MENSCH-MASCHINE-DIALOGE IN NATÜRLICHER SPRACHE

Hitzenberger, L., Kritzenberger, H. FG: Linguistische Informationswissenschaft Universität Regensburg

1. Einleitung

Der Anwendungskontext der Speech-Technologie, z.B. Dialogsysteme, Bürokommunikation oder telefonische Auskunftssysteme, bestimmt jeweils die speziellen Anforderungen an Spracherkennung und -verarbeitung. Festlegungen der Leistungsanforderungen können nur empirisch abgeleitet werden. Eine ad-hoc-Modellierung mit Beispielsätzen hilft hier nicht weiter, da sie über die Zufälligkeit der Auswahl nicht hinausgeht und zudem die auftretenden pragmatischen Faktoren (z.B. Kooperativität) nicht berücksichtigt. Auch die Ableitung aus dem Mensch-Maschine-Dialog ist unzureichend, da Mensch-Maschine-Dialoge und Mensch-Mensch-Dialoge unterschiedlich verlaufen. Hinweise hierauf finden sich in /1/, /2/, /3/. Die Ergebnisse der DICOS-Simulationsstudien /4/, /5/, /6/ zeigen aber nicht nur, daß Sprache im Dialog mit Computersystemen anders eingesetzt wird als in der zwischenmenschlichen Kommunikation. Sie können auch nachweisen, inwieweit das Computerbild des Benutzers oder die Restriktionen (Beschränkungen des Leistungsumfangs) das Sprachverhalten des Benutzers beeinflussen. Diese Erkenntnisse haben Auswirkungen auf das Design von Informationssystemen, weil die Einbeziehung des Wissens über Art und Umfang der Beeinflussungsfaktoren zu einem sehr frühen Zeitpunkt des Designprozesses die Akzeptanz des Systems entscheidend beeinflußt. Da die heutige Spracherkennungstechnologie noch nicht ausgereift genug ist, um die notwendigen Daten mit einem einsatzfähigen System bekommen zu können, bietet sich hier die Hidden-Operator-Simulation an.

2. Methode

Die Daten der Hidden-Operator Simulationsstudie von DICOS wurden mit drei qualitativ unterschiedlichen hypothetischen Informationssystemen (im folgenden als Sys 2, Sys 3 und Sys 4 bezeichnet) gewonnen. Es handelt sich um ein optimales Informationssystem (Sys 2), das sich exakt wie ein menschlicher Dialogpartner verhielt und um zwei Systeme mit unterschiedlicher Restriktionsmenge (Sys 3, Sys 4). Parallel zu System 2 wurde ein verhaltensgleiches System 1 getestet, bei dem ein Mensch-Mensch-Dialog über den Computer als Vermittlungsmedium stattfand. Dieses Konzept läßt eine direkte Vergleichsmöglichkeit zwischen Mensch-Mensch-Dialog und Mensch-Maschine-Dialog zu.

Die o.g. vier Dialogsituationen (Sys 1 bis 4) wurden jeweils einmal mit getipptem und einmal mit gesprochenem Input untersucht. Die Ausgabe der Antwort erfolgte über Bildschirm. Aus diesem Versuchsaufbau ergibt sich ein zweifaktorieller Versuchsplan mit zwei mal vier Faktorstufen (Sys 1 bis 4, getippter und gesprochener Input) und 40 Versuchspersonen (acht unabhängige Stichproben zu je fünf Versuchspersonen). (Systembeschreibungen und Versuchsaufbau in /7/, /8/, /9/.)

3. Ergebnisse

Die folgenden Ergebnisse beziehen sich auf zwei Testreihen mit der Domäne Bundesbahnauskunft und der Domäne Bibliotheksauskunft. Das Korpus umfaßt 518 Sätze (getippt) und 620 Sätze (gesprochen) bei der Bundesbahnauskunft, sowie 599 (getippt) und 615 Sätze (gesprochen) bei der Bibliotheksauskunft.

Die Auswertung der Versuchsprotokolle erfolgte sowohl mit den Mitteln der statistischen Varianzanalyse als auch explorativ.

Die statistische Analyse zeigt, daß im Mensch-Maschine-Dialog die Anzahl der Types (Worttypen) gegenüber dem Mensch-Mensch-Dialog absinkt und daß die Einführung von Restriktionen ein weiteres Absinken bewirkt. Für die Tokens (Anzahl der insgesamt gebrauchten Wörter) konnte hingegen kein Einfluß des Computerbildes oder der Restriktionen festgestellt werden. (Im Gegensatz dazu stehen /10/, /11/.) Unterschiedlich bezüglich der Tokenanzahl erwiesen sich die Inputkanäle. Mit Ausnahme von System 4 war bei allen Systemen die Tokenanzahl im gesprochenen Input höher als für das korrespondierende System im getippten Input. Ebenso konnte auch eine höhere Fehlerrate bei getipptem als bei gesprochenem Input festgestellt werden (genaues in /9/).

Die explorative Analyse der Protokolle ergab, daß sich der Mensch-Maschine-Dialog (Sys 2, Sys 3, Sys 4) vom Mensch-Mensch-Dialog (Sys 1) durch die Variationsbreite syntaktischer Strukturen, die Menge der Fehlleistungen in gesprochener Sprache, den Fragestil, die Mittel der Dialogorganisation und partnerorientierter Dialogsignale, sowie durch Abweichungen und Überspezifikationen gegenüber dem normalen Sprachgebrauch unterscheidet.

3.1 Strukturunterschiede

In der Domäne Bundesbahnauskunft schränkten Versuchspersonen die Variationsbreite von Satzmustern im Dialog mit dem Computer ein. Die Tendenz zu solchen Einschränkungen wurde bereits ab Sys 2 beobachtet. Als Ursache für dieses Verhalten muß das Computerbild gesehen werden. Die Reduktion der oberflächenstrukturellen Vielfalt setzt sich aber in noch stärkerem Ausmaß bei der Einführung von Restriktionen (Sys 3, Sys 4) fort. Je eingeschränkter die sprachlichen und kooperativen Fähigkeiten des Systems werden um so größer wird die Tendenz, bereits erfolgreich eingeführte Syntaxmuster im Verlauf des Dialogs beizubehalten. Oftmals werden lediglich die zur Spezifizierung des neuen Informationswunsches notwendigen Slots (z.B. für Zeit- und Ortsangaben etc.) ausgetauscht.

Neben den Satzmustervariationen wurden im Mensch-Maschine-Dialog Konstruktionen verwendet, die gegenüber dem normalen Sprachgebrauch im Mensch-Mensch-Dialog als abweichend oder überspezifiziert bezeichnet werden können. Beispielsweise ist eine Formulierung wie "Zeigen Sie mir die Zugverbindungen zwischen Abfahrtsort Regensburg und Ankunftsort Dortmund mit Zwischenaufenthalt in Köln." als überspezifiziert, bzw. eine Formulierung wie "Ich möchte nächsten Dienstag von Regensburg nach Kiel fahren. Geben Sie mir die Zeitpläne." als abweichend (in der Wortwahl) zu bewerten. Die o.g. Beobachtungen stammen aus der Domäne Bundesbahnauskunft. Für die Domäne Bibliotheksauskunft liegen zur Verwendung der Syntaxmuster noch keine Ergebnisse vor.

Für den gesprochenen Input konnte in beiden Domänen beobachtet werden, daß im Dialog mit dem Computer Fehlleistungen beim Sprechen vermieden werden, die im allgemeinen im Mensch-Mensch-Dialog unkontrolliert sind. Bei diesen Phänomenen gesprochener Sprache handelt es sich um verschiedene Arten von Anakoluthen wie Abbruch (z.B. "Kann ich da mit diesem / Gilt dies dann auch oder muß ich dann eine neue Karte lösen?"), oder Verbesserungen (z.B. "Können Sie mir sagen, wann heute abend ein Zug / Züge nach Frankfurt fahren?"). Weiter zählen zu den Merkmalen gesprochener Sprache auch Nachträge, Ausklammerungen und Durchbrechung von Satzkonstruktionen (z.B. "Ist das dann eigentlich praktisch egal Hauptsache ein Wochenende liegt da dazwischen. äh an welchem Wochentag man da abfährt?"). Die genannten Phänomene kommen bei den Systemen 1 bis 4 der Bundesbahnauskunft in einem absoluten Zahlenverhältnis von 51:15:9:2 und bei der Bibliotheksauskunft in einem Zahlenverhältnis von 31:18:10:4 Belegen vor. Dieser Rückgang in der Beleglage bereits bei dem optimalen, im Leistungsumfang nicht beschränkten Simulationssystem (Sys 2) zeigt, daß hier ebenfalls der Einfluß des Computerbildes wirksam wird.

3.2 Fragestil

Als weiterer Unterschied zwischen Mensch-Mensch- und Mensch-Maschine-Dialogen ist die Veränderung des Fragestils zu nennen. Dieser verhält sich jedoch domänenspezifisch unterschiedlich.

Bei der Domäne Bahnauskunft nimmt der indirekte Fragestil mit Imperativsätzen (z.B. "Nennen Sie mir eine Zugverbindung am 23.12.1989 von Regensburg nach Oldenburg."), Aussagesätzen (z.B. "Ich suche eine Verbindung, mit der ich am 23.12. um 8 Uhr in Oldenburg sein kann.") und Wunschsätzen (z.B. "Ich hätte gerne eine Verbindung von Etterzhausen nach Hindelang für Samstag möglichst früh hin und Sonntags möglichst spät zurück.") ab, während der direkte Fragestil zunimmt. Dies bedeutet einen Anstieg der W-Fragen (z.B. "Wann geht am Samstag der erste Zug von Regensburg nach Nürnberg und am Sonntag abend der späteste zurück.") und der Entscheidungsfragen (z.B. "Gibt es eine Zugverbindung zwischen Regensburg und Hamburg zwischen 8 Uhr und 15 Uhr.").

Auffällig war auch der gehäufte Gebrauch von Ellipsen bei den Versuchspersonen, die mit dem nicht-restriktiven Computermodell (System 2) arbeiteten. Der Anteil der Ellipsen liegt bei System 2 bei ca. 22% (getippter Input) bzw. bei ca. 35% (gesprochener Input). Kennzeichnend für einen Großteil dieser Ellipsen ist, daß sie auf überflüssige oberflächenstrukturelle Organisation (d.h. Einbettung der domänenspezifischen Informationsfelder in "wohlgeformte", vollständige Satzstrukturen) verzichten und die Informationsabfrage auf die Nennung der relevanten Felder im domänenspezifischen Informationsgerüst reduzieren. Beispielsweise treten Anfragen auf wie "Preise" oder "Bahnsteig".

Bei der Domäne Bibliotheksauskunft nimmt der direkte Fragestil im Dialog mit dem Computer ab (gesprochener Input: 53:27:2:2 Belege, getippter Input: 37:41:19:30 Belege), ebenso z.T. der indirekte Fragestil (gesprochener Input: 68:40:100:46 Belege, getippter Input: 48:13:3:0 Belege). Eine besonders starke Zunahme ist hingegen bei den Ellipsen zu beobachten. Bereits bei System 2 neigten die Benutzer dazu, nur noch das Schlagwort oder den Themenbereich der Suche als isolierte Nominalphrase anzugeben (gesprochener Input: 19:53:51:77 Belege, getippter Input 19:57:95:80 Belege).

3.3 Dialogorganisation und partnerorientierte Dialogsignale

Die Einschränkungen dialogorganisierender Mittel und partnerorientierter Dialogsignale ist ebenfalls auf das Computerbild zurückzuführen. Sie haben im Mensch-Mensch-Dialog eine vorwiegend pragmatische Funktion. Sie sollen u.a. Höflichkeit signalisieren, sozialen Kontakt herstellen, die Aussage abzutönen oder dienen als Kohärenzmittel. Da der pragmatische Gehalt von Äußerungen ein inhärenter Bestandteil der sprachlichen Kommunikation ist, findet sich hier ein deutlicher Hinweis auf einen anderen Gebrauch von Sprache.

Benutzer verwenden im Dialog mit dem Computer weniger häufig eine explizite Dialogeröffnung durch Gruß und indirekte metathematische Einführung (z.B. "Grüß Gott. Ich hätte eine Frage und zwar..."). Auch die Dialogbeendigung durch Bedanken und Verabschiedung am Ende des Dialogs (z.B. "Vielen Dank.") wird verringert. Bei den partnerorientierten Dialogsignalen werden Modalpartikeln (z.B. allerdings, eventuell etc.), satzeinleitende Partikeln (z.B. satzeinleitendes "und", "und außerdem", "dann" etc.) und Höflichkeitsindikatoren wie konjunktivisch formulierten Frageäußerungen (z.B. "Könnten Sie mir das Gleis sagen?") und Frageäußerungen mit bitte unter dem Einfluß des Computerbildes reduziert.

Die Vorkommenshäufigkeit der Partikeln bei den Systemen der Bahnauskunft beträgt 225:48:20:2 Belege (gesprochener Input) bzw. 63:37:12:5 Belege (getippter Input) und bei den Systemen der Bibliotheksauskunft 13:8:1:0 Belege (gesprochener Input) bzw. 36:6:1:0 Belege (getippter Input). Die genannten Höflichkeitsindikatoren kommen bei der Bahnauskunft mit 47:37:27:0 Belegen (gesprochener Input) bzw. 37:9:2:1 Belegen (getippter Input)

und bei der Bibliotheksauskunft mit 38:19:0:1 Belegen (gesprochener Input) bzw. 18:2:5:2 Belegen (getippter Input) vor.

4. Fazit

Die Ergebnisse des Projekts DICOS zeigen deutlich, daß Sprache im Mensch-Maschine-Dialog anders gebraucht wird als im Mensch-Mensch-Dialog. Benutzer verändern ihr Sprachverhalten gegenüber dem Dialogpartner Computer entweder, um ihm das Verstehen zu erleichtern (z.B. Einschränkung der syntaktischen Vielfalt oder gegenüber dem normalen Sprachgebrauch Verwendung abweichender oder überspezifizierter Strukturen) oder weil sie ihm hinsichtlich einiger Faktoren als inkompetenten Dialogpartner einstufen (z.B. hinsichtlich der Fähigkeit sprachliche Fehlleistungen auszufiltern und auch hinsichtlich der pragmatischen Kompetenz).

Benutzervorstellungen über Deutlichkeit der Eingabe und Erleichterungen des Sprachverstehens sind nur schwer vorhersehbar. Auch müssen sie nicht immer tatsächlich eine Erleichterung für den Sprachanalysealgorithmus darstellen. Es ist offensichtlich, daß Introspektion und direkte Ableitung aus der Mensch-Mensch-Kommunikation nur sehr unzureichende Mittel sind, die Leistungsanforderungen an Spachanalysealgorithmen für spezifische Informationssysteme zu beschreiben. Es bedarf einer soliden empirischen Absicherung, die über die intuitive ad-hoc Setzung von Beispielsätzen hinausgeht. Nur so kann ein Subset der natürlichen Sprache festgelegt werden, das auf der einen Seite die Fähigkeiten der heutigen Spacherkenner und linguistischen Analysealgorithmen nicht überfordert, auf der anderen Seite aber natürlich genug ist, um den Benutzer in seinen Formulierungsmöglichkeiten kaum einzuschränken. Unter diesen vorgegebenen Bedingungen wird der Benutzer sich auf diesen Subset von sich aus einstellen, ohne ihn eigens erlernen zu müssen.

Literatur

- /1/ Capellmann, C., Franzke, A. Krause, J. (1988): Materialien zum Computertalk. Microficheausgabe. MCS Verlag. Nürnberg
- /2/ Falzon, P., Amalberti, R., Carbonell, N. (1988): Man-Machine Voice Dialogue: The Role of Interlocutor Models. In: Man-Machine Systems. Analyses, Design and Evaluation. Proceedings of IFIP/IFAC. Oulu, Finnland, 14.-16. Juni, 511-516
- /3/ Morel, M.-A. (1988) (Ed.): Analyse Linguistique d'un Corpus de Dialogues Homme/Machine. Paris: Université de Paris. Centre de Recherche en Morphosyntaxe du Français Contemporain.
- /4/ Kritzenberger, H. (1990): Eigengesetzlichkeiten der MMI. Empirische Untersuchungen der Protokolle des DICOS-Bahnauskunftssystems. DICOS-Arbeitspapier No. 10. Regensburg, masch.
- /5/ Kritzenberger, H. (1990): Zustandsbeschreibung der Benutzereingaben zum DICOS-Bahnauskunftssystem. Auswertung und Erweiterung der Merkmalsbeschreibungen zum Computertalk. DICOS-Arbeitspapier No.11. Regensburg, masch.
- /6/ Kritzenberger, H. (1990): Empirische Untersuchungen der Protokolle des DICOS-Bibliotheksauskunftssystems. DICOS-Arbeitspapier No. 12. Regensburg, masch.
- /7/ Kritzenberger, H. (1989): Aufbau und Beschreibung der Simulationssysteme zum DICOS-Bahnauskunftssystem. DICOS-Arbeitspapier No. 4. Regensburg, masch.
- /8/ Krischker, S. (1989): Versuchsbericht DICOS. DICOS-Arbeitspapier No. 6. Regensburg, masch.
- /9/ Hitzenberger, L., Kritzenberger, H. (1989): Simulation Experiments and Prototyping of User Interfaces in a Multimedial Environment of an Information System. In: Tubach, J.P., Mariani, J.J. (Eds.): Eurospeech 89. European Conference on Speech Communication and Technology. Paris, Sept. 1989, 597-600
- /10/ Chapanis, A. et al. (1977): Studies in Interactive Communication: II. The Effects of Four Communication Modes on the Linguistic Performances of Teams during Cooperative Problem Solving. In: Human Factors 19 (2), 101-126
- /11/ Zoltan, E., Weeks, G.D., Ford, W.R. (1982): Natural-Language Communication with Computers: A Comparison of Voice and Keyboard Input. In: Johannsen, G., Rijnsdorp, J.E. (Eds): Analysis, Design and Evaluation of Man-Machine-Systems. Proceedings of the IFAC/IFIP/IFORS/IEA Conference, Baden-Baden, 27-29 September 1982. Oxford:Pergamon Press, 255-260.