

Ruhr-Universität Bochum

Fakultät für Elektrotechnik und Informationstechnik
Institut für Kommunikationsakustik

Fakultät für Philosophie und Publizistik
Sektion für Publizistik und Kommunikationswissenschaften

Mensch-Maschine Kommunikation

**Eine Rezipientenstudie zur Identifizierung und Bewertung
qualitätsbeeinflussender Faktoren der Mensch-Maschine Kommunikation
am Beispiel von BoRIS, einem automatischen sprachbasierten Restaurant-
Auskunftssystem**

Schriftliche Hausarbeit zur Erlangung des Grades Magister Artium

vorgelegt von:
Andreea Niculescu
Laerholz-Strasse. 84, C110
44801 Bochum

Referenten:
Prof. Dr. Jens Blauert
Prof., Dr. Franz Stuke

2.01.2002

Inhaltsverzeichnis

	Seite
Vorwort	4
1. Einleitung	6
1.1 Ziel	6
1.2 Aufbau der Arbeit	7
2. Anwendungsbereiche der Sprachtechnologie	8
3. Kommunikation	10
3.1 Der Kommunikationsprozess	10
3.2 Der Dialog	12
3.2.1. Theoretische Darstellung einer Dialogstruktur	12
3.2.2 Mensch-Mensch Dialog	13
3.2.3 Mensch-Maschine Dialog.....	14
3.3 Die Akteure.....	15
3.3.1 Darstellung eines natürlichsprachlichen Dialogsystems.....	15
3.4 Elemente einer interaktiven Sprachtheorie.....	18
4. Qualitätsbeurteilung einer Dialogsteuerung	25
4.1 Qualität: Versuch einer Definition	25
4.2 Evaluierungsmethoden	26
4.3 Kommunikationsprinzipien eines kooperativen Mensch-Maschine Dialogs	29
4.3.1 Verletzung der Kommunikationsprinzipien	31
4.4 Gestaltung der Qualitätsfaktoren	46
5. Gestaltung der Experimente	52
5.1 Das NSD-System BoRIS	52
5.1.1 Bedienung des CSLU RAD Programms	53
5.1.2 Der Dialogablauf mit BoRIS	54
5.1.3 Logfile	55
5.2 Wizard-of-Oz Technik	56
5.2.1 Kritische Betrachtung der Wizard-of-Oz Technik	57
5.3 Der Experimentaufbau	57
5.3.1 Versuchslayout	57

5.3.2.	Versuchsablauf	58
5.3.3	Szenarien	59
5.3.3.1	Kommunikatives Verhalten der Versuchspersonen	60
6.	Auswertung der Experimente	62
6.1	Qualitative versus quantitative Befragung	62
6.2	Allgemein.....	63
6.3	Das qualitative Interview	64
6.3.1	Aufbau.....	64
6.3.1.1	Aufbau des ersten Fragebogens	64
6.3.1.2	Aufbau des zweiten Fragebogens	65
6.3.2	Auswertung	67
6.3.2.1	Auswertung des ersten Fragebogens	67
6.3.2.2	Auswertung des zweiten Fragebogens	70
6.4	Entwurf des quantitativen Interviews	80
6.5	Das quantitative Interview	86
6.5.1	Aufbau	87
6.5.1.1	Aufbau des ersten Fragebogens	87
6.5.1.2	Aufbau des zweiten Fragebogens	87
6.5.2	Auswertung	91
6.5.2.1	Auswertung des ersten Fragebogens	91
6.5.2.2	Auswertung des zweiten Fragebogens	98
6.5.2.2.1	Qualitätsbeurteilung des Gesamtsystems	98
6.5.2.2.2	Diagnostikstatements.....	102
6.5.2.2.2.1	Systemkooperativität	111
6.5.2.2.3	Rangordnungsstatements	114
7.	Diskussion der Ergebnisse	116
8.	Ausblick	122
9.	Literatur	124

Vorwort

Die Kommunikation zwischen Menschen zeichnet sich dadurch aus, dass man mit einem Partner Inhalte austauscht um dadurch ein gemeinsames Verständnis herbeizuführen. Wir kommunizieren mit unseren Mitmenschen, indem wir uns an jeweilige Gesprächspartner anpassen und ihnen unsere Vorstellungen mitteilen [Hoepfner, 1999]. Nur durch intensive Kommunikation kann sich Wissen und letztlich auch Kultur entwickeln. Dabei bleibt die Sprache das bevorzugte Kommunikationsmittel mit unserer Umwelt. Die Sprache dient nicht nur als Instrument der Kommunikation, sondern sie ist auch selbst Teil des gemeinsamen Wissens und der Verarbeitung dieses Wissens. Weil Computer mittlerweile zu unserem täglichen Leben gehören, versucht heutzutage die Technik in die Geheimnisse unserer natürlichen Sprachkommunikation einzudringen und damit den Computer zu einem intelligenten Werkzeug zu machen, das viel mehr kann, als lediglich fest vorgegebene Algorithmen auszuführen.

Die Mensch-Maschine Kommunikation ist in den letzten Jahren ein viel diskutiertes Thema geworden, denn sie trat stark in den Vordergrund zusammen mit der weiteren Entwicklung der Informations- und Kommunikationstechnik. Bereits in den 60er Jahren machten sich mit der Entwicklung der EDV Kommunikationsforscher Gedanken darüber, was für eine Art Kommunikation und Interaktion zwischen Menschen und Computern stattfindet, und welche Rolle die Sprache dabei spielt [Ungeheuer, 1972]. Als zentrales Problem stellt sich die Frage, wie sich menschliche und maschinelle Informationsverarbeitung trotz ihrer grundlegenden Verschiedenheit zu einer Art „Dialog“ zusammenführen lassen, um ein Zusammenwirken von menschlichen Fähigkeiten und maschinellen Funktionsmechanismen zu bewirken [Paetau, 1990]. Diese Frage wird gegenwärtig vor allem unter dem Aspekt der Anpassung der Maschine an den Menschen diskutiert. Im Mittelpunkt steht die Grundidee, dass der Mensch sich nicht mehr wie bisher der vorbestimmten Funktionsweise der Maschine zu unterwerfen hat, sondern umgekehrt sich die Maschine in bestimmtem Maße an den Menschen, an seine Handlungs- und Denkgewohnheiten anpassen sollte. In dieser Diskussion spielen Methoden der Künstlichen Intelligenz eine große Rolle. Nach Auffassung der Vertreter dieser computerwissenschaftlichen Forschungsrichtung lassen sich mit KI¹-Methoden in weitaus angemessenerer Weise als mit traditionellen Programmiermethoden die Schwierigkeiten der Mensch-Maschine Kommunikation überwinden. Ihrer Meinung nach sollte der Computer nicht nur als Simulationsinstrument angesehen werden, sondern als kooperativer, dialogfähiger und intelligenter Assistent, der während seiner Interaktion mit einem Menschen programmierte mentale Prozesse durchläuft. Kognitive Leistungen sollen im Computer modelliert und verstanden werden. Das bedeutet schließlich, dass man mit dem Computer in eine Form des Informationstausches treten kann, die durchaus humane

¹KI ist die Abkürzung für Künstliche Intelligenz.

Züge annimmt. Diese Auffassung ist immer noch Gegenstand heftiger Diskussionen, nicht nur wegen des problematischen Analogiebegriffs zwischen Mensch und Computer, sondern auch wegen der umstrittenen Fähigkeit des Computers, als intelligenter Gesprächspartner zu fungieren. Es gibt natürlich viele einfache Fälle, bei denen entweder die Informationsübertragung sehr einseitig ist, oder die Maschine sich auf einem um Größenordnungen niedrigeren Informationsstand befindet. Die Maschine wird dann nur zu einem passiven Glied in der Informationskette. Dazu gehören Systeme, bei denen nur eine Worterkennung und keine sprachverstehenden Komponenten eingesetzt werden. Terminologisch steht noch nicht fest, ob derartige Methoden des Datenaustausches als Kommunikation oder als Interaktion zu bezeichnen sind.

Die zukünftige Entwicklung geht allerdings deutlich in Richtung Kommunikation. Dies geschieht durch eine verstärkte Einbindung höherer semantischer Ebenen in den Informationstausch, was auf der Seite der Maschine wiederum höheres Wissen erfordert. Gleichzeitig werden die Fortschritte bei der Erkennung, beim Verstehen, bei Synthese und Generierung von Sprache immer mehr eine echte, akustische Sprachkommunikation erlauben [Mangold, 1992]. Eine Übersicht über die verschiedenen Arten von Dialogsystemen, die es heutzutage gibt, wird im zweiten Kapitel geliefert.

Aus diesem Grund kann man hier die Schlussfolgerung ziehen: gleichgültig wie wir die Interaktion zwischen Mensch und Computer auffassen wollen, sie bleibt ein Teil der Kommunikationsforschung und wird die in der Zukunft eine Vielzahl neueren Anwendungen erschließen.

1. Einleitung

1.1 Ziel

Obwohl es heutzutage Systeme gibt, die intelligent auf die Fragen antworten können, sind sie meistens nicht in der Lage die gleichen Bedingungen zu schaffen, wie bei einer Unterhaltung zwischen zwei Menschen. Im Mittelpunkt der Überlegung dieser Arbeit steht der Versuch, Lösungen zur Verbesserung der Dialogsteuerung eines NSD-Systems zu finden, die das System als Gesprächspartner „menschlicher“ erscheinen lassen. Ihr konkretes Ziel ist die Identifizierung und Bewertung, aus kommunikationswissenschaftlicher Perspektive, der qualitätsbeeinflussender Faktoren eines Dialogs mit einem NSD-Systems². Die Festlegung der Faktoren soll Allgemeincharakter haben und für künftige Untersuchungen ebenfalls relevant sein. Die Untersuchung soll gleichzeitig eine Rangordnung der Wichtigkeit dieser Faktoren aus Benutzersicht aufstellen.

Die Arbeit hat einen interdisziplinären Charakter: sie interpretiert aus dem Blickwinkel der Kommunikationswissenschaft die Sichtweise und Problemstellungen anderer Wissensgebiete (Sprachwissenschaften, angewandte Linguistik, Psycholinguistik und Ingenieurwissenschaften), die an dem Kommunikationsprozess zwischen Mensch und Maschine beteiligt sind. Dabei ergibt sich ein Gesamtbild, welches sowohl Aussagekraft für die Beurteilung von Systemen aus Benutzersicht (Nutzerqualität) als auch für die Gestaltung von Systemen (Qualitätsgestaltung) hat.

Die Arbeit besteht aus einem theoretischen und einem empirischen Teil. Die Theorie behandelt kommunikations- und sprachwissenschaftlich relevante Ansätze, die den Kommunikationsprozess erklären und Hinweise zur Optimierung eines Dialogablaufes liefern. Dabei wird versucht, ein für den empirischen Teil geeignetes Untersuchungsschema zu entwerfen, welches Aussagekraft für die Qualitätsbeurteilung der Interaktion zwischen Mensch und Maschine hat.

Im Rahmen des empirischen Teils werden Experimente mit Versuchspersonen zur Erkundung der in der Theorie dargestellten Konzepte organisiert. Für eine einheitliche Durchführung der Experimente werden sechs verschiedene Dialogszenarien entwickelt. Die Versuchspersonen werden vor und nach dem Versuch interviewt. Die Interviews werden je nach Befragungsart in zwei Etappen durchgeführt: zuerst werden acht Versuchspersonen nach einer qualitativen Befragungsmethode interviewt. Die Ergebnisse dieses Interviews werden eingesetzt, um ein quantitatives Interview auszuarbeiten, welches mit weiteren 53 Personen durchgeführt wird.

Damit versucht man von den Vorteilen beider Befragungsmethoden zu profitieren und dem Untersuchungsobjekt vielseitig zu analysieren.

² NSD-System ist eine Abkürzung für natürlichsprachliches Dialogsystem (siehe Kapitel 2).

1.2. Aufbau

Das zweite Kapitel bietet einen Überblick über die bestehenden Sprachsysteme und ihre Anwendungsbereiche. Dabei werden die neuesten Sprachtechnologien kurz präsentiert.

Das dritte Kapitel befasst sich mit der Darstellung des Kommunikationsprozesses und des Dialogs als Form der Kommunikation. Untersucht werden die Gemeinsamkeiten und Unterschiede der Mensch-Mensch und Mensch-Maschine Dialog, wobei eine kurze Beschreibung der Funktionsweise eines natürlichsprachlichen Dialogsystems präsentiert wird. Das Kapitel endet mit der Darstellung einer Sprachtheorie, die Struktur, Inhalt und Dynamik einer Mensch-Maschine Kommunikation beschreibt.

Das vierte Kapitel setzt sich mit dem zentralen Forschungsobjekt dieser Arbeit auseinander: die Qualitätsbeurteilung einer Dialogsteuerung. Dabei wird der Versuch unternommen, den Begriff „Qualität“ allgemein und im bezug auf den Dialog zu definieren und dabei passende Untersuchungskriterien zu bestimmen. Zu diesem Zweck werden verschiedene Ansätze zur Evaluierung von NSD-Systemen dargestellt.

Für eine gelungene Kommunikation zwischen Menschen und Maschinen wurden in der Literatur Prinzipien festgelegt. Diese Prinzipien sprechen qualitätsrelevante Dialogaspekte an, die im Rahmen dieser Arbeit als Evaluierungskriterien einer Dialogsteuerung betrachtet werden. Die Kommunikationsprinzipien werden mit konkreten Beispielen aus eigenen Experimenten veranschaulicht werden.

Aufgrund all dieser Ansätze wird versucht, eine Zusammenfassung möglicher für die Qualität der Dialogsteuerung wichtiger Faktoren zu erstellen.

Das fünfte Kapitel beschäftigt sich mit dem Ablauf der Experimente. Es wird zeigt, mit welchen Instrumenten und in welcher Art die Versuche gestaltet wurden. Dabei werden die Funktionsweise des verwendeten NSD-Systems, die angewandte Experimenttechnik, sowie den Experimentaufbau präsentiert. Das Kapitel endet mit einem kurzen Abschnitt, in dem das kommunikative Verhalten und Reaktionen der Versuchspersonen während der Experimente dargestellt wird.

Die Auswertung der Experimente nach qualitativen und quantitativen Befragungsmethoden ist das Thema des sechsten Kapitels. Es wird gezeigt, welche Vorteilen und Nachteile mit diesen Methoden verbunden sind, wie die Fragebögen entstanden sind und zu welche Ergebnissen man bei diesen Befragungen gekommen ist.

Im siebten Kapitel werden die Ergebnisse der Untersuchung zusammenfassen kommentiert und Vorschläge zur Systemverbesserung gemacht.

Die Arbeit schließt mit einem Ausblick über die erreichte Ziele dieser Arbeit.

Im Anhang der Magisterarbeit befinden sich, die Szenarien, die Transkription der qualitativen Interviews, die Fragebögen sowie die statistische Auswertung der quantitativen Fragebogen.

2. Anwendungsbereiche der Sprachtechnologie

Heute befinden wir uns mitten in einer Revolution der Kommunikationswelt, die es uns ermöglicht, Systeme zu erstellen, in denen Sprachcomputer intelligent auf sprachliche Befehle reagieren können. So z. B. sind Diktier- und Kommandosysteme, Voice Mails und Voice SMS Systeme schon auf dem Markt erhältlich. Diese Systeme haben jedoch nicht die Fähigkeit, die Spracheinheiten zu analysieren, ihren Sinn zu interpretieren und aufgrund dieser Interpretation zu reagieren. Ihre Aufgabe beschränkt sich auf das Umsetzen des akustischen Signals durch die Spracherkennung in ein anderes Medium. Als ein Beispiel ist hier Spechtel von VOCALIS zu nennen, ein Voice-calling System, das bei manchen Handy-Modellen eingebaut wird und dem Benutzer sprachgesteuerte Wahl beim Telefonieren ermöglicht.

Die maschinelle Übersetzung erfuhr in der Zeit modernen Sprachtechnologien eine neue Dimension: die Sprache fungiert als Mittel im Dialog zwischen Menschen. Das prominenteste Beispiel ist im deutschsprachigen Raum das Verbundprojekt VERBMOBIL [Wahlster, 1993]. Dabei wurde ein System entwickelt, das zwei englischsprechenden Gesprächspartnern erlaubt, sich in unterschiedlichen Muttersprachen (Deutsch und Japanisch) am Telefon zu unterhalten. Ein solches System ist hilfreich in Situationen, in denen einer der Gesprächspartner nicht weiß, wie er eine Äußerung auf Englisch ausdrücken kann. Er kann das System aktivieren und die Äußerung in seiner Muttersprache sprechen. Das System analysiert die Äußerungen und übersetzt sie ins Englische. Dabei muss das System fähig sein, nicht nur akustische Eingaben zu erkennen, sondern sie auch im Kontext des Gesprächs zu interpretieren, so daß die Übersetzung sich nicht von ihrem ursprünglichen Sinn entfernt.

Einen weiteren technologischen Fortschritt stellen dabei die NSD-Systeme dar. In Forschungslaboratorien wurden in den letzten Jahren Systeme entwickelt, die über die reine Erkennung akustischer Eingaben und ihre kontextuelle Interpretation hinausgehen und dem Benutzer erlauben, in natürlicher Sprache mit einem Computer komplex zu interagieren. Solche Dialogsysteme ermöglichen die Formulierung einer konkreten Problemstellung in natürlicher Sprache. Darauf bietet das System Lösungen an, die wiederum in natürlicher Sprache formuliert sind. Diese Problemstellungen beziehen sich dann immer auf ein bestimmtes Themengebiet, z.B. Fahrpläne, Wetter, Bankkonten, Kinokartenbestellung usw.

Man unterscheidet aufgrund unterschiedlicher Dialogfähigkeit zwischen mehreren Arten von Dialogsystemen. Die einfachsten Dialogsysteme sind die Frage-Antwort Systeme, die schon zu Beginn der '60er Jahre gebaut wurden und nur in der Lage waren, einfache Fragen zu beantworten, wie z.B. Spieldaten, Spielstätte oder Namen von Spielmannschaften. Man hat versucht, solche Systeme zu erweitern, indem man sie mit Datenbankzugriff ausgestattet hat. Dadurch konnten sie auf die Datenbank beliebig

zugreifen und gleichzeitig Querverweise erstellen. Diese Art von Systemen werden von Benutzer gesteuert und verhalten sich passiv.

Die interaktiven Sprachreaktionssysteme bilden eine andere Kategorie von Sprachdialogsystemen, bei denen das System aktiv mit dem Benutzer arbeitet. Solche Systeme werden z.B. von Telefongesellschaften angeboten, wobei das System ein Menü anbietet und der Benutzer eine Option über die Telefontastatur auswählt (z. B. „für nein drücken Sie die 1; für ja drücken Sie die 2; um mit unserem Kundenservice verbunden zu werden, drücken Sie die 3“ usw.). Diese Systeme hat man heute modernisiert und die Betätigung der Telefontasten wurde durch die Spracheingabe ersetzt.

Die sprachgesteuerten Dialogsysteme gehören zu der dritten Kategorie. Solche Systeme verhalten sich aktiv und führen den Benutzer ans Ziel. Sie sind fähig Missverständnisse zu beseitigen, indem sie Fehler erkennen und korrigieren können. Als Beispiel soll hier ist das Bahnauskunftssystem von Philips genannt werden, das seit Februar 1994 aus dem Festnetz der Deutschen Telekom unter der Telefonnummer 0241/604020 zugänglich ist, oder das ATIS System [Zue et al., 1992], das Flugreservierungen anbietet. Andere nennenswerte Projekte sind das dänische Dialogprojekt, das ein System für Flugreservierung in dänischer Sprache darstellt [Bernsen et al., 1998], sowie das SUNDIAL-Projekt, eines der größten und bekanntesten Projekte in Europa auf diesem Gebiet; im Rahmen dieses Projekts wurde ein natürlichsprachiges Dialogsystem in vier verschiedenen Sprachen für Informationen und Flugreservierungen (England und Frankreich) und Bahnauskünfte (Deutschland und Italien) entwickelt.

Die vierte und letzte Kategorie von NSD-Systemen bilden die Konversationssysteme. Die Entwicklung solcher Systeme liegt noch in der Zukunft denn dafür benötigt man künstliche Intelligenz, die bis heute noch nicht zur Verfügung steht.

3. Kommunikation

3.1 Der Kommunikationsprozess

Kommunikation wird als ein Vorgang bezeichnet, der auf bestimmten Gemeinsamkeiten zwischen verschiedenen Subjekten beruht. Das Wort stammt aus dem Lateinischen *communis* und bedeutet so viel wie gemeinsam. Gemeinsam haben die am Gespräch beteiligten Partner eine physische oder energetische Verbindung zur Verarbeitung von Signalen. Zweitens besteht zwischen den Gesprächspartnern eine Beziehung, die durch Erwartungen gekennzeichnet ist; aus dieser Beziehung entsteht dann die Information. Drittens sind ihnen bestimmte übereinstimmende Kognitionen, d.h. Wissen, Erfahrung, Bewertungen usw. gemeinsam, aus denen sich die Erwartungen ableiten und die den Signalen Bedeutungen verleihen. Viertens verbinden die beteiligten Gesprächspartner bestimmte Absichten, die im Laufe ihrer Interaktion verfolgt werden.

Kommunikation wird oft mit Begriffen wie Verständigung, Mitteilung, Interaktion, Übertragung umschrieben, was den dynamischen Charakter der Kommunikation unterstreicht: die Kommunikation ist ein Prozess, der aus mehreren Bestandteilen zusammengesetzt wird.

Diese Betrachtung des Kommunikationsprozesses wurde das erste Mal von dem amerikanischen Soziologe Harold D. Lasswell in seiner berühmten Lasswell-Formel zum Ausdruck gebracht: „*A convenient way to describe an act of communication is to answer the following question: Who says what in which channel to whom with what effect?*“ [Lasswell, 1948]. Anhand dieser Elemente des Kommunikationsprozesses werden die wichtigsten Forschungsfelder der Kommunikationswissenschaft identifiziert:

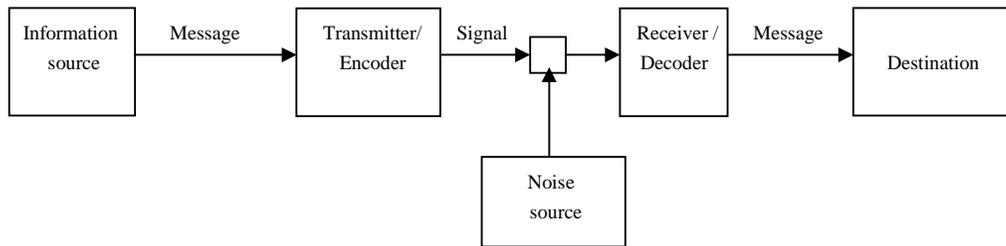
Tab.1 Die Forschungsfelder des Kommunikationsprozesses³

Who	says what	in which channel	to whom	with what effect
Kommunikator	Mitteilung	Medium	Rezipient	Wirkung
Kommunikator-Forschung	Inhaltsanalyse	Medienkunde	Publikumsforschung	Wirkungsforschung

Eine Vertiefung dieses Schemas bringt Shannon [Shannon und Weaver, 1948] mit seiner mathematischen Informationstheorie. Dabei differenziert er auf der Kommunikatorseite zwischen Informationsquelle und dem Sender sowie auf Rezipientenseite zwischen Adressat und Empfänger. Diese Betrachtung ist an technisch vermittelter Kommunikation orientiert, wie z. B. an einem Telefongespräch. Dabei muss die Mitteilung zum Zwecke der Übertragung in Signale umgewandelt werden.

³ Siehe [Fischer Lexikon der Publizistik, 1996].

Abb. 1 Dynamik eines Kommunikationsprozesses ⁴

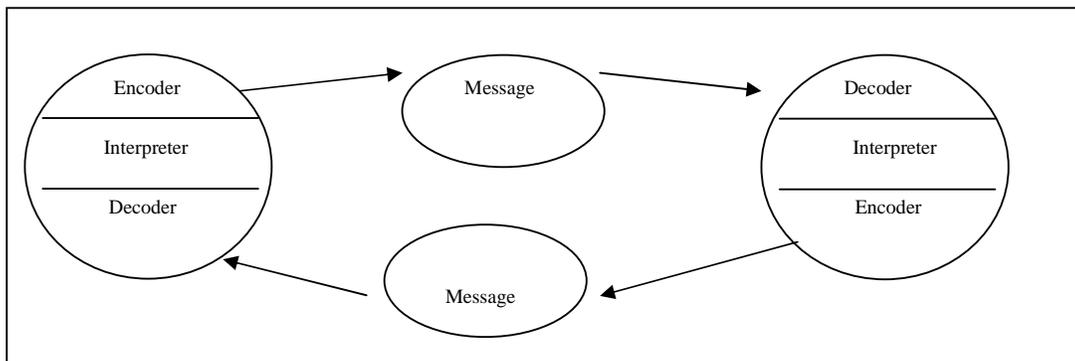


Die Umwandlung der Mitteilung in Signale auf Seiten des Kommunikators wird **Enkodierung** genannt und die Rückübersetzung der empfangenen Signale auf Seiten der Rezipienten **Dekodierung**. Sowohl die Enkodierung als auch die Dekodierung sind konstitutive Merkmale eines Kommunikationsprozesses, nicht nur bei technisch gestützten Kommunikationsprozessen, die sich der Telefonleitung als Übertragungskanal bedienen, sondern auch bei interpersonellen Gesprächen. Hier ist der Kommunikator gleichzeitig Quelle und Enkoder. Die Mitteilung wird als akustisches Signal, als Schallschwingung übertragen und wird vom Rezipienten dekodiert; in diesem Fall ist der Rezipient Adressat und Dekoder.

Shannon berücksichtigt in seinem Modell noch ein Element, das er **Störquelle (Noise Source)** nennt. Damit werden alle Arten von Veränderung der Mitteilung von ihrer ursprünglichen Form gemeint, die vor allem bei technisch gestützten Übertragungen auftreten können. Hier kann man zwischen zwei verschiedenen Arten von Störquellen unterscheiden: nämlich zwischen der Veränderung der Mitteilung durch das Hinzufügen irrelevanter Information einerseits (bezeichnet in der Fachsprache als „Äquivokation“) und durch den Verlust von relevanten Informationen verursacht, z.B. durch Rauschen, andererseits.

Bei der interpersonellen Kommunikation übernehmen beide Partner abwechselnd die Rolle des Kommunikators und des Rezipienten. Die Kommunikation hat in einer solchen Situation eine symmetrische, reziproke Struktur, wie in der folgenden Darstellung:

Abb. 2 Reziproke Struktur der Kommunikation ⁵



⁴ Siehe [Shannon und Weaver, 1948].

⁵ Siehe[Fischer Publizistik Lexikon, 1996].

3.2 Der Dialog

Der Dialog wird als Form der interaktionalen Kommunikation beschrieben, bei der thematische oder situativ bestimmte, intentional gesteuerte Äußerungen an einen Partner gerichtet werden [Lewandowski, 1979]. Dabei werden unkonventionelle Inhalte (Gedanken) in einer konventionellen Form (Sprache) getauscht. Der Dialog wird durch das Informationsniveau, die emotionale Aufladung und die Interessen der Dialogpartner beeinflusst und hat als wesentliches Merkmal einzelne aufeinander bezogene Beiträge (Frage und Antwort). Meistens haben Dialoge eine relativ kurze Form und eine syntaktisch einfache Baustruktur.

Damit ein Dialog zustande kommt, müssen einige Voraussetzungen erfüllt werden: ein wichtiges Kriterium ist die Bereitschaft der potentiellen Sprecher Informationen auszutauschen, damit man überhaupt ein Dialog zustande kommt; zweitens benötigt man ein gemeinsames Thema; drittens müssen die Dialogpartner mindestens eine minimale Sprachkompetenz vorweisen und viertens müssen sie sich in einen permanenten Rollentausch verwickeln lassen, damit der Dialog nicht zu einem Monolog wird.

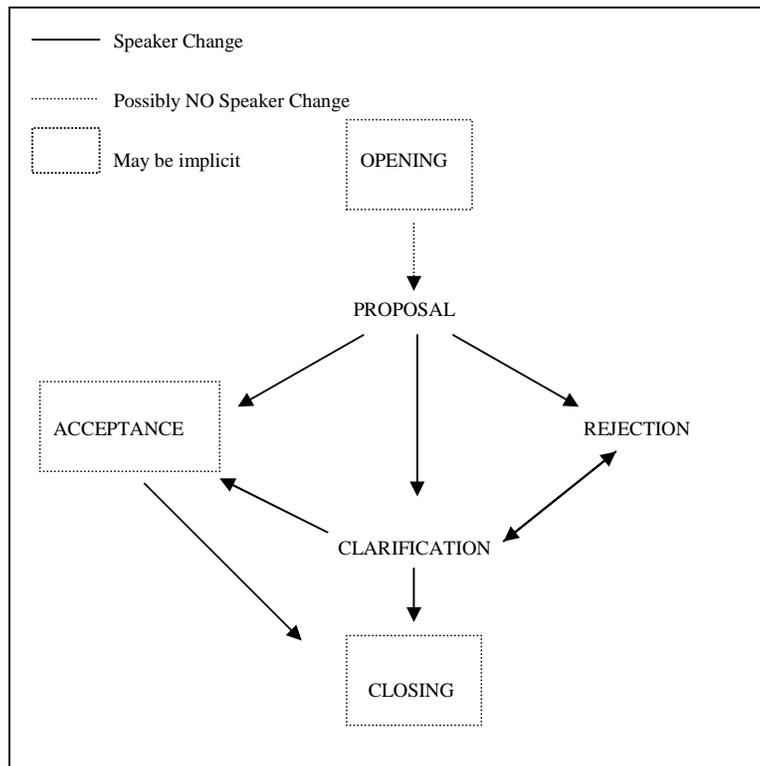
3.2.1 Theoretische Darstellung einer Dialogstruktur

Wenn man davon ausgeht, dass der Dialog eine einzige Aufgabe (*Task*) erfüllt, kann er als Transaktionskette angesehen werden, wobei jede Transaktion einen kleinen Teil der Aufgabe erfüllt, eine sogenannte Subaufgabe (*Subtask*). Jede Subaufgabe kann in kleinere Segmente zerlegt werden. Diese werden Dialogsequenzen genannt. Jede Transaktion enthält darüber hinaus ein oder mehrere komplementäre *Turns* (Frage/Antwortpaare). Ein *Turn* steht für einen illokutiven Akt, der auch als *Speech Act* bezeichnet wird. Er beginnt in dem Moment, wo einer der Dialogpartner anfängt zu reden und endet, wenn der andere Gesprächspartner die Rede übernimmt. In der linguistischen Fachsprache herrscht keine terminologische Übereinstimmung für die Begriffsbezeichnung *Speech Act*. Dieser wird auch als *Dialog Act* [Bunt, 1994], [Stein und Maier, 1994] oder als *Move* [Carletta et al., 1997] bezeichnet. Die Autorin benutzt für ihre Arbeit einheitlich den Begriff *Turn*.

Der Dialog selbst besteht aus mehreren *Turns* und kann bis zum Dialogende verschiedene Zustände durchlaufen. Die Zustände werden wie folgt klassifiziert: *Proposal, Acceptance, Rejection, Clarification, Openings* und *Closing* (Abb. 3, S.13).

Auch die inhaltliche Gestaltung der Dialogzustände variiert. Abhängig ist dies je nach Dialogsstrategie, Dialoginitiative, Informationsmenge, Ausmaß der Ressourcen und Kommunikationsstil.

Abb. 3 Dialogzustände⁶



3.2.2 Mensch-Mensch Dialog

Dialoge bilden die Grundlage menschlicher Kommunikation und erfüllen sowohl eine informelle als auch eine soziale Funktion. Um erfolgreich miteinander kommunizieren zu können, sind gewisse Fertigkeiten der beiden Gesprächspartner gefragt, wie z.B.:

1. Erkennung von spontanen sprachlichen Äußerungen einschließlich ihrer intentionalen Bedeutung, unabhängig von Geschlechts-, Alters-, Dialektunterschieden, Hintergrundgeräuschen oder Signalintensität.
2. Beherrschung eines breiten Wortschatzes aus verschiedenen Bereichen.
3. Weite Sprachkompetenz, die es den Gesprächsteilnehmern ermöglicht, komplexe, prosodische, elliptische oder anaphorische Konstruktionen sowie Unterbrechungen („ahm“, „ehm“), Wortwiederholungen, Eingabekorrekturen innerhalb desselben Sprachinputs („...ich möchte am Samstag,... nein... am Sonntag fahren“) und Hintergrundgeräusche (Husten, Räuspern usw.) in einem Kontext zu verstehen und zu interpretieren.
4. Fähigkeit kohärente Äußerungen zu produzieren.
5. Fähigkeit, semantische Beziehungen herzustellen und das Gesprächsthema in Verbindung mit anderen verwandten Themen zu bringen.

⁶ Siehe [Walker, 1993].

6. Wahrnehmung der Kontextdimensionen, in denen das Gespräch stattfindet, gefolgt von einer entsprechenden stilistischen und semantischen Anpassung des Ausdrucksvermögens an Situation, Thema und Gesprächspartner.
7. Fähigkeiten, Korrekturen und Aufklärungen problemlos durchzuführen (die Durchführung von Korrekturen und Aufklärungen werden in der linguistische Fachsprache Metakommunikation genannt).
8. Fähigkeit, einen Dialog weiter zu führen, auch wenn spontane Reaktionen des Gesprächspartner mitten im Gespräch eintreten.
9. Fähigkeit, die Sprachausgabe, einschließlich Tonfall und Betonung in unterschiedlicher Form zu realisieren [Bernsen et al., 1998].

3.2.3 Mensch-Maschine Dialog

Der Dialog zwischen Mensch und Maschine wird als wechselseitiger Informationsaustausch zur Erfüllung einer Aufgabe verstanden. Aus diesem Grund nennt man das Interaktionsmodell aufgabenorientiert.

Man muss von vornherein klar stellen, dass es im Falle der Mensch-Maschine Kommunikation mit einem asymmetrischen Verhältnis zu tun hat. Auf der einen Seite steht der Mensch, als kognitives Wesen, das sich in sprachlicher Interaktion fließend, mühelos und effizient über fast jedes Thema ausdrücken kann, und auf der anderen Seite die Maschine, als vorgegebene, leistungsbegrenzte Instanz.

Durch den Vergleich der Charakteristika des Mensch-Mensch und des Mensch-Maschine Dialoges wird der Unterschied noch deutlicher. Die menschlichen Fertigkeiten, die auch als Charakteristika der zwischenmenschlichen Sprachinteraktionen fungieren, können nur sehr begrenzt auf die Mensch-Maschine-Kommunikation übertragen werden. Die Kommunikationsfähigkeit des Computers ist begrenzt und die Sprachinteraktionen, die daraus entstehen, haben selten eine soziale Funktion - zumindest nicht für den Computer.

Zuerst ist die Erkennung der kontinuierlichen Sprache für solche Systeme eine große Herausforderung: die Grenzen zwischen einzelnen Wörtern verschwinden teilweise oder sogar vollständig, einmal durch Sprechgeschwindigkeit und einmal durch die sogenannte Koartikulation – dabei können die für die Artikulation oder Betonung erforderlichen Sprechpausen innerhalb eines Wortes größer sein als die Pause zwischen einzelnen Wörtern, oder die Aussprache eines Wortes ist von dem vorangehenden oder nachfolgenden abhängig (z.B. Drucker-Zeugnis oder Druck-Erzeugnis). Ein solches Problem wird verständlicher, insofern auch Menschen damit konfrontiert werden, wenn sie eine unbekannte Fremdsprache hören und versuchen, einzelne Wörter herauszuverstehen.

Zweitens wird der Wortschatz auf ein Minimum reduziert, einerseits um die Chancen der Spracherkennung zu steigern und andererseits, um den Speicherplatz nicht vollkommen zu belegen [Boner, 1994].

Drittens kann es durch den begrenzten Wortschatz schon vorkommen, dass komplexe oder stichwortartige Äußerungen vom Spracherkenner nicht erkannt werden.

Viertens sind laut ihrer Definition aufgabenorientierte Systeme auf Dialoge rund um ein einziges Thema begrenzt und können nur bedingt innerhalb dieses Themas semantische Beziehungen herstellen.

Fünftens können semantisches Ausdrucksvermögen, stilistische Angemessenheit sowie variable Realisierung des Sprachoutputs leider von dem meisten Systemen nicht erwartet werden.

Zusammenfassend ist die Mensch-Maschine Kommunikation auf folgende Charakteristika beschränkt:

1. bedingte Erkennung kontinuierlicher, aufgabenbezogener Sprachäußerungen, abhängig von Aussprachebesonderheiten (undeutlicher Aussprache, Dialektunterschiede) und Hintergrundgeräuschen.
2. Beherrschung eines thematisch begrenzten Wortschatzes.
3. Verfügen über ein rudimentäres syntaktisch-semantisches Parsersystem.
4. Herstellung semantischer Beziehungen nur bedingt durch Querverweise.
5. Möglichkeit, je nach Systemart meta-kommunikative Strategien durchzuführen (Korrekturen, Aufklärungen und Wiederholungen).

Es wird deutlich, dass Dialoge zwischen Menschen und Maschinen andere Regeln befolgen als menschliche Gespräche und dass die Maschine riesige Hürden zu bewältigen hat, um mit einem menschlichen Gesprächspartner konkurrieren zu können. Und teilweise sind diese Hürden unüberwindbar.

3.3 Die Akteure

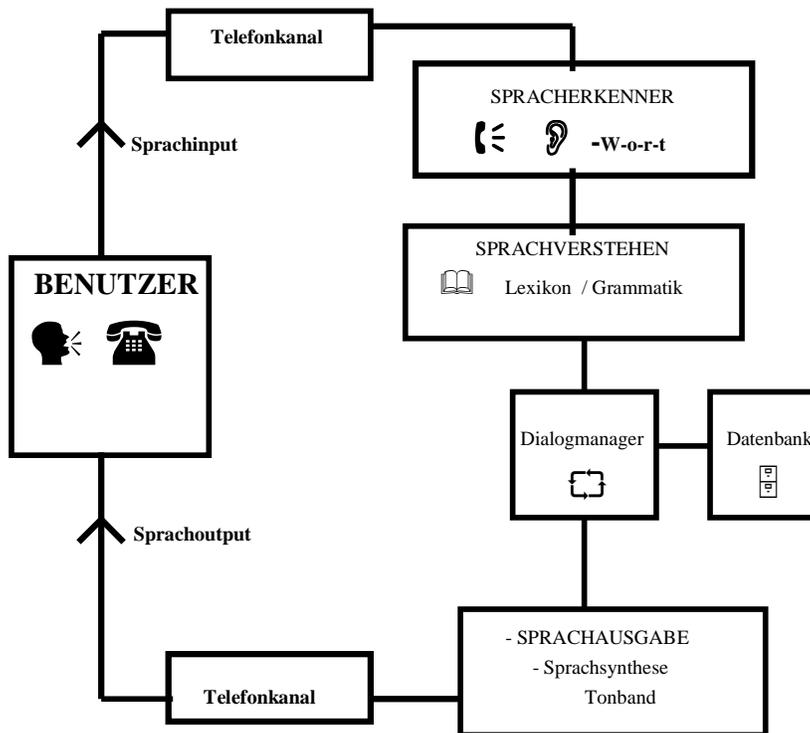
Um einen Dialog in seinen qualitativen Aspekten untersuchen zu können ist es notwendig, sich zuerst die am Gespräch beteiligten „Akteure“ detailliert anzuschauen. Der „Mensch“ als Gesprächspartner braucht an dieser Stelle nicht weiter erläutert zu werden. Dagegen stellt die „Maschine“ einen ungewöhnlichen Gesprächspartner dar, der erklärungsbedürftig ist.

3.3.1 Darstellung eines natürlichsprachlichen Dialogsystems

NSD-Systeme verfügen über ein komplexes Softwaresystem, das es ihnen ermöglicht sprachliche Äußerungen zu analysieren und weiter selbst zu produzieren. Hier unterscheidet

man zwischen unimodalen und multimodalen Systemen, wobei die multimodalen zusätzlich die Information dem Benutzer in geschriebener Form bieten können, und somit für jeden Zweck ein optimales Eingabe- bzw. Ausgabemedium darstellen.

Abb. 4: Architektur eines unimodalen sprachbasierten Dialogsystems



Das Strukturschema der Abbildung 4 zeigt eine grobe Darstellung der Hauptkomponenten eines natürlichsprachlichen Dialogsystems und ihrer Interaktion bei der Verarbeitung einer Benutzereingabe. Als Übertragungskanal wird hier die Telefonleitung verwendet.

Der Sprachinput des Benutzers besteht aus akustischen Signalen, die im Rahmen der **Spracherkennung** analysiert und einzelnen Phonemen zugeordnet werden. Aus diesen Phonemen versucht man nun Wortketten zu rekonstruieren. Die von einem Aussprachelexikon und einem Sprachmodell unterstützte Erkennung wird dadurch zu einem Puzzlespiel, wobei alle möglichen Wortkombinationen geprüft werden, bevor man eine endgültige Entscheidung für die wahrscheinlichste Kette trifft. Durch den Vergleich akustischer Eigenschaften der gesprochenen Wörter mit denen ihrer Referenzmustern und durch den Einsatz statistischer Methoden (wie z.B. Hidden-Markov Modelle⁷) werden Wort- und Satzhypothesen erzeugt, die dem **Sprachverstehensmodul** zur Verarbeitung weitergereicht werden.

In dem **Sprachverstehensmodul** wird die Wortrekonstruktion syntaktisch, semantisch und pragmatisch analysiert. Das Ziel dabei ist die Bedeutung der einzelnen Wörter zu ermitteln. Durch die syntaktische Analyse wird den sequentiellen Eingabemustern eine hierarchische Analyse aufgrund von bestimmten Kombinationsregeln zugewiesen. Dieser Vorgang nennt sich Parsing. In der Regel benutzt man für das Parsingverfahren eine Grammatik, die

⁷ Die statistischen Hidden Markov-Modelle werden seit den 60'er Jahren zur Spracherkennung eingesetzt [Rabiner, 1986].

beschreibt soll, wie Wörter in einem Satzbau miteinander kombiniert werden können. Für die syntaktische Analyse ist in der Computerlinguistik eine große Zahl von Analyseverfahren entwickelt worden [Pereira, 1997].

Syntax allein reicht nicht aus, um die Bedeutung der Benutzereingabe zu dekodieren. Es gibt Äußerungen, welche syntaktisch korrekt sind, die aber semantisch keinen Sinn ergeben. Die grammatikalisch korrekten Strukturen werden zunächst in semantische Bäume umgewandelt. Die semantische und pragmatische Analyse wird dann überprüfen, welche Bedeutung die sprachlichen Einheiten haben und wie sie vom Benutzer im Dialogzusammenhang gebraucht werden. Dafür werden die einzelnen Wörter der Äußerung mit dem Lexikon⁸ abgeglichen, um Grundformen und syntagmatische Einheiten zu ermitteln. Interessante Darstellungen zu syntaktischen und semantischen Analysen findet man bei [Görtz et al., 1993], [Wahlster et. al., 1987] und [Dengel, 1994].

Nun wird die dekodierte Information an den **Dialogmanager** weitergeleitet. Ihm fällt eine zentrale Rolle zu, nämlich die Steuerung des Dialogs. Außerdem muss er Lösungen zu verschiedenen Problemen finden, die durch die untergeordneten Komponente (Spracherkennung und Sprachverstehen) entstehen können. So z. B. muss das System eine kontinuierlich gesprochene Sprache verarbeiten, die ihm oft unklare und verwaschene Daten liefert. Diese Daten können fragmentiert, falschgeformt und von speziellen linguistischen Besonderheiten geprägt sein [Loderer, 1998]: Hintergrundgeräusche, Husten, unterschiedliche Betonung, Wortakzente, undeutliche und unvollständige Aussprache eines Wortes, Wort- und Silbenviederholungen, sowie unterschiedliche akustische Realisierungen der Phoneme können die Spracherkennung erheblich beeinträchtigen.

Aus den genannten Gründen kann es schon mal vorkommen, dass die Spracheingabe des Benutzers fehlerhaft dem Parser zugeführt wird. Wenn die Information vom Parser nur teilweise interpretiert werden kann, ist es die Aufgabe des Dialogmanagers, dieses Problem zu lösen.

Der Dialogmanager bedient sich der Voraussagen von Dialogen, die in der Planungsphase des Systems aufgestellt werden. Diese Voraussagen sind Erwartungen über potentielle Benutzereingaben, die im System vorprogrammiert sind [Loderer, 1998]. Der Dialogmanager muss sicherstellen, dass vom Benutzer ausreichende Information zur Verfügung gestellt wird, und wenn dies nicht der Fall ist, durch zusätzliche Fragen die Datenlücken ausfüllen.

Wenn die vollständige Information bereit liegt, kommt dem Dialogmanager eine andere wichtige Aufgabe hinzu: die Interpretation des gesamten Outputs innerhalb eines dynamisch aktualisierenden Kontextes [Loderer, 1998]. Der Dialogmanager muss aufgrund dieser Interpretation eine Antwortstrategie ausarbeiten. Dazu wird er eine Datenbank

⁸ Das Lexikon setzt sich aus einer Wortliste zusammen, die mit syntaktischen, morphologischen und semantischen Eigenschaften ausgestattet ist.

verwenden, die Informationen je nach Aufgabengebiet über Fahrpläne, Abfahrt- und Ankunftszeiten, Reservierungsmöglichkeiten, Adressen usw. beinhaltet.

Die erzeugte Antwort wird in Textform umgewandelt und weiter an die **Sprachausgabekomponente** geführt. Der Text wird dort entweder von einer Sprachsynthesen (*Text to Speech*), oder es werden vorher aufgezeichnete natürlichsprachliche Äußerungen abgespielt. Somit erreicht die vom System verarbeitete Information in akustischer Form den Benutzer.

3.4 Elemente einer interaktiven Sprachtheorie

Es gibt unzählige Theorien, welche die zwischenmenschliche Kommunikation erforschen, aber fast keine, die sich mit der Mensch-Maschine Kommunikation auseinandersetzt. Die Gründe dafür werden ausgezeichnet von [Fraser et al., 1991] zusammengefasst: „The designer is caught in a vicious circle - it is necessary to know the characteristic of dialogues between people and automata in order to be able to built the system, but it is impossible to know what such dialogues would be like until such a system has been built“.

Die Erforschung menschlicher Dialoge liefert eine gute Basis für die Konstruktion eines Mensch-Maschine Interaktionsmodells. Sie kann aber aufgrund der bereits erwähnten Unterschiede zwischen menschlichen und „computerischen“ Gesprächspartnern nicht alle Aspekte dieser Interaktion berücksichtigen. Aus diesem Grund haben dänische Wissenschaftler [Bensen et al., 1998] versucht, eine interaktive Sprachtheorie zu entwickeln, die Struktur, Inhalt und Dynamik einer Mensch-Maschine Interaktion beschreibt. Die Theorie soll systematisch alle Komponenten dieser Interaktion erfassen und ihre gegenseitige Wechselwirkung analysieren. Dadurch sollen wertvolle Hinweise zur optimalen Dialoggestaltung geliefert werden.

Die Theorie ist, nach der Meinung der Autoren, nicht vollständig. Sie hat einen Entwurfcharakter und stellt ein Modell dar, welches sich aus mehreren konzeptuellen Elementen zusammensetzt. Das Modell organisiert elf Elemente auf fünf pyramidenartige Stufen. Das Bild entstand aus der Analyse verschiedener interaktiver Sprachsysteme und berücksichtigt Komponenten zwischenmenschlicher Kommunikationstheorien [Aust et. al., 1995; Bunt, 1994; Carberry, 1990, Bild 1.1; Eckert und McGlashan, 1993, Bild 1 und 2; Grau et al., 1994, Bild 1; Groz und Sidner, 1986; Jönsson, Bild 7.1 und 7.2; Smith und Hipp, 1994, Bild 3.3.; Zue, 1994]. Diese Quellenangaben unterstreichen den interdisziplinären Charakter dieses Modells, welches von verschiedenen Fachdisziplinen, wie Linguistik, Computerwissenschaften und kognitiven Wissenschaften abgeleitet wurde.

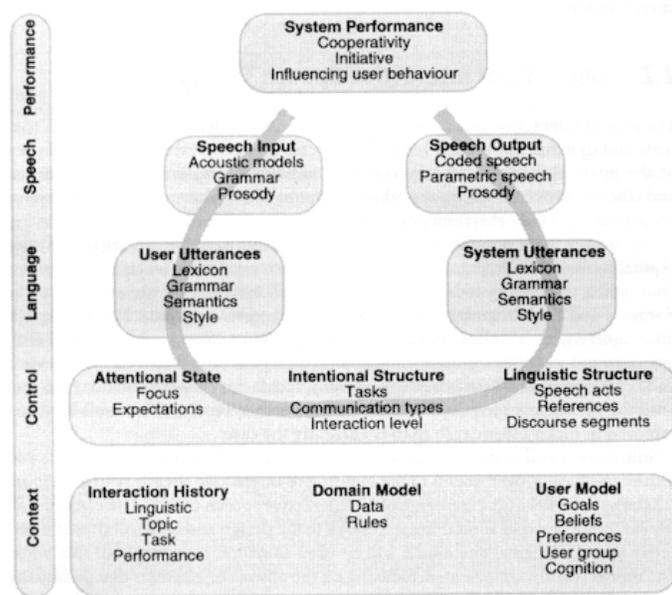
I. Auf der ersten Stufe der Pyramide befindet sich das **Context**-Modul⁹. Hier wird der Inhalt folgender Komponenten festgelegt: des Lexikons, der Interpretation der Äußerungen

⁹ Die Begriffe wurde absichtlich nicht übersetzt, um einen höheren Kohärenzgrad mit der englischsprachigen Literatur zu erreichen.

und der Planung und Durchführung der Aufgabe. Der *Context* setzt sich aus drei Elemente zusammen: *Interaction History*, *Domain Model*, *User Model*.

Interaction History ist eine selektive Speicherung von Informationen, die während der sprachlichen Interaktionen ausgetauscht werden. Die Theorie unterscheidet hier zwischen vier Kategorien: *Linguistic History*, die sich auf die linguistische „Sprachoberfläche“ bezieht; *Topic History*, welche die Reihenfolge der Subaufgaben angibt; *Task History*, die alle relevanten im Laufe der Interaktion ausgetauschten Informationen speichert; *Performance History*, die den Interaktionsfortschritt zwischen Benutzer und System berücksichtigt.

Tab. 5 Elemente einer interaktiven Sprachtheorie¹⁰



Domain Model bezieht sich auf den Informationsbereich, der vom System angeboten wird. Das *Domain Model* soll sowohl relevante Daten enthalten, die in direkter Beziehung zu den künftigen Aufgaben stehen als auch allgemeine Informationen, die es dem System ermöglichen, logischen Schlussfolgerungen zu ziehen, (z.B. das System soll wissen, dass ein Rückflugdatum zeitlich immer nach dem Abflugdatum stehen soll).

User Model soll im voraus die Ziele, Annahmen und Präferenzen des Benutzers berücksichtigen. Dabei werden die potentiellen Benutzer in Gruppen unterteilt (z.B. erfahrene und unerfahrene Benutzer) und dem System unterschiedliche kommunikative Strategien je nach den Bedürfnissen einer jeden Benutzergruppe implementiert.

II. Die zweite Stufe wird vom *Interaction Control*-Modul belegt. Dies ist für die Dialogsteuerung die wichtigste Modellstufe. Unterschieden wird hier zwischen drei Elementen: *Attentional State*, *Intentional Structure* und *Linguistic Structure*.

¹⁰ Siehe [Bensen et. al., 1998]

Das *Attentional State*-Element wird vom System als *Focus Set* repräsentiert. Darunter ist eine Reihe thematischer Subaufgaben zu verstehen, die im Laufe eines Dialogs auftreten können (z.B. das Thema „Fahrkarten reservieren“ kann in verschiedenen Subaufgaben fokussiert werden, wie z.B. Tag der Abfahrt, Uhrzeit, Preis der Fahrkarte, Datum usw.)

Expectation beziehen sich auf die erwarteten lexikalischen Inputs. Aufgrund von *Expectations* wird die Selektierung eines relevanten akustischen Modells, Lexikons und Grammatik durchgeführt.

Intentional Structure beschreibt die Interaktionsmöglichkeiten, die Benutzer und System je nach kommunikativen Absichten zur Verfügung stehen. Sie stellt gleichzeitig eine gewisse Kontrollinstanz dar, die dafür sorgt, dass jeder ausgetauschte Informationsteil eines Dialogs zur Erfüllung der Aufgabe beiträgt. Innerhalb der *Intentional Structure* wird zwischen drei Interaktionsarten unterschieden: die aufgabenorientierte Interaktion, als Kernpunkt eines Dialogs, die meta-kommunikative Interaktion als Strategie zur Lösung und Korrektur von Missverständnissen und die Rahmeninteraktion, die sich auf die Formulierungen am Anfang und am Ende des Gesprächs bezieht.

Die Interaktionen können sich auf verschiedenen Niveaus abspielen, je nach Systemkomplexität. Auf dem „niedrigsten“ Niveau kann das System vom Benutzer verlangen, dass er seine Antwort buchstabiert; auf dem „höchsten“ wird die Frage sehr allgemein formuliert, so dass der Benutzer seine Antwort offen, ohne Einschränkungen gestalten kann.

Tab. 2 Interaktionsniveaus und ihr Einfluss auf die erwarteten Benutzerinputs¹¹

Niveau	Ausgabeinput
<u>1. Buchstabieren</u> Wie buchstabiert man Ihren Namen?	<u>1. Die Antwort wird buchstabiert</u> M-ü-l-l-e-r.
<u>2. Ja/Nein</u> Möchten Sie auch eine Rückfahrkarte?	<u>2. Ja oder Nein</u> Ja.
<u>3. Multiple choice</u> Möchten Sie ein <i>one-way</i> Flugticket oder ein Rückflugticket ?	<u>3. Liste von akzeptablen Werten</u> Hin und zurück bitte!
<u>4. Generell, fokussiert</u> Wann möchten Sie hinfliegen?	<u>4. Die Antwort wird fokussiert</u> Am Montag, nächste Woche.
<u>5. Generell nicht fokussiert</u> Was kann ich für Sie tun?	<u>5. Jede Eingabe ist möglich</u> Ich möchte gerne wissen, ob ich meine Katze im Flieger mitnehmen darf?

Diese Interaktionsniveaus sind Teil einer meta-kommunikativen Strategie, die eingeführt werden, um das Verstehen zu erleichtern. Wichtig ist dabei, dem Benutzer klarzustellen, welches Input genau von ihm erwartet wird: z.B. die Frage „Möchten Sie Ihre Eingaben

¹¹ [Bensen et al., 1998]

Fachliteratur findet man ausführliche Beschreibungen von Dialogsequenzen bei [Clark und Schaefer, 1989], [Ahrenberg et al., 1995], [Carletta et al., 1997], sowie bei [Mann und Thompson, 1987] und speziell für Dialogssysteme bei [Fischer et al., 1994] und [Stein und Maier 1994].

III. Auf der dritten Stufe befindet sich das *Language-Modul*, die sich aus zwei Elementen zusammensetzt: *User Utterances* und *System Utterances*. *Utterances* sind sprachliche Äußerungen, die sich den starren Regeln der Schriftsprache nicht unterwerfen. Die Syntax [Rupp, 1965] der spontanen Rede stimmt oft mit der normativen Syntax der geschriebenen Sprache nicht überein. Das situativ spontane Gespräch folgt eigenen Gesetzen, wonach einfache, elliptische und sogar fehlerhafte Sätze erlaubt sind [Lewandowski, 1979]. Solche Bedingungen, die erheblich die maschinelle Verarbeitung erschweren, müssen bei der Systemmodellierung beachtet werden.

Um Äußerungen verstehen und selbst produzieren zu können braucht das System ein **Lexikon** und eine **Grammatik**. Der Inhalt des Lexikons wird aufgrund empirischer Simulationen mit Versuchspersonen festgelegt; obwohl die Grammatik auf den Regeln der Schriftsprache beruht wird sie in erster Reihe alle grammatikalischen korrekten Konstrukte der Spontanäußerungen berücksichtigen, die von Benutzer immer wieder gebraucht werden.

Nicht selten kommt es vor, dass Systeme durch die am Anfang des Dialogs gegebene Anleitung oder durch die eigene Sprachausgabe das kommunikative Verhalten der Benutzer beeinflussen. Der Benutzer wird versuchen, sich seinem „Gesprächspartner“ anzupassen und ungeeignete Ausdrucksweise (z.B. zu lange Sätze bilden, zu viel sprechen, zu höflich sein usw.) zu vermeiden, um einen optimalen Informationsaustausch zu erreichen.

Im Rahmen der *System Utterances*-Komponente werden die Bestätigungsstrategien festgelegt. Man kann entweder eine explizite Bestätigungsstrategie anwenden, wobei das System die Benutzereingaben wiederholt und somit überprüft, ob ein Fehler vorliegt:

BEISPIEL: SYSTEM: „Sie wollen also von Hamburg nach Berlin fahren?“,

Die Bestätigung kann auch implizit durch Verknüpfung mit einer neuen Frage erfolgen:

BEISPIEL: SYSTEM: „Wann wollen Sie von Hamburg nach Berlin fahren?“.

IV. Auf der vierte Stufe befindet sich das *Speech-Modul*, welches sich aus zwei weiteren Elementen zusammensetzt: *Speech Input* und *Speech Output*.

Typische Bewertungseinheiten für den *Speech Input* sind die Wort- und Satzgenauigkeit. Gemessen wird dabei, wie viele Wörter bzw. Sätze korrekt erkannt wurden. Die Erkennung ist von einer guten Übereinstimmung des akustischen Modells mit den beobachteten Daten abhängig. Dafür ist es notwendig ein geeignetes akustisches Modell

abhängig von Aufgabe und Benutzergruppe zu wählen und das Lexikon auf eine gewisse Anzahl von Vokabel zu reduzieren. Je kleiner der Wortschatz, desto geringer ist die Gefahr Erkennungsfehler, denn bei einer kleineren Anzahl von Vokabel können verschiedene Aussprachevarianten von Phonemen berücksichtigt werden. Obwohl sinnvoll, generiert die Begrenzung des Vokabulars eine Konfliktsituation: das System soll, trotz seines kleinen Vokabulars, fähig sein, eine Vielzahl von spontanen Äußerungen zu verarbeiten, in denen die verschiedensten Wortkombinationen erscheinen können. Heute ist es möglich Systeme mit einem Vokabular von bis zu 2000 Wörtern auszustatten.

Die Erkennung kann in *real-time* (z.B. in Dialogsituationen) oder als Aufeinanderfolgen von Prozessen (es wird jeweils ein *Keyword* erkannt) ablaufen. Für interaktive Dialogsysteme ist eine *real-time* Erkennung absolut notwendig.

Die Erkennung kann sprecherabhängig (z.B. für alternatives Nutzen desselben PCs), sprecherunabhängig (z.B. beim Einsatz im öffentlichen Bereich), sprecheradaptiv (verschieden Benutzergruppen, je nach Geschlecht, Dialekt, Sprache usw. angepasst) oder individuell trainiert sein (bei sprecherabhängigen Systemen).

Der ***Speech Output*** kann aufgrund von zwei unterschiedlichen Verfahren erzeugt werden: durch Verkettung oder Parametrisierung. Die Verkettung besteht aus im voraus aufgenommenen Silben, Wörtern und Sätzen; sie werden von einer natürlichen Stimme gelesen, anschließend zusammen verkettet und abgespielt. Nachteilig dabei ist, dass der Satzrhythmus und die Satzmelodie nicht immer richtig eingehalten werden. Außerdem wirken sich die Verkettungsstellen negativ auf Qualität der erzeugten Sprache aus. Man geht deshalb zu längeren Einheiten (z.B. Halbsätzen) über. Die Erzeugung der dazu erforderlichen großen Datenbank ist jedoch kostspielig: für neue Wortaufnahmen muss derselbe Sprecher mit gleichem Ton und gleicher Stimmqualität die neu hinzukommenden Wörter bzw. Sätze aussprechen. Bei einem parametrischen Verfahren generiert ein Synthetisator ein akustisches Signal, auf Grundlage eines menschlichen Spracherzeugungsmodells. Ein Anregungssignal wird entweder in Form eines Rauschens (stimmlose Anteile) oder eines gefilterten Impulskamms (stimmhafte Abschnitte) generiert. Diesem Signal wird anschließend mittels eines Filtersystem die gewünschte Lautinformation aufgeprägt. Dieses Verfahren benötigt keine große Datenbank und ermöglicht die Berücksichtigung prosodischer Merkmale, wie Betonung, Pausen usw. Der Nachteil dabei ist die oft schlechte Sprachqualität.

V. Die fünfte und letzte Stufe der Pyramide wird durch das ***System Performance***-Modul belegt; darunter werden folgende Elemente gezählt: ***Cooperativity, Initiative, Influencing User Behaviour.***

Cooperativity wird von [Bersen et al., 1998] als „*a key to a succesfull interaction design*“ bezeichnet und spielt eine zentrale Rolle bei der Qualitätsbeurteilung von NSD-Systemen. Unter ***Cooperativity*** (Kooperativität) soll die Fähigkeit eines Dialogssystems verstanden

werden, den richtigen Informationssinn aus der Benutzereingabe zu extrahieren (z.B. die Antwort auf die Frage wie: „Wissen Sie wie spät es ist?“ könnte „Ja, ich weiß es!“ lauten; eine kooperative Antwort wäre aber „Ja, sicher, es ist 14.30“ [Sadek und De Mori, 1997]). In den zwischenmenschlichen Gesprächen sind kooperative Antworten meist die Regel. Für Dialogsysteme sind solche Gesprächssituationen immer noch problematisch.

Kooperativität bezieht sich noch auf die Fähigkeit des Systems informativ passende Lösungen zu bieten, auch wenn zu den angeforderten Informationswünschen keine direkte Verknüpfung in der Datenbank gefunden wird: z.B. wenn die Anfrage eines Benutzers lautet: "Ich möchte zum Italiener im Stadtteil X essen gehen", und das System da kein italienisches Restaurant findet, sollte es überprüfen, ob es im angegebenen Ort ein anderes Restaurant gibt, das kein italienisches ist, aber dafür italienische Gerichte anbietet.

Initiative (Dialoginitiative) bezieht sich auf die Anregung zum Dialog. Die Modellierung der Dialoginitiative ist ein wichtiger Punkt des Systemdesign, die Auswirkungen auf das kommunikative Verhalten des Benutzers hat. Für ein benutzerfreundliches System eignet sich am besten eine gemischte Dialoginitiative. Eine solche Dialoginitiative soll die vollständige Übernahme der Dialogkontrolle durch den Benutzer (wie z.B. im Falle eines Frage-Antwortsystems) oder durch das System (in diesem Fall wäre der Benutzer ein passiver, der Maschine vollkommen unterworfenen Gesprächspartner) verhindern. Das Bedürfnis nach einer gemischten Dialoginitiative ergibt sich auch dadurch, dass Benutzer einen unterschiedlichen Informationsbedarf haben: in manchen Fällen sind Benutzer nur an einem Teil der Information interessiert und möchten manche vorgeplante Dialogsegmente überspringen, was aufgrund einer gemischten Dialoginitiative ermöglicht wird. Eine gemischte Dialoginitiative erlaubt, sowohl dem Benutzer als auch dem System beliebig in den Dialogfluss einzugreifen und schafft somit die Bedingungen einer optimalen Interaktion.

Influencing User Behavior ist das letzte Element des **System Performance**-Moduls.

Im Gegensatz zum System stellt der Mensch einen Faktor dar, dessen Reaktionen nicht direkt kontrollierbar sind. Das System ist als Gesprächspartner dem Menschen weit unterlegen. Dieses asymmetrische Verhältnis soll dem Benutzer signalisiert werden, um eventuelle Missverständnisse und falsche Erwartungen dem System gegenüber zu vermeiden. Das System kann das Verhalten der Benutzer indirekt, durch die Vorgabe seiner Sprachausgabemuster oder direkt durch Systemanweisungen vor dem Gespräch beeinflussen.

4. Qualitätsbeurteilung einer Dialogsteuerung

Bevor die Qualitätsbeurteilung einer Dialogsteuerung einer genauen Analyse unterzogen wird, soll hier der Begriff „Qualität“ definiert werden. Dabei ist festzustellen, in welchem Bezug die „Qualität“ zum Dialog steht.

4.1 Qualität: Versuch einer Definition

Das Wort Qualität wird als Synonym sowohl für „positiv“ als auch für „Güte“ gebraucht und steht im Zusammenhang mit der Beurteilung von Gegenständen, wie z.B. Waren, Industrieprodukten und Dienstleistungen. Man spricht von der Qualität eines Produktes und meint damit seine positiven Eigenschaften. Der Kunde ist derjenige, der darüber entscheidet, ob ein Produkt Qualität besitzt oder nicht. Es hängt von dem Nutzen ab, der ihm die Ware bringt.

Die Qualität bezieht sich auf die Einhaltung bestimmter ästhetischer, funktionaler oder inhaltlicher Standardnormen. Diese Standardnormen setzen sich aus den Qualitätsmerkmalen zusammen, die das Produkt zu erfüllen verspricht. Aus diesem Grund ist die Qualität eine relative Eigenschaft: sie hängt von den Standardnormen ab, nach denen sie beurteilt wird und von den Erwartungen, die Beurteiler den Produkten gegenüber haben. Sie kann zeitlich und örtlich variieren.

Um die Beurteilung der Qualität zu vereinheitlichen müssen Maßstäbe gefunden werden, die Allgemeingültigkeit besitzen.

Die Standardnormen für Produkte, die z.B. aus dem Industriebereich stammen, werden, je nach Produkt, in Anforderungskatalogen aufgelistet und sind als verbindliche Referenz für die Qualitätsbeurteilung zu verstehen. Die erfahrene Qualität wird ermittelt, indem man die Produkteigenschaften mit ihrer aus dem Katalog verbindlichen Referenz vergleicht.

Kompliziert wird es für Produkte, die nicht aus dem industriell-wirtschaftlichen Bereich kommen und die etwa künstlicher oder immaterieller Natur sind. Dafür gibt es keine gelisteten Anforderungskataloge, an denen man sich orientieren kann. Man muss sich selbst Maßstäbe setzen, um die Qualität dieser Produkte ermitteln zu können.

Wie kann man einen solchen Maßstab ermitteln? Zuerst muss festgelegt werden, in welcher Hinsicht man die Qualität definieren möchte: ob sie inhaltliche, funktionale oder ästhetische Aspekte bewerten soll. Die Eigenschaften eines Produktes setzen sich aus mehreren Merkmalen zusammen, die je nach Bewertungsaspekt qualitätsrelevant sind.

Unter einem qualitätsrelevanten Merkmal versteht man ein Merkmal, welches einen Beitrag zur optimalen Erreichung der Produktbestimmung leistet.

Nach der Festlegung der qualitätsrelevanten Merkmale eines Produktes stellt sich die Frage, inwieweit diese den gewünschten Merkmalen sich annähern: Je kleiner die Unterschiede, umso höher ist die Qualität des Produkts zu bewerten.

Die Qualität kann entweder aus der Perspektive des wahrnehmenden Benutzers betrachtet werden, um Verhaltenweise und grundlegende Prinzipien aufzudecken, oder aus dem Blickwinkel der Systementwickler, deren Ziel ist Produkte zu gestalten, die vom Benutzer als qualitativ hochwertig eingestuft werden [Jekosch, 2000].

Für NSD-Systeme stellt die Dialogsteuerung das Produkt dar, dessen Qualität wir beurteilen wollen. Dafür müssen die Merkmale und vor allem die Werte dieser Merkmale erfasst werden. Zuerst sollen wir uns das Objekt unserer Qualitätsbeurteilung - die Dialogsteuerung- näher anschauen: sie erfüllt die zentrale Aufgabe des Systems. Sie ermittelt abhängig von den Eingaben des Benutzers die jeweils adäquaten Antworten. Ihr Ziel ist funktional zu definieren, nämlich dem Benutzer die gewünschte Information zu überbringen. Die qualitätsrelevanten Merkmale werden dementsprechend auf die Dialogfunktionen bezogen sein und Hinweise zur optimalen kommunikativen Zielerreichung liefern.

4.2 Evaluierungsmethoden

Die Perspektive der Kommerzialisierung von NSD-Systemen stellt das Problem der Interaktionsqualität in den Vordergrund. Dabei wird das Bedürfnis nach geeigneten Evaluierungsmethoden immer größer.

Grundsätzlich können Dialogsysteme methodologisch mit Hilfe objektiver und subjektiver Kriterien evaluiert werden. Die objektive Evaluierung misst metrische Parameter, wie z.B. Dialogdauer, durchschnittliche Anzahl der *Turns* pro Subaufgabe, allgemeine Fehlerrate. Diese Daten werden aufgrund von *Logfiles* (siehe Abschnitt 5.1.3) ermittelt und sind von der Meinung der Benutzer unabhängige Messeinheiten. Die subjektive Evaluierung ermittelt Bewertungen, die von den Versuchspersonen im Laufe eines Experiment vorgenommen werden. Die Parameter, die dadurch zu Stande kommen, bilden die Grundlage der Qualitätsbeurteilung des Systems. Sie geben Auskunft über die allgemeine Akzeptanz und Zufriedenheit der Benutzer dem System gegenüber und messen die Einfachheit der Bedienung, subjektiv empfundene Schnelligkeit des Systems, Erwartungen, inhaltliche und sprachliche Verständlichkeit, subjektiv empfundener Stress. Viele Untersuchungen kombinieren beide Methoden, um präzisere Ergebnisse zu erhalten. In der Fachliteratur über Dialogsysteme findet man unzählige Darstellungen verschiedener Evaluierungsarten. Die wichtigsten werden hier unten kurz zusammengefasst.

[Hirschman und Thompson, 1994] schlagen drei Typen von Evaluierungsmethoden vor:

Adequacy Evaluation, die die Tauglichkeit eines Systems für einen bestimmten Zweck ermittelt: es soll überprüft werden, inwieweit bestimmte Systeme für eine spezielle Aufgabe geeignet sind.

Diagnostic Evaluation stellt, wie der Name andeutet, eine Bewertung der gesamten Systemleistung anhand von Mustereingaben dar. Diese Evaluierung zeigt, welche Änderungen das System verbessern können.

Performance Evaluation misst die Leistungen einer oder mehrerer Komponenten des Systems. Dabei werden die zu untersuchenden Komponenten nacheinander in verschiedene Systemkonfigurationen getestet, um herauszufinden, welche Konfiguration sich am besten eignet.

[Simpson und Fraser, 1993] unterscheiden im Rahmen des SUDIAL-Projektes zwischen zwei Evaluierungskonzepten: **Black Box** und **Glas Box Evaluation**.

Bei der **Black Box Evaluation** werden drei Kriterien überprüft: **Contextual Appropriateness**, **Turn Correction Ratio** und **Transaction Success**.

Das Konzept der **Contextual Appropriateness** stammt von dem britischen Philosoph und Linguisten Paul Grice [Grice, 1998]. Dieses Konzept misst die Angemessenheit der Systemaussagen im Dialogkontext. Für die Bewertung dieses Konzeptes ist ein Komitee Fachkundiger erforderlich, welche die Angemessenheit jeder einzelnen Systemäußerung überprüfen soll. Die Analyse unterscheidet zwischen fünf verschiedenen Niveaus der Angemessenheit:

Total failure: wenn das System nicht in der Lage ist, einen sprachlichen Output zu produzieren.

Appropriate: wenn die Systemäußerung gegen kein Grice'sches Kommunikationsprinzip verstößt (siehe Abschnitt 4.3).

Inappropriate: wenn die Systemäußerung gegen ein Grice'sches Kommunikationsprinzip verstößt.

Appropriate/Inappropriate: wenn die Fachkundigen sich nicht einigen können, ob eine Äußerung als **Appropriate** oder **Inappropriate** zu bewerten ist.

Incomprehensible: wenn eine Aussage als unverständlich bezeichnet wird.

Turn correction ratio ist ein Maßstab für die Anzahl der zusätzlichen *Dialogturns*, die man benötigt, um Missverständnisse im Dialogverlauf zu korrigieren.

Transaction success zeigt, inwieweit das System die angeforderte Information geliefert hat. Die Merkmale dieses Kriteriums werden unter folgenden Aspekten untersucht: **Succeed**, **Succeed with constraint relaxation**, **Succeed in spotting that no answer exists** und **fail**.

Die **Glass Box-Evaluation** bezieht sich auf die Auswertung der einzelnen Komponente im gesamtem Systemkontext. Dabei werden zwei bestimmbare Größen definiert: **Word**

Accuracy (ein Maß für die Anzahl richtig erkannter Wörter) und **Information Content** (die Fähigkeit des Systems, die Information aus den Benutzeraussagen zu entnehmen).

[Danieli und Gerbino, 1996] erweitern die Evaluierungskriterien von Simpson und Fraser um eine weitere Größe: **Implicit Recovery**. Diese Maßzahl steht für die Fähigkeit des Dialogsystems missverstandene Aussagen zu korrigieren.

Ein anderes Evaluierungskriterium ist die Benutzerfreundlichkeit des Systems, die aufgrund subjektiv empfundener Merkmale, wie z.B. Freundlichkeit der Stimme, Verständlichkeit der vom System gelieferten Informationen usw., ermittelt wird. Die Untersuchung von [Love et. al., 1994] stellte eine Liste von 18 sogenannten *Service Attributes* dar, die in fünf Kategorien eingeteilt wurden:

1. **die Qualität des Systems** in Bezug auf seine Effizienz und Zuverlässigkeit. Dabei sollen eventuelle Verbesserungen des Systems aufgezeigt werden.
2. **der Stress**; dazu zählen Aspekte, wie die Zufriedenheit der Benutzer, ihre Stimmung während des Gesprächs, die erforderliche Konzentration, die Schnelligkeit des Systems.
3. **die Freundlichkeit des Systems**; hier wird untersucht, wie schwierig das Gespräch mit dem System empfunden wird.
4. **die Gewandtheit des Systems** bezieht sich auf, wie höflich das System wirkt und welche Erwartungen die Benutzer dem System gegenüber haben.
5. **die Transparenz des Systems** bezieht sich auf, wie leicht es ist, das System zu bedienen und wie hilfreich es sich zeigt, wenn die Benutzer kommunikative Schwierigkeiten haben.

Interessante Untersuchungen, die objektive und subjektive Evaluierungskriterien kombinieren, findet man noch bei [Delogu et al., 1993], [Larsen, 1999], [Relaño et al., 1999], [Walker et. al., 1997], [Goodine et. al., 1992] und [Gleiss, 1992].

Bei einer kommunikationswissenschaftlicher Evaluierung steht die "menschliche" Perspektive in Vordergrund: die Interaktion soll einem Dialog zwischen Menschen ähnlich sein, wobei das System an seinen menschlichen Gesprächspartner angepasst werden soll. Für eine solche Evaluierung eignen sich einerseits Benutzerfreundlichkeitskriterien, die über die Zufriedenheit der Benutzer mit dem System berichten und andererseits Kriterien, die mit den Regeln der sprachlichen Kommunikation verbunden sind (diese Regeln werden im folgenden Abschnitt ausführlich dargestellt). Aus diesem Grund wird sich die Evaluierung im Rahmen dieser Arbeit zum einen auf die Untersuchung der qualitativen Aspekte eines Dialogs beziehen, die durch Verletzung eines Kommunikationsprinzip, im Sinne des *Appropriateness* Konzeptes, festgestellt und analysiert werden können. Zum

anderen werden Benutzerfreundlichkeitkriterien untersucht, die Hinweise auf Systemmängel aus Perspektive der Benutzer liefern. Eine Untersuchungsschema zur kommunikationswissenschaftlicher Evaluierung, die qualitativrelevante Dialogaspekte und Benutzerfreundlichkeitkriterien kombiniert wird im Abschnitt 4.4 vorgeschlagen.

4.3 Kommunikationsprinzipien eines kooperativen Mensch-Maschine Dialogs

Der Bedarf an Prinzipien und Regeln, an denen sich die Dialogstrategie orientieren soll, wird im Zuge der Entwicklung neuer Technologien immer größer.

In diesem Sinne versuchten die dänischen Wissenschaftler Niels O. Bernsen, Hans Dybkjær und Laila Dybkjær im Rahmen des sogenannten *Danish Dialog Project*¹² eine Liste von insgesamt 24 Richtlinien zusammenzustellen, die in Anlehnung an die Grice'schen Kommunikationsprinzipien [Grice, 1998] entstand, und im Laufe der empirischen Untersuchungen vervollständigt wurde. Ursprünglich waren diese Prinzipien von Grice als Richtlinien für die zwischenmenschliche Kommunikation gedacht; sie wurden durch zusätzliche Maximen an die Mensch-Maschine Kommunikation angepasst.

Die Prinzipien sollen als Regel der sprachlichen Mensch-Maschine Kommunikation verstanden werden, die qualitätsrelevante Dialogaspekte hervorheben. Im Mittelpunkt der Überlegung steht die Systemkooperativität als zentraler Schlüsselbegriff einer erfolgreichen Interaktion. Die zusammengestellten Richtlinien sollten dabei als Operationalisierungsinstrument der Systemkooperativität dienen [Bernsen et al., 1998].

Die Richtlinien unterscheiden zwischen generischen und spezifischen Prinzipien, wobei ein generisches Prinzip (GP) mehrere spezifische (SP) zusammenfassen kann. Alle Prinzipien werden in sieben Kategorien (sogenannten *Aspects*) unterteilt und jede Kategorie stellt eine von einem Prinzip angesprochene Dialogeigenschaft dar (Grice's Prinzipien sind mit einem Stern gekennzeichnet).

Tab. 3 Prinzipien einer kooperativen Systeminteraktion¹³

Interaction aspect	GG	SG	Generic or Specific Guideline
Aspect 1: Informativeness	GG1		*Make your contribution as informative as is required (for the current purpose of the exchange)
		SG1	Be fully explicit in communicating to users the commitments they have made
		SG2	Provide feedback on each piece of information provided by the user

¹² Im Rahmen dieses Projektes wurde ein Dialogsystem für Flugreservierungen in dänischer Sprache entwickelt. Siehe [Bernsen et al., 1998]

¹³ Siehe [Bernsen et al., 1998]

Aspect 2 Truth and evidence	GG2		*Do not make your contribution more informative than is required
	GG3		*Do not say what you believe to be false
	GG4		*Do not say that for which you lack adequate evidence.
Aspect 3: Relevance	GG5		*Be relevant, i.e. be appropriate to the immediate needs at each stage of the transaction
Aspect 4: Manner	GG6		*Avoid obscurity of expression
	GG7		*Avoid ambiguity
		SG3	Provide same formulation of the same question (or address) to users everywhere in the system's interaction turns.
	GG8		*Be brief (avoid unnecessary prolixity)
	GG9		* Be orderly
Aspect 5: Partner asymmetry	GG10		Inform the users of important non-normal characteristics which they should take into account in order to behave cooperatively in spoken interaction. Ensure the feasibility of what is required of them.
		SG4	Provide clear and comprehensible communication of what the system can and cannot do.
		SG5	Provide clear and sufficient instruction to the users on how to interact with the system.
Aspect 6 Background knowledge	GG11		Take partners' relevant background knowledge into account.
		SG6	Take into account possible (and possibly erroneous) user interference by analogy from related task domains.
		SG7	Separate whenever possible between the needs of novice and expert users (user-adaptive interaction)
	GG12		Take into account legitimate partner expectations as to your own background knowledge.
		SG8	Provide sufficient task domain knowledge and interference.
Aspect 7: Repair and clarification	GG13		Enable repair or clarification meta-communication in case of communication failure.
		SG9	Initiate repair meta-communication if system understanding has failed
		SG10	Initiate repair meta-communication in case of inconsistent user input
		SG11	Initiate repair meta-communication in case of ambiguous user input

4.3.1 Verletzung der Kommunikationsprinzipien

Es wird im Folgenden der Versuch unternommen, die Kommunikationsprinzipien mit konkreten Beispielen zu veranschaulichen und dabei zu zeigen, welche Konsequenzen die Verletzung eines Prinzips mit sich bringt. Die Beispiele stammen aus Konversationsversuchen, die im Rahmen des empirischen Teils dieser Arbeit mit Hilfe von BoRIS (das Bochumer Restaurantsauskunftssystem) durchgeführt wurden. Die Transkription der Dialoge basiert auf den originalen *Logfiles* (siehe Abschnitt 5.1.3.).

Die Beispiele erhalten auf der rechten Seite einen Kommentar, der die Prinzipverletzung kommentiert und einen Korrekturvorschlag macht. Die Eingaben, die zur Prinzipverletzung führen, werden im Text unterstrichen. Die Abkürzung „S“ und „B“ stehen für System und Benutzer. Die auf der linken Seite eingetragenen Zahlen beziehen sich auf die Interviewnummer (erste Zahl) und auf die Szenarionummer (zweite Zahl).

Interaction Aspect 1: Informativness

Dieser Aspekt und die dazu gehörigen Prinzipien beziehen sich auf die Informationsmenge, die im Laufe der Interaktion mit dem System ausgetauscht wird.

GG1: Say enough

Make your contribution as informative as is required (for the current purposes of the exchange).

Wenn das System nicht ausreichende Informationen bietet, können Missverständnisse im späteren Verlauf der Interaktion entstehen. Im besten Fall wird der Benutzer selbst nach den fehlenden Informationsteilen fragen, was für das System kommunikative Schwierigkeiten bereiten kann, im schlechtesten Fall wird der Dialog am Informationsmangel scheitern.

<p>[Nr.] 1.3 S: Guten Tag, willkommen bei BoRIS, dem Bochumer Informationssystem. BoRIS erlaubt Ihnen eine Recherche nach den folgenden Kriterien durchzuführen: der Art der Küche, der Tageszeit oder dem Preis der Mahlzeit, dem ungefähren Ort des Restaurants oder nach dessen Öffnungszeiten. Wie lautet Ihre Anfrage? B: Ort. S: Sie haben folgende Auswahlmöglichkeiten: Stadtmitte, Unicenter, andere Möglichkeiten oder keine Präferenz. B: andere Möglichkeiten S: Die anderen Möglichkeiten sind Querenburg, Langendreer und Wiemelhausen. B: <u>Altenbochum</u> S: Suchen Sie ein Restaurant für das Mittagessen oder für</p>	<p><u>Verletzung der GG1</u> Das System liefert nicht ausreichende Informationen. Um seine Ansagen möglichst kurz zu gestalten, gibt das System nicht alle Eingabemöglichkeiten an, die man zur Auswahl hat. Trotzdem versucht der Benutzer eine Art der Küche und einen Ort des Restaurant einzugeben, die vom System nicht genannt wurden. Er findet heraus, dass das System ihm kein passendes Restaurant nennen kann. Dabei kann er dabei nicht wissen, worauf diese Informationsverweigerung zurückzuführen ist: ob das System die <i>Keywords</i> „arabisch“ (Küche) und „Altenbochum“ (Ort)</p>
---	---

<p>das Abendessen? B: Für das Abendessen. S: Bevorzugen Sie eine bestimmte Küche, zum Beispiel italienisch, griechisch, deutsch? B: <u>Arabisch</u> S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p>nicht in seiner Datenbank hat, ob ein Erkennungsfehler vorliegt oder ob es am eingegeben Ort kein arabisches Restaurant gibt. VORSCHLAG: das System sollte den Benutzer informieren, warum es kein passendes Restaurant gefunden hat.</p>
<p>[Nr.] 8.1 S: Guten Tag [...]Wie lautet Ihre Anfrage? B: Art der Küche. S: Sie können zwischen den folgenden Möglichkeiten wählen: italienische, griechische, chinesische oder keine Präferenz. B: <u>Italienisch</u>. S: Suchen Sie ein Restaurant für das Mittagessen oder für das Abendessen? B: Mittagessen. S: An welchem Tag beabsichtigen Sie essen zu gehen? B: Samstag. S: Wie viel möchten Sie ungefähr pro Person ausgeben? B: 25 DM. S: In welchem Stadtteil möchten Sie essen gehen? B: <u>Querenburg</u>. B: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p><u>Verletzung der GG 1</u> Hier wird noch ein Fall von unzureichenden Information festgestellt. Der Benutzer fragt nach italienischer Küche im Stadtteil Querenburg, für 25 DM. Das System kann in seiner Datenbank nur preisgünstige dieser Art finden. Aus diesem Grund gibt es an, dass es kein passendes Restaurant gefunden hat. VORSCHLAG: das System sollte sich kooperativ verhalten und den Benutzer informieren, dass er italienische Gerichte in Querenburg für einen günstigeren Preis findet.</p>

SG1 (GG1) State commitments explicitly.

Be fully explicit in communicating to users the commitments they have made.

Es ist nicht ungewöhnlich, dass es zu einer Anfrage keine passende Information gibt und dass das System den Benutzer auffordert, seine Eingaben zu modifizieren. Es kann nur vorkommen, dass der Benutzer, nach einer ganzen Reihe misslungener Modifikationen vergisst, wonach er genau gefragt hat. Aus diesem Grund ist es sinnvoll, bevor das System die Restaurantadresse ausgibt, die vorab gemachten Benutzereingaben kurz zusammenzufassen. So können eventuelle Eingabefehler rechtzeitig korrigiert werden.

BoRIS ist mit einer solchen Funktion nicht ausgestattet.

<p>[Nr.] 22.1 S: Guten Tag [...] Wie lautet Ihre Anfrage? B: Küche. S: Sie können zwischen den folgenden Möglichkeiten wählen: [...] B: <u>Italienisch.</u> S: Suchen Sie ein Restaurant für das Mittagessen oder für das Abendessen? B: Abendessen. S: An welchem Tag beabsichtigen Sie essen zu gehen? B: Freitag. S: Wie viel möchten Sie ungefähr pro Person ausgeben? B: 40 DM. S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren? B: Ja. S: Sie können die Art der Küche, die Tageszeit [...] B: <u>Italienisch.</u> S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p><u>Verletzung der SG 1.</u> VORSCHLAG: das System sollte die Benutzeräußerungen zusammenfassen und ihm sagen, warum es kein passendes Restaurant gefunden hat. In diesem Beispiel hat der Benutzer entweder vergessen, dass er schon einmal „italienisch“ eingegeben hat oder er weiß einfach nicht, wie er seine Eingaben modifizieren soll.</p>
--	--

SG2 (GG1) Provide immediate feedback.

Provide feedback on each piece of information provided by the user

Bei vielen interaktiven Sprachsystemen, die über einen schwachen Spracherkennung verfügen, ist es notwendig, eine Bestätigungsstrategie zu implementieren, die zu jeder Eingabe des Benutzers ein Feedback liefert. So können Missverständnisse frühzeitig erkannt und korrigiert werden. Der einzige Nachteil ist, dass die immer wiederkehrenden Bestätigungen den Kommunikationsfluss erschweren. Bei **BoRIS** wurden keine Bestätigungsmöglichkeiten implementiert.

<p>[Nr.] 28.1 S: An welchem Tag beabsichtigen Sie essen zu gehen? B: <u>Am 16 Mai.</u> S: <u>[cheap]</u>¹⁴ Es gibt mehrere auf Ihre Anfrage passende Restaurants.</p>	<p><u>Verletzung der SG 2</u> Das System interpretiert das Datum 16. Mai fälschlicherweise als Preiseingabe. VORSCHLAG: Eine Bestätigung hätte wahrscheinlich dieses Missverständnis aufgeklärt.</p>
---	---

GG 2 Don't say too much

Do not make your contribution more informative than is required.

Der Benutzer kann durch Informationsüberfluss verwirrt werden und leicht den Interaktionsfokus aus den Augen verlieren. Der Informationsüberfluss kann nur in Extremfällen den Dialog zum Scheitern bringen; meistens führt er zu einer Verlängerung

¹⁴ In Klammern, fett gedruckt, steht die Variable (auf englisch oder auf französisch), nach der das System die Datenbankanfrage macht.

der Dialogdauer: der Benutzer wird das System zur Eingabewiderholung auffordern. Bei **BoRIS** konnte keine Informationsüberfluss festgestellt werden.

Interaction Aspect 2. Truth and Evidence

Diese Richtlinie ist für ein Informationssystem von fundamentaler Bedeutung. Eine Aussage, egal welcher Art, hat nicht den geringsten Wert, wenn sie nicht der Wahrheit entspricht. Es ist die einzige Richtlinie, die sich mit einem extrakommunikativen moralischen Aspekt auseinandersetzt.

GG 3 Don't lie

Do not say what you believe to be false.

Es ist äußerst wichtig, dass der Benutzer die Systemäußerungen für glaubwürdig hält, sonst wird die Inanspruchnahme eines solchen Informationssystems sinnlos. Die Verletzung dieses Prinzips kann durch mangelnde Pflege der Datenbank verursacht werden.

<p>[Nr.] 20.5 S: Guten Tag [...] Wie lautet Ihre Anfrage? B: Ich würde gerne serbisch essen gehen und zwar Preis so um die 40 Mark, das ganze an der Uni. S: Es gibt mehrere auf Ihre Anfrage Passende Restaurants: Brinkhoff's Stammhaus, Casa portuguesa[...]</p>	<p><u>Mögliche Verletzung der SG 3</u> Das System liefert eindeutig falsche Informationen: „Casa portuguesa“ (portugiesisches Restaurant, im Stadtteil Altenbochum) wird als serbisches Restaurant, in Uni-Nähe (Stadtteil Querenburg) angeboten. Das Prinzip der Wahrheit wird dadurch verletzt. Die Verletzung ist entweder auf mangelnde Datenbankpflege oder auf einen Erkennungsfehler zurückzuführen.</p>
--	---

GG4. Check what you say

Do not say that for which you lack adequate evidence

Im Falle einer unklaren Benutzereingabe soll das System durch zusätzliche Fragen überprüfen, ob es den Informationswunsch richtig verstanden hat. So kann die Verarbeitung von Eingaben vermieden werden, die das System zur falschen Informationenausgabe veranlassen.

<p>[Nr.] 7.1. S: Guten Tag [...]Wie lautet Ihre Anfrage? B: ich würde chinesisch, am Bahnhof nicht all zu teuer S:[chinese expensive centre] Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p><u>Verletzung der GG4</u> Die Benutzereingaben wurden falsch verstanden. „Nicht all zu teuer“ wird als „teuer“ interpretiert. Bei BoRIS wurde keine Grammatikkomponente implementiert, was erklärt, warum das System nicht weiß, dass ein Negationsadverb („nicht“) den Sinn der nachstehenden Wortkombination in ihr Gegenteil umwandelt. Dementsprechend ist die Antwort falsch.</p>
---	--

<p>[Nr.] 8.2</p> <p>S: Guten Tag [...]. Wie lautet Ihre Anfrage?</p> <p>B: Ich möchte heute chinesisches essen gehen.</p> <p>S: In welchem Stadtteil möchten Sie essen gehen?</p> <p>B: Egal.</p> <p>S: Es gibt mehr als drei passende Restaurants auf Ihre Anfrage. Möchten Sie Ihre Frage präzisieren?</p> <p>B: <u>Innenstadt oder Querenburg.</u></p> <p>S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p><u>Verletzung der GG 4</u></p> <p>Das System erwartet bei der Frage nach Modifikationen ein „Ja“ oder ein „Nein“. Der Benutzer weiß es nicht und stellt sofort seine Anfrage.</p> <p>Statt den Benutzer aufzufordern, seine Eingaben entsprechend zu modifizieren, kündigt es an, es gäbe kein passendes Restaurant. Diese Aussage widerspricht dem Informationsinhalt seiner Datenbank und ist insofern als falsch zu beurteilen.</p>
---	--

Interaction Aspect 3: Relevance

Eine relevante Information bezieht sich unmittelbar auf das laufende *Subtask*. Die Interaktionsdesigner müssen daher bei der Modellierung der Dialogstrategie irrelevante und inkohärente Systemäußerungen vermeiden.

GC 5. Be relevant

Relevanzmängel können durch Spracherkennungsfehler verursacht werden und führen meistens zu meta-kommunikativen Aufklärungsfragen. Komplizierter wird es, wenn die Spracherkennungsfehler unentdeckt bleiben, und die Systemantworten anhand von falschen Daten ermittelt werden.

Irrelevante Äußerungen ergeben sich auch durch Verletzung des Kohärenzprinzips¹⁵ und können den Benutzer verwirren, den Dialog durch unnötige *Turns* verlängern und den gesamten Kommunikationserfolg erheblich beeinträchtigen.

<p>[Nr.] 1.1</p> <p>S: Guten Tag, Guten Tag willkommen bei BoRIS, dem Bochumer Informationssystem. BoRIS erlaubt es Ihnen eine Recherche nach den folgenden Kriterien durchzuführen: der Art der Küche, der Tageszeit oder der Preis der Mahlzeit, dem ungefähren Ort des Restaurants oder nach dessen Öffnungszeiten. Wie lautet Ihre Anfrage?</p> <p>B: Nach dem Ort.</p> <p>S: Sie haben folgende Auswahlmöglichkeiten: Stadtmitte, <u>Unicenter</u>, andere Möglichkeiten oder keine Präferenz.</p> <p>B: <u>Uni-Center.</u></p> <p>S: <u>[hors contexte.]</u></p> <p>Vom derzeitigen System können nur die folgenden</p>	<p><u>Verletzung der GG5.</u></p> <p>Das Keyword „Uni-Center“ wird vom System nicht erkannt, was auf einen Tippfehler des Versuchsleiters¹⁶ zurückzuführen ist. Die Benutzereingabe wird als irrelevant (<i>hors_contexte</i> [franz.]) betrachtet und das System bricht den Kommunikationsfluss ab und kehrt zum Gesprächsanfang zurück.</p> <p>Dabei überspringt das System wichtige Dialogschritte einer konformen Dialogstrategie:</p> <ol style="list-style-type: none"> 1. Der Abbruch des Kommunikationsflusses wird dem Benutzer nicht erklärt. 2. Das System fordert, obwohl es die irrelevante Äußerung erkennt den Benutzer nicht auf, seine Eingabe zu wiederholen, sondern sie neu zu
--	--

¹⁵ Das Kohärenzprinzip besagt, dass in einem Dialog die Antworten an die Elemente der davor stehenden Frage anknüpfen müssen.

¹⁶ Um Spracherkennungsfehler zu vermeiden, wurde im Laufe der Experimente mit **BoRIS** die Spracherkennung ausgeschaltet. Die Benutzereingaben wurden von der Versuchsleiterin eingetippt (siehe Abschnitt 5.2)

<p>Informationen bearbeitet werden: die Art der Küche, die Tageszeit und der Preis der Mahlzeit, der ungefähren Ort des Restaurants oder dessen Öffnungszeiten. Wie lautet Ihre Anfrage?</p>	<p>formulieren. Das Kohärenzprinzip wird hier dadurch verletzt, dass die Systemantwort nicht explizit an die Benutzereingabe knüpft. VORSCHLAG: das System sollte kohärenter Weise signalisieren, dass es die Eingabe „Uni-Center“ nicht verstanden hat und nach einer Wiederholung fragen.</p>
<p>[Nr.] 3.4 S: Guten Tag willkommen bei BoRIS, dem Bochumer Informationssystem. BoRIS erlaubt es Ihnen eine Recherche nach den folgenden Kriterien durchzuführen: der Art der Küche, der Tageszeit oder der Preis der Mahlzeit, dem ungefähren Ort des Restaurants oder nach dessen Öffnungszeiten. Wie lautet Ihre Anfrage? B: Tageszeit. S:[<u>hors contexte</u>] Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden [...]. B: Tageszeit. S: Suchen Sie ein Restaurant für das Mittagessen oder für das Abendessen?</p>	<p><u>Verletzung der GG5.</u> Das System schlägt ein Keyword („Tageszeit“) vor, das es im nachhinein, überraschenderweise nicht erkennen kann. Dadurch dass die Eingabe erst beim zweiten Versuch erkannt wird, kann man die erste Systemreaktion als irrelevant und inkohärent bezeichnen.</p>

Interaction Aspect 4: Manner

Der vierte Interaktionsaspekt bezieht sich auf den Ausdruck der Dialogäußerungen. Die Verletzung der Prinzipien, die unter diesem Aspekt zusammengefasst werden, ergibt sich meistens durch ungeschickte Formulierungen.

GG 6. Avoid obscurity

Avoid obscurity of expression

Undeutliche Äußerungen sollen sowohl vom System als auch vom Benutzer vermieden werden. Sie sorgen für Verwirrung und können zu meta-kommunikativen Fragen nach Aufklärungen führen. Die „außerplanmäßigen“ meta-kommunikativen Fragen können für das System „unerwartete“ Begriffe enthalten, die es nicht verstehen kann. Solche Situationen generieren weitere Missverständnisse, die am Ende den Dialog zum Scheitern bringen können.

<p>[Nr.] 3.6 S: Vom derzeitigen System können nur die folgende Informationen bearbeitet werden [...] B: Art der Küche.</p>	<p><u>Verletzung der GG 6</u> Der Benutzer möchte eine Art der Küche eingeben, die vom System nicht genannt wurde. Er verwendet den Ausdruck „gar nicht“ um zu sagen, dass er</p>
--	---

<p>S: Sie können zwischen den folgenden Möglichkeiten wählen: italienische, griechische, chinesische oder keine Präferenz.</p> <p>B: Gar nicht.</p> <p>S: [analyse de reponse neutral]</p> <p>Suchen Sie ein Restaurant für das Mittagessen oder für das Abendessen?</p> <p>B: Abendessen.</p>	<p>keine der angegebenen Küchenarten möchte. Das System interpretiert fälschlicherweise „gar nicht“ als „keine Präferenz“. Die ungeschickte Benutzerformulierung verursacht eine Verletzung des GG 6 Prinzips und führt zu einem unentdeckten Missverständnis, welches erst im späteren Dialogverlauf für Verwirrung sorgt.</p>
---	---

GG 7 Avoid ambiguity

Zweideutige Äußerungen müssen zur Gewährleistung eines reibungslosen Dialogablaufs vermieden werden. Zweideutigkeit kann dazu führen, dass der Benutzer sich an einer falsch verstandenen Information orientiert und dementsprechend seine Äußerungen gestaltet, was den Erfolg der Interaktion gefährden würde.

<p>[Nr.] 6.3</p> <p>S: Möchten Sie noch weitere Antworten hören?</p> <p>B: Wiederholen Sie bitte!</p> <p>S: Es gibt noch andere Antworten. Möchten Sie sie hören?</p> <p>B: Ich möchte den Preis modifizieren.</p> <p>S: Vielen Dank, dass Sie BoRIS benutzt haben. Auf Wiederhören und einen schönen Tag</p>	<p>Verletzung der GG 7</p> <p>Der Benutzer möchte offensichtlich zu einem früheren Menü zurückkehren, wird aber vom System nicht verstanden. Das Missverständnis entsteht dadurch, dass der Benutzer sein Wunsch nicht eindeutig genug formuliert. Der Ausdruck „Wiederholen Sie bitte!“ wird vom System als Aufforderung zur Wiederholung des letzten Satzes interpretiert, was schließlich zum Scheitern des Dialoges führt.</p>
--	---

SG3 Ensure uniformity

Provide the same formulation of the same question (or address) to users everywhere in the system's interaction turns.

Um zweideutige Äußerungen vermeiden zu können, empfiehlt sich dieselbe Formulierung für gleiche Sachverhalte zu verwenden. Dabei wird der potentielle Wortschatz sowohl des Systems, als auch des Benutzers¹⁷ eingeschränkt und die Gefahr möglicher Erkennungsfehler verringert.

Bei **BoRIS** konnte keine Verletzung dieses Prinzips festgestellt werden.

GG 8 Be brief (avoid unnecessary prolixity)

Langatmige Dialogsequenzen können den Benutzer langweilen und seine Aufmerksamkeit

¹⁷ Dadurch, dass die Benutzer dazu tendieren, die Systemäußerungen zu imitieren, wird auch ihr Wortschatz dementsprechend eingeschränkt.

ablenken. Die Verletzung dieser Richtlinie bereitet keine ernsthaften Interaktionsprobleme und wird erst nach den Experimenten durch die Kommentare der Benutzer signalisiert.

<p>[Nr.] 3.1 S: Guten Tag willkommen bei BoRIS, dem Bochumer Informationssystem. BoRIS erlaubt Ihnen eine Recherche nach den folgenden Kriterien durchzuführen: der Art der Küche, der Tageszeit oder dem Preis der Mahlzeit, dem ungefähren Ort des Restaurants oder nach dessen Öffnungszeiten. Wie lautet Ihre Anfrage? B: <u>Art des Restaurants.</u> S: <u>[hors context]</u> Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden [...]</p> <p>[Nr.] 8.6 S: Guten Abend [...] Wie lautet Ihre Anfrage? B: <u>Art.</u> S: <u>[hors contexte]</u> Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden [...]</p>	<p><u>Verletzung des GP 8</u> Die Systemäußerungen sind zu lang, so dass die Benutzer sich an die genaue Bezeichnung jedes Auswahlkriteriums nicht mehr erinnern können. Sie versuchen ähnliche Ausdrücke (z.B. „Art des Restaurants“ oder „Art“) für die Gestaltung ihrer Anfrage zu finden, die aber vom System abgelehnt werden. VORSCHLAG: Sinnvoller wäre es den Benutzern nur ein einziges Auswahlkriterium zu bieten (z.B. „Art der Küche“) und sie anschließend nach ihren Präferenzen bezüglich des Preises, des Ortes, der Tageszeit usw. zu befragen.</p>
---	---

GG 9. Be orderly

Das System sollte den Dialogablauf erwartungsgemäß strukturieren. Wenn der Benutzer einen thematischen *Subtask* im Dialogverlauf früher erwartet, als dieser tatsächlich kommt, wird er wahrscheinlich die Dialoginitiative ergreifen und dem System nach den fehlenden Informationsteilen fragen. Die Verletzung dieses Prinzips kann durch die Implementierung einer erwartungsgemäßen Dialogstrategie vermieden werden, die man anhand einer gründlichen Analyse zwischenmenschlicher Dialoge ermitteln kann.

<p>[Nr.] 5.3 S: Guten Tag [...] Wie lautet Ihre Anfrage? B: <u>Arabisch.</u> S: <u>Es gibt mehrere auf Ihre Anfrage passende Restaurants:</u> Du Cote de chez Fred, Al Sahara, Arabesque. Vielen Dank, dass Sie BoRIS benutzt haben. Auf Wiederhören und einen schönen Tag.</p>	<p><u>Verletzung der GG 9</u> Das System bleibt in seiner eigenen Dialogstruktur nicht konsequent. Der Benutzer erwarte, aufgrund seiner Erfahrung der früheren Dialoge mit BoRIS, nach Tageszeit, Preis und Stadtteil gefragt zu werden. Statt dessen gibt das System nach dem ersten eingegebenen Keyword alle arabischen Restaurants an, die es in der Datenbank hat, ohne den Benutzer nach den üblichen Kriterien zu fragen.</p>
--	--

Interaction Aspect 5: Partner asymmetry

Mit dem Spracherwerb lernen wir gleichzeitig die Charakteristika eines sogenannten „Standardpartners“ kennen. Im Laufe unserer Sozialisation treffen wir viele Menschen, mit denen wir ins Gespräch kommen. Bei jedem Gespräch wird dann unbewusst überprüft, inwieweit der Gesprächspartner mit dem Standardmuster übereinstimmt. Wenn der Gesprächspartner dem erwarteten „Standard“ nicht entspricht, versuchen wir uns ihm anzupassen: so z.B. reden wir mit einem Kind anders als mit einem Erwachsenen oder wir werden plötzlich leiser sprechen, wenn wir mit jemandem telefonieren, der selber leise spricht.

Der Computer stellt, in diesem Zusammenhang, einen ungewöhnlichen Gesprächspartner dar, dem man sich anpassen muss, um erfolgreich kommunizieren zu können.

SP 10 Highlight asymmetries

Inform interaction partners of important non-normal characteristics which they should take into account in order to behave cooperatively during interaction. Ensure the feasibility of what is required of them.

Es ist eine wichtige und komplexe Richtlinie, die signalisiert, dass es zwischen den Gesprächspartnern wichtige Unterschiede gibt, die den Gesprächsablauf und den eventuellen Erfolg der Interaktion beeinflussen können. Der Benutzer muss im Kauf nehmen, dass er sich in einer ungewöhnlichen Dialogsituation mit einem außerordentlichen Dialogpartner befinden wird.

<p>[Nr.] 25.2 S: Wie viel möchten Sie ungefähr pro Person ausgeben? B: Ich weiß es nicht. S: In welchem Stadtteil möchten Sie essen gehen? B: <u>In Essen</u>. Ich kenne mich da nicht aus. S: Es gibt mehr als drei passende Restaurants auf Ihre Anfrage. Möchten Sie Ihre Frage präzisieren?</p>	<p><u>Verletzung der GG 10</u> Der Benutzer hat nicht gemerkt, dass BoRIS Informationen über Bochumer Restaurants liefert. Außerdem versucht er sich mit dem System zu unterhalten, wie in einem normalen Gespräch mit einem Menschen. Er scheint sowohl über den Aufgabenbereich als auch über die Grenzen des Systems nicht informiert zu sein. VORSCHLAG: Das System sollte über diese zwei Aspekte Klarheit schaffen., bevor es zu weiteren Fragen kommt</p>
--	--

SG4 State your capabilities

Provide clear and comprehensible communication of what the system can do and cannot do.

Der Benutzer soll informiert werden, welche Art von Informationen das System liefern kann und, vor allem, wo die Systemgrenzen liegen.

<p>[Nr.] 3.6</p> <p>S: Guten Tag [...]Wie lautet Ihre Anfrage?</p> <p>B: Ort</p> <p>B: Sie haben folgende Auswahlmöglichkeiten: Stadtmitte, Unicenter, andere Möglichkeiten oder keine Präferenz.</p> <p>S: Andere Möglichkeiten.</p> <p>B: Die anderen Möglichkeiten sind am Schloss, am Stadion, am Amphitheater oder an der Burg.</p> <p>B: <u>Keine von denen.</u></p> <p>S: <u>[hors contexte]</u> Vom derzeitigen System können nur die folgenden Informationen</p>	<p><u>Verletzung der SG4</u></p> <p>Der Benutzer wird nicht informiert, dass die Datenbank über mehr Information verfügt, als das System im Menü angibt. Die Aufzählung einer geringen Anzahl von Restaurantsorten wird ihn verwirren. Er versucht dem System zu sagen, dass er keine der angegebenen Möglichkeiten auswählen möchte.</p> <p>VORSCHLAG: das System sollte an dieser Stelle entweder die ganze Liste der Restaurantsorte angeben oder dem Benutzer klarmachen, dass es auch die Eingabe anderer Restaurantsorte erlaubt, die im Vorfeld vom System nicht genannt wurden.</p>
---	--

SG 5. State how to interact

Provide clear and sufficient instructions to users on how to interact with the system.

Dialogsysteme sind meistens einem menschlichem Interaktionspartner weit unterlegen. Dadurch ergibt sich ein asymmetrisches Verhältnis, welches man bewusst wahrnehmen muss, um den Interaktionserfolg zu sichern. Aus diesem Grund sollte der Benutzer über alle unstandardisierten Vorgehensweisen des Systems informiert werden.

Obwohl von grundlegender Bedeutung wird dieses Prinzip oft bei der Dialogmodellierung vernachlässigt, vor allem, weil eine komplette „Gebrauchsanweisung“ zeitspielig und schwer realisierbar ist.

<p>[Nr.] 3.2</p> <p>B: Ich möchte chinesisches essen gehen.</p> <p>S: Suchen Sie ein Restaurant für das Mittagessen oder für das Abendessen?</p> <p>B: Abendessen.</p> <p>S: An welchem Tag beabsichtigen Sie essen zu gehen?</p> <p>B: Am Montag.</p> <p>S: Wie viel möchten Sie ungefähr pro Person ausgeben?</p> <p>B: <u>50 DM.</u></p> <p>S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p> <p>B: Ja.</p> <p>S: Sie können die Art der Küche[...]. Geben Sie bitte Ihre Änderung an.</p> <p>B: <u>Preis.</u></p> <p>S: Sie haben folgende Auswahlmöglichkeiten: untere, mittlere, obere Preisklasse oder keine Präferenz.</p> <p>B: <u>Keine Präferenz.</u></p> <p>S: Es gibt kein passendes Restaurant auf Ihre Anfrage.</p>	<p><u>Verletzung der SG 5</u></p> <p>Es ist nicht klar, wie das System das Keyword „keine Präferenz“ interpretiert: als einen globalen Wert, bei welchem alle Preisklassen in Frage kommen oder als neutrale Eingabe, welche einen Modifikationswert gleich Null setzt, bzw. den vorher eingegebenen Wert (50 DM) nicht ändert. Die zweite Variante scheint in diesem Fall plausibler zu sein: das System übernimmt beim Einsatz des Keywords „keine Präferenz“ automatisch den vorher eingegebenen Wert und kann dem entsprechend kein passendes Restaurant finden. Der Benutzer geht von der ersten Variante aus und versucht durch die Anwendung des Keywords „keine Präferenz“ an die Liste aller möglichen chinesischen Restaurants zu gelangen.</p> <p>Die verschiedenen Keywordinterpretationen</p>
--	---

<p>Möchten Sie die Anfrage modifizieren? B: Ja. S: Sie können die Art der Küche [...]. Geben Sie bitte Ihre Änderung an.[...]. B: Nach dem Ort. S: Sie haben folgende Auswahlmöglichkeiten, Stadtmitte, Unicenter [...]. B: <u>Keine Präferenz.</u> S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p>führen schließlich zum scheitern des Dialogs. VORSCHLAG: Solche Konfliktsituationen können vermieden werden, indem man dem Benutzer alle Bedeutungen der Keywords erklärt, die von der üblichen, logischen Interpretation abweichen.</p>
<p>[Nr.] 1.1 S: Guten Abend, [...]. BoRIS erlaubt es Ihnen eine Recherche nach den folgenden Kriterien durchzuführen: der Art der Küche, der Tageszeit oder der Preis der Mahlzeit, dem ungefähren Ort des Restaurants oder nach dessen Öffnungszeiten. Wie lautet Ihre Anfrage B: <u>Wo kann ich essen gehen?</u> S: <u>[hors contexte]</u> Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden [...]. B: Preis.</p> <p>[Nr.] 27.1 S: Guten Tag [...].Wie lautet ihrer Anfrage? B: <u>Die Speisekarte hätte ich bitte.</u> S: Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden [..].</p>	<p><u>SG 5</u> Der Benutzer wird unterrichtet, nach welchen Kriterien man eine Recherche bei BoRIS durchführen kann. Er ignoriert diesen Hinweis und versucht offene Fragen zu stellen. Das System reagiert richtig und weist noch einmal darauf hin, wie man seine Äußerungen zu gestalten hat.</p>

Interaction Aspect 6: Background Knowledge

Bei der Dialogmodellierung soll berücksichtigt werden, dass Benutzer unterschiedliches Hintergrundwissen besitzen. Dabei soll zwischen erfahrenen und unerfahrenen Benutzern unterschieden werden.

GG 11 Be aware of users' background knowledge.

Take user's relevant background knowledge into account.

Die Interaktionsmodellierung sollte das Hintergrundwissen potentieller Benutzer berücksichtigen.

<p>[Nr.] 21.2 S: Guten Tag [...] Wie lautet Ihre Anfrage? B: Art der Küche. S: Sie können zwischen den folgenden Möglichkeiten wählen [...].</p>	<p><u>Verletzung der GG 11</u> Der Benutzer weißt wahrscheinlich nicht, dass chinesische Restaurants allgemein preiswert sind und fragt nach einem teureren chinesischen Restaurant. Das System hat in seiner Datenbank nur chinesische</p>
---	---

<p>B: Chinesische Küche. S: Wie viel möchten Sie ungefähr pro Person ausgeben? B: <u>Teuer</u>. S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren?</p>	<p>Restaurant mittlerer oder unteren Preisklasse und findet daher kein passendes Restaurant. VORSCHLAG: Die Modellierung der Dialogstrategie sollte dem System erlauben, sich kooperativ zu verhalten und dem Benutzer darauf hinzuweisen, dass solche Preise für die Küchenkategorie untypisch sind. Anschließend sollte das System dem Benutzer eine Alternative vorschlagen, z.B. ein chinesisches Restaurant mittlerer Preisklasse.</p>
---	--

SG 6 Be aware of user inferences

Take into account possible (and possibly erroneous) user inferences by analogy from related task domains.

Bei der Implementierung der Dialogstrategie muss damit gerechnet werden, dass der Benutzer, durch Analogien thematische Verbindungen herstellen könnte, die sich außerhalb des aufgabenorientierten Kontextes befinden..

<p>[Nr.] 4.5 S: Guten Tag [...]. B: Ich möchte in eine <u>Cocktailbar</u> gehen. S: Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden[...] B: Lateinamerikanische Küche. S: Hier ist das passende Restaurant auf Ihre Anfrage: Havanna.</p>	<p><u>Verletzung der SG 6</u> Der Benutzer stellt aufgrund semantischer Verwandtschaft eine synonymische Beziehung zwischen den Begriffen „Cocktailbar“ und „lateinamerikanisch“ her und ordnet analogischerweise den Begriff „Cocktailbar“ als Unterkategorie dem Keyword „Art der Küche“ zu. Das System kann diesen Analogievorgang nicht nachvollziehen, dadurch, dass es in seinem Lexikon den Begriff nicht gibt. Alle vorhandenen Gaststättentypen werden unter dem Begriff „Restaurant“ zusammenfasst und weiter nach Küchenarten geordnet. Aus diesem Grund erklärt sich, warum das System erst auf die Küchenart „lateinamerikanisch“ reagiert. VORSCHLAG: Um die Verletzung dieses Prinzips zu vermeiden, könnte man im Systemlexikon Synonyme zum Begriff „Restaurant“ einfügen.</p>
--	--

SG7 Adapt to novice and experts

Separate whenever possible between the needs of novice and expert users (user-adaptive interaction).

Erfahrene Benutzer haben andere kommunikative Bedürfnisse als unerfahrenen Benutzer. Die Erfahrene kennen die Systemvorgehensweise und wissen durch welche Befehle sie am schnellsten ans Ziel kommen. Sie bevorzugen knappe, implizite Bestätigungsstrategien und stichwortartige Eingaben.

Dagegen können Benutzer in der Anfangsphase nur bedingt mit dem System umgehen. Nach [Kamm, 1997] bevorzugen sie immer wiederkehrende, explizite Bestätigungen, um

sicher zu sein, dass der Dialog in die gewünschte Richtung abläuft und sind oft auf die Systemhilfe angewiesen.

Wenn die Gestaltung der Dialogstrategie sich nach erfahrenen Benutzern orientiert werden unerfahrene Benutzer wahrscheinlich nicht damit klarkommen; sollte sie dagegen die Bedürfnisse der Anfänger in erster Reihe ansprechen wird sie den erfahrenen Benutzer wahrscheinlich schwerfällig und redundant erscheinen [Kamm, 1997]. Es muss einen Weg gefunden werden, der für beide Benutzerkategorien geeignet ist. Man könnte z.B. eine Sonderoption im Anfangsmenü für Systemanfänger implementieren. Diese Option sollte man nur nach Wunsch aktivieren und eine kurze, jedoch sinnvolle Anleitung der Systemvorgehensweise für unerfahrene Benutzer bieten.

BoRIS unterscheidet nicht zwischen erfahrenen und unerfahrenen Benutzer.

GG12 Be aware of user expectations

Take into account legitimate user expectation as to your own background knowledge.

Bei einer gut modellierten Dialogstrategie wird das Hintergrundwissen der Benutzer antizipiert und entsprechend implementiert.

<p>[Nr.] 8.6 S: Guten Tag, [...] Wie lautet Ihre Anfrage? B: Können Sie die <u>Adresse von Coco Loco wiederholen?</u> S: Sie haben folgende Auswahlmöglichkeiten [...]. B: Die Adresse von Coco Loco in Weitmar, hätte ich gern. S: Suchen Sie ein Restaurant für das Mittagessen oder für das Abendessen?</p>	<p><u>Verletzung der GG 12</u> Der Benutzer hat vom System eine Restaurantadresse erhalten, die er nicht verstanden hat. Er telefoniert erneut mit BoRIS und bittet nach einer Wiederholung der Adressenansage. Das Systems kann aufgrund seiner Dialogstruktur diese Aufgabe nicht erfüllen und der Benutzer muss das ganze Telefonat noch einmal durchführen. Der Benutzer weiß, dass das System die Adresse in seiner Datenbank hat und erwartet, dass es diese auch direkt ausgeben kann, ohne den ganzen Dialog durchlaufen zu müssen. VORSCHLAG: Eine Modifizierung der Dialogstrategie wäre in diesem Fall angebracht.</p>
--	--

SG8 Cover the domain

Provide sufficient task domain knowledge and inference.

Ein interaktives Sprachsystem sollte in seiner Datenbank nicht nur ausreichende Informationen haben, sondern auch über Allgemeinwissen verfügen, das es ihm ermöglicht logische Schlussfolgerungen innerhalb desselben Themenbereichs zu ziehen (z.B. dass ein Rückflugdatum immer nach dem Hinflugdatum stehen soll).

<p>[Nr.] 3.4 S: An welchem Tag beabsichtigen Sie essen zu gehen? B: <u>Heute.</u> S: [<u>Friday</u>] Wie viel möchten Sie ungefähr pro Person ausgeben?</p>	<p><u>SG8</u> Das System verfügt über ein Kalender, wonach das System ein Datum mit einem Wochentag gleichsetzen kann; in diesem Fall wird der Begriff „heute“ mit dem laufenden Wochentag Freitag korrekt identifiziert.</p>
---	---

Interaction Aspect 7: Repair and Clarification

Dieser Aspekt bezieht sich auf die Notwendigkeit der Implementierung meta-kommunikativer Strategien in einem Dialogsystem. Meta-kommunikative Strategien beziehen sich auf Aufklärungen, Korrekturen und Wiederholungen, die man gelegentlich in einen Dialog zur Beseitigung von Missverständnissen einführt. Ihr Einsatz ist auf verschiedene Ursachen zurückzuführen (z.B. auf die Verletzung eines Kommunikationsprinzips, auf die sinkende oder abgelenkte Aufmerksamkeit des Benutzers, auf Verständigungsprobleme aufgrund von widersprüchlichen oder zweideutigen Inputs). Meta-kommunikative Strategien sollte man nicht im Übermaß anwenden, um die Effizienz der Interaktion nicht zu beeinträchtigen.

GG13 Enable meta-communication

Enable repair or clarification meta-communication in case of communication failure.

Wenn der Benutzer feststellt, dass seine Aussagen missverstanden wurden, wird er versuchen eine meta-kommunikative Strategie zur Korrektur einzuführen.

<p>[Nr.]2.6 S: Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden: die Art der Küche, die Tageszeit und der Preis der Mahlzeit, der ungefähren Ort des Restaurants oder dessen Öffnungszeiten. Geben Sie bitte Ihre Änderung ein. B: <u>Wiederholen Sie bitte!</u> S: Sie können eine Anfrage durchführen nach der Art der Küche, nach der Tageszeit oder dem Preis der Mahlzeit, nach dem ungefähren Ort des Restaurants oder nach seinen Öffnungszeiten. Wie lautet Ihre Anfrage? B: Ort.</p>	<p><u>GC 13</u> Der Benutzer kann sich nicht an alle Modifikationskriterien erinnern und fordert das System auf, ihm diese Kriterien noch einmal aufzulisten.</p>
---	---

SG 9 Enable system repair

Initiate repair meta-communication if system understanding has failed.

Wenn das System den Benutzerinput nicht in einem sinnvollen Kontext interpretieren kann, sollte es dies dem Benutzer sofort signalisieren und eine Korrekturstrategie einführen.

<p>[Nr.] 2.4 S: Guten Tag [...]Wie lautet Ihre Anfrage? B: Es ist 12. 00 Uhr, und ich möchte preiswert essen, deutsch. S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren? B: <u>Ich möchte dann mehr zahlen.</u> S: <u>[pas de reponse transitions]</u> Vielen Dank, dass Sie BoRIS benutzt haben. Auf Wiederhören und einen schönen Tag.</p> <p>[Nr.] 6.3 S: Guten Tag [...]Wie lautet Ihre Anfrage? B: Ich suche ein arabisches Restaurant in Altenbochum. S: Es gibt kein passendes Restaurant auf Ihre Anfrage. Möchten Sie die Anfrage modifizieren? B: <u>Vielleicht in Grumme</u> S: <u>[pas de reponse transitions]</u> [M35] Vielen Dank, dass Sie BoRIS benutzt haben. Auf Wiederhören und einen schönen Tag.</p>	<p><u>Verletzung der SG 9</u> Das System hat die Äußerungen „ich möchte mehr zahlen“ (Nr. 2.4) oder „Vielleicht in Grumme“ (Nr. 6.3) nicht verstanden, denn an dieser Stelle, laut Struktur der Dialogstrategie, war ein „Ja“ oder ein „Nein“ als Antwortmöglichkeit vorgesehen. Das System bricht den Dialog ab, ohne dem Benutzer zu signalisieren, dass es seine Eingaben nicht verstanden hat und ohne versucht zu haben, eine meta-kommunikative Strategie einzuführen, was zur Prinzipverletzung führt.</p>
<p>[Nr.] 34.4 S: Sie haben folgende Auswahlmöglichkeiten: untere, mittlere und obere Preisklasse. B: <u>Mittlere Preisklasse</u> S: <u>Könnten Sie bitte den gewünschten Preis der Mahlzeit wiederholen?</u> B: Zwischen 25 und 30 DM. S: Hier ist das passende Restaurant auf Ihre Anfrage: Alt Nürnberg.</p>	<p><u>GG 13</u> Das System hat die Benutzeraussagen nicht verstanden und führt eine meta-kommunikative Strategie ein, indem es den Benutzer auffordert, seine Aussage zu modifizieren.</p>

SG 10 Enable inconsistency clarification

Initiate clarification meta-communication in case of inconsistent user input.

Wenn der Benutzerinput widersprüchlich ist, sollte das System nicht versuchen, selbst die Äußerung zu deuten, sondern direkt nach einer Aufklärung fragen.

<p>[Nr.] 34.1 S: Vom derzeitigen System können nur die folgenden Informationen bearbeitet werden [...]. Stelle Sie bitte Ihre Frage? B: <u>Öffnungszeiten 10 00 bis 23 00.</u> S: <u>[cheap,average]</u> S: Könnten Sie bitte den gewünschten Preis der Mahlzeit wiederholen?</p>	<p><u>SG 10</u> Der Benutzer macht eine für das System widersprüchliche Äußerung: er nennt sowohl das Keyword (Öffnungszeiten) ein, als auch die entsprechenden Zeiten mit Zahleneingaben. Die Zahlen werden vom System als Preiseingaben interpretiert. In diesem Fall signalisiert das System indirekt, dass es die Eingaben nicht richtig verstanden hat, so dass der Benutzer rechtzeitig mitbekommt, dass das System seine Eingaben falsch verstanden hat.</p>
--	---

SG11 Enable ambiguity clarification

Initiate clarification meta-communication in case of ambiguous user input.

Im Falle zweideutiger Benutzeräußerung sollte das System erst nach Aufklärungen fragen und nicht versuchen, selber den Inhalt zu deuten, denn eine falsche Interpretation könnte den Erfolg der Interaktion gefährden.

<p>[Nr.] 6.2 S: Wie viel möchten Sie ungefähr pro Person ausgeben? B: <u>30 DM, nicht mehr als 30DM.</u> S: <u>[analyse de reponse: neutral]</u>_reponse_ok [M45] In welchem Stadtteil möchten Sie essen gehen? B: Innenstadt.</p>	<p><u>Verletzung der SG 11</u> Die Äußerung „30 DM, nicht mehr als 30 DM“ ist für das System zweideutig: es kann einen Wert gleich 30 bedeuten oder einen, der weniger als 30 ist. Das System interpretiert diesen Wert als „neutral“; dies ist ein zusätzlicher Fehler, denn „neutral“ soviel wie „keine Präferenz“ bedeutet. Das Missverständnis bleibt dabei unentdeckt und wird später für ein unzufrieden stellendes Ergebnis sorgen.</p>
---	---

4.4 Gestaltung der Qualitäts- und Zufriedenheitsfaktoren

Der folgende Abschnitt soll als Entwurf verstanden werden, der aufgrund des bisher präsentierten Wissenstandes in dieser Arbeit versucht, ein auf verschiedenen theoretischen Ansätzen beruhendes Qualitäts- und Zufriedenheitsfaktorenschema zusammenzustellen, nach dem die Dialogsteuerung qualitativ bewertet werden soll. Da es keine kommunikations-wissenschaftlichen Evaluierungsmodelle zu diesem Thema gibt, stellt die Autorin ein selbst entwickeltes Untersuchungsschema nach bekannten Evaluierungskriterien zusammen.

Mit dieses Schema wird keinen Anspruch auf Vollständigkeit erhoben: es stellt nur einen ersten Schritt auf der Suche nach einem geeigneten theoretischen Hintergrund dar.

Wenn die Dialogsteuerung ihr Ziel optimal erreicht hat, d. h. wenn schnell, einfach, fehlerlos, verständlich und vollständig eine Information ermittelbar ist, dann kann der Dialog mit der Maschine als qualitativ hochwertig bewertet werden. Als Gesprächspartnermuster dient hier der Mensch.

Welche Standardnormen müssen eingehalten werden, um diese Ansprüche zu erfüllen? Eine Antwort auf diese Fragen bieten die im Abschnitt 4.3 dargestellten Aspekte der Kommunikationsprinzipien: sie sollten für den Erfolg eines Dialogs ausschlaggebend sein. Bei der Betrachtung der Kommunikationsprinzipien zeigt sich die Kooperativität ein zentraler Indikator für die Qualitätsbeurteilung einer Dialogsteuerung. Das Ausmaß der Kooperativität gibt an, inwieweit das System sich einem menschlichen Gesprächspartner anpasst. Die Operationalisierung der Kooperativität erfolgt nach aufgrund sieben heterogener Aspekte, welche die Grundlage für die qualitative Evaluierung einer Dialogsteuerung aus kommunikationswissenschaftlicher Sicht bilden.

Die Untersuchungen, die man im Rahmen dieser Arbeit durchgeführt hat, versuchen aufgrund dieser Prinzipien die Analyse der Kooperativität zu erweitern. Nach Ansicht der Autorin kann die Kooperativität mit Hilfe von 11 Faktoren analysiert werden, die inhaltlich in drei Untersuchungsebenen eingeteilt werden können: Frage-Antwort Aspekte, Systemmöglichkeiten und Zufriedenheit der Benutzer. Die ersten zwei Ebenen setzen sich aus Kriterien zusammen, die einerseits von den Aspekten der im Abschnitt 4.3 dargestellten Kommunikationsprinzipien [Bernsen et. al., 1998] stammen und andererseits in Anlehnung an die Benutzerfreundlichkeitskriterien von [Love et. al., 1994] entstanden sind. Die auf dieser Ebene gezählten Kriterien werden im Rahmen dieser Untersuchung qualitätsbeeinflussende Faktoren, oder kürzer Qualitätsfaktoren, genannt. Auf der Frage/Antwort Ebene befinden sich Qualitätsfaktoren, die sich auf Merkmale der Äußerungen beziehen. Darunter werden folgende Faktoren gezählt: Relevanz, Ausdruckform und Korrektur- und Aufklärungsstrategien. Die zweite Ebene enthält Qualitätsfaktoren, die sich auf Systemmöglichkeiten beziehen, die eine qualitativ hochwertige Dialogsteuerung sichern sollen, wie z.B. Transparenz, Berücksichtigung des Hintergrundwissens und der Erwartungen der Versuchspersonen, Dialoginitiative, Schnelligkeit und Hilfefunktion.

Die dritte Ebene untersucht Benutzerfreundlichkeitskriterien, die sich auf die allgemeine Zufriedenheit der Benutzer dem System gegenüber beziehen. Dabei werden folgende Kriterien angesprochen: die Erreichung des Informationswunsches, das System als Gesprächspartner und das Schwierigkeitsgrad des Dialogverlaufs. Diese Kriterien werden im weiteren Zufriedenheitsfaktoren genannt.

Die Untersuchungsschema der Faktoren wurde nach der qualitativen Befragung teilweise modifiziert: die Versuchspersonen sprachen Kriterien an, die sich außerhalb des ersten

Faktorenschemaentwurfs befanden. So wurde die Untersuchungsliste um folgende Faktoren erweitert: Gestaltung der Äußerungen (2.3), Wiederholungen (3.1), Möglichkeit das Gespräch zu führen, solange man Fragen hat (6.2), Hilfefunktion (8.1) und Einschätzung des Gesprächsaufbaus (11).

Die Unterteilung der Untersuchungskriterien in Faktoren und Unterfaktoren erwies sich aufgrund ihres inhaltlichen Kohäsionsgrades als schwer: die aus der Literatur entnommenen Aspekte der Kommunikationsprinzipien und die Benutzerfreundlichkeitskriterien überschneiden sich teilweise. Sie mussten aneinander angepasst und entsprechend neustrukturiert werden, um ein einheitliches Untersuchungsschema zu ergeben. Außerdem sind die Qualitätsfaktoren mit den Zufriedenheitsfaktoren in einer Ursache/Effekt-Kette korreliert: die Qualitätsfaktoren liefern die Gründe, warum die Benutzer mit dem System, zufrieden oder weniger zufrieden waren.

I. Ebenen: Frage-Antwort Aspekte

Die Relevanz/*Relevance* stellt, als dritter Interaktionsaspekt der Kommunikationsprinzipien eine zentrale inhaltliche Eigenschaft einer Dialogäußerung dar. Sie überprüft, welche Äußerungen einen passenden Beitrag zum kommunikativen Ziel leisten. Eine relevante Äußerung ist vor allem informativ. Aus diesem Grund entschloss sich die Autorin den Informationsgehalt/*Informativness*, dem Faktor Relevanz zuzuordnen. Der Informationsgehalt bezieht sich auf die Menge und die Vollständigkeit der ausgetauschten Dialogäußerungen und ist somit für die Dialogsteuerung qualitätsrelevant. Die Sprachtheorie berücksichtigt sie im *Focus Set* auf dem *Control*-Niveau. Dieser Faktor wird mit Fragen nach qualitativen und quantitativen Aspekten der Informationsmenge und nach der Angemessenheit verschiedener Äußerungen operationalisiert.

Die Ausdrucksform/*Manner* zählt zwei Unterfaktoren: Verständlichkeit und Gestaltung der Benutzereingaben. Die Verständlichkeit nimmt zwei Ausprägungen an, die sowohl sprachakustisch als auch inhaltlich-kognitiv zu verstehen sind. Die sprachakustische Verständlichkeit wird mit Fragen nach akustischer Verstehbarkeit untersucht, während die inhaltlich-kognitive sich auf Fragen nach der Eindeutigkeit der Äußerungen konzentriert. Die Sprachtheorie erfasst diesen Faktor unter *Language*- und *Speech*-Niveau.

Die Gestaltung der Benutzereingaben untersucht, wie die Benutzer ihre Eingaben formuliert haben. Dieser Unterfaktor wird nicht ausdrücklich von der Sprachtheorie behandelt, gehört aber dem *Language*-Modul an und wird unter der Kategorie der „Grammatik“ bei der *System Utterances* - Komponente berücksichtigt.

Die Geschicklichkeit, mit der Korrekturen und Aufklärungen durchgeführt werden, bildet den Kernpunkt des Faktors Korrektur- und Aufklärungsstrategien. Die Sprachtheorie lokalisiert diesen Faktor innerhalb der *Intentional Structure* auf dem *Interaction Control* - Modul. Die Korrektur- und Aufklärungsstrategien untersuchen im Rahmen dieser Arbeit zwei Unterfaktoren: Wiederholungen, sowohl der System- als auch der Benutzereingaben,

und Bestätigungen, die man mit entsprechenden Fragen nach derselben Kategorie operationalisiert. Unter Bestätigung wird hier die Bestätigung der Benutzereingaben und der Informationsverarbeitung verstanden (z.B. vor der Antwortausgabe kommt es gelegentlich zu Pausen, die den Benutzer mit einem „Bitte warten Sie einen Moment“ signalisieren sollte). Die Kommunikationsprinzipien zählen die Bestätigung unter dem Aspekt des Informationsgehaltes. Nach Meinung der Autorin gehört die Bestätigung, aufgrund ihrer aufklärenden Funktion, zu dieser Kategorie.

II. Ebene: Systemmöglichkeiten

Auf der zweiten Untersuchungsebene befinden sich die Möglichkeiten, die das System bietet, um die Kooperativität zu steigern. Sie sollen qualitativ hochwertige Interaktionsbedingungen schaffen.

Unter dem Faktor Transparenz wird der Interaktionsaspekt *Partner Asymmetry* aufgefasst. Er signalisiert, dass zwischen den Gesprächspartnern kommunikativ bezogene Unterschiede liegen, die bewusst wahrgenommen werden müssen, um reibungslos interagieren zu können. Aus diesem Grund bildet es ein unerlässliches Kriterium für die qualitative Gestaltung einer Dialogsteuerung. Die Sprachtheorie berücksichtigt diesen Faktor auf dem letzten Niveau unter der *Influencing User Behaviour*-Komponente. Dieser Faktor wird in zwei Unterfaktoren eingeteilt: Einfachheit der Bedienung und Systemgrenzen. Dabei wird überprüft, wie einfach es für die Benutzer war, mit dem System umzugehen, ob ihnen die Systemgrenzen bekannt waren, ob sie wussten welche Informationen das System liefern konnte und wie sie ihre Eingaben formulieren mussten.

Nicht nur *Backgroundknowledge*/Hintergrundwissen sondern auch die Erwartungen der Benutzer sind für die qualitative Beurteilung der Dialogsteuerung relevant (selbst unerfahrene Benutzer verfügen immer, ob bewusst oder unbewusst, über vorgegebene Vorstellungen, die kulturell-, sozial-, alters- oder geschlechtsbedingt sind). Diese Kriterien werden unter dem Faktor „Kontextuelle Bedingungen“ zusammengefasst. Sie variieren je nach Gesprächspartner und werden außerhalb des Systems bestimmt. Diese Faktoren werden durch direkte Fragen operationalisiert. Die Sprachtheorie berücksichtigt diese Faktoren auf dem *Context*-Niveau im User Model.

Die Dialoginitiative ist ein Merkmal der implementierten Dialogstrategie und kein Interaktionsaspekt. Die Sprachtheorie berücksichtigt diesen Faktor als eine wichtige qualitative Komponente einer Interaktion und lokalisiert ihn auf dem *System Performance* Niveau. Die Wichtigkeit der Dialoginitiative kann überprüft werden, indem man den Benutzer z.B. fragt, ob es für ihn wichtig ist, dem System Fragen stellen zu können und zu bestimmen, wann das Gespräch zu Ende gehen soll.

Die Schnelligkeit und die Implementierung einer Hilfefunktion sind Faktoren, die von technischen Komponenten abhängen und die sich außerhalb des eigentlichen Kommunikationsprozesses befinden. Somit werden sie von der Sprachtheorie nicht

mitberücksichtigt. In der Literatur [Love et al., 1994] werden sie aufgrund ihres Einflusses auf das Qualitätsurteil untersucht. Diese Unterfaktoren werden durch direkte Fragen überprüft. Die Schnelligkeit wird zusätzlich durch die Überprüfung der Unterbrechungen untersucht, die sich im Dialogverlauf ergeben können.

III Ebene: Zufriedenheit

Auf der dritten Untersuchungsebene befindet sich die Zufriedenheit. Sie ist ein Indikator für die Kooperativität und zugleich für die Gesamtqualität des Systems, indem sie Auskunft über die Benutzerfreundlichkeit des Systems gibt. Die Zufriedenheit liefert indirekt aussagekräftige Hinweise auf die Qualität der Dialogsteuerung, indem sie überprüft, wie die Befragten die kommunikative Zielerreichung, den Computer als Gesprächspartner und den Gesprächsaufbau einschätzen. Die kommunikative Zielerreichung wird mit folgenden Kriterien ermittelt: ob sich der Anruf gelohnt und ob man die gewünschte Information bekommen hat.

Der Computer als Gesprächspartner wird unter folgenden Aspekten untersucht: inwieweit er kommunikativ einem Menschen ähnelt und ob er als vertrauenswürdig eingeschätzt wird. Die Vertrauenswürdigkeit spricht einen extrakommunikativen Interaktionsaspekt – *Truth and Evidence*/Wahrheit und Nachweis– an. Dieser Aspekt ist für die Informationsnutzung von fundamentaler Bedeutung. Die Sprachtheorie berücksichtigt diesen Aspekt im *Domain Model* auf dem *Context*-Niveau.

Die Einschätzung des Gesprächsaufbaus wird mit einer einzigen Frage nach der Komplexität der Dialogstruktur untersucht: es wird überprüft, ob der Weg zur gewünschten Information umständlich war.

Tab. 4 Qualitäts- und Zufriedenheitsfaktoren zur Operationalisierung der Kooperativität

I. Frage-Antwort Aspekte

- 1. Relevanz : 1.1 Informationsgehalt
 - 1.1.1 Informationsmenge
 - 1.1.2 Vollständigkeit der Informationen

- 2. Ausdrucksform 2.1 Akustische Verständlichkeit
 - 2.2 Inhaltliche Verständlichkeit
 - 2.3 Gestaltung der Benutzereingaben

- 3. Korrektur- und Aufklärungsstrategien
 - 3.1 Wiederholung
 - 3.2 Bestätigung
 - 3.2.1 Bestätigung der Eingaben
 - 3.2.2 Bestätigung der Pausen

II. Systemmöglichkeiten

- 4. Transparenz 5.1 Einfachheit der Bedienung
 - 5.2 Systemgrenzen
 - 5.2.1 Einfachheit der Eingabe- und Modifikationformulierung

- 5. Berücksichtigung kontextueller Bedingungen
 - 4.1. Hintergrundwissen
 - 4.2. Erwartungen

- 6. Dialoginitiative 6.1 Möglichkeit Fragen stellen zu können
 - 6.2 Möglichkeit das Gespräch zu führen, solange man möchte

- 7 Schnelligkeit : 7.1 Schnelligkeit der Datenverarbeitung
 - 7.2 Dialogunterbrechungen

- 8. Hilfefunktion

III. Zufriedenheit

- 9. Einschätzung der kommunikativen Zielerreichung

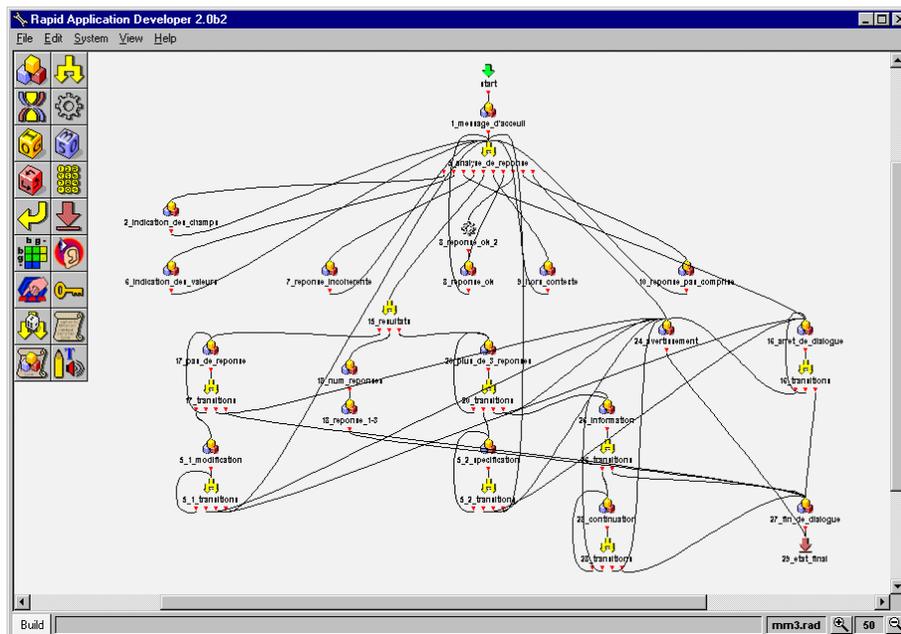
- 10. Einschätzung des Computers als Gesprächspartner
 - 10.1 Ähnlichkeit des Systems mit einem menschlichen Gesprächspartner
 - 10.2 Vertrauenswürdigkeit
 - 10.3 Stimmung während des Gesprächs

- 11. Einschätzung des Gesprächsaufbaus

5. Gestaltung der Experimente

Um die Qualität der Dialogsteuerung eines NSD-Systems und ihren Einfluss auf die vom Benutzer wahrgenommene Systemqualität beurteilen zu können, ist es notwendig, Dialogsituationen mit Versuchspersonen in konkreten Anwendungssituationen zu testen. Am Institut für Kommunikationsakustik der Ruhr-Universität Bochum wurde ein Simulationssystem entwickelt, mit dessen Hilfe die empirische Untersuchung dieser Arbeit durchgeführt werden konnte.

Abb. 6 Das Hauptfenster von BoRIS



5.1 Das NSD-System BoRIS

Der Abschnitt 5.1 entstand in Anlehnung an die von Christine Dudda im Mai 2001 an hiesigen Institut für Kommunikationsakustik angefertigte Diplomarbeit [Dudda, 2001].

BoRIS (Bochumer Restaurants Informationssystem) ist ein unimodales natürlichsprachliches Dialogssystem, welches über die Telefonleitung abrufbar ist und eine umfassende Auswahl an Gaststätten, Kneipen, Bistros, Pizzerias und Cafés bietet. Sein Konzept beruht auf InfoVox, einem System, das vom Institut dalle Molle d'Intelligence Artificielle Perceptive in schweizerischer Martigny erstellt wurde. BoRIS wurde mit Hilfe des Programmpackets *Speech toolkits* des *Centre for Spoken Language Understanding* des *Oregon Graduate Institute* implementiert. Das Programm RAD (*Rapid Application Developer*),¹⁸ welches die Implementierung verschiedener interaktiver Sprachsysteme unterstützt wird für wissenschaftliche Zwecke kostenlos angeboten.

¹⁸ Das Programm auf der Website des Instituts unter <http://cslu.cse.ogi.edu/toolkit> erhältlich.

5.1.1 Bedienung des CSLU RAD Programms

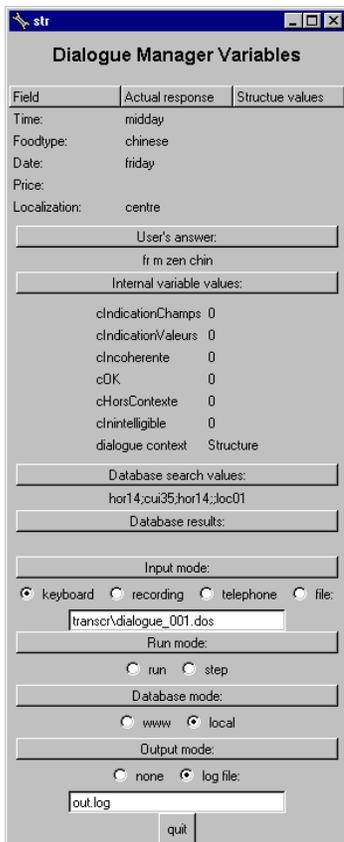
Das Programm muss zunächst auf einem PC gestartet werden. Außerdem muß noch eine Verbindung vom Computer zu einer Sprachsynthese hergestellt werden. Während des ganzen Dialogs werden drei Ausgabefenster geöffnet: im Hauptfenster werden die Zustände gezeigt, die das System zum Erfüllen der Dialogaufgabe durchlaufen muss (siehe Abb. 6). Im zweiten Fenster werden die Systemäußerungen angezeigt (siehe Abb. 7)

Abb. 7 Fenster mit Systemäußerungen



und im dritten die aktuellen Kriterienvorgaben (siehe Abb. 8), die als Text und als Abkürzungen (hor14;cui35) erscheinen. Im dritten Fenster werden auch die gefundenen Restaurants angezeigt. Zusätzlich bietet das Fenster die Möglichkeit, eine bestimmte Art der Eingabe auszuwählen, z.B. per Tastatur, Mikrophon, Telefon oder Datei.

Abb. 8 Fenster mit aktuellen Kriterienvorgaben



5.1.2 Der Dialogablauf mit BoRIS

Die implementierte Dialogstrategie weist eine ziemlich rigide Struktur auf: das System übernimmt die Gesprächsführung, bietet keinerlei Art von Bestätigungen und gibt maximal drei Restaurants aus, deren Namen und Adressen es nach der Ausgabe nicht mehr wiederholen kann (sollten mehr als drei Restaurants gefunden werden fragt das System nach dem dritten, ob es noch weitere ansagen soll).

Die Dialoginitiative kann jedoch als gemischt bezeichnet werden, insofern der Benutzer an jeder Stelle des Dialogs sämtliche Informationen zu den gewünschten Restaurants liefern kann und das System zur Eingabewiederholung (ausgenommen der Restaurantnamen und der Adressenansagen) auffordern kann.

Das System verfügt über einen Spracherkenner, der als Schlüsselworterkenner arbeitet und prinzipiell einen kontinuierlichen Sprachinput ermöglicht.

Nach einer kurzen Begrüßung stellt das System fünf Kriterien zur Verfügung (Art der Küche, Tageszeit, Preis der Mahlzeit, Ort des Restaurants und Öffnungszeiten), nach denen der Benutzer seine Informationssuche gestalten soll. Das Sprachverstehensmodul des Systems verfügt über keine Grammatikkomponente und keinen Parser. Der „Verstehensprozess“ vollzieht sich aufgrund eines Lexikons mit 395 Wörtern, wobei jedem Wort ein Klassentyp (*Food type*), eine Kategorie (*Fish*) und ein Wert (*cui8*) zugeordnet werden.

Tab. 5 Auszug aus der Keyword-Liste¹⁹

WORT	KLASSENTYP	KATEGORIE	WERT
Montag	date	monday	hor01,hor02,hor3
mittags	time	midday	hor02,hor5,hor8,hor11,hor14,hor17,hor20
zwanzig_Uhr	time	evening	hor3,hor6,hor9,hor12,hor15,hor18,hor21
Innenstadt	localization	centre	loc01
Bahnhof	localization	centre	loc01
Querenburg	localization	querenburg	loc12
billig	price	cheap	prx1,prx2,prx3
teuer	price	expensive	prx5,prx6,prx7,prx8
Fisch	foodtype	fish	cui8
Muscheln	foodtype	fish	cui8
südländisch	foodtype	mediterranean	cui19
italienisch	foodtype	italian	cui20
rustikal	foodtype rustic		cui13
nein	logical	no	0

Wenn die Information des Benutzers nicht alle fünf Kriterien enthält, so fragt das System der Reihe nach die Fehlenden ab. Enthält die Information keine der fünf Kriterien und keine Bitte nach Wiederholung oder Hilfe erklärt BoRIS noch einmal seine Arbeitsweise. Nachdem alle Kriterien eingegeben werden, sucht das System nach einem passendem Restaurant aus seiner Datenbank. Wird dabei kein passendes Restaurant gefunden, wird der Benutzer gefragt, ob er seine Anfrage modifizieren möchte. An dieser Stelle ist das

¹⁹ Siehe[Dudda, 2001].

System unflexibel: es erwartet ein „Ja“ oder ein „Nein“, ansonsten verabschiedet es sich. Dadurch, dass das Interaktionsniveau sich auch aus Fragen in *Multiple choice* Form (Art der Küche, Preis, etc.) zusammensetzt, wird es für manche Benutzer schwierig die richtige Antwort zu geben.

Die Modifikationen gestalten sich nach denselben Kriterien der ursprünglichen Informationssuche, wobei die unveränderten Eingaben zusammen mit den neumodifizierten ausgewertet werden.

Das System ist so programmiert, dass, wenn der Benutzer nach einer Wiederholung der Ansagen fragt, es sich nicht wortwörtlich wiederholt. Hiermit wird berücksichtigt, dass, wenn eine Ansagen nicht deutlich war, die Wahrscheinlichkeit größer ist, dass es verstanden wird, wenn die Informationen mit anderen Worten formuliert werden.

Für die Sprachausgabe wurden zwei unterschiedliche Varianten gewählt: die Namen und Adressen der Restaurants wurden aus synthetisch generierten Spracheinheiten erzeugt; für alle anderen Systemansagen wurden fertige Audiodateien benutzt, die in den einzelnen Zuständen abgespielt werden.

5.1.3 Logfile

Für jeden Dialog erzeugt BoRIS ein Protokoll, in dem die Äußerungen des Systems, die vom Spracherkenner erkannten Wörter und die daraus extrahierten Vorgaben für die Datenbankabfrage eingetragen werden. Die Datei, in der diese Informationen gespeichert werden, nennt sich *Logfile*.

Tab. 6 Logfile

[No]	21.1	13	51	53	
1_message_d'accueil	13	52	10	[M6]	Guten Tag [U0]
5_analyse_de_reponse	13	52	18	[U1]	ich möchte heute Abend Pizza essen gehen evening italian monday
8_reponse_ok	13	52	21	[M47]	Wieviel möchten Sie ungefähr pro Person ausgeben?
5_analyse_de_reponse	13	52	27	[U2]	15 m midday cheap
7_reponse_incoherente	13	52	30	[M18]	Könnten Sie bitte die gewünschte Tageszeit der Mahlzeit wiederholen?
5_analyse_de_reponse	13	52	40	[U3]	ja ich möchte abends ausgehen evening
8_reponse_ok	13	52	43	[M45]	In welchem Stadtteil möchten Sie essen gehen?
5_analyse_de_reponse	13	52	50	[U4]	äh in der Stadtmitte centre 3
15_resultats	13	52	51		
18_reponse_1-3 num reponses	13	52	53	[M59]	Es gibt mehrere auf Ihre Anfrage passende Restaurants.
18_Synthese	13	53	05		Spaghetteria San Marco Il Tortellino Die Nudelküche Italia
18_reponse_1-3	13	53	53	[M98]	
27_fin_de_dialogue	13	53	21	[M35]	Vielen Dank, dass Sie BoRIS benutzt haben. Auf Wiederhören und einen schönen Tag

Das *Logfile* hat eine spezielle Struktur: zu Anfang steht die Dialognummer, die aus einer zweistelligen Zahl für die Versuchsperson und das Szenario besteht. Dann wird der Zeitpunkt eingetragen, an dem die Eingabe der Dialognummer erfolgt. Die Anfangszeit des eigentlichen Dialogs befindet sich auf der ersten Zeile, hinter der Eingabe des Systemzustands „1_message_d’acceuil“. Die erste Zeile enthält außerdem noch die Nummer der ersten Systemäußerung [M7] und die Systemansagen in abgekürzter Form.

In der zweiten Zeile wird die Eingabe des Systemzustandes „5_analyse_de_reponse“ wieder eingegeben, gefolgt von der Zeitangabe und der ersten *User*-Äußerung. Jede *User*-Äußerung wird gezählt und bekommt eine Nummer zugeordnet [U1]. Dazu gehören in derselben Zeile die Variablen, z.B. "evening", "centre", nach denen das System die Datenbankanfrage macht (aus Platzgründen werden sie eine Zeile darunter dargestellt). Am Ende der *Logfile*-Datei stehen Messgrößen, wie *Transaction Succes* [TS], [„helpRequest“], *System Corrrrection Turns* [SCT] und *User Corrrrection Turns* [UCT].

Die *Logfile*-Dateien sind ein hilfreiches Instrument für die Evaluierung des Systems und wurden im Rahmen dieser Arbeit zur Veranschaulichung der Kommunikationsprinzipien von Grice (siehe Abschnitt 4.3) gebraucht.

5.2 Wizard-of-Oz Technik

Wizard-of-Oz (WOZ) ist eine Technik, bei der das Gespräch mit einem Computer durch einen Menschen (*Wizard*) simuliert wird. Der Name dieser Technik stammt aus dem gleichnamigen Buch, in dem ein Mann hinter einem Vorhang sitzt und eine Maschine derart manipuliert, dass alle glauben, es handle sich um einen Zauberer.

Bei der *Wizard-of-Oz* Simulation eines interaktiven Sprachsystems wird den Versuchspersonen erzählt, sie würden mit einem Computer telefonieren. In Wirklichkeit steht hinter dem angegebenen Computer ein Mensch. Der *Wizard* übernimmt dabei einen oder mehrere Teile der Computeraufgaben oder kann sogar den Computer ganz ersetzen wie z.B. bei den *Wizard-of-Oz* Versuchen, die im Rahmen der Entwicklung eines dänischen Fluginformationssystems unternommen wurden [Dybkjær, L. und Dybkjær, H., 1993]; dabei übernahm ein *Wizard* die Rolle des Computers; für die Sprachausgabe wurde keine Sprachsynthese verwendet, sondern man verzerrte die Stimme des *Wizards* elektronisch.

Die Sprachverstehenskomponenten wurden im Laufe der Versuche mit BoRIS ausgeschaltet und die Benutzereingaben wurden manuell eingetippt. Dadurch wollte man Spracherkennungsfehler vermeiden, die eventuell die Testergebnisse beeinflussen könnten und so die Versuche gezielt auf die Dialogsteuerung konzentrieren. Mit dieser Simulation konnten keine Korrekturstrategien im Spracherkennungs- oder Sprachverstehensmodul untersucht werden, weil gerade diese Komponente simuliert wurden.

5.2.1 Kritische Betrachtung der Wizard-of-Oz-Technik

Gegen die Verwendung der *WOZ* Technik lässt sich als Argument anführen, dass diese aus ethischen Gründen bedenklich sei, da die Versuchspersonen zum Zwecke des Versuchs getäuscht und belogen werden. Diese Täuschung ist aus psychologischer Sicht jedoch notwendig, da sich die Versuchspersonen anders verhalten, wenn sie meinen, mit einem Computer zu sprechen. Dieser Konflikt lässt sich lösen, indem man den Versuchspersonen nach Durchführung des Experiments erklärt, wie das System funktioniert und aus welchen Gründen sie dies nicht wissen durften.

Ein weiterer möglicher Einwand gegen diese Technik besteht darin, dass es für den *Wizard* insbesondere schwierig sei, während der ganzen Simulation konsistent zu bleiben. Es ergibt sich die Notwendigkeit, dem Versuchsleiter (*Wizard*) feste Dialogstrategie und Reaktionsschablonen vorzugeben. An diese muss sich der *Wizard* während des Verlaufs aller Versuche konsequent halten, um das Verhalten eines computergesteuerten Systems möglichst gut zu approximieren [Rehmann, 1999].

Trotzdem bleibt die *Wizard-of-Oz* Technik eine sinnvolle Methode zur Evaluierung von technisch beschränkten Dialogsystemen, deren Untersuchung Aussagekraft für künftige Systeme haben soll, die dem gegenwärtigen Stand der Technik einen Schritt voraus sind.

5.3 Der Experimentaufbau

5.3.1 Versuchslayout

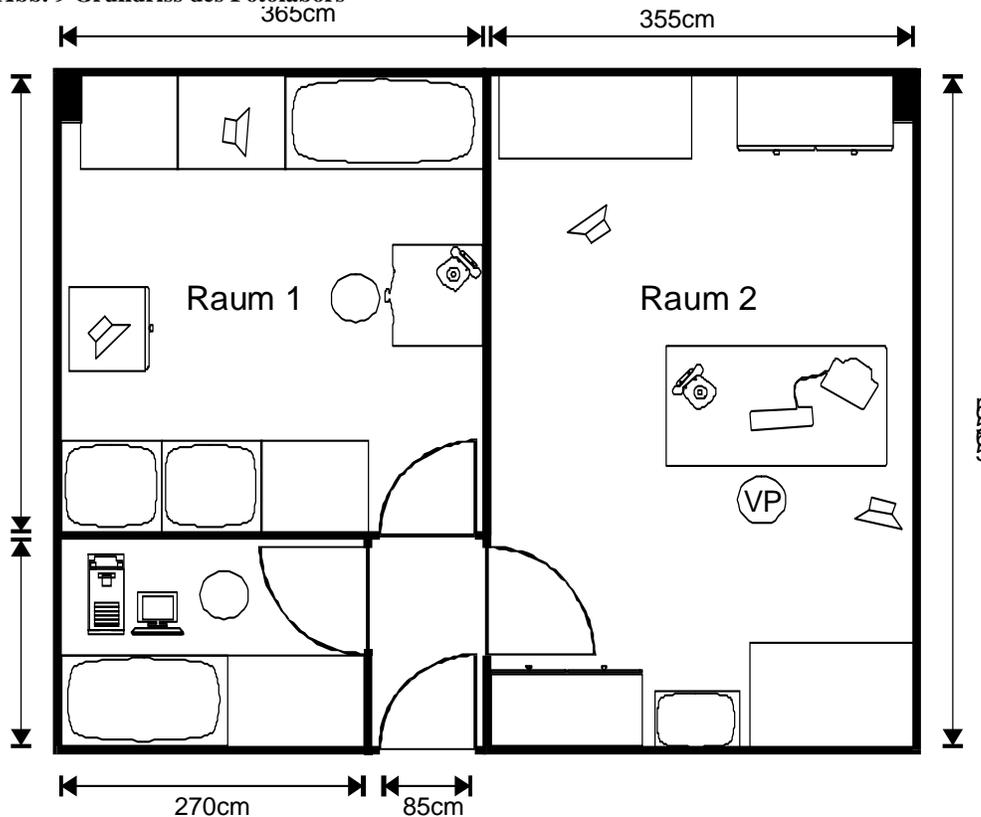
Die Abschnitte 4.3.1 und 4.3.2 entstanden in Anlehnung an [Dudda, 2001].

Die Versuche wurden am Institut für Kommunikationsakustik der Ruhr-Universität Bochum durchgeführt. Die Simulation lief in den Räumen des Fotolabors, wo der Computer an den vorhandenen Telefonsimulator und für die Online-Synthese an das Netzwerk des Instituts angeschlossen wurde.

Das Fotolabor besteht aus drei Räumen: in dem ersten wurde der Simulationsrechner aufgestellt, am dem die Versuchsleiterin saß; in dem zweiten Raum stand das Telefon für die Versuchspersonen und in dem dritten befand sich der Leitungssimulator und ein Dat-Recorder für die Audioaufzeichnung der Versuche.

Die Versuchspersonen interagierten mit dem System über eine simulierte Telefonleitung. Diese simulierte Verbindung, die mit Hilfe des Leitungssimulators implementiert wurde, ermöglichte die Übertragung des Sprachoutputs vom Computer zum Benutzer, soweit die Übertragung der Benutzeräußerung zum Dialogsystem bzw. zur Versuchsleiterin erfolgte. Die Antworten der Versuchspersonen wurden von der Versuchsleiterin über einen Kopfhörer in dem Raum 1 angehört. Die Lautstärke konnte in dem Raum 3 vor dem Versuch eingestellt werden.

Abb. 9 Grundriss des Fotolabors



5.3.2. Versuchsablauf

Die Versuchspersonen wurden, der Reihe nach, in den Raum 2 des Fotolabors eingeführt. Man hat darauf geachtet, dass die Türen der anderen zwei Räumen während der Versuche geschlossen blieben. Die Versuchspersonen wurden gebeten sich an den Tisch zu setzen und sich ein kurzes Informationsblatt über BoRIS durchzulesen. Dann wurden anhand eines standardisierten Fragebogens vor dem Versuch interviewt (siehe Kap. 6). Am Ende bekamen sie die Szenarien und wurden von der Versuchsleiterin unterrichtet, wie sie die Telefonleitung bedienen sollen: um den Startsignal zu geben, sollten die Versuchspersonen den Telefonhörer abnehmen und den Knopf am Telefon drücken, welcher im benachbarten Raum 1 das Klingeln auslöste. So konnte der Versuchsleiterin den bevorstehenden Gesprächsanfang signalisiert werden. Nachdem sie das Gespräch mit BoRIS beendet haben, sollten sie den Hören wieder auflegen und dieselbe Vorgehensweise bei der Durchführung jedes Szenarios wiederholen. Nach den Experimenten wurden sie erneut interviewt. Die Versuche wurden zum Zwecke der Analyse auf Dat-Kassen aufgenommen. Die qualitativen Interviews (siehe Kap. 6) zeichnete man auf einem Tonbandrecorder mit Hilfe eines Mikrofons auf, das man im Raum 2 neben den Versuchspersonen aufstellte.

5.3.3 Szenarien

Um die Experimente einheitlich durchführen zu können, wurden sechs verschiedene Szenarien entwickelt. Die Versuchspersonen wurden aufgefordert sich an diese Szenarien zu halten und dabei zu notieren, welche Restaurants BoRIS für sie gefunden hat und eventuell einen Kommentar zum jeweiligen Szenario abzugeben. Die Durchführung der Szenarien dauerte insgesamt, je nach kommunikativem Verhalten der Versuchspersonen, zwischen 15 und 25 Minuten.

Die Vorgabe der Szenarien sorgte für vergleichbare Ergebnisse, aber gleichzeitig hemmte es die Spontaneität der Versuchspersonen. Um ihnen den Einstieg bei BoRIS zu erleichtern und ihre Phantasie zu anzuregen, wurde bei dem ersten Szenario den Versuchspersonen die Freiheit gelassen, im Rahmen der vom System angebotenen Informationsmöglichkeiten zu fragen, wonach sie möchten. Die Gestaltung des ersten Szenarios ist immer essentiell für den ersten Eindruck über das System und hat Auswirkungen auf das Verhalten der Versuchspersonen im späteren Verlauf der Experimente. Indem man den Versuchspersonen erlaubt, den Dialog mit dem System nach eigenen Vorstellungen zu gestalten, gibt man ihnen die Möglichkeit das System auf eigene Art kennen zu lernen und selbst herauszufinden, welches Vokabular benutzt werden soll, wie man seine Äußerungen gestalten darf usw. Dadurch können auch Lerneffekte erzielt werden.

Erst bei dem zweiten Szenario wurden die Versuchspersonen aufgefordert präzise Eingaben zu formulieren. Sie sollten am Montagabend, zu einem chinesischen Restaurant in die Innenstadt essen gehen und sich dementsprechend bei BoRIS erkundigen. Die Aufgabe sollte eigentlich den Versuchspersonen keine Probleme bereiten (BoRIS hat in seiner Datenbank eine Menge Adressen chinesischer Restaurants in der Innenstadt, so dass der Dialog erfolgreich ablaufen sollte ohne Modifikationen nötig zu haben). Schwierigkeiten damit bekamen nur Versuchspersonen, die sich mit Restaurantpreisen nicht auskannte und als Preisklasse für chinesische Restaurants „teuer“ eingaben (chinesische Restaurants sind allgemein preiswert).

Die Durchführung des dritten Szenarios erwies sich, im Vergleich zu den ersten zwei, als problematischer. Laut der Aufgabe sollten die Versuchspersonen, nach einem arabischen Restaurant in Altenbochum suchen. Fall es dort keins gibt, sollte es in Stadtteil Grumme gesucht werden. Das System gibt weder „arabisch“ noch „Altenbochum“, noch „Grumme“ als Antwortmöglichkeit an, so dass die Versuchspersonen zwei Möglichkeiten hatten: entweder die Szenarioaufgabe komplett zu ignorieren und sich strikt an den Systemanweisungen zu halten und eine andere Art der Küche oder einen anderen Ort des Restaurants auszuwählen, die vom System genannt werden; oder einfach zu versuchen, die Eingaben aus dem Szenario zu machen und dabei zu riskieren, vom System nicht verstanden zu werden. Das Szenario bringt an dieser Stelle ein kommunikatives Problem: nach dem Kohärenzprinzip müssen die Antworten an Frageelemente knüpfen. Dadurch, dass BoRIS nicht preisgibt, was in seiner Datenbank steckt, wird eine Konfliktsituation

generiert, die oft den Dialog zum Scheitern führt: einerseits versuchen die Versuchspersonen die Szenarioeingaben zu respektieren, andererseits müssen sie kohärent bleiben.

Das vierte Szenario stellte als Aufgabe ein preiswertes deutsches Restaurant für das Mittagessen zu suchen. Die Versuchspersonen hatten dabei die Möglichkeit, den Preis zu modifizieren, um ein passendes Restaurant finden zu können. Die Schwierigkeit dieses Szenario ergab sich dadurch, dass die Versuchspersonen unterschiedliche Vorstellungen über die Preiskategorien hatten, die zum Teil mit denen des Systems nicht übereinstimmten. Für manche war schwer einzuschätzen, was das System unter „teuer“ oder „billig“ verstand. Einige der Teilnehmer gaben bei den Preismodifikationen sehr niedrige Werte ein (z.B. 5 DM), mit der Absicht in die Preiskategorie „billig“ zu gelangen und so vom System die gewünschte Restaurantinformation zu bekommen. Da das System die Preiswerte den dazu passenden Restaurants zuordnet, war eine solche Überlegung verkehrt (ein Wert von 5 DM ist für Restaurantpreisverhältnisse unrealistisch).

Bei dem fünften Szenario mussten die Versuchspersonen nach einer lateinamerikanischen Cocktailbar fragen. Die Versuchspersonen wurden mit demselben Problem konfrontiert, wie beim dritten Szenario: unter den aufgelisteten Küchearten gibt es keine „lateinamerikanische“ Kategorie. Wenn die Versuchspersonen sich trotzdem traute, diese Kategorie einzugeben fand das System mit einem einzigen *Turn* das passende Restaurant. So konnte die Lernfähigkeit der Versuchspersonen dabei getestet werden.

Die Struktur des sechsten Szenarios unterscheidet sich wesentlich vom Rest. Die Versuchspersonen bekamen den Namen eines spanischen Restaurants („Coco Loco“) in Weimar und sollten sich bei BoRIS nach dessen Adresse erkundigen. Obwohl in seiner Datenbank vorhanden, kann das System aufgrund der implementierten Dialogstrategie, nicht den Namen „Coco Loco“ mit der Kategorie „spanische Küche“ verknüpfen. Um die Aufgabe durchführen zu können mussten die Versuchspersonen die Kategorien „spanisch“ und/oder „Weimar“ eingeben, was wiederum das System bei der Auflistung der Küchenarten und Stadtteile nicht zur Wahl gestellt hat. Die Lernfähigkeit der Benutzer wurde dadurch noch einmal geprüft.

5.3.3.1 Kommunikatives Verhalten der Versuchspersonen

Im Laufe der Untersuchungen hat man festgestellt, dass die Versuchspersonen unterschiedliche Dialogstrategien hatten. Eine grobe Unterscheidung teilt sie in zwei Gruppen. Testpersonen aus der ersten Gruppe formulierten alle Eingaben, die sie machen wollten, in einen Satz. Typisch waren solche Dialogstrategien für unerfahrene und oft für weibliche Versuchspersonen. Der Vorteil einer solchen Dialogstrategie ist, dass die oftmals verwirrende Abfrage und Aufzählung der einzelnen Kategorien durch das System übersprungen wird und dadurch das Dialogziel schneller erreicht werden kann. Der

Nachteil besteht darin, dass in dem Fall, in welchem das System kein passendes Restaurant findet, eine Modifizierung der Eingaben riskant wird. Das System gibt nicht an, welche Eingabe falsch ist und übernimmt dazu noch alle unmodifizierten Kategorien aus dem vorherigen *Turn*. Unter diesen Bedingungen kann es vorkommen, dass die Testperson genau die Eingaben modifiziert haben, die korrekt waren. Durch die automatische Übernahme der unmodifizierten Kategorien war es der Versuchsperson nicht möglich zu erkennen, in welchem *Turn* der Fehler auftrat. Der Dialog konnte sich unendlich in die Länge ziehen, bis die Versuchsperson ihn abbrach.

Zu der zweiten Gruppe gehören vor allem erfahrene, meist männliche, Testpersonen, die angaben, Erfahrungen im Umgang mit computergestützten Systemen zu haben. Sie gestalteten ihre Äußerungen stichwortartig und erwarteten vom System nach jeder einzelnen Kategorie gefragt zu werden. Sie versuchten stets sich dem System anzupassen und antworteten nur mit den *Keywords*, die das System vorgab. Eine solche Dialogstrategie hätte die Spracherkennung erheblich erleichtert. Der Nachteil dabei ist, dass das Gespräch länger dauert und die Versuchspersonen, durch die Auflistung der möglichen Antworteingaben, sich nicht trauen, andere Eingaben zu machen, die vom System nicht genannt werden. Versuchspersonen aus dieser Kategorie hatten vor allem Schwierigkeiten beim dritten, fünften und sechsten Szenario, bei denen sie aufgefordert wurden, Kategorien einzugeben, die im Vorfeld das System nicht aufgelistete.

6. Auswertung der Experimente

Aufgrund der Tatsache, dass die Untersuchungen bezüglich der Qualitätsfaktoren eines Dialogs Pioniercharakter haben und dass das Untersuchungsobjekt schwer quantifizierbare Merkmale hat, traf man die Entscheidung, die Befragung in zwei Etappen und mit Hilfe von zwei verschiedenen Methoden durchzuführen: zuerst wurden Daten aufgrund einer qualitativen Methode ermittelt, die mit Hilfe einer quantitativen Methode, statistisch gemessen wurden.

Welches sind die Charakteristika dieser Methoden? Mit welchen Vor- und Nachteilen sind sie verbunden? Der folgende Abschnitt versucht eine Antwort darauf zu geben.

6.1 Qualitative versus quantitative Befragung

In einer qualitativen Analyse werden verbale bzw. nicht numerische Daten interpretativ verarbeitet. Diese Daten können z.B. Beobachtungsprotokolle, Interviewtexte, Briefe oder Zeitungsartikel. Ziel einer solchen Analyse ist durch den inhaltlichen Reichtum der individuellen Antworten die Erfahrungsrealität zu rekonstruieren und sie dann interpretativ auszuwerten.

Dagegen analysiert die quantitative Forschung statistische Messwerte. Im Vordergrund steht bei einer quantitativen Methode die Frage, wie die zu erhebenden Merkmale operationalisiert bzw. quantifiziert werden sollen. Die Merkmale können quantitativ (Körpergröße, Pulsfrequenz, Klausurleistungen) oder qualitativ (Geschlecht, Augenfarbe, Konfession) sein.

Eine der wichtigsten Grundtechniken zur Erhebung qualitativer Daten und die gleichzeitig in den empirischen Sozialwissenschaften am häufigsten angewandte Methode ist die Befragung. Qualitative Befragungen arbeiten mit offenen Fragen, lassen dem Befragten viel Spielraum beim Antworten und berücksichtigen die Interaktionen zwischen Befragten und Interviewer sowie die Eindrücke und Deutungen des Interviewers als Informationsquelle. Im Unterschied dazu sind quantitative Befragungen an einer Antwortvorgabe gebunden: sie fordern vom Befragten eine Auswahl von zutreffenden Antworten und vom Interviewer ein gleichförmiges Verhalten [Bortz und Döring, 1995].

Der Vorteil der qualitativen Methode ist, dass der Befragte durch offene Fragen innerhalb seines eigenen Referenzsystems antworten kann, ohne in eine bestimmte Richtung gelenkt zu werden. Dadurch wird viel mehr an Information gewonnen als bei quantitative Untersuchung. Nachteilig dabei ist, dass nicht alle Befragten ein gleich gutes Ausdrucksvermögen bezüglich ihrer Einstellung und Meinungen haben. Antwortunterschiede können somit nicht auf Einstellungsunterschiede zurückgeführt werden, sondern ergeben sich aus der unterschiedlichen Art und Weise der Befragten, ihre Einstellungen in Worte zu fassen. Qualitative Befragungen beanspruchen in der Regel

mehr Zeit, können entsprechend mit weniger Personen durchgeführt werden und erfordern einen erheblichen zusätzlichen Auswertungsaufwand, da im nachhinein Auswertungskategorien gebildet werden müssen [Schnell et. al., 1989].

Quantitative Untersuchungen dagegen schränken die Ausdrucksmöglichkeiten der Befragten durch standardisierte Fragebögen ein, sind leichter auszuwerten, können mit mehreren Personen durchgeführt werden und liefern statistisch interpretierbare Werte.

Aus qualitativen Daten lassen sich auch quantitative Daten erzeugen, indem die Texte in geordnete Kategorien thematisch sortiert, zusammengefasst und pro Kategorie auf eine einzige Aussage reduziert werden. Die Aussagen können als Frage oder Behauptung mit Rating-Skala als Antwortvorgabe formuliert werden.

Viele Forschungsprojekte, darunter auch diese Arbeit zählt, kombinieren die Herangehensweise der qualitativen und quantitativen Forschung, um von den Vorteilen beider Methoden zu profitieren [Bortz und Döring, 1995].

6.2 Allgemein

Die qualitative Befragung wurde mit acht Personen durchgeführt. Bei der Auswertung der Interviews ergaben sich wiederkehrende Eingaben, die man, wie oben beschrieben, thematisch gruppiert und jeweils auf eine einzige Kernaussage reduziert hat. Die Kernaussagen wurden im Laufe der zweiten Etappe der Experimente zu geschlossenen Fragen (*Multiple choice*-Form) und Behauptungen (auf fünf Stufen skaliert), nach dem Muster einer quantitativen Methode bewertet. Für die gesamte Gestaltung der Experimente wurden folgende Annahmen getroffen:

Form der Befragung: Je nach Art der Befragung wurde ein standardisiertes qualitatives bzw. quantitatives Interview ausgearbeitet.

Für jedes Interview wurden zwei Fragebögen entworfen: der eine wurde vor dem Versuch durchgeführt, der andere danach. Sowohl die Versuche als auch die qualitativen Interviews wurden auf DAT- bzw. Audiokassetten aufgenommen.

Zielgruppe: Die Zielgruppe der Befragten wurde aus potentiellen Kunden des Systems in einer Altersgruppe zwischen 19 und 42 Jahren zusammengesetzt. Der Kreis der Befragten umfasste größtenteils Studenten und Mitarbeiter der Ruhr-Universität Bochum.

Ort und Dauer der Befragung: die Befragungen wurden am Institut für Kommunikationsakustik in einem speziell eingerichteten Raum durchgeführt (siehe Abschnitt 5.3.1). Die Durchführung der qualitativen Befragung erfolgte an einem Tag. Die quantitative Befragung dagegen erstreckten sich über eine Woche.

6.3 Das qualitative Interview

Zu Beginn der Experimente ging es um die Entwicklung eines Fragebogens, der eine konkrete Identifizierung und pauschale Bewertung der Qualitätsfaktoren durch die

Versuchspersonen ermitteln sollte. Die Fragen wurden offen gestellt, um den Befragten zu ermöglichen, sich, obwohl an einem Interviewfaden knüpfend, frei zu äußern. Es ergaben sich Situationen, in denen die Versuchspersonen die Fragen nicht genau verstanden. Die Versuchsleiterin stellte zusätzliche Fragen, um die erzielte Antwort zu bekommen. Diese Fragen wurde aufgrund ihres wichtigen Informationsgehalts bei der Auswertung mitberücksichtigt.

Die Interviews wurden anhand der Aufnahmen ordentlich transkribiert (siehe Anhang B) und dauerten zwischen 40 und 60 Minuten. Nachdem das akustische „Material“ in eine lesbare Form gebracht wurde (dabei wurden auch nonverbale Aspekte wie Pausen, Lachen und Unterbrechungen berücksichtigt), begann man mit der Analyse der einzelnen Interviews. Zunächst wurden die Nebensächlichkeiten aus den Texten entfernt, und so die wichtigsten Passagen der Texten hervorgehoben. Die prägnantesten Textstellen wurden dann aus dem Text entnommen und der Frage entsprechend einer untersuchten Kategorie zugeordnet. Auf eine detaillierte Analyse der einzelnen Interviews, hat man verzichtet, denn sie würde den Rahmen dieser Arbeit sprengen. Vielmehr konzentrierte sich die Analyse auf die Darstellung der aufgetretenen Gemeinsamkeiten und Unterschiede zwischen den Befragten, die zusammengefasst Indizien für Grundtendenzen der Auswertung liefern sollen.

Die Fragen (7.0), (8.0) und (13.0) wurden in Anlehnung an [Loderer, 1998] gestellt.

6.3.1 Aufbau

6.3.1.1 Aufbau des ersten Fragebogens

Der erste Teil des Interviews, den man vor dem Versuch durchgeführt hat, kann in drei thematische Einheiten eingeteilt werden: Personalien, allgemeine Einstellungen zum Thema Restaurantsuche und zu Dialogsystemen als Sprach-Informationssystemen.

Die Personalien erfassen einige Daten zu den Versuchspersonen, wie Geschlecht, Alter, und Beruf/Ausbildung. Die Versuchspersonen wurden noch gebeten, ihre Herkunftsregion anzugeben, um eventuelle Unterschiede im Sprachgebrauch nachzuvollziehen, die für die Untersuchung relevant sein könnten. Dazu wurden sie zusätzlich nach ihren jetzigen Wohnort gefragt, um so festzustellen, welche potentielle Benutzer des Systems sein könnten.

Die ersten zwei Fragen des Interviews beziehen sich auf die Einstellungen der Versuchspersonen zum Thema Restaurantsuche und erkunden die Informationsquelle (Frage 1.0), die man in Anspruch nimmt, wenn man in einem fremden Ort auswärts essen gehen möchte (man ging davon aus, dass vor allem Ortsfremde die Dienste eines NSD-Informationssystems beanspruchen würden) und die Kriterien, die bei der Restaurantsuche eine Rolle spielen (2.0). Die Informationen können bei der Auswertung Aufschluss darüber

geben, warum eine Versuchsperson auffällig anders urteilt als der Durchschnitt [Dudda, 2001].

Der Fragebogen soll weiter die Erfahrungen (3.0) der Befragten mit Sprachdialogsystemen und ihre Erwartungen (4.0) diesen gegenüber dokumentieren.

Bei Frage (3.0) sollten die Versuchspersonen Auskunft geben, ob sie wissen, was ein Sprach-Informationssystem ist. Sollte die Antwort positiv sein, wurden sie zur Kontrolle (3.1) gebeten, zu beschreiben, was sie unter einem solchen System verstehen. Hier konnte die Interviewerin eingreifen, falls die Versuchspersonen falsche Vorstellungen hatten, um ihnen zu erläutern, was damit gemeint ist. Zusätzlich sollten die Versuchspersonen über ihre Erfahrungen mit Sprach-Informationssystemen berichten (3.1.1) und angeben, wie oft (3.1.2) und zu welchem Zweck (3.1.3) sie die Dienste eines solchen Systems in Anspruch genommen haben.

Sollten die Versuchspersonen keinerlei Erfahrungen mit Sprach-Informationssystemen haben, wurde ihnen die Systemfunktionsweise kurz erklärt und anschließend gefragt, ob sie sich vorstellen können, damit umgehen zu können (3.2).

Frage (4.0) stellt eine hypothetische Situation dar: die Befragten sollten sich vorstellen, dass sie bei einem Sprach-Informationssystem anrufen. Sie sollten sich dabei Gedanken machen, welche Reaktionen vom System zu erwarten sind (4.1), wie die Stimme eines solchen Systems klingt (4.1.1), wie sie sich den Gesprächsablauf (4.2) und das Umgehen mit dem Sprachcomputer (4.3) vorstellen, welche Informationen und Servicemöglichkeiten das System bieten könnte (4.4, 4.5) und wie sie das System bedienen soll, damit sie zufrieden sind (4.6). All diese Fragen wurden gestellt, um die Erwartungen der Versuchspersonen an das System zu ermitteln.

Zum Schluss wurden die Versuchspersonen gefragt, welche Vor- und Nachteile ein solches System haben kann.

6.3.1.2 Aufbau des zweiten Fragebogens

Der zweite Fragebogen enthält noch 17 andere Fragen, die das Gespräch mit BoRIS detailliert dokumentieren sollen.

Zuerst wurden die Versuchspersonen gefragt, ob das System ihren Vorstellungen entsprach; und wenn nicht, warum dies der Fall war (6.0). Die Antwort auf diese Frage lieferte die ersten Eindrücke, die das System bei den Versuchspersonen hinterließ und steht in Zusammenhang mit den Erwartungen und dem Hintergrundwissen der Versuchspersonen.

Die Fragen (7.0) und (8.0) versuchen von den Befragten herauszufinden, was sie bei BoRIS gestört hat und was sie gut gefunden haben. Die Fragestellung ist allgemein und lässt den Befragten die Freiheit, Sachverhalte anzusprechen, die für ihre eigene Erfahrung mit BoRIS repräsentativ sind. An Frage (8.0) anknüpfend wird die Frage (9.0) gestellt, ob

das System den Befragten Probleme bereitet hat und an welchen Stellen sie Schwierigkeit hatten. Damit soll herausgestellt werden, ob BoRIS schwer zu bedienen war.

Man kann gewiss die Versuchspersonen nicht direkt nach den Qualitätsfaktoren fragen. Die Fragestellung soll so geschickt formuliert sein, dass die Versuchspersonen selbst im Laufe des Interviews die Qualitätsfaktoren selber ansprechen. Aus diesem Grund überlegte sich die Autorin, welche möglichen „Problemstellen“ eventuell den Dialogablauf erschweren könnten. Frage (10.0) thematisiert die Transparenz des Systems; die Versuchspersonen wurden gefragt, ob ihnen von Anfang an klar war, welche Information BoRIS liefern kann und ob sie nach dem ersten Kontakt mit dem System wussten, nach welchen Kriterien sie ein Restaurant suchen sollten.

Die Fragen (11.0) und (12.0) sprechen den Qualitätsfaktor "Verständlichkeit" an: die Versuchspersonen wurden gefragt, ob sie das Gefühl hatten, vom System verstanden worden zu sein und ob das System deutlich gesagt hat, was es von ihnen wissen wollte.

Frage (13.0) bezieht sich auf den Gesprächsaufbau. Die Versuchspersonen wurden dabei aufgefordert, den Gesprächsaufbau zu charakterisieren. Die Antworten auf diese Frage sollten zeigen, wie zufrieden die Versuchspersonen mit dem Gesprächsaufbau waren.

Frage (14.0) bezieht sich auf eventuelle Unterbrechungen, die sich ergeben können; die Versuchspersonen sollten darüber Auskunft geben, ob der Gesprächsfluss unterbrochen wurde, wie oft es geschah und ob sie diese Unterbrechungen als störend empfunden haben. Die Dialogunterbrechungen wirken sich auf die vom Benutzer wahrgenommene Schnelligkeit des Systems aus und werden unter dem Qualitätsfaktor "Schnelligkeit" berücksichtigt.

Frage (15.0) sollten die Versuchspersonen angeben, wie sie ihre Äußerungen formuliert haben: ob sie Stichwörter benutzt haben oder sich wie in einem normalen Gespräch geäußert haben. Die Antwortformulierung spielt eine große Rolle für den guten Verlauf des Dialogs und wirkt sich auf die Qualitätsbeurteilung des Gesamtsystems aus.

Frage (16.0) soll den Erfolg des Dialogs mit BoRIS überprüfen. Die Befragten sollen angeben, ob sie die gewünschten Informationen bekommen haben und wie schnell es empfunden wurde. Dann wurden sie gebeten, den Informationsgehalt zu beurteilen und anzugeben, ob sie damit zufrieden waren (Frage 17.0). Der Informationsgehalt wird unter dem Qualitätsfaktor „Relevanz“ untersucht und bildet somit einen wichtigen Punkt der Analyse.

Frage (18.0) überprüft, wie vertrauenswürdig das System auf die Befragten wirkt und liefert Hinweise auf die Zufriedenheit der Versuchspersonen mit dem System.

Die Systemkonfiguration von BoRIS, die für die Versuche benutzt wurde, verfügt über keine Bestätigungsmöglichkeit der Eingaben. Frage (19.0) versucht herauszufinden, ob die Versuchspersonen eine solche Bestätigung für wichtig halten.

Frage (20.0) untersucht die Stimmung der Versuchspersonen während des Gesprächs mit BoRIS und Frage (21.0) die allgemeine Zufriedenheit der Versuchspersonen mit dem

System: dabei wurden sie gefragt, ob der Anruf bei BoRIS sich gelohnt hat und ob sie die Dienste eines solchen Sprach-Informationssystems nochmals im Anspruch nehmen würden.

Die letzte Frage (22.0) versucht herauszufinden, ob die Versuchspersonen der Versuchsleiterin über ihre Erfahrung mit BoRIS noch was mitteilen wollen. Die Befragten haben dadurch die Möglichkeit, sich zu Sachverhalten bezüglich des Systems zu äußern, die im Laufe des Interviews eventuell nicht angesprochen wurden.

6.3.2 Die Auswertung

6.3.2.1 Auswertung des ersten Fragebogens

Die Versuchspersonen, vier weibliche und vier männliche, waren zwischen 22 und 39 Jahren alt. Die Untersuchung strebte nach einem gleichmäßigem Anteil von Männern und Frauen. Es wird sich hier zeigen, ob es geschlechtsbedingte Unterschiede in der Bewertung des System gibt.

Fünf von ihnen hatten eine ingenieurwissenschaftliche bzw. technische Ausbildung, zwei waren Naturwissenschaftler und einer Geisteswissenschaftler.

Die meisten Befragten (7 Nennungen) gaben an, Computerkenntnisse zu haben.

Wohnsitz: Alle waren in Bochum ansässig, drei von ihnen waren Ausländer.

Durch die ersten Frage (1.0) des Interviews stellt heraus, dass die meisten Befragten einen menschlichen Informanten bevorzugen: sechs von insgesamt acht gaben an, neben schriftlichem Informationsmaterial (Gelbe Seiten, Reiseführer, Magazine, Internet), Bekannte, Freunde oder sogar unbekannte Passanten -„*einfach Leute fragen, die rumlaufen*“ (VP.²⁰ 3)- zu konsultieren. Nur zwei Befragte gaben an, sie würden alleine etwas suchen: „...*ich gehe einfach hin und schaue mir an, was es da gibt*“ (VP. 6).

Die wichtigsten Kriterien bei der Restaurantsuche (2.0) scheinen der Preis (7 Nennungen) und das Ambiente (6 Nennungen) zu sein. Die Art der Küche, die Erreichbarkeit und die Qualität des Essens erhielten jeweils 3 Nennungen. Erwähnt wurden außerdem die Schnelligkeit der Bedienung (2 Nennungen) und die Öffnungszeiten (1 Nennung). Interessant war die Bemerkung einer weiblichen Versuchsperson, die angab: „... *es müssen Leute drin sein, sonst traue ich mich nicht rein zu gehen. Ein Tisch muss schon mit Leuten besetzt sein...*“ (VP. 7). Dieses Argument spricht nicht nur für die Schüchternheit der Versuchsperson, sondern auch für ihre Präferenz sich von Menschen beraten zu lassen: die Anwesenheit anderer Menschen im Restaurant kann auch als Hinweis auf die Qualität des Restaurants interpretiert werden (bemerkenswert wäre noch, dass diese Versuchsperson als einzige Informationsquelle menschliche Informanten angab : „*Ich versuche irgendwen auf der Straße zu treffen, wenn ich unterwegs bin, ich frage irgendeinen Passanten und frage,*

²⁰ VP. steht für Versuchsperson. Der Text aus dem Interview wurde kursiv in Einführungsstrichen angegeben.

ob er irgendwas kennt, wo man lecker essen kann. Wenn ich in einem Hotel wohne, dann würde ich die Leute, die da arbeiten fragen. Ich würde normalerweise Leute einfach fragen“ (VP. 7).

Dann wurden die Versuchspersonen gefragt, ob sie wussten, was ein Sprach-Informationssystem ist (3.0). Die Frage ist im bezug des QF.²¹ "Hintergrundwissen" relevant. Man hat absichtlich das Wort „Dialogsystem“ nicht benutzt, da es allgemein nicht bekannt ist. Dadurch konnte man an den Wissenstand der Versuchspersonen anknüpfen und ihnen erklären, was eigentlich unter einem Sprach-Informationssystem zu verstehen ist.

Nur eine Versuchsperson hatte eine einmalige Erfahrung mit einem Kinokartenreservierungssystem und eine andere hatte bereits von einem solchen System gehört, es aber selbst noch nicht benutzt. Alle unerfahrenen Versuchspersonen konnten sich vorstellen mit einem Dialogsystem umzugehen (3.2), obwohl manche sich über ein Gespräch mit einem Computer kritisch äußerten (3 Nennungen): *„Wenn ich mir eine ganz konkrete Situation vorstelle, dass ich eine richtige Frage habe oder so, ich müsste irgendwo anrufen und da würde sich ein Sprachcomputer oder dieses System melden, statt einer menschlichen Stimme da wäre ich nicht besonders begeistert, da steh ich also kritisch“ (VP. 5).*

Die Erwartungen bezüglich der Systemreaktionen sind sehr verschieden (4.1): Die Hälfte der Befragten hat keine konkreten Vorstellungen über die Reaktionen des Systems: *„..ich hoffe, dass es nicht so dumm wird [...], dass ich eine Frage stelle und dann eine Antwort bekomme, die überhaupt nichts damit zu tun hat“ (VP.1); „ich habe keine Ahnung, ich glaube wie ein Computer“ (VP.7) (2 Nennungen) sind dagegen negativ eingestellt: „...Computer verstehen immer mies was man möchte. Also Missverständnisse erwarte ich.“ (VP. 8). Nur zwei Versuchspersonen machten intuitiv eine kurze Beschreibung eines Dialogs mit einem Sprachcomputer: „ [...] er hat zwei Möglichkeiten, also mich reden lassen, was ich will, oder aber er fragt, was ich suchen will, „Wo wollen Sie hin, was für ein Restaurant ?“ Oder was ich für Wünsche habe und dann sagt man einen Text, wo er sich die Wörter aussucht und selber zuordnet. Die beiden Möglichkeiten stelle ich mir vor, wobei wahrscheinlich die eine mit den einzelnen Abfragen, die üblichere sein könnte“ (VP. 3).*

Die meisten Befragten (5 Nennungen) erwarten eine menschliche freundliche Computerstimme (4.1.1). Nur zwei Befragte gaben an, sie würden eine synthetische Sprachausgabe erwarten; zwei männliche Versuchspersonen fügten noch hinzu, sie würden eine weibliche Stimme bevorzugen.

Interessant ist die Tatsache, dass obwohl die Erfahrungen mit dem System bei den meisten Befragten gleich null sind, ihre intuitiven Vorstellungen über den Gesprächsablauf

²¹ QF. steht für Qualitätsfaktor

tatsächlich der Realität entsprechen: fast alle Versuchspersonen (7 Nennungen) sind der Meinung, der Computer würde ihnen Fragen stellen, wobei manche sich den Dialog sehr positiv „*wie ein Wechselspiel, wie bei einer normalen Unterhaltung zwischen Menschen*“ (VP. 2) vorstellen, und andere eher negativ, als eintönige Abfragen: „... [ich stelle mir vor]²²... *dass der Computer mir Fragen stellt, auflisten wird, welche Art von Restaurants ...also griechisch, italienisch und jeweils eine Zahl sagen würde und ich muss dann die Zahl nennen, dann werde ich weiter vermittelt zu einem italienischen Restaurant, als sozusagen Untermenü*“ (VP. 8). Eine sehr realitätsnahe Beschreibung eines Dialogablaufes mit dem System wird vom (VP.4) geschildert: „*Das beste wäre, wenn man sich so unterhalten würde oder könnte wie mit einem Menschen. Aber wenn das nicht geht, dann, dass er z.B. sagt : „Was wollen sie wissen?“, erst mal in Kategorien, in Oberbegriffe ... Art des Essen oder Preise oder was weiß ich... und dann muss man entsprechend eine Zahl angeben, wo das Gespräch hinführen soll und danach wird gefragt, ja von dieser Kategorie von Essen gibt's wiederum das und das und das. „Worüber möchten Sie etwas näheres erfahren?“ usw.. Dann entscheidet man sich für irgendetwas, er stellt so zu sagen verschiedene Richtungen vor, in die das Gespräch gehen kann und dann hat man die Entscheidung dazu. Ja gut, ich möchte über das erfahren. Dann erzählt er irgendwas...*“ (VP. 4)

Die Hälfte der Befragten fügten noch hinzu, sie würden gerne Fragen stellen.

Die Mehrheit der Befragten (7 Nennungen) glaubt, es sei unkompliziert mit einem Sprachsystem zu interagieren (4.3), aber sie äußern gleichzeitig ihr Misstrauen dem Computer gegenüber (3 Nennungen): „... *ich zweifle, ob ja... ein sinnvolles Gespräch sein wird, ob ich dann nicht wirklich in der Mitte auflegen werde und meine, ich rede jetzt bloß nur mit einem Computer, ich komme nicht weiter*“ (VP.1); „*Es kann den Menschen nicht ersetzen glaube ich, irgendwo bleibt es mangelhaft.*“ (VP. 5).

Gute Hinweise zur Informationsgestaltung (4.4) und zu den Servicemöglichkeiten (4.5) eines Dialogsystems liefern die Antworten auf die nächsten beiden Fragen. Die Versuchspersonen gaben an, sie wollen vom System über folgende Kategorien informiert werden: Art der Küche, Preisklasse, Adresse, Telefonnummer, Ort, Wegbeschreibung, Reservierung, Sonderangebote und Öffnungszeiten. Als Servicemöglichkeiten (4.6) soll das System den Kunden 24 Stunden zur Verfügung stehen und ihnen die Möglichkeit bieten mit einem Callservice verbunden zu werden; es soll die Informationen auch schriftlich liefern können (per Post oder Fax). Das System soll vor allem schnell sein (4 Nennungen), nicht abstürzen und über mehrere telefonische Leitungen verfügen, um lange Wartezeiten zu vermeiden. Die Systemantworten sollen kohärent und sinnvoll sein. Das System bräuchte eine flexible Dialoggestaltung, die dem Benutzer ermöglichen sollte, sich diejenige Information herauszusuchen, die er hören will. Im Idealfall sollte man gar nicht

²² Um den Sinn des Zitates zu verdeutlichen wurde den Satz vervollständigt. Die Vervollständigung steht in Klammern .

merken, dass man mit einem Computer spricht. Die Freundlichkeit des Systems soll für ein angenehmes Gespräch sorgen.

Als Vorteile sehen die Versuchspersonen folgende Aspekte: durch die Inanspruchnahme der Systemdienstleistungen (5.0) können Zeit (1 Nennung) und Kosten (5 Nennungen) gespart werden; auch für den Betreiber könnte die Anwendung eines solchen Systems kostenreduzierend sein. Das System wird als direkte Informationsquelle (1 Nennung) immer auf dem Laufenden sein (1 Nennung) und könnte 24 Stunden zur Verfügung stehen (3 Nennungen).

Als nachteilig wird die vermutete Begrenztheit des Systems bezeichnet, seine Unfähigkeit z.B. spontane Fragen zu beantworten. Gefürchtet werden auch Kommunikationsprobleme, unzureichende Informationen, Unverständlichkeit, Spracherkennungsprobleme und vor allem das Fehlen des persönlichen Kontaktes (vier der Befragten ziehen den Menschen einem Sprachcomputer vor).

6.3.2.2 Auswertung des zweiten Fragebogens

Der zweite Fragebogen soll die Erfahrungen der Befragten mit dem System dokumentieren. Die Versuchspersonen wurden zuerst aufgefordert, ihre Telefonate mit BoRIS_detailliert zu beschreiben und zu unterscheiden, was ihnen dabei positiv auffiel und was sie gestört hat.

Die erste Frage (6.0) bezieht sich rückblickend auf die Erwartungen der Versuchspersonen dem System gegenüber. Die Befragte sollten entscheiden, inwieweit ihre Erwartungen, sofern sie überhaupt welche hatten, erfüllt wurden. Wenn man das Erwartungsbild, welches im ersten Teil des Fragebogens konkretisiert wurde, mit den Antworten vergleicht, die bei der erneuten Frage über die Erwartungen gegeben wurden, wird man feststellen, dass diejenigen, die sehr diffuse und nicht richtig definierte Erwartungen hatten, nachher gemischte oder positive Eindrücke über das System äußerten. Dagegen sind diejenigen, die schon gewisse Vorstellungen über die Funktionsweise des Systems vor dem Versuch hatten, nachher negativ eingestellt.

Die Befragte werden zum Zwecke der Analyse in zwei Hälften, je nach Erwartungsbild aufgeteilt: die erste Hälfte hatte vordefinierte Erwartungen, die gleichzeitig als Vergleichsmuster fungiert. Die an diesem Vergleichsmaßstab gemessenen Systemleistungen wurden allgemein als schlecht bewertet (zwei Befragte gaben an, ihre Erwartungen würden dadurch erfüllt, dass sie sich von vorneherein das System als unzureichend vorgestellt haben). Eine Versuchsperson erklärte: „ *es war interessant und hat sich es herausgestellt, dass manche Dinge auch so waren, wie ich es mir vorgestellt habe. Im Prinzip ja ..., von dem Ablauf her, von der Struktur, von dem Prinzip her, aber ich hatte es mir in dem Sinne vorgestellt, dass die Information besser ist.*“ (VP. 4).

Die andere Hälfte hatte keine vordefinierten Erwartungen; dabei hat BoRIS (mit zwei positiven und zwei teilweise positiven Urteilen) besser abgeschnitten. Diese Gruppe konnte, dadurch, dass sie keine konkreten Vorstellungen vor dem Versuch über das System hatte, nur feststellen, ob das System gut oder schlecht funktionierte und war am Ende der Experimente von der Systemleistungen positiv überrascht: „... *dass es funktioniert hat, in den meisten Fälle. Das hat mich beeindruckt.*“ (VP. 5).

Nur eine einzige Versuchsperson gab an, ihre Erwartungen seien nicht erfüllt worden: „... *es entsprach nicht meiner Vorstellungen, denn ich habe mehr klare Frage erwartet, dass man gut weiß, was man antworten muss. Am Anfang erzählt er so viel Sachen und weißt du nicht, ob er dich verstehen wird, ob du normal sprechen sollst oder nicht. Ich habe eigentlich kürzere Fragen erwartet*“ (VP. 6)

Tab. 7 Erwartungen und Bewertungen der Systemleistungen

VP. nr.	Vordefinierte Erwartungen	VP. nr.	Keine vordefinierte Erwartungen
3	<p>Vor dem Versuch: negative vordefinierte Erwartungen</p> <p>Nach dem Versuch: Erwartungen wurden erfüllt</p> <p>Bewertung: das System wird als schlecht beurteilt</p>	1	<p>Vor dem Versuch: keine Erwartungen</p> <p>Nach dem Versuch: gemischten Eindruck über das System</p> <p>Bewertung: das System wird als teilweise gut beurteilt</p>
4	<p>Vor dem Versuch: positive vordefinierte Erwartungen</p> <p>Nach dem Versuch: teilweise erfüllte Erwartungen</p> <p>Bewertung: das System wird teilweise gut /teilweise schlecht beurteilt</p>	2	<p>Vor dem Versuch: keine Erwartungen</p> <p>Nach dem Versuch: gemischten Eindruck über das System</p> <p>Bewertung: das System wird als teilweise gut beurteilt</p>
8	<p>Vor dem Versuch: negative vordefinierte Erwartungen</p> <p>Nach dem Versuch: erfüllte Erwartungen</p> <p>Bewertung: das System wird als schlecht beurteilt</p>	7	<p>Vor dem Versuch: keine Erwartungen</p> <p>Nach dem Versuch: guten Eindruck über das System</p> <p>Bewertung: das System wird als gut beurteilt</p>
6	<p>Vor dem Versuch: -positive vordefinierte Erwartungen</p> <p>Nach dem Versuch: -nicht erfüllte Erwartungen</p> <p>Bewertung: das System wird als schlecht beurteilt</p>	5	<p>Vor dem Versuch: keine Erwartungen</p> <p>Nach dem Versuch: guten Eindruck über das System</p> <p>Bewertung: das System wird als gut beurteilt</p>

Der Fragebogen konzentriert sich des weiteren auf die pauschale Bewertung des Systems. Die Versuchspersonen sollten einerseits erklären, was ihnen am System gefiel (7.0), und andererseits, was sie beim System störte (8.0) und an welchen Stellen genau sie Probleme

hatten. In diesem Teil der Auswertung kommen die meisten Qualitätsfaktoren zum Ausdruck.

Von der Hälfte der Teilnehmer wurde neben der angenehmen Stimme der Ansagen die Menüführung sehr geschätzt, die auf dem QF. "Informationsgehalt" beruht : „...*dass das Menü wirklich fast alles umfasst, was man von einem Auskunftssystem erwartet*“ (VP. 1). Die Schnelligkeit der Verarbeitung, wurde auch positiv hervorgehoben, jedoch nur von drei Befragten (dadurch, dass die Sprachverstehenskomponente ausgeschaltet und die Eingaben von der Versuchsleiterin eingetippt wurden, führte dies zu unterschiedlich schnellen Systemreaktionen).

Interessant ist, dass manche Versuchspersonen Attribute dem System zuschreiben, die normalerweise Menschen zugesprochen werden; so wird BoRIS als „*recht sympathisch aufgezogen*“ (VP. 2) oder als „*sehr höflich*“ (VP. 7) beschrieben. Positiv betrachtet wird die Möglichkeit überhaupt von einem solchen computergestützten Sprachsystem Informationen zu bekommen (1 Nennung).

Am meisten von allen Befragten wurde die Sprachsynthese (QF. "Verständlichkeit") kritisiert: sie sei zu schlecht verständlich (7 Nennungen) „*unsympathisch*“ (VP. 2) (1 Nennung) und zu schnell (5 Nennungen). Andererseits musste man oft mit langen Pausen zwischen den Eingaben und den Antworten rechnen (3 Nennungen); zwei der Befragten schlugen vor, der Computer sollte diese Pausen verbal signalisieren (QF. „Bestätigung“): „...*das dauert ziemlich lang, bis er was sagt und dann fühle ich mich verunsichert, weil ich denke, ich habe was falsches gemacht..., dass man vielleicht etwas dazwischen schaltet, wie „Warten Sie ein Moment!“ oder so... Dann ist man mindestens sicher, es ist in Ordnung*“ (VP. 7).

Manche Befragten (3 Nennungen) beklagten mit zu vielen Kriterien der Informationssuche überschüttet zu sein (QF. „Verständlichkeit“) und hinterher mit zu wenigen Auswahlmöglichkeiten (4 Nennungen): Probleme ergaben sich vor allem bei der Ortauswahl und Küchenart: „*Bei dem dritten [Szenario] kam ich nicht weiter. Arabisch gab's nicht. Ich hätte fragen sollen, nach arabisch, ja... ich habe aber nicht gemacht, weil er mir das nicht angeboten hat und hat er mir auch kein „zum Beispiel“ gegeben. Deswegen habe ich das nicht so richtig gefunden. Es war eine Irreführung*“ (VP. 3).

Für die Hälfte der Befragten ist es nicht klar, wie sie ihre Äußerungen gestalten sollten (5 Nennungen insgesamt): ob man langsam, betont deutlich, satzartig oder stichwortartig sprechen sollten (1 Nennung), wie man eine Aussage modifiziert (4 Nennungen), ob man als PreisEingabe „billig“ oder direkt eine Zahl (z.B. 10 DM) sagen darf (1 Nennung) - hier wird der QF „Transparenz“ angesprochen. Dieser Kritikpunkt wird auch damit begründet, dass das System den Versuchspersonen keine Hinweise lieferte, warum kein passendes Restaurant gefunden wurde (3 Nennungen): „*Auf einmal, als ich mich um preiswerte Restaurants erkundigen wollte, schien es irgendwo, dass es eine untere Grenze gab, unter der er nicht verstanden hat. Ich habe wirklich eine dumme Summe gesagt, von 4 DM, aber*

das muss er doch versteht können und dann angeben, dass es nichts dazu gibt.[...]Er sollte sagen, „es gibt's nichts um diesen Preis, wollen Sie Ihre Daten verändern?“ (VP. 1). Die Befragten bekamen keinen Einblick in die Vorgehensweise des Systems und kein Feedback, ob das System ihre Eingaben verstanden hat (4 Nennung)- QF. "Transparenz" und "Bestätigung". Eine Versuchsperson beschwerte sich, dass das System sich nicht behilflich zeigte: „Und wenn man sich einmal vertan hat, dann weiß man irgendwie nicht mehr weiter. Dann hängt man so in einer Schleife...“ (VP. 5).

Die Befragten waren unzufrieden, weil das System überraschend Gespräche abbrach- (3 Nennungen), seine Restaurantansagen nicht wiederholen konnte (3 Nennungen) und keine Möglichkeit bietet, Nachfragen zu stellen (1 Nennung).- hier werden die QF. "Dialogunterbrechungen", "Wiederholung" und "Dialoginitiative" angesprochen.

Die Hälfte der Befragten hätte sich eine einfachere Menüführung gewünscht (ZF.²³ "Einschätzung des Gesprächsaufbaus"): *„Es wäre besser, wenn er getrennte Fragen gestellt hätte.“ (VP. 6).*

Kritisiert wurde auch der Informationsinhalt (4 Nennungen insgesamt): das System gab nicht an, wo genau das Restaurant lag, ob es z.B. in der Nähe der Innenstadt lag (2 Nennung); es lieferte auch keine Indizien über Service und Ambiente (1 Nennung) und für manche Küchenarten, z.B. „arabisch“ oder „lateinamerikanisch“ gab es keine Auswahl der Ortschaften (3 Nennungen) – QF. "Informationsgehalt".

Ein Lerneffekt machte sich nach den ersten zwei, drei Telefonaten mit BoRIS spürbar: die Versuchspersonen merkten im Laufe der Experimente, dass *„man gezielt was fragen muss, ansonsten klappt das nicht“ (VP. 2)*, dass man nicht Fragen wie *„Was empfehlen Sie mir?“ (VP. 6)* stellen darf, als ob man mit einer Person sprechen würde (1 Nennung) und wie sie ihre Eingaben modifizieren sollten (QF. "Einfachheit der Systembedienung"): *„Hinterher aber hatte ich keine Probleme mit dem System mehr. Hinterher wusste ich schon ... z.B. bei diesem „Coco Loco“, das letzte, da wusste ich inzwischen, wie er denkt. Ich wusste, dass ich nicht nach „Coco Loco“ fragen darf, sondern, dass ich nach einem spanischen Restaurant in Weitmar fragen muss, weil ich dann wusste, dass er das nicht kann. Aber wenn das mein erstes Gespräch mit BoRIS gewesen wäre, hätte ich einfach gefragt: Wo ist das „Coco Loco“? Und das hätte er nicht gekonnt“ (VP. 7).*

Des weiteren versucht der Fragebogen noch andere Kritikpunkte zu entdecken, indem er konkrete Fragen zu möglichen Systemmängeln (**9.0**) stellt.

Die Hälfte der Befragten gab an, sie waren über die Systemgrenzen informiert und wussten welche Art von Informationen BoRIS ihnen liefern konnte (QF. "Systemgrenzen"). Zwei Befragten äußerten sich negativ: *„Nein, in dem ersten Gespräch ist dir nicht klar, wie viel Information er hat. Nach ein paar Anrufe merkst du schon, wie viel Information er hat“ (VP.6).* Die anderen zwei gaben an, nur teilweise gewusst zu haben, welche Informationen

²³ Z.F. steht für Zufriedenheitsfaktor.

BoRIS bieten kann und meinten, sie hätten noch Informationen über Restauranttelefonnummer und Reservierungsmöglichkeiten erwartet.

Tab. 8 Kritikpunktematrix

Kritikpunkte	Interview- Nr.							
	1	2	3	4	5	6	7	8
1. Sprachsynthese - zu schlecht	*	*	*	*	*		*	*
2. Sprachsynthese - zu schnell				*	*	*	*	*
3. Pausen – zu lang			*			*	*	
4. es werden zu viele Kriterien der Informationssuche genannt		*	*			*		
5. es werden nicht alle Auswahlmöglichkeiten bei der einzelnen Kategorien genannt		*	*	*				*
6. man weiß nicht, wie man seine Äußerungen gestalten soll		*			*	*	*	*
7. Keine Hinweise, warum das System eine ein Restaurant nicht gefunden hat	*				*			*
8. Schwierigkeiten bei den Modifikationen	*	*			*	*	*	*
9. kein Feedback vom System, ob es die Benutzeraussagen verstanden hat	*					*	*	*
10. keine Wiederholmöglichkeit			*	*		*		
11. Keine Nachfragemöglichkeit							*	
12. Abbruch des Gesprächs	*					*	*	
13. Kritik zum Infogehalt		*	*	*				*
12. Wunsch nach einer anderen Menüführung .	*			*		*	*	

Die meisten Versuchspersonen (7 Nennungen) wussten, nach dem ersten Kontakt mit BoRIS, wie sie ihre Informationssuche gestalten sollten (10.0) – nur eine Testperson erklärte, erst nach dem zweiten Anruf mehr darüber gewusst zu haben (QF. "Transparenz").

Das System äußert deutlich (11.0), was es von den Befragten wissen wollte (5 Nennungen), nur zwei Befragten hatten teilweise Probleme damit und einer beschwerte sich, dass ihm die Systemfragen nicht klar waren und insofern wusste er nicht, wie man seine Antworten entsprechend formulieren sollte (QF. "Verständlichkeit"): „am Anfang hat er –zig Sachen aufgezählt, so dass ich nicht wusste, wie ich antworten soll. Und bei den Modifikationen war mir auch nicht klar, ob ich die Frage mit „Ja“ oder „Nein“ beantworten soll oder einfach meinen Wunsch direkt lassen kann“ (VP. 6).

Die meisten Befragten (6 Nennungen) haben nur teilweise den Eindruck vom System verstanden zu sein (12.0); ein einziger Befragter beantwortete die Frage mit „Ja“ darüber

und ein anderer wollte sich dazu nicht äußern (QF. "Verständlichkeit"): „*ich kann manche Sachen nicht überprüfen: den Ort, die Uhrzeit, Tag. Er hat etwas mit den Informationen gemacht, aber was er gemacht hat, weiß ich es nicht*“ (VP. 3).

Der Fragebogen versucht folglich die Gesprächsführung (**13.0**) zu dokumentieren. Zuerst wurden die Versuchspersonen aufgefordert, ihren Gesprächsaufbau mit BoRIS zu charakterisieren (ZF. "Einschätzung des Gesprächsaufbaus"). Dabei ergaben sich sowohl positive auch als negative Bewertungen; auffällig ist, dass jede Versuchsperson mindestens eine negative Bewertung geäußert hat, aber nicht immer eine positive. Der Gesprächsaufbau wurde als logisch, sinnvoll, relativ einfach, relativ schnell und der Aufgabe angepasst beurteilt, aber auch als begrenzt, gewöhnungsbedürftig, schlecht strukturiert und nicht besonders angenehm.

Fünf Befragten gaben an, dass der Gesprächsfluss durch Unterbrechungen gestört wurde, aber nur drei Personen empfanden diese Störung als lästig (**14.0**).

Tab. 9 Gesprächunterbrechungen

Versuchsperson Nr.	Gespräch wurde unterbrochen	Wie oft	Als lästig empfunden
1.	Ja	2 x mal	Ja
2.	Nein	0	Nein
3.	Ja	1 x mal	Nein
4.	Nein	0	Nein
5.	Nein	0	Nein
6.	Ja	Mehrmals	Ja
7.	Ja	2-3 x mal	Ja
8.	Ja	Selten	Nein

Die Unterbrechungen, die den Gesprächsfluss stören, sind insofern für die Qualitätsbeurteilung der Dialogsteuerung relevant, wenn sie wahrgenommen und als lästig beurteilt werden. Die Tabelle zeigt, dass weniger als die Hälfte der Befragten (3 Nennungen) sich von solchen „Pannen“ im Gespräch stören ließen (QF. "Dialogunterbrechungen").

Ein wichtiges Kriterium für den Erfolg der Interaktion ist auch die Formulierung der Eingaben auf der Benutzerseite, die von Erwartungen und Hintergrundwissen bzw. Erfahrungen mit dem System bestimmt sind. Sie variieren vom einem Szenario zum anderen und deuten auf den Lerneffekt hin und auf das Streben der Versuchspersonen sich dem System anzupassen.

Die Versuchspersonen können hinsichtlich der Gestaltung ihrer Äußerungen (**15.0**) in drei Gruppen eingeteilt werden: die erste Gruppe gestaltete ihre Äußerungen ausschließlich wie in einem normalen Gespräch mit Sätzen (2 Nennungen), die zweite bildete ursprünglich

Sätze und, nach den ersten Erfahrungen, formulierten sie ihre Äußerungen stichwortartig (3 Nennungen) und die dritte Gruppe benutzte nur Stichwörter (3 Nennungen).

Tab. 10 Gestaltung der Äußerungen

Versuchsperson Nr.	Geschlecht	Sätze	Sätze/Stichwörter	Stichwörter
1.	Weiblich		*	
2.	Weiblich	*		
6.	Weiblich	*		
7.	Weiblich		*	
3.	Männlich			*
4.	Männlich		*	
5	Männlich			*
8	Männlich			*

Aus der Tabelle wird deutlich, dass die Gruppe der Männer eine deutliche Präferenz für Stichwörter hat; dagegen entschieden sich die Frauen Sätze zu bilden und erst wenn sie merkten, dass sie damit nicht weiter kommen, versuchten sie Stichwörter zu benutzen. Diese Vorgehensweise erklärt sich dadurch, dass die Versuchspersonen unterschiedliche Erwartungen und Hintergrundwissen haben.

Die Frauen hatten allgemein keine genauen, vordefinierten Erwartungen und kein Hintergrundinformationen zu Dialogssystemen. Zwei von ihnen waren vor dem Versuch zum System positiv eingestellt. Dieser Zusammenhang führte dazu, dass das Sprachverhalten der Frauen spontaner war als das der Männer. Die Frauen versuchten sich dem System anzupassen, aber gleichzeitig ihre normalen kommunikativen Gewohnheiten zu behalten :*„Ich habe Sätze gemacht, weil man in einem Gespräch daran gewöhnt ist, Sätze zu bilden, und weil es sich nicht wie eine Maschine am Anfang anhört, weil er ganz normal spricht. Dann machst du automatisch auch Sätze“* (VP. 6).

Die Gruppe der Männer hatte, bis auf zu einen Befragten, vordefinierte Erwartungen und zum Teil (2 Nennungen) gewisse Hintergrundinformationen: *„...ich kenne mich ein bisschen mit Computern aus und ich weiß, dass die Spracherkennung nicht optimal ist und dass sie wahrscheinlich auf Stichwörter reagieren wird.“* (VP. 8). Eine interessante Feststellung ist zu bemerken, dass diejenigen, die zum System negativ eingestellt waren, Stichwörtern benutzten. Die (VP. 4) war der einzige männliche Befragte, der positive Einstellungen hatte und der einzige seiner Gruppe, der sowohl Sätze als auch Stichwörter benutzte.

Tab. 11 Erwartungen, Einstellungen zum System nach dem Versuch und Hintergrundwissen

VP. nr.	Geschlecht	Erwartungen/Einstellungen zum System	Hintergrundwissen
1.	weiblich	- keine vordefinierte Erwartungen - positiv eingestellt	Nein
2.	weiblich	- keine vordefinierte Erwartungen - positiv eingestellt	Nein
6.	weiblich	- keine vordefinierte Erwartungen - keine Einstellung	Nein
7.	weiblich	- keine vordefinierte Erwartungen - keine Einstellung	Nein
3.	männlich	- vordefinierte Erwartungen - negativ eingestellt	Teilweise
4.	männlich	- vordefinierte Erwartungen - positiv eingestellt	Teilweise
5.	männlich	- keine Erwartungen - negativ eingestellt	Nein
8.	männlich	- vordefinierte Erwartungen - negativ eingestellt	Nein

Der Fragebogen konzentriert sich des weiteren auf Informationen, die das System lieferte (16.0). Über die Hälfte der Versuchspersonen (6 Nennungen) gab an, zum größten Teil oder sogar ganz (1 Nennung) die gewünschte Information bekommen zu haben und beurteilten den Dialog als ziemlich schnell (7 Nennungen); nur eine einzige Versuchsperson beschwerte sich: „*Das war überhaupt nicht deutlich. Wenn ich wirklich in einer richtigen Situation wäre, wo ich die Adresse und den Namen des Restaurants will, dann habe ich das nicht bekommen*“ (VP. 1). Kritisiert wurde das Fehlen der Reservierungsmöglichkeit, der Telefonnummer, der Auskunft über Service und Ambiente und Informationen über Kneipen. Obwohl der Informationsgehalt (17.0) als „*irgendwie begrenzt*“ (VP. 3, Seite, 6), „*mäßig*“ (VP. 4) und „*rudimentär*“ (VP. 5) bezeichnet wurde ist über die Hälfte (insgesamt 5 Nennungen) der Befragten damit zufrieden (3 Nennungen) und teilweise zufrieden (2 Nennungen).

Das System wirkt auf die Versuchspersonen vertrauenswürdig (18.0): sechs Befragten glauben daran, dass BoRIS ihnen richtige Informationen geliefert hat (ZF. "Einschätzung des Computers als Gesprächspartner"), ein Befragter ist nur teilweise davon überzeugt und ein anderer gab an: „*...das kann man nicht überprüfen. Oder ich kann es nicht nachprüfen*“ (VP.7).

Die Meinungen der Befragten, ob BoRIS eine Bestätigung (QF. "Bestätigung") zu den Benutzeraussagen liefern soll, sind geteilt (19.0): manche lehnten es kategorisch ab (2 Nennungen), andere stimmten nur bedingt zu (3 Nennungen) und begründeten ihre Einstellung damit, dass das Gespräch ziemlich lange dauern würde: „*Das wäre gut zum*

Schluss, aber nicht [...] im Laufe des Gesprächs“ (VP. 3); drei Befragten waren bedenkenlos dafür.

Das Gespräch mit BoRIS löste bei den Befragten unterschiedliche Reaktionen aus (**20.0**): die Befragten fühlten sich zum Teil wohl (2 Nennungen), angenehm überrascht (1 Nennung) aber auch als komisch (1 Nennung) und genervt (4 Nennungen); drei Befragten gaben an, das Gespräch mit BoRIS sei angenehm gewesen, einer entschied sich nur für „teilweise angenehm“ und drei andere beurteilten den Dialog als lästig und nervig (ZF. "Stimmung während des Gesprächs"). Ein Befragter äußerte sich nicht dazu.

Zum Schluss wurden die Versuchspersonen gefragt, ob sich der Anruf bei BoRIS gelohnt hat (**21.0**) und ob sie die Dienste eines solchen Systems noch in Anspruch nehmen wollten. Damit versuchte man herauszufinden, wie zufrieden die Befragten mit dem System waren (ZF. "Einschätzung der kommunikativen Zielerreichung"). Die Antworten waren wiederum gemischt: nur zwei Befragten gaben an, sie waren mit BoRIS zufrieden; die Hälfte der Befragten war, vor allem wegen der schlechten Sprachsynthese (3 Nennungen), teilweise zufrieden und zwei Versuchspersonen erklärten sich für unzufrieden. Positiv zu bemerken ist, dass nur ein einziger Befragter eindeutig ablehnte, die Dienste eines solchen Dialogsystems nochmals in Anspruch zu nehmen. Alle anderen zeigten sich bereit (3 Nennungen) oder teilweise bereit (1 Nennung), BoRIS für Auskunftsdienste wieder anzurufen, unter der Bedingung, dass das System hinsichtlich Sprachqualität, Informationsgehalt und Modifizierung der Eingaben verbessert wird.

Der Fragebogen endet mit der Frage (**22.0**) „Gibt es noch etwas, was Ihnen am BoRIS aufgefallen ist?“. Die Hälfte der Befragten hatten nichts mehr zu berichten. Die andere Hälfte machte einige zusätzliche Kommentare und Verbesserungsvorschläge über die gesamte Interaktion mit BoRIS: als positiv wurde die Struktur des Gesprächs, die Tonbandstimme und die kurze Gesprächsdauer genannt; negative Bemerkungen wurde über die Gesprächsunterbrechungen und über die englische Begrüßung von BoRIS am Ende des Gesprächs²⁴. Die Sprachsynthese sollte verbessert werden, das System sollte die Auswahl zwischen Kneipen und Restaurants erlauben und es sollte für Ausländer die Möglichkeit geben, die Ansagen auf Englisch zu hören.

FAZIT

1. Die meisten Befragten bevorzugten als Informationsquelle zur Restaurantsuche einen menschlichen Informanten oder schriftliches Material (Gelbe Seiten, Magazine).
2. Die wichtigsten Kriterien für die Restaurantentscheidung sind: Preis, Ambiente, Art der Küche, Ort des Restaurants und Qualität des Essens.
3. Fast alle Versuchspersonen, bis auf eine, hatten keinerlei Erfahrungen mit

²⁴ Das Programm spielt automatisch nach der Eingabe der „Transaction Success“ (siehe Abschnitt 5.1.3) noch ein „Good bye“ mit einer eigenen Synthesestimme.

Sprachauskunftssystemen.

4. Das Erwartungsbild vor dem Versuch ist sehr verschieden: die Hälfte der Befragten hatte keine konkreten Vorstellungen, die meisten erwarteten eine menschliche, freundliche Stimme, sie glaubten eher, das System sei nicht kompliziert zu bedienen, sie erwarten Missverständnisse.
5. Ein Restaurantsprachauskunftssystem sollte Informationen über Art der Küche, Preisklasse, Adresse, Telefonnummer, Ort, Wegbeschreibung, Reservierung, Sonderangebote und Öffnungszeiten enthalten.
6. Als mögliche Systemvorteile sehen die Befragten folgende Aspekte: es wird Zeit und Geld gespart, das System kann immer auf dem Laufenden sein, das System steht einem 24 h zu Verfügung.
Nachteilhaft wird vermutet: die Unfähigkeit des Systems spontane Fragen zu beantworten, Kommunikationsprobleme/Spracherkennungsprobleme, unzureichende Informationen, Unverständlichkeit, das Fehlen des persönlichen Kontaktes.
7. Je nach Erwartungsbild können die Versuchspersonen in zwei Kategorien eingeteilt werden: Versuchspersonen, die keine vordefinierte Erwartungen vor dem Versuch hatte und welche, die gewisse Vorstellungen aufzeigten.
8. Das System wurde unter folgenden Aspekten als gut bewertet: Menüführung, die Ansagestimme, teilweise die Schnelligkeit der Verarbeitung.
9. Die meist erwähnte Kritikpunkte waren: zu schnelle und zu schlechte Sprachsynthese, es werden nicht alle Auswahlmöglichkeiten der einzelnen Kategorien genannt, Unklarheit, wie man die Eingaben gestaltet und modifiziert werden sollen, das Fehlen jeglicher Bestätigungsmöglichkeiten, Kritik zum Informationsgehalt.
10. Die männlichen Versuchspersonen gestalteten ihre Äußerungen meistens stichwortartig. Die weiblichen Versuchspersonen kombinierten Sätzen mit Stichwörtern oder bildeten nur Sätze.
11. Die Hälfte der Befragten gab an, über die Systemgrenzen informiert zu sein.
12. Die meisten Befragten hatten nur teilweise das Gefühl vom System verstanden zu sein.
13. Die meisten Befragten haben ihren Informationswunsch in der Regel bekommen.
14. Das System wirkt vertrauenswürdig.
15. Eine Bestätigung der Eingaben ist nur teilweise erwünscht.
16. Der Anruf bei BoRIS hat sich nur teilweise gelohnt, vor allem wegen der schlechten Sprachsynthese.

6.4 Der Entwurf des quantitativen Interview

Die erste Befragungsrunde ermöglichte die Feststellung und Evaluierung nach qualitativen Methoden der Qualitäts- und Zufriedenheitsfaktoren und lieferte Hinweise auf relevante Aspekte der Kommunikation mit BoRIS. Mit ihrer Hilfe konnte Behauptungen und Fragen formuliert, welche die direkten Erfahrungen der Versuchspersonen mit BoRIS widerspiegeln.

Für den Entwurf des quantitativen Interviews gruppierte man diese Aspekte um die Qualitäts- bzw. Zufriedenheitsfaktoren. Die repräsentativste Aussagen (im kursiv) wurde aus den Interviews entnommen und unter den entsprechenden Faktor in die untere Tabelle (linke Spalte) eingetragen. Diese Aussagen dienten als Muster der Frageformulierung im Rahmen des quantitativen Interview (rechte Spalte).

I. Frage/Antwort Aspekte

1. Relevanz

Qualitatives Interview	Quantitatives Interview
<p>1.1 Informationsgehalt</p> <p>1.1.1 Informationsmenge</p> <p>[Der Informationsgehalt war] „ausreichend, im Sinne von genug, es reichte aus.“ (VP. 2.)</p> <p>[Der Informationsgehalt war] „mäßig“ (VP. 4) .</p> <p>„Das Menu umfasst fast alles, was man von einem Auskunftssystem erwartet“ (VP. 1).</p> <p>„Er [der Computer] hat mir Orte angeboten, in denen gar kein Restaurant gab, was ich haben wollte“ (VP. 8).</p> <p>1.1.2 Vollständigkeit der Informationen</p> <p>„Ich konnte nirgendwo erfahren, etwas über den Restaurantservice oder wie es drin aussieht.“ (VP. 4).</p> <p>„Ich habe ganz am Anfang erwartet, dass es irgendwelche Reservierungen gab und gab's nicht“ (VP. 1).</p>	<p>1.1 Informationsgehalt</p> <p>1.1.1 Informationsmenge</p> <p>9.2. BoRIS bietet umfangreiche Information über Bochumer Restaurants.</p> <p>11.6 Es ist wichtig, dass BoRIS passende Information gibt.</p> <p>1.1.2 Vollständigkeit der Informationen</p> <p>10.4 BoRIS bietet keine detaillierten Informationen über die gewünschten Restaurants.</p>

2. Ausdrucksform

Qualitatives Interview	Quantitatives Interview
<p>2.1 Akustische Verständlichkeit</p> <p>„Die Adresse und Namenansage der Restaurants war total schlecht“ (VP. 1).</p> <p>„Die wichtige Information ist nicht deutlich.“ (VP. 1).</p> <p>„Die Restaurantansage war zu schnell und schlecht zu verstehen, es kam zu viel und zu schnell hintereinander“ (VP. 8).</p>	<p>2.1 Akustische Verständlichkeit</p> <p>10.8 Die Ansagen der Restaurantnamen und die Adresse war nicht deutlich.</p>

<p>2.2 Inhaltliche Verständlichkeit [Das System hat mich] „nicht immer verstanden“ (VP. 4). „Ich weiß es nicht, ob er alle meine Wünsche verstanden hat“ (VP. 6). „Am Anfang kommen so viele Möglichkeiten, man kann sich nicht alles merken.“ (VP. 3) „Es sind zu viele Informationen am Anfang“ (VP. 3).</p>	<p>2.2 Inhaltliche Verständlichkeit 11.2 Es ist wichtig, dass BoRIS meine Eingaben gut versteht. 11.3 Es ist wichtig, dass BoRIS klare Fragen stellt 10.6 Ich konnte mir nicht die Auswahlkriterien am Anfang des Gesprächs nicht merken.</p>
<p>2.3 Gestaltung der Benutzereingaben [Ich habe die Fragen] „ganz normal“ [beantwortet]. „Also nicht stichwortartig“ (VP.2). Ich habe die Fragen] „stichwortartig [beantwortete], weil ich nur die wichtigsten Dingen erzählen wollte“ (VP.3).</p>	<p>2.3 Gestaltung der Benutzereingaben 9.11 Ich habe die Fragen von BoRIS beantwortet : <input type="checkbox"/> stichwortartig <input type="checkbox"/> satzartig <input type="checkbox"/> teils-teils</p>

3. Korrektur- und Aufklärungsstrategien

Qualitatives Interview	Quantitatives Interview
<p>3.1 Wiederholung 3.1.1. Wiederholung der Systemeingaben „Dann hat mich sehr gestört, das trifft auf alles zu, dass man keine wiederholende Möglichkeit der Auskunft hat“ (VP. 3). „Es gibt keine Wiederholungsmöglichkeit für die Adresse“ (VP. 8). „Es wäre nicht schlecht, wenn es ein Kommando gäbe, wie z.B. „Bitte wiederholen!“ (VP. 3). 3.1.2 Wiederholung der Benutzereingaben „Einmal habe ich nach deutschem Essen gefragt und hat er mich nicht verstanden und hat er nochmals gefragt, vielleicht, weil ich Ausländer bin und nicht deutlich ausspreche, aber das ist O.K., Du kannst natürlich wiederholen.“ (VP. 6).</p>	<p>3.1 Wiederholung 3.1.1 Wiederholung der Systemeingaben 10.3 BoRIS konnte seine Ansagen nicht wiederholen. 11.1 Es ist wichtig, dass es seine Fragen wiederholt, wenn ich nicht mehr weiter weiß. 3.1.2 Wiederholung der Benutzereingaben 9.5 Ich musste oft meine Eingaben wiederholen.</p>
<p>3.2 Bestätigung 3.2 1. Bestätigung der Eingaben [Das System soll] „nicht alle Antworten bestätigen, sondern nur zum Schluss. (VP. 3). „Wenn er mir eine Adresse gibt, weiß ich nicht, ob die Adresse ist, die ich brauche; hier wäre eine Bestätigung nicht schlecht.“ (VP. 6). 3.2.2 Bestätigung der Pausen [Die Pausen zwischen den Fragen und Antworten sind zu lang]. „Man sollte vielleicht etwas dazwischen schalten, wie „Warten Sie ein Moment bitte!“ oder so ..., dann weißt man, es ist in Ordnung“ (VP. 7).</p>	<p>3.2 Bestätigung 3.2 1. Bestätigung der Eingaben 11.11 Es ist wichtig, dass ich von BoRIS eine Bestätigung meiner Eingaben bekomme. 3.2.2 Bestätigung der Pausen 11.12 Es ist wichtig, dass die Pausen nach meinen Eingaben von einem „Warten Sie einen Moment“ ausgefüllt werden.</p>

II. Systemmöglichkeiten

4. Berücksichtigung kontextueller Bedingungen

Qualitatives Interview	Quantitatives Interview
<p>4.1 Hintergrundwissen</p> <p>4.1.1 Erfahrung mit dem System</p> <p>3.0 Wissen Sie, was ein Sprach-Informationssystem ist?</p> <p>3.1 Wenn ja, was verstehen Sie darunter (näher beschreiben)?</p> <p>3.1.1 Haben Sie ein solches System schon benutzt?</p> <p>3.1.2 Wie oft?</p> <p>3.1.3 Zu welchem Zweck?</p>	<p>4.1.Hintergrundwissen</p> <p>4.1.1 Erfahrung mit dem System</p> <p>3.0 Haben Sie bereit ein Sprach-Informationssystem benutzt?</p> <p>3.1. wenn ja, bei welcher Gelegenheit? _____</p> <p>3.1.1 Wie würden Sie Ihre Erfahrung charakterisieren?</p> <p>ausgezeichnet <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/> schlecht</p>
<p>4.1.2 Erfahrung mit dem Thema</p> <p>[4.4 Welche Informationen soll ein Restaurantsauskunftssystem anbieten?]</p> <p>„Art der Küche, Restaurants in meiner Nähe[…], die Adresse, die Telefonnr.“ (VP. 8)</p> <p>„Öffnungszeiten, was man essen kann, wie teuer ungefähr, vielleicht ob man reservieren muss, ob noch Tische frei sind, vielleicht Sonderangebote.“ (VP.6, Seite 2)</p> <p>„Öffnungszeiten und das Ambiente, wie dort aussieht, ob es geräumig ist oder eher klein. Die Lage, wie es zu erreichen ist, mit welchen Verkehrsmitteln.“ (VP. 4)</p>	<p>4.1.2 Erfahrung mit dem Thema</p> <p>4.0 Welche Informationen sollte ein solches Restaurantsauskunftssystem liefern?</p> <p>4.1 Name, Adresse, Telefonnr. <input type="checkbox"/></p> <p>4.2 Öffnungszeiten <input type="checkbox"/></p> <p>4.3 Preise <input type="checkbox"/></p> <p>4.4 Art der Küche <input type="checkbox"/></p> <p>4.5 Spezielle Speiseangebote <input type="checkbox"/></p> <p>4.6 Äußeres Ambiente <input type="checkbox"/></p> <p>4.7.Reservierungsmöglichkeiten <input type="checkbox"/></p> <p>4.8 Wegbeschreibung <input type="checkbox"/></p>
<p>4.2 Erwartungen</p> <p>4.2.1 Vor dem Test</p> <p>„Ich hoffe, dass es nicht so dumm wird, dass ich eine Frage stelle und keine Antwort bekommen.“ (VP. 1.)</p> <p>[Ich erwarte die Reaktion]“ eines Computers, der den Menschen ersetzen möchte“. (VP. 2)</p> <p>„Es kann dem Menschen nicht ersetzen, glaube ich, irgendwo bleibt es mangelhaft“ (VP. 5)-</p>	<p>4.2. Erwartungen</p> <p>4.2.1 Vor dem Test</p> <p>5.0 Wenn Sie bei BoRIS anrufen, erwarten Sie:</p> <p>5.1 ein vernünftiges Gespräch wie mit einem Menschen</p>
<p>„Ich habe mir gedacht, ich stelle Fragen und er antwortet mir, das ist , was ich erwarten würde“ (VP.7)</p>	<p>5.2 Sie stellen dem Computer Fragen</p>
<p>[Ich erwarte] „wahrscheinlich eher, dass das System Fragen stellt und dass ich die Antwort gebe“. (VP. 5)</p>	<p>5.3 Der Computer stellt Ihnen Fragen</p>
<p>„Maximal erwarte ich, dass ich eine Auskunft kriege, wie ich in Seiten auch finden würde, nur akustisch“ (VP.8)</p>	<p>5.4 Sie bekommen präzise Antworten</p>
<p>„Computer verstehe immer mies, was man möchte. Also Missverständnisse erwarte ich“. (VP. 8)</p>	<p>5.5 Es werden keine Missverständnisse entstehen.</p>
<p>„Es wird kein richtiges Gespräch statt finden, nur eine Auswahl-situation und das ist etwas umständliches. Man muss sich durch viele Menüs durchschlagen.“ (VP. 8)</p>	<p>5.6 Sie werden schnell die gewünschte Information bekommen</p>

[Ich erwarte], <i>dass der Computer ziemlich menschlich klingt.</i> (VP. 6) [Ich erwarte] <i>„eine freundliche Stimme“.</i> (VP. 4)	5.7 Der Computer hat eine freundliche Stimme
[Ich glaube], <i>„es wird nicht sehr kompliziert.“</i> (VP. 2,) <i>„Es wird komplizierter als mit einer Person zu reden“.</i> (VP. 3) <i>„Ich glaube nicht, dass es schwer zu bedienen ist“.</i> (VP. 5)	5.8 Das System ist leicht zu bedienen.
4.2.2.Nach dem Test <i>„Es war interessant und hat sich herausgestellt , dass manche Dinge auch so waren wie ich es mir vorgestellt habe.“</i> (VP.4)	4.2.2 Nach dem Test 9.1. BoRIS hat genau so funktioniert, wie ich es mir vorgestellt habe.

5. Transparenz

Qualitatives Interview	Quantitatives Interview
5.1 .Einfachheit der Systembedienung <i>„Es war nicht so kompliziert aber auch nicht so einfach“</i> [zu benutzen]. (VP. 6)	5.1 Einfachheit der Systembedienung 9.9 Jeder kann mit BoRIS ohne Probleme umgehen. 11.5 Es ist wichtig, dass BoRIS einfach zu bedienen ist.
5.2 Systemgrenzen 5.2.1 Eingabeformulierung und -modifizierung Eingabeformulierung <i>„Man merkt richtig schnell, dass man gezielt fragen muss, ansonsten klappt das nicht“.</i> (VP. 2) <i>„Man muss genau wissen, was man für Fragen stellt; man darf nicht ausschweifen“.</i> (VP. 2) <i>„Es war mir nicht klar, wie ich antworten soll.“</i> (VP. 6) Eingabemodifizierung <i>„Bei den Modifikationen war mir auch nicht klar, ob ich die Frage mit „Ja“ oder „Nein“ beantworten soll, oder einfach mein Wusch loslassen.“</i> (VP. 6)	5.2 Systemgrenzen 5.2.1 Eingabeformulierung und -modifizierung Eingabeformulierung 11.9 Es ist wichtig, dass ich von Anfang an weiß, wie ich meine Eingaben formulieren soll Eingabemodifizierung 9.10 Es war mir nicht klar, wie ich meine Eingaben modifizieren soll, um schnell die gewünschte Information zu bekommen. 11.13 Es ist wichtig, dass es einfach ist, die Eingaben zu modifizieren
5.2.2 Welche Art von Informationen das System bieten kann. <i>"Nein, in dem ersten Gespräch ist dir nicht klar, wie viel Information er hat. Nach ein paar Anrufen merkst Du schon, wie viel Information er hat"</i> (VP.6).	5.2.2. Welche Art von Informationen das System bieten kann 11.10 Es ist wichtig, dass ich von Anfang an weiß, welche Art von Informationen mir BoRIS liefern kann

6. Dialoginitiative

Qualitatives Interview	Quantitatives Interview
6.1 Möglichkeit Fragen stellen zu können <i>„Es ist ziemlich eingeengt, weil man keine Nachfrage stellen.“ (VP. 7)</i>	6.1 Möglichkeit Fragen stellen zu können 11.7 Es ist wichtig, dass ich BoRIS Fragen stellen kann.
6.2 Möglichkeit das Gespräch zu führen solange man möchte <i>„Er hat das Gespräch, so zu sagen wirklich beendet, das Gespräch wurde nicht von einem selbst beendet und es sollte es nicht so sein.“ (VP. 4)</i>	6.2 Möglichkeit das Gespräch zu führen solange man möchte 11.8 Es ist wichtig, dass nachdem ich die gewünschte Information bekommen habe, das Gespräch weiter führen kann, solange ich noch zusätzliche Fragen habe

7. Schnelligkeit

Qualitatives Interview	Quantitatives Interview
7.1 Schnelligkeit der Datenverarbeitung <i>[Ich fand es gut, dass] „dass System relativ schnell die Daten verarbeitet.“ (VP. 1)</i>	7.1 Schnelligkeit der Datenverarbeitung 9.7 Das Gespräch mit BoRIS hat schnell zur gewünschten Information geführt.
<i>„Bis die Antwort kam hat irrelevant gedauert.“ (VP. 3)</i> <i>„Er machte so lange Pausen“ (VP. 3)</i> <i>„Man musste zwischendurch ziemlich lange warten“ (VP. 7)</i>	11.4 Es ist wichtig, BoRIS schnell auf meine Eingaben reagiert. 10.7 Die Pausen zwischen meinen Eingaben und den Antworten von BoRIS waren zu lang.
7.1 Dialogunterbrechungen <i>„Wenn er etwas nicht erkannt hat, hat er das Gespräch abgebrochen, hat sich sofort verabschiedet. Das ist natürlich auch komisch für die Person, die anruft“ (VP. 6)</i> <i>„Bei den Modifikation hat sich das Gespräch unterbrochen, weil er mich nicht verstanden hat“ (VP. 6)</i>	7.1 Dialogunterbrechungen 10.1 BoRIS hat plötzlich das Gespräch unterbrochen und ist spontan zu einem vorherigen Menu zurückgekehrt. 10.2 BoRIS hat das Gespräch oft abgebrochen. 11.12 Es ist wichtig, dass das Gespräch ohne Unterbrechungen verläuft.

8. Hilfefunktion

<i>„Am Anfang ist schon unangenehm, weil da nicht eine Person ist und du nicht weißt, wie du dich verhalten musst.“ (VP.6)</i> <i>„Wenn man sich einmal vertan hat, dann weiß man irgendwie nicht mehr weiter. Dann hängt man so in einer Schleife“ (VP. 5)</i> <i>„Beim letzten war ich in einer Schleife, in diesem spanischen... da könnte ich alles modifizieren und es gab nie ein Restaurant. Bis ich dann aufgegeben habe.“ (VP. 8)</i>	11.15 Es ist wichtig, dass BoRIS mit einer Hilfefunktion ausgestattet ist
--	---

III. Zufriedenheit

1 Einschätzung der kommunikativen Zielerreichung

Qualitatives Interview	Quantitatives Interview
<p>„Ich habe im Prinzip die Information bekommen, die ich wollte.“ (VP. 7)</p> <p>„Bei mir hat es sich nicht gelohnt, weil ich die Adresse nicht verstanden habe“ (VP.6)</p>	<p>9.3 Ich bekam immer die gewünschte Antwort auf meine Eingaben</p> <p>9.8 Der Anruf bei BoRIS hat sich gelohnt</p>

2. Einschätzung des Computers als Gesprächspartner

Qualitatives Interview	Quantitatives Interview
<p>2.1 Ähnlichkeit des Systems mit einem menschlichen Gesprächspartner</p> <p>„Man merkt schon, es ist keine richtige Kommunikation; das ist so, als ob man auf Knöpfe drücken würde.“ (VP. 5)</p> <p>„Es war langsamer als mit einer Person zu sprechen und nicht so angenehm.“ (VP. 6)</p>	<p>2.1 Ähnlichkeit des Systems mit einem menschlichen Gesprächspartner</p> <p>9.4 Ich spreche lieber mit einem Menschen als mit BoRIS.</p> <p>9.5. Ich konnte mich äußern, wie in einem normalen Gespräch mit einem Menschen</p>
<p>2.2 Vertrauenswürdigkeit</p> <p>„Ich glaube schon, dass die Information richtig war“. (VP. 1)</p> <p>„Ich traue Computer nicht so viel zu“. (VP. 8)</p>	<p>2.2 Vertrauenswürdigkeit</p> <p>9.6 Ich bin davon überzeugt, dass BoRIS mir richtige Information geliefert hat</p>
<p>2.3. Stimmung während des Gesprächs</p> <p>[Ich empfand das Gespräch als] <i>lästig</i>. (VP. 3)</p> <p>„Am Anfang war ich total genervt.“ (VP. 7)</p> <p>„Es war auf jedem Fall nicht unangenehm“ [mit BoRIS zu telefonieren]. (VP.7)</p> <p>[Das Fehlen einer Bestätigung] „verunsichert einen vielleicht.“ (VP.2)</p> <p>[Ich war] „nicht immer“ zufrieden. (VP.6)</p>	<p>9.10 Durch das Gespräch mit BoRIS fühlte ich mich:</p> <p>wohl <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/> unwohl</p> <p>sicher <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/> unsicher</p> <p>ruhig <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/> aufgeregt</p> <p>zufrieden <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/> unzufrieden</p>

3.2 Einschätzung des Gesprächsaufbaus

Qualitatives Interview	Quantitatives Interview
<p>„Es war ein typischer Ablauf, es war logisch aber nicht voraussagbar“ (VP. 8)</p> <p>[Der Gesprächsaufbau] „war unkompliziert.“ (VP. 3)</p> <p>[Der Gesprächsaufbau] „war gewöhnungsbedürftig.“ (VP. 4)</p> <p>„Was mich aber mehr gestört hat waren die Modifikationen: bei letzten musste ich alle durchgehen und am Ende hat er mir gesagt: kein Restaurant“. (VP.8)</p>	<p>10.9 Der Weg bis zu der gewünschten Information war umständlich.</p>

6.5 Das quantitative Interview

Die zweite Befragungsrunde wurde nach den Prinzipien der quantitativen Verfahren gestaltet und, wie das qualitative Interview, in zwei Etappen durchgeführt: vor und nach dem Versuch.

Das quantitative Interview wurde als Kombination von offenen und geschlossenen Antwortvorgaben gestaltet. Es wurden mehreren Typen von Fragen gestellt: Multiple Choice mit verschiedenen Antwortvorgaben (Fragen 1.0, 2.0, 4.0, 6.0, 2.11) und Mehrfachnennungen (dieselben Fragen außer Frage 6.0); „ja-nein“ Fragen (3.1. und 3.2) mit zusätzlichen Kategorien „teilweise“ und „weiß-nicht“ (7.0); offene Fragen (3.2, 8.2 und 8.3). Frage 8.1 hat als Antwortmöglichkeit die Bodden-Jekosch Skala [Bodden und Jekosch, 1996].

Die Fragestellung des Interviews bestand hauptsächlich aus Behauptungen (Statements), die die Versuchsperson zur Erkundung ihrer Einstellungen und Erfahrungen mit dem System einsetzte. Als Antwortmöglichkeit wurde eine fünfstufige Skala verwendet (Fragen 3.2.1, 5.0 und 9.0-9.10 und 10.1-11.15):

trifft zu trifft nicht zu

Man hat die Entscheidung getroffen, die Interviewfragen als Behauptungen zu formulieren, dadurch, dass mit ihrer Hilfe die zu untersuchenden Merkmale sich prononcierter und differenzierter erfassen lassen (siehe Abschnitt 6.3). Die Behauptungen veranlassen durch eine zum Teil provozierende Wortwahl auch zweifelnde, unsichere Befragungspersonen zu einer eindeutigen Stellungnahmen.

Man hat darauf geachtet, dass die Fragen und Behauptungen kurz formuliert sind, einfache Worte enthalten, dies im Sinne von nicht gebräuchlichen Fachausdrücken, Fremdwörtern, Abkürzungen oder Slangausdrücken, und dass sie sich nur auf einen einzigen Sachverhalt beziehen (Bortz und Döring, 1995).

Die letzten zwei Fragen des ersten Fragebogens, 6.0 und 7.0, sowie 8.3 entstanden in Anlehnung an [Loderer, 1998].

6.5.1 Aufbau

6.5.1.1 Aufbau des ersten Fragebogens

Der erste Fragebogen, vor dem Versuch, wurde nach der Struktur und Fragestellung dem ersten qualitativen Fragebogen nachempfunden. Die ersten Fragen zu persönlichen Daten der Versuchspersonen (Geschlecht, Alter, Beruf, Herkunftsregion und jetziger Wohnort) sowie Frage 1.0 und 2.0 zur Einstellungen der Versuchspersonen zum Thema Restaurantsuche sind mit ihrem qualitativen Muster identisch.

Bei der Frage 3.0 wurden die Versuchspersonen gefragt, ob sie Erfahrungen mit Sprach-Informationssystem hatten; sie sollten auch angeben, bei welcher Gelegenheit sie ein solches System benutzt haben und mit welchem Erfolg. Diese Informationen liefern Hinweise über die Erwartungen und das Hintergrundwissen der Versuchspersonen.

Die Frage 4.0 fasst die Themen der ersten zwei Fragen zusammen; die Versuchspersonen wurden aufgefordert sich Gedanken zu machen, welche Auskünfte ein Sprach-Informationssystem geben soll [Dudda, 2000]. Die Frage soll zeigen, ob die Datenbank alle relevanten Informationen enthält und ob das System geeignete Auswahlmöglichkeiten bietet.

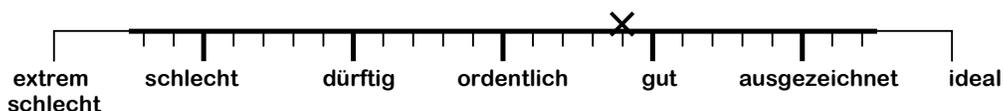
Die Frage 5.0 ist für das Erwartungsbild der Versuchsperson relevant: dabei sollten sie einschätzen, inwieweit das System einem menschlichen Gesprächspartner ähneln wird, ob sie Fragen dem System stellen werden oder vom System befragt werden; ob sie präzise Antworten und keine Missverständnisse erwarten; ob sie glauben, dass sie die gewünschte Information schnell bekommen werden und dass das System leicht zu bedienen ist; ob sie erwarten, dass der Computer eine freundliche Stimme hat.

Frage 6.0 erkundet, welche Gesprächsart die Versuchspersonen bevorzugen: ob sie lieber Fragen stellen oder Fragen beantworten. Daraus kann eine Schlussfolgerung über die Benutzerfreundlichkeit des Systems aus der Sicht der Versuchsperson gezogen werden, die sich auf die Qualitätsbeurteilung des Gesamtsystems auswirken kann.

Am Ende wurden die Versuchspersonen gefragt, ob die Sprach-Informationssysteme eine Alternative zu derzeitigen Informationssystemen darstellen könnten (7.0). Die Frage soll abschließend die positiven bzw. negativen Einstellungen der Versuchsperson dem System gegenüber aufzeigen.

6.5.1.2 Aufbau des zweiten Fragebogens

Der Aufbau des zweiten Fragenbogens, nach dem Versuch, unterscheidet sich wesentlich von der qualitativen Variante. Der Fragebogen besteht aus vier Teilen: der erste Teil soll Auskunft über das allgemeine Qualitätsurteil über BoRIS geben; bei der Frage 8.1 wurden die Versuchspersonen aufgefordert, den gesamten Eindruck von BoRIS auf einer siebenstufigen Skala einzuschätzen (Bodden-Jekosch Skala), die Ähnlichkeiten mit Temperaturskalen aufweist. Die Skala besteht aus 30 Einheiten und die Versuchspersonen mussten auf der Skala ihr Urteil durch einen Kreuz angeben:



Danach wurden die Versuchspersonen gebeten, das System einem Freund zu beschreiben (Frage 8.2) und Vorschläge zur Systemverbesserung zu machen (Frage 8.3).

Diese drei Fragen wurden absichtlich auf der ersten Seite gestellt, damit sich die Befragten bei ihrem Qualitätsurteil nicht von den andern Statements beeinflussen lassen. Sowohl die Systembeschreibung als auch die Verbesserungsvorschläge liefern Hinweise auf die Gründe, welche die Versuchspersonen zu ihrem Qualitätsurteil veranlasst haben und sind somit für die Evaluierung des Systems sehr wichtig.

Der zweite und dritte Teil bilden zusammen eine Art Diagnostiktest, in dem eine Reihe von elf bzw. zehn Behauptungen aufgestellt werden. Die Behauptungen schildern eine Situation, zu denen sich die Befragten äußern sollten: dabei werden sie entscheiden, inwieweit die dargestellten Sachverhalte ihrer Erfahrung mit BoRIS entsprechen. Die Behauptungen wurden aufgrund von positiven und negativen Bewertungen zusammengestellt, welche sich aus dem qualitativen Interview ergaben.

Die Behauptungen des ersten Teils wurden, außer Frage 9.4, als positive Äußerungen formuliert. Die Frage 9.4 („Ich spreche lieber mit einem Menschen als mit BoRIS“) wurde aus strategischen Gründen dem System gegenüber eher negativ formuliert; die Befragten können sich durch übertrieben provokative Behauptungen irritiert fühlen, was sich auf ihr Urteil auswirken könnte.

Bei der Frage 9.1 sollten sich die Befragten dazu äußern, ob das System so funktioniert hat, wie sie es sich vorgestellt haben. Dieses Statement wird im Zusammenhang mit der Frage 5.0 des ersten Fragebogens ausgewertet. Sie wurde insofern gestellt, weil die Ergebnisse der Auswertung des qualitativen Interviews gezeigt haben, wie wichtig die Erwartungen für das Qualitätsurteil der Versuchspersonen ist.

Die Frage 9.2 behauptet, dass BoRIS umfangreiche Informationen über Restaurants in Bochum bietet. Sie wurde gestellt, um zu überprüfen, für wie gut der Informationsgehalt gefunden wird.

Die Fragen 9.3 und 9.8 untersuchen, ob die Befragten immer die gewünschte Antwort auf ihre Eingaben bekamen und inwieweit sich der Anruf bei BoRIS gelohnt hat; dieses Statement bezieht sich auf die kommunikative Zielerreichung und zeigt, wie zufrieden die Versuchspersonen mit dem System waren.

Der Fragebogen stellt weitere Äußerungen, nach denen BoRIS mit einem menschlichen Gesprächspartner verglichen wird (Fragen 9.4 und 9.5). Der Mensch ist als Gesprächspartner ein Referenzpunkt für die qualitative Gestaltung eines maschinellen Dialogpartners und somit ist jeder Annäherungsversuch ein weiteren Schritt in die Qualitätsrichtung.

Die Frage 9.6 will herausstellen, ob die Befragten davon überzeugt sind, dass BoRIS ihnen richtige Information geliefert hat. Damit möchte man untersuchen, wie vertrauenswürdig das System den Befragten erscheint.

Im zweiten Teil wurden noch Fragen zur Schnelligkeit des Systems (Frage 9.7), zur Einfachheit der Bedienung (Frage 9.9) und zur Stimmung der Befragten während der Gespräche mit BoRIS (Frage 9.10) gestellt. Das Statement der Frage 9.9 („Jeder kann mit BoRIS ohne Probleme umgehen“) soll darüber Auskunft geben, wie leicht sich mit BoRIS umgehen lässt. Man hat absichtlich eine direkte Formulierung, wie „BoRIS ist leicht zu bedienen“ vermieden, um dadurch falsche Aussagen zu verhindern: es könnte vorkommen, dass manche Versuchspersonen zu stolz wären, um zuzugeben, dass sie Schwierigkeiten mit dem System hatten.

Für die Diagnostizierung der Stimmung der Testpersonen während des Versuchs wurden aufgrund der Ergebnissen qualitativer Interviews vier Antonympaare gefunden: wohl/unwohl, sicher/unsicher, ruhig/aufgeregt und zufrieden/unzufrieden. Die Versuchspersonen sollten ihre Stimmung auf einer fünfstufigen Skala definieren:

wohl unwohl

Der psychische Zustand der Versuchspersonen ist ein Messinstrument für den Stress, welcher mit den Gesprächen verbunden ist. Er liefert somit Anhaltspunkte für die Zufriedenheit der Versuchspersonen ihrem Dialogpartner gegenüber.

Zuletzt sollten die Versuchspersonen angeben, wie sie ihre Antworten formuliert haben: satzartig, stichwortartig oder teils-teils.

Der dritte Teil enthält negativ formulierte Behauptungen, die sich auf die Systemschwächen beziehen. Die Mängel wurden im Laufe der Experimente durch die Versuchspersonen selbst festgestellt und bei der qualitativen Befragung angesprochen. Die Behauptungen enthalten die häufigsten erwähnten Kritikpunkte (siehe Abschnitt 6.3.3); die Befragten wurden aufgefordert, anzugeben, ob sie von den vorgestellten Situationen sehr gestört wurden.

Die Statements 10.1 und 10.2 sprechen die Dialogunterbrechungen an und entsprechen dem Kritikpunkt 12.

10.3 bezieht sich auf die Beschwerde, dass BoRIS die Restaurantadressen nicht wiederholen konnte (Kritikpunkt 10).

10.4 spricht die Mängel bezüglich detaillierter Informationen an (Kritikpunkt 13).

10.5 behauptet, dass die Versuchspersonen oft ihre Eingaben wiederholen musste (abgeleitet aus dem Kritikpunkt 6).

10.6 gibt an, dass die Befragten sich die Auswahlkriterien am Anfang des Gesprächs nicht merken konnten. Damit möchte man herausfinden, ob die Systemansagen zu lang waren (Kritikpunkt 5).

10.7 bezieht sich auf die Länge der Pausen zwischen den Eingaben und den Antworten (Kritikpunkt 3).

10.8 spricht die Undeutlichkeit der Sprachsynthese an (Kritikpunkt 1 und 2).

10.9 behauptet, dass der Weg bis zur gewünschten Information umständlich war und wurde indirekt aus dem Kritikpunkt 12 abgeleitet.

10.10 wurde aus zwei verschiedenen Kritikpunkten zusammengestellt (Kritikpunkt 7 und 8) und bezieht sich auf einen viel kritisierten Sachverhalt: den Benutzern ist es nicht klar, wie sie ihre Eingaben modifizieren sollen.

Der vierte Teil des Fragebogens enthält eine Reihe von fünfzehn Statements und jedes Statement entspricht einem Qualitätsfaktor. Die Behauptungen haben einen hypothetischen Charakter und beziehen sich auf eine Idealsituation. Die Befragten sollten aufgrund ihrer Erfahrung mit BoRIS entscheiden, welche Kriterien für einen guten Gesprächsverlauf wichtig sind. Die Behauptung wurden absichtlich trivial formuliert (z. B. „Es ist wichtig, dass BoRIS meine Eingaben immer gut versteht), mit dem Ziel eine statistische Rangordnung der Qualitätsfaktoren herzustellen.

11.1 spricht die Notwendigkeit der Versuchspersonen zur Aufklärung an: BoRIS soll seine Fragen wiederholen, wenn der Benutzer nicht mehr weiter weiß.

11.2 und 11.3 beziehen sich auf die Verständlichkeit der Eingaben: BoRIS soll die Benutzereingaben immer gut verstehen und selber klare Fragen stellen.

11.4 und 11.5 sprechen die Schnelligkeit und die Einfachheit der Bedienung an.

11.6 bezieht sich auf die Relevanz der Informationen; BoRIS soll passende Informationen liefern.

11.7 und 11.8 beziehen sich auf die Dialoginitiative: die Versuchspersonen sollten beurteilen, wie wichtig für einen guten Gesprächsverlauf ist, Fragen und Nachfragen stellen zu können.

11.9 und 11.10 wurde gestellt, um die Wichtigkeit der Systemtransparenz zu überprüfen. Die Versuchspersonen sollten entscheiden, ob und wie wichtig es ist, von Anfang an zu wissen, wie sie ihre Eingaben formulieren sollen und welche Art von Informationen sie von BoRIS bekommen können.

11.11 und 11.12 beziehen sich auf Möglichkeiten, die das System nicht anbietet: es soll dabei herausgestellt werden, inwieweit eine Bestätigung der gemachten Eingaben, sowie das Signalisieren der Verarbeitungszeit der Daten mit einem Satz wie „Warten Sie ein Moment!“ erwünscht sind. Darauf wurde im qualitativen Interview von einigen Versuchspersonen hingewiesen und hier versucht man zu erfahren, ob die meisten Versuchspersonen sie tatsächlich vermisst haben.

11.13 und 11.14 beziehen sich wieder auf die Bedienung, diesmal aus der Perspektive der Einfachheit der Modifikationen (11.13) und Dialogunterbrechungen (11.14).

Zuletzt wurden die Versuchspersonen aufgefordert, zu entscheiden, wie wichtig das Vorhandensein einer Hilfefunktion für einen guten Gesprächsverlauf ist (11.15).

Der Fragebogen wurde so konzipiert, dass jedem Qualitätsfaktor mindestens ein Diagnostik- und ein Rangordnungsstatement zugeordnet wurden. Jedoch konnten Qualitätsfaktoren, die Komponenten ansprechen, welche bei BoRIS nicht implementiert wurden, unmöglich mit Diagnostikstatements operationalisiert werden. Sie konnten nur mit

Hilfe der Rangordnungsstatements bewertet. Diese Faktoren sind: „Bestätigung“, „Dialoginitiative“ (Fragen stellen zu können) und „Hilfefunktion“.

Andere Faktoren, wie „Gestaltung der Benutzereingaben“, „akustische Verständlichkeit“ sowie „Erwartungen“ und „Hintergrundwissen“ sowie die Zufriedenheitsfaktoren, konnten dagegen bei den Diagnostikstatements berücksichtigt werden. Die Wichtigkeit der Faktoren „Gestaltung der Eingaben“, „Erwartungen“ und des „Hintergrundwissens“ für einen guten Gesprächsablauf konnte von den Befragten selbst nicht beurteilt werden; es ist die Aufgabe der Autorin durch eine detaillierte Analyse zu entscheiden, inwieweit sie die Qualität eines Dialogs beeinflussen. Die Bedeutung der akustischen Verständlichkeit bleibt für den Erfolg jeder Interaktion unbestritten. Sie ist neben der inhaltlichen-kognitiven Verständlichkeit die wichtigste Eigenschaft eines Dialogs. Jedoch gehört sie zu dem Sprachausgabemodul und ist somit von der Dialogsteuerung unabhängig. Weil die Zahl der Rangordnungsstatements nicht zu groß sein durfte, um die Befragten nicht übermäßig zu beanspruchen, traf man die Entscheidung die akustische Verstehbarkeit nur bei den Diagnostikstatements zu berücksichtigen.

Die Auswirkungen dieser Faktoren auf die Qualitätsbeurteilung werden im Kapitel 8 ausführlich besprochen.

6.5.2 Auswertung

6.5.2.1 Auswertung des ersten Fragebogens

Der erste Fragebogen beginnt mit der Frage nach einigen persönlichen Daten der Versuchspersonen. Dabei werden die ersten zwei qualitätsbeeinflussenden Faktoren: das Hintergrundwissen und die Erwartungen der Versuchspersonen untersucht.

Die meisten Testpersonen (83%) waren jung, zwischen 20 und 29 Jahren, nur 10 % gehörten zu der Gruppe der 30-39 jährigen. Um ein gewisses Gleichgewicht einzuhalten versuchte man möglichst gleich soviel Frauen wie Männer zu befragen: von insgesamt 53 Testpersonen waren 23 weiblich und 30 männlich.

Tab. 12 Alter der Versuchspersonen

Alter	Weiblich	Männlich	Total
Unter 20	0	1 (3,3%)	1 (1,88%)
Zwischen 20-29	23 (100%)	25 (83%)	48 (90,5%)
Zwischen 30-39	0	3 (10%)	3 (5,6%)
Zwischen 40-49	0	1 (3,3%)	1 (1,88%)
Total	23	30	53

88,6% der Versuchspersonen waren Studenten, nur 11,3% hatten einen Abschluss.

Tab.13 Beruf der Versuchspersonen

Studenten	47 (88,6%)
Absolventen	6 (11,3%) -4 Ingenieure -2 Geisteswissenschaftler

Die meisten kamen aus Nordrhein-Westfalen (58,4%) und 90,5% wohnten in Bochum. Man versuchte auch einen gewissen Anteil von Ausländern zu involvieren: 15,09% der Versuchspersonen waren Ausländer.

Ausschlaggebend für die Restaurantsuche (**1.0**) war für die meisten Testpersonen (94,33%) der Tip von Bekannten. Die Präferenzen variieren dann geschlechtsbedingt: die Frauen bevorzugten den Stadtführer (78,26%) das Internet (34,%), die Magazine (21,73%), die Gelben Seiten. Für die Männer war das Internet zuverlässiger (43,33%) und „Gelben Seiten“ (40%); für die Magazine zeigten, ähnlich wie bei den Frauen, nur 20% von ihnen Interesse.

Tab. 14 Kriterien für Restaurantsuche

FRAUEN		MÄNNER	
1. Tip von Bekannten	91, 3%	1. Tip von Bekannten	96,66%
2.Stadtführer	78,26%	2. Internet	43,33%
3. Internet	34,78%	3. Gelbe Seiten	40%
4. Magazine	21, 73%	4. Stadtführer	36,66%
5. Gelbe Seiten	17,40%	5. Magazine	20%
6. Werbebroschüren	13,04%	6. Werbebroschüren	13,33%

Keine der Versuchspersonen würde die Dienste eines automatischen Sprachinformationssystems in Anspruch nehmen (**3.0**). Das erklärt sich dadurch, dass für die meisten Versuchspersonen solche Systeme relativ fremd waren: nur 17% aller Versuchspersonen hatten ein automatisches Sprachinformationssystem schon benutzt.

Tab. 15 Erfahrungen mit dem System

Kategorie	Frauen	Männer	Total
Nein	20 (87%)	24 (80%)	44 (83%)
Ja	3 (13,04%)	6 (20%)	9 (17%)

Die Mehrheit bezeichnet ihre Erfahrung als dürftig:

Tab. 16 Qualitätsbeurteilung der verwendeten Systeme

<u>Erfahrung</u>	<u>Qualitätsurteil</u>	<u>Geschlecht</u>
<i>1. Filminformation;</i>	<i>Gut</i>	<i>weiblich</i>
<i>2. Bestellung von Kinokarten</i>	<i>schlecht</i>	<i>weiblich</i>
<i>3. Erfahrung mit BoRIS</i>	<i>dürftig</i>	<i>weiblich</i>
<i>4. Handyinformation</i>	<i>dürftig</i>	<i>männlich</i>
<i>5. Telefónica, Service Espania directo</i>	<i>dürftig</i>	<i>männlich</i>
<i>6. Bahnauskunft</i>	<i>dürftig</i>	<i>männlich</i>
<i>7. Bahnauskunft</i>	<i>dürftig</i>	<i>männlich</i>
<i>8. keine Eingabe</i>	<i>dürftig</i>	<i>männlich</i>
<i>9. Telefonbanking</i>	<i>gut</i>	<i>männlich</i>

Die wichtigsten Kriterien, die bei der Entscheidung für die Auswahl eines Restaurants eine Rolle spielen (**2.0**), sind: Qualität des Essens (81,13%), Freundlichkeit des Personals (79,24%) , Preis (75,47%), Art der Küche (73%) und äußeres Ambiente (52,83%).

Tab. 17 Entscheidungskriterien für die Wahl eines Restaurants

Kategorien	Frauen	Männer
2.1 Preis	18 (78,26%)	22 (73,33%)
2.2 Art der Küche	18 (78,26%)	21 (70%)
2.3 Qualität des Essens	17 (74%)	26 (86,66%)
2.4 Breitgefächertes Speiseangebot	2 (8,70%)	4 (13,33%)
2.5 Ort des Restaurants	6 (26,08%)	7 (23,33%)
2.6 Äußeres Ambiente	13 (56,52%)	15 (50%)
2.7 Öffnungszeiten	3 (13,04%)	6 (20%)
2.8 Schnelligkeit der Bedienung	4 (7,4%)	6 (20%)
2.9. Freundlichkeit des Personals	18 (78,26%)	24 (80%)
2.10 Sonstiges: vegetarisch essen	1 (4,35%)	0 (0%)

Erwähnenswert wäre noch, dass die Männer viel mehr Wert auf die Schnelligkeit legen als Frauen. Das beweist z.B. die unterschiedliche Verteilung der Prozentzahlen bei der Kategorie „Schnelligkeit der Bedienung“: 20% der Männer und nur 7,4% von den Frauen kreuzte diese Kategorie an.

Die Frage, welche Auskünfte von einem Sprachauskunftssystem zu erwarten sind (**4.0**), soll zeigen, ob die Datenbank erwartungsgemäß relevante Informationen enthält. Die meist angekreuzten Kategorien waren: Name, Adresse, Telefonnummer (100%), Öffnungszeiten, (100%), Art der Küche (85%), Wegbeschreibung (69,81%), Reservierungsmöglich (54,71%), Preis(49,05%). Es konnten keine bedeutenden Unterschiede zwischen Männer

und Frauen festgestellt werden: ihre Präferenzen unterscheiden sich nur um einige Prozente.

Tab. 18 Informationen, die ein Sprachauskunftssystem liefern sollte

Kategorien	Frauen	Männer	Total
4.1 Name, Adresse, Tel.	23 (100%)	30 (100%)	53 (100%)
4.2 Öffnungszeiten	23 (100%)	30 (100%)	53 (100%)
4.3 Preis	12 (52,17%)	14 (46,66%)	26 (49,05%)
4.4 Art der Küche	22 (95,65%)	23 (76,66%)	45 (85%)
4.5 Spez. Speiseangebote	7 (30,43%)	5 (16,67%)	12 (22,64%)
4.6 Äußeres Ambiente	8 (34,78%)	7 (23,33%)	15 (28,30%)
4.7 Reservierungsmöglich.	12 (52,17%)	17 (56,66%)	29 (54,71%)
4.8. Wegbeschreibung	16 (69,56%)	21 (70%)	37 (69,81%)

Der nächste Abschnitt des Bogens beschäftigt sich mit den Erwartungen der Versuchspersonen (5.0). Die Auswertung der ersten Behauptung bezüglich eines vernünftigen Gesprächs mit BoRIS liefert weit verteilte Daten, was die Interpretation erheblich erschwert. Aus diesem Grund wurden die Werte des positiven („trifft voll zu“ und „trifft zu“) und negativen („trifft nicht zu“ und „trifft gar nicht“) Bereichs kumuliert. Durch die Kumulierung stellte man heraus, dass die Erwartungen in dieser Hinsicht ziemlich niedrig waren: über Hälfte der Versuchspersonen (54,71%) erwarten kein vernünftiges Gespräch mit BoRIS (5.1).

Tab. 19 Erwartungen bezüglich eines vernünftigen Gesprächs

	trifft voll zu	trifft zu	Kumuliert positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ	Total
Frauen	4 (17,4%)	3 (13,04%)	=7 (30,43%)	5 (21,73%)	5 (21,73%)	6 (26,08%)	=11 (47,81%)	23
Männer	3 (10%)	2 (6,66%)	=5 (16,67%)	7 (23,33%)	13 (43,33%)	5 (16,67%)	=18 (60,00%)	30
Total	7 (13,20%)	5 (9,43%)	=12 (22,63%)	12 (22,64%)	18 (34%)	11 (20,75%)	=29 (54,71%)	53

Zu bemerken wäre hier, dass die Werte der Frauengruppe allgemein "optimistischere" Tendenzen aufweisen als die der Männergruppe: die positiven kumulierten Werte liegen hier bei 30,43% im Vergleich zu 16,66% bei den Männer.

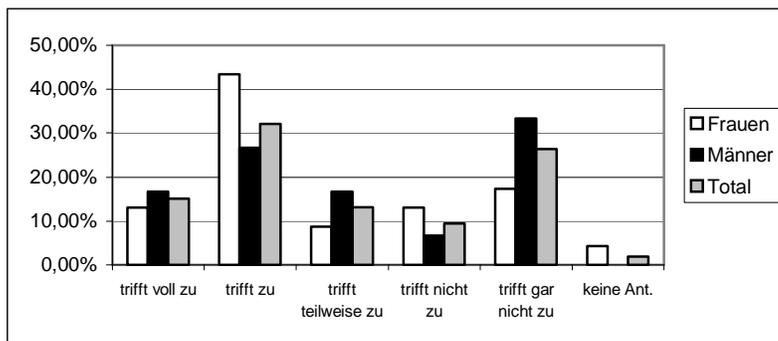
Die Erwartungen konzentrieren sich auf die Vorstellungen der Versuchspersonen über das Frage-Antwortspiel mit dem System. Dazu wurden zwei komplementäre Behauptungen (5.1/5.2): „Sie stellen dem Computer fragen“/„Der Computer stellt Ihnen Fragen“ gestellt. Die Auswertung der ersten Behauptung liefert ziemlich verwirrende Daten, deren Kumulierung auch nicht Klarheit verschafft :

Tab. 20 Sie stellen dem Computer Fragen

	trifft voll zu	trifft zu	Kumuliert positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ	Keine Antwort	Total
Frauen	3 (13,04%)	10 (43,47%)	=13 (56,51%)	2 (8,70%)	3 (13,04%)	4 (17,39%)	=7 (30,43%)	1 (4,35%)	23
Männer	5 (16,67%)	8 (26,66%)	=13 (43,33%)	5 (16,67%)	2 (6,66%)	10 (33,33%)	=12 (40%)	0	30
Total	8 (15,09%)	17 (32,07%)	=25 (47,16%)	7 (13,20%)	5 (9,43%)	14 (26,41%)	=19 (35,84%)	1 (1,88%)	53

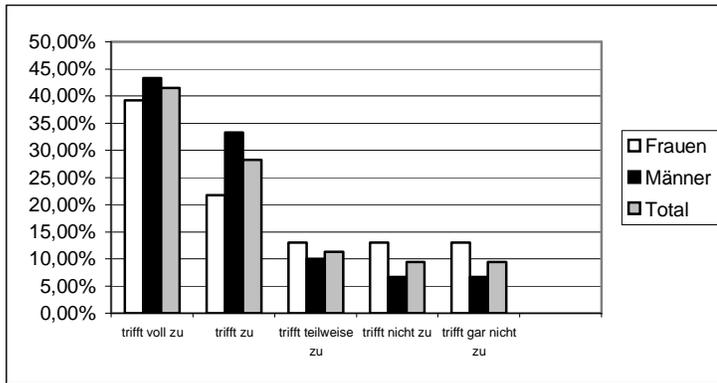
Es lässt sich nur im Falle der Frauengruppe ein eindeutiges Resultat ablesen: der Modus wird bei "trifft zu" mit 43,47% bestimmt und kumuliert ergibt sich eine Mehrheit von 56,51%. Das heißt, dass die Mehrheit der Frauen erwartet, dem Computer Fragen stellen zu können. Die Schwierigkeiten der Interpretation tauchen erst bei der Gruppe der Männer auf. Der Modus wird bei der Kategorie "trifft gar nicht zu" mit 33,33% bestimmt. Der Wert weist aber keine eindeutige Tendenz auf: er wird von dem Gegenwert "trifft zu" mit 26,66% gefolgt; zwischen den beiden gegenseitigen Werten liegt also nur eine Differenz von 6,67%. Bei der Kumulierung der Werte kommt man zu einem ganz umgekehrten Resultat, welches keine eindeutige Tendenz aufweist: die positiven Werte (trifft voll zu + trifft zu) betragen 43,33% im Vergleich zu den negativen Werten (trifft nicht zu + trifft gar nicht zu), die 40% betragen; die Differenz von 3% kann nicht als relevant betrachtet werden.

Abb. 10 Sie stellen dem Computer Fragen



Dagegen stellt sich die Auswertung der zweiten komplementären Behauptung. Die Testpersonen sind mit eindeutigen Werten davon überzeugt (Frauen: 39,13%; Männer: 43,33%; Total: 42,50%), dass das System ihnen Fragen stellen wird (5.3) In diesem Fall ist eine Kumulierung nicht notwendig

Abb.11 Der Computer stellt Ihnen Fragen



Die Auswertung dieser zwei Statements führt zu der Schlussfolgerung, dass die Versuchspersonen sich nicht sicher waren, ob das System ihnen Fragen stellen wird und glaubten eher, dass sie Fragen beantworten sollen.

43,47% der Versuchspersonen waren davon überzeugt, dass sie präzise Antworten bekommen werden (5.4).

Tab. 21 Sie bekommen präzise Antworten.

	trifft voll zu	trifft zu	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Total
Frauen	10 (43,47%)	9 (39,13%)	2 (8,70%)	1 (4,35%)	1 (4,35%)	23
Männer	13 (43,33%)	6 (20%)	3 (10%)	5 (16,67%)	3 (10%)	30
Total	23 (43,40%)	15 (28,30%)	5 (9,43%)	6 (11,32%)	4 (7,55%)	53

Während die Werte der Kategorie "trifft voll zu" bei Männer und Frauen fast identisch sind, weisen die Werte anderer Kategorien große Unterschiede zwischen den männlichen und weiblichen Befragten auf: der Wert von "trifft zu" ist bei den Frauen doppelt so groß als bei den Männern (39,13 % > 20%); dagegen sind die Werte der negativen Kategorien bei Männer viel höher als bei den Frauen: bei „trifft nicht zu“ fast viermal höher (16,67 > 4,35%) und bei „trifft gar nicht zu“ mehr als doppelt (10% > 4,35%). Daraus lässt sich schließen, dass Frauen etwas optimistischere Erwartungen als Männer haben.

Die Auswertung des fünften Statements über die Erwartungen bezüglich der Entstehung möglicher Missverständnisse im Laufe des Gesprächs mit BoRIS (5.5) erwies sich wiederum als schwierig wegen der homogenen Verteilung der Werte auf die einzelnen Kategorien.

Tab 22 Es werden keine Missverständnisse entstehen

	trifft voll zu	trifft zu	Kumuliert positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ	Total
Frauen	5 (21,74%)	3 (13,04%)	=8 (34,78%)	5 (21,74%)	7 (30,43%)	3 (13,4%)	=10 (43,47%)	23
Männer	6 (20%)	8 (26,67%)	=14 (46,67%)	8 (26,67%)	5 (16,67%)	3 (10%)	=8 (26,67%)	30
Total	11 (20,75%)	11 (20,75%)	=22 (41,50%)	13 (24,53%)	12 (22,64%)	6 (11,32%)	=18 (34%)	53

Die kumulierten Werten zeigen, dass die Frauen mit 43,47 % eher Missverständnisse erwartet haben. Im Vergleich dazu tippen die Männern, die mit 46,79% auf „keine Missverständnisse“. Der Gesamtwert (Total) zeigt eine positive Tendenz (41,50% trifft zu-kumuliert), jedoch unterscheidet er sich nur mit 7,50% von seinen negativen Gegenwert (34 % „trifft nicht zu“ -kumuliert).

Die Mehrheit der Versuchspersonen (64,15%-kumuliert) erwartetet, dass sie die gewünschte Information schnell bekommen werden (5.6).

Tab. 23 Sie werden schnell die gewünschte die gewünschte Information bekommen

	trifft voll zu	trifft zu	Kumuliert positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ	Total
Frauen	12 (52,17%)	5 (21,74%)	=17 (74%)	4 (17,4%)	1 (4,35%)	1 (4,35%)	=2 (9,70%)	23
Männer	11 (36,67%)	6 (20%)	=17 (56,67%)	7 (23,33%)	3 (10%)	3 (10%)	=6 (20%)	30
Total	23 (43,40%)	11 (20,75%)	=34 (64,15)	11 (20,75%)	4 (7,55%)	4 (7,55%)	=8 (15,1%)	53

Bemerkenswert ist es, dass die Frauen größere positive Erwartungen als die Männer (vgl. die positiven kumulierten Werte 74% > 56,67%) hatten.

Die Mehrheit der Testpersonen (66,08%-kumuliert) glaubte, dass der Computer eine freundliche Stimme hat (5.7). Der positive kumulierte Wert bei Frauen liegt mit 6,23% über den gleichen Wert bei Männern.

Tab.24 Sie werden schnell die gewünschte Information bekommen

	trifft voll zu	trifft zu	Kumuliert Positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ	Total
Frauen	6 (26,08%)	10 (43,48%)	=16 69,56%	3 (13,04%)	2 (8,70%)	2 (8,70%)	=4 (17,40%)	23
Männer	13 (43,33%)	6 (20%)	=19 63,33%	9 (30,00%)	0 (0,00%)	2 (6,67%)	=2 (6,67%)	30
Total	19 (35,85%)	16 (30,19%)	=35 66,08%	12 (22,64%)	2 (3,77%)	4 (7,55%)	=6 (11,32%)	53

77,36% der Versuchspersonen erwarteten, dass BoRIS leicht zu bedienen ist (**5.8**). Wiederum sind die Prozentzahlen bei den Frauen im positiven Bereich (kumuliert) um 1,6% höher als die der Männer. Es ist selbstverständlich kein wichtiger Unterschied, aber trotzdem in diesem Kontext erwähnenswert, da die Werte der Frauen fast immer positivere Tendenzen zeigen als die der Männer.

Tab. 25 Das System ist leicht zu bedienen

	trifft voll zu	trifft zu	<u>Kumuliert Positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert Positiv</u>	Tot al
Frauen	12 (52,17%)	6 (26,09%)	<u>= 18</u> <u>(78,26%)</u>	3 (13,04%)	2 (8,70%)	0	<u>2</u> <u>(8,70%)</u>	23
Männer	13 (43,33%)	10 (33,33%)	<u>=23</u> <u>(76,66%)</u>	4 (13,33%)	3 (10%)	0	<u>3</u> <u>(10%)</u>	30
Total	25 (47,17%)	16 (30,19%)	<u>=41</u> <u>(77,36%)</u>	7 (13,21%)	5 (9,43%)	0	<u>5</u> <u>(9,43%)</u>	53

Frauen unterscheiden sich deutlich von Männern, wenn es um die Art des Gesprächs geht (**6.0**): 39,13% der Frauen möchten lieber Fragen beantworten als Fragen stellen (8,70%). Am liebsten würden die Frauen ein gemischtes Gespräch mit BoRIS führen: der Modus liegt bei Frauen bei der Kategorie „Gemischt“, mit 43,48%.

Dagegen möchten 40% der Männer lieber die Initiative ergreifen und dem System Fragen stellen oder ein gemischtes Gespräch führen, 33,33%. Nur 8,70% der Männer bevorzugen BoRIS Fragen zu stellen.

Tab. 26 Art des Gesprächs

Art des Gespräch	Frauen	Männer	Total
Sie stellen Fragen an das System	2 (8,70%)	12 (40%)	14 (26,42%)
Sie beantworten Fragen	9 (39,13%)	6 (20%)	15 (28,30%)
Gemischt	10 (43,48%)	10 (33,33%)	20 (37,74%)
Egal	2 (8,70)	2 (6,67%)	4 (7,55%)

Über 60% aller Versuchspersonen sind der Meinung, dass Sprachinformationssysteme teilweise eine Alternative zu den derzeitigen Informationssystemen darstellen (**7.0**). Es wäre an dieser Stelle interessant anzumerken, dass 83% aller Versuchspersonen noch nie ein solches System benutzt haben. Aus dieser Perspektive ist diese Antwort als „positiv“ zu interpretieren. Vor allem die Männergruppe neigte, im Vergleich zur Frauengruppe dazu, Sprachauskunftssysteme als Alternative zu sehen.

Tab.27 Sprachauskunftssysteme als Alternative

Kategorie	Frauen	Männer	Total
Ja	4 (17,4%)	8 (26,67%)	12 (22,64%)
Teilweise	14 (60,87%)	19 (63,33%)	33 (62,26%)
Nein	1 (4,35%)	0 (0%)	1 (1,89%)
Ich weiß es nicht	4 (17,39%)	3 (10%)	7 (13,21%)

6.5.2.2. Auswertung des zweiten Fragebogens

6.5.2.2.1 Qualitätsbeurteilung des Gesamtsystems

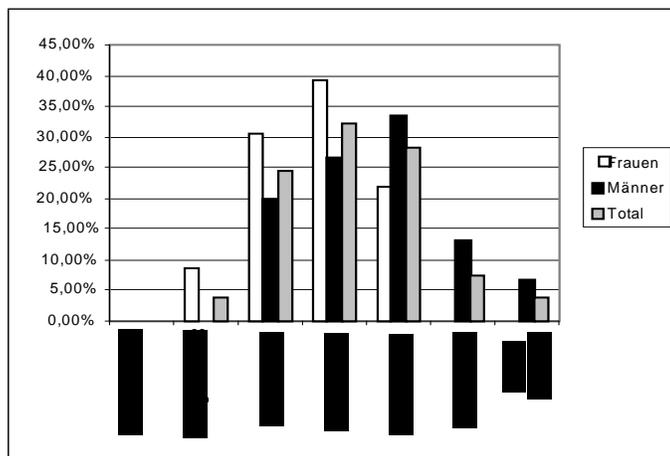
Nach den Versuchen bekamen die Befragten den zweiten Fragebogen. Ihre erste Aufgabe bestand darin, ihr Qualitätsurteil über das System auf der Bodden-Jekosch Skala zu platzieren (8.1).

Bei den Frauen fiel das Qualitätsurteil positiver als bei den Männern aus: der Modus bei den Frauen liegt mit 39,13% bei „ordentlich“; bei den Männern beträgt das Modus 33,33% bei der Kategorie „dürftig“. 30,43% der Frauen beurteilten das System als „gut“ und 8,7% sogar als „ausgezeichnet“; dagegen haben nur 20% der Männer BoRIS „gut“ gefunden. Keiner der Männer kreuzte die Kategorie „ausgezeichnet“ an, aber dafür gab es insgesamt 6 Nennungen für „schlecht“ (13,33%) und für „extrem schlecht“ (6,67%).

Tab. 28 Qualitätsbeurteilung des System

Kategorie	Frauen	Männer	Total
Ideal	0 0%	0 0%	0 0%
Ausgezeichnet	2 8,70%	0 0%	2 3,77%
Gut	7 30,43%	6 20,00%	13 24,53%
Ordentlich	9 39,13%	8 26,67%	17 32,08%
Dürftig	5 21,74%	10 33,33%	15 28,30%
Schlecht	0 0%	4 13,33%	4 7,55%
Extrem schlecht	0 0%	2 6,67%	2 3,77%

Abb. 12 Qualitätsbeurteilung des Systems



Danach wurden die Versuchspersonen indirekt aufgefordert ihr Qualitätsurteil zu begründen, indem sie einem Freund das Systems beschreiben sollten (8.2). Dadurch bekamen sie die Möglichkeit, sich offen über ihre Erfahrung mit dem System zu äußern. Es kamen eine Anzahl von ganz schlichten und neutralen bis komplizierten und kritischen Systembeschreibungen heraus. Das System wurden gleichzeitig kritisiert und gelobt, es wurden „Problemstellen“ angesprochen und Ratschläge gegeben, wie man am besten mit BoRIS umgeht und was dabei vermieden werden soll.

Zusammenfassend wurden folgende Kritikpunkte genannt, die sich im größten Teil mit denen überschneiden, die sich aus der qualitativen Befragung ergaben (siehe Abschnitt 6.3.3).

Kritikpunkte

1. unverständliche und zu schnelle Sprachausgabe.
2. umständliches, langwieriges Frage-Antwortspiel.
3. keine Wiederholmöglichkeit.
4. keine Nachfragemöglichkeit: das System beendet das Gespräch zu früh; für spezielle
5. Wünsche sei das System nicht geeignet.
6. keine Möglichkeit, konkret eigene Fragen zu stellen.
7. die Fragenstellung war nicht präzise genug.
8. es ist unklar, wie man die „richtigen“ Fragen stellen soll.
9. die Pausen nach dem Input waren zu lang.
10. sehr begrenzte Möglichkeiten der Modifizierung.
11. keine Feedbackmöglichkeit: es werden keinerlei Hinweise gegeben, warum das System das passende Restaurant nicht gefunden hat.
12. ist verwirrend: das System gibt wenige Antwortmöglichkeiten an, reagiert aber auf andere Antworten, die es vorher nicht nennt.
13. es werden keine Angaben über sonstige Eigenschaften des Restaurants gemacht.
14. am Anfang wurden zu viele Kriterien der Informationssuche genannt.
15. man hat das Gefühl, dass man mit einer Wand spricht, dass man nicht verstanden wird.
16. irgendwie fehlt der persönliche Charakter eines Gesprächs.
17. Das System hat eine kalte Roboterstimme.

BoRIS wurde unter anderem als ein Informationssystem beschrieben, „das überhaupt nicht funktioniert“ (VP 6), „total unsympathisch“ (VP 49), wobei es einem „keinen Spaß macht, da anzurufen“ (VP 25).

Andererseits wurden auch positive Eigenschaften des Systems genannt:

Positive Eigenschaften

1. flexible Wahl.
2. gute Spracherkennung.
3. einleuchtende Benutzerführung.
4. das System sei unkompliziert und relativ schnell.
5. interessant auch unter wirtschaftlichen Aspekten.
6. präzise Systemantworten.
7. nette, freundliche Bandstimme.
8. einfache Bedienung.
9. gute Auswahl an Restaurants.
10. man kann seine Wünsche verändern, d.h. die Wahl modifizieren.

Das System wurde auch als „hilfreicher“ (VP 37), „intelligenter Sprachcomputer“ (VP 52) mit „recht hohem Verständnis“ (VP 44) bezeichnet, ein „cooles System“, das man als „Alternative zu Telefonbüchern“ benutzen kann (VP 47).

Die nächste Frage, „Was würden Sie bei BoRIS verbessern?“ (8.3) bringt eine Vertiefung der obengenannten Kritikpunkte. Dabei wird die Phantasie der Versuchspersonen stimuliert und gleichzeitig werden Hinweise zur Systemverbesserung geliefert, an die man vorher vielleicht nicht gedacht hat.

Verbesserungsvorschläge

Das erste, was die Testpersonen verbessern würden, wäre selbstverständlich die Sprachausgabe. Dazu wurde den Wunsch geäußert, den Benutzern eine Alternative zur Sprachsynthese zu bieten, z.B. die Information schriftlich zu liefern.

Die Modifikationen sollten in kleinen Schritten erfolgen: das System sollte eventuell einzelne Fragen stellen, wie z.B.: „Möchte Sie die Art der Küche modifizieren?“. Wenn man seine Eingaben zwei-dreimal geändert hat, weiß man nicht mehr genau, was eigentlich modifiziert wird; es wäre sinnvoll, den Benutzer auf seine letzte Änderung hinweisen, oder ihnen die Möglichkeit geben, über eine Resetfunktion die Frage neu zu stellen. Das sollte allerdings auch in dem Falle gelten, in dem man eine falsche Eingabe gemacht hat.

Am Ende des Gesprächs sollte das System die Adressenansage wiederholen können. Der Wunsch nach einer Wiederholung sollte verbal oder z.B. durch Bestätigung der Sterntaste signalisiert werden.

Manche Versuchspersonen beschwerten sich, dass BoRIS kein „Nachfragen“ erlaubt und schlugen vor, dass man während eines Telefonats mehrere Anfragen macht.

Die Auswahlkriterien sollten erweitert werden: es fehlten Auskünfte über Telefonnummer und Wegbeschreibung. Die Fragestellung hinsichtlich des Ortes und der Küche sollte man

ausweiten; für einige Versuchspersonen war es nicht eindeutig, ob man auch andere Varianten zur Auswahl hatte (z.B. bei der Art der Küche konnte man nur zwischen drei Typen auswählen).

Für manche Versuchspersonen war es nicht klar, ob das System die Eingabe aller Kategorien in einem Satz erlaubte. Aus diesem Grund sollte man im Startmenü die Benutzer entsprechend aufklären. Ein Vorschlag einer möglichen Ansage kam von einem männlichen Versuchsperson (VP 25): „Hallo, hier Ihr Berater für Ihre Wünsche. Nach welchen Kriterien soll ich für Sie suchen?“ oder allgemeiner formuliert: „Nach welchen Kriterien möchten Sie suchen?“.

Es sollte eine bessere Beschreibung geben, was der Anrufer machen muss. Sollte das System „kein passendes Angebot“ zu seiner Anfrage finden, wäre es angebracht dem Benutzer die Gründe dafür zu nennen.

Die Zwischenpausen sollten gekürzt werden. Die langen Pausen verwirren die Benutzer, die einfach nicht wissen, ob das System ihre Eingaben verstanden hat. Eventuell sollte man eine Ansage so wie „Ihre Eingaben werden bearbeitet“ zwischenschalten. Dagegen sollten die Pausen zwischen den verschiedenen Restaurantadressen vergrößert werden.

Das System sollte die Eingaben der Benutzer bestätigen und das Anfangsmenü kürzen. Die Datenbasis könnte vergrößert werden und die Menüführung sollte auch spezielle Wünsche berücksichtigen, wie z.B. die Adressen nach Restaurantnamen suchen oder die Möglichkeit die Ansagen für Ausländer auf Englisch zu halten.

Ein männlicher Teilnehmer (VP 25) machte noch einen phantasievollen Vorschlag: die Stimme von BoRIS sollte wählbar sein, z.B. männlich oder weiblich; im Hintergrund sollte eine angenehme Musik gespielt werden. Dazu sollte das System die Benutzer noch mit ihrem Namen ansprechen, um eine persönlichere Atmosphäre schaffen.

6.5.2.2 Diagnostikstatements

Der Fragebogen konzentriert sich weiter auf die Evaluierung der Qualitäts- und Zufriedenheitsfaktoren der Dialogsteuerung. Für eine übersichtliche Darstellung wird die Reihenfolge der Statements aus dem Fragebogen nicht eingehalten. Die Auswertung wird sich nach dem Untersuchungsschema der Kooperativität richten, welches im Abschnitt 4.2 dargestellt wurde.

I. Frage-Antwort Aspekte

1. Relevanz

1.1. Informationsmenge

Die Behauptung (9.2), dass das System umfangreiche Informationen über Bochumer Restaurants bietet, liefert sehr weit verteilte Werte, die sich schwer interpretieren lassen

Tab. 29 "BoRIS bietet umfangreiche Information über Bochumer Restaurants"

	trifft voll zu	trifft zu	<u>Kumuliert</u> <u>Positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert</u> <u>negativ</u>
Frauen	3 (13,04%)	6 (26,09%)	<u>≙9</u> <u>39,13%</u>	6 (26,09%)	5 (21,74%)	3 (13,04%)	<u>≙8</u> <u>(34,78%)</u>
Männer	6 (20%)	6 (20%)	<u>≙12</u> <u>(40%)</u>	6 (20%)	9 (30%)	3 (10%)	<u>≙12</u> <u>(40%)</u>
Total	9 (17%)	12 (22,64%)	<u>≙21</u> <u>(39,62%)</u>	12 (22,64%)	14 (26,42%)	6 (11,32%)	<u>≙20</u> <u>(37,74%)</u>

In diesem Fall kann der Modus bei der Gruppe der Frauen nicht eindeutig bestimmt werden, weil dass zwei Merkmalsausprägungen mit gleicher Häufigkeitsverteilung (26,09% bei „trifft zu und „trifft teilweise zu“) an der Spitze liegen. Die Kumulierung kann auch nicht die Interpretation der Daten vereinfachen: bei den Männer zeigen die kumulierten Werte die gleichen Prozentzahlen im positiven wie im negativen Bereich (40%). Die Totale zeigt durch die Kumulierung ihren Höchstwert (39,13%) im positiven Bereich; jedoch unterscheidet sich dieser Wert nur um 1,39% vom negativen Bereich „trifft nicht zu“ kumuliert (37,74%). Aus diesem Grund liefert die Auswertung dieser Behauptung kein eindeutiges Ergebnis.

1.2 Vollständigkeit

Dagegen liefert die Behauptung „BoRIS bietet keine detaillierte Informationen über die gewünschten Restaurants“ (10.4) eine eindeutige Verteilung: 74% die Frauen und 56,66 % wählten die positiven Kategorien („trifft zu“ und „trifft voll zu“). Die Totale hat einen kumulierten positiven Wert von 64,15%. Man kann daraus schließen, dass die Mehrheit der Befragten mit dem Informationsgehalt von BoRIS nicht zufrieden waren.

Tab. 30 "BoRIS bietet keine detaillierte Informationen über die gewünschten Restaurants"

	trifft voll zu	trifft zu	<u>Kumuliert</u> <u>positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert</u> <u>negativ</u>
Frauen	12 (52,17%)	5 (21,74%)	<u>≙17</u> <u>(74%)</u>	2 (8,70%)	1 (4,35%)	3 (13,04%)	<u>≙4</u> <u>(30,80%)</u>
Männer	7 (23,33%)	10 (33,33%)	<u>≙17</u> <u>56,66%</u>	6 (20%)	5 (16,67%)	2 (6,67%)	<u>≙7</u> <u>(23,34%)</u>
Total	19 (35,85%)	15 (28,30%)	<u>24</u> <u>64,15%</u>	8 (15,09%)	6 (11,32%)	5 (9,43%)	<u>≙11</u> <u>20,75%</u>

2. Ausdrucksform

2.1 Akustische Verständlichkeit

95,42% der Frauen und 90% der Männer (die Werte gehören der Kategorie „trifft voll zu“) beschwert sich über die undeutliche Ansage der Restaurantnamen (10.8).

2.2 Inhaltliche Verständlichkeit

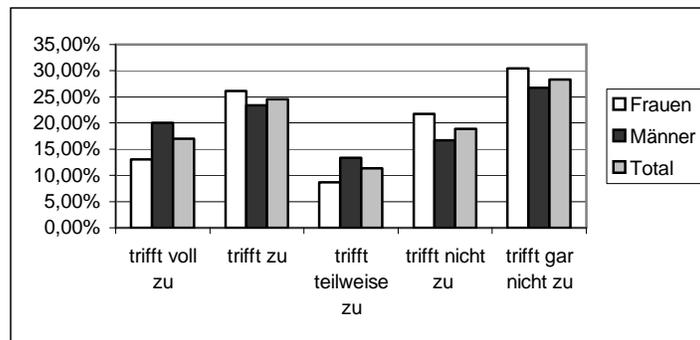
Einer der oft erwähnten Kritikpunkte war, dass das System am Anfang des Gesprächs zu viele Auswahlkriterien auflistete. Man hat versucht, diesen Kritikpunkt auszuwerten, indem man die Befragten aufforderte, sich zum Statement „Ich konnte mir die Auswahlkriterien am Anfang des Gesprächs nicht merken“ (10.6) zu äußern. Dabei hat man herausgestellt, dass, statistisch gesehen, eine solche Maßnahme nicht mit absoluter Mehrheit begrüßt worden wäre:

Tab. 31 "Ich konnte mir die Auswahlkriterien am Anfang des Gesprächs nicht merken"

	trifft voll zu	trifft zu	Kumuliert Positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ
Frauen	3 (13,04%)	6 (26,09%)	=9 (39,13%)	2 (8,70%)	5 (21,74%)	7 (30,43%)	=12 (52,17%)
Männer	6 (20%)	7 (23,33%)	=13 (43,33%)	4 (13,33%)	5 (16,67%)	8 (26,67%)	=13 (43,33%)
Total	9 (17%)	13 (24,53%)	=22 (41,53%)	6 (11,32%)	10 (18,87%)	15 (28,30%)	=25 (47,17%)

Der Modus lässt sich bei jeder Gruppe aus dem Diagramm bei der Kategorie „trifft gar nicht zu“ ablesen..

Abb. 12 "Ich konnte mir die Auswahlkriterien am Anfang des Gesprächs nicht merken"



Die kumulierten Werte bei der Frauengruppe liegen im negativen Bereich bei 52,17%. Bei den Männern bringt die Kumulierung kein eindeutiges Ergebnis: die gleichen Prozentzahlen (43,33%) liegen im positiven, wie im negativen Bereich. Der Höchstwert der Totalen liegt im negativen Bereich bei 47,17%, jedoch unterscheidet es sich vom positiven Bereich nur um 5,66%. Es ist in diesem Fall schwer, einen eindeutigen Trend festzustellen; man kann erschließen, dass, obwohl viele Befragten die Ansagen im Anfangsmenü zu lang fanden, sie von den meisten als nicht besonders lästig empfunden wurden.

2.3 Gestaltung der Benutzereingaben

Mehr als die Hälfte der Frauen hat, laut eigener Angaben die Fragen von BoRIS sowohl stichwortartig als auch satzwortartig beantwortet (**9.11**). Die Männer (50%) dagegen haben ihre Antworten nur stichwortartig konzipiert..

Tab. 32 Eingabeformulierung

	Frauen	Männer	Total
stichwortartig	7 (30,43%)	15 (50%)	22 (41,51%)
satzartig	3 (13,04%)	3 (10%)	6 (11,32%)
teils-teils	13 (56,52%)	11 (36,67%)	24 (45,28%)

3. Korrektur und Aufklärungsstrategien

3.1 Wiederholung

3.1.1 Wiederholung der Systemeingaben

Die Mehrheit der Testpersonen (75,47%) wurden dadurch unangenehm überrascht, dass BoRIS seine Eingaben nicht wiederholen konnte (**10.3**): 87% der Frauen und 66,67% der Männer kreuzten die Kategorie „trifft voll zu“ an. Bemerkenswert ist die Differenz von fast 20% zwischen Männern und Frauen. Dies ist ein Zeichen dafür, dass die Männer weniger von diesem Problem betroffen wurden als die Frauen.

3.1.2 Wiederholung der Benutzereingaben

Nur ein Viertel (26,41% kumuliert) der Versuchspersonen wurden im Laufe des Gesprächs mit BoRIS dadurch gestört, dass sie immer wieder ihre Eingaben wiederholen mussten (**10.5**). 56,52% der Frauen und 63,33 der Männer beurteilten die Aussage als nicht zutreffend.

Tab. 33 "Ich musste oft meine Eingaben wiederholen"

	trifft voll zu	trifft zu	Kumuliert Positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ
Frauen	2 (8,70%)	4 (17,40%)	=6 (26,10%)	4 (17,40%)	6 (26,09%)	7 (30,43%)	=13 (56,52%)
Männer	1 (3,33%)	7 (23,33%)	=8 (26,66%)	3 (10%)	10 (33,33%)	9 (30%)	=19 (63,33%)
Total	3 (5,66%)	11 (20,75%)	=14 (26,41%)	7 (13,21%)	16 (30,19%)	16 (30,19%)	=36 (60,38%)

3.2 Bestätigung

Das System verfügt über keine Bestätigungsmöglichkeit. Aus diesem Grund wird dieser Qualitätsfaktor erst bei den Rangordnungsstatements berücksichtigt.

II. Systemmöglichkeiten

4. Berücksichtigung kontextueller Bedingungen

4.1 Erwartungen

Die Befragten sollten entscheiden, inwieweit ihre Erwartungen an BoRIS erfüllt wurden (9.1). Die kumulierten Höchstwerte liegen im negativen Bereich (Frauen: 47,82%, Männer: 53,33%, Total 51%), was ein Zeichen dafür ist, dass die meisten Versuchspersonen ein Erwartungsbild hatten, welches der realen Gesprächssituation nicht entsprach (die Gründe dafür werden im nächsten Kapitel ausführlich kommentiert).

Tab. 34 "BoRIS hat genau so funktioniert, wie ich es mir vorgestellt habe"

	trifft voll zu	trifft zu	<u>Kumuliert positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert negativ</u>
Frauen	3 (13,04%)	5 (21,74%)	<u>=8</u> 34,78%	4 17,40%	8 34,78%	3 13,04%	<u>=11</u> 47,82%
Männer	1 (3,33%)	11 (36,67%)	<u>=12</u> 40%	2 6,67%	10 33,33%	6 20%	<u>=16</u> 53,33%
Total	4 (7,55%)	16 (30,19%)	<u>=20</u> 37,74%	6 11,32%	18 34%	9 17%	<u>=17</u> 54%

4.2 Hintergrundwissens

Dieser Faktor wurde im ersten Fragebogen bereits untersucht (siehe Abschnitt 6.5.1, Frage 3.0).

5. Transparenz

5.1 Einfachheit der Bedienung

Die Mehrheit der Befragten erwarteten vor dem Versuch eine leichte Bedienung des Systems. Ob diese Behauptung tatsächlich stimmt, soll aufgrund des Statements „Jeder kann mit BoRIS ohne Probleme umgehen“ (9.9) festgestellt werden. Die Auswertung zeigt, dass nur im Falle der Frauen diese Erwartung erfüllt wurde: durch Kumulierung erreicht man bei der Frauengruppe einen Wert von 55,17% im positiven Bereich. Die Männergruppe vertritt mit einem kumulierten Wert von 46,67% die gegengesetzte Meinung. Bemerkenswert ist dabei, dass, aufgrund der homogenen Verteilung der Prozentzahlen, die Auswertung ohne Kumulierung nicht denkbar ist.

Tab.35 "Jeder kann mit BoRIS ohne Probleme umgehen"

	trifft voll zu	trifft zu	<u>Kumuliert Positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert negativ</u>
Frauen	6 (26,09%)	6 (26,09%)	<u>=12</u> (52,18%)	3 (13,04%)	5 (21,74%)	3 (13,04%)	<u>=8</u> (34,78%)
Männer	2 (6,67%)	7 (23,33%)	<u>=9</u> (30%)	7 (23,33%)	7 (23,33%)	7 (23,33%)	<u>=14</u> (46,67%)
Total	8 (15,09%)	13 (24,53%)	<u>=21</u> (39,62%)	10 (18,87%)	12 (22,64%)	10 (18,87%)	<u>=22</u> (41,51%)

5.2 Systemgrenzen

Für die meisten Versuchspersonen (62,27% kumuliert), vor allem für die Männer (70% kumuliert) war es nicht klar, wie sie ihre Eingaben modifizieren sollen, um schnell die gewünschte Information zu bekommen (**10.10**). Die kumulierte Werte zeigen, dass mehr Männer als Frauen Probleme an dieser Stelle hatten. (70% > 52,18%).

Tab. 35 "Es war mit nicht klar, wie ich meine Eingaben modifizieren soll [...]"

	trifft voll zu	trifft zu	<u>Kumuliert</u> <u>Positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert</u> <u>negativ</u>
Frauen	8 (34,78%)	4 (17,40%)	<u>=12</u> <u>(52,18%)</u>	5 (21,74%)	2 8,70%	4 (17,40%)	<u>=6</u> <u>(26,1%)</u>
Männer	12 (40%)	9 (30%)	<u>=21</u> <u>(70%)</u>	2 (6,67%)	5 (16,67)	2 (6,67%)	<u>=7</u> <u>(23,34%)</u>
Total	20 (37,74%)	13 (24,53%)	<u>=33</u> <u>(62,27%)</u>	7 (13,21%)	7 (13,21%)	6 (11,32%)	<u>=13</u> <u>(24,53%)</u>

6. Dialoginitiative

6.1 Fragen stellen

Die Versuchspersonen haben leider keine Möglichkeit dem System Fragen zu stellen. Sie können nur bedingt die Initiative ergreifen (indem sie z.B. das System auffordern seine Angaben zu wiederholen). Aus diesem Grund konnte dieser Qualitätsfaktor nicht diagnostiziert werden. Es wird erst bei den Rangordnungsstatements berücksichtigt.

6.2. Gespräch führen solange man möchte

BoRIS erlaubt am Ende eines Dialogs keine neue Anfrage zu starten und keine weitere Fragen zu stellen. Aus diesem Grund wird dieser Qualitätsfaktor erst bei den Rangordnungsstatements berücksichtigt.

7. Schnelligkeit

7. Schnelligkeit der Datenverarbeitung

Im Laufe der Versuche hat man festgestellt, dass Frauen und Männer unterschiedliche kommunikative Verhaltensweisen haben (siehe Abschnitt 5.3.3.1), die Auswirkungen auf den Dialogerfolg hatten. Die Tatsache, dass Frauen allgemein schneller ans Ziel kamen als Männer beweist die Auswertung der nächsten Statements: "Das Gespräch mit BoRIS hat schnell zur gewünschten Information geführt" (**9.7**): der Modus der Frauengruppe liegt mit 34,78% bei „trifft zu“; bei den Männer dagegen liegt der Modus mit 36,67% bei „trifft nicht zu“. Die Kumulierung der Werte zeigt die gleiche Tendenz: die Mehrheit der Frauen (65,21%) bekam schneller die gewünschte Antwort im Vergleich zur Männergruppe, die mit 40% die negativen Kategorien angekreuzt haben. Die kumulierten Werte der Totale zeigen mit 45,28% eine Präferenz für den positiven Bereich:

Tab. 36 "Das Gespräch mit BoRIS hat schnell zur gewünschten Information geführt"

	trifft voll zu	trifft zu	Kumuliert positiv	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ
Frauen	7 (30,43%)	8 (34,78%)	=15 (65,21%)	3 (13,04%)	4 (17,40%)	1 (4,35%)	=5 21,75%
Männer	4 (13,33%)	5 (16,67%)	=9 (30%)	9 (30%)	11 (36,67%)	1 (3,33%)	=12 (40%)
Total	11 (20,75%)	13 (24,53%)	=24 (45,28%)	12 (22,64%)	15 (28,30%)	2 (3,77%)	=17 (32,07%)

7.2 Dialogunterbrechungen

Obwohl BoRIS für lange Pausen zwischen Fragen und Antworten kritisiert wurde, sind diejenigen, die sich darüber beschwert haben, statistisch in der Minderheit: die kumulierten Werte zeigen, dass 65,21% der Frauen und 53,34% der Männer von diesen angeblich langen Pausen nicht gestört wurden (**10.7**). Bemerkenswert wäre hier noch, dass innerhalb der Gruppe, die doch zu dieser Behauptung tendierten, 10% mehr Männer (36,67% kumuliert) als Frauen (26,09% kumuliert) waren.

Plötzliche Unterbrechungen im Gesprächsverlauf (**10.1**) haben die meisten Versuchspersonen gar nicht gestört: der Modus der Frauengruppe (56,52%) und der Männergruppe (40%) liegt bei der Kategorie „trifft gar nicht zu“. Durch Kumulation werden eindeutige Werte erreicht: 78% der Frauen und 60% der Männer kreuzten die negativen Kategorien „trifft nicht zu“ und „trifft gar nicht zu“ an (wiederum besteht zwischen Männern und Frauen eine Differenz von 18%).

Allgemein wurde das Gespräch nicht mittendrin abgebrochen (**10.2**). Dies entschied über die Hälfte der Testpersonen (54,72%), indem sie die Kategorie „trifft gar nicht zu“ ankreuzten. Eine interessante Auskunft über die unterschiedlichen Erfahrungen, die Männer und Frauen mit dem System machten, zeigen die Werte der positiven Kategorien „trifft zu“ und „trifft voll zu“: während keine einzige Frau diese Kategorien wählte, gaben 33,33% (kumuliert) der Männer an, von ständigen Unterbrechungen während des Gesprächs mit BoRIS gestört worden zu sein.

8. Hilfefunktion

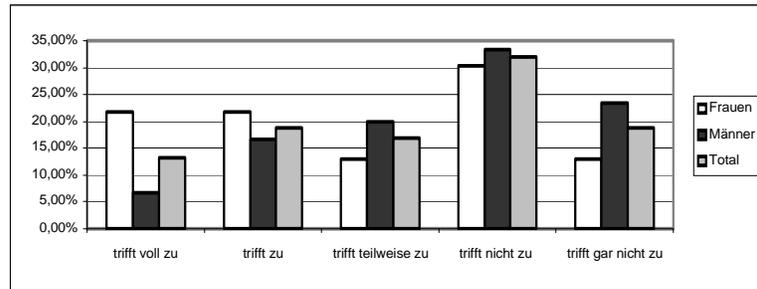
Das System verfügt über keine Hilfefunktion. Aus diesem Grund wird dieser Qualitätsfaktor erst bei den Rangordnungsstatements berücksichtigt.

III. Zufriedenheit

9. Einschätzung der kommunikativen Zielerreichung

Die Auswertung der Behauptung „Ich bekam immer die gewünschte Antwort auf meine Eingaben“ (**9.3**) erwies sich als problematisch. Aus dem Diagramm wird deutlich, dass der Modus aller drei Gruppen (Frauen: 30,43% , Männer: 33,33%, Total: 32,08%) bei „trifft nicht zu“ liegt:

Abb. 13 "Ich bekam immer die gewünschte Antwort auf meine Eingaben"



Durch Kumulierung kommt es im Falle der Frauen zu einer gleichmäßigen Verteilung der Werte (43,47%) im positiven und negativen Bereich:

Tab. 37 "Ich bekam immer die gewünschte Antwort aus meine Eingaben"

	trifft voll zu	trifft zu	<u>Kumuliert positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert negativ</u>
Frauen	5 (21,74%)	5 (21,74%)	<u>=10</u> <u>(43,47%)</u>	3 (13,04%)	7 (30,43%)	3 (13,04%)	<u>=10</u> <u>(43,47%)</u>
Männer	2 (6,67%)	5 (16,67%)	<u>=7</u> <u>(23,34%)</u>	6 (20%)	10 (33,33%)	7 (23,33%)	<u>(17)</u> <u>56,66%</u>
Total	7 (13,21)	10 (18,27)	<u>=17</u> <u>(31,48%)</u>	9 (17%)	17 (32,08%)	10 (18,87%)	<u>(27)</u> <u>(51%)</u>

Daraus kann man schließen, dass Frauen häufiger die gewünschte Antwort auf ihre Frage bekamen als Männer. Das wird durch die Auswertung der Behauptung „Der Anruf bei BoRIS hat sich gelohnt“ (9.8) bestätigt: bei der Frauengruppe liegt der Modus mit 30,43% bei „trifft zu“ und die kumulierten Werte zeigen 47,82% im positiven Bereich. Für Männer scheint, sich der Anruf bei BoRIS weniger gelohnt zu haben: der Modus liegt bei „trifft teilweise zu“ mit 30% und durch die Kumulierung werden 36,67% im positiven Bereich erreicht. Der positive Bereich unterscheidet sich nur mit knapp 3,34% vom negativen Bereich. **Tab. 38 "Der Anruf bei BoRIS hat sich gelohnt."**

	trifft voll zu	trifft zu	<u>Kumuliert positiv</u>	trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	<u>Kumuliert negativ</u>
Frauen	4 (17,40%)	7 (30,43%)	<u>=11</u> <u>(47,83%)</u>	6 (26,09%)	5 (21,74%)	1 (4,35)	<u>1</u> <u>(26,09%)</u>
Männer	5 (16,67%)	6 (20%)	<u>=11</u> <u>(36,67%)</u>	9 (30%)	7 (23,33%)	3 (10%)	<u>=10</u> <u>(33,33%)</u>
Total	9 (17%)	13 (24,53%)	<u>=22</u> <u>(41,53%)</u>	15 (28,30)	12 (22,64%)	4 (7,55%)	<u>=16</u> <u>(30,19%)</u>

10. Einschätzung des Computers als Gesprächspartner

10.1 Ähnlichkeit des Systems mit einem menschlichen Gesprächspartner

Menschen bevorzugen Menschen als Gesprächspartner. So lässt sich die Auswertung der Behauptung „Ich spreche lieber mit einem Menschen als mit BoRIS“ (9.4) interpretieren. Die Kumulierung zeigt, dass 68% der Befragten (52,18% der Frauen und 80% der Männer) der Behauptung zustimmten. Das Interessante hier ist nicht das Ergebnis an sich, das durchaus zu erwarten war, sondern der Unterschied von 27,82%, der sich zwischen Männern und Frauen im positiven Bereich ergibt und die angegebenen Werte im negativen Bereich: 21,75% der Frauen und nur 10% der Männer kreuzten die Kategorie „trifft nicht zu“ und „trifft gar nicht zu“ an. Das bedeutet, dass Frauen eine bessere Einstellung zum System als Gesprächspartner hatten als Männer.

Ähnliche Ergebnisse liefert die Auswertung des nächsten Statements: „Ich konnte mich äußern, wie in einem normalen Gespräch mit einem Menschen“ (9.5). Der Modus liegt bei „trifft nicht zu“, sowohl bei Frauen (39,13%) als auch bei Männern (36,67%). Durch die Kumulierung werden eindeutige Werte im negativen Bereich erreicht: 60,87% bei Frauen und 46,67% bei Männern. Erstaunlich ist, dass, obwohl die Frauen spontanere kommunikative Reaktionen hatten, sie das Gespräch mit BoRIS als unnatürlicher als die Männer empfanden. Dies wird durch die positiven kumulierten Werte bestätigt: 17,4% der Frauen kreuzten „trifft zu“ und „trifft voll zu“ im Vergleich zu 26,67% der Männer an.

10.2 Vertrauenswürdigkeit

BoRIS wirkt auf die meisten Befragten vertrauenswürdig: die kumulierten Werte zeigen, dass 68% aller Befragten davon überzeugt sind, dass BoRIS ihnen richtige Informationen geliefert hat (9.6). Dabei war das Urteil der Frauen mit 74% (kumuliert) auffällig positiver als bei den Männern (63,33%) kumuliert

10.2 Stimmung während des Gesprächs mit BoRIS

Frauen fühlten sich wohler (57,52% kumulierter), sicherer (56,52% kumulierter) und ruhiger (74% kumulierter Wert) als Männer (9.10). Die Höchstwerte der Männergruppe zeigen, dass Männer sich nur teilweise „wohl“ (40%) und „sicher“ (53%) fühlten. Bei der Kategorie „ruhig“ kreuzten jeweils 36,67% der Männer den neutralen Bereich („teilweise“) und den positiven Bereich „trifft zu“ (kumuliert) an.

Zufrieden mit BoRIS während des Gesprächs fühlten sich nur 22,64% (kumuliert) aller Befragten. 46,67% (kumuliert) der Männer und 39,13% Frauen (kumuliert) erklärten, sie fühlten sich dabei unzufrieden. Bei den Frauen können die Werte nicht eindeutig interpretiert werden, da die Bereiche „teilweise“ und „trifft zu“ (kumuliert) den gleichen Wert (39,13%) aufweisen.

11. Einschätzung des Gesprächsaufbaus

Für 47,17% (kumuliert) aller Testpersonen war der Weg bis zur gewünschten Information nicht umständlich (**10.9**). Bei den Frauen haben die Kategorien „trifft teilweise zu“ und „trifft zu“ (kumuliert) den gleichen Wert (43,50%). Bei den Männern erreicht man durch Kumulierung einen Höchstwert von 50% im negativen Bereich:

Tab. 39 "Der Weg zur gewünschten Information war umständlich"

	trifft voll zu	trifft zu	Kumuliert Positiv	Trifft teilweise zu	trifft nicht zu	trifft gar nicht zu	Kumuliert negativ
Frauen	1 (4,35%)	2 (8,70%)	=3 (13,05%)	10 (43,50%)	2 (8,70%)	8 (34,80%)	=10 (43,50%)
Männer	5 (16,67%)	5 (16,67%)	=10 (33,34%)	5 (16,67%)	9 (30%)	6 (20%)	=15 (50%)
Total	6 (11,32%)	7 (13,21%)	=13 (24,53%)	15 (28,30%)	11 (20,75%)	14 (26,42%)	=25 (47,17%)

6.5.2.2.1 Systemkooperativität

Um entscheiden zu können, ob das System als kooperativ bezeichnet werden darf, müssen die Ergebnisse der untersuchten Faktoren zusammengefasst werden. Es muss bei jedem Statements überprüft werden, in welcher Kategorie und in welchem Bereich (negativ oder positiv) der Höchstwert liegt. Dabei wird man sich an die Werte der Totale richten.

Es wird festgestellt, ob das Qualitätsurteil für BoRIS negativ oder positiv ausfiel. Wenn die positiven Bewertungen die negativen übertreffen, kann man das System als kooperativ bezeichnen. Einige Faktoren konnten dabei nicht berücksichtigt werden: zum einen, weil sie Komponente ansprechen, die im System nicht implementiert wurden, wie z.B. "Bestätigung" und "Dialoginitiative"; zum anderen, weil sie keine „Bewertungsdaten“ im Sinne eines positiven oder negativen Urteils liefern, wie "Hintergrundwissen", "Erwartungen" vor dem Versuch, und "Gestaltung der Eingaben". Der Einfluss dieser Faktoren auf die Qualitätsbeurteilung kann erst bei einer genauere Analyse im gesamten Interaktionskontext ermittelt werden (siehe Kap.7)

Kooperativität

Tab. 40 Qualitätsfaktoren

<u>Qualitätsfaktoren</u>		<u>Nr.</u>	<u>Statements</u>	<u>Bewertung</u>
1. Relevanz	1.1 Informationsmenge	1.	9.2. BoRIS bietet umfangreiche Information über Bochumer Restaurants	positiv

	1.2 Vollständigkeit der Informationen	2.	10.4 BoRIS bietet keine detaillierte Informationen über die gewünschten Restaurants	negativ
2.Ausdrucksform	2.1. Akustische Verständlichkeit	3.	10.8 Die Ansagen der Restaurantnamen und die Adresse war nicht deutlich	negativ
	2.2. Inhaltliche Verständlichkeit	4.	10.6 Ich konnte mir nicht die Auswahlkriterien am Anfang des Gesprächs nicht merken.	positiv
	2.3 Gestaltung der Äußerung	5.	9.11 Ich habe die Fragen von BoRIS beantwortet : <input type="checkbox"/> stichwortartig, <input type="checkbox"/> satzartig, <input type="checkbox"/> teils-teils	/
3. Korrektur- und Aufklärungsstrategien	3.1.1 Wiederholung der Systemeingaben	6.	10.3 BoRIS konnte seine Ansagen nicht wiederholen	negativ
	3.1.2 Wiederholung der Benutzereingaben	7.	10.5 Ich musste oft meine Eingaben wiederholen.	positiv
	3.2.1 Bestätigung der Systemeingaben	/	/	/
	3.2.2 Bestätigung der Benutzereingaben	/	/	/
4.Berücksichtigung kontextueller Bedingungen	4.1.1 Hintergrundwissen über System	8.	3.0 Haben Sie bereit ein automatisches sprachbasiertes Auskunftssystem benutzt?	/
	4.1.2 Hintergrundwissen über Thema der Infosuche	9.	4.0 Welche Informationen sollte ein solches Restaurantsauskunftssystem liefern?	/
	4.2.1 Erwartungen vor dem Test	10	5.0 Wenn Sie bei BoRIS anrufen, erwarten Sie: Statements 5.1 5.8	
	4.2.2 Erwartungen nach dem Test	11	9.1. BoRIS hat genau so funktioniert, wie ich es mir vorgestellt habe	negativ
5. Transparenz	5.1 Einfachheit der Bedienung	12.	9.9 Jeder kann mit BoRIS ohne Probleme umgehen	negativ

	5.2 Systemgrenzen	13.	10.10 Es war mir nicht klar, wie ich meine Eingaben modifizieren soll, um schnell die gewünschte Information zu bekommen	negativ
6. Dialoginitiative	6.1 Möglichkeit Fragen stellen zu können	/	/	/
	6.2 Möglichkeit das Gespräch zu führen, solange man möchte	/	/	/
7 Schnelligkeit	7.1 Schnelligkeit der Datenverarbeitung	14	9.7 Das Gespräch mit BoRIS hat schnell zur gewünschten Information geführt	positiv
		15	10.7 Die Pausen zwischen meinen Eingaben und den Antworten von BoRIS waren zu lang	positiv
	7.2 Dialogunterbrechungen	16	10.1 BoRIS hat plötzlich das Gespräch unterbrochen und ist spontan zu einem vorherigen Menu zurückgekehrt.	positiv
		17	10.2 BoRIS hat das Gespräch oft abgebrochen	positiv
8. Hilfefunktion		/	/	/

Tab. 41 Zufriedenheitsfaktoren

<u>Zufriedenheitsfaktoren</u>		Nr.	<u>Statements</u>	<u>Bewertung</u>
1. Einschätzung der kommunikativen Zielerreichung		17.	9.3 Ich bekam immer die gewünschte Antwort auf meine Eingaben	negativ
		18.	9.8 Der Anruf bei BoRIS hat sich gelohnt	positiv
2. Einschätzung des Computers als Gesprächspartner	2.1 Ähnlichkeit des Systems mit einem menschlichen Gesprächspartner	19	9.4 Ich spreche lieber mit einem Menschen als mit BoRIS.	negativ
		20.	9.5. Ich konnte mich äußern, wie in einem normalen Gespräch mit einem Menschen	negativ
	2.2 Vertrauenswürdigkeit	21	9.6 Ich bin davon überzeugt, dass BoRIS mir richtige Information geliefert hat	positiv

	1.3 Stimmung während des Gesprächs	22.	9.10 Durch das Gespräch mit BoRIS fühlte ich mich:	
		23.	wohl <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> unwohl	positiv
		24.	sicher <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> unsicher	positiv
		25.	ruhig <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> aufgeregt	positiv
		26.	zufrieden <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> unzufrieden	negativ
3. Einschätzung des Gesprächsaufbaus		27.	10.9 Der Weg bis zu der gewünschten Information war umständlich	positiv

Von insgesamt 35 Statements konnten 27 ausgewertet werden. Dabei stellte sich man heraus, dass 13 Statements positiv und 9 Statements negativ bewertet wurden. Somit kann man den Aufschluss ziehen, dass das System nach der Meinung der Versuchspersonen als kooperativ zu bezeichnen ist.

6.5.2.2.3 Rangordnungsstatements

Der letzte Teil des Fragebogens enthielt eine Reihe von fünfzehn Statements, die sich auf die Einhaltung von Kriterien beziehen, die einen guten Ablauf eines Dialogs sichern sollten. Die Kriterien entsprechen den Qualitätsfaktoren. Ziel dabei ist eine Rangordnung der Qualitätsfaktoren herauszuarbeiten, indem man sich bei jeder Behauptung an den Totalwert der höchsten positiven Kategorie „trifft vollkommen zu“ orientiert: der Totalwert, der sich bei dieser Kategorie ergibt, bestimmt die Rangordnung der Qualitätsfaktoren. Im Falle, dass der Totalwert bei „trifft vollkommen zu“ die gleichen Prozentzahlen bei mehreren Kriterien aufweist, ist es sinnvoll noch eine andere Kategorie in Betracht zu ziehen, um festzustellen, welche der Kriterien auf der Rangliste Priorität genießt: zu der Kategorie „trifft vollkommen zu“ wird die Kategorie „trifft zu“ addiert und die höhere Summe hat dabei den Vorzug. Wenn zwischen den Kriterien immer noch Gleichheit besteht, wird aus dieser Summe die Summe der negativen „trifft nicht zu“ und „trifft gar nicht zu“ subtrahiert, um den höheren Wert zu ermitteln und dabei die Rangordnung festzustellen. In der Reihenfolge ihrer Wichtigkeit haben die Befragten die Kriterien wie folgt beurteilt:

Tab. 43 Rangordnung der Qualitätsfaktoren

Rang	Kriterien / Behauptungen	trifft voll zu + trifft zu -[trifft nicht zu + trifft gar nicht zu] =	Qualitätsfaktor
1	11.2 es ist wichtig, dass BoRIS meine Eingaben immer gut versteht	83,02%.	Verständlichkeit /Relevanz
2	11.3 es ist wichtig, dass BoRIS	79,25	Verständlichkeit/Relevanz

	klare Fragen stellt			
3	11.5 es ist wichtig, dass BoRIS einfach zu bedienen ist.	73,58% + 20,75%	= 94,33%	Einfachheit der Bedienung/ Transparenz/
4	11.6 es ist wichtig, dass BoRIS passende Informationen auf meine Anfrage liefert.	73,58% + 18,87%	= 92,54%	Informationsgehalt/Relevanz
5	11.1 es ist wichtig, dass BoRIS seine Fragen wiederholt, wenn ich nicht mehr weiter weiß.	52,83%		Wiederholung/Korrektur- und Aufklärungsstrategie
6	11.13 es ist wichtig, dass es einfach ist, die Eingaben zu modifizieren.	51%		Systemgrenze/ Transparenz
7	11.8 es ist wichtig, dass nachdem ich die gewünschte Information bekommen habe, das Gespräch weiter führen soll, solange ich noch zusätzliche Fragen habe.	50,94%		Gespräch führen , solange man möchte/Dialoginitiative
8	11.7 es ist wichtig, dass ich BoRIS Fragen stellen kann.	35,85% + 35,85%	= 71,70	Fragen stellen zu können /Dialoginitiative
9	11.10 es ist wichtig, dass ich von Anfang an weiß, welche Art von Information mir BoRIS liefern kann und welche nicht.	35,85% + 30,19% - [5,66% + 9,43%]	= 50,95%	Systemgrenze/ Transparenz
10	11.9 es ist wichtig, dass ich von Anfang an weiß, wie ich meine Eingaben formulieren soll	35,85% + 30,19% - [13,21% + 9,43%]	= 43,40%	Systemgrenze/ Transparenz
11	11.14 es ist wichtig, dass das Gespräch ohne Unterbrechungen verläuft.	30,19% + 30,19%	= 60,38%	Dialogunterbrechungen/Schnelligkeit
12	11.15 es ist wichtig, dass BoRIS mit einer Hilfefunktion ausgestattet ist.	30,19% + 20,75%	= 50,94%	Hilfefunktion
13	11.4 es ist wichtig, dass BoRIS schnell auf meine Eingaben reagiert	22,64% + 52,85%	=75,49%	Schnelligkeit der Datenverarbeitung/ Schnelligkeit
14	11.11 es ist wichtig, dass ich von BoRIS eine Bestätigung meiner Eingaben bekomme.	22,64% + 20,75%	= 43,39%	Bestätigung der Eingaben/ Korrektur- und Aufklärungsstrategien
15	11.12 es ist wichtig, dass die Pausen nach meinen Eingaben von einem "Warten Sie einen Moment" ausgefüllt werden.	13,21%		Bestätigung der Pausen! Korrektur- und Aufklärungsstrategien

Die Rangordnung zeigt, dass die Verständlichkeit, Einfachheit der Bedienung, Informationsgehalt, Wiederholung, Einfachheit der Modifikationen und Dialoginitiative die wichtigsten Qualitätsfaktoren eines guten Dialogablaufes aus Benutzersicht sind. Sie wurden von über 50% aller Befragten gewählt und sollen für die Qualitätsgestaltung einer Dialogsteuerung als Referenzmuster verstanden werden.

7. Diskussion der Ergebnisse

In diesem Kapitel soll nun eine Zusammenfassung der Ergebnisse gemacht werden, die aufgrund der Auswertung beider Befragungsrunden am sinnvollsten erscheinen.

Die Versuchspersonen wählten aus 7 vorgeschlagenen Qualitätsfaktoren die wichtigsten aus. Diese sind: Verständlichkeit, Einfachheit der Bedienung, Wiederholungsmöglichkeiten, Transparenz der Systemgrenzen und Dialoginitiative. Alle diese Kriterien wurden von über 50% der Befragten als sehr wichtig für einen guten Gesprächsablauf bezeichnet. Diese Faktoren beeinflussen, aus der Sicht der Befragten, am meisten die Qualitätsbeurteilung des Gesamtsystems und implizit der Dialogsteuerung. Die Reihenfolge dieser Faktoren ist nicht zufällig: sie hängt mit den Erfahrungen zusammen, welche die Versuchspersonen mit BoRIS machten. Es wird nun hier in einer verkürzten Form gezeigt, wie das System aufgrund dieser Faktoren beurteilt wurde und welche Verbesserungsvorschläge in Frage kommen.

Die inhaltlich-kognitive Verständlichkeit wird als der wichtigste Qualitätsfaktor angesehen: das System sollte einerseits die Eingaben immer gut verstehen und andererseits klare Fragen stellen. Die qualitative Befragung zeigte, dass für die meisten Befragten das System deutliche Fragen stellte. Sie hatten nur teilweise Probleme, sich die Kriterien am Anfang des Gesprächs zu merken, was sich bei der quantitativen Befragung bestätigt hat. In diesem Fall empfiehlt sich, die Menüführung zu vereinfachen, indem man den Versuchspersonen ein bis maximal drei Suchkriterien anbietet.

Bei der qualitativen Befragung hat man festgestellt, dass die meisten Befragten nur teilweise das Gefühl hatten, verstanden zu werden. Es war schwierig zu überprüfen, ob das System die Eingaben richtig verstanden hat. Aus diesem Grund wurde von einer Versuchsperson vorgeschlagen, eine Bestätigung der Eingaben einzuführen. Die Untersuchungen der quantitativen Befragung zeigte, dass eine solche Maßnahme nicht mit eindeutiger Mehrheit begrüßt würde: viele Versuchspersonen befürchten eine zusätzliche Verlängerung des Gesprächs. Eine Lösung dazu wäre eine implizite Bestätigung zu liefern (siehe Abschnitt 3.3, Seite 25), wie z.B.: „In welchem Stadtteil möchten Sie *chinesisch* essen gehen?“. Dadurch können unnötige *Turns* vermieden und gleichzeitig sicher gestellt werden, dass der Input richtig verstanden wurde.

Verwirrend war für viele Versuchspersonen, dass das System nur eine geringe Zahl von Küchenarten- und Stadtteilmöglichkeiten zur Auswahl stellt. Manche trauten sich nicht Eingaben zu machen, die vom System nicht vorgeschlagen wurden. Diese Fragestellung sollte man hier offen lassen: „Welche landesspezifische Art der Küche bevorzugen Sie?“ bzw. „In welchem Stadtteil möchten Sie essen gehen?“

Die am meisten kritisierten Systemmängel, sowohl bei der qualitativen als auch bei der quantitativen Befragung war die schlechte akustische Verständlichkeit der Sprachsynthese:

sie sei zu schnell und zu schlecht verständlich. Die qualitative Untersuchung hat gezeigt, dass sich für die meisten Versuchspersonen der Anruf bei BoRIS nur teilweise gelohnt hat; als Ursache dafür wurde die schlechte Sprachsynthese genannt. Die Synthese wurde nur für Restaurantansagen verwendet; für alle anderen Äußerungen setzte man eine menschliche Stimme ein. Dadurch, dass die menschliche Stimme bei der qualitativen Befragung als angenehm bezeichnet wurde, wäre es ratsam alle Ansagen mit einer menschlichen Stimme zu gestalten.

Die Einfachheit der Systembedienung spiegelt sich in dem vom Benutzer wahrgenommenen Schwierigkeitsgrad des Gesprächs und in dem damit verbundenen Stress wieder. Der Stress ist ein Zufriedenheitsindikator, der mit der Frage nach der Stimmung während des Gesprächs operationalisiert wird.

Die Ergebnisse beider Befragungen zeigten, dass den meisten Befragten das Gespräch mit dem System Probleme bereitet hat und dass sie geteilte Meinungen bezüglich der Behauptung „Jeder kann mit BoRIS ohne Probleme umgehen“ hatten. Eine Bestätigung dazu brachten die Angaben, welche die Befragten zur Beschreibung ihres seelischen Zustandes während des Gesprächs machten: die weit verteilten Werte, die sich daraus ergaben zeigten, dass die Versuchspersonen während des Gesprächs teilweise beansprucht wurden.

Die Einfachheit der Bedienung hängt noch mit der Einfachheit der Eingabeformulierung und -modifizierung zusammen. Viele Befragte beklagten, sie hätten nicht gewusst, wie sie ihre Eingaben formulieren bzw. modifizieren sollten. Das System sollte im Anfangsmenü dem Benutzer entgegen kommen und ihnen erklären, welche Formulierungsmöglichkeiten sie haben: ob sie satzartig oder stichwortartig ihre Eingaben gestalten sollen und ob sie ihren Informationswunsch komplett in einem Satz formulieren können oder ob sie warten müssen, bis die einzelnen Kategorien abgefragt werden.

Als schwierigste Modifikation erwies sich die Eingabe des Preises. Die Versuchspersonen waren unsicher, was das System unter „teuer“ oder „billig“ verstand. Hier sollte man vielleicht nur nach Zahleneingaben fragen, z.B. 20 DM. Wenn die Werte modifiziert werden, wäre sinnvoll anzugeben: „es gibt kein teureres chinesisches Restaurant in Altenbochum.“ So können die Benutzer erkennen, warum das System kein passendes Restaurant gefunden hat und gleichzeitig erfahren, unter welcher Preisklasse das System den angegebenen Preiswert, z.B. 20 DM, ordnet.

Eine Vereinfachung der Dialogstrukturierung sollte auch bei der Frage: „Möchten Sie ihre Frage präzisieren?“ stattfinden. Diese Frage kann nur mit „Ja“ oder „Nein“ beantwortet werden. Das System sollte die Versuchspersonen explizit auffordern, die Frage entsprechend zu beantworten (z.B. „Möchten Sie ihre Frage präzisieren? Bitte Ja oder Nein

angeben“) oder die Möglichkeit bieten, die Frage direkt mit der Präzisierung beantworten zu können.

Die Relevanz einer Äußerung hängt von ihrem Informationsgehalt ab. Eine passende Aussage liefert genau die Information, nach welcher gefragt wurde. Die Relevanz wird einerseits durch die Verständlichkeit der Eingaben bestimmt (falsch verstandene Eingaben führen zu unpassenden Antworten) und andererseits durch die Information aus der Datenbank. Was die Relevanz anbelangt, hat sich bei der qualitativen Befragung gezeigt, dass die Versuchspersonen die Telefonnummer des Restaurants, eine Wegbeschreibung und Reservierungsmöglichkeiten vermisst haben. Diese Wünsche wurden durch die Erwartungen der Versuchspersonen vor dem Versuch bestätigt. Inwieweit es möglich ist diese Informationen in der Datenbank zu berücksichtigen und entsprechend im Gesprächsaufbau zu implementieren wird sich sicherlich bei künftigen Systemverbesserungen zeigen.

Die Informationsmenge und die Ansagegeschwindigkeit verwirrten oft die Versuchspersonen, die sich dabei eine Wiederholung der Eingaben gewünscht haben. Das System kann wohl auf Wunsch, bis auf die Adressenansage, alle seine Angaben wiederholen, nur war dies für die meisten Befragten unklar. Aus diesem Grund sollte man eine Wiederholung der Restaurantansagen bieten und den Versuchspersonen dabei signalisieren, dass sie die Möglichkeit haben, die Systemangaben nochmals zu hören. Allerdings wird in der Literatur [Bernsen, et. al., 1998] empfohlen, die Wiederholungen der Systemfragen mit anderen Worten zu formulieren, um eventuelle Unklarheiten der ursprünglichen Frage zu beseitigen.

Die Untersuchungen zeigten, dass es für viele Versuchspersonen äußerst wichtig war, dass der Dialog nicht abrupt endet. Sie hätten sich gewünscht, das Gespräch weiter führen zu können, solange sie Fragen hatten und eventuell sogar mit einem Anruf mehrere Anfragen durchzuführen. Die Gestaltung der Dialogstrategie sollte künftig flexibler werden und mehr Spielraum für eventuelle Nachfragen lassen. Die Versuchspersonen sollen aus jedem Dialogzustand ins Hauptmenü zurück kehren und entsprechend ihre Suchkriterien neu formulieren können.

Ein anderer Kritikpunkt, der im letzten Teil der quantitativen Befragung aufgrund seiner sekundären Bedeutung nicht berücksichtigt wurde, war die einheitliche Gestaltung der Dialogstruktur: wenn zu einer Anfrage ein einziges Restaurant gefunden wird, gibt das System sofort die Adresse aus, ohne nach den anderen Kriterien (Preis, Ort, Öffnungszeiten) zu fragen. Diese Vorgehensweise führt zu keinem Dialogabbruch, verunsichert aber die Befragten, die sich an ein bestimmtes Dialogschema gewöhnt hatten.

Das System sollte den Benutzern signalisieren, dass es in seiner Datenbank nur ein einziges passendes Restaurant gibt.

Die Untersuchung zeigte auch, dass Dialogunterbrechungen von den meisten Befragten als nicht besonders störend empfunden wurden. Das kann dadurch erklärt werden, dass solche Pausen zu üblichen Gesprächen zwischen Menschen gehören. Dies führte zur Einstufung des Faktors „Schnelligkeit“ auf den elften Rangplatz.

Bei der Rangordnung, die sich bei der Auswertung der letzten 15 Statements ergab, wurden vier qualitätsbeeinflussende Faktoren nicht überprüft – das Hintergrundwissen, die Erwartungen der Versuchspersonen, die Gestaltung der Eingaben und die akustische Verständlichkeit²⁵. Es wird sich hier zeigen, welche Rolle sie bei der Qualitätsbeurteilung der Interaktion spielen.

Die Befragten verfügten größtenteils über wenig Hintergrundinformationen über das System: Erfahrungen damit hatten nur sehr wenige Befragte (1 Versuchsperson bei dem qualitativen Interview bzw. insgesamt 17% aller Befragten bei dem quantitativen Interview). Aus diesem Grund ist es unwahrscheinlich, dass die Berücksichtigung dieses Faktors die Qualitätsbeurteilung besonders beeinflusst hat.

Fehlen an Hintergrundinformationen über das Thema „Restaurantessen“ wurde nur bedingt festgestellt, vor allem wenn es darum ging, eine Preisklasse anzugeben: einige Versuchspersonen versuchten teure griechische oder chinesische Restaurants zu finden, die allgemein als preiswert bekannt sind. Solche Fälle kamen selten zu Stande, verursachten aber erhebliche Schwierigkeiten im Dialogverlauf und führten schließlich zum Dialogabbruch.

Ganz anders ist das Verhältnis des Faktors „Erwartungen“ zur Qualitätsbeurteilung des Systems. Die Analyse der qualitativen Interviews haben gezeigt, dass Versuchspersonen, die vom System keine konkreten Vorstellungen hatten und entsprechend ein diffuses Erwartungsbild abgaben, BoRIS allgemein besser beurteilt haben als Versuchspersonen, die gut vordefinierte Vorstellungen vom System hatten.

Bei der quantitativen Befragung wurde festgestellt, dass je „optimistischer“ die „Prognosen“ BoRIS gegenüber waren, desto besser fiel das Qualitätsurteil aus. Vor allem die Frauengruppe wies, im Vergleich zur Männergruppe allgemein positivere Erwartungswerte auf. Auch wenn die meisten Befragten relativ niedrige Erwartungen bezüglich eines der zu untersuchenden Merkmale hatten, waren die Werte der Frauengruppe im negativen Bereich deutlich kleiner als die der Männer.

Die Auswertung lieferte zum Teil sehr weit verteilte Werte, deren Interpretation ohne Kumulierung undenkbar gewesen wäre. Das erklärt sich dadurch, dass die

²⁵ Die Gründe dafür wurden im Abschnitt 6.5.1 bereits erklärt.

Versuchspersonen extrem unterschiedliche Erfahrungen mit BoRIS machten, die einerseits durch ihre Erwartungen bedingt waren und andererseits durch ihr kommunikatives Verhalten. Die Untersuchung zeigte, dass sich Frauen allgemein viel lockerer im Gespräch mit BoRIS zeigten als Männer. Sie versuchten sich zu äußern, wie in einem normalen Gespräch und formulierten ihre Äußerungen hauptsächlich als Kombination von Stichwörtern und Sätzen. Sie bekamen deutlich schneller die gewünschte Information von BoRIS. Dagegen erschien das sprachliche Verhalten der Männer unnatürlich und etwas verkrampft; sie schienen mit BoRIS kein Gespräch führen zu wollen, sondern versuchten nur durch entsprechende Bedienung Informationen zu bekommen. Männer orientierten sich hauptsächlich an den Systemanweisungen und führten ein „computermässiges“, stichwortartiges Gespräch durch, indem sie glaubten, den richtigen Funktionsmechanismus des Systems verstanden zu haben. Dadurch kamen sie langsamer und oft nicht ohne Schwierigkeiten ans Ziel. So erklärt sich, warum Frauen den Weg bis zur gewünschten Informationen leichter fanden als Männer, BoRIS als Gesprächspartner besser beurteilten und sich allgemein wohler, ruhiger und sicherer als Männer fühlten. Entsprechend beurteilten die Frauen das System deutlich besser als die Männer: prozentual wählten die Kategorien „ausgezeichnet“ und „gut“ doppelt so viele Frauen als Männer. Bemerkenswert ist noch, dass die Kategorien „schlecht“ und „extrem schlecht“ von keiner weiblichen Versuchsperson angekreuzt wurden; diese Kategorien wurden aber von 20% der Männer gewählt.

Noch ein Grund, warum die Frauen das System besser beurteilt haben als die Männer, ist die Tatsache, dass laut Statistik Frauen im Vergleich zu Männern lieber Fragen beantworten als Fragen stellen. So erklärt sich, warum das System den Männern starrer erschien als den Frauen, was sich auf die Qualitätsbeurteilung auswirkte.

Die Ergebnisse dieser Arbeit haben gezeigt, dass obwohl das System mit knapper Mehrheit als kooperativ bezeichnet werden konnte, das Qualitätsurteil des Gesamtsystems überwiegend schlecht war. Dies erklärt sich dadurch, dass die Qualitätsfaktoren in unterschiedlichem Maße die Qualitätsbeurteilung beeinflussen (z.B. der Faktor „Verständlichkeit“ wird im Vergleich zum Faktor „Bestätigung“ als viel wichtiger eingeschätzt) und dass die am wichtigsten für den Dialogablauf empfundenen Qualitätsfaktoren zum Teil fehlten (Dialoginitiative) und zum Teil als negativ bewertet wurden (Systemgrenze, Einfachheit der Bedienung, Wiederholungen). Ein Faktor, welcher sich auf die Qualitätsbeurteilung sehr negativ auswirkte, war die schlechte akustische Verständlichkeit der synthetischen Stimme, die bei dem Anruf bei BoRIS die Fragen stellte. Dieser Faktor ist technischer Natur und von der Dialogsteuerung unabhängig. Für künftige Untersuchungen, die sich gezielt mit der Qualitätsbeurteilung der Dialogsteuerung auseinandersetzen, sollte man diese technische Komponente verbessern.

Es konnte festgestellt werden, dass es verschiedenen Zusammenhänge zwischen den Faktoren gibt, welche der Untersuchung einen einheitlichen Charakter boten. Welcher Natur diese Zusammenhänge sind, inwieweit sich die Faktoren gegenseitig beeinflussen und wie sich das auf die Qualitätsbeurteilung wirkt, sollte noch untersucht werden.

8. Ausblick

Trotz unbestrittener Fortschritte und dem damit verbundenen kommerziellen Erfolg sind die natürlichsprachlichen Dialogsysteme noch verbesserungsbedürftig. Diese Arbeit hat versucht einen Beitrag zur qualitativen Verbesserung der Dialogsteuerung aus kommunikationswissenschaftlicher Perspektive zu leisten, indem sie Maßstäbe zur Qualitätsbeurteilung vorschlug. Dabei wurde festgestellt, welche Faktoren der Mensch-Maschine Interaktion für die Dialogssteuerung qualitätsrelevant sind. Diese Faktoren teilte man zum Zwecke der Analyse in drei verschiedene Aspekte: Frage-Antwort Aspekt, Systemmöglichkeiten und Zufriedenheit der Benutzer. Die Zusammenfassung der Ergebnisse dieser drei Untersuchungsebenen sollte darüber Auskunft geben, wie kooperativ das System sich verhält. Das Ausmaß der Kooperativität zeigt, inwiefern das System sich einem menschlichen Gesprächspartner nähert. In diesem Zusammenhang versteht sich die Kooperativität als Indikator für die Qualitätsbeurteilung einer Dialogsteuerung.

Anschließend wurden die Faktoren aufgrund ihrer Wichtigkeit geordnet. Dabei stellte man eine Rangordnung der Faktoren fest, die als Maßstab der Qualitätsbeurteilung gelten soll. Die zu evaluierenden Aspekte wurden aus der Literatur übernommen [Bernsen, 1998] und [Love et al., 1994] und für die Untersuchung im Rahmen dieser Arbeit entsprechend neu strukturiert.

Für die empirische Untersuchung der Faktoren wurde zwei Arten von Befragungen durchgeführt: eine qualitative und eine quantitative Befragung. Die Kombination qualitativer und quantitativer Verfahren zeigte durchaus positive Effekte und ist künftigen Untersuchungen im diesen Bereich zu empfehlen. Vor allem die Anwendung qualitativer Aspekte brachte ein unglaublich reiches Material an Erfahrungen, Eindrücken, Beobachtungen, Kritiken und Verbesserungsvorschlägen, die wertvolle Hinweise zur Qualitätsbeurteilung der Dialogssteuerung, sowie des Gesamtsystems lieferten. Sehr hilfreich zeigte sich diese Methode für die Aufbereitung der quantitativen Methode: aus den Interviews konnten wiederkehrende Äußerungen festgestellt werden, die man thematisch gruppierete und auf eine einzige Aussage reduzierte. Diese Aussagen wurden als Behauptungen bei der Gestaltung der quantitativen Befragung verwendet. Der Vorteil dabei war, dass dadurch vermieden wurde, nach Sachverhalten zu fragen, an welche die Befragten nicht gedacht hatten.

Die Befragungen bezogen sich auf empirische Versuche mit BoRIS, dem Bochumer Restaurantinformationssystem. Für eine einheitliche Gestaltung der Versuche wurden sechs Szenarios entwickelt, die man mit insgesamt 61 Personen durchführte. So konnten die im theoretischen Teil festgelegten Qualitäts- und Zufriedenheitsfaktoren im empirischen Teil untersucht werden. Die Befragungen stellten eine Art Diagnostiktest dar, wobei die Faktoren an den Systemleistungen überprüft wurden. Bei der quantitative

Befragung wurde eine Rangordnung der Qualitätsfaktoren hergestellt. Diese Rangordnung besitzt Allgemeincharakter und, obwohl sie aufgrund der Erfahrungen der Versuchspersonen mit BoRIS entstanden ist, kann sie auch bei anderen Untersuchungen gleicher Art eingesetzt werden.

Aufgrund der Analyse dieser Faktoren wurde das System als kooperativ bewertet. Der Befund der Systemkooperativität stimmt nicht mit dem Qualitätsurteil überein: das System wurde von den meisten Versuchspersonen als ziemlich schlecht beurteilt. Die Übereinstimmungslosigkeit erklärt sich dadurch, dass obwohl die für die Kooperativität minimalen qualitativen Anforderungen erfüllt wurden, die wichtigsten Qualitätsfaktoren teilweise mangelhaft blieben.

Obwohl die Arbeit auf einer umfangreichen Literatur aufbaut, ist sie was ihr konkretes Untersuchungsobjekt betrifft eine Pionierarbeit im Bereich der Kommunikationswissenschaften, die sich auf keine vorherbestimmten Modelle berufen konnte. Infolgedessen hebt sie keine Ansprüche auf Vollständigkeit und Endgültigkeit der gewonnenen Ergebnisse. Eventuelle unzureichend strukturierte Artikulationen der einzelnen Teile sind auf dem Mangel wegweisender, gleichartiger Unternehmungen zurückzuführen. Die theoretischen und experimentellen Schwierigkeiten, die man dabei zu überwinden hatte, erwiesen sich als sehr groß. Unter diesen Umständen können die Resultate dieser Studie als Ansatz zur Entwicklung eines künftigen Modells dienen, welches die Qualitätsbeurteilung der Dialogsteuerung von NSD-Systemen untersuchen soll.

9 Literatur²⁶

Ahrenberg, L. Dahl, N. & Jönsson, A. (1995). Coding Schemes for Studies of Natural Language Dialogue. (Proceedings of the 1995 AAAI Spring Symposium on empirical methods in discourse interpretation and generation) (S. 8-13). Stanford, Menlo Park, CA: AAAI Press.

American Psychological Association (1994). Publication Manual of the American Psychological Association (4. Aufl.). Washington, D.C.: APA.

Aust, H., Oerder, M., Seide, F. & Stenbiss, V. (1995). The Philips Automatic Train Timetable Information System. Speech Communication, 17, 249-262.

Bersen, N. O., Dybkjær H. & Dybkjær L. (1998). Designing Interactive Speech Systems. London: Springer Verlag.

Bodden, M & Jekosch, U. (1996). Entwicklung und Durchführung von Tests mit Versuchspersonen zur Verifizierung von Modellen zu Berechnung der Sprachübertragungsqualität. Abschlußbericht. Institut für Kommunikationsakustik, Ruhr-Universität Bochum.

Boner, A. (1994). Spracherkennung mit Computer. Schweiz: AT Verlag.

Bortz, J. & Döring, N. (1995). Forschungsmethoden und Evaluation (2.Auflage). Berlin: Springer Verlag.

Bunt, H. (1994). Context and Dialog Control. Think, 3, 19-31.

Carberry, S. (1990). Plan recognition in natural language dialogue. Cambridge, MA: Harvard University Press.

Carletta, J., Isard, A., Isard, S., Kowtko, J.C., Doherty-Sneddon, G. & Anderson, A.H. (1997). The Reability of a Dialog Structure Coding Scheme. Computational Linguistics, 23 (1), 13-31.

²⁶ Die Literaturangaben wurden nach APA-Vorschriften [American Psychological Association, 1994] zusammengestellt.

Clark, H. H. & Schaefer, E. F. (1989). Collaborating on Contributions to Conversation. In R. Dietrich & Graumann, C. F. (Hrsg.), Language Processing in Social Context (S.123-152). Amsterdam.

Danieli, M., Gerbino, E. (1995). Metrics for evaluating dialogue strategies in a spoken language system. (Proceedings of the 1995 AAAI Spring Symposium on empirical methods in discourse interpretation and generation) (S. 34-39).Stanford, Menlo Park, CA: AAAI Press.

Delogu, C., Carlo, A. D., Sementina, C. & Stecconi, S. (1993). A Methodology for Evaluating Human-Machine Spoken Language Interaction (Proceedings 3rd Europ. Conf. An Speech Communication and Technology, EUROSPEECH '93, Nr.3) (S.1427-1430). Berlin.

Dengel, A. (1994). Künstliche Intelligenz. Mannheim [u.a.]: Meyers Verlag.

Dudda, Ch. (2001). Evaluierung eines natürlichen Dialogsystem für Restaurantauskünfte. Diplomarbeit. Institut für Kommunikationsakustik, Ruhr-Universität Bochum.

Dybkjær, H. & Dybkjær, L. (1993). Wizard of Oz Experiments in the Development of the Dialogue Model for Pl. Report 3 from the Danish Project in Spoken Language Dialogue Systems. Roskilde: University Press.

Eckert, W. & McGlashan, S. (1993). Managing Spoken Dialogues for Information Services (Proceedings 3rd Europ. Conf. An Speech Communication and Technology, EUROSPEECH '93, Nr.3) (S.141-144). Berlin.

Fraser, G. & Gilbert., N.(1991). Simulating Speech Systems. Computer Speech and Language, 5 , 81-99.

Fischer, M., Maier, E. & Stein, A. (1994). Generating Kooperative Systems Responses in Information Retrieval Dialogue (Proceedings of the International Workshop on Natural Language Generation INLGW '94) (S. 207-216.). Kennbunkport, Maine.

Gleiss, N. (1992). Usability-Concepts and Evaluation. TELE (English edition), Swedish Telecommunication Administration, 2, 24-30.

Goodine, D., Hirschman, L., Polifroni, J., Seneff, S., & Zue, V. (1992). Evaluating Interactive Spoken Language Systems (Proceedings of the 1992 International Conference on Spoken Language Processing, ICSLP '92) (S. 197-200). Banff.

Görtz, G. (Hrsg.) (1993). Einführung in die künstliche Intelligenz. Bonn: Addison-Wesley Verlag.

Grau B., Sabah, G. & Vilnat, A. (1994): Control in Man-Machine Dialogue. Think, 3, 32-55.

Grice, P. (1998). Studies in the Way of Words. Cambridge MA: Harvard University Press.

Grosz, B. J. & Sidner, C. L.(1986). Attention, Intention, and the Structure of Discourse. Computational Linguistics, 12 (3), 175-204.

Hansen, P. M., Holtse, P., Nielsen, H. & Petersen, N. R.: Speech Synthesis - Teaching a Computer Spoken Language. Teletechnik, 1/2, 153-156.

Hirschman, L., Thompson, H. (1996). Overview of Evaluation in Speech and Natural Language Processing. In R. Cole (Hrsg.), Survey of the State of the Art in Human Language Technology. Cambridge: Cambridge University Press.

Hoepfner, W. (1999). Der Mensch-Maschine-Dialog.
http://www.uni-duisburg.de/FB3/CL/seiten_d/MM-Dialog.html

Jönsson, A. (1993). Dialogue Management for Natural Language Interfaces. An Empirical Approach. Ph.D. thesis. Linköping Studies in Science and Technology, Nr.312. Linköpings universitet.

Jekosch, U. (2000). Sprache hören und beurteilen: Ein Ansatz zur Grundlegung der Sprachqualität. Universität/Gesamthochschule Essen.

Kamm, C., Walker, M. A. and Rabiner, L.. (1997). The role of speech processing in human-computer intelligent communication. Speech Communication, 23, 263-278.

Lamnek, S., (1998). Qualitative Sozialforschung. München: Beltz, Psychologie Verlag.

Laswell, H. D., (1948). The Structure and Function of Communication in Society. In L. Bryson (Hrsg.): The Communication of Ideas (S. 37-51). New York.

Loderer, G. (1998). Evaluierung von Dialogstrategien eines natürlchsprachlichen Dialogssystem durch Wizard-of-Oz Experimente. Diplomarbeit. Institut für med. Kybernetik und Artificial Intelligence, Universität Wien.

Love, S., Dutton, R. T., Jack, M. A. & Stentiford, F. W. M. (1994). Identifying Salient Usability Attributes for Automated Telephone Services (Proceedings of the International Conference on Spoken Language Processing, ICSLP '94) (S. 1307-1311). Yakohama.

Mangold, H. (Hrsg.) (1992). Sprachliche Mensch-Maschine Kommunikation. München: R. Oldenbourg Verlag.

Mann, W. C., and S. A. Thompson. (1987). Rhetorical Structure Theory: A Theory of Text Organization. Information Sciences Institute, University of Southern California.

Paetau, M. (1990). Mensch-Maschine Kommunikation, Software, Gestaltungspotentiale, Sozialverträglichkeit. Frankfurt [u.a.]: Campus Verlag.

Pereira, F. (1997): Sentence Modeling and Parsing. In: Cole, Ronald (Hrsg.): Survey of the State of the Art in Human Language Technology (S. 111-121). Cambridge.

Rabiner, L. R. & Juang, B. H. (1986). An introduction to the Hidden Markov Model, IEEE ASSP Magazine, 4-16.

Rehmann, S. (1999). Entwurf und Implementierung einer Versuchsumgebung zum Testen von Sprachsynthese und Spracherkennung in einer Dialogsituation. Arbeitsbericht, Institut für Kommunikationsakustik, Ruhr-Universität Bochum.

Relaño, G.L., Tapias, D., Villar, J.M. (1999). Flexible Mixed-Initiative Dialogue for Telephone Services (Proceedings of the Europ. Conf. An Speech Communication and Technology, EUROSPEECH '99, Nr.3) (S. 1179-1182). Budapest.

Sadek, D., De Mori, R. (1997). Dialogue Systems. In R. De Mori (Hrsg.), Spoken Dialogues with Computers (S. 523-561). San Diego: Academic Press.

Schnell, R., Hill, P. B., Esser, E. (1989). Methoden der empirischen Sozialforschung (2.Auflage). München: R. Oldenbourg Verlag.

Searle, J. R. (1969). Speech Acts. An Essay in the Philosophy of Language. Cambridge. Cambridge University Press.

Searle, J. R. (1979). Expression and Meaning. Studies in the Theory of Speech Acts. Cambridge: Cambridge University Press.

Shannon, C. E., Weaver, W. (1948). The Mathematical Theory of Communication. Urbana: University of Illinois Press

Simpson, A., Fraser, N.M. (1993): Black-Box and Glass-Box Evaluation of the SUDIAL-System, (Proceedings 3rd Europ. Conf. An Speech Communication and Technology, EUROSPEECH '93, Nr.3) (S.1423-1426). Berlin.

Smith, R. W., Hipp, R. D. & Biermann, A. W. (1994). Spoken Natural Language Dialog Systems: A Practical Approach. Oxford University Press, 1994.

Stein, A & Maier, E. (1994). Structuring collaborative information-seeking dialogues. Knowledge Based Systems, 8 (2/3), 82-93.

Ungeheuer, G. (1972). Sprache und Kommunikation. (2. Auflage). Hamburg.

Wahlster, W. (Hrsg.) (1987). Künstliche Intelligenz. Bonn: Addison-Wesley Verlag.

Walker, M. (1993). Informational Redundancy and Resource Bounds in Dialogue. PhD thesis, University of Pennsylvania.

Walker, M. A., Litman, D. J., Kamm, C. A., Abella, A. (1997). PARADISE: A Framework for Evaluating Spoken Dialogue Agents, (Proceedings of ACL/EACL, '97) (S.271-281). Madrid.

Zue, V. W.(1994): Toward systems that understand spoken language. IEEE Expert, 9 (1), 51-59.

Anhang B

Die Szenarios

Szenario 1.

Sie sind vor kurzem nach Bochum gezogen, Sie kennen sich in der Stadt überhaupt nicht aus aber möchten auswärts essen. Ein Bekannter gibt Ihnen einen guten Tip: BoRIS, die Bochumer Restaurantauskunft kann Ihnen dabei helfen.

Bitte führen Sie das Telefonat durch und notieren Sie was BoRIS für Sie gefunden hat, bzw. das/die Restaurant(s) und die Adresse(n).

Restaurant:

Adresse:

Restaurant:

Adresse:

Kommentar:

Szenario 2

Es ist Montag Abend und Ihre Freundin möchte unbedingt zum Chinesen essen gehen. Es wird wahrscheinlich schwierig sein etwas zu finden, denn bekanntlich haben viele Restaurants montags zu. Um von ungünstigen Öffnungszeiten nicht überrascht zu werden, entscheiden Sie sich für BoRIS. Falls das System doch nichts passendes findet, versuchen Sie Ihre Angaben zu modifizieren. Schließlich wollen Sie heute Abend doch nicht hungern! Notieren Sie anbei , was BoRIS für Sie gefunden hat:

Restaurant:
Adresse:

Restaurant:
Adresse:

Kommentar:

Szenario 3

Sie wohnen in Bochum-Grumme und Sie wollen heute Abend (Mittwoch) griechisch essen gehen, jedoch nicht so weit entfernt von Ihrem Heim. Sie rufen BoRIS an, um zu herauszufinden, welches griechische Restaurant sich in der Nähe Ihres Hauses befindet. Sollte BoRIS kein passendes Restaurant auf Ihrer Anfrage ausfindig machen, versuchen Sie es erneut, diesmal in einem anderen Stadtteil.

Restaurant:

Adresse:

Restaurant:

Adresse:

Kommentar:

Szenario 4

Es ist Mittagszeit und Sie haben großen Hunger. Sie möchten egal was essen, schnell und möglichst preiswert. Der Stadtteil spielt für Ihre Präferenz keine allzu große Rolle, denn Sie besitzen ein Auto. Die Art der Küche ist Ihnen auch nicht besonders wichtig, Hauptsache es soll lecker sein. Sollte BoRIS doch nichts für Sie finden, sind Sie bereit einige Mark darauf zulegen.

Restaurant:

Adresse:

Restaurant:

Adresse:

Kommentar:

Szenario 5

Ihr Freund Martin aus Düsseldorf hat eine große Vorliebe für Südamerika. Er kommt heute Abend nach Bochum und Sie möchten ihm eine Freude machen und ihn zu einem Cocktail in einer lateinamerikanischen Kneipe einladen. Das Problem ist, daß Sie gar nicht wissen , ob es in Bochum überhaupt eine solche Kneipe gibt. Die Lösung wäre bei Boris anzurufen, um sich informieren zu lassen.

Restaurant:

Adresse:

Restaurant:

Adresse:

Kommentar:

Szenario 6

Ein Bekannter hat Ihnen über ein sehr feines spanisches Restaurant namens "Coco Loco" erzählt. Sie möchten unbedingt dieses Restaurant ausprobieren, egal was es kosten sollte und Sie brauchen dafür die Adresse. Ihnen steht nur eine einzige Auskunftsmöglichkeit zur Verfügung, nämlich BoRIS. Führen Sie das Gespräch durch und notieren Sie dabei die Adresse des Restaurants.

Restaurant:

Adresse:

Restaurant:

Adresse:

Kommentar:

Anhang B

Der qualitative Fragebogen

Fragebogensnr.

Datum:

Qualitatives Interview 1.

PERSONALIA

Geschlecht: weiblich männlich
Alter:
Beruf /Ausbildung:
Computerkenntnisse:
Herkunftsregion /-stadt:
Jetziger Wohnort:

EINSTELLUNGEN ZUM RESTAURANT- SUCHE

1.0 *Wie erkundigen Sie sich nach einem Restaurant , wenn Sie in einem fremden Ort essen gehen möchten?*
2.0 *Was ist für Sie wichtig, wenn Sie sich für ein Restaurant entscheiden?*

EINSTELLUNGEN ZUM DIALOGSYSTEM

3.0 *Wissen Sie, was ein automatisches Sprachauskunftssystem ist?*
3.1 *Wenn ja, was verstehen Sie darunter (näher beschreiben)?*
3.1.1 *Haben Sie ein solches System schon benutzt?*
3.1.2 *Wie oft?*
3.1.2 *Zu welchem Zweck?*
3.2 *Wenn nein-[Erklärung]-, können Sie sich vorstellen mit einem solchen System umzugehen?*
4.0 *Stellen Sie sich vor, Sie rufen bei einem automatischen Restaurantauskunftssystem an und Sie werden von einem Sprachcomputer bedient.*
4.1 *Wie glauben Sie, dass der Sprachcomputer reagieren wird?*
4.2 *Wie stellen Sie sich den Gesprächsablauf [Frage-Antwortspiel] mit diesem Sprachcomputer vor?*
4.3 *Wie stellen Sie sich das Umgehen mit einem solchen Sprachcomputers vor?*
4.4 *Welche Informationen soll ein Restaurantauskunftssystem anbieten?*
4.5 *Welche Servicemöglichkeiten soll ein Restaurantauskunftssystem anbieten?*
4.6 *Wie soll Sie das System bedienen, damit Sie zufrieden sind?*
5.0 *Welche Vorteile / Nachteile sehen Sie bei einem Computer-Auskunftssystem?*

Qualitatives Interview 2

ERWARTUNGEN	<p>1.0 Nachdem Sie den Test mit BORIS gemacht haben, können Sie mir sagen, ob das System Ihren Vorstellungen entsprach? 1.1. Wenn nein, warum nicht?</p>
PAUSCHALE BEURTEILUNG	<p>2.0 Können Sie zusammenfassend sagen, was Sie am BoRIS gut gefunden haben? 3.0 Was hat Sie am BoRIS gestört?</p>
<p>POSITIV</p> <p>NEGATIV</p>	<p>4.0 Hat Ihnen das Gespräch mit BoRIS Probleme bereitet ? 4.1 Wenn ja, nennen Sie bitte genau die Stellen, wo Sie im Laufe des Gesprächs Schwierigkeiten bemerkt haben.</p>
DETAILLIERTE BEURTEILUNG	<p>5.0 War Ihnen von Anfang an klar, welche Informationen Sie vom bekommen könnten und welche nicht ? 5.1 Wussten Sie schon nach dem ersten Kontakt mit BoRIS, nach welchen Kriterien Sie ein Restaurant aussuchen sollten?</p>
TRANSPARENZ	<p>6.0 Hatten Sie das Gefühl, dass das System Ihren Informationswunsch richtig verstanden hat? 7.0 Hat das System, Ihrer Meinung nach deutlich gesagt, was es von Ihnen wissen wollte?</p>
VERSTÄNDLICHKEIT	<p>8.0 Wie würden Sie den Gesprächsaufbau mit BORIS charakterisieren? 9.0 Hatten Sie den Eindruck, dass der Gesprächsfluss unterbrochen wurde, dass es mal zu „Pannen“ kam ? 9.1 Wenn ja wie häufig? 9.2 Haben Sie diese „Pannen“ gestört?</p>
GESPRÄCH-DARSTELLUNG	<p>10.0 [Erinnern Sie sich noch], wie Sie die Fragen des Computers beantwortet haben? 11.0 Haben Sie allgemein die gewünschten Informationen bekommen? 11.1 Wenn ja, wie schnell ging es ? 11.2 Wenn nein oder teilweise, erinnern Sie sich noch welche Information fehlte?</p>
ZUFRIEDENHEIT	<p>12.0 Wie würden Sie den Informationsgehalt beurteilen? 12.1 Waren Sie damit zufrieden?</p>

**VERTRAUENS-
WÜRDIGKEIT**

13.0 *Glauben Sie, dass BoRIS Ihnen richtige Informationen geliefert hat?*

BESTÄTIGUNG

14.0 *Bei manchen Computersystemen ist eine Bestätigung der Eingaben vorgesehen. Dadurch kann man sicher stellen, ob der Computer die richtigen Informationswünsche bearbeiten wird. Hätten Sie sich auch so etwas gewünscht?*

ZUFRIEDENHEIT

15.0 *Wie haben Sie sich im Laufe des Gesprächs mit BoRIS gefühlt?*

15.1 *Wie haben Sie das Gespräch empfunden?*

15..2 *War es angenehm mit BoRIS zu telefonieren?*

16.0 [Trotz vorhandenen Schwierigkeiten] *Hat sich insgesamt Ihr Anruf bei BoRIS gelohnt?*

16.1 *Würden Sie die Dienste eines solchen Auskunftssystems noch mal im Anspruch nehmen?*

**ABSCHLIESSENDE
BEMERKUNGEN**

17.0 *Gibt es noch etwas, was Ihnen am BoRIS aufgefallen ist?*

Qualitative Interviews

Fragebogennr. : 1.1

Datum 23.03.2001

Qualitatives Interview 1 (vor dem Versuch)

Geschlecht: weiblich männlich

Alter: 26

Beruf /Ausbildung: Ingenieur

Computerkenntnisse: ja

Herkunftsregion /-stadt: Ägypten, Alexandria

Jetziger Wohnort: Bochum, seit 4 1/2 Monaten

1.0 *Wie erkundigen Sie sich nach einem Restaurant, wenn Sie in einem fremden Ort essen gehen möchten?*

Entweder über 's Internet, wenn ich den Zugriff über 's Internet habe oder über die Gelben Seiten.

2.0 *Was ist für Sie wichtig, wenn Sie sich für ein Restaurant entscheiden?*

Hm, wenn ich in einem fremden Land bin, für mich ist es wichtig, dass es nicht allzu weit weg von mir ist und dann entsprechend, dass die Öffnungszeiten passend sind

Auch andere Kriterien?

Ja, es gibt natürlich auch andere Kriterien, vom Geschmack her. Ich würde also nicht in ein indisches Restaurant, oder so was hingehen. Andere Kriterien gibt's nicht. Wenn mir z.B. jmd. sagt, dass ein besonderes Restaurant schlecht ist und ich ganz allein in der Stadt bin, dann sind dies die Kriterien. Die Art der Küche ist also noch ein Kriterium

3.0 *Wissen Sie, was ein automatisches Auskunftssystem ist?* **Ja.**

3.1 *Wenn ja, was verstehen Sie darunter (näher beschreiben)?*

Darunter verstehe ich dass z.B. genau wie beim Fahrkartenbestellung, dass ich einfach sage, was ich genau will, dass es mir die entsprechenden Daten dazu gibt.

3.1.1 *Haben Sie ein solches System schon benutzt?* **Per Internet ja, aber per Telefon nein.**

3.1.2 *Wie oft?* **Per Internet sehr oft, es ist kein Gespräch, man muss alles anklicken**

3.1.3 *Zu welchem Zweck?* **Fahrkarten reservieren, oder eine bestimmte Reise planen.**

3.2 *Wenn nein-[Erklärung]-, können Sie sich vorstellen mit einem solchen System umzugehen?*

4.0 *Stellen Sie sich vor, Sie rufen bei einem automatischen Restaurantauskunftssystem an und Sie werden von einem Sprachcomputer bedient.*

4.1 *Wie glauben Sie, dass der Sprachcomputer reagieren wird?* **Da bin ich gespannt, ich bin an keiner solchen Situation gewöhnt; hm, ich weiss es nicht also, ich hoffe, dass es**

nicht so dumm sein wird, ha, hah, dass es wirklich dann auch die Fragen beantwortet werden, dass ich nicht eine Frage stelle und dann eine Antwort bekomme, die überhaupt nicht damit zu tun hat oder eine zu allgemeine Antwort bekomme, mit der ich auch nichts anfangen kann.

Glaubst Du, dass Du merken wirst, dass du mit einem Computer sprichst?

Hmm, ja es kommt darauf an, wenn es mit einem natürlichen Sprecher ist, ich glaube, ich werde schon merken, es kommt darauf an, wie intelligent das System ist.

4.2 Wie stellen Sie sich den Gesprächsablauf [Frage-Antwortspiel] mit diesem Sprachcomputer vor?

Hm, ich rufe an, ich werde von jmd. begrüßt ja und wenn ich mit einem Computer bin, dann würde ich mich vorstellen, dass er die Fragen stellt, wie hm, weil es dann wahrscheinlich einfacher für ihn ist, wenn ich die Fragen stelle, dann kann es dazu kommen, dass ich Fragen stelle, die ein bisschen unklar ist oder er nicht richtig begreift aber wenn er die Fragen stellt, dann ist es einfacher.

Was würdest du bevorzugen?

Mir wären lieber Fragen zu beantworten. Wenn ich weiss, dass es ein Computer ist, wäre mir lieber, wenn ich Fragen beantworten würde.

4.3 Wie stellen Sie sich das Umgehen mit einem solchen Sprachcomputer vor?

Hmm, ich kann mir das schwer vorstellen, muss ich sagen ahmm es wird nicht kompliziert aber ich zweifle, ob, ja ein sinnvolles Gespräch sein wird, ob ich dann nicht wirklich in der Mitte auflegen werde und ich meine, ich rede jetzt bloss nur mit einem Computer, ich komme nicht weiter.

Glaubst Du, dass ein solcher Computer einfach zu bedienen ist?

Hm, ich glaube nicht sehr einfach zu bedienen sein wird, es ist doch nicht menschlich.

4.4 Welche Informationen soll ein Restaurantauskunftssystem anbieten?

Ja, zuerst muss eine vernünftige Aufteilung von Restaurants, nach bestimmten Kriterien anbieten, wo ich dann wählen kann, welches Kriterium am wichtigsten ist und dann, so stelle ich mir aufgebaut vor, so wie man bei einem Computer klickt man auf das aller wichtigste und dann kommt man in ein anderes Menu, wo da zweit wichtigste zu besprechen und man darf dann auch wählen usw., so stelle ich es mir vor; das erste Hauptmenu sind die wichtigsten Kriterien gesehen und dann wählt man sich ein Kriterium aus und dann kommt auf ein zweites Menu, drittes Menu usw. Also, alles in Form von Auswahl, er gibt dir verschiedene Auswahlmöglichkeiten.

Für mich wäre geeignet, wenn ich mir zuerst einen bestimmten Stadtteil aussuchen soll und dann in dieser Zeit die Art der Küche und dann entsprechend die Öffnungszeiten, Preise und Reservierungsmöglichkeiten wären auch nett.

4.5 Welche Servicemöglichkeiten soll ein Restaurantauskunftssystem anbieten?

24 Stunden pro Tag, auf jedem Tag, soll schnell sein, soll nicht irgendwie ein halbstündiges Gespräch, ich möchte es in 5 Minuten haben; natürlich ein deutliches, sinnvolles Gespräch und möglichst, am Ende möchte ich auch eine Reservierung.

4.6 *Wie soll Sie das System bedienen, damit Sie zufrieden sind?*

Schnell, sinnvoll, am Ende eine Reservierung, wenn es geht.

5.0 *Welche Vorteile / Nachteile sehen Sie bei einem Computer- Auskunftssystem?*

Hm Vorteile... oder vielleicht fangen wir lieber mit den Nachteilen. Wie ich schon vielleicht mehrmals gesagt habe, dass es vielleicht zu begrenzt sein wird, zu beschränkt, dass die Information..... dass Du jetzt einige Fragen stellen willst, die er nicht beantworten kann, das wäre schlecht. Vorteile wären dann, dass er mir die Zeit spart, selbst nachzugucken, er hat dann schon die ganzen Daten vorbereitet und wenn es ein gutes schnelle System ist, einfacher ist.

Qualitatives Interview 1 (nach dem Versuch)

6.0 *Nachdem Sie den Test mit BoRIS gemacht haben, können Sie mir sagen, ob das System Ihren Vorstellungen entsprach?*

Hmm, also, ich finde, dass in manchen Szenarien schon meiner Vorstellungen entsprachen, aber nicht in allen; es gab schon manche Situationen, wo es überhaupt nicht geklappt hat.

6.1 *Wenn nein, warum nicht?*

Was hat einfach nicht geklappt? Was entsprach nicht deiner Vorstellungen?

Die Art und Weise, wie die Adressen und Namen angeboten sind. Es ist sehr schlecht. Die Synthese ist nicht schlecht, ist sehr schlecht

7.0 *Können Sie zusammenfassend sagen, was Sie am BoRIS gut gefunden haben?*

Dass das Menu wirklich fast alles umfasst, was man vom einem Auskunftssystem erwartet, dass es relativ schnell die Daten verarbeitet, dass es auch ..., also bei mir ist es nur in 2 Fälle hängen geblieben, als er nicht verstanden hat, aber ansonsten hat er immer richtig reagiert.

8.0 *Was hat Sie am BoRIS gestört?*

Wie gesagt, diese Adressen- und Namendarbietung war total schlecht, so dass man nicht also ich ruf an, weil ich eine Auskunft über einen Namen und über eine Adresse haben will und da komme ich überhaupt nicht weiter, das stört total, erfüllt in diesem Sinne gar nicht die Aufgabe. Ansonsten gibt es einen Teil, wo er meine Aussprache nicht verstanden hat, obwohl ich ganz langsam und deutlich, meiner Meinung nach, gesagt habe.

9.0 *Hat Ihnen das Gespräch mit BoRIS Probleme bereitet?*

Ja, zum Teil schon.

9.1 *Wenn ja, nennen Sie bitte genau die Stellen, wo Sie im Laufe des Gesprächs Schwierigkeiten bemerkt haben.*

Auf einmal, als ich mich um preiswerte Restaurants erkundigen wollte, schien es irgendwo, dass es eine untere Grenze gab, unter der er nichts verstanden hat ich habe wirklich eine dumme Summe gesagt, von 4 DM, aber das muss er doch verstehen können und dann angeben, dass es nichts dazu gibt. Aber er hat immer wieder wiederholt. Aber das sollte nicht sein, er sollte sagen, es gibt nichts um diesen Preis, wollen Sie ihre Daten verändern und auf der anderen, was ich sehr schlecht fand, war im ersten Gespräch, da hatte ich mich zuerst nach dem Ort entschieden und das lief ganz gut und auf einmal wurde es unterbrochen und fing von ganz vorne an. Ich habe auch meine Datenbank geändert, aber das fand ich.....

10.0 *War Ihnen von Anfang an klar, welche Informationen Sie vom System bekommen könnten und welche nicht?*

Hm.....ja, aber.. Ja schon, ich habe ganz am Anfang erwartet, dass es irgendwelche Reservierungen gibt und gab's nicht. Das habe ich erwartet

10.1 *Wussten Sie schon nach dem ersten Kontakt mit BoRIS, nach welchen Kriterien Sie ein Restaurant aussuchen sollten?* **ja**

11.0 *Hatten Sie das Gefühl, dass das System Ihren Informationswunsch richtig verstanden hat?*

Ja, im größten Teil, nicht überall aber, wie gesagt, was das stört, ich bin in keiner von den sechs Szenarien eine Adresse oder einen Namen herausbekommen, bei allen war synthetisch so schlecht, dass das Gespräch gar nicht genutzt hat. Das stört.

12.0 *Hat das System, Ihrer Meinung nach deutlich gesagt, was es von Ihnen wissen wollte?*
Ja.

13.0 *Wie würden Sie den Gesprächsaufbau mit BoRIS charakterisieren?*

Vom Aufbau her war sehr sinnvoll, wie ich am Anfang schon gesagt habe, mir hat's gefallen, dass man auswählen soll von einem bestimmten Menu und dann in ein anderes Menu reingeführt wurde. Das fand ich sinnvoll. Es war auch übersichtlich. In einem Teil hat er auf das Hauptmenu wieder zurückgegriffen, obwohl dass es überhaupt nicht so sein sollte. Ich habe meine Auswahl angegeben und das hat er falsch verstanden und hat irgendwie von Anfang an angefangen.

14.0 *Hatten Sie den Eindruck, dass der Gesprächsfluss unterbrochen wurde, dass es mal zu "Pannen" kam?*

Ja

14.1 *Wenn, ja wie häufig?* **2 mal.**

14.1.1 *Haben Sie diese „Pannen“ gestört?* **Einmal nicht so sehr, aber bei dem zweiten schon, weil ich schon mitten im Gespräch war, bestimmte Kriterien gewählt hatte und er plötzlich von Anfang an begonnen und das nervt.**

15.0 *[Erinnern Sie sich noch], wie Sie die Fragen des Computers beantwortet haben?*

Ich habe am Anfang ganze Sätze benutzt. Nachher habe ich festgestellt, dass es so wie so gar keinen Unterschied gibt und habe ich nur Stichwörter benutzt. Ich habe beides ausprobiert um festzustellen, ob es dann schwieriger wird, wenn ich ganze Sätze benutze, aber es war kein Problem. Das war nicht, dass ich jetzt um die Arbeit für ihn zu vereinfachen, sondern für mich zu vereinfachen.

16.0 *Haben Sie allgemein die gewünschten Informationen bekommen?* **Nein!**

16.1 *Wenn ja, wie schnell ging es?*

16.2 *Wenn nein oder teilweise, erinnern Sie sich noch welche Information fehlte?*

Reservierung und eine deutlich Auskunft. Das war überhaupt nicht deutlich. Wenn ich wirklich in einer richtigen Situation wäre, wo ich die Adresse und den Namen des Restaurants will, dann habe ich das nicht bekommen.

17.0 *Wie würden Sie den Informationsgehalt beurteilen?* **Hm....weiß ich nicht...**

17.1 *Waren Sie damit zufrieden?* **Hm ...ja..**

18.0 *Glauben Sie, dass BoRIS Ihnen richtige Informationen geliefert hat?*

Es war ... also in zwei Fälle konnte ich überprüfen, wo die Namen des Restaurants da waren, da hat er richtige Informationen gegeben, von einem Havanna, bei dem

lateinamerikanischen Essen, da weiß ich schon das es richtig war, von anderen bei diesem Coco Loco, wo ich den Namen des Restaurants wusste, da hat er auch etwas ähnliches wie Coco Loco gesagt, aber ich leider die Adresse nicht mitbekomme. Ich glaube schon, dass die Information richtig war, aber es war nicht deutlich

19.0 *Bei manchen Computersystemen ist eine Bestätigung der Eingaben vorgesehen. Dadurch kann man sicher stellen, ob der Computer die richtigen Informationswünsche bearbeiten wird. Hätten Sie sich auch so etwas gewünscht?*

Nein!

20.0 *Wie haben Sie sich im Laufe des Gesprächs mit BoRIS gefühlt?*

Besser als erwartet. Ich habe die Stimme synthetisch schlechter erwartet und dass meine Daten langsamer verarbeitet werden, dass er mehr nicht verstehen wird, obwohl ich nicht nur Stichwörter benutzt habe, sondern ganze Sätze.

20.1 *Wie haben Sie das Gespräch empfunden?* **Angenehm.**

20.1.1 *War es angenehm mit BoRIS zu telefonieren?*

21.0 [Trotz vorhandenen Schwierigkeiten] *Hat sich insgesamt Ihr Anruf bei BoRIS gelohnt?*

Soll ich ehrlich sein? Hhhh, wenn ich die Fragestellung oder die Aufgaben im Hinterkopf habe, hat es sich nicht gelohnt, das ganze Gespräch ist sehr gut, nur die wichtige Information ist nicht deutlich.

21.1 *Würden Sie die Dienste eines solchen Auskunftssystems noch mal im Anspruch nehmen?*

Nur wenn es synthetisch deutlich ist, wenn es so weiter ist, dann nicht, nein!

22.0 *Gibt es noch etwas, was Ihnen am BoRIS aufgefallen ist?*

Das Gespräch selbst ist es gut, es sind diese Pannen, die stören immer. Der Sprecher ist sehr gut. Ich hätte natürlich erwartet, dass es eine synthetisierte Sprache sein wird, wo man nichts verstehen wird, aber das ist schon gut. Der allgemeine Kommentar ist, dass es natürlich deutlicher sein soll, weil sonst keinen Sinn hat.

Hh .. Was ich hier noch gesagt habe, das war bei Szenario 5, als ich nach einer Kneipe gesucht habe, vielleicht sollte man ganz am Anfang auch, bevor man das Hauptmenu angibt, einfach zwischen Restaurants und Kneipen wählen. , z.B. zuerst könnte man sagen, wollen Sie eine Auskunft über Restaurants oder Kneipen haben und danach nach Preis oder nach was ich fragen.

Adresse und Name habe ich nicht bekommen, oder bekommen doch, aber nicht verstanden.

Qualitatives Interview 2 (vor dem Versuch)

Geschlecht: weiblich männlich
Alter: 39
Beruf /Ausbildung: technische Rechnerin
Computerkenntnisse: ja
Herkunftsregion /-stadt: NRW
Jetziger Wohnort: Bochum

1.0 *Wie erkundigen Sie sich nach einem Restaurant, wenn Sie in einem fremden Ort essen gehen möchten?*

Ich gucke mir die Stadt an, was mir gefällt.

Andere Informationsmittel nehmen Sie nicht Anspruch?

Nein, bis jetzt habe ich noch nie gemacht.

2.0 *Was ist für Sie wichtig, wenn Sie sich für ein Restaurant entscheiden?*

Ja... Ob, z.B. mediteranee Küche gibt, weil ich diese Küche am liebsten mag.

Was anderes?

Also, ich mag nicht Oberklasse oder Mittelklasse Restaurants, sondern normale...

Was anderes? Was ist für Sie ausschlaggebend, wenn Sie sich für ein Restaurant entscheiden?

Ja. Die Einrichtung, ob sie mich dann anspricht... Ambiente

3.0 *Wissen Sie, was ein automatisches Sprachauskunftssystem ist?*

Eine Art Call Center?

Nein, nicht so ganz!! Ein Sprachauskunftssystem ist ein Dialogsystem, das, in unserem Fall, von einem Sprachcomputer gesteuert wird. Der Dialog läuft, mehr oder weniger, wie in einer normalen Unterhaltung. Man muß keine Tastenkombinationen eingeben, alles läuft natürlich durch Spracheingabe.

3.1 *Wenn ja, was verstehen Sie darunter (näher beschreiben)? -*

3.1.1 *Haben Sie ein solches System schon benutzt? -*

3.1.2 *Wie oft?*

3.1.2 *Zu welchem Zweck?*

.2 Wenn nein-[Erklärung]-, können Sie sich vorstellen mit einem solchen System umzugehen? Ja.... ich könnte mir schon vorstellen, aber alles ist ziemlich abstrakt; wenn ich bei der Bahnauskunft eine Computerstimme höre und sagt, die und die Züge fahren dann und dann... dann unterhalte ich mich lieber persönlich.

4.0 *Stellen Sie sich vor, Sie rufen bei einem automatischen Restaurantauskunftssystem an und Sie werden von einem Sprachcomputer bedient.*

4.1 *Wie glauben Sie, dass der Sprachcomputer reagieren wird?*

Ich glaube, er wird mich bedienen, wie man jemanden bedient, der nach einem Restaurant fragt. Was man gerne ißt, wo man gerne hin geht, ob es Ambiente gibt, wo ich mich informieren kann.

Erwarten Sie die Reaktion eines Computers oder eines Menschen?

Ja.. eigentlich eines Computers, der den Menschen ersetzen will.

4.2 *Wie stellen Sie sich den Gesprächsablauf [Frage-Antwortspiel] mit diesem*

Sprachcomputer vor?

Ja, vielleicht stellt er Fragen und wenn er keine Antwort bekommt, dann geht er weiter... Ich stelle es mir wie ein Wechselspiel vor, wie bei einer normalen Unterhaltung zwischen Menschen. Es ist doch ein Dialogsystem, sonst wäre es ein bisschen einseitig.

4.3 Wie stellen Sie sich das Umgehen mit einem solchen Sprachcomputer vor?

Nein, glaube ich nicht. Es wird nicht sehr kompliziert sein.

4.4 Welche Informationen soll ein Restaurantauskunftssystem anbieten?

Ich denke, Küche, Preisklasse, Ambiente, Ruf, was für ein Ruf das Restaurant hat. Kategorie.

4.5 Welche Servicemöglichkeiten soll ein Restaurantauskunftssystem anbieten?

Ja, Tag- und Nachtservice immer zur Verfügung stehen, egal wo man sich befindet.

4.6 Wie soll Sie das System bedienen, damit Sie zufrieden sind?

Soll vielleicht nicht abstürzen, es wäre unangenehm.

5.0 Welche Vorteile / Nachteile sehen Sie bei einem Computer- Auskunftssystem?

Nachteil Hm, ich weiß nicht, ob das zum Nachteil sein sollte, ich würde lieber Sie fragen nach einem Restaurant, ob Sie was kennen würden, als den Computer. Aber Nachteile kenne ich keine...

Welche Vorteile können solche Computer haben?

Ja, dass, wenn man in einer Stadt gerade fremd ist, dass man eine direkte Informationsquelle wählt, ohne sich umzuhören zu müssen und sich selbst einzubringen... kann ich ja hier gut essen.

Qualitatives Interview 2 (nach dem Versuch)

6.0 Nachdem Sie den Test mit BoRIS gemacht haben, können Sie mir sagen, ob das System Ihren Vorstellungen entsprach?

Mal ging er auf mich ein, mal nicht. Zum Schluß bei dem vorletzten. Wenn man z.B. zu... hm ich hatte das Gefühl, dass man zum Ausschleifen wird, konnte er vielleicht die Informationen nicht sofort irgendwie verarbeiten und dann hackt das System ein bißchen aus und dann wird das alles ungenau. Oder er geht nicht mehr auf das ein, was man fragt, hatte ich bei dem vorletzten gefühlt.

6.1 Wenn nein, warum nicht?

7.0 Können Sie zusammenfassend sagen, was Sie an BoRIS gut gefunden haben?

Dass es überhaupt diese Möglichkeit gibt, diese Art von Information zu bekommen, wie bei BoRIS. Dass es recht sympathisch aufgezoogen ist, es soll natürlich nicht abschreckend wirken.. und dass die 4 Hauptkategorien - Preis, Ambiente, Tag, Öffnungszeiten, alle bietet er an.

8.0 Was hat Sie am BoRIS gestört?

Die Adressenansagen fand ich unsympathisch. Schon beim ersten mal, als glaube ich, nach dem Araber fragte, sollte ein bißchen arabisch klingeln... es ist eine andere Stimme, die die Adressen durchgibt.

Ja, es ist eine Computerstimme.

Ja, fand ich nicht so schön.

Und was anderes?

Hmmm... ja, dass man genau wissen muß, was man für Fragen stellt und nicht davon ausgeht, also man darf nicht so sehr ausschweifen und sich nicht vorstellen, dass z.B. uns gegenüber sitzen und ich Sie jetzt frage, wo kann ich essen gehen. Man merkt richtig schnell, dass man gezielt was fragen muß, ansonsten klappt das nicht.

Und was anderes?

Was anderes... wüßte ich nicht.

9.0 Hat Ihnen das Gespräch mit BoRIS Probleme bereitet?

9.1 Wenn ja, nennen Sie bitte genau die Stellen, wo Sie im Laufe des Gesprächs

Schwierigkeiten bemerkt haben.

Bei dem 4. Szenario habe ich keine Information erhalten, wobei ich gesagt habe, dass ich ein Auto besitze und ein paar Mark drauflegen würde. Er konnte damit nichts anfangen.

Bei Szenario 5 war halt auch der Info zur Kneipe... hat auch weg fallen lassen... also im Grunde nur Havanna, Kreuzstrasse, das südamerikanische Restaurant und da war basta, fini...

Bei Coco Loco hat er erst beim zweiten mal geschaltet, da war alles dazwischen, außer spanisch, obwohl ich nach spanisch gefragt hatte. War so, ne?

Und am Anfang, was man das erste mal hört, dann ist es sehr schnell was man an Kategorien für..., na ja wenn man vorbereitet ist, man denkt, ja ich möchte jetzt den Preis wissen, Ambiente, wo, kann in der Nähe sein, ich habe ein Auto, wenn man sich das im Vorfeld überlegt und darauf vorbereitet ist, kommt man gut damit zurecht. Am Ende ist es ein bißchen verwirrend, die Kategorien schmeißt er sehr schnell raus.

10.0 *War Ihnen von Anfang an klar, welche Informationen Sie vom System bekommen könnten und welche nicht?*

Welche Art von Information kann Ihnen BoRIS liefern?

Eigentlich schon.

10.1 *Wussten Sie schon nach dem ersten Kontakt mit BoRIS, nach welchen Kriterien Sie ein Restaurant aussuchen sollten?*

Ja.

11.0 *Hatten Sie das Gefühl, dass das System Ihren Informationswunsch richtig verstanden hat?*

Die ersten Male ja, die letzten nicht. Warum auch immer.

12.0 *Hat das System, Ihrer Meinung nach deutlich gesagt, was es von Ihnen wissen wollte?*

Ja.

13.0 *Wie würden Sie den Gesprächsaufbau mit BORIS charakterisieren?*

Am Anfang war es mehr Frage Antwort und hinterher, wenn man denkt, wenn man sich anvertraut hat, stellt man vielleicht mehr Fragen, ob dann das System begreift, weiß ich nicht..... aber das habe ich zum Schluß war ja im Grunde so war so ein bißchen verschachtelt hat und dann wenn man nach mehreren Nationalitäten gefragt und hat plötzlich die eine ausgelassen, nach dem er gefragt hat. Das habe ich nicht so richtig verstanden.

Wie würden sie den Gesprächsaufbau beschreiben?

Noch recht einfach, vielleicht kann ein bißchen mehr strukturieren.

14.0 *Hatten Sie den Eindruck, dass der Gesprächsfluss unterbrochen wurde, dass es mal zu "Pannen" kam?*

Nein. Obwohl eine Panne war ja da, bei dem letzten, als er die verschiedenen Nationalitäten, da war kurz Einbruch drin. Einmal, als ich nach einem spanischen Restaurant gefragt hatte und hat er mir alle Nationalitäten ausgeworfen.

14.1 *Wenn, ja wie häufig?*

14.1.1 *Haben Sie diese „Pannen“ gestört?*

15.0 *[Erinnern Sie sich noch], wie Sie die Fragen des Computers beantwortet haben?*

Ah... es geht, eigentlich ganz normal. Weil die Stimme eigentlich keine Computerstimme ist. Also, nicht stichwortartig

16.0 *Haben Sie allgemein die gewünschten Informationen bekommen?*

Ja

16.1 *Wenn ja, wie schnell ging es ?*

Recht zügig.

16.2 *Wenn nein oder teilweise, erinnern Sie sich noch welche Information fehlte?*

17.0 *Wie würden Sie den Informationsgehalt beurteilen?*

Ausreichend, im Sinne von genug, es reichte aus.

17.1 *Waren Sie damit zufrieden? ja*

18.0 *Glauben Sie, dass BoRIS Ihnen richtige Informationen geliefert hat?*

Ja, bin ich ganz blöd naiv... Wobei eigentlich bei Coco Loco, das kenne ich nämlich, hat er die richtige Straße angegeben, aber man konnte sehr schlecht verstehen. Die Adressenauskunft waren ein bißchen undeutlich, alle Fragen waren deutlich nur bei den Adressen muss man die Ohren doppelt ausspannen.

19.0 *Bei manchen Computersystemen ist eine Bestätigung der Eingaben vorgesehen. Dadurch kann man sicher stellen, ob der Computer die richtigen Informationswünsche bearbeiten wird. Hätten Sie sich auch so etwas gewünscht?*

Stimmt, das verunsichert einen vielleicht erst; obwohl ich ihm, so nach dem Informationsquelle, die er abgegeben hat war O.K. reicht aus.

Nein, ich hätte mir so was nicht gewünscht.

20.0 *Wie haben Sie sich im Laufe des Gesprächs mit BoRIS gefühlt?*

Gut.

20.1 *Wie haben Sie das Gespräch empfunden? **Angenehm***

20.1.1 *War es angenehm mit BoRIS zu telefonieren? **Ja.***

21.0 [Trotz vorhandenen Schwierigkeiten] *Hat sich insgesamt Ihr Anruf bei BORIS gelohnt?*

Ja.

21.1 *Würden Sie die Dienste eines solchen Auskunftssystems noch mal im Anspruch nehmen? **Ja, ich würde das schon machen.***

22.0 *Gibt es noch etwas, was Ihnen am BoRIS aufgefallen ist?*

Hmmm... ja, wie gesagt, erstmals beim zweiten Anlauf erst irgendwie Informationen gibt, schlecht zu verstehen war die Adressenvermittlung. Ansonsten wüßte ich nichts.

Qualitatives Interview 3 (vor dem Versuch)

Geschlecht: weiblich männlich

Alter: 28

Beruf /Ausbildung: Ingenieur

Computerkenntnisse: ja

Herkunftsregion /-stadt: Bonn

Jetziger Wohnort: Bochum

1.0 *Wie erkundigen Sie sich nach einem Restaurant, wenn Sie in einem fremden Ort essen gehen möchten?*

Wenn ich wirklich da niemanden kenne, ich hatte eigentlich noch nie so ein Problem, also dann..... gehe ich ins Zentrum und ich sehe zu, dass ich irgendwas finde; ich könnte auch Gelbe Seite mal aufschlagen. Obwohl in Gelbe Seiten suche ich nie nach... ich würde einfach Leute fragen, die herumlaufen.

2.0 *Was ist für Sie wichtig, wenn Sie sich für ein Restaurant entscheiden?*

Das Essen!!!! Normalerweise geht's bei mir um die Art der Küche, Italiener, Grieche, irgend etwas sucht man sich vorher aus, irgendwie. Qualität... ich meine, ich bin derjenige, der das Luxusrestaurant sucht, aber... klar soll nicht schlecht sein; Preis wäre noch, Erreichbarkeit.

Was anderes?

Hmmm... ich möchte ein Restaurant, wo man nicht stundenlang warten muß, also...

3.0 *Wissen Sie, was ein automatisches Sprachauskunftssystem ist?*

Ja

3.1 *Wenn ja, was verstehen Sie darunter (näher beschreiben)?*

Ich werde wahrscheinlich, irgendwo übers Telefon irgendwo anrufen und dann habe ich einen Spracherkenner auf der anderen Seite, der nach meinen Wünschen mich befragt und erfüllt dann ein Dialog mit mir.

3.1.1 *Haben Sie ein solches System schon benutzt?*

Nein, aber ich habe schon zugehört.

3.1.2 *Wie oft?*

3.1.2 *Zu welchem Zweck?*

3.2 *Wenn nein-[Erklärung]-, können Sie sich vorstellen mit einem solchen System umzugehen?*

Also, ich habe in Griechenland Leute gesehen, die mit einem solchen System gearbeitet haben. Es war eine Art Notruf, die man wählen konnte. Ich fand es nervig, weil es nie verstanden hat. Also, ich kann mir vorstellen, dass es schlecht wird?????????

4.0 *Stellen Sie sich vor, Sie rufen bei einem automatischen Restaurantauskunftssystem an und Sie werden von einem Sprachcomputer bedient.*

4.1 *Wie glauben Sie, dass der Sprachcomputer reagieren wird?*

Er wird erstmals blöd hallo sagen!!! Nachher hat er zwei Möglichkeiten, also mich reden lassen, was ich will, oder aber er fragt, was ich suchen will, wo wollen Sie hin,

was für ein Restaurant; oder was ich für Wünsche habe und dann sagt man einen Text, wo er sich die Wörter aussucht und selber zuordnet. Die beiden Möglichkeiten stelle ich mir vor, wo wahrscheinlich die eine mit den einzelnen Abfragen, die üblicher sein könnte.

Nicht, dass Du selber Fragen stellen kannst?

Also,... das ist ein bißchen blöd, weil normalerweise eine Person würde ich durchaus fragen, wenn sie da sitzen würde, aber ich würde nicht davon ausgehen, dass die Maschine mir Tipps geben wird;

4.2 Wie stellen Sie sich den Gesprächsablauf [Frage-Antwortspiel] mit diesem Sprachcomputer vor?

4.3 Wie stellen Sie sich das Umgehen mit einem solchen Sprachcomputer vor?

Es wird komplizierter als mit einer Person zu reden. Weil viel Information ausgetauscht wird

Welche Stimme erwartest Du denn?

Ich gehe davon aus, dass er eine Sprachsynthese benutzt, dann eine synthetische Sprache; Also, entweder einzelne Worte gesprochen, die zusammengefügt zu Sätzen werden. Oder es kommt darauf an, eine menschliche Stimme, bei der man vielleicht einzelne Wörter geschnitten hat, um einen ganzen Satz zu bilden.

Welche Erwartungen hast Du bezüglich der Qualität?

Es sollte qualitativ einiger Maße sein. Ich würde keine synthetische Stimme unbedingt erwarten sondern dass Intonation und so was ein bißchen Sprecher ähnlich ist.

4.4 Welche Informationen soll ein Restaurantauskunftssystem anbieten?

Adresse, Telefonnummer, nicht schlecht wäre, wie man dahin ankommt, Wegbeschreibung; Reservieren wäre auch nett; Alternativen: es gibt zwei Situationen, die man sich hier vorstellen kann: entweder weiß ich genau was ich will, dann suche ich genau das; oder ich weiß nicht genau was ich will und gucke was es in der Nähe gibt. Eine Preisliste würde ich nicht erwarten, ich sitz eine Stunde da, bis er die Preise durchgibt.

4.5 Welche Servicemöglichkeiten soll ein Restaurantauskunftssystem anbieten?

Es wäre schön, wenn das System mir etwas ausdrucken könnte, dann brauche ich nicht alles aufzuschreiben. Ich gehe davon aus, dass das System 24 h pro Tag erreichbar ist, dass es nicht überlastet sein wird, dass ich nicht lange warten muß, und dass ich nicht eine Stunde brauche um das Gespräch durchzuführen.

4.6 Wie soll Sie das System bedienen, damit Sie zufrieden sind?

5.0 Welche Vorteile / Nachteile sehen Sie bei einem Computer- Auskunftssystem?

Vorteile 24h Erreichbarkeit, Preis, billiger als andere Auskunftssysteme, Tempo.

Nachteile dass der Computer nicht ausreichende Information gibt, Unverständlichkeit, Kommunikationsprobleme.

Qualitatives Interview 3 (nach dem Versuch)

6.0 *Nachdem Sie den Test mit BoRIS gemacht haben, können Sie mir sagen, ob das System Ihren Vorstellungen entsprach?*

Ja. Teilweise war besser als ich dachte. Aber ich hatte sehr geringe Ansprüche ehrlich gesagt. Obwohl ich manche Sachen nicht überprüfen kann. Ich habe nach der Uhrzeit gefragt ich habe „gleich“ „jetzt“ oder „heute“ oder so was, er hat irgendwas erkannt. Was er genau erkannt hat, weiß ich natürlich nicht.

6.1 *Wenn nein, warum nicht?*

7.0 *Was haben Sie Ihnen am BoRIS gut gefunden?*

Die Menüführung war gut, es ist ein bißchen widersprüchlich, am Anfang kommen so viele Möglichkeiten, man kann sich nicht merken, und wenn man ungefähr sagt was man will, kennt er das nicht. Das war am Anfang, mit dem Restauranttyp, glaube ich, da habe ich gesagt die Küche oder so was gesagt, hat er aber nicht verstanden...

Art des Restaurants hast Du gesagt ?

Ja stimmt, Art des Restaurants, genau, weil ich nicht mehr wußte, welche Worte er genau gebraucht hat... und dann wenn mal das hat dann ist die Menüführung eigentlich relativ gut erklärt, wenn er die Möglichkeit bietet, die man möchte, wenn er unterstützt, was man eigentlich will, dann war die Führung gut. Kann man aber nicht erreichen, ohne weiteres, versucht man auf andere Wege und dann ist umständlich.

Die Stimme war sehr ansprechend, sehr freundlich, ist gut, man fühlt sich schon bedient.

Er hat mir schon was ausgegeben ich aber leider nichts verstanden.

Bei dem 5 hatte ich kein Erfolg gehabt. Nader hat es doch irgendwie geklappt. Ich fand das gut, dass er nicht von vorne herein angefangen hat, sondern gefragt hat, was ich verändern möchte. Das fand ich gut.

Wenn ich den Weg gefunden hab, wenn ich so wußte, das hat gepaßt, war ja schon gut, weil man dann, war so gut wie ein Sprecher, habe ich eigentliche nichts vermißt.

8.0 *Was hat Sie am BoRIS gestört?*

Erstmals, wenn man mit BoRIS telefoniert dann gibt er sieben sechs oder Möglichkeiten, was man machen kann, was man aussuchen darf, dann hat man die ersten Möglichkeiten schon vergessen, man hört zu und will wissen, wenn man was falsches sagt, ich glaube ich habe Art der Küche gesagt oder irgendwie so was, dann hat er mir keine Möglichkeit gegeben, hat er hingenommen, und nochmals wiederholt, konnte ich darauf genauer achten. Es sind einfach zu viele Informationen am Anfang und das hat mich gestört.

Es wäre nicht schlecht, wenn es ein Kommando gäbe, wie z.B. "Bitte wiederholen!"

Dann hat mich sehr gestört, das trifft auf alles zu, dass man keine wiederholende Möglichkeit der Auskunft hat, er gibt die Auskunft und legt auf sagt "Tschüs!"

Dann die Sprachsynthese also die Antwort selber ist sehr schlecht.

9.0 *Hat Ihnen das Gespräch mit BoRIS Probleme bereitet?*

9.1 *Wenn ja, nennen Sie bitte genau die Stellen, wo Sie im Laufe des Gesprächs Schwierigkeiten bemerkt haben.*

Bei 1. Und 2 hatte ich keine Probleme, nur Innenstadt gab's nicht. Genau das war's. Ich hatte keine Chance herauszufinden, was in der Nähe von Innenstadt gab's. Es ging ansonsten. Es war blöder Weise das gleiche als ich mir am Anfang ausgesucht habe. Bis die Antwort kam hat irrelang gedauert. Ich dachte er hat schon vergessen. Er machte so lange Pausen.

Bei der dritten kam ich nicht weiter. Arabisch gab's nicht. Ich hätte fragen sollen, nach arabisch ja, ich habe aber nicht gemacht, weil er mir das nicht angeboten hat und hat er mir auch kein "zum Beispiel" gegeben, deswegen habe ich das nicht so richtig gefunden. Es war eine Irreführung, dass er mir nichts aus der Nähe des Restaurants angeboten, wo ich mir was anderes aussuchen könnte. Ich habe mich naher für griechische Küche entschieden. Ich wollte wissen, ob das in der Nähe ist. Er hat mir irgendwelche Adresse gegeben. Konnte ich, also nichts in der Nähe kriegen, hat er nichts gefunden

Bei der 4. Typisch deutsche Küche, war ich erfolgreich, obwohl ich sagen muß, hätte ich "Summa Cum Laudae" nicht gekannt, hätte ich es nicht verstanden

Bei Szenario 5 war einfach Schwachsinn da anzurufen, habe ich von Anfang an bemerkt, dass ich das falsche System angerufen habe,... weil es keine die Auskunft geben kann und deswegen hätte ich aufgelegt und ich dann das Gespräch zu Ende geführt, weil ich irgendwie das enden... alles was egal, was ich da gemacht habe, ich hätte einfach aufgelegt, das habe ich auch aufgeschrieben. Keine Chance gehabt...

Und 6 hmm..., ich habe zwei Wege versucht, die mir so aufgefallen wären, wie ich das versuchen könnte, beide haben nicht zum Ziel geführt, dann hat er angeboten, dass er... aus Weitmar tausend Restaurants anbieten würde und... dann habe ich es en defekt doch gefunden, aber es gab viele und zufällig ich hab's... normalerweise hätte ich nicht mehr gefragt ich hätte aufgelegt, hätte aufgegeben aber zufällig hat er beim letzten Mal, als ich mitgemacht, gesagt... Das war also die Grenze, die ich so mitgemacht hätte.

10.0 *War Ihnen von Anfang an klar, welche Informationen Sie vom System bekommen könnten und welche nicht? Ich hätte gehofft, dass ich er mir eine Telefonnummer gäbe. Damit ich bei dem Restaurant anrufen kann, und dort weiteres zu klären. Mir war es klar, dass er mir die Telefonnummer anbietet. Es kam nicht vor, es kam nur die Adresse vor. Was ich natürlich erwartet habe, war Adresse, Name, Telefonnummer*

10.1 *Wussten Sie schon nach dem ersten Kontakt mit BoRIS, nach welchen Kriterien Sie ein Restaurant aussuchen sollten?*

Ja, schon...

11.0 *Hatten Sie das Gefühl, dass das System Ihren Informationswunsch richtig verstanden hat?*

Ich kann manche Sachen nicht überprüfen: den Ort, die Uhrzeit, Tag. Er hat etwas mit der Information gemacht, aber was er gemacht weiß ich es nicht. Was ich überprüfe, ja.. Das hat er, weil ich es nichts falsches eingegeben habe.. Ich hab nicht mit arabisch probiert oder irgendwas.

12.0 *Hat das System, Ihrer Meinung nach deutlich gesagt, was es von Ihnen wissen wollte?*
Ja..

13.0 *Wie würden Sie den Gesprächsaufbau mit BoRIS charakterisieren?*

Der Aufgabe angesprochen... angepaßt.. Ich glaube, es war gut... irgendwie begrenzt.. wenn man was ausgefallenes wollte...

Es ging relativ schnell... Es war unkompliziert...

14.0 *Hatten Sie den Eindruck, dass der Gesprächsfluss unterbrochen wurde, dass es mal zu „Pannen“ kam?*

Nur vor den Antworten.

14.1 *Wenn, ja wie häufig? Am Anfang war echt lang, nachher wußte ich, dass es irgendwas kam. Erstens habe ich gedacht, das Gerät würde nicht mehr antworten.. Das erste Mal weiß man das und reagiert entsprechend..*

14.1.1 *Haben Sie diese „Pannen“ gestört?*

Beim ersten Mal Ja.. und dann nicht mehr..

15.0 *[Erinnern Sie sich noch], wie Sie die Fragen des Computers beantwortet haben?*

Stichwortartig.. weil ich nur die wichtigsten Dinge erzählen wollte... Ich weiß, dass eine Spracherkennung nicht gerade zugetexten werden will.

16.0 *Haben Sie allgemein die gewünschten Informationen bekommen? Zum Teil.*

16.1 *Wenn ja, wie schnell ging es? Wenn es eine Antwort gab ja. Wenn es keine Antwort gab, dann habe ich lange rumgestochen und nachher aufgegeben.*

16.2 *Wenn nein oder teilweise, erinnern Sie sich noch welche Information fehlte?*

Telefonnummer. Und ich hätte gerne weiter verbunden werden. Bei 5 und 6 hatte ich Probleme, wobei bei 6, was ich gut fand, war, dass er andere Möglichkeiten angeben hat.

17.0 *Wie würden Sie den Informationsgehalt beurteilen? Momentan stark begrenzt*

17.1 *Waren Sie damit zufrieden? Nein*

18.0 *Glauben Sie, dass BoRIS Ihnen richtige Informationen geliefert hat?*

Wenn er mir was gegeben hat, glaube ich. Ja.

19.0 *Bei manchen Computersystemen ist eine Bestätigung der Eingaben vorgesehen. Dadurch kann man sicher stellen, ob der Computer die richtigen Informationswünsche bearbeiten wird. Hätten Sie sich auch so etwas gewünscht?*

Zum zu sagen, z.B. "Sie haben folgendes eingegeben, daraus folgen die und die Daten". Das wäre gut zum Schluß aber nicht während dessen. Nicht alle Antworten bestätigen, sondern nur zum Schluß, was es ausgesucht wurde. Das hätte ich mir gewünscht, also nicht im Laufe des Gesprächs. Es wäre zu überprüfen, deswegen hier ist es nicht ein 100% tiges Ja, sondern nur teilweise.

20.0 *Wie haben Sie sich im Laufe des Gesprächs mit BoRIS gefühlt?*

Während die Szenarien einiger Maßen die Anforderungen erwarten konnte aus dieser Menüführung da eigentlich schon war ganz gut aufgehoben, wenn diese komplizierten Fälle, ich glaube 5 und 6, da hätte ich aufgelegt. Das hat mich genervt.

20.1 *Wie haben Sie das Gespräch empfunden? Bei 5 als lästig, das war mir eigentlich von Anfang an klar...*

20.1.1 *War es angenehm mit BoRIS zu telefonieren? Doch, kann man schon sagen.*

21.0 *[Trotz vorhandenen Schwierigkeiten] Hat sich insgesamt Ihr Anruf bei BORIS gelohnt? Es ist doch, ich bin in 6 verschiedenen... Teilweise ja.*

21.1 *Würden Sie die Dienste eines solchen Auskunftssystems noch mal im Anspruch nehmen?*

Wenn ich weiß, was ich zu erwarten habe und denke, dass es erfüllen kann, ja.

22.0 Gibt es noch etwas, was Ihnen am BoRIS aufgefallen ist?

Weiß ich nicht, wüßte ich nicht.

Anhang C:

Quantitatives Interview

VP-Nr: _____

Datum: _____

Fragebogen 1

Persönliche Daten

Geschlecht: weiblich männlich

Alter: _____

Beruf / Ausbildung: _____

Herkunftsregion /-stadt: _____

Jetziger Wohnort: _____

1.0 Wie erkundigen Sie sich nach einem Restaurant, wenn Sie in einem fremden Ort essen gehen möchten (mehrfache Nennung möglich)?

- | | | | |
|---------------------|--------------------------|--|--------------------------|
| 1.1. Magazine | <input type="checkbox"/> | 1.6. Tip von Bekannten | <input type="checkbox"/> |
| 1.2 Werbebroschüren | <input type="checkbox"/> | 1.7. Anruf bei einem Sprach-Informationssystem | <input type="checkbox"/> |
| 1.3 Stadtführer | <input type="checkbox"/> | 1.8 Sonstiges _____ | <input type="checkbox"/> |
| 1.3 Stadtführer | <input type="checkbox"/> | | |
| 1.4 Gelbe Seiten | <input type="checkbox"/> | | |
| 1.5 Internet | <input type="checkbox"/> | | |

2.0 Was ist Ihnen wichtig, wenn Sie sich für ein Restaurant entscheiden (mehrfache Nennung möglich) ?

- | | | | |
|------------------------------------|--------------------------|----------------------------------|--------------------------|
| 2.1 Preis | <input type="checkbox"/> | 2.6 Äußeres Ambiente | <input type="checkbox"/> |
| 2.2 Art der Küche | <input type="checkbox"/> | 2.7 Öffnungszeiten | <input type="checkbox"/> |
| 2.3 Qualität des Essens | <input type="checkbox"/> | 2.8 Schnelligkeit der Bedienung | <input type="checkbox"/> |
| 2.4 Breitgefächertes Speiseangebot | <input type="checkbox"/> | 2.9 Freundlichkeit des Personals | <input type="checkbox"/> |
| 2.5 Ort des Restaurants | <input type="checkbox"/> | 2.10 Sonstiges: _____ | <input type="checkbox"/> |

3.0 Haben Sie bereits ein automatisches sprachbasiertes Auskunftssystem benutzt? ja

nein

3.1. wenn ja, bei welcher Gelegenheit? _____

3.1.1 Wie würden Sie Ihre Erfahrung charakterisieren?

ausgezeichnet schlecht

4.0 Welche Informationen sollte ein solches Restaurantauskunftssystem liefern können (mehrfache Nennung möglich) ?

- | | | |
|-------------------------------|--------------------------|--------------------------------|
| 4.1 Name, Adresse, Telefonnr. | <input type="checkbox"/> | 4.5 Spezielle Speiseangebote |
| <input type="checkbox"/> | | |
| 4.2 Öffnungszeiten | <input type="checkbox"/> | 4.6 Äußeres Ambiente |
| <input type="checkbox"/> | | |
| 4.3 Preise | <input type="checkbox"/> | 4.7 Reservierungsmöglichkeiten |
| <input type="checkbox"/> | | |
| 4.4 Art der Küche | <input type="checkbox"/> | 4.8 Wegbeschreibung |
| <input type="checkbox"/> | | |

5.0 Wenn Sie bei BoRIS anrufen, erwarten Sie:

5.1 ein vernünftiges Gespräch wie mit einem Menschen
trifft zu trifft nicht zu

5.2 Sie stellen dem Computer Fragen
trifft zu trifft nicht zu

5.3 Der Computer stellt Ihnen Fragen
trifft zu trifft nicht zu

5.4 Sie bekommen präzise Antworten
trifft zu trifft nicht zu

5.5 Es werden keine Missverständnisse entstehen
trifft zu trifft nicht zu

5.6 Sie werden schnell die gewünschte Information bekommen
trifft zu trifft nicht zu

5.7 Der Computer hat eine freundliche Stimme
trifft zu trifft nicht zu

5.8 Das System ist leicht zu bedienen.
trifft zu trifft nicht zu

6.0 Was ist Ihnen lieber?:

- | | |
|--|--------------------------|
| 6.1 Sie stellen Fragen an das System | <input type="checkbox"/> |
| 6.2 Sie beantworten die Fragen des Systems | <input type="checkbox"/> |
| 6.3 Gemischt | <input type="checkbox"/> |
| 6.4 Egal | <input type="checkbox"/> |

7.0 Finden Sie, dass solche Auskunftssysteme eine Alternative zu derzeitigen Informationssystemen darstellen?

- ja teilweise nein ich weiß es nicht

Fragebogen 2

I. Ihr Qualitätsurteil über BoRIS:

- 1.1 Wie würden Sie den gesamten Eindruck von BoRIS einschätzen?
(ankreuzen)



- 1.2 Wie würden Sie BoRIS einem Freund beschreiben?

-
-
-
-
-

- 1.3 Was würden Sie bei BoRIS verbessern?

-
-
-
-
-

II. Bitte beurteilen Sie die folgenden Aussagen über Ihre Erfahrung mit BoRIS:

- 2.1 BoRIS hat genau so funktioniert, wie ich es mir vorgestellt habe.
trifft zu trifft nicht zu
- 2.2 BoRIS bietet umfangreiche Informationen über Bochumer Restaurants.
trifft zu trifft nicht zu
- 2.3 Ich bekam immer die gewünschte Antwort auf meine Eingaben.
trifft zu trifft nicht zu
- 2.4 Ich spreche lieber mit einem Menschen als mit BoRIS .
trifft zu trifft nicht zu
- 2.5 Ich konnte mich äußern, wie in einem normalen Gespräch mit einem Menschen.
trifft zu trifft nicht zu
- 2.6 Ich bin davon überzeugt, dass BoRIS mir richtige Informationen geliefert hat.
trifft zu trifft nicht zu
- 2.7 Das Gespräch mit BoRIS hat schnell zur gewünschten Information geführt.
trifft zu trifft nicht zu
- 2.8 Der Anruf bei BORIS hat sich gelohnt.
trifft zu trifft nicht zu
- 2.9 Jeder kann mit BoRIS ohne Probleme umgehen.
trifft zu trifft nicht zu
- 2.10 Durch das Gespräch mit BoRIS fühlte ich mich :
wohl unwohl
sicher unsicher
ruhig aufgereg
zufrieden unzufrieden
- 2.11 Ich habe die Fragen von BoRIS beantwortet:
stichwortartig satzartig teils-teils

III. WAS Sie bei BoRIS sehr gestört hat:

- 3.1 BoRIS hat plötzlich das *Gespräch* unterbrochen und ist spontan zu einem vorherigen Menu zurückgekehrt.
trifft zu trifft nicht zu
- 3.2 BoRIS hat das *Gespräch* oft mittendrin abgebrochen.
trifft zu trifft nicht zu
- 3.3 BoRIS konnte die Restaurantadressen nicht wiederholen.
trifft zu trifft nicht zu
- 3.4 BoRIS bietet keine detaillierteren Informationen über die gewünschten Restaurants.
trifft zu trifft nicht zu
- 3.5 Ich musste oft meine Eingaben wiederholen.
trifft zu trifft nicht zu
- 3.6 Ich konnte mir die Auswahlkriterien am Anfang des *Gesprächs* nicht merken.
trifft zu trifft nicht zu
- 3.7 Die Pausen zwischen meinen Eingaben und die Antworten von BoRIS waren zu lang.
trifft zu trifft nicht zu
- 3.8 Die Ansage der Restaurantnamen und die Adressen war nicht deutlich.
trifft zu trifft nicht zu
- 3.9 Der Weg bis zu der gewünschten Information war umständlich.
trifft zu trifft nicht zu
- 3.10 Es war mir nicht klar, wie ich meine Eingaben modifizieren soll, um schnell die gewünschte Information zu bekommen
trifft zu trifft nicht zu

IV. Für einen guten Gesprächsverlauf mit BoRIS ist es sehr wichtig;

- 4.1 ,dass BoRIS seine Fragen wiederholt, wenn ich nicht mehr weiter weiß.
trifft zu trifft nicht zu
- 4.2 ,dass BoRIS meine Eingaben immer gut versteht.
trifft zu trifft nicht zu

- 4.3 ,dass BoRIS klare Fragen stellt.
trifft zu trifft nicht zu
- 4.4 ,dass BoRIS schnell auf meine Eingaben reagiert.
trifft zu trifft nicht zu
- 4.5 ,dass BoRIS einfach zu bedienen ist.
trifft zu trifft nicht zu
- 4.6 ,dass BoRIS passende Informationen auf meine Anfragen gibt.
trifft zu trifft nicht zu
- 4.7 ,dass ich BoRIS Fragen stellen kann.
trifft zu trifft nicht zu
- 4.8 ,dass, nachdem ich die gewünschte Information bekommen habe, das Gespräch weiterführen kann, solange ich noch zusätzliche Fragen habe.
trifft zu trifft nicht zu
- 4.9 ,dass ich von Anfang an weiß, wie ich meine Eingaben formulieren soll.
trifft zu trifft nicht zu
- 4.10 ,dass ich von Anfang an weiß, welche Art von Informationen mir BoRIS liefern kann und welche nicht.
trifft zu trifft nicht zu
- 4.11 ,dass ich von BoRIS eine Bestätigung meiner Eingaben bekomme.
trifft zu trifft nicht zu
- 4.12 ,dass die Pausen nach meinen Eingaben von einen "Warten Sie einen Moment" ausgefüllt werden.
trifft zu trifft nicht zu
- 4.13 ,dass es einfach ist, die Eingaben zu modifizieren.
trifft zu trifft nicht zu
- 4.14 ,dass das Gespräch ohne Unterbrechungen verläuft.
trifft zu trifft nicht zu
- 4.15 4.15 ,dass BoRIS mit einer Hilfefunktion ausgestattet ist.
trifft zu trifft nicht zu

Datenerhebung

FRAGEBOGEN 1

Persönliche Daten

Geschlecht

Alter	Frauen	Männer	Total
Unter 20	0	1 (3,3%)	1... (1,88%)
zwischen 20-29	23 (100%)	25 (83%)	48 (90,5%)
zwischen 30-39		3 (10%)	3 (5,6%)
Zwischen 40-49		1 (3,3%)	1 (1,88%)
Total	23	30	53

Beruf /Ausbildung:

Studenten	47 (88,6%)
Absolventen	6 (11,3%) (4 Ingenieure & 2 Geisteswissenschaftler)

Herkunftsregion /-stadt:

Nordrhein Westfalen	31 (58,4%)
Rheinland - Pfalz	2 (3,7%)
Hessen	2 (3,7%)
Niedersachsen	3 (5,6%)
Brandenburg	1 (1,88%)
Sachsen	1 (1,88%)
Sachsen Anhalt	1 (1,88%)
Baden Württemberg	4 (7,54%)
Ausland	8 (15,09%) (Kirgistan: 1 Ecuador: 1 Spanien: 2 Griechenland: 1 Mazedonien: 1 Frankreich: 1 Rumänien: 1)

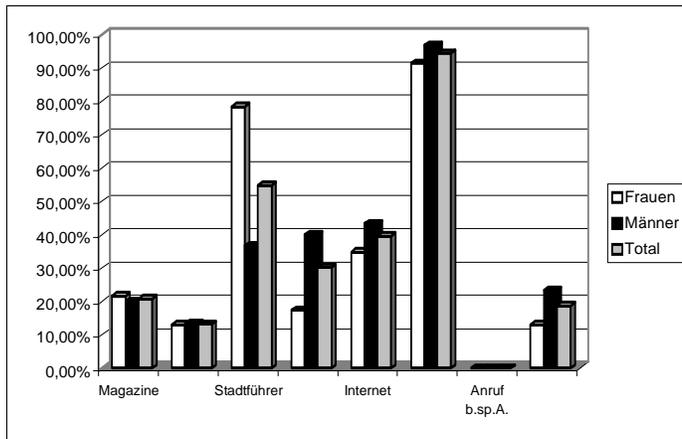
Jetziger Wohnort:

Bochum	48 (90,5%)
Essen	1 (1,88%)
Dortmund	1 (1,88%)
Gelsenkirchen	2 (3,7%)
Moers	1 (1,88%)

1. Wie erkundigen Sie sich nach einem Restaurant, wenn Sie in einem fremden Ort essen gehen möchten?

	Frauen	Männer	Total
1.1. Magazine	5 (21,73%)	6 (20%)	11 (20,75%)
1.2. Werbebroschüren	3 (13,04%)	4 (13,33%)	7 (13,20%)
1.3. Stadtführer	18 (78,26%)	11 (36,66%)	29 (54,71%)
1.4. Gelbe Seiten	4 (17,39%)	12 (40%)	16 (30,18%)
1.5. Internet	8 (34,78%)	13 (43,33%)	21 (39,62%)
1.6 Tip von Bekannten	21 (91,30%)	29 (96,66%)	50 (94,33%)
Anruf bei einem Sprach-Informationssystem.	0	0	0
1.8. Sonstiges	3 (13,04%) a) Anwohner fragen 1	7 (23,33%) a) Anwohner fragen 2	10 (18,86%) 3
	b) Selbst suchen 2	b) Selbst suchen 7	9

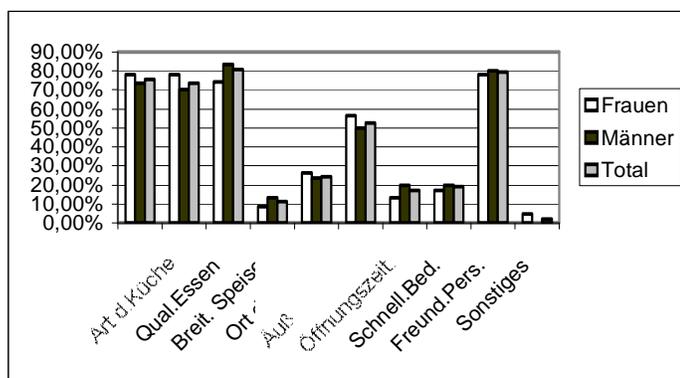
Tab.1



2. Was ist Ihnen wichtig, wenn Sie sich für ein Restaurant entscheiden sollen?

	Frauen		Männer		Total	
2.1. Preis	18	(78,26%)	22	(73,33%)	40	(75,47%)
2.2. Art der Küche	18	(78,26%)	21	(70%)	39	(73,58%)
2.3. Qualität des Essens	17	(73,91%)	26	(86,66%)	43	(81,13%)
2.4. Breitgefächertes Speiseangebot	2	(8,70%)	4	(13,33%)	6	(11,32%)
2.5. Ort des Restaurants	6	(26,08%)	7	(23,33%)	13	(24,52%)
2.6. Äußeres Ambiente	13	(56,52%)	15	(50%)	28	(52,83%)
2.7. Öffnungszeiten	3	(13,04%)	6	(20%)	9	(16,98%)
2.8. Schnelligkeit der Bedienung	4	(17,39%)	6	(20%)	10	(18,86%)
2.9. Freundlichkeit des Personals	18	(78,26%)	24	(80%)	42	(79,24%)
2.10 Sonstiges vegetarisch essen	1	(4,35%)	0		1	(1,88%)

Tab.2



3.0 Haben Sie schon ein automatisches sprachbasiertes Auskunftssystem benutzt?

	Frauen		Männer		Total	
Nein	20	(86,95%)	24	(80%)	44	(83%)
Ja	3	(13,04%)	6	(20%)	9	16,98%

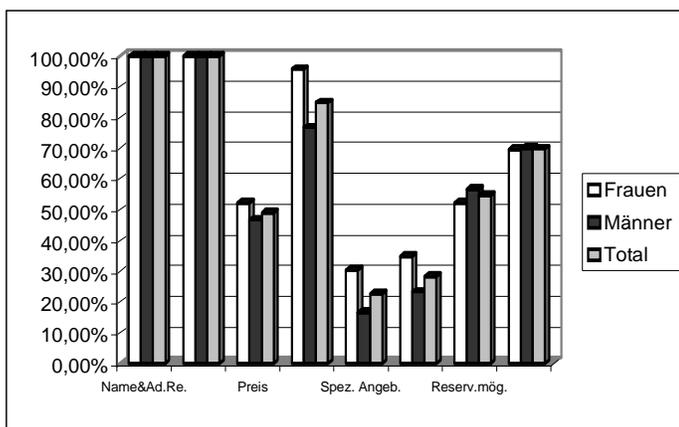
3.1. Wenn ja , bei welcher Gelegenheit ?

Frauen	Männer
1: kannte BoRIS; <i>dürftig</i>	1: beim Handy; <i>dürftig</i>
2: Bestellung von Kinokarten und Filminfo; <i>schlecht; gut</i>	1: Telefónica, Service Espania directo; <i>dürftig</i>
	2: Bahnauskunft ; <i>dürftig; ordentlich</i>
	1: Telefonbanking; <i>gut</i>
	1: erinnert sich nicht mehr; <i>dürftig</i>

4. Welche Informationen soll ein automatisches Restaurantauskunftssystem anbieten (mehrfache Nennung möglich)?

	Frauen	Männer	Total
4.1. Name, Adresse, Tel.	23 (100%)	30 (100%)	53 (100%)
4.2. Öffnungszeiten	23 (100%)	30 (100%)	53 (100%)
4.3. Preis	12 (52,17%)	14 (46,66%)	26 (49,05%)
4.4. Art der Küche	22 (95,65%)	23 (76,66%)	45 (84,90%)
4.5. Spezielle Speiseangebote	7 (30,43%)	5 (16,67%)	12 (22,64%)
4.6. Äußeres Ambiente	8 (34,78%)	7 (23,33%)	15 (28,30%)
4.7. Reservierungsmöglichkeit	12 (52,17%)	17 (56,66%)	29 (54,71%)
4.8. Wegbeschreibung	16 (69,56%)	21 (70%)	37 (69,81%)

Tab.3

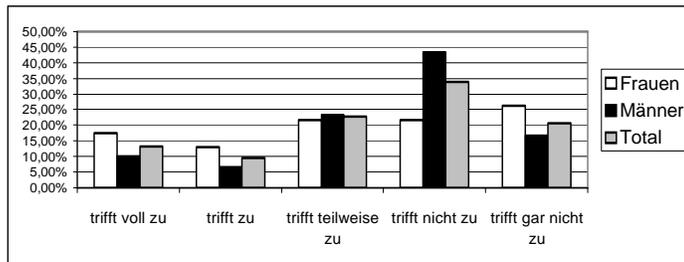


5. Wenn Sie bei BoRIS anrufen, erwarten Sie:

5.1. ein vernünftiges Gespräch wie mit einem Menschen

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	4 (17,39%)	3 (13,04%)	5 (21,73%)	5 (21,73%)	6 (26,08%)	23
Männer	3 (10%)	2 (6,66%)	7 (23,33%)	13 (43,33%)	5 (16,67%)	30
Total	7 (13,20%)	5 (9,43%)	12 (22,64%)	18 (33,96%)	11 (20,75%)	53

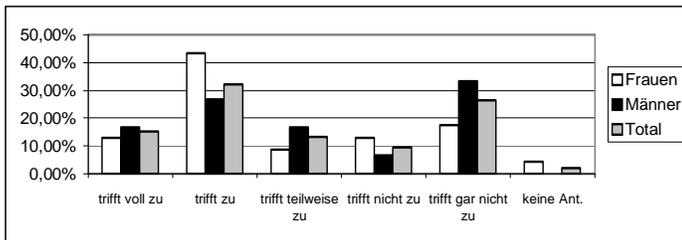
Tab 4.



5.2. Sie stellen dem Computer Fragen.

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Keine Antwort	Total
Frauen	3 (13,04%)	10 (43,47%)	2 (8,70%)	3 (13,04%)	4 (17,39%)	1 (4,35%)	23
Männer	5 (16,67%)	8 (26,66%)	5 (16,67%)	2 (6,66%)	10 (33,33%)	0	30
Total	8 (15,09%)	17 (32,07%)	7 (13,20%)	5 (9,43%)	14 (26,41%)	1 (1,88%)	50

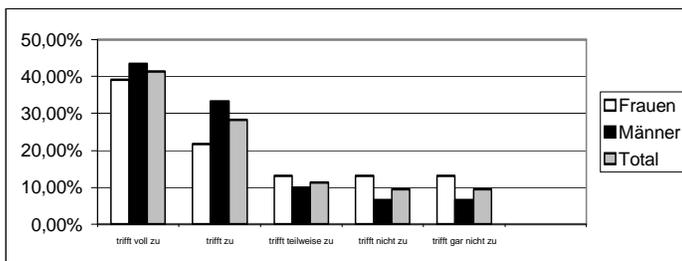
Tab 5.



5.3. Der Computer stellt Ihnen Fragen

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	9 (39,13%)	5 (21,73%)	3 (13,04%)	3 (13,04%)	3 (13,04%)	23
Männer	13 (43,33%)	10 (33,33%)	3 (10%)	2 (6,66%)	2 (6,66%)	30
Total	22 (41,50%)	15 (28,30%)	6 (11,32%)	5 (9,43%)	5 (9,43%)	53

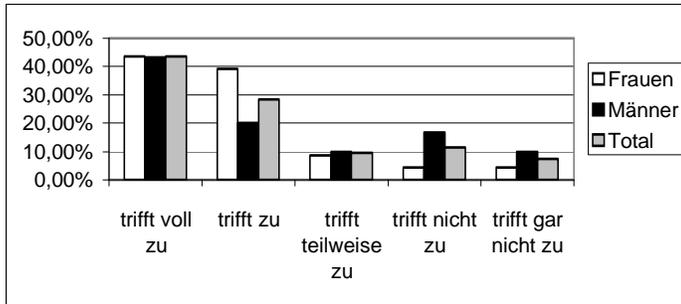
Tab. 6



5.4 Sie bekommen präzise Antworten.

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	10 (43,47%)	9 (39,13%)	2 (8,70%)	1 (4,35%)	1 (4,35%)	23
Männer	13 (43,33%)	6 (20%)	3 (10%)	5 (16,67%)	3 (10%)	30
Total	23 (43,40%)	15 (28,30%)	5 (9,43%)	6 (11,32%)	4 (7,55%)	53

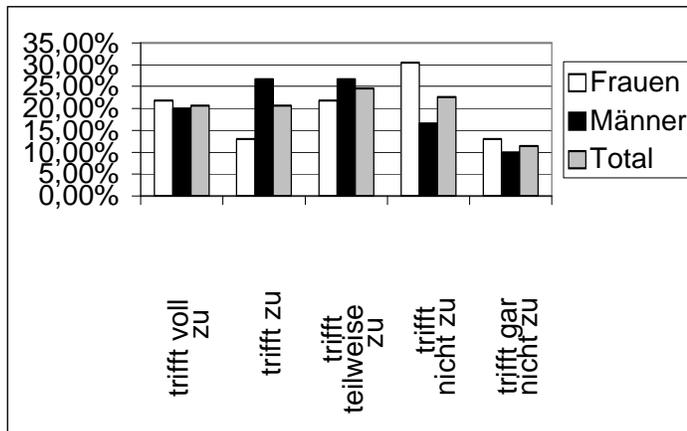
Tab.7



5.5. Es werden keine Mißverständnisse entstehen

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	5 (21,74%)	3 (13,04%)	5 (21,74%)	7 (30,43%)	3 (13,4%)	23
Männer	6 (20%)	8 (26,67%)	8 (26,67%)	5 (16,67%)	3 (10%)	30
Total	11 (20,75%)	11 (20,75%)	13 (24,53%)	12 (22,64%)	6 (11,32%)	53

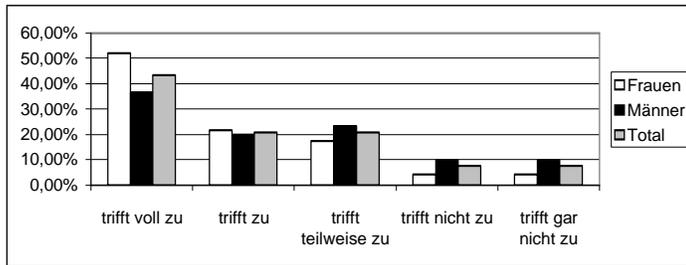
Tab.8



5.6. Sie werden schnell die gewünschte Information bekommen.

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	12 (52,17%)	5 (21,74%)	4 (17,39%)	1 (4,35%)	1 (4,35%)	23
Männer	11 (36,67%)	6 (20%)	7 (23,33%)	3 (10%)	3 (10%)	30
Total	23 (43,40%)	11 (20,75%)	11 (20,75%)	4 (7,55%)	4 (7,55%)	53

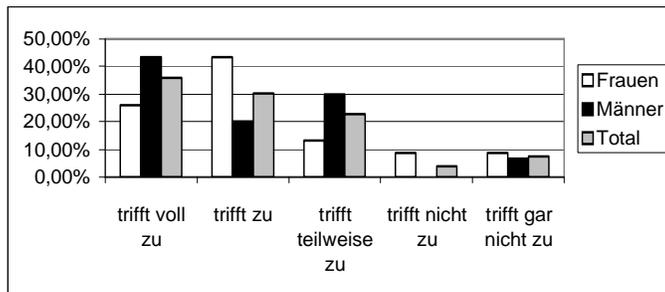
Tab. 9



5.7 Der Computer hat eine freundliche Stimme

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	6 (26,08%)	10 (43,48%)	3 (13,04%)	2 (8,70%)	2 (8,70%)	23
Männer	13 (43,33%)	6 (20%)	9 (30,00%)	0 (0,00%)	2 (6,67%)	30
Total	19 (35,85%)	16 (30,19%)	12 (22,64%)	2 (3,77%)	4 (7,55%)	53

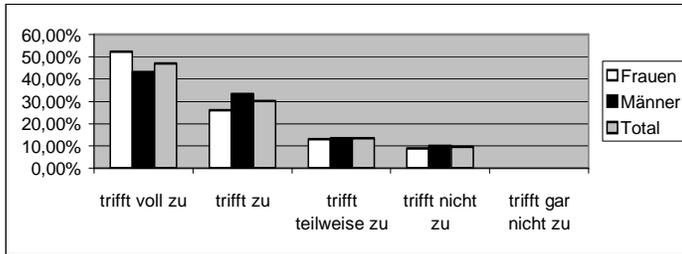
Tab. 9



5.8. Das System ist leicht zu bedienen

	Trifft voll zu	Trifft zu	Trifft teilweise zu	Trifft nicht zu	Trifft gar nicht zu	Total
Frauen	12 (52,17%)	6 (26,09%)	3 (13,04%)	2 (8,70%)	0	23
Männer	13 (43,33%)	10 (33,33%)	4 (13,33%)	3 (10,00%)	0	30
Total	25 (47,17%)	16 (30,19%)	7 (13,21%)	5 (9,43%)	0	53

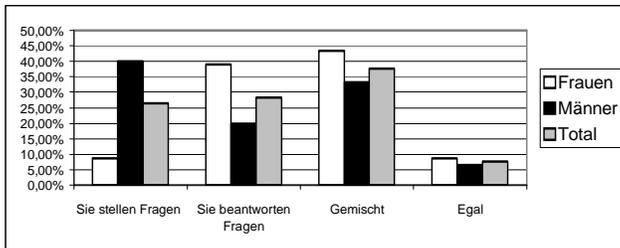
Tab.10



6.0 Was ist Ihnen lieber?

	Frauen	Männer	Total
Sie stellen Fragen an das System	2 (8,70%)	12 (40,00%)	14 (26,42%)
Sie beantworten Fragen	9 (39,13%)	6 (20,00%)	15 (28,30%)
Gemischt	10 (43,48%)	10 (33,33%)	20 (37,74%)
Egal	2 (8,70)	2 (6,67%)	4 (7,55%)

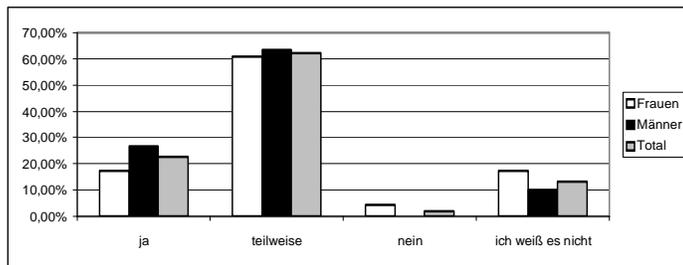
Tab.11



10. Finden Sie, dass natürliche Dialogssysteme für Restaurantauskünfte eine Alternative zu derzeitigen Systemen darstellen?

	Frauen	Männer	Total
Ja	4 (17,39%)	8 (26,67%)	12 (22,64%)
Teilweise	14 (60,87%)	19 (63,33%)	33 (62,26%)
nein	1 (4,35%)	0	1 (1,89%)
ich weiß es nicht	4 (17,39%)	3 (10%)	7 (13,21%)

Tab.12



FRAGEBOGEN 2

1.2 Was würden Sie bei BoRIS verbessern?

MÄNNER

1. **M** VP. 9 - Die Stimme bzw. Aussprache, da sonst alles übrig sinnlos ist.
2. **M** VP. 10 - Wenn der Computer bei der letzten Frage nichts findet, dann könnte man da zu ändern versuchen und erst dann alles zu modifizieren.
-Modifizieren, aber in kleinen Schritten, statt alles von Anfang anzufangen.
-Die Datenbasis könnte umfangreicher sein.
3. **M** VP. 12 Computergesteuerte Gesprächsführung. Mehr pausen zwischen Antworten. Abspielen deutlicher Antwort - Sprachdateien.
4. **M** VP. 13 bessere akustische Verständlichkeit der Computerstimme
 - langsame und freundliche menschliche Stimme
 - Möglichkeit, die Adresseninformation noch einmal anzuhören, zum Mitschreiben.
 - Bessere Möglichkeiten für spezielle Wünsche
 - In Anfangsmenü zu viele Auswahlkriterien genannt, die für einige Leute bestimmt schwer zu behalten sind.
5. **M** VP 2. Auf jeden Fall die Angaben über Namen + Adresse des Restaurants deutlicher / verständlicher aussprechen
 - Es sollte auch eine Option "Andere Präferenzen" geben.
6. **M** VP 1 Ausweitung der Fragestellung hinsichtlich des Ortes, der Küche.
 - Ausweitung aus Auskünfte für Objekte mit bekannten Namen um anderen Info wie Ort, Öffnungszeiten etc. zu bekommen.
 - Verbesserte Stimme, welche die Antworten gibt keine Computerstimme. Langsame Durchsage der Suchergebnisse.
7. **M** VP6 Das Navigationssystem, die Ansagenqualität, die Begrüßung am Ende ist nervig

M VP 49 Die Stimme, man sollte auch Nachfrage können; Adresse sollte man wiederholen.

9. **M** VP 46 Die Sprachsynthese am Ende. Die Zwischenpausen waren zu lang.

10 **M** VP 45 Sprachqualität der Auskunft. Es fehlte die Öffnungszeiten oder evt. Telefonnr. Zur Tischbestellung. Mehr Zeit zwischen den einzelnen Adressen, wenn es mehrere sind.

11 **M** VP 42 Es war nicht eindeutig, ob auch andere Varianten bei der Art der Küche Möglich waren; man weiß nicht, wie das funktioniert; einmal auf Cocktailbar nicht reagiert, ich wußte nicht ob er das verstanden hat; das System sollte vielleicht sagen "ich habe verstanden, Ihre Eingaben werden bearbeitet" oder so was ähnliches.

- Verbessern sollte man die Sprachsynthese.
- Am Ende des Gesprächs sollte er Nachfragen erlauben - man bekommt 3 Adressen am Ende und dann ist Schluss.
- Man sollte auch die Möglichkeit haben zusätzliche Fragen stellen zu können.
- Man sollte eine Tastenmöglichkeit einführen als Alternative
- bei "keine Präferenz" ist es nicht deutlich, ob ich per Zufall weiter geleitet werde oder ob ich auch eine Änderungsmöglichkeit habe.
- bei "arabisch" wird man einfach nicht mehr nach dem Ort gefragt.

12 **M** VP 44 - Sprachausgabe schwer verständlich

- - etwas zu unflexibel
- - besseres Feedback
- - Korrekturmöglichkeit

13 **M** VP 14 - Die Eingabemöglichkeit z. B. Eingabe von Art der Küche, Ort, usw. in einem Anfangsdialog (z.B. "ich möchte am Mittwoch um 20.00 Uhr spanisch essen gehen.

- die Ansage von Restaurants und Adresse.

14 **M** VP 15 Aussprache der Namen des Restaurants

- andere Formulierung der Alternativen
- bei "keinem Angebot" sagen, warum es kein "passendes" gibt

15 **M** VP 16 die Verständlichkeit der Angaben

16 **M** VP 17 Antwortzeiten von BoRIS verkürzen.

- beim erstem Gebrauch war ich ein wenig orientierungslos
- - Restaurantnamen und Adresse waren teilweise undeutlich.

17 **M** VP 21 Geschwindigkeit der Adressenaussagen (zu schnell und undeutlich)

- deutlich machen, daß auch nicht erwähnte Antworten möglich sind.
- nach Restaurant Namen suchen lassen

18 **M** VP 22 Ergebnisansage (Leerlauf davor kürzer, deutlichere Ansage)

- bessere Beschreibung, was der Anrufer machen soll

19 **M** VP 24 Computerstimme

- Resetfunktion, um sofort neu starten zu können
- Wiederholung vom Namen und Adresse der Restaurants

20 **M** VP 26 TTS = Text to Speech als Alternative zu Sprachsynthese

- Wiederholmöglichkeit

21. **M** VP 28 Restaurantnamen Aussprache.

22 **M** VP 29 Adresse und Restaurantnamen wiederholen zum mitschreiben

- Wiederholung der Dateninformation, mit welcher BoRIS denn man genau sucht;
- Ich weiß es nicht mehr - bei den Modifikationen- was er sucht, wenn ich mehr als 2 -3 Mal meine Eingaben modifiziert habe, vielleicht einen Hinweis geben- wenn man aus dem Rhythmus aus ist, daß man seine Frage komplett neu stellen kann.

23 **M** VP 30 Die Aussprache muß deutlicher werden.

24 **M** VP 31 Ausgabe muß viel deutlicher werden.

- Ausgabe muß wiederholt werden
- - wenn er kein Restaurant findet, muß die Möglichkeit bestehen neu zu beginnen (er arbeitet mit Verknüpfungen die Fehler verursachen)

25 **M** VP 33 besser/ deutlicher ausgesprochenen Restaurantnamen und Adressen.

26 **M** VP 36 Möglichkeit der Wiederholung

27 **M** VP 53 - die ganze Struktur des Systems

- die Stimme
- die Geschwindigkeit

28 **M** VP 25 nicht schlecht wäre eine bessere Anleitung (z.B. "Hallo, hier ihr Berater für Ihre Wünsche. Nach welchen Kriterien soll ich für Sie suchen?"), sollte man persönlicher sein

- er sollte auf Wunsch die Fragen wiederholen, wenn man die Sterntaste betätigt
- eine andere Stimme welche wählbar ist, z.B. Frau /Mann
- - eine Hintergrundmusik
- - Mit Anrede ansprechen, z.B. "sagen sie mir ihren Namen....in dem Fall, Herr Blat, Guten Tag Herr Blat.
- Allgemeine Fragen formulieren, z.B.: "nach welchen Kriterien möchte Sie suchen, dann als Anrufer die Fragen stellen.

29 **M** VP 4 Ansagetext vor der gleiche Stimme wie Fragesteller vorlesen lassen

- Ansagetext langsamer vorlesen
- - Ansagetext nach Vorbild der Telefonauskunft wiederholen lassen

30. **M** VP51 Einzelne Fragen stellen: z.B. 1) Art der Küche; 2) Ort usw.
Auswahlkriterien am Anfang waren zu lang

FRAUEN

1 **F** VP 35 Adresse mit mehr Pausen / Abstand sagen;
- immer nach genauen Wünschen fragen

2 **F** VP. 18 Die Auswahlkriterien erweitern.
- die Totalzeit verkürzen.
- Die Adressen deutlicher und langsamer nennen.

3. **F** VP 48 Wenn man mehrere Angaben gibt, dann sollte größere Pausen zwischen den einzelnen Adressen sein, da man sonst nicht mitschreiben kann.
- die Angaben des Restaurants sind sehr undeutlich

4. **F** VP7. Die Restaurantaufzählung ist sehr schlecht zu verstehen- es mußte langsam und deutlicher gesprochen werden.

5 **F** VP 39 Verständlichkeit der Restaurantaussagen.
- Fragenaufbau
- Wiederholen der Antworten.

6 **F** VP 11 mehr Möglichkeiten für spezielle Nachfragen Wiederholung der Angaben (Adressen + Namen des Restaurants)
- deutlicher aussprechen.

7 **F** VP 43 Sprachqualität
- Restaurantangabe wiederholen

8. **F** VP 47 die Ansagen des Systems und die Restaurantnamen
- vielleicht mehr Auswahlmöglichkeiten an Restaurantansagen lassen
- man sollte am Ende noch einen Hinweis darauf geben, daß BoRIS auch die Restaurantadresse wiederholen kann.

9.**F** VP 38 Die Auskunftsstimme sollte die Adresse und Namen der Restaurants langsamer und deutlicher nennen, so daß die Möglichkeit besteht in Ruhe mit zu schreiben. Außerdem sollte es eine Wiederholungsmöglichkeit geben, falls man etwas nicht verstanden hat.

10. **F** VP 41 Die Adressennennung besonders die Hausnummer.

11. **F** VP 3 Sprachqualität
- Auswahlmenü.

12. **F** VP 32 Adressenansagen: deutlicher, langsamer, wiederholen!

- Es muß deutlich werden, daß es neben den genannten Alternativen (z. B. Art der Küche) noch andere Möglichkeiten gibt.
- Die Ansage muß durch frühere Beantwortung der Frage verkürzt werden.

13. **F** VP 34 Die Ansagen des Namens und der Adresse sollten deutlicher sein und wiederholt werden.

- es sollte die Möglichkeit geben mehrere Anfragen während eines Telefonats zu machen und nicht nur eine
- auf Wunsch kurze Wegbeschreibung
-

14. **F** VP 27 Name und Adresse von der gleichen Stimme; und ruhig und langsam aussprechen.

- daß man die Möglichkeit hat, wenn man etwas falsches angegeben hat zurück zu dem Menu zu kommen, dem man falsch angegeben hat und nicht alles von Anfang an wiederholen.
- Die Sprachsynthese

15 **F** VP 40 Die Sprachsynthese

- die Ansagen langsamer und deutlicher aussprechen

16 **F** VP 37 Restaurant und Straßen deutlicher ansagen und wiederholen

- nach Restauranttyp Programm nicht automatisch beenden, sondern Möglichkeit für weitere Suche. Geben.

17 **F** VP 8. Flexibilität

- deutlichere Aussprache, leichtere Verständlichkeit der Adressenangaben

18 **F** VP 19 Die Antworten etwa deutlicher geben.

19 **F** VP 23 Keine Fragen, wenn nicht mit "Ja " oder "Nein" geantwortet werden kann

- längere pausen zwischen den einzelnen Angaben (Restaurants)
- deutlichere (nicht elektronische) Aussprache der Suchergebnisse

20. **F** VP 52 Wenn man eine Auswahl modifiziert wird und mal wieder nicht gefunden wurde, sollte man die letzte Modifikation nach einmal ändern können und nicht wieder von Anfang an (siehe Szenario 1)

- mehrfach auswählen

21 **F** VP 50 Adresse deutlicher machen

- langsamer sprechen
- die Frage - Antwort zuvor gut, nur evtl. Verwirrung durch zu viele Präzisionen

22. **F** VP 20 öfter nach Spezifizierung fragen.

- vielleicht eine Art von Wiederholung "Sie suchen also ein spanisches, preiswertes, etc. ...", man weiß nicht ob man auch wirklich das gesagt bekommt, was man sucht.
- Ansage der Adressen sehr viel deutlicher evtl. Wiederholung auf Wunsch.

23. **F** VP 5 - Möglichkeit die Adresse zu wiederholen
- Pausen zwischen mehreren Restaurants
 - Die Computerstimme, die die Restaurants nennt eventuell Möglichkeit auf Englisch zu sprechen für Ausländer

1.3. Wie würden Sie BoRIS beschreiben

MÄNNER

1. **M** VP.9: - keine Alternative zu herkömmlichen Informationsquellen.
- kalte Roboterstimme
2. **M** VP.10: Maschine immer mit gleichen Fragen, die aber auch nicht dabei einem hilft mit kleinen Angaben oder Informationen (z.B. Preise etc.) und wenn sie nichts findet dann legt sie auf. Es war mir nicht klar warum bei der "Art der Küche", nur ein paar ital., deutsch, griechisch; man sagt "arabisch" z.B. und dann gibt er was an; spanisch gibt er nicht an, aber es ist unklar ob es "spanisch" nicht gibt oder habe ich andere Daten falsch eingegeben und die Kombination passt nicht.
3. **M** VP 12. Info über Art der Küche, Uhrzeit, ungefähren. Ort, Preis bei Restaurants
Vorsicht! Antworten zu schnell und zu undeutlich
4. **M** VP 13. Computergesteuerte Auskunft über die Restaurants in Bochum. Schlechte Verständlichkeit der Auskunft. Nur für nicht spezielle Wünsche geeignet.
5. **M** VP 2 Du kannst nach Preisklassen, Tageszeiten, Art der Küche und Ort des Restaurants fragen. Es geht, daß Du "quasi" "rückblättest" und dich neu entscheidest zwischen den Möglichkeiten einer vorher gestellten Frage.
Die Antworten sind leider manchmal undeutlich, was Name und Adresse betrifft. Prüft nach, ob das stimmt, was er Dir erzählt. Verlaß Dich nicht darauf!
6. **M** VP 1 SVEN Unpräzise Ergebnisse. Bei genauer _____ einigermaßen erfolgreich.
Schlecht verständliche Ergebnisse. Umständliches Frage / Antwort Spielchen.
- 7 **M** VP 6. Ein Informationssystem, das überhaupt nicht funktioniert. Die Ansagen waren zu schnell, die Qualität des Systems war schlecht.
8. **M** VP 49 BoRIS spricht zu schnell, ich habe nichts verstanden. Er konnte nichts wiederholen, auch keine Nachfrage war möglich. Das System ist total unsympathisch.
9. **M** VP 46 Ziemlich komisch, sehr fremd für mich, ich habe mich dumm gefühlt, ich war genervt, besonders am Ende, wegen den schlechten Ansagen;

10. **M** VP 45 Eine gute Idee, man kann genau definieren was man sucht; man kann auch gezielt nach Adressen fragen. Aber bei den Restaurant - Adressen muß man ziemlich gut hinhören und flink schreiben können. Mußt Du ausprobieren.
11. **M** VP 42 Informationssystem mit Restaurantdatenbank, flexible Wahl und Adressennennung.
12. **M** VP 44 recht hohes Verständnis
- gute Sprachausgabe
 - einleuchtende Benutzerführung
13. **M** VP 14 BoRIS ist ein Auskunftssystem für Bochumer Restaurants.
14. **M** VP 15 man darf sich nicht irritieren lassen 8 durch Wiederholungen
- relativ einfach
 - man muß immer wieder probieren und darf nicht schnell aufgeben.
15. **M** VP 16 benutze es wenn Du Rätsel magst
16. **M** VP 17 - Computergestütztes Informationssystem, das Auskünfte über Bochumer Restaurants und Kneipen erteilt
17. **M** VP 21 für einfache Infos O.K.; für spezielle Wünsche irreführend
18. **M** VP 22 ein Versuch ist es wert; Ergebnis kommt ziemlich zügig
- 19 **M** VP 26 Beendet das Gespräch zu früh;
- die eigentliche Auskunft ist schwer verstehbar
 - - keine Wiederholung möglich.
 -
- 20 **M** VP 28 telefonisches Auskunftssystem über Restaurants in Bochum
- bei der Antwort sollte dieser genau zuhören, falls er es testet.
- 21 **M** VP 29 Ruf an! Laß Dich nicht vollquatschen und lege, wenn Du aus dem Rhythmus bist, auf
- ordentliche bis schlechte Auskünfte.
- 22 **M** VP 30 relativ schnelles Info - System
- 23 **M** VP 31 weiß viel, sagt es aber nicht immer;
- schlecht zu verstehen
 - - das System an sich finde ich super, auch unter wirtschaftlichen Gesichtspunkt sehr interessant (Einführung +Pflege des Systems)

24 **M** VP 33 Restaurantnamen sehr schnell und schlecht ausgesprochen; Straßennamen konnte man verstehen.

- die Ausgabe schlecht , deswegen das System ist als dürftig zu bezeichnen.

25 **M** VP 36 Unkompliziert und schnell

26 **M** VP 53 Computersystem; mangelhafte Informationen.

27 **M** VP 25 - Keine qualitative Informationsquelle. BoRIS ist kein ausgereiftes System. Vor allen Dingen ist die Stimme nicht angenehm.

-macht keinen Spaß da anzurufen; man spricht mit einer Wand

- ich hatte das Gefühl, nicht richtig verstanden zu sein;
- die Pausen waren zu lang, man wußte es nicht mehr, was da abläuft.

28 **M** VP 4 Restaurantführer auf Computerbasis;

- etwas umständlich geht aber.
- - Ansagen schnell, teilweise zu schnell

29. **M** VP 51 Zuerst aufschreiben, welche Fragemöglichkeiten man hat

- schneller Ablauf
- - Präzise Antworten
- - Probleme die richtigen Fragen zu stellen.

30. **M** VP 24 einfaches System, was keine genauen Daten angeben kann (??

- Resetfunktion, um sofort neu starten zu können
- Wiederholung von Namen und Adresse der Restaurants
-

FRAUEN

1. **F** VP 35 Eigentlich gut. Mit eigener Frage. Aber lieber am Anfang genau sagen, was man will. Adresse wird zu schnell und undeutlich gesagt; klappt nicht immer mit dem genauer Nachfragen.

2. **F** VP 18 Falls meine Wünsche außerhalb der von BoRIS genannten Angaben liegen, konnte ich nicht weit.

- Die genannten Adressen sind schwer verständlich.
- Die Möglichkeiten der Modifizierung sind begrenzt. Man fühlt sich eingeschränkt.

3. **F** VP48 Man bekommt Informationen über Restaurants in Bochum, nach Art der Küche. Ort, Tageszeit.

- die genaue Adresse ist jedoch schlecht zu verstehen, d.h. die Straße kann man verstehen, die Namen überhaupt nicht!

4. **F** VP 7 Du rufst an und wirst von einem Computerstimme nach Deinen Interessen gefragt, z.B. Art der Küche oder nach dem Ort. Zum Schluß wertet er Deine Aussagen aus und schlägt Dir ein passendes Restaurant mit Adresse vor.

5. **F** VP 39 nette Stimme

- geht schnell

6. **F** VP 11. Informativ über Adressen (Nennung) des Restaurants schlecht zu verstehen
- keine eigenen konkreten Fragen möglich
 - keine Angaben über sonstige Eigenschaften des Restaurants, nur grobe Preiskategorie, Art der Küche und Lage.
7. **F** VP 43 einfache Bedienung; es war einfach etwas raus zu finden nur war schwer zu verstehen
8. **F** VP 47 Als eine Restaurantsuchsystem, bei dem man den Ort d. Restaurants, dem Preis sowie die Art aussuchen kann.
- cooles System; ich würde es nutzen als Alternative zu Telefonbücher.
9. **F** VP 38 Ein elektronisches Auskunftssystem, das einem hilft innerhalb von Bochum das jeweils passende Restaurant zu finden.
- 10 **F** VP 41 sehr freundliche Stimme; Spracherkennung war gut.
11. **F** VP 3 automatisches Programm; Endlosewortesschleife; Restaurant - Berater für Bochum.
- 12 **F** VP 32 Das System ist unpräzise: nur eine Orientierungsmöglichkeit; für genaue Auskünfte ungeeignet
- die Adresse sind schlecht zu verstehen.
 - Es werden nicht immer alle Alternativen geboten.
- 13 **F** VP 34 kurze, präzise Fragen stellen
- zuvor gut überlegen nach welcher Art Restaurant man sucht und in welcher Preisklasse
 - - deutlich sprechen
 - - Zeit nehmen, um Informationen von BoRIS zu bekommen.
- 14.- **F** VP 27 am Anfang verwirrend, weil ziemlich wenig Antwortmöglichkeiten angegeben werden, reagiert aber auf andere Antworten auch.
- die Adresse: zu schnell und zu undeutlich; Umstellung wegen anderen Stimme
 - Das System wirkt intelligent, gut konzipiert
 - Die Stimme ist ruhig und freundlich.
15. **F** VP 40 Die Adressenansage ist sehr undeutlich; man muss sich sehr konzentrieren um etwas zu verstehen
- wenn man genaue Angaben macht, dann ist es O.K.; nur bei ungenauen Angaben hat man Probleme damit.
 - es ist nicht schlecht, aber es wäre besser mit einem Menschen zu sprechen
 - - ich würde aber das System benutzen
 - - es ist komisch mit einem Computer zu reden; ich habe versucht so deutlich wie möglich zu sprechen;irgendwie fehlt da der persönliche Charakter
 - -

16. **F** VP37 hilfreich
Zusätzlich zu anderen Informationsquellen verwenden
- Ansage schwer verständlich.
17. **F** VP 8 Langwierig
- nervig, wenn BoRIS die Anfrage nicht versteht
 - - ich würde mich lieber bei einem Menschen nach einem Restaurant erkundigen.
 - - Adressenangabe der jeweiligen Restaurants schwer verständlich.
18. **F** VP 19 gutes System mit den Abfragen
- undeutliche Antworten
 - nicht zu lange dauert bis man die Information bekommt (trotz Pause)
19. **F** VP 23 Typisch "elektrisch"
20. **F** VP 52 intelligenter Sprachcomputer mit angenehmer Stimme (siehe Szenario 1)
21. **F** VP 50 Frage- Antwortspiel sehr gut! Klar u. deutlich; kam heraus, was ich wollte
- die Bearbeitung sehr schnell, wenn man gezielte Fragen stellt, wie z.B. "20 DM, chinesisch, Zentrum"
 - - es war freundlich, nur er redet zu schnell und zu undeutlich, Straßennamen nicht verständlich- die Stimme war ausländisch, deswegen habe ich vielleicht nicht verstanden
22. **F** VP 20 - gut, wenn man eine genaue Adresse von einem Restaurant wissen möchte; es wird schwierig bei speziellen Wünschen;
- man bekommt nicht unbedingt ausreichend Informationen.
23. **F** VP 5 - gute Auswahl an Restaurants
- nette Stimme (also die erste Stimme)
 - - man kann seine Wünsche verändern, Wahl modifizieren
 - versteht oft die Adresse nicht
 - geht etwas zu schnell.