

Mensch-Computer-Interaktion über gesprochenen Dialog

Felix Wiedemann

Proseminar Künstliche Intelligenz 2006

Universität Ulm

Abstract

Sprache gilt als das natürlichste Mensch-Computer Interface. Bisher wird es nur in sehr rudimentärer Form eingesetzt. Ein großes Ziel der Forschung in diesem Bereich ist es, ein Dialogsystem zu entwickeln, über das der Mensch mit dem Computer wie mit einem menschlichen Gegenüber kommunizieren kann. Welche Schwierigkeiten bei der Entwicklung eines solchen Systems zu überwinden sind, und welche Rolle künstliche Intelligenz dabei spielt, wird im Folgenden erörtert.

1. Einleitung

2. Sprache als Computer-Interface

2.1 Historische Entwicklung

2.2 Spracherkennung im Einsatz heute

2.2.1 Textverarbeitung

2.2.2 Voice Portals

2.2.3 Künstliche Intelligenz bei Textverarbeitung und Voice Portals

2.3 Dialogsysteme

2.3.1 Was ist ein Dialogsystem

2.3.2 Vergleich Dialogsysteme – GUIs

2.3.3 Praktischer Dialog

2.3.4 Künstliche Intelligenz in Dialogsystemen

3. TRIPS (Beispielsystem)

3.1 Was ist TRIPS

3.2 Systemarchitektur

4. Ausblick

1. Einleitung

Die Mensch-Computer-Interaktion über den gesprochenen Dialog ist ein beliebtes Thema in der Science Fiction. Der Kapitän der Enterprise spricht mit seinem Computer als wäre dieser ein Mensch, und der Computer antwortet so als wäre er ein Mensch. Dieser Computer wirkt auf uns viel “intelligenter” als unser Heimcomputer. Bis auch unser Heimcomputer unsere Sprache spricht ist es sicherlich noch ein weiter Weg, doch die großen Fortschritte der Spracherkennungstechnologie und die rasant steigende Leistungsfähigkeit der Computer könnten solche Systeme möglich machen. Doch bessere Technik allein wird nicht reichen – auch der geschickte Einsatz von künstlicher Intelligenz wird notwendig sein.

2. Sprache als Computer-Interface

2.1 Historische Entwicklung

Bereits in den 1960er Jahren wurden von privaten Firmen erste Forschungen zum Thema Spracherkennung durchgeführt. Erste Systeme waren allerdings nicht in der Lage mehr als einige hundert Worte zu erkennen. Weitere Fortschritte gab es erst in den 1980er Jahren als IBM ein System präsentierte, das 5000 englische Einzelworte erkennen konnte. Allerdings war dieses System nur über einen Großrechner zu nutzen und sehr langsam.

In den 1990er Jahren waren bereits Systeme in der Entwicklung, die bis zu 30000 deutsche Wörter erkennen konnten. Das „IBM VoiceType Diktiersystem“ zum Beispiel war für den Privatanwender erschwinglich und lief auf einem normalen PC.

2.2 Spracherkennung heute

Heute wird Spracherkennung vor allem in zwei Bereichen angewendet: in der Textverarbeitung und bei sogenannten Voice Portals.

2.2.1 Textverarbeitung

Der Anwender diktiert dem System einen Text über das Mikrofon. Die Spracherkennung ist hier also in ein Textverarbeitungsprogramm (wie zum Beispiel Microsoft Word) integriert. Im Idealfall muss der User keinen Text mehr per Tastatur eingeben. Vor allem für weniger geübte User bringt das eine erhebliche Zeitersparnis. Derzeit gibt es zwei unterschiedliche Ansätze:

- „sprecherabhängige Spracherkennung“: Der Benutzer „trainiert“ das System vor dem Einsatz. Das heißt, er spricht über Mikrofon verschiedene Worte ein um das System auf die Eigenarten seiner Stimme und Aussprache zu kalibrieren. Ein solches System kann nicht mit Erfolg von mehreren häufig wechselnden Usern verwendet werden.
- „sprecherunabhängige Spracherkennung“: Jeder Benutzer kann sofort ohne Trainingsphase mit der Spracherkennung beginnen. Allerdings ist der Wortschatz vergleichsweise begrenzt und die Wahrscheinlichkeit von Fehlern ist höher.

2.2.2. Voice Portals

Über ein Voice Portal können Anrufer über Telefonanlagen Informationen erfragen. Die rudimentäre Form eines Voice Portals wird auch IVR (Interactive Voice Response) genannt. Bei einem IVR handelt es sich um eine simple Einzelworterkennung. Der Dialog läuft automatisiert ab, dem User wird nur eine kleine Anzahl von Antwortmöglichkeiten vorgegeben. Nur wenn er eines dieser

Schlüsselwörter spricht kann das System fortfahren. Beispiel: „*Um ein Produkt zu kaufen sagen sie bitte „Verkauf“, haben sie eine Frage zu einem Produkt sagen sie bitte „Service“ ...*“

Durch die extreme Einschränkung der Antwortmöglichkeiten wird die Antwort in solchen Systemen einfach vorhersagbar und mit fast hundertprozentiger Sicherheit richtig erkannt. Allerdings werden auch die Interaktionsmöglichkeiten des Benutzers erheblich eingeschränkt.

Komplexere Voice Portals ermöglichen etwa die Abfrage von Aktienkursen oder Fahrplänen über das Telefon. Bei manchen Systemen können die Fragen in ganzen Sätzen formuliert werden.

Natürlich werden bei Voice Portals aufgrund der vielen unterschiedlichen Anrufer sprecherunabhängige Spracherkennungssysteme eingesetzt.

2.2.3 Künstliche Intelligenz bei Textverarbeitung und Voice Portals

Künstliche Intelligenz wird sowohl bei der Textverarbeitung als auch bei Voice Portals nur sehr begrenzt für die Spracherkennung eingesetzt.

Bei der Fehlerkorrektur wird mit Wahrscheinlichkeiten gearbeitet. Ein Wort des Sprechers wird digitalisiert und in das Frequenzspektrum zerlegt. Dieses Signal wird dann mit den Signalen von Wörtern, die in einer Datenbank gespeichert sind, verglichen. Ergibt sich eine Übereinstimmung, so ist das Wort sofort erkannt. Im Normalfall wird es aber Abweichungen zu den Wörtern in der Datenbank geben, da jede Person einen anderen Sprachduktus hat. Jetzt gilt es über Wahrscheinlichkeiten herauszufinden, welches Wort wohl gesagt wurde.

Hierzu ein einfaches Beispiel auf der Wortebene: Eine telefonische Fahrplanauskunft fragt den Benutzer nach dem Zielbahnhof. Der Benutzer antwortet undeutlich, der Computer „versteht“ das Wort „*Müchen*“. Dieses Wort findet sich nicht in der Datenbank, es unterscheidet sich aber nur durch einen fehlenden Konsonanten vom Wort „*München*“. Also interpretiert der Computer das Wort richtigerweise als „*München*“ und sucht nach Reisemöglichkeiten zu diesem Zielbahnhof.

2.3 Dialogsysteme

2.3.1 Was ist ein Dialogsystem ?

In einem Dialogsystem basiert die Mensch-Computer-Interaktion auf der menschlichen Sprache. Der Mensch spricht mit dem Computer wie mit einem anderen Menschen, mit dem er gemeinsam ein Problem lösen will. Das heißt der Mensch muss keine vorgefertigten Formeln aussprechen, sondern kann seine Fragen und Wünsche in seinen eigenen Worten formulieren. Auch der Computer kann in das Gespräch eingreifen und zum Beispiel noch benötigte Informationen erfragen. Keine Partei kontrolliert den gesamten Dialog komplett.

Wichtig ist, dass ein solches System nicht einfach nur ein gesprochenes Navigieren durch Menüs darstellt.

Ein solches System wäre das natürlichste und intuitivste Computerinterface. Im Idealfall könnte ein Benutzer es ohne Einarbeitung und Lernphase sofort benutzen und damit komplexe Aufgaben bewältigen.

Im Moment sind solche Systeme Zukunftsmusik, es gilt noch einige Probleme zu überwinden, aber durch immer leistungsfähigere Computer und die Verbesserung der Spracherkennungstechnologien scheinen solche Systeme möglich zu werden.

2.3.2 Vergleich Dialogsysteme –GUIs

Es stellt sich die Frage ob es entscheidende Vorteile von Dialogsystemen gegenüber Graphical User Interfaces (GUIs) gibt. GUIs sind im Moment das als Standard verwendete Mensch-Computer Interface. Allerdings wird es in Zukunft mehr und mehr Systeme geben, für die ein GUI nicht realisierbar ist, da das Gerät zu klein ist oder der Benutzer die Augen und/oder Hände für andere Aufgaben braucht ([1]). In solchen Fällen ist ein Dialogsystem eine gute Lösung.

Aber auch wenn ein GUI möglich ist, kann ein Dialogsystem eine sinnvolle Alternative darstellen. Sehr einfache und schnell erlernbare GUIs ermöglichen keine komplexeren Aufgaben. Sobald die zu erledigen Aufgaben sehr komplex werden, wird der Bildschirm mit vielen Optionen, Menüs und Untermenüs schnell unübersichtlich. Ein zeitintensives Training ist für den Benutzer unerlässlich.

Bei einem GUI muss sich ein Benutzer über mehrere Untermenüs zu einer Funktion (zum Beispiel Entsättigen eines Fotos in einem Grafikprogramm im Untermenü „Filter“ im Menü „Bild Bearbeiten“) durchklicken, bei einem Dialogsystem könnte er dem Computer einfach nur sein Ziel nennen („entsättige das Foto“). Dieser kann dann ggf. durch Nachfragen das Problem spezifizieren und lösen. Der User muss also nicht genau wissen wie das Problem letztendlich Schritt für Schritt zu lösen ist – es reicht wenn er sein Ziel formuliert. So sind Probleme einfacher und schneller zu lösen. Auch Lösungen, bei denen GUIs und Dialogsysteme kombiniert werden, sind denkbar. So könnte man so die Stärken beider Interfaces sinnvoll nutzen.

2.3.3 Praktischer Dialog

Die menschliche Sprache ist in ihrem gesamten Spektrum viel zu komplex, um in naher Zukunft von einem Computer komplett „verstanden“ zu werden. Deshalb schränkt man den Dialog in Dialogsystemen auf den sogenannten praktischen Dialog ein. Praktischer Dialog wird definiert als der Dialog, der darauf abzielt ein konkretes Ziel zu erreichen ([1]). Diese Einschränkung hat für den Benutzer kaum negative Folgen, da er mit Hilfe eines solchen Systems immer ein Problem lösen will, also ein Ziel erreichen will. Innerhalb dieses Rahmens kann er seine Fragen und Ziele in seinen eigenen Worten formulieren.

Abgesehen von einer geringeren Komplexität, hat der praktische Dialog noch einen weiteren Vorteil: Man geht davon aus, dass der praktische Dialog in den verschiedensten Anwendungsgebieten dieselben Grundstrukturen aufweist, auch wenn der Dialog oberflächlich betrachtet je nach Anwendungsgebiet ganz unterschiedlich abläuft ([1]). Diese Ähnlichkeit in der Grundstruktur könnte man sich zu Nutze machen, um ein anwendungsunabhängiges Dialogsystem zu entwickeln, dass dann durch Hinzufügen einzelner Module an die jeweilige spezifische Aufgabe angepasst wird.

2.3.4 Künstliche Intelligenz bei Dialogsystemen

Bei der Entwicklung eines Dialogsystems müssen einige Herausforderungen bewältigt werden.

Um mit der Komplexität der Sprache besser umgehen zu können werden semantische und anwendungsspezifische Restriktionen verwendet. Semantische Restriktionen schließen bestimmte Bedeutungen von mehrdeutigen Wörtern aus, indem andere Wörter des Satzes analysiert werden.

Beispiel für semantische Restriktion: Man könnte als Restriktion implementieren, dass nach dem englischen Wort „eat“ nur essbare Objekte folgen. Der Computer kann also in dem Satz „*He eats chips*“ ausschließen, dass es sich um Computerchips handelt. Gerade im praktischen Dialog lassen sich oft sinnvolle semantische Restriktionen finden.

Anwendungsspezifische Restriktionen hingegen schließen bestimmte Bedeutungen von mehrdeutigen Wörtern aus, indem Rücksicht auf das Anwendungsgebiet genommen wird.

Beispiel für eine anwendungsspezifische Restriktion: Ein System, das Nahrungsmitteltransporte zu verschiedenen Supermärkten organisiert, kann bei der Bestellung „*send chips to Chesterfield*“ ausschließen, dass es sich um Computerchips handelt, da es nur im Bereich Nahrungsmittel arbeitet.

Eine große Herausforderung für ein Dialogsystem ist es auch die Absicht des Benutzers zu verstehen. Das System muss die Semantik des Gesagten richtig deuten. Hier versagen die wahrscheinlichkeitsbasierten Methoden derzeitiger Spracherkennungssysteme.

Ein Beispiel: Der Benutzer sagt: „*Can we use a helicopter to get people from Delta?*“ Es ist hier ohne Kontext nicht klar, ob der Benutzer eine Planänderung vornimmt (also statt eines Trucks einen Helikopter senden will) oder ob er nur nach der Machbarkeit dieser Möglichkeit fragt. Bei Interpretation 1 könnte das System mit „*ok*“ antworten und den Plan ändern. Bei Interpretation 2 wäre eine angemessene Antwort. „*Yes, a helicopter would be available and that would save us 10 hours*“.

Das Beispiel zeigt, dass es unterschiedliche Interpretationsmöglichkeiten zu einer einzigen Äußerung des Benutzers gibt. Um die richtige Interpretation zu finden muss

das System herausfinden welche unter den gegebenen Umständen und in der jetzigen Situation rational sinnvoll ist. Hier ist also künstliche Intelligenz im Einsatz. Eine Methodik, die nur die Syntax der Sprache analysiert, kann hier nicht zum Ziel führen. Eine mögliche Herangehensweise an eine solche Problematik besteht aus zwei Phasen ([3]): In der ersten Phase werden über den Abgleich von gespeicherten Regeln mit der Eingabe verschiedene mögliche Absichten unterschieden. In der zweiten Phase werden diese Möglichkeiten dann unter Berücksichtigung der aktuellen Situation und des bisherigen Verlaufs des Gespräches untersucht. So können insbesondere die Möglichkeiten, die in dieser Situation keinen Sinn machen, aussortiert werden. Beispielsweise würde ein Benutzer nicht ein Ziel erneut nennen wenn es vorher bereits erreicht wurde.

Eine weitere große Herausforderung bei der Entwicklung von Dialogsystemen ist es, den sogenannten „mixed-initiative dialogue“ umzusetzen. Ein praktischer Dialog ist ein „mixed-initiative dialogue“, das heißt dass sich der Dialog dynamisch zwischen zwei Parteien entwickelt, die beide zum Dialog beitragen und ihn vorantreiben bis die Bedürfnisse beider Parteien erfüllt sind. Das Gegenteil dazu wäre der „fixed initiative dialogue“, bei dem eine Partei die gesamte Interaktion kontrolliert. Voice Portals basieren auf „fixed initiative dialogue“ da der Benutzer immer genau die Frage beantworten muss, die das System gerade gestellt hat. Solche Systeme sind sinnvoll für einfache Aufgaben, die immer genau demselben Muster folgen. Sie bieten aber kaum die Flexibilität die ein Dialogsystem benötigt.

Für ein flexibles Dialogsystem ist demnach „mixed initiative dialogue“ zu bevorzugen. Sowohl der Benutzer als auch der Computer haben Bedürfnisse. Der Benutzer äußert einen Plan. Benötigt der Computer jetzt noch eine bestimmte spezifische Information um fortfahren zu können, so fragt er danach. Das System sollte erkennen welche Informationen es genau in diesem Moment benötigt und nur diese dann auch erfragen. So wird verhindert, dass der Benutzer eine automatisierte Kette von allen möglichen Informationen eingeben muss, die im Moment zum Teil gar nicht relevant sind. So kann Zeit gespart werden.

Bei einem „mixed initiative dialogue“ System kann der Computer den Verlauf des Dialogs aber auch noch drastischer ändern, indem er vielleicht selber ein neues, im Moment dringenderes Problem, vorstellt.

Ein Beispiel: Es geht um ein System, das Rettungseinsätze koordiniert.

Benutzer: „*Wie weit sind die Räumungsarbeiten nach dem Unfall auf der Landstraße 8?*“

System: „*Einen Moment. Ich erhalte gerade Meldung dass es eine Massenkarambolage auf der Autobahn gibt.*“

Hier wird der Benutzer zwar unterbrochen, aber es ist durchaus in seinem Interesse die dringendere Nachricht sofort zu erhalten. Bei der Massenkarambolage besteht sofortiger Handlungsbedarf, während die Räumungsarbeiten auf der Landstraße ja schon eingeleitet sind und es nur noch um die Statusabfrage geht. Das System muss in einem solchen Fall also abwägen, ob es wichtiger ist der Initiative des Users zu folgen oder die neue Notfallmeldung vornanzustellen. Ein solches Abwägen erfordert künstliche Intelligenz, die teilweise sehr komplexe Aufgaben zu lösen hat, beispielsweise wenn mehrere Notfallmeldungen gleichzeitig registriert werden.

Die obigen Ausführungen belegen, dass künstliche Intelligenz in verschiedenen Bereichen von Dialogsystemen sinnvoll eingesetzt werden kann.

3. TRIPS

3.1 Was ist TRIPS ?

TRIPS steht für „The Rochester Interactive Planning System“ und wird an der Universität von Rochester in den USA entwickelt. Es handelt sich um einen Prototypen eines Dialogsystems, das mit einem menschlichen Manager bei der Erstellung von Notfallplänen in Krisensituationen zusammenarbeitet ([2]). Es gibt bereits eine lauffähige Version, die in einem Beispielszenario einsetzbar ist.

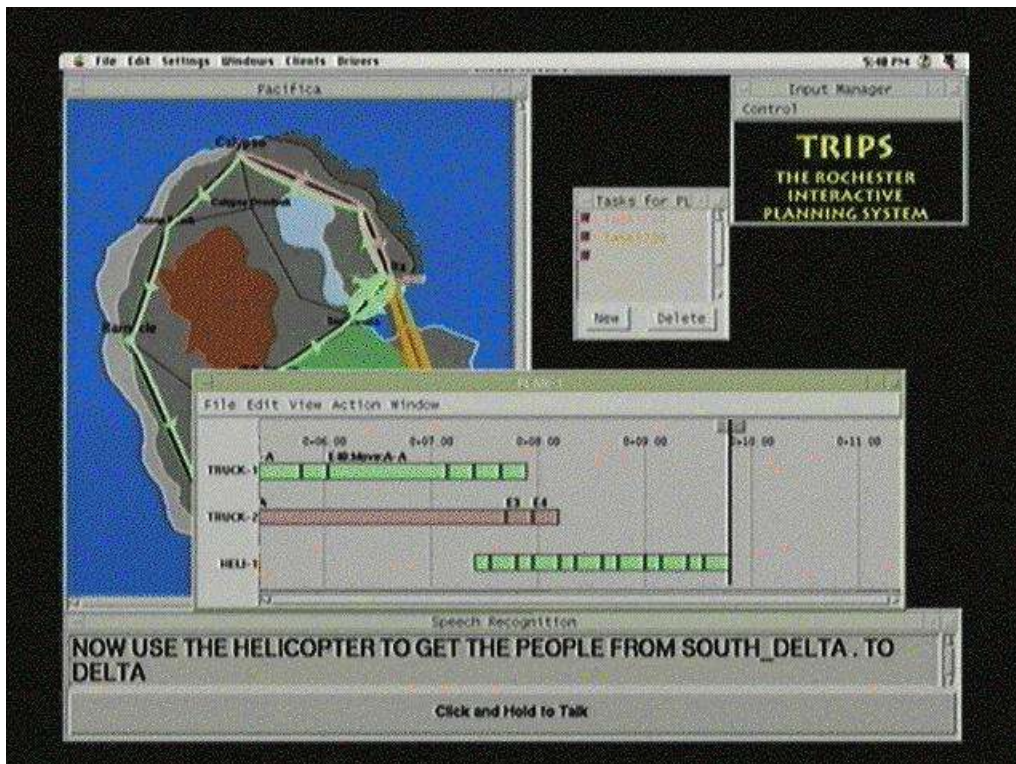
Das Beispielszenario: Ein Hurricane bedroht die Insel Pazifika. Deshalb müssen die Bewohner der verschiedenen Städte evakuiert werden und in eine abgelegene Stadt in Sicherheit gebracht werden. Die logistische Planung dieser Evakuierung wird von einem menschlichen Manager zusammen mit TRIPS durchgeführt. Dabei hat er eine begrenzte Anzahl verschiedener Fahrzeuge und Hubschrauber zur Verfügung. Die

Wahl des endgültigen Planes hängt von verschiedenen Faktoren ab, wie zum Beispiel Zeitaufwand, Kosten, Wetter usw .

Vom wissenschaftlichen Standpunkt aus gesehen bietet dieses Beispielszenario drei klare Vorteile:

1. Es ist klar was das Ziel ist und wann es erreicht wurde.
2. Die Effizienz verschiedener Lösungen kann leicht gemessen werden indem man Faktoren wie zum Beispiel die benötigte Zeit vergleicht.
3. Es ist einfach die Komplexität des Szenarios zu verändern indem man zum Beispiel die Anzahl der Städte oder das Straßennetz verändert.

So lässt sich dieses Beispielszenario leicht in verschiedenen Komplexitätsstufen wissenschaftlich analysieren. TRIPS ist also ein geeigneter Prototyp für Dialogsysteme. Anhand der Forschungsergebnisse lassen sich auch Rückschlüsse ziehen, die für Dialogsysteme im Allgemeinen gelten.



Screenshot der TRIPS Benutzeroberfläche: Karte der Insel und Status der Transportfahrzeuge

3.2 Systemarchitektur

TRIPS besteht aus einer Vielzahl von Modulen und Agenten. Durch Veränderungen an einzelnen Modulen lässt sich das System theoretisch an andere Einsatzgebiete anpassen. Die Komponenten von TRIPS können in 3 übergeordnete Gruppen eingeteilt werden.

1. modality processing: Hier findet die Interaktion (Input und Output) mit dem Benutzer statt. Dies beinhaltet Spracherkennung und Sprachgenerierung, graphische Anzeigen, Texteingaben usw.
2. dialogue management: Diese Komponenten bilden den Kern von TRIPS. Hier wird der gesprochene Dialog mit dem Benutzer im Kontext analysiert, die Semantik untersucht, die Bedürfnisse des Systems in Bezug auf die Problemlösung erfragt und die Antworten des Systems generiert. Hier werden die sogenannten „specialized reasoners“ koordiniert. Die wichtigsten Komponenten sind der „conversational agent“ und der „problem solving manager“. Vor allem hier kommt die künstliche Intelligenz (wie in 2.3.4 beschrieben) zum Einsatz.
3. specialized reasoners: Diese Module sind speziell an das Aufgabenfeld des Systems angepasst. Hier werden die eigentlichen Rechenoperationen zur Lösung der Subprobleme bearbeitet, zum Beispiel die Berechnung von Routen oder das Erstellen von Zeitplänen. Es reicht diese Komponenten auszutauschen um das System für ein ganz anderes Aufgabenfeld nutzbar zu machen.

4. Ausblick

Wie bei jeder neuen technischen Entwicklung sollte man sich auch bei Dialogsystemen fragen, ob solche Systeme die Arbeit wirklich erleichtern und verbessern können, oder ob hier die Technik nur um der Technik willen erforscht wird. Würde ein Logistiker, der eine Evakuierung planen muss, lieber mit einem Dialogsystem arbeiten oder doch mit einem Grafischen Interface? Diese Frage ist im Moment noch schwer zu beantworten, da heute noch nicht ersichtlich ist, ob die hochgesteckten Ziele der Forschung wirklich in Zukunft erreicht werden können.

Sollte aber die Vision von einem voll funktionalen Dialogsystem, das schnell und fehlerfrei läuft, Realität werden, so würde die Mensch-Computer Interaktion revolutioniert werden. Die Kommunikation zwischen Mensch und Computer würde unmittelbarer, direkter, einfacher und schneller werden. Und dies wäre sicher auch im Sinne eines Logistikexperten.

Referenzen:

- [1]. J. F. Allen, D. K. Byron, M. Dzikovska, G. Ferguson, L. Galescu, A. Stent 2001; „Towards Conversational Human-Computer Interaction“. *AI Magazine 2001*
- [2]. G. Ferguson, J. F. Allen; „TRIPS: An Integrated Intelligent Problem-Solving Assistant“; Department of Computer Science, University of Rochester, Rochester, NY.
- [3]. E. Hinkelman, J. F. Allen, „Two Constraints on Speech Act Ambiguity“, *Proc. of the Association for Computational Linguistics (ACL)*, Vancouver, Canada, 1989.