschwindigkeitsbedingung unterschiedlich und sie gleichen sich etwas aus. In Abbildung 3.14 werden die Tendenzen der Einsatzlatenz dargestellt.

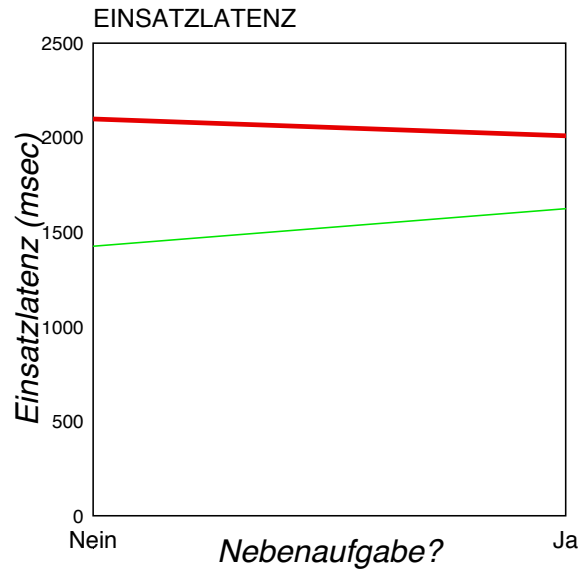


Abbildung 3.14: Durchschnittliche Werte für die Einsatzlatenz in den vier Bedingungen

An den konvergierenden Linien in beiden Abbildungen ist zu erkennen, dass der Effekt von Qualität/Geschwindigkeit unter Belastung deutlich zurückgeht. Wir gehen davon aus, dass dies teilweise auf eine Verringerung der Partnerbezogenheit unter Belastung zurückzuführen ist. Dies ist konform mit den Schlußfolgerungen von Oberauer und Hockl (2000) und Roßnagel (1996), bei denen die Phase der Partnerorientierung als belastungssensitiv angesehen wird. Während sich die Sprecher in der unbelasteten Bedingung bemühen, ihre Sprache an den Partner anzupassen, ist dies unter Belastung weniger der Fall. Das kann zwei mögliche Ursachen haben:

1. Aufgrund der zusätzlichen Belastung vergisst der Sprecher die gegebenen Anforderungen.
2. Der Sprecher ist sich zwar nach wie vor der Anforderungen bewusst, kann sie jedoch aufgrund der Belastung nicht erfüllen.

In Abbildung 3.15 werden die Tendenzen der Wiederholungen dargestellt. Die fehlende Signifikanz ist möglicherweise auf die geringe Häufigkeit der Variablen zurückzuführen.

Faktor drei: Begrüßungen und Anreden Wie bereits angedeutet, ist die Häufigkeit von Begrüßungen und Anreden (Faktor 3) zu gering für eine Interpretation. Durch eine entsprechende multivariate Analyse wurde dies bestätigt. Weder die beiden Haupteffekte, noch die Interaktion sind signifikant. Die Tendenzen sind jedoch konsistent mit dem Faktor der Zeitsymptome (s. Abbildung 3.16).

Faktor vier: Silbendehnungen und Zögerungen Als nächstes wurde eine multivariate Analyse des Faktors der Silbendehnungen und Zögerungen durchgeführt. Der Haupteffekt

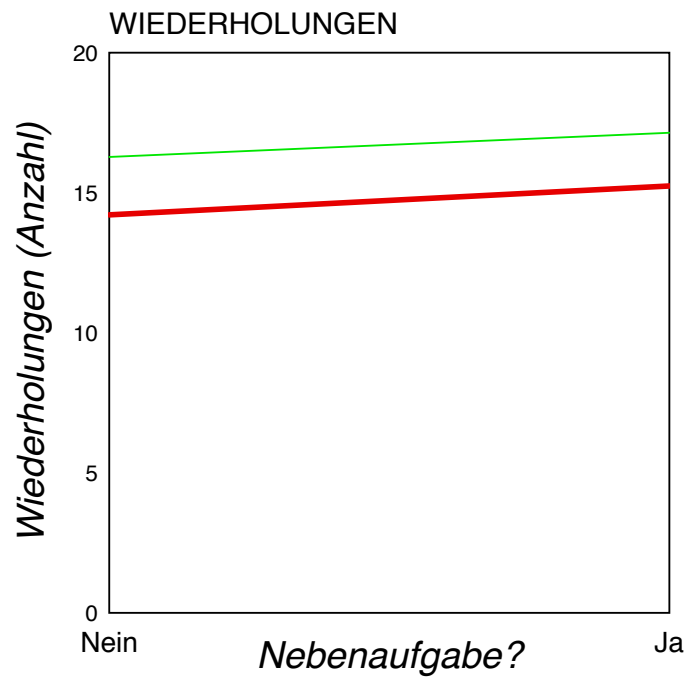


Abbildung 3.15: Durchschnittliche Werte für die Wiederholungen in den vier Bedingungen

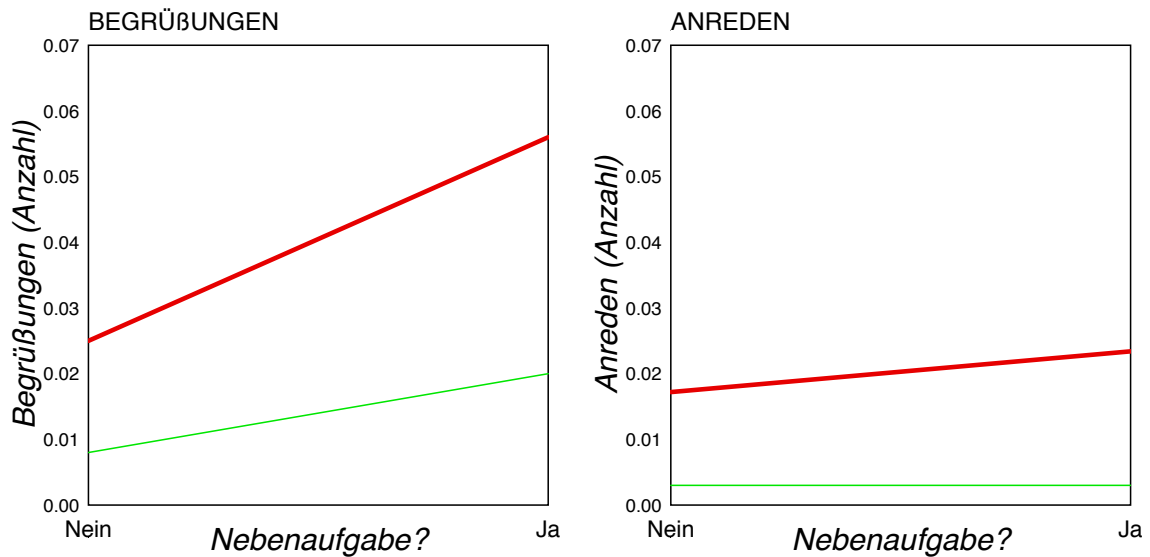


Abbildung 3.16: Durchschnittliche Werte für die Begrüßungen und Anreden in den vier Bedingungen

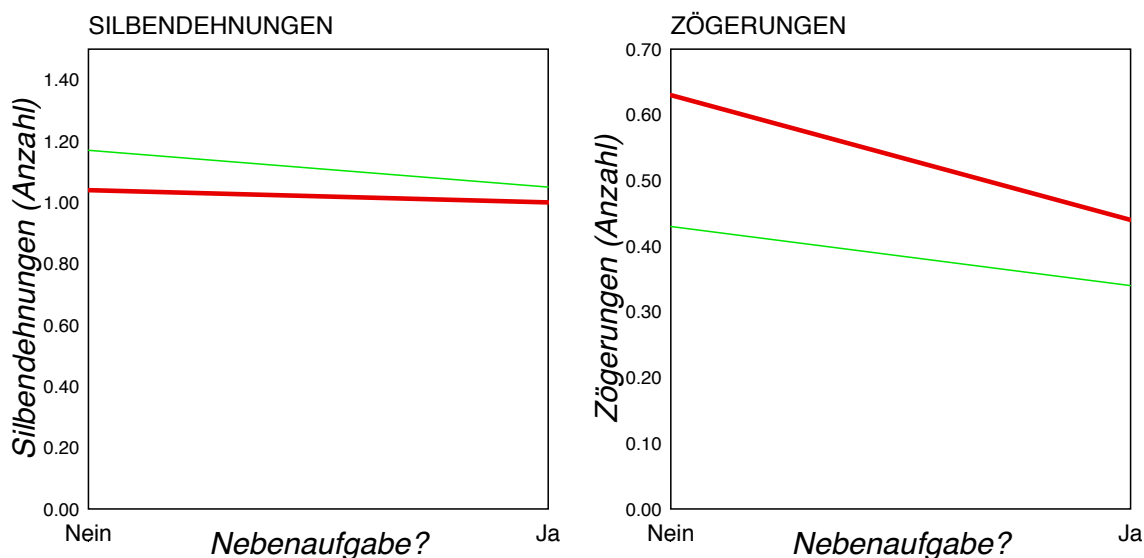


Abbildung 3.17: Durchschnittliche Werte für die Silbendehnungen und Zögerungen in den vier Bedingungen

von Qualität vs. Geschwindigkeit ist signifikant mit $F(2, 30) = 23,466, p < 0,01$. Nicht signifikant dagegen ist auch in diesem Fall der Haupteffekt von Belastung sowie die Interaktion. Die Häufigkeit der Phänomene ist anscheinend bei diesem Faktor ebenfalls nicht groß genug, um zu signifikanten Ergebnissen zu führen. Von einer weiteren Analyse der beiden Faktoren durch univariate Tests sehen wir deshalb ab. Die Tendenzen werden in Abbildung 3.17 dargestellt. Auch hier sind deutlich die konvergierenden Linien zu erkennen, die auf eine rückläufige Partnerbezogenheit unter Belastung schließen lassen.

3.4.5 Qualitätssymptome

Mit den Qualitätssymptomen als zweiter Hauptkategorie wurde zunächst ebenfalls eine Faktoranalyse vorgenommen, deren Ergebnisse in Tabelle 3.6 dargestellt werden.

Tabelle 3.6: Faktoranalyse aller Qualitätssymptome

	1	2	3
Syntaktische Selbstkorrekturen	0	0	0,450
Fehlansätze	0,303	0	0
Wortfragmente	0,569	-0,249	0,304
Satzfragmente	0	-0,113	0
Inhaltliche Selbstkorrekturen	0,195	-0,627	0

Auf Faktor eins laden die inhaltlichen Selbstkorrekturen, Fehlansätze und Wortfragmente. Auf Faktor zwei laden Wort- und Satzfragmente, auf Faktor drei Wortfragmente und syntaktische Selbstkorrekturen. Die Tatsache, dass Wortfragmente in jeder der

drei Faktoren vorkommen, lässt sich z.T. auf eine überlappende Kodierung der Daten zurückführen. Selbstkorrekturen enthalten z.B. häufig Wortabbrüche, die ihrerseits eigene Symptome darstellen (s. Beispiel 3.7).

(3.7) <18/220300/G+/5/001/6.16>

<EL:1.88> Ich möchte <SP:0.34> mir einen Mietwagen <+i/> <-/> kauf <-/>
</+i> <äh:0.24> mieten. Wo gibts den? <S> <S>

Die multivariate Analyse der Qualitätssymptome ergab lediglich einen marginalen Haupteffekt von Belastung mit $F(5, 27) = 20,374, p < 0,066$. Der Haupteffekt von Qualität vs. Geschwindigkeit ist nicht signifikant. Wie wir sehen werden, gehen die Effekte auf die sechs abhängigen Variablen in unterschiedliche Richtungen: bei dreien ist er positiv, bei dreien negativ. Die Interaktion ist ebenfalls nicht signifikant.

Die Vorhersagen bezüglich der Qualitätssymptome waren weniger spezifisch, als es bei den Geschwindigkeitssymptomen der Fall war (s. Abschnitt 3.3.4). Der erwartete Haupteffekt von Belastung wurde beobachtet. Obwohl die Ergebnisse der MANOVA in diesem Fall marginal waren, wurde eine Analyse der einzelnen Faktoren vorgenommen.

3.4.5.1 Weitere Analysen der einzelnen Faktoren

Wie wir aus den Ergebnissen der MANOVA der Qualitätssymptome bereits vermuten konnten, ergaben die multivariaten Analysen der einzelnen Faktoren ebenfalls keine signifikanten Effekte. Im Folgenden werden daher lediglich die Ergebnisse aus den univariaten Tests dargestellt.

Fehlansätze Die univariate Analyse der Fehlansätze ergab einen signifikanten Haupteffekt von Belastung mit $F(1, 31) = 40,173, p < 0,05$. Die Tendenzen werden in Abbildung 3.18 dargestellt. Sie bestätigen die früheren Ergebnisse, die in Abschnitt 2.1 zusammengefasst wurden. Wie z.B. bei Berthold (1998) berichtet wird, können Fehlansätze als Symptome kognitiver Belastung angesehen werden.

Satzfragmente Eine ANOVA der Satzfragmente ergab einen marginal signifikanten Effekt von Belastung mit $F(1, 31) = 30,683, p < 0,064$. Die Tendenz geht jedoch in die vorhergesagte Richtung (s. Abbildung 3.19). Dies bestätigt frühere Ergebnisse, wie sie von Berthold (1998) zusammengefasst wurden.

Inhaltliche Selbstkorrekturen Die univariate Analyse der inhaltlichen Selbstkorrekturen ergab keine signifikanten Effekte. Auf die Problematik bei der Interpretation dieser Variablen wurde bereits in Abschnitt 3.3.4 eingegangen.

3.4.6 Längensymptome

Für die Längensymptome wurde durch die Faktoranalyse ein einziger Faktor gefunden (s. Tabelle 3.7).

Für diesen Faktor wurde ebenfalls eine multivariate Analyse durchgeführt. Beide Haupteffekte sowie die Interaktion sind signifikant mit $F(4, 28) = 28,308, p < 0,01$

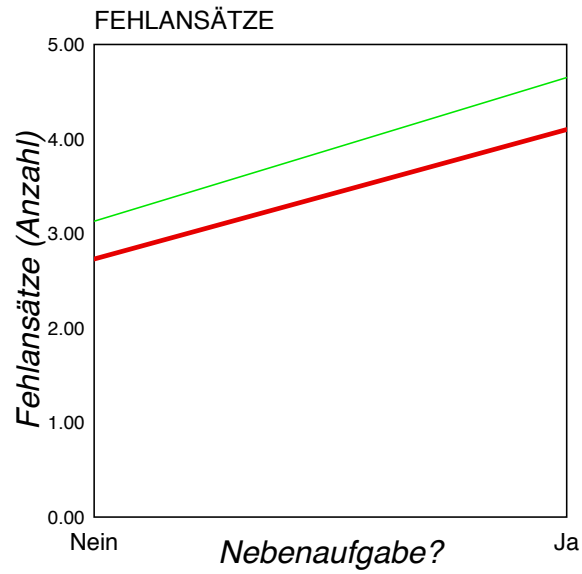


Abbildung 3.18: Durchschnittliche Werte für die Fehlansätze in den vier Bedingungen

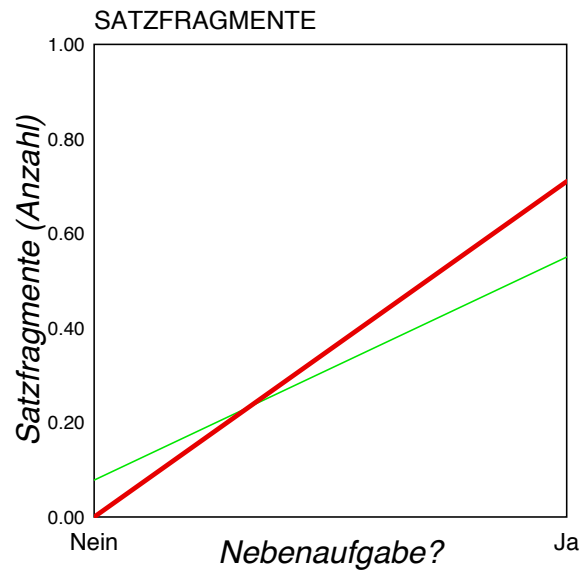


Abbildung 3.19: Durchschnittliche Werte für die Satzfragmente in den vier Bedingungen

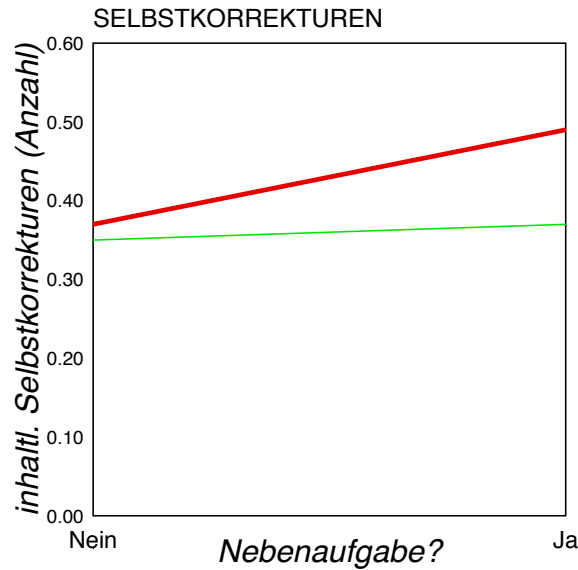


Abbildung 3.20: Durchschnittliche Werte für die inhaltlichen Selbstkorrekturen in den vier Bedingungen

Tabelle 3.7: Faktoranalyse der Längensymptome

	1
Gesamtlänge	0,760
Anzahl der Silben	0,995
Anzahl der Sätze	0,572
Anzahl der Wörter	0,978

(Haupteffekt von Qualität vs. Geschwindigkeit), $F(4, 28) = 2,861, p < 0,05$ (Haupteffekt von Belastung), bzw. $F(4, 28) = 5,838, p < 0,01$ (Interaktion). Die Analyse der individuellen Variablen ist demnach ohne Weiteres gerechtfertigt.

Die univariate Analyse ergab einen signifikanten Haupteffekt von Qualität vs. Geschwindigkeit auf die Gesamtlänge mit $F(1, 31) = 94,530, p < 0,01$, auf die Anzahl der Silben mit $F(1, 31) = 49,236, p < 0,01$, auf die Anzahl der Sätze mit $F(1, 31) = 13,802, p < 0,01$, und auf die Anzahl der Wörter mit $F(1, 31) = 47,491, p < 0,01$. In Abbildung 3.21 werden die durchschnittlichen Werte für die Längensymptome in den vier Bedingungen dargestellt. Der Haupteffekt von Belastung auf diese Variablen ist durchweg ebenfalls signifikant mit $F(1, 31) = 2,565, p < 0,01$ (Gesamtlänge), $F(1, 31) = 6,782, p < 0,05$ (Anzahl der Silben), $F(1, 31) = 7,700, p < 0,01$ (Anzahl der Sätze), $F(1, 31) = 5,662, p < 0,05$ (Anzahl der Wörter). Signifikant ist auch die Interaktion mit $F(1, 31) = 13,101, p < 0,01$, $F(1, 31) = 9,117, p < 0,01$, $F(1, 31) = 11,052, p < 0,01$, und $F(1, 31) = 8,204, p < 0,01$.

Wie erwartet, sind die Äußerungen in der Qualitätsbedingung deutlich länger, als in der Geschwindigkeitsbedingung. Die Tendenzen der Längensymptome werden in Abbildung

3.21 dargestellt. Die Linien zeigen die typische Konvergenz, die wir bereits bei den stillen Pausen beobachtet haben.

3.4.7 Die Artikulationsgeschwindigkeit

Die Artikulationsgeschwindigkeit wurde als einzelnes Symptom analysiert, da es keiner der drei Hauptkategorien angehört (s. Abschnitt 3.4.2.1). Eine univariate Analyse wurde durchgeführt mit der Artikulationsgeschwindigkeit als abhängige Variable. Beide Haupteffekte sind signifikant, der Haupteffekt von Qualität vs. Geschwindigkeit mit $F(1, 31) = 73,945, p < 0,01$, und der Haupteffekt von Belastung mit $F(1, 31) = 19,958, p < 0,01$. Die Interaktion ist signifikant mit $F(1, 31) = 4,297, p < 0,05$. In Abbildung 3.22 werden die durchschnittlichen Werte für die Artikulationsgeschwindigkeit in den vier Bedingungen dargestellt. Diese unterstützen die Annahme, dass Sprecher unter Zeitdruck schneller artikulieren, was meines Wissens bislang nicht explizit untersucht worden ist. Unter Belastung geht die Artikulationsgeschwindigkeit tendenziell zurück.

3.4.8 Inhaltliche Qualität

Da die oben aufgeführten Qualitätssymptome, die an der Oberfläche der Äußerung erkennbar sind, durchweg die formale Qualität einer Äußerung betreffen, wurde die inhaltliche Qualität separat gemessen (s. Abschnitt 3.3.3.2). Das Ergebnis ist der durchschnittliche Rang, den eine Äußerung von den vier Bewertern erhalten hat. Mit dieser Variablen wurde ebenfalls eine univariate Analyse durchgeführt. Der Haupteffekt von Qualität vs. Geschwindigkeit ist signifikant mit $F(1, 31) = 54,986, p < 0,01$. Der Haupteffekt von Belastung ist ebenfalls signifikant mit $F(1, 31) = 12,562, p < 0,01$. Die Interaktion ist nicht signifikant. In Abbildung 3.23 werden die Tendenzen dargestellt.

3.4.9 Belastung durch die Aufgabenkomplexität

In Abschnitt 3.4.1 wurde darauf hingewiesen, dass der Einfluß der Äußerungskomplexität auf das Vorkommen von Symptomen nicht vernachlässigt werden sollte. Aus diesem Grund wurde die Korrelation der Komplexität mit der Häufigkeit der einzelnen abhängigen Variablen untersucht. Als Maß für die Komplexität diente die eingeschätzte Schwierigkeit der Bilder (s. Abschnitt 3.4.1). Die Ergebnisse werden in Tabelle 3.8 dargestellt. Aufgrund ihres seltenen Vorkommens wurden Begrüßungen und Anreden nicht berücksichtigt. Die Sterne markieren die signifikanten Korrelationen.

Insgesamt sind die Korrelationen hoch. Bis auf die inhaltlichen Selbstkorrekturen sind alle Ergebnisse signifikant. Auf die Probleme, die die Interpretation dieses Symptoms mit sich bringt, wurde in Abschnitt 3.4.4 bereits eingegangen.

3.4.10 Zusammenfassung der Ergebnisse

Insgesamt konnten eine Reihe signifikanter Haupteffekte nachgewiesen werden. Ein signifikanter Haupteffekt von Belastung wurde für gefüllte Pausen, Fehlansätze, Länge der Äußerung sowie Artikulationsgeschwindigkeit beobachtet. Einen marginal signifikanten

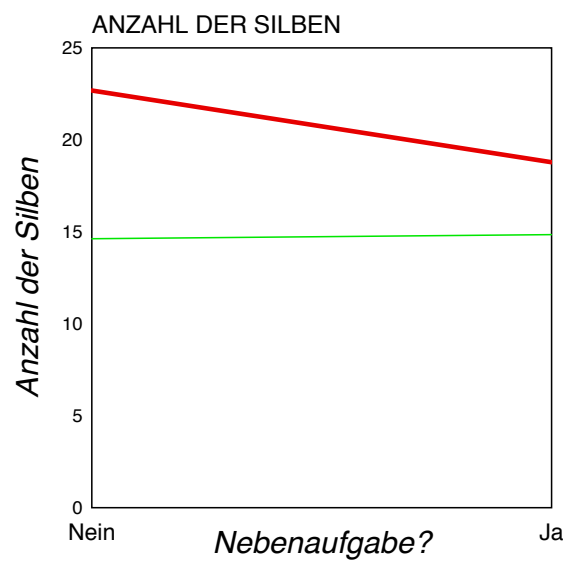
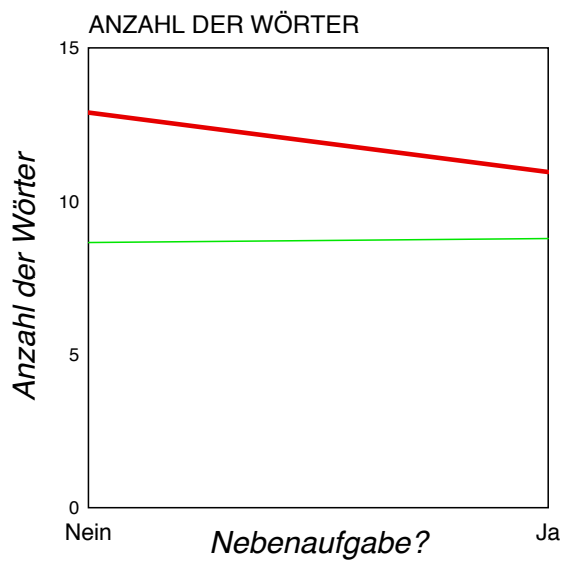
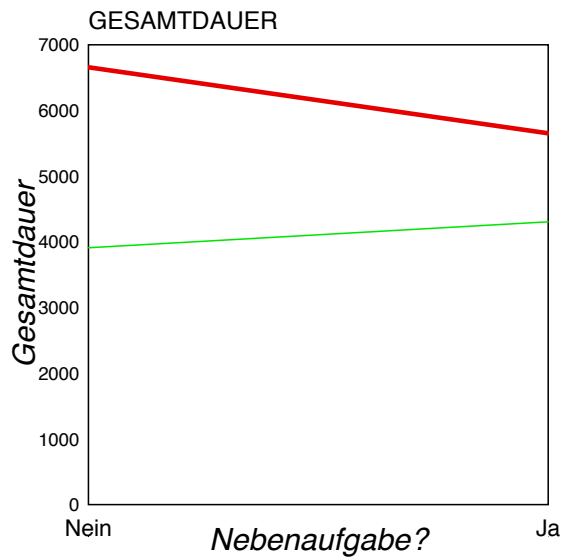


Abbildung 3.21: Durchschnittliche Werte für die Längensymptome in den vier Bedingungen

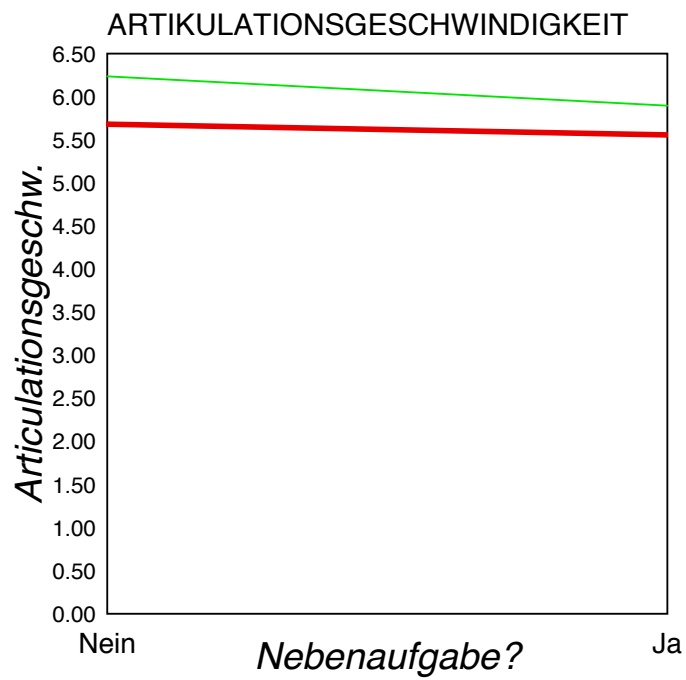


Abbildung 3.22: Durchschnittliche Werte für die Artikulationsgeschwindigkeit in den vier Bedingungen

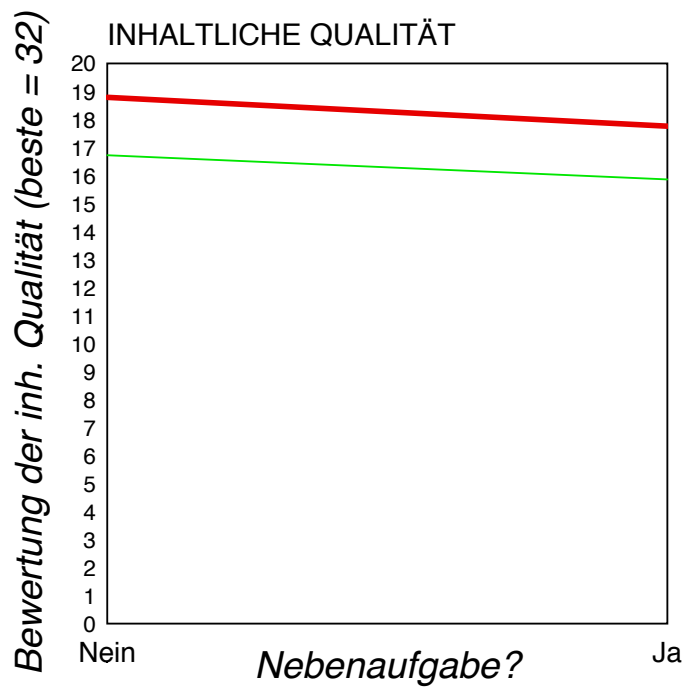


Abbildung 3.23: Bewertung der inhaltlichen Qualität

Tabelle 3.8: Korrelation der Aufgabenkomplexität mit den einzelnen Symptomen

<i>Zeitsymptome</i>	
Anzahl der gefüllten Pausen	0,502(**)
Dauer der gefüllten Pausen	0,472(**)
Anzahl der stillen Pausen	0,531(**)
Dauer der stillen Pausen	0,503(**)
Einsatzlatenz	0,503(**)
Zögerungen	0,344(**)
Silbendehnungen	0,504(**)
Wiederholungen	0,296(**)
<i>Qualitätssymptome</i>	
Fehlansätze	0,424(**)
inhaltliche Selbstkorrekturen	0,176
syntaktische Selbstkorrekturen	0,534(**)
Wortabbrüche	0,284(*)
Syntaktische Fehler	0,287(**)
Satzabbrüche	0,231(*)
<i>Längensymptome</i>	
Gesamtzeit	0,748(**)
Anzahl der Silben	0,489(**)
Anzahl der Wörter	0,549(**)
Artikulationsgeschwindigkeit	-0,462(**)

Effekt hat die Belastung auf stille Pausen und Satzfragmente. Ein signifikanter Haupteffekt von Qualität vs. Geschwindigkeit konnte für stille Pausen, Silbendehnungen, Zögerungen, Länge der Äußerungen und Artikulationsgeschwindigkeit nachgewiesen werden. Neben den Haupteffekten konnte in einigen Fällen eine signifikante Interaktion nachgewiesen werden, was für die Interpretation ebenfalls bedeutsam ist. Dies betrifft die Variablen stille Pausen, Einsatzlatenz, Länge der Äußerungen und Artikulationsgeschwindigkeit. Für die meisten der abhängigen Variablen konnte eine starke Korrelation zwischen Häufigkeit und Aufgabenkomplexität festgestellt werden⁸.

3.5 Belastung durch die Nebenaufgabe

Insgesamt betrachtet ist der Haupteffekt von Belastung geringer als der von Qualität vs. Geschwindigkeit. Es scheint also so zu sein, dass die Versuchspersonen trotz Belastung durch die Nebenaufgabe, die Hauptaufgabe ohne gravierende Probleme durchführen konnten. Wir sind jedoch nicht der Meinung, dass die Nebenaufgabe nicht belastend genug war. Durch die Erfahrung aus einer Reihe von Vorversuchen wurde die Aufga-

⁸Die Korrelation war nicht signifikant bei den inhaltlichen Selbstkorrekturen. Begrüßungen und Anreden wurden nicht berücksichtigt (s. Abschnitt 3.4.9)

be deutlich erschwert. Während des Experimentes konnte beobachtet werden, dass die Versuchspersonen die Navigationsaufgabe sowie das Memorieren der Ziffern nicht ohne Schwierigkeiten bewältigen konnten. Das wurde auch durch die Berichten der Versuchspersonen selbst bestätigt. Es scheint vielmehr so zu sein, dass es den Sprechern trotz Belastung gelingt, die Qualität der Sprache weitgehend aufrechtzuerhalten. Beim Entwurf des Experimentes wurde bewusst auf eine standardisierte Nebenaufgabe verzichtet, da die Auswirkungen der Belastung in einem möglichst realistischen Szenario untersucht werden sollten. Die Ergebnisse legen den Schluß nahe, dass die Sprecher auch in einer realen Belastungssituation die Sprachproduktion weitgehend aufrechterhalten können.

3.6 Bewertung der Ergebnisse

Wie erwartet, konnte insgesamt festgestellt werden, dass die untersuchten abhängigen Variablen statisch signifikante Effekte von Zeitdruck und/oder kognitiver Belastung zeigten. Bezüglich der Fragestellungen dieses Experimentes, die in Abschnitt 3 diskutiert wurden, können daraus folgende Schlußfolgerungen gezogen werden (die Fragestellungen werden hier nochmals aufgeführt):

Wie sind die Auswirkungen von Zeitdruck auf die sprachlichen Phänomene, die bislang zumeist nur im Zusammenhang mit Arbeitsgedächtnisbelastung untersucht wurden?

Neben der kognitiven Belastung hat Zeitdruck einen deutlichen Einfluß auf die Sprache, was sich an einer Reihe von Effekten zeigen lässt. Wir konnten feststellen, dass die Sprecher dazu neigen, schneller zu artikulieren. Diese Tatsache ist keineswegs trivial, zumal sie in dieser expliziten Art bislang nicht nachgewiesen wurde. Darüberhinaus beginnen Sprecher unter Zeitdruck schneller mit der Äußerung und machen weniger (stille) Pausen, was darauf schließen lässt, dass sie weniger Zeit mit der Planung verbringen. Dafür nehmen sie in Kauf, dass die Qualität der Äußerung zurückgeht, zu erkennen an der geringeren durchschnittlichen Bewertung der Äußerungen, die unter Zeitdruck gemacht wurden. Die formale Qualität leidet ebenfalls, was an der Oberfläche vor allem an dem verstärkten Auftreten von Qualitätssymptomen zu erkennen ist.

Können die Ergebnisse aus den bisherigen Studien repliziert werden, wenn die Symptome alle gleichzeitig in einem Experiment untersucht werden?

Durch die Ergebnisse dieses Experimentes konnten wir die Korrelation zwischen einer großen Anzahl von abhängigen Variablen feststellen und die relative Stärke der Effekte einschätzen. Auf Basis dieser Daten ist es möglich, Aussagen über eine mögliche Erkennung von Arbeitsgedächtnisbelastung und Zeitdruck zu machen. Abschnitt 4 wird sich mit der Frage beschäftigen, wie ein System aussehen könnte, das eine solche Erkennungsleistung erbringt.

Können die Ergebnisse aus den bisherigen Studien repliziert werden, wenn die Sprachaufgabe und die Art der Belastung möglichst realistisch in Bezug auf das Flughafenszenario gestaltet werden?

Die bisher bekannten Effekte wurden in dieser Domäne weitgehend repliziert. Durch den möglichst realistischen Entwurf von Haupt- und Nebenaufgabe ist ein hoher Grad von Anwendungsrelevanz gewährleistet.

Inwieweit existiert das von Berthold und Jameson (1999) postulierte Speed-Accuracy-Tradeoff in der Sprachproduktion?

Wie bereits in obiger Diskussion über den Einfluß von Zeitdruck angedeutet, tendieren

Sprecher dazu, bei Erhöhung der Geschwindigkeit der Sprachproduktion, Äußerungen von geringerer formaler und inhaltlicher Qualität zu produzieren. Umgekehrt ist bei einem Schwerpunkt auf die Qualität der Äußerung ein Rückgang der Äußerungsgeschwindigkeit zu beobachten. Dies entspricht der Idee eines Speed-Accuracy-Tradeoffs in der Sprachproduktion.

Kapitel 4

Modellierung

Im vorangehenden Kapitel wurde von einem Experiment berichtet, dem sogenannten *Flughafenexperiment*, bei dem die Auswirkungen von kognitiver Belastung und Zeitdruck auf die Sprachproduktion untersucht werden sollten. Dabei ging es unter anderem um die Frage, ob es möglich ist, anhand der Beobachtung bestimmter sprachlicher Phänomene, eine eventuelle Belastung bzw. einen eventuellen Zeitdruck des Sprechers zu erkennen. Diese Phänomene werden in der Domäne von Sprachverarbeitungssystemen häufig als „Disfluencies“ bezeichnet, weil sie als störende Unterbrechungen des Sprachflusses bei einer automatischen Verarbeitung zu zusätzlichen Problemen führen. Im Zusammenhang der Behandlung derselben als Anzeichen für eine Belastung des Sprechers, wurde von Berthold (1998) der Begriff des *Symptoms* eingeführt. Auf Basis dessen vergleichender Literaturstudie sowie den Ergebnissen anderer Untersuchungen wurde ein Katalog von Symptomen aufgestellt, deren Korrelation mit kognitiver Belastung und Zeitdruck im Flughafenexperiment explizit untersucht wurde.

Die Ergebnisse wurden im Projekt READY bereits für mehrere Analysen verwendet, in denen Bayssche Netze gelernt wurden. Zwei dieser Analysen betreffen eher methodische Fragen; diese werden hier nur kurz zusammengefasst. Der nächste Abschnitt geht dann auf die inhaltliche Fragestellung, wie gut Ressourcenbeschränkungen aufgrund von Merkmalen der Sprache erkannt werden können.

4.1 Methodische Analysen

Lernen theoretisch interpretierbarer Baysschen Netze. Wittig und Jameson (2000b) entwickelte eine Methode, um Bayssche Netze mit verborgenen Variablen so zu lernen, dass die quantitativen Beziehungen zwischen den verborgenen und den beobachteten Variablen theoretisch interpretierbar sind. Im ursprünglichen Artikel wurde nur über Tests dieser Methode anhand synthetischer Daten berichtet. Seitdem haben die Autoren die Methode auch anhand der empirischen Daten des Flughafenexperiments (sowie eines anderen Experiments) erfolgreich getestet (s. Wittig & Jameson, 2000a).

Vergleich verschiedener Methoden für die Anpassung allgemeiner Benutzermodelle an einen individuellen Benutzer. Wittig und Jameson (2000a) gehen außerdem auf die folgende Frage ein: Wenn ein System anhand von Daten vieler Benutzer ein allgemeines Benutzermodell (in Form eines Baysschen Netzes) gelernt hat, wie kann es dieses Modell auf einen einzelnen Benutzer zuschneiden, während es Daten über diesen Benutzer sam-

melt und analysiert? Die Autoren testen vier Methoden empirisch, anhand der Daten des vorliegenden und eines anderen Experiments. Die Vielzahl und die Unterschiedlichkeit der abhängigen Variablen im vorliegenden Experiment ermöglichten einen differenzierten Vergleich der Methoden.

4.2 Erkennung von Ressourcenbeschränkungen aufgrund von Merkmalen der Sprache

Um ein System zu entwickeln, das in der Lage ist, die Ressourcenbeschränkungen des Benutzers im Sinne kognitiver Belastung und Zeitdruck zu erkennen, ist es zunächst notwendig, ein Modell zu entwerfen, das die Verbindung zwischen den Symptomen in der Sprache und den relevanten Ressourcenbeschränkungen herstellt (s. Müller, Großmann-Hutter, Jameson, Rummer & Wittig, 2001). Dies stellt ein Problem des maschinellen Lernens dar, wobei die experimentellen Daten als Trainingsdaten betrachtet werden können. Innerhalb des Projektes READY erfolgt eine solche Modellierung mithilfe von Baysschen Netzen als zentraler Inferenzmechanismus. Abschnitt 4.2.1 wird sich deshalb zunächst kurz mit der Technik der Baysschen Netze auseinandersetzen. In Abschnitt 4.2.3 wird dargelegt, welche Techniken des Lernens von Baysschen Netzen genutzt werden können, um ein Modell zur Erkennung von Ressourcenbeschränkungen aufzubauen. Ein solches Modell kann bei der Interaktion mit dem Benutzer angewandt werden, wobei die Merkmale seiner Sprache als Evidenz dienen. Abschnitt 4.2.4 untersucht die Leistungen des hier beschriebenen Modells bei der Erkennung unter Verwendung verschiedener Parameter.

4.2.1 Bayssche Netze

Ein Bayssches Netz (BN) (s. Pearl, 1991) besteht aus zwei Komponenten. Der erste ist ein gerichteter azyklischer Graph der die kausalen Zusammenhänge darstellt, die in der entsprechenden Domäne gelten. Die Knoten repräsentieren beliebige Variablen. Die Verbindungen zwischen den Knoten werden interpretiert als kausale Einflüsse zwischen den Variablen. Die zweite Komponente eines BN ist eine Menge von bedingten Wahrscheinlichkeitstabellen (engl. *conditional probability tables*, CPTs). Die CPT eines Knoten repräsentiert die (unsicheren) Beziehungen zwischen diesem Knoten und seinen Eltern. Sie besteht aus bedingten Wahrscheinlichkeiten für jeden Zustand des Knoten abhängig von der Konfiguration der Zustände der Eltern. Ein BN repräsentiert eine gemeinsame Wahrscheinlichkeitsverteilung über die Zustände seiner Variablen.

Ein Vorteil von BNs ist die Möglichkeit der Entwicklung eines expliziten, theoretisch interpretierbaren Modells der relevanten kausalen Zusammenhänge. Am besten theoretisch interpretierbar sind BNs dann, wenn sie erklärende versteckte Variablen enthalten. Das BN aus Abbildung 3.1 stellt ein solches Netz dar. Von theoretischem Interesse ist in diesem Fall besonders die versteckte Variable TATSÄCHLICHE ARBEITSGEDÄCHTNISBELASTUNG. Diese Art von Knoten entsprechen einer Theorie, wie Ressourcenbeschränkungen die Symptome in der Sprache beeinflussen. Im Rahmen dieser Arbeit wird eine theoretisch neutralere Struktur zugrundegelegt, die nur Variablen enthält, die im Experiment voll be-

obachtbar waren ¹

4.2.2 Die Struktur des Baysschen Netzes

In Abbildung 4.1 wird die Struktur des Baysschen Netzes dargestellt, das auf der Basis der experimentellen Daten gelernt werden soll. Die beiden Knoten NEBENAUFGABE und ZEITDRUCK auf der linken Seite entsprechen den unabhängigen Variablen aus dem Experiment. Die fünf oberen Variablen auf der rechten Seite in ZEITSCHLEIBE 1 entsprechen einzelnen Symptomen. Aus der Kategorie der Geschwindigkeitssymptome wurden die stillen Pausen und die gefüllten Pausen ausgewählt. Die Anzahl der Silben repräsentiert die Kategorie der Längensymptome. Die Artikulationsgeschwindigkeit sowie die inhaltliche Qualität wurden als einzelne Symptome ebenfalls ausgewählt. Als Vertreter der Kategorie der Qualitätssymptome wurde die Variable FORMALE QUALITÄT eingeführt. Sie stellt die logische Disjunktion von einzelnen binären Variablen dar, die jeweils das Vorkommen eines der folgenden Qualitätssymptome anzeigen: inhaltliche und syntaktische Selbstkorrekturen, Fehlansätze sowie Satz- und Wortabbrüche. Diese Variablen wurden deshalb in einer Disjunktion zusammengefasst, weil sie einzeln selten vorkommen. Die Auswahl ist repräsentativ für die Gesamtmenge der Symptome. Die sechs Knoten auf der rechten Seite entsprechen den individuellen *Baselines* für die sechs Symptomvariablen. Sie wurden eingeführt, um die individuellen Unterschiede in der Ausprägung der Symptomvariablen einfließen zu lassen. Um z.B. Aussagen über den aktuellen Zustand einer Person auf Basis ihrer derzeitigen Artikulationsgeschwindigkeit zu machen, ist es sinnvoll zu wissen, ob sie im Allgemeinen eher schnell oder langsam spricht. In diesem BN ist der Wert dieser Variablen eine Konstante für jeden Benutzer. Sie stellt den Durchschnittswert der entsprechenden Variable über alle Messungen im Experiment dar. Der letzte Knoten in dem dargestellten BN, ÄUSSERUNGSKOMPLEXITÄT, entspricht der geschätzten Komplexität einer Äußerung auf Basis des Stimulus-Bildes (s. 3.3.3.2).

Es gilt zu überprüfen, wie gut ein System \mathcal{S} eine mögliche Ressourcenbeschränkung eines Sprechers erkennen kann. Das heisst, wenn ein Sprecher eine Folge von Äußerungen in einer bestimmten experimentellen Bedingung produziert, kann \mathcal{S} die Bedingung erkennen? Die Variablen NEBENAUFGABE und ZEITDRUCK können also als statische Variablen betrachtet werden, die über die gesamte Folge von Äußerungen gleich bleiben. Ebenfalls gleich bleiben die Baseline-Variablen. Die Variablen innerhalb der Rechtecke ZEITSCHLEIBE 1 bzw. ZEITSCHLEIBE 2 sind temporäre Variablen und entsprechen den Eigenschaften einer einzelnen Äußerung. Für jede Äußerung muß eine solche Zeitscheibe erzeugt werden. Bayssche Netze wie dieses, die sich über mehrere Zeitscheiben erstrecken, werden als *dynamische Bayssche Netze* (DBN) bezeichnet (s. Schäfer & Weyrath, 1997).

4.2.3 Lernen eines Bayssches Netzes

Das gelernte BN sollte anhand von Daten einer bestimmten Versuchsperson getestet werden. Es war deshalb darauf zu achten, dass diese nicht bereits in der Trainingsphase verwendet wurde. Aus diesem Grund wurden für jede Versuchsperson ein eigenes Bayssches Netz gelernt, jeweils mit den Daten der übrigen 31 Versuchspersonen. Die gelernten Baysschen Netze hatten die Struktur, wie sie in Abbildung 4.1 dargestellt wird, abzüglich

¹Zur Zeit wird im Projekt READY mit verschiedenen theoretisch interpretierbaren Modellen gelernt und verglichen (s. Wittig & Jameson, 2000a)

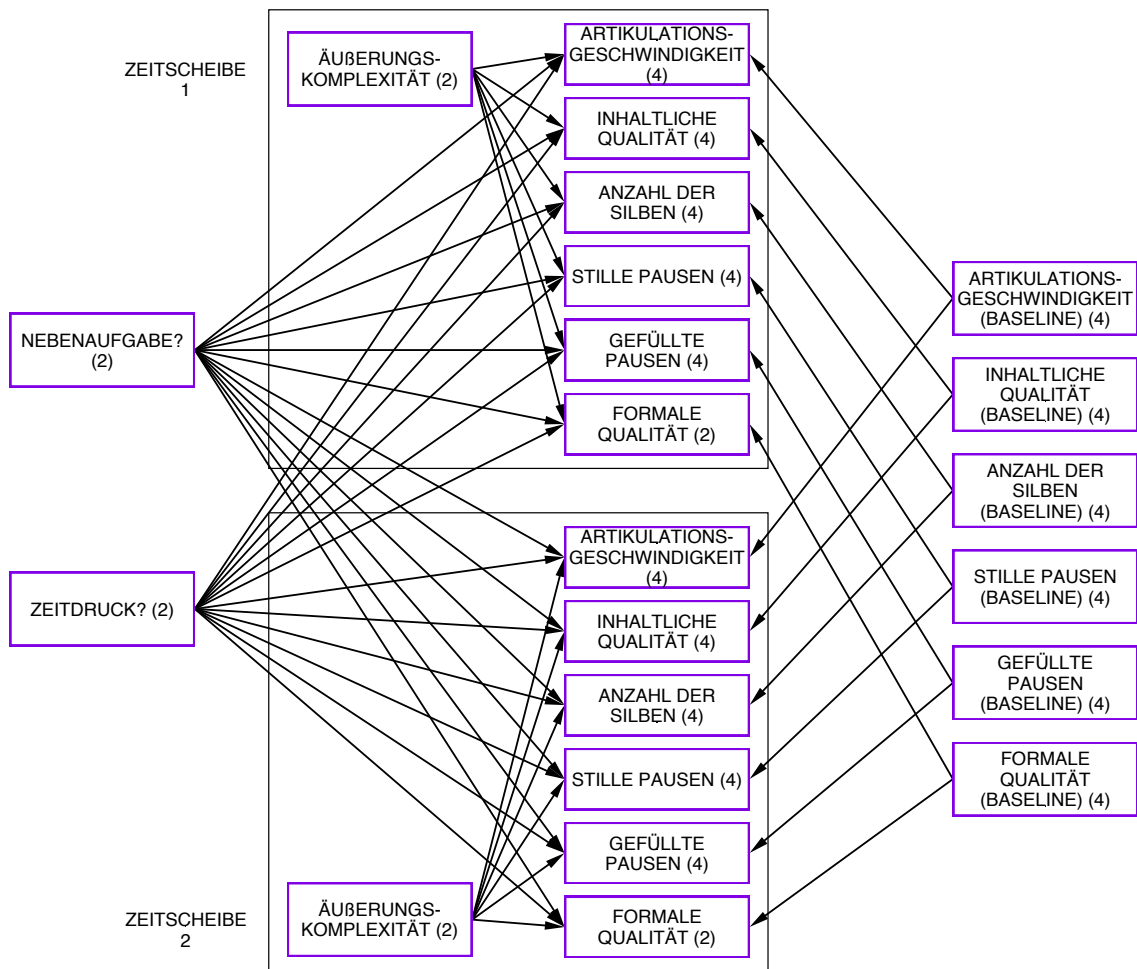


Abbildung 4.1: Struktur des Baysschen Netzes, das auf Basis der experimentellen Daten gelernt werden soll

der Knoten für ZEITSCHEIBE 2. Die CPTs für die temporären Variablen in jeder Zeitscheibe sind die gleiche wie sie in ZEITSCHEIBE 1 gelernt wurden.

Die Lernmethode, die hier implementiert wurde, entspricht der „Maximalen-Wahrscheinlichkeitsmethode“ für das Lernen von voll beobachtbaren Baysschen Netzen (s. z.B. Buntine, 1996). Dabei wird die Schätzung jeder (bedingten) Wahrscheinlichkeit aus der (relativen) Häufigkeit in den Daten berechnet.

4.2.4 Evaluation

4.2.4.1 Verfahren

Zur Evaluation des gelernten Bayssches Netzes wurde die folgende Prozedur zugrundegelegt:

Relevante Variablen und ihre Werte

- Ein Benutzer \mathcal{B}
- Werte z und n der Bool'schen Variablen Z (ZEITDRUCK?) und N (NEBENAUFGABE?)

Aufgabe

- Inferiere die Werte von Z und N auf der Basis von Symptomen in der Sprache von \mathcal{B}

Vorbereitung der Testdaten

- Wähle 20 Beobachtungen für \mathcal{B} mit $T = t$ und $N = n$, in der Reihenfolge, in der sie im Experiment auftraten

Evaluation der Erkennungsgenauigkeit

Initialisierung des Modells

1. Erstelle die erste Zeitscheibe des BN für \mathcal{B} .
2. Instanziiere alle individuellen Baseline-Variablen mit den richtigen Werten für \mathcal{B} . Lasse die Variablen T und N uninstanziiert.

Für jede Evidenz E aus der Menge der Evidenzen für \mathcal{B}

1. In der neuesten Zeitscheibe des BN, leite Annahmen über Z und N ab
 - Instanziiere alle temporären Variablen für diese Zeitscheibe mit ihren Werten in E
 - Evaluere das BN, um die Wahrscheinlichkeiten von Z und N zu erhalten
 - Speichere die Wahrscheinlichkeiten, die zu diesem Zeitpunkt mit den Werten von Z und N verbunden sind
2. Füge eine neue Zeitscheibe zu dem dynamischen BN hinzu für die nächste Evidenz

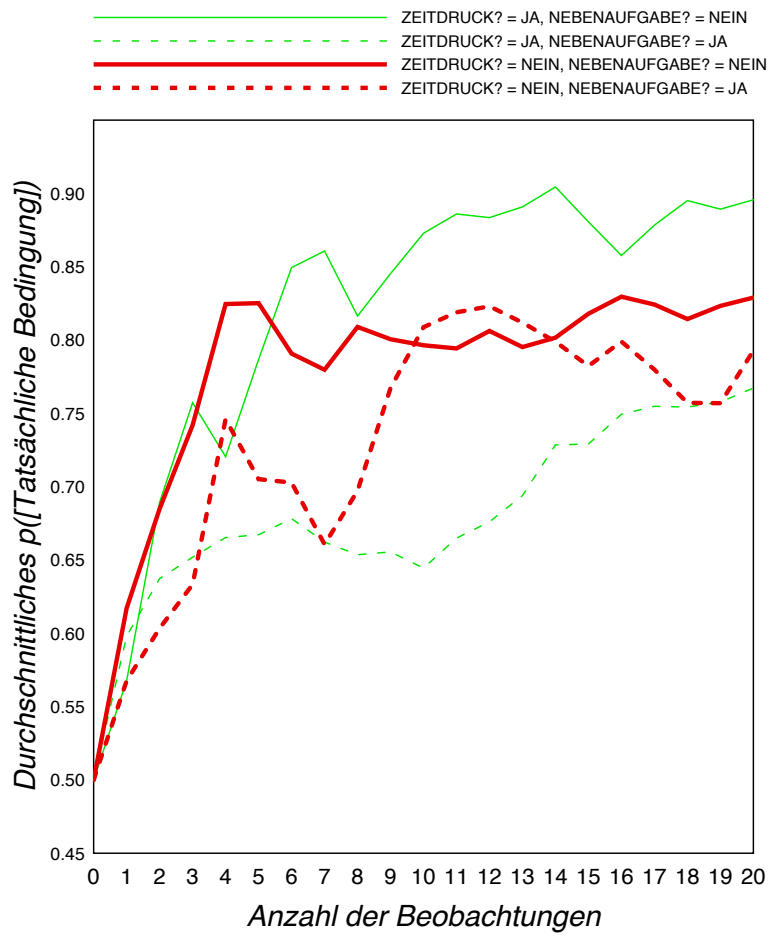


Abbildung 4.2: Genauigkeit der gelernten Baysschen Netze bei der Inferenz des korrekten Wertes für ZEITDRUCK

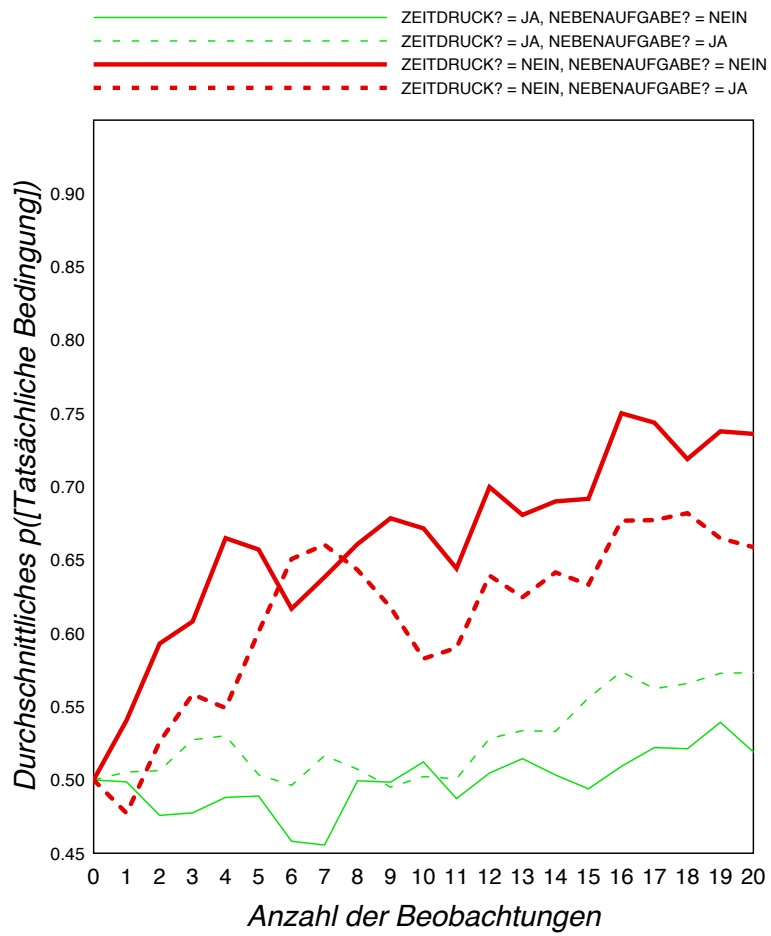


Abbildung 4.3: Genauigkeit der gelernten Baysschen Netze bei der Inferenz des korrekten Wertes für NEBENAUFGABE

4.2.4.2 Ergebnisse

In Abbildung 4.2 und 4.3 werden die Ergebnisse der Evaluation dargestellt, die zuvor über alle 32 Versuchspersonen aggregiert wurden.

Zeitdruck In Abbildung 4.2 werden die Ergebnisse für die Erkennung von Zeitdruck dargestellt. Die durchschnittliche Wahrscheinlichkeit, die der Bedingung zugewiesen wird, die zu diesem Zeitpunkt tatsächlich gegeben war, steigt während der ersten Beobachtungen stark an. Dabei ist die Erkennung von Zeitdruck einfacher, wenn es sich um eine Bedingung ohne Nebenaufgabe handelt. Dieses Ergebnis ist plausibel, wenn man die insgesamt zu beobachtende Tendenz bedenkt, dass die Effekte von Zeitdruck ohne Nebenaufgabe am deutlichsten waren (s. z.B. Abbildung 3.13). Eine mögliche Erklärung war, dass die Partnerbezogenheit durch die zusätzliche Belastung durch die Nebenaufgabe reduziert wurde ².

Nebenaufgabe Die Erkennung der Belastung, d.h. der Bedingung mit Nebenaufgabe, ist deutlich weniger erfolgreich: die höchste Kurve in Abbildung 4.3 ist während der letzten 10 Beobachtungen signifikant geringer als die niedrigste Kurve auf der linken Seite ($p < 0.01$). In Betrachtung der Hauptergebnisse des Experimentes ist auch dieser Befund plausibel: die Effekte von Zeitdruck waren insgesamt substantieller als die Effekte von Belastung (s. Abschnitt 3.4). Besonders in der Zeitdruck-Bedingung sind die gemessenen Effekte von Belastung gering. Jedoch auch ohne Zeitdruck ist der Qualität Erkennung von Belastung lediglich auf einem mittleren Niveau: nach ca. 5 Beobachtungen, wählt das System eine Wahrscheinlichkeit von 0.60 bis 0.70 für die korrekte Hypothese.

Wie bereits in Abschnitt 3.5 diskutiert wurde, konnten die Versuchspersonen trotz Belastung die Leistungen in der Hauptaufgabe weitestgehend aufrechterhalten. Dies spiegelt sich auch in der Erkennung des Belastungszustandes wider. Wenn es jedoch wie hier der Fall ist, dass die Belastung einen relativ geringen Einfluß auf die Ausführung einer anderen Aufgabe hat, ist die Erkennung derselben von weniger großer Bedeutung. Diese Frage wird derzeit noch in READY untersucht.

²Der Zeitdruck wurde im Experiment dadurch erzeugt, dass ein Partner beschrieben wurde, der über wenig Zeit verfügt. In der Bedingung ohne Zeitdruck handelte es sich dagegen um einen Partner mit mehr Zeit (s. Abschnitt 3.3.2.9)

Kapitel 5

Bewertung und Ausblick

5.1 Bewertung

Mit dieser Arbeit sollte ein Beitrag zu der interdisziplinären Arbeit des Sonderforschungsbereichs 378 geleistet werden, was ein wesentliches Ziel desselben ist. Der nachfolgende Abschnitt stellt eine Zusammenfassung der Arbeit dar, bei der - neben der Bewertung inhaltlicher Aspekte - unter anderem der Frage nachgegangen wird, inwieweit dieses Ziel erreicht werden konnte.

Beim Überblick über den Stand der Forschung in Kapitel 2, wurden zwei Literaturstudien gegenübergestellt, die sich mit - wie wir es nennen - Symptomen kognitiver Belastung und Zeitdruck auseinandersetzen. Den Ergebnissen von Berthold (1998), dessen Arbeit ebenfalls im Rahmen des Projektes READY durchgeführt wurde, konnte durch eine Gegenüberstellung der Ergebnisse von (Oberauer & Hockl, 2000) eine neue Sichtweise hinzugefügt werden. Dadurch konnte das theoretische Verständnis des Zusammenhangs zwischen Belastung bzw. Zeitdruck und den Symptomen an der Oberfläche vertieft und auf dieser Basis die experimentellen Hypothesen formuliert werden. Aus interdisziplinärer Sicht ist diese Arbeit im Überschneidungsbereich von Computerlinguistik und Psychologie anzusiedeln.

In Kapitel 3, das von der eigenen empirischen Arbeit berichtet, wurden zunächst die Ergebnisse einer Feldstudie am Frankfurter Flughafen beschrieben. Diese ist durchgeführt worden, um einen Katalog authentischer Anfragen an einem Informationskiosk am Flughafen zusammenstellen zu können. Der Katalog und die Erfahrungen, die dabei gewonnen werden konnten, haben dazu beigetragen, dass das Experiment dem Anspruch der Realitätsnähe gerecht werden konnte. Die Ergebnisse der Feldstudie sind jedoch nicht nur für das Experiment relevant, sondern können auch für andere Arbeiten innerhalb des Flughafenszenarios verwendet werden.

Wie bereits in Abschnitt 3.6 beschrieben wurde, konnten im Experiment signifikante Haupteffekte von Qualität vs. Geschwindigkeit und Belastung sowie Interaktionen festgestellt werden. Insgesamt war der Einfluß von Qualität vs. Geschwindigkeit größer als der von Belastung. Wir schlossen daraus, dass Sprecher unter den gegebenen Voraussetzungen auch unter Belastung ihre normale Art zu Sprechen weitgehend aufrechterhalten können.

Die Ergebnisse des Experimentes konnten bereits in einer Reihe von Untersuchungen innerhalb des Projektes READY verwendet werden. In Kapitel 4 wurde von einer Untersuchung berichtet, in der auf Basis der experimentellen Daten ein System zur Erkennung

von Arbeitsgedächtnisbelastung und Zeitdruck des Benutzers entwickelt wurde. Die Entwicklung eines solchen Systems auf der Basis von empirischen Daten verbindet die Informatik mit der Psychologie und kann insofern ebenfalls als Beitrag zur interdisziplinären Zusammenarbeit angesehen werden.

5.2 Ausblick

Es bestehen konkrete Pläne, die Daten und die Umgebung des Flughafenexperiments im Sonderforschungsbereich 378 weiter zu verwenden. In Zusammenarbeit mit Prof. Lindenberger, von dessen Arbeiten im Bereich der Altersforschung bereits in Abschnitt 1 berichtet wurde (s. auch Kray & Lindenberger, 2000), ist eine Durchführung des Experiments mit älteren Versuchspersonen geplant. Ein Vergleich der Ergebnisse könnte Aufschluß darüber geben, inwieweit sich unter den gegebenen Voraussetzungen die Auswirkungen der Belastung bei älteren Menschen von denen bei jungen Menschen unterscheidet ¹.

Unabhängig davon ist derzeit geplant, ein weiteres Experiment im Rahmen einer Diplomarbeit im Fachbereich Psychologie durchzuführen, bei der es um „Differenzierte Modelle des Arbeitsgedächtnisses“ geht. Dabei soll ein dritter Faktor eingeführt werden, nämlich eine Ablenkung durch Lautsprecher-Aufrufe.

¹Wie in Abschnitt 3.3.2.1 erwähnt, waren die Versuchspersonen bei diesem Experiment durchweg Studenten der Universität des Saarlandes

Literaturverzeichnis

- Baddeley, A. (1993). Working memory or working attention? In A. Baddeley & L. Weiskrantz (Hrsg.), *Attention: Selection, awareness, and control* (S. 152-170). Oxford: Clarendon Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Barch, D. & Berenbaum, H. (1994). The relationship between information processing and language production. *Journal of Abnormal Psychology, 103*, 241–250.
- Berthold, A. (1998). *Repräsentation und Verarbeitung sprachlicher Indikatoren für kognitive Ressourcenbeschränkungen*. Unveröffentlichte Diplomarbeit, Fachbereich Informatik, Universität des Saarlandes.
- Berthold, A. & Jameson, A. (1999). Interpreting symptoms of cognitive load in speech input. In J. Kay (Hrsg.), *UM99, User modeling: Proceedings of the Seventh International Conference* (S. 235–244). Wien, New York: Springer Wien New York.
- Bock, J. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.
- Bock, J. & Levelt, W. (1994). Language production. Grammatical encoding. In M. Gernsbacher (Hrsg.), *Handbook of psycholinguistics* (S. 945-984). NY: Academic Press, NY.
- Bosshardt, H.-G. (1999). Effects of concurrent mental calculation on stuttering, inhalation, and speech timing. *Journal of Fluency Disorders, 24*, 43–72.
- Bredenkamp, J. & Dilger, S. (1998). Erkundungsexperimente über den Zusammenhang von Versprechern und Arbeitsgedächtnis. *Zeitschrift für Experimentelle Psychologie, 45*, 72–79.
- Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on Knowledge and Data Engineering, 8*, 195–210.
- Craik, F., Govoni, R., Naveh-Benjamin, M. & Anderson, N. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology, 125*, 159–180.
- Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1–18.

- Eklund, R. (1999). A comparative analysis of disfluencies in four Swedish travel dialogue corpora. In *Proceedings of Workshop on Disfluency in Spontaneous Speech*. Berkeley, CA.
- Engle, R., Conway, A., Tuholski, S. & Shisler, R. (1995). A resource of inhibition. *Psychological Science*, 6, 122–125.
- Eysenck, M. W. & Keene, M. T. (1990). *Cognitive psychology. A student's handbook*. Hove - London - Hillsdale: Ablex Pub. Corp. Lawrence Erlbaum Associates.
- Gopher, D. (1994). Dual-task performance. In M. Eysenck (Hrsg.), *The Blackwell dictionary of cognitive psychology* (S. 111-116). Oxford: Blackwell.
- Harley, T. A. (1995). *The psychology of language*. Hove, UK: Erlbaum (UK) Taylor&Francis.
- Herrmann, T. & Grabowski, J. (1994). *Sprechen: Psychologie der Sprachproduktion*. Heidelberg: Spektrum Akademischer Verlag.
- Horton, W. S. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Jou, J. & Harris, R. J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society*, 30, 301–304.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kray, J. & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, 15, 126–147.
- Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- McClelland, J. & Elman, J. (1986). The trace model of speech perception. *Cognitive Psychology*, 18, 1–81.
- Modley, R. (1974). *Handbook of pictorial symbols*. N.Y.: Dover.
- Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R. & Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. In J. Vassileva & P. Gmytrasiewicz (Hrsg.), *UM2001, User modeling: Proceedings of the Eighth International Conference*. Berlin: Springer. (Im Druck)
- Oberauer, K. & Hockl, I. (2000). Sprachproduktion und kognitive Ressourcen. In T. Herrmann & J. Grabowski (Hrsg.), *Enzyklopädie der Psychologie*. Hogrefe. (in press)
- Oviatt, S. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction*, 12, 93–129.
- Pearl, J. (1991). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

- Roßnagel, C. (1995). Kognitive Belastung und Hörerorientierung beim monologischen Instruieren. *Zeitschrift für Experimentelle Psychologie*, 42, 94–110.
- Roßnagel, C. (1996). Kognitive Beanspruchung und Hörerorientierung beim Sprechen: Zur Rolle der Motivation. In R. H. Kluwe & M. May (Hrsg.), *Proceedings der 2. Fachtagung der Gesellschaft für Kognitionswissenschaft* (S. 143–144). Hamburg: Universität der Bundeswehr.
- Rummer, R. (1996). *Kognitive Beanspruchung beim Sprechen*. Weinheim, Germany/: Beltz.
- Schäfer, R. & Weyrath, T. (1997). Assessing temporally variable user properties with dynamic Bayesian networks. In A. Jameson, C. Paris & C. Tasso (Hrsg.), *User modeling: Proceedings of the Sixth International Conference, UM97* (S. 377–388). Wien, New York: Springer Wien New York.
- Schmaucks, D. (1999). *Semiotische Aspekte der Fortbewegung in Großflughäfen: Teil I Raumstruktur und Handlungsphasen* (Tech. Rep.). SFB 378, Universität des Saarlandes.
- Starkweather, W. (1995). A simple theory of stuttering. *Journal of Fluency Disorders*, 20, 91–116.
- Townsend, J. T. & Ashby, F. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Wieland-Eckelmann, R. (1992). *Kognition, Emotion und psychische Beanspruchung*. Göttingen: Hogrefe.
- Wittig, F. & Jameson, A. (2000a). *Exploiting prior empirical and theoretical knowledge in the learning of individual user models*. (Manuscript submitted for publication)
- Wittig, F. & Jameson, A. (2000b). Exploiting qualitative knowledge in the learning of conditional probabilities of Bayesian networks. In C. Boutilier & M. Goldszmidt (Hrsg.), *Uncertainty in artificial intelligence: Proceedings of the sixteenth conference* (S. 644–652). San Francisco: Morgan Kaufmann.