

Harald H. Zimmermann

Linguistische Informationswissenschaft

Im Rahmen des Studienfachs „Allgemeine Sprachwissenschaft“ wird seit kurzem an der Universität Regensburg neben der theoretisch-systematischen und der empirisch-pragmatischen Sprachwissenschaft ein Teilgebiet vertreten, das eine Verknüpfung von Sprachwissenschaft und Informationswissenschaft zum Ziel hat: die Linguistische Informationswissenschaft. Gegenstandsbereich, Ausbildungskonzept und Studienmöglichkeiten dieser Disziplin sollen im folgenden vorgestellt werden. Die Informationswissenschaft, wie sie hier verstanden wird, hat ihre Grundlage in der kybernetischen Informationstheorie. Im Mittelpunkt stehen dabei Probleme der Informationserzeugung, -übertragung und -verarbeitung. Das dem Menschen geläufigste System ist vor allem im Zusammenhang menschlicher Kommunikation seine eigene Sprache. Die theoretische Disziplin, die sich mit der Beschreibung der natürlichen Sprache befasst, ist die Linguistik. In erster Näherung kann man daher sagen, dass die Linguistische Informationswissenschaft die Theorie natürlichsprachiger Informationssysteme zum Gegenstand hat.

Mit der raschen und noch lange nicht abgeschlossenen Entwicklung der elektronischen Datenverarbeitung ist eine Möglichkeit gegeben, die mit bisher üblichen intellektuellen Methoden nicht mehr zu bewältigende Menge sprachlicher Informationen, die in allen Gesellschaftsbereichen (Wirtschaft, Recht, Politik, Kultur u. a.) produziert werden, zu verarbeiten im (positiven) Sinne einer sinnvollen Regulierung, Kanalisierung und Informationsreduktion. In der dadurch notwendig gewordenen Kommunikation Mensch-Maschine ist der Mensch bisher weitgehend gezwungen, sich solcher (abstrakter) formaler Sprachen zu bedienen, die der Computer 'verstehen' kann. Erst in jüngster Zeit und in kleinen Schritten hat eine Entwicklung in Richtung auf ein dem Menschen gewohntes Kommunikationsmittel, seine natürliche Sprache, begonnen. Erst dadurch werden viele - und nicht nur Spezialisten - den Dialog mit dem Computer aufnehmen können, wobei auch ein Teil der letztlich unbegründeten Reserviertheit gegenüber diesem technischen Instrument verlorengehen dürfte.

Mensch-Maschine-Systeme im Mittelpunkt

Mensch-Maschine-Systeme, denen Informationssysteme mit natürlichsprachigem Inhalt zugrunde liegen, die über natürliche Sprache abgefragt und verändert werden können, stehen aufgrund ihrer sozialen Relevanz im Mittelpunkt der Linguistischen Informationswissenschaft. Damit wird eine engere (auch stärker ingenieurwissenschaftlich geprägte) Definition des Gegenstandsbereichs ermöglicht: Die Linguistische Informationswissenschaft (im engeren Sinne) hat die Entwicklung und Problematisierung sprachlicher Mensch-Maschine-Informationssysteme zum Gegenstand. Eine wichtige Komponente der Linguistischen Informationswissenschaft hat sich aus der mehr oder minder wissenschaftlichen Bearbeitung natürlichsprachiger Daten ergeben, vor allem aus den früher meist unsystematischen Versuchen einer maschinellen Sprachübersetzung. Von verschiedenartigen ad-hoc-Ansätzen ausgehend hat sich eine Teildisziplin vor allem im Rahmen der Sprach- und Literaturwissenschaft herausgebildet, die als ein wesentlicher Bestandteil der Linguistischen Informationswissenschaft gelten kann. Dieser Bereich ist zumindest unter dem Begriff 'Linguistische Datenverarbeitung' (LDV, gelegentlich auch 'Computerlinguistik' nach dem

englischen Computational Linguistics) gefasst worden. Damit soll die wechselseitige Relation zwischen Linguistik und elektronischer Datenverarbeitung umschrieben werden, wobei die Anwendung linguistischer Methoden in der Datenverarbeitung bisher in der Praxis, quantitativ gesehen, zumeist im Hintergrund stand: im Vordergrund standen sprach- und literaturwissenschaftliche Anwendungen; etwa maschinelle Sprachanalyse und -übersetzung, Textindices und Textkordanzanzen. Eine Orientierung der Linguistischen Datenverarbeitung auf die Kommunikation Mensch-Computer wurde allerdings schon früher gefordert.

Linguistische Datenverarbeitung ist ein Teilgebiet

Aus dieser Sicht stellt die Linguistische Datenverarbeitung ein Teilgebiet der Linguistischen Informationswissenschaft dar; die allgemeine und einzelsprachliche Linguistik ist in diesem Sinne eine Hilfswissenschaft. Da die Linguistische Datenverarbeitung mehr und mehr auch für die theoretische Sprachwissenschaft an Bedeutung gewinnt - u. a. als mögliches Korrektiv linguistischer Modelle über ihre Anwendung in konkreten Systemen - ist eine wichtige Hilfsfunktion der Linguistischen Informationswissenschaft für die Linguistik selbst gegeben.

Aber auch aus anderer Richtung hat sich eine Entwicklung ergeben, die zu einer wesentlichen Komponente der Linguistischen Informationswissenschaft geworden ist. In der maschinellen Dokumentation wurden schon früh - bescheidene - Ansätze unternommen, (meist neben formalisierten Angaben) natürlichsprachige Daten: Volltexte, Abstracts u. a. zur Informationsspeicherung und -wiedergewinnung heranzuziehen. Diese Entwicklung verstärkt sich zusehends, der Einsatz linguistischer Methoden bei maschinellen Informationssystemen wird zu einer Notwendigkeit, will man die langwierige, fehleranfällige oder zumindest wenig systematisierbare formale Beschreibung von (Text-) Dokumenten und Fakten über intellektuelle, d. h. von Menschen ausgeführte Prozesse überwinden oder wenigstens teilweise maschinell unterstützen.

Die systematische Behandlung maschineller Informationssysteme ist Aufgabe eines Teilbereichs der Informatik. Die Informatik - selbst noch eine junge Wissenschaft - ist eine vorwiegend mathematisch-technische Disziplin, deren Gegenstandsbereich in Anlehnung an die Computer Science (USA) die Grundlagenforschung und allgemeine Anwendung der Datenverarbeitung darstellt. Doch nicht nur für den Bereich der Informationssysteme stellt die Informatik wichtige Grundlagen bereit. Die Theorie der formalen Sprachen und die Automatentheorie, wesentliche Schwerpunkte der „Kern“ Informatik, liefern wertvolle Hinweise für die Analyse natürlicher Sprachen. Intensive Kenntnisse über die Programmierung von Rechenanlagen sind daneben für die Linguistische Informationswissenschaft eine notwendige Voraussetzung. Wie die Linguistik stellt die Informatik damit eine wesentliche Hilfswissenschaft dieser Disziplin dar. Der Teilbereich 'Informationssysteme' ist in der Orientierung auf Mensch-Maschine-Informationssysteme wiederum Teil der Linguistischen Informationswissenschaft. Es ist hier nicht der Raum, auf weitere wichtige Grenzbereiche einzugehen, von denen für die Linguistische Informationswissenschaft Impulse ausgehen können: zur Informationspsychologie, Artificial Intelligence, Philosophie (speziell: formale Logik), Statistik, Kommunikationswissenschaft und allgemeine Informationstheorie bestehen z.B. besonders enge Beziehungen.

Zwei Hauptgebiete

Versucht man, die Teilbereiche der Linguistischen Informationswissenschaft zum Zwecke einer Übersicht zu differenzieren und in eine gewisse Ordnung zu bringen, so lassen sich aufgrund einer schwerpunktorientierten Aufgabenstellung zwei Hauptgebiete unterscheiden:

Die Analyse der verschiedenartigsten sprachlichen Äußerungen steht im Mittelpunkt der Textanalyse. Dazu rechnet die Beschreibung oder Manipulation (im positiven Sinn!) kleinerer sprachlicher Einheiten (Zeichen, Wörter, Sätze) ebenso wie die Deskription von Texten oder (Literatur-) Werken. Dieser Bereich ist noch am ehesten mit dem herkömmlichen Begriff 'Linguistische Datenverarbeitung' zu benennen. Einer Beschreibung auf vorwiegend statistischer Grundlage stehen dabei Analysen auf grammatischer Basis (Syntax, Semantik, jeweils mit lexikalischer und algorithmischer Komponente) gegenüber, erweitert um den Bereich der Maschinellen Sprachübersetzung und ergänzt um die sprachanalytischen Teilbereiche der Artificial Intelligence.

Das zweite Hauptgebiet umfasst den eigentlichen Dialogbereich Mensch-Computer, d.h. den Aufbau und die Strukturierung natürlichsprachiger Informationssysteme. Die Unterscheidung herkömmlicher Informationssysteme nach Dokumenten-Retrieval (das den Nachweis von Dokumenten - Literatur, Textsammlungen, Normen u. a. - anhand von Schlüsselwörtern/Klassifikationen aufgrund einer entsprechenden Anfrage zum Ziel hat) und Fakten-Retrieval (das sprachlich/logische Zusammenhänge von Daten - inhaltliche Fakten - aus Dokumenten erfragbar machen soll) bildet auch die Grundlage der Differenzierung entsprechender Teilbereiche der Linguistischen Informationswissenschaft: Maschinelles Indexing und Abstracting auf linguistischer Basis einerseits, Entwicklung von Frage-Antwort-Systemen über natürlichsprachige Informationen andererseits sind eine Voraussetzung für die heute noch mehr oder weniger utopisch anmutende Entwicklung von automatischen Problemlösungsverfahren, etwa in der Rechtsprechung oder in der Programmierung von Rechenanlagen über natürlichsprachige Kommunikation.

Weitergehende Differenzierungen sind möglich

Diese - eher vorläufige - Gliederung soll und darf nicht weitergehende oder andersartige Einteilungen ausschließen, sondern ist als eine für die Übersicht nützlich erscheinende Strukturierung zu verstehen. Die einzelnen Teilbereiche sollten nach Möglichkeit in ihrem Zusammenhang verstanden werden, eine weitergehende Differenzierung ergibt sich aber ggf. aus Anforderungen theoretischer und praktischer Art, wie sie sich bei Anwendungen stellen mögen, z. B. im Bereich des Computerunterstützten Unterrichts das Problem der Freiantwortanalyse (in der Kommunikation des Lernenden mit dem Computer) oder bei Einzelfragen, etwa im Zusammenhang mit Forschungsprojekten.

Der Themenbereich der Linguistischen Informationswissenschaft wurde bisher in den weitgehend etablierten Ausbildungsdisziplinen, wie etwa Linguistik oder Informatik, nur unzureichend abgehandelt; dies betrifft vor allem die Problematik der Schnittstellen zwischen Linguistik und Informatik bzw. Informationswissenschaft. Eine eigenständige fachgebietspezifische Entwicklung, wie sie sich auch anderweitig, etwa in der Etablierung von Wirtschafts- und Rechtsinformatik andeutet, ist auch für die Linguistische Informationswissenschaft durchaus denkbar; mit der engen Verflechtung zur Allgemeinen Sprachwissenschaft, wie sie das Regensburger Konzept vorsieht, erscheint in der Linguistischen Informationswissenschaft neben der fachspezifischen Qualifikation eine grundlegende Ausbildung in Linguistik gesichert; die notwendige Kombination mit ei-

nem weiteren Studienfach gibt zudem die Möglichkeit, hier eine anwendungspraktische Ergänzung zu wählen (z. B. Medizin, Recht, Betriebswirtschaft, Soziologie). Damit sind zusätzliche Qualifikationsmöglichkeiten geboten, die bei dem Bedarf an Fachkräften mit eingehenden (auch praktischen) Kenntnissen in DV-Anwendung - hier vor allem im Bereich der Nichtnumerischen Datenverarbeitung mit den Schwerpunkten Textverarbeitung und Dokumentation - auch gute Berufschancen eröffnen.

Das Regensburger Konzept sieht Hochschulabschlüsse (Magister und Promotion) im Fach Allgemeine Sprachwissenschaft (AS) und einem weiteren Hauptfach vor. Im Rahmen der AS sind von drei Teilgebieten (theoretische und systematische Sprachwissenschaft; empirische und pragmatische Sprachwissenschaft; linguistische Informationswissenschaft) zwei (nach der Zwischenprüfung) auszuwählen, wobei bei einem Abschluss mit AS als 1. Hauptfach die schriftliche Arbeit einem dieser Teilgebiete zu entnehmen ist.

Daneben wird versucht, Elemente aus den in Regensburg vorhandenen Schwerpunkten der angewandten Informatik (Wirtschaftsinformatik und Rechtsinformatik), aber auch die Erfahrungen der (weitgehend automatisierten) Bibliothek sowie des Rechenzentrums in das Konzept einzubeziehen. Das Studienkonzept für das Teilfach 'Linguistische Informationswissenschaft' ist dabei wie folgt gegliedert:

Grundstudium

(+) Einführung in die Linguistische Informationswissenschaft (2-std./W); (+) Grundlagen der Programmierung und EDV (2-std./W); (+) Programmiersprache(n) I und II (4-std./W); (+) Proseminare I und II „Textanalyse“ (4-std./W); (/) Statistik für Linguisten (2-std./W); (+) Praxis des Programmierens 1 und II (4-std./W); (/) Logik für Linguisten (2-std./W); (-) Einführung in das Bibliothekswesen (2-std./W); (-) Automatisierung des Bibliothekswesens (2-std./W).

Hauptstudium

(+) Hauptseminare I und II (4-std./W); (/) DV in der Schule (2-std./W); (-) Datenschutz (2-std./W); (/) Neuere Entwicklungen in der Angewandten Informatik (2std./W); (-) Verwaltungsdokumentation (2-std./W); (-) Wirtschaftlichkeitsanalyse (2-std./W); (/) Simulation (2std./W); (-) Betriebssysteme (2-std./W); (-) Rechtsdokumentation (2-std./W); (/) Datenbanksysteme (2-std./W).

Die Linguistische Informationswissenschaft wird organisatorisch und inhaltlich von der Abteilung Nichtnumerische Datenverarbeitung im Fachbereich Sprach- und Literaturwissenschaft der Universität Regensburg betreut. Die Abteilung ist dabei dem Institut für Allgemeine und Indogermanische Sprachwissenschaft zugeordnet.

Die vorhandene Mindestkapazität von 12 Semesterwochenstunden ist für die obligatorischen Veranstaltungen (+) ausreichend; fakultative Seminare können größtenteils von der Abteilung betreut werden (/), die übrigen werden durch andere Disziplinen (-) bei Bedarf angeboten.

Mit der Einrichtung des Teilfaches 'Linguistische Informationswissenschaft' im Rahmen der Allgemeinen Sprachwissenschaft sollen nicht nur Studierenden von Fächern im Fachbereich Sprach- und Literaturwissenschaften, sondern auch von anderen Fächern zusätzliche Qualifikationsmöglichkeiten angeboten werden. Eine Einbindung einzelner Bereiche in das Angebot lehramtsorientierter Fächer (Bereich 'Rechnergestützter Unterricht' u. a.) erscheint möglich. Der Bedarf an

Fachkräften - etwa in den Bereichen Dokumentation, Textverarbeitung, Datenbanken, Rechner-Kommunikation - mit einer Orientierung an natürlichsprachlichen Informationen ist in Wirtschaft und Verwaltung stark gestiegen. Dieser Bedarf hat bereits einige Hochschulen (etwa Hamburg) veranlasst, im Rahmen der Informatikausbildung ein Nebenfach „Linguistik“ zuzulassen. Die Universität Bonn bietet im Studium der Kommunikationswissenschaft seit kurzem ein Teilfach 'Linguistische Datenverarbeitung' an. Das Regensburger Modell stellt zugleich eine Möglichkeit dar, spezifische Komponenten der Informationswissenschaft im Rahmen vorliegender Ausbildungsstrukturen kontinuierlich zu entwickeln. Es hat sich immer wieder gezeigt, dass die Kommunikation zwischen Linguisten, Dokumentations- und Informationswissenschaftlern noch erschwert ist durch das mangelnde Verständnis der jeweiligen (theoretischen und praktischen) Probleme der (Nachbar-) Disziplinen. Die Linguistische Informationswissenschaft will daher keine Alternative zu den Konzeptionen einer „Informationswissenschaft“, etwa von Berlin (Wersig) oder Düsseldorf (Henrichs), sein, sondern zwischen Linguistik, Informatik und Informationswissenschaft vermitteln.

Harald Zimmermann

Wiss. Rat und Professor Abt. Nichtnumerische Datenverarbeitung
an der Universität Regensburg

Literatur (Auswahl)

- (1) Damerau, F. J.: Automatic Language Processing. In: Ann. Review of Information Science and Technology 11 (1976) (in Vorber.).
- (2) Dietrich, W.; Klein, W.: Computerlinguistik - Eine Einführung. Stuttgart 1974.
- (3) Kunz, W.; Rittel, H. W. J.: Die Informationswissenschaften. (München 1972).
- (4) Sparck Jones, K.; Kay, M.: Linguistics and Information Science. N. Y.; London 1973 (dt. Linguistik und Informationswissenschaft. München 1976).
- (5) Ungeheuer, G.: Linguistische Datenverarbeitung - Die Realität und eine Konzeption. In: IBM Nachrichten 21 (1971) S. 688-694.
- (6) Walker, D. E.: Automated Language Processing. In: Ann. Review of Information Science and Technology 8 (1973) S. 67-119.
- (7) Wersig, G.: Information - Kommunikation - Dokumentation. München 1971.