

Kapitel 6

Zusammenfassung und Ausblick

Das menschliche Sprachverstehen arbeitet inkrementell.

Diese Beobachtung bildete den Ausgangspunkt der vorliegenden Arbeit. Wir haben versucht darzulegen, daß die Anwendung analoger Prinzipien in der maschinellen Sprachverarbeitung ebenso wünschenswert wie sinnvoll ist. Sie ist wünschenswert, da erst durch die durchgängige Bearbeitung von mehr oder weniger kleinen Teilen der Eingabe fortgeschrittene Systeme zur Mensch-Maschine-Kommunikation oder zur Unterstützung der Kommunikation zwischen Menschen möglich sind. Als Beispiele hierfür mögen erneut Dialogsysteme dienen, die — genau wie Menschen dies untereinander zu tun pflegen — den Partner in wichtigen Fällen unterbrechen können sollten, oder aber Programmsysteme, die eine Simultanübersetzung gesprochener Sprache anstreben.

Sowohl die Leistungsfähigkeit von Computern allgemein, als auch die von Spracherkennungssystemen dringen seit einiger Zeit in Bereiche vor, die eine durchgängig inkrementelle Arbeitsweise möglich machen. Es ist heutzutage durchführbar, eine Menge von Workstations zusammenzuschalten, um mit Hilfe ihrer kombinierten Rechenleistung den Mehraufwand, den die Einführung inkrementeller Verfahren zunächst mit sich bringt, auf ein erträgliches Maß zu reduzieren. Unsere Experimente zeigen, daß eine inkrementelle Vorgehensweise um etwa eine Größenordnung schwieriger ist als vergleichbare nichtinkrementelle Verfahren. Trotzdem liegt die Verarbeitungszeit um lediglich einen geschätzten Faktor drei über der Analysezeit hoch optimierter konventioneller Ansätze.

Die Genauigkeit spracherkennender Systeme auf der anderen Seite ist in den letzten Jahren so gut geworden, daß immer mehr spontan gesprochene Sprache als Eingabe benutzt wird. Die im Rahmen des Projektes Verbmobil aufgenommenen Dialoge stellen dabei jedoch lediglich einen Meilenstein dar, da sie auf einen engen Themenbereich, die Terminvereinbarung, eingeschränkt sind. Die Erkennungsqualität, die den Eingaben in unser System zugrunde liegt, ist mit ca. 70% Wortakkuratheit deutlich geringer als die momentan besten Werte für Offline-Erkennen im Verbmobilkontext. Sie entspricht in etwa den Ergebnissen, die man für weitgehend unrestringierte Sprache, z.B. beim Telefonieren, erwarten könnte. Eine derart hohe Wortfehlerrate wird durch die Benutzung von Wortgraphen relativiert, durch die eine kompakte Darstellung großer Anzahlen von Alternativen möglich

ist. Unter diesen kann dann im Rahmen einer linguistischen Verarbeitung eine gezielte Suche nach plausiblen Interpretationen durchgeführt werden.

Die inkrementelle Art des menschlichen Sprachverstehens zeigt sich auf nahezu allen Ebenen. Wir haben versucht, dies zumindest für die Analyse in den Bereichen Worterkennung sowie der syntaktischen und semantischen Präferenzmodellierung anhand psycholinguistischer Forschung zu demonstrieren. Ebenso "profitiert" ein Konferenzdolmetscher eminent von der Inkrementalität, indem sie ihm hilft, quellsprachlichen Äußerungen zuzuhören und doch gleichzeitig dazu in der Lage zu sein, schon früh zielsprachliche Sätze zu formulieren. Inkrementalität ist offensichtlich in mindestens dreifacher Hinsicht unabdingbar für die menschliche Sprachperformanz. Zuerst ermöglicht sie die gleichzeitige Verarbeitung nahezu identischer Äußerungsausschnitte auf verschiedenen Ebenen der Analyse. Darüberhinaus können Ergebnisse auf höheren linguistischen Ebenen zur Blockierung unwahrscheinlicher Alternativen in anderen Bereichen der Verstehensprozesses dienen. Und drittens schließlich findet eine Rückkopplung mittels Prädiktionen statt, die erwartungsgesteuert auf einige Bereiche des Sprachverstehensapparates wirkt.

Die Untersuchung der Möglichkeit einer durchgängig inkrementellen Analyse im Rahmen des Dolmetschens spontan gesprochener Sprache bildet folglich die Hauptmotivation zur Anfertigung dieser Arbeit.

Dabei sollte allerdings vermieden werden, ein allzu statisches System zu erstellen, das in der Zukunft nur mit hohem Aufwand modifizierbar ist. Vielmehr war das Ziel, neben einer qualitativen und groben quantitativen Untersuchung inkrementeller Algorithmen innerhalb einer vollständigen Anwendung gleichzeitig auch ein offenes, dynamisches System zu konzipieren, das als Grundlage für weitergehende architektonische Studien im Rahmen inkrementeller Sprachverarbeitung dienen kann. Insbesondere im Bereich der Rückkopplungen zwischen unterschiedlichen Modulen weiß man bisher viel zu wenig, als daß sich bereits eine allgemein akzeptierte architektonische Ausrichtung hätte ausprägen können. Wir haben folglich mit der Implementierung die Maxime verfolgt, sowenig Einschränkungen für das Design alternativer Module wie möglich aufzuprägen.

Die wesentlichen Beiträge dieser Arbeit sehen wir vor allem in:

- Der Charakterisierung wesentlicher Eigenschaften von Wortgraphen sowie der Einführung von Algorithmen, die eine Bearbeitung auch umfangreicher Graphen innerhalb der automatischen Sprachverarbeitung ermöglichen. Neben einer veränderten Sicht auf die Beurteilung der Größe von Graphen, die nicht ausschließlich erkenntniszentriert ist, sondern sich an der Verwendung der Graphen in einem größeren Zusammenhang orientiert, sind hier außerdem Methoden zur Reduktion der Komplexität von Wortgraphen zu nennen. Für die Verarbeitung inkrementeller Graphen schließlich ist die theoretische Aufarbeitung und praktische Anwendung von Hypergraphen-Mechanismen ausgesprochen nützlich, hilft sie doch, bislang außerhalb der Reichweite aktueller Verarbeitungsmethoden liegende Eingaben auf ein Größenmaß zu reduzieren, das in vielen Fällen eine erfolgreiche Benutzung erlaubt.
- Die Konzeption eines Architekturschemas, das umfangreich und zugleich offen genug ist, um in der Zukunft Forschungsarbeiten im Gebiet der inkrementellen, interaktiven Sprachver-

beitung zu ermöglichen. Dieses Architekturschema besteht im wesentlichen aus drei Komponenten, welche die wichtigen Bereiche Datenstrukturen, Repräsentationen und Systemintegration abdecken. Auf der Ebene der Datenstrukturen haben wir die Mehr-Ebenen-Chart eingeführt, eine weitgehende Verallgemeinerung bislang üblicher Charts. Mehr-Ebenen-Charts sind zur verteilten, synchronisationsarmen Speicherung mannigfaltiger Aspekte einer Eingabeäußerung in der Lage. Sie sind unabhängig von einer zentralen Speicherung oder Kontrolle und sorgen durch die integrierte Darstellung aller Hypothesen innerhalb eines Systems für eine unkomplizierte Interaktion von Komponenten. Jedweder Austausch von Information findet über Hyperkanten der Chart statt, so daß keine Schnittstellenproblematik auftritt. Zugleich stellt die Vereinigung aller Kanten stets ein kohärentes Bild des aktuellen Bearbeitungsstandes eines Systems dar. Mehr-Ebenen-Charts dehnen die bisher übliche Form der graphenartigen Datenhaltung für Worthypothesen auf alle Ebenen eines Systems aus, sie sorgen mithin für die Etablierung eines integrierten Interpretationsgraphen für eine Anwendung.

Auf der Ebene der Repräsentationen ist durch die Konzeption und Realisierung eines verschiebungsinvarianten getypten Merkmalformalismus mit Zulässigkeitsfunktion eine hocheffiziente, deklarative Sprache zur Beschreibung linguistischer Objekte entstanden. Diese Sprache erlaubt — wie einige ihrer Vorgänger — eine graphenartige Spezifikation getypter Merkmalstrukturen. Der Ausgangspunkt für die Entwicklung des in der vorliegenden Arbeit benutzten Formalismus war die Ansicht, daß für hochgradig verteilte, parallele Systeme — womit wir sowohl die intermodulare wie auch die intramodulare Parallelität meinen — eine stark verzeigerte Speicheraufteilung herkömmlicher Ansätze ungeeignet ist. Stattdessen wurde hier eine Implementierung verfolgt, die eine blockorientierte Allokation vorsieht. Sie ist extrem kompakt und erlaubt den Transport von Merkmalstrukturen in andere Adreßräume ohne aufwendige Linearisierung und Rekonstruktion. Der Formalismus ist allen Komponenten direkt zugänglich und realisiert eine uniforme Darstellung unterschiedlichster Arten linguistischen Wissens.

Auf der Ebene der Systemintegration schließlich ist ein Architekturschema zur Erstellung heterogener Systeme von Komponenten entstanden, das eine einheitliche Sicht auf die Kommunikation innerhalb eines Systems zuläßt. Der kanalorientierte Entwurf des Infrastruktursystems ICE garantiert einen hochgradig effizienten und konzeptionell geradlinigen Informationsaustausch. Die hier entwickelte Klasse von Anwendungen besteht aus weitgehend unabhängigen, heterogenen Modulen, die ohne Rückgriff auf eine zentrale Instanz miteinander zusammenwirken können. Eine flexible Konfiguration erlaubt die weitgehende Kontrolle über die tatsächliche Topologie eines Systems, ohne auf die Arbeitsweise einzelner Module übermäßigen Einfluß zu nehmen. Die Übermittlung getypter Nachrichten befreit den Entwickler von der Notwendigkeit, ein hohes Maß an Aufmerksamkeit auf infrastrukturelle Belange zu verwenden. Der Ablauf einer individuellen Komponente erfolgt hierbei datengetrieben; die verteilte Überwachung der Terminierung wird automatisch vom architektonischen Rahmen geleistet.

- Der vollständigen Implementierung eines Dolmetschsystems zur Übersetzung spontan gesprochener deutscher Äußerungen in englische Oberflächenrepräsentationen in der Domäne der Terminvereinbarung. Dieses System (MILC, *Machine Interpreting with Layered Charts*)

ist innerhalb des vorher beschriebenen Architekturrahmens verwurzelt. Es erbringt zum einen den Nachweis, daß durchgängig inkrementelles automatisches Dolmetschen unter Beibehaltung des Chart-Paradigmas möglich ist. Jedwede Form der Hypothese wird dabei als potentielle Interpretation eines Ausschnittes der Eingabeäußerung aufgefaßt. Dies gilt für Analyse und Generierung genauso wie für den Transfer, der in dieser Form bisher nicht mit der Unterstützung chartbasierter Verfahren durchgeführt wurde. Die Inkrementalität wird bis in die Konstruktion von englischen Oberflächenhypothesen beibehalten, was sich als sich zeitlich ändernder Strom von größer werdenden Fragmenten äußert. Ebenso wichtig wie die Bereitstellung eines funktionsfähigen, vollständigen Systems ist allerdings die Möglichkeit, an diesem Änderungen vorzunehmen. Exemplarisch wird dies in der vorliegenden Arbeit durch die Integration von Idiomen in den Übersetzungsprozeß gezeigt. Sie ist orthogonal zur Hauptverarbeitungsrichtung angelegt und dient in erster Linie dazu, aufzuzeigen, welche Form und Anwendungsweise zukünftige Komponenten besitzen können. Erkannte Idiome werden im restlichen System bevorzugt behandelt und modifizieren zugleich Suchräume, deren Beschreiten andernfalls unnötigen Aufwand bedeuten würde.

Innerhalb dieses Experimentierrahmens bieten sich zahlreiche Erweiterungsmöglichkeiten an, welche die Effizienz und Bandbreite der abgedeckten Phänomene betreffen. Unmittelbar durchführbar ist dies für die Ausnutzung weiterer im Sprachsignal vorhandener Information, die von der Worterkennung nicht unmittelbar ausgewertet wird. Insbesondere die Integration prosodischer Evidenz zur Bestimmung von Phrasengrenzen und Akzentuierungen wird einen relevanten Einfluß auf die Qualität der Übersetzungen und die Systemperformanz haben. Modusinformation kann zur Disambiguierung dienen. Darüberhinaus erscheint die Etablierung einer Dialogkomponente der nächste logische Schritt, um eine Bearbeitung vollständiger Gespräche durchführen zu können; dies führt nicht nur dazu, daß eine weitere Quelle von Wissen bezüglich der Auflösung von Mehrdeutigkeiten zur Verfügung steht, sondern auch zur Erhöhung der Ausgabequalität durch die Anwendung dialogischer Phänomene. Eine zusätzliche, ebenso einfache wie potentiell lohnende Erweiterung ergibt sich aus der Inkorporation beispielorientierter Übersetzungen. Diese können weitgehend analog zu Idiomen behandelt werden und sorgen im Erfolgsfall für eine extrem schnelle und zuverlässige Verarbeitung zumindest von Teilen der Eingabe.

Aus prinzipiell architektonischen Erwägungen heraus ist eine Kopplung des Worterkenners mit Modulen zur linguistischen Analyse lohnend, um den Einfluß linguistischer Prozesse auf die Worterkennung zu erforschen. Insbesondere gilt dies für die Anwendung von Restriktionen aus der Semantik und der Dialogverarbeitung, die bisher wenig untersucht wurden.

Schließlich sollte versucht werden, eine dynamische Ausweitung von Suchräumen nur dann vorzunehmen, wenn mit der bislang zur Verfügung stehenden Information eine erfolgreiche Analyse nicht möglich ist. Der Ansatz, den wir in dieser Richtung verfolgen, operiert inkrementell stets auf dem besten Pfad durch den bisher besuchten Graphen. Die Bewertungen von Kanten sind so gestaltet, daß für den Fall einer unwahrscheinlichen Analyse auf bisher zurückgehaltene akustische Hypothesen rekurriert werden kann.

Aus einem größeren Zusammenhang betrachtet, ist die Anwendungsbandbreite von Systemen, die

auf Mehr-Ebenen-Charts beruhen bzw. die auf dem in dieser Arbeit entwickelten System MILC aufbauen, sehr groß. Das automatische Dolmetschen stellt nur einen Ausschnitt der Möglichkeiten dar, wenngleich die Übersetzung gesprochener Sprache einen sehr hohen Anspruch an Systeme stellt. Desgleichen sind etwa Zugangsmechanismen zu Informationssystemen (etwa Datenbanken) oder Anwendungen in der natürlichsprachlichen Steuerung (z.B. Robotik) denkbar.

Neben dem dezidierten Einsatz für gesprochene Sprache ist mit dem vorliegenden Modell jedoch auch die Behandlung geschriebener Sprache leicht möglich. Durch die graphenartige Strukturierung aller Hypothesenräume ist die Darstellung von Ambiguität ohne konzeptionellen Mehraufwand erreichbar. Gerade im Bereich von hochgradig modularen Systemen, welche die Behandlung multilingualer Eingaben zum Ziel haben, bietet sich ein einheitliches Architekturschema an, wenn ein Anwendungssystem in kurzer Zeit auf unterschiedliche Quellsprachen adaptiert werden soll.

Eine Architektur, wie sie in dieser Arbeit dargelegt wurde, macht es zugleich möglich, unterschiedliche Aspekte von Information zu speichern und in einem Systemzusammenhang konsistent und effizient nutzbar zu machen. Auf den Bereich der Sprachverarbeitung beschränkt, ist angedeutet worden, wie dialogische Information in einer Anwendung, die äußerungsübergreifende Phänomene modelliert, durch eine einfache Erweiterung genutzt werden kann. In ähnlicher Weise kann prosodisches Wissen als Hypothesen über zeitliche Ausschnitte einer Äußerung repräsentiert werden.

Ein weiterer Blickwinkel macht deutlich, daß der Effekt einer graphenorientierten Darstellungsform nicht nur auf eigentlich sprachinhärente Merkmale restringiert ist. Die Erweiterung auf multimodale Ein- und Ausgaben sowie die Repräsentation von Ereignissen und Objekten sind ebenfalls denkbar. Unmittelbar einsichtig ist dies für alle Arten von Eingabeinformation, die zeitlich und thematisch mit der Spracheingabe im engeren Sinn korreliert sind. Zeigegesten eines Benutzers, seien sie durch Zeigeinstrumente wie der Maus, durch Augen- oder Handbewegungen realisiert, sind in einfacher Weise in den Systemverlauf integrierbar. Sie lassen sich etwa — analog zur Behandlung von Idiomem im vorliegenden System — in einer spezialisierten Komponente erkennen, die als Ausgabe eine Handlungsbeschreibung in Form einer Merkmalstruktur zusammen mit ihrer zeitlichen Ausdehnung liefert. Selbstverständlich müssen die restlichen Komponenten durch eine geeignete Modifikation ihrer Grammatiken dazu in der Lage sein, diese Information auch auszunutzen.

Eine weitere Form eines zusätzlichen Eingabekanals bildet die Auswertung von Lippenbewegungen während des Sprechens. Hier ist die zeitliche und konzeptuelle Assoziation mit der Originaleingabe in Form eines Wortgraphens unmittelbar einleuchtend. Die Information über die sichtbare Stellung der Lippen kann zur Disambiguierung dienen, indem ein Artikulationsmodell für die sich aus dem Wortgraphen ergebenden Wörter aufgestellt und mit der visuellen Eingabe verglichen wird. Auch hier stellen die zusätzlichen Informationen Hypothesen über zeitliche Ausschnitte des Eingabesignals dar.

Komplizierter, aber dennoch prinzipiell möglich, ist die Integration arbiträrer Objekte und ihrer sich zeitlich ändernden Zustände. Ausgangspunkt für ein derartiges Verfahren muß stets eine Ausdehnung der Abdeckung der Mehr-Ebenen-Chart über isolierte Äußerungen hinaus in Richtung auf eine kontinuierliche Repräsentation der Zeitachse sein. Um die Datenmenge dadurch nicht in unhandhabbare Dimensionen wachsen zu lassen, ist zunächst erforderlich, daß ein Fokuskonzept ausgearbeitet

wird. Der Wechsel des Fokus innerhalb des Graphen hat dann zur Folge, daß irrelevant gewordene Kanten entfernt werden und lediglich zentrale Analysen in der Chart verbleiben.

Das zweite Problem betrifft die Einführung einer Methode zur Kombination linguistischen und extralinguistischen Wissens. Dies kann in der vorliegenden Arbeit nicht geleistet werden. Allerdings ist die Art der hier verwendeten Graphenstrukturen außerordentlich gut geeignet, zeitliche Abhängigkeiten durch topologische Eigenschaften von Graphen abzubilden. Im einfachsten Fall lassen sie sich als Schnittoperationen auf Knotenmengen bzw. als Überprüfung der Erreichbarkeitsrelation modellieren.

Die Repräsentation nichtsprachlichen Wissens muß außerdem in den zugrundeliegenden formalen Beschreibungen, also dem verwendeten Typenverband, reflektiert werden. Es ist zu verhindern, daß nichtsprachliche Objekte und Zustandsdarstellungen unkontrolliert mit linguistischen Operationen interagieren. Vielmehr muß die Verarbeitung solchermaßen heterogenen Wissens stets explizit gemacht werden. Das schließt jedoch nicht aus, daß dieselben Mechanismen, wie sie in dieser Arbeit entwickelt wurden, nicht auch auf Kanten anderer Herkunft angewendet werden können, zumal durch die multiplen Ebenen einer Mehr-Ebenen-Chart unterschiedliche Aspekte von Information gut gegeneinander abgeschirmt werden können, ohne dadurch ihre integrierte Verarbeitung zu stören. Um nur ein Beispiel zu nennen, können Objekt- und Handlungsreferenzen in einem Wartungsszenario, in dem ein Benutzer von einem System instruiert wird (wie etwa dem Nachfüllen von Motoröl durch technisch nicht versierte Autofahrer), dadurch stark erleichtert werden. Deiktische Ausdrücke ("Nicht den!", wenn der Benutzer statt des Öldeckels den Deckel des Kühlers öffnen will) lassen sich so einfacher generieren und disambiguieren. In diesem hypothetischen Beispiel war bisher von einem Öldeckel die Rede, die Objektmodellierung kennt jedoch auch andere Deckel innerhalb des Motorraumes. Durch die zeitliche Abhängigkeit der Aufforderung, den Öldeckel zu öffnen und der Handlung des Benutzers kann eine Eingrenzung der Referenz erfolgen und z.B. korrekt auf das eigentlich gemeinte Objekt referiert werden. In diesem Anwendungsfeld zeigt sich außerdem erneut die Notwendigkeit einer inkrementellen, im Idealfall zeitsynchronen, Verarbeitung. Das System sollte beispielsweise dazu in der Lage sein, den Benutzer unmittelbar zu warnen, um Fehler bei der Bedienung zu verhindern.

Alle hier aufgeführten hypothetischen Einsatzbereiche zeigen einige gemeinsame Eigenschaften: Die wesentliche Dimension der Eingabe ist die Zeit. Komplexe Objekte (linguistische Beschreibungen, Repräsentationen von Handlungen) dienen zur Darstellung von Wissen. Der Einsatz inkrementeller Techniken ist notwendig oder zumindest wünschenswert. Eine an Graphen orientierte Datenstruktur wie die Mehr-Ebenen-Chart bildet zusammen mit einem architektonischen Systemumfeld wie dem in dieser Arbeit beschriebenen eine gute Ausgangsbasis zur Untersuchung solcher Phänomene.

It is a curious fact, and one to which no one knows quite how much importance to attach, that something like 85% of all known worlds in the Galaxy, be they primitive or highly advanced, have invented a drink called jynnan tonnyx, or gee-N'N-T'N-ix, or jinond-o-nicks, or any one of a thousand or more variations on the same phonetic theme. The drinks themselves are not the same, and vary between the Sivolvian 'chinanto/mnigs' which is ordinary water served at slightly above room temperature, and the Gagrakackan 'tzjin-anthony-ks' which kills cows at a hundred paces; and in fact the one common factor between all of them, beyond the fact that the names sound the same, is that they were all invented and named *before* the worlds concerned made contact with any other worlds.

What can be made of this fact? It exists in total isolation. As far as any theory of structural linguistics is concerned it is right off the graph, and yet it persists. Old structural linguists get very angry when young structural linguists go on about it. Young structural linguists get deeply excited about it and stay up late at night convinced that they are very close to something of profound importance, and end up becoming old structural linguists before their time, getting very angry with the young ones. Structural linguistics is a bitterly divided and unhappy discipline, and a large number of its practitioners spend too many nights drowning their problems in Ouisghian Zodahts. (Adams, 1980, p. 138)