

Sprachdialogsysteme – nach der pragmatischen Wende

Günther Görz, Universität Erlangen-Nürnberg

1 Sprachdialogsysteme: Zum Stand der Kunst und seinen Wurzeln

Die Computerlinguistik verdankt ihr Entstehen und auch ihre heutige Gestalt vor allem einer aus dem Strukturalismus herrührenden Sichtweise, die man mit "Sprache als System" charakterisieren kann. Vorherrschend ist dabei das Verständnis von "Sprache" als einer Abstraktion im Unterschied zu der anfänglich von Saussure postulierten Gegenständlichkeit der 'langue'. Dieser steht seit der durch Wittgenstein eingeleiteten pragmatischen Wende "Sprache als Handlung" komplementär gegenüber. Diese Wende nachzuvollziehen, steht der Computerlinguistik noch bevor. Schon sind erste Anzeichen zu erkennen und es sprechen nicht nur methodische Gründe, sondern auch praktische dafür – gerade im Zusammenhang mit Sprachdialogsystemen.

1.1 Der traditionelle Ansatz der Computerlinguistik

Die durch den Strukturalismus beförderte Systemsicht der Sprache, genauer: der Sprache als Zeichensystem, und die durch Leibniz konzipierte und durch Frege und andere vollzogene Begründung der formalen Logik führte zur Entfaltung der theoretischen Grundlagen der Computerlinguistik. Auch wenn das Leibniz-Programm mit der *characteristica universalis* als einer universalen Kalkülsprache und dem *calculus ratiocinator* als Konzeption des Logikkalküls auf die Sprache der Wissenschaft gerichtet war, sind in ihm gleichwohl doch die Mittel zur Behandlung natürlicher Sprachen als formale angelegt. Kurz gesagt wird in diesem Ansatz, den wir im folgenden auch *logozentriert* nennen, "Sprechen-Können" ("knowing how") als "Sprach-Wissen" ("knowing that") repräsentiert und die Saussuresche Unterscheidung von *langue* und *parole* findet sich wieder in der von Schema und Aktualisierung und in einem gewissen Sinn auch in der von Kompetenz und Performanz bei Chomsky. Damit dienen formale Kalküle und die mit ihnen begründete operative Logik als Basis der linguistischen Modellierung. Dies umfasst auch den Einsatz sog. subsymbolischer Verfahren wie künstlicher neuronaler Netzwerke, denn diese haben zu einem Teil ihren Ursprung in (symbolischen) semantischen Netzen. Dieser Modellierungsansatz schließt an die Konzeption der Philosophie der idealen Sprache an und arbeitet sie bis zur maschinellen Umsetzung aus. Sie bleibt damit weithin der Einheit Satz, der Wahrheitswertsemantik und der Abbildtheorie verhaftet. In letzterer wird aber die Deutungsproblematik der Kalküle offensichtlich: Die formalen Beziehungen können nicht unmittelbar auf die "Welt" bezogen werden, sondern nur vermittelt über die immer schon vorliegende Weltdeutung der Umgangssprache als letzter Metasprache.

Der heute schon als traditionell zu bezeichnende Aufbau der meisten computerlinguistischen Systeme bildet in seiner Schichten- bzw. Modularchitektur die mit den Mitteln der Kalkülierung entwickelte Unterscheidung in die linguistischen Abstraktionsebenen *Lexikon/Morphologie*, *Syntax*, *Semantik* und *Pragmatik* nach. Die durch Chomsky aufgeworfene Frage, inwieweit diese Ebenen als in sich abgeschlossene autonome Module zu verstehen sind, spielt dabei eine nachgeordnete Rolle. Festzuhalten bleibt, dass die Konstruktion der linguistischen Theorie in der formalistischen Tradition einen inhärenten Ansatz zur Operationalisierung und damit zu ihrer prinzipiellen Implementierbarkeit in sich trägt, die eine maschinelle Verarbeitung in folgender Weise ermöglicht: Signale werden in Zeichenfolgen transformiert, die dann segmentiert und Hand in

Hand mit morphologischer Analyse durch Lexikoneinträge annotiert werden. Darauf setzt eine grammatische Analyse auf, die Strukturbäume erzeugt – in den meisten Fällen Konstituentenstrukturen und, weil für die folgende semantische Verarbeitung wichtig, auch Abhängigkeitsstrukturen. Sodann werden diese grammatischen Strukturen in eine logische Form transformiert, die, semantisch ausgewertet – u.a. zur Auflösung von Referenzen – dann als Darstellung der Bedeutung der anfänglichen sprachlichen Eingabe gilt. In Dialogsystemen wird auf der pragmatischen Ebene zumeist mit Sprechakten gearbeitet; diese Stufe ist aber oft nur aufgesetzt und spielt dann eine ergänzende Rolle. Auch wenn verschiedentlich Zweifel an der Adäquatheit einer streng sequentiellen Verarbeitungsordnung geäußert wurden, was sich dann in verschiedenen Formen der inkrementell verzahnten Verarbeitung und der Rückkopplung zwischen Komponenten niederschlug, ist doch festzuhalten, dass eine solchermaßen konstruierte Klasse von Verarbeitungssystemen als Realisierung einer bestimmten theoretischen Konzeption zu verstehen ist.

1.2 Eine pragmatische Wende für die Computerlinguistik?

Nach der durch Wittgenstein eingeleiteten pragmatischen Wende in der Linguistik stellt sich auch für die Computerlinguistik die Frage nach einer pragmatischen Fundierung. Gerade bei Dialogsystemen liegt es nahe, die Mensch-Computer-Interaktion aus der Perspektive der "Sprache als Handlung" zu verstehen, und dies um so mehr, weil ja sprachliche Handlungen zu Aktionen in den mit ihnen verbundenen technischen Anwendungssystemen führen sollen.

Nun ist unbestritten, dass Konzepte der linguistischen Pragmatik in die Computerlinguistik Eingang gefunden haben. An erster Stelle ist die im Anschluß an Austin von Searle entwickelte Theorie der Sprechakte zu nennen; eine Reihe computerlinguistischer Systeme verfügt über pragmatische Komponenten im Sinne des oben beschriebenen Architekturrahmens, die Sprechakte verarbeiten. Des Weiteren finden sich öfter Hinweise auf die Griceschen Kommunikationsprinzipien, auch wenn hier zumeist eine Lücke zwischen dem Anspruch und ihrer praktischen Umsetzung klafft. Festzuhalten bleibt aber, dass die Aufnahme pragmatischer Konzepte – wie etwa der Sprechakte – üblicherweise im Sinne einer Ergänzung erfolgte, sie aber nicht zu einer grundsätzlichen Neuorientierung im Sinne eines Fundierungsverhältnisses geführt hat. In welcher Weise wäre denn eine solche pragmatische Neuorientierung überhaupt möglich?

Was die pragmatische Fundierung im Kern ausmacht, hat H.J. Schneider¹ treffend formuliert: "Die Sprache ist nur dann angemessen beschrieben, wenn sie von vornherein unter dem Aspekt der Handlung gesehen wird, d.h., wenn die Kategorien und Klassifizierungen pragmatisch fundiert sind. Konkreter gesagt: Wenn sich über Bedeutungen nur in Kategorien des Gebrauchs sprechen läßt, und wenn man ferner zeigen kann, daß es nicht sinnvoll ist, syntaktische Kategorien unabhängig von semantischen zu bilden, dann ergibt sich ein Fundierungsverhältnis, dessen Basis die Pragmatik bildet, auf der zunächst die Semantik und schließlich die Syntax aufruht." Mit dem Primat der Pragmatik ernst zu machen heißt also, dass sprachliche Strukturen nicht als Objekte, sondern als Teil einer rationalen Praxis zu verstehen sind; die Rekonstruktion von Bedeutung beruht ausschließlich auf Unterscheidungen, die zu einer Theorie des Handelns und der Interaktion ohne jegliche mentalistische oder ontologische Annahmen gehören.

Nun geht es uns an dieser Stelle aber auf dem Hintergrund der eingangs gestellten Frage nach operationalisierbaren Ansatzpunkten nicht um die Ausführung eines sprachphilosophischen Begründungsprogramms. Vielmehr ist unser Interesse darauf gerichtet, die pragmatische Wende im Rahmen der formalen Mittel, die der Computerlinguistik zur Verfügung stehen², umzusetzen. Das heißt, dass wir im Rahmen des logoszentrierten Paradigmas notabene an der Prämisse der "Sprache" als abstraktes Konstrukt "hinter" dem Sprechen³ festhalten müssen, allerdings nicht im Sinne der idealsprachlichen Konzeption. Es verschiebt sich jedoch das Gewicht des formalsprachlich-logischen Ansatzes auf die pragmatische Ebene. Dabei ist darauf zu achten,

¹H.J. Schneider: Pragmatik als Basis von Semantik und Syntax. Frankfurt 1975

²Als formale Mittel setzen wir die konstruktive Logik, Elementarmathematik, prädikative Mengenlehre und Funktionentheorie voraus.

³s. S. Krämer, E. König (Hg.): Gibt es eine Sprache hinter dem Sprechen? Frankfurt, 2002

dass wir nicht den Fallstricken einer intentionalistischen Semantik erliegen, wie sie von Meggle u.a. ausgearbeitet wurde: In der Auszeichnung von Intentionen zu Objekten, wie sie auch von Searle in seinen jüngeren Arbeiten vorgenommen wird, sehen wir einen mentalistischen Rückfall hinter die pragmatische Wende; dies trifft in ähnlicher Weise auch auf den realistischen Repräsentationalismus der Situationstheorie nach Barwise und Perry zu.

Die Umsetzung der pragmatischen Wende in der Computerlinguistik bedeutet, den durch Chomsky angelegten Weg umzukehren: Methodisch steht die Pragmatik am Anfang. Bedeutung konstituiert sich im Sprachhandeln, in der Sprachverwendung. Es werden nicht nur die semantischen und syntaktischen Kategorien aus pragmatischen Unterscheidungen abgeleitet, sondern auch die Inferenzprozesse werden von der pragmatischen Ebene aus angestoßen und gesteuert. Dies steht in deutlichem Kontrast zur üblichen Sichtweise, in der Bedeutungen als abstrakte Objekte mit rein linguistischen Objekten in einer funktionalen Weise verbunden werden und in der die Pragmatik lediglich in einer Untersuchung der Abhängigkeiten zwischen Bedeutungsfunktionen und Funktionen der Verwendung linguistischer Ausdrücke besteht. Dass aus dem Strukturalismus heraus schon in den 1920er Jahren durch den Funktionalismus der Prager Schule eine Alternative bestand, ist in der Computerlinguistik weitgehend folgenlos geblieben: Die Analyse beginnt demnach bei der durch eine sprachliche Äußerung ausgedrückten Sprecherintention, d.h. bei der Funktion, um die Form zu beschreiben. Die gemäß der "funktionalen Satzperspektive" herausgearbeitete Thema-Rhema-Strukturierung von Satzfolgen als hauptsächliches Strukturierungsprinzip kann dabei eine ausgezeichnete Brückenfunktion zur pragmatischen Ebene übernehmen. Mit der Pragmatik im Mittelpunkt ergibt sich zugleich auch ein unmittelbarer Ansatz zur Performanzmodellierung, die für Dialogsysteme unerlässlich ist. Performanzaspekte auszuschließen ist nicht nur praktisch, sondern auch methodisch ein Holzweg.

2 Ein Ansatz für pragmatisch fundierte Dialogsysteme

Wie die Umsetzung des Primats der Pragmatik in einem Dialogsystem vorgenommen werden kann, sei nun am Beispiel unserer eigenen Arbeiten kurz umrissen. Wir gehen davon aus, dass im Unterschied zu freien Konversationen *zielorientierte, rationale* – d.h., gewisse Rationalitätskriterien einhaltende – Dialoge geführt werden sollen. Jeder Beitrag in einem solchen Dialog wird als differentielle kommunikative Handlung verstanden, die dazu beiträgt, schrittweise eines oder mehrere aus den Intentionen der Dialogpartner resultierende Ziele in Kooperation mit einem Anwendungssystem zu erreichen. Soll ein Dialogsystem die Rolle eines Partners spielen, ist es notwendig, dass es die Intentionen seiner Dialogpartner erkennen kann – eine genuin pragmatische Fragestellung. Um die daraus abgeleiteten Ziele erfüllen zu können bzw. ihre Erfüllung zu überprüfen, müssen Inferenzen über Sprechakte ausgeführt werden. Diese Interpretationsprozesse finden in dynamischen Kontexten statt, zu denen die Dialogsituation, die Anwendungs- oder Aufgabensituation sowie Benutzerpräferenzen und -attribute beitragen. Dies heißt, dass wir mit einer komplexen Planungsaufgabe konfrontiert sind, zu der sprachliche Handlungen, d.h. Dialogakte, und Aktionen in dem jeweiligen Anwendungsbereich beitragen. Im übrigen liefert die plan-basierte Herangehensweise erst die Voraussetzungen zur Führung von Dialogen mit gemischter Initiative und zum Einbezug weiterer Modalitäten wie z.B. Zeigehandlungen.

2.1 Entwurfskriterien für die Systemarchitektur

Für die Architektur *plan-basierter* Dialogsysteme ergibt sich damit die zentrale Stellung einer Komponente zur Dialogführung, dem sog. *Dialogmanager*, der in enger, aber arbeitsteiliger Koppelung mit einem Anwendungssystem kooperiert. Es ist wichtig, festzuhalten, dass sich eine Zielerreichung erst aus dieser Kooperation ergibt: Während der Dialogmanager für die Erfüllung der diskursiven Ziele zuständig ist, hat das Anwendungssystem die Erfüllung der Benutzerziele in der Anwendungsdomäne zu planen. Diese Funktionsteilung resultiert in kanonischer Weise aus den verschiedenen zu berücksichtigenden Wissensarten, dem linguistischen und dem

Anwendungswissen, und sie bietet darüber hinaus einen ausgezeichneten technischen Ansatzpunkt zur Rekonfigurierbarkeit für unterschiedliche Anwendungen. Aus architektonischer Sicht ist der skizzierte Aufbau – und dies wird sich auch bei der feineren Strukturierung der linguistischen Verarbeitungskomponenten zeigen – gegenüber traditionellen Ansätzen dennoch ziemlich konservativ. Wo liegt also die entscheidende Differenz?

Im Hinblick auf die statischen Wissensquellen bestehen signifikante Unterschiede nicht so sehr in ihrem Aufbau, sondern in den nunmehr von der Pragmatik her konstruierten Kategoriensystemen. Das zweite wesentliche Unterscheidungskriterium sind die semantisch-pragmatisch dominierten Verarbeitungsstrategien und Inferenzmethoden. Im Mittelpunkt der Verarbeitung stehen nun die Planung, der sich entwickelte Informationsstand des Systems – epistemisch gesprochen: das dynamische Situationswissen – und die Verarbeitung von Sprechakten mittels diskursiver Regeln und Rationalitätskriterien im Hinblick auf die zu erreichenden Ziele. Erst mit der plan-basierten Vorgehensweise kann systemseitig mit einem kooperativen Antwortverhalten und der Möglichkeit für Verhandlungsstrategien ernst gemacht werden.

Die arbeitsteilige Planung zwischen Dialogmanager und Anwendungssystem setzt voraus, dass beide “eine gemeinsame Sprache sprechen”, d.h., im Hinblick auf die von beiden gemeinsam genutzten Konzepte muss semantische Kompatibilität garantiert werden. Dies wird dadurch erreicht, dass beide auf dasselbe Anwendungsmodell in der Gestalt einer “formalen Ontologie” zurückgreifen. Zur Modellierung des Anwendungswissens – übrigens ebenso wie des größten Teils des linguistischen Wissens – haben wir uns für eine uniforme logische Repräsentation mit den Mitteln entscheidbarer Logiksprachen, den sog. Beschreibungslogiken, entschieden, wofür auch effiziente Inferenzmechanismen bereitstehen. Die Wissensbasis des Dialogsystems besteht aus einem Konzeptgraphen, der aus folgenden Teilgraphen besteht: Lexikalische Konzepte (“Begriffe”) – sie repräsentieren die Synonymmengen (Synsets) aus dem hierarchischen Lexikon EuroWordNet –, Konzepte der Diskurspragmatik, Konzepte zur Benutzermodellierung, und, je nach Anwendung austauschbar, die Konzepte der Anwendungsdomäne.

2.2 Linguistische Verarbeitung

Die pragmatisch gesteuerte kann an dieser Stelle nur kurz umrissen werden.⁴ Abgesehen von trivialen Fällen ist eine direkte Abbildung einer Äußerung des Benutzer in Aktionen eines Anwendungssystem nicht möglich. Im allgemeinen liegen komplexe Sprechakte vor, wobei die Interpretation des propositionalen Inhalts der Äußerung zunächst durch ihren (lokalen) linguistisch-pragmatischen Kontext bestimmt wird. Diese wiederum wird beeinflusst durch (globale) diskurspragmatische Merkmale, die Restriktionen aus der Dialoggeschichte und der faktischen Position der Äußerung im Dialog einbringen – beispielsweise, wenn es sich um die erwartete Antwort auf eine Frage handelt. Darüber hinaus sind durch die Anwendung weitere Restriktionen gegeben, die die Bedeutung von Wörtern und Konstituenten auf ihren speziellen Gebrauch innerhalb der Anwendungsdomäne einschränken. Daher müssen mehrere ineinander verwobene Ebenen bei der Verarbeitung von Benutzeräußerungen unterschieden werden:

- Linguistische Analyse auf der äußerungs-lokalen Ebene, die ihrerseits in weitere Ebenen der syntaktischen Strukturanalyse und der Konstruktion semantischer Repräsentationen unterteilt ist;
- Semantische Auswertung, d.h. Auswertung semantischer Operatoren, Referenzauflösung und weitere Transformationen der logischen Form, die sich aus speziellen Berechnungen ergeben;
- Anwendungsspezifische Spezialisierung der ausgewerteten semantischen Repräsentation, d.h. Abbildung der lexikalischen auf Domänenkonzepte;
- Diskurspragmatische Analyse, d.h. Einbettung in den Dialogkontext durch den Dialogmanager;

⁴Für technische Details sei auf weiterführende Publikationen verwiesen: [<http://www8.informatik.uni-erlangen.de/IMMD8/staff/Goerz/goerzpubl.html>]

- Sprachgenerierung nach Ausführung der Aktionen im Anwendungssystem und entsprechenden Rückmeldungen, die ihrerseits aus Inhaltsplanung und Verbalisierung besteht.

Die Analyseseite sei nun noch etwas genauer betrachtet, wobei für alle Ebenen gilt, daß die Verarbeitung weitgehend inkrementell erfolgt. D.h., daß Teilergebnisse einer Analysestufe bereits von der nächsten weiterverarbeitet werden können, während erstere noch andere Teile ihrer Eingabe verarbeitet.

Syntaktische Ebene. Sie führt eine Strukturanalyse auf einem vom Spracherkenner erzeugten Graphen aus zeitlich konkurrierenden Worthypothesen durch, wobei ein zweiphasiges Verfahren zur syntaktischen Ableitung benutzt wird. In der ersten Phase werden Kanten dieses Graphen in sog. Chunks zusammengefasst. Die zweite Phase kombiniert diese Fragmente zu größeren Einheiten unter Ausnutzung semantischer und pragmatischer Restriktionen, deren Anwendbarkeit durch die Valenz- und Tiefenkasus-Rahmen der lexikalischen Einheiten bestimmt wird.

Semantische Ebene. Die semantische Repräsentation erfolgt im Rahmen der von Kamp entwickelten Diskursrepräsentationstheorie (DRT), die von vorneherein satzübergreifend angelegt ist. Während die semantischen Repräsentationen der Wörter bereits durch den Lexikonzugriff eingesetzt werden, erfolgt ihre Komposition durch die Anwendung semantischer Operatoren, die mit den Regeln der Grammatik verbunden sind. Dieser Schritt ist inkrementell mit der syntaktischen Analyse verzahnt: Der Parser baut Diskursrepräsentationsstrukturen (DRSen) auf und nach jedem Kompositionsschritt wird deren Erfüllbarkeit bezüglich einer gegebenen Wissensbasis durch einen Konsistenztest überprüft.

Semantische Auswertung. Erfüllbare DRSen müssen ausgewertet werden im Hinblick auf die Auflösung von Referenzen, insbesondere Anaphern, sowie spezielle DRT-Operatoren. Die Anaphernresolution wird durch linguistische und pragmatische Heuristiken unterstützt. DRSen können logische Operatoren enthalten, z.B. Disjunktion, bedingte Ausdrücke und sog. Duplexbedingungen, die natürlichsprachliche Quantoren repräsentieren. Auswertung von DRSen bedeutet, bestimmte Transformationen auf sie anzuwenden. Disjunktion führt zu zwei alternativen DRSen, analog die Auflösung von Skopusambiguitäten bei Quantoren. Weiterhin werden einige generische Berechnungen ausgeführt, z.B. Kalenderrechnung zur genauen Zeit- und Datumsbestimmung bei Zeitangaben.

Anwendungsdomänen-spezifische Spezialisierung. Im nächsten Schritt wird die ausgewertete semantische Repräsentation in eine domänenspezifische logische Form umgesetzt, wobei die allgemeinen lexikalischen Konzepte mittels gegebener Abbildungsbeziehungen und Restriktionen durch Konzepte der Anwendungsdomäne aus der formalen Ontologie der der Anwendung ersetzt werden.

Damit ist eine logische Form erreicht, die auf ihre Erfüllbarkeit im Diskurskontext zu überprüfen ist und dann zur Weiterverarbeitung and das Anwendungssystem übergeben werden kann, z.B. als Anfrage. Die Einbettung in den Diskurskontext ist ein Inferenzprozess, der die Anwendung von Diskursoperationen umfaßt, welche die durch den jeweiligen Sprechakt bestimmten Performativa gegeben sind.

Wir sind überzeugt, dass der Vollzug der pragmatischen Wende in der Computerlinguistik auf dem aufgezeigten Weg zu flexiblen Systemarchitekturen führt, die gute Ansatzpunkte für eine Lösung vieler bisher nur teilweise oder gar nicht gelöster Fragestellungen bietet und damit der konventionellen Herangehensweise überlegen ist.