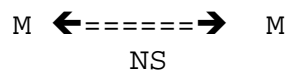


H. Zimmermann

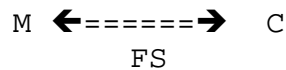
### Was leistet die Linguistik für die Informationswissenschaft und die Informatik?

Zu Beginn ihrer Einleitung zu 'Linguistics and Information Science' sprechen Sparck Jones/Kay<sup>1</sup> von Linguistik und Informationswissenschaft als "natürlichen Bettgenossen". Mir scheint, mit der Informatik teilt sich mindestens noch eine weitere Disziplin in diesen Raum. Es ist daher kein Eskapismus, wenn im folgenden die Bemerkungen zu möglichen Nutzenwendungen von Verfahren und Ergebnissen der (modernen) Sprachwissenschaft nicht streng nach Leistungen für "Informatik" und für "Informationswissenschaft" getrennt werden. Vielmehr ergibt sich dies, wenn man die drei Disziplinen unter dem meines Erachtens wesentlichen Aspekt ihrer Kommunikationsfunktion betrachtet. Die kommunikationsorientierte Betrachtungsweise erlaubt zugleich über eine - wenn auch vereinfachende - Differenzierung nach Kommunikationspartner und Kommunikationsmittel eine Beschreibung der zentralen Berührungspunkte dieser drei Disziplinen:

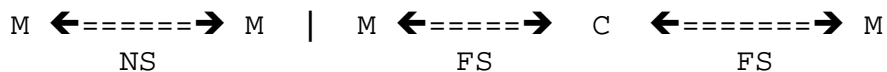
Die Linguistik beschäftigt sich mit Funktion und Strukturierung der natürlichen Sprache(n) in der unmittelbaren Kommunikation Mensch-Mensch.



Die Informatik befasst sich (u.a. - und dies ist hier von Interesse -) mit Funktion und Strukturierung der formalen (künstlichen) Sprache(n) in der Kommunikation Mensch-Computer.



Beide Disziplinen stellen damit wesentliche Grundlagen bereit für die Informationswissenschaft, die im Rahmen ihrer vielfältigen Probleme der Informationsübermittlung und -strukturierung dafür geeignete Kommunikationsmittel benötigt.



Die natürlichen Schnittstellen und Überschneidungsbereiche der drei Disziplinen haben zu wechselseitigen Beziehungen geführt. Dass die moderne theoretische Linguistik bereits wichtige Beiträge zur Theorie der formalen Sprachen geleistet hat, die wiederum in der theoretischen Informatik eine zentrale Stelle einnimmt, zeigen beispielhaft die grundlegenden Arbeiten Chomskys zu formalen Eigenschaften von Grammatiken.<sup>3</sup> Die ersten Überlegungen Chomskys sind bereits Mitte der fünfziger Jahre veröffentlicht, also etwa zeitgleich mit seinen ersten grundlegenden Publikationen zur Linguistik, so dass man daran zweifeln mag, ob sie ursprünglich "linguistisch" sind oder aber von der (damals noch nicht so benannten) Informatik her die Linguistik beeinflusst haben.

Wie dem auch sei, zwischen Linguistik und Informatik bestehen in diesem Bereich enge Wechselbeziehungen.

Die theoretische Linguistik vermochte und vermag sicherlich auch - so vermerken es etwa Kunz/Rittel<sup>4</sup> - die Informationswissenschaft zu beeinflussen, beispielsweise im Rahmen theoretischer Überlegungen zu Begriffssystemen, zur Sprachübersetzung oder zur allgemeinen Zeichentheorie.

Im folgenden soll das Thema jedoch stärker auf die praktischen Beziehungen dieser Disziplinen zueinander ausgerichtet werden. Damit soll einem spürbaren Mangel begegnet werden: bislang war vielfach m.E. der Informationsfluss zwischen Linguistik einerseits und Informatik/Informationswissenschaft andererseits zu sehr auf die theoretischen Aspekte beschränkt. Auch dies lässt sich weitgehend historisch erklären: für die Anwendung linguistischer Verfahren schien zunächst kein Raum vorhanden. In der Informatik bestand kein Bedarf an der Entwicklung einer der natürlichen Sprache angenäherten Programmier- oder Dialogsprache. Zu sehr war man mit der Bewältigung der ersten Schritte zur Kommunikation Mensch-Computer in Anspruch genommen. Andererseits haben die Misserfolge der maschinellen Übersetzung in den 60er Jahren und die unzureichenden Ergebnisse der automatischen Indexierung die Informationswissenschaftler, die zunächst sehr auf die angewandte Linguistik bauten, weitgehend verprellt oder verunsichert; viele große Versprechungen von Linguisten haben sich hier als Seifenblasen erwiesen.

Dass sich trotz mancher Rückschläge so etwas wie eine neue Disziplin in der angewandten Linguistik entwickeln konnte, bei der automatische Sprachverarbeitungsmethoden im Mittelpunkt stehen, ist wohl in erster Linie dem ständig zunehmenden Anwendungsspektrum der Datenverarbeitung zu verdanken. Immer mehr Nicht-Spezialisten werden mit dem Computer konfrontiert, somit wird das natürliche Kommunikationsmittel, die menschliche Sprache, als Verständigungsmittel mit dem Computer ständig attraktiver. Für die Informations- und Dokumentationswissenschaft stellt sich - ebenfalls durch das ständige Anwachsen der zu dokumentierenden Bereiche (etwa Recht, Presse, Wirtschaft) das Problem einer automatischen Bewältigung großer Mengen natürlichsprachigen Dokumentationsmaterials, da menschliche Kräfte zur Bearbeitung vielfach nicht mehr in ausreichendem Maß zur Verfügung stehen. Man kann also davon ausgehen, dass sowohl in der Informatik als auch in der Informations- und Dokumentationswissenschaft ein grundsätzlicher Bedarf an linguistischen Verfahren gegeben ist. Vor diesem Hintergrund konnte sich die maschinelle Sprachverarbeitung weitgehend kontinuierlich entwickeln. Sie beginnt sich dabei mittlerweile von den engen Bindungen zur Linguistik zu lösen und sich stärker auf die Anwendungsgebiete zu orientieren. Dies wird etwa schon an den Begriffen deutlich, unter denen diese Teildisziplin subsumiert wird: neben Termini wie Automatische Sprachverarbeitung (Automatic/Automated Language Processing), Linguistische Datenverarbeitung und Computerlinguistik (Computational Linguistics) treten die Anwendungsorientiertheit verdeutlichende Begriffe wie Algorithmische Linguistik, Linguistische Informationsverarbeitung und Linguistische Informationswissenschaft.

Im Vordergrund stehen nun nicht mehr allein Probleme der Anwendung der Datenverarbeitung als Hilfsmittel für die Linguistik (das klassische Beispiel für diese Betrachtung ist der Grammatik-Tester von J. Friedman<sup>5</sup>; aber auch die statistische Linguistik war zunächst vorwiegend auf die DV-Anwendung zur Erweiterung linguistischer Erkenntnisse ausgerichtet): immer häufiger wird versucht, die erarbeiteten - vorerst noch bescheidenen - Ergebnisse für die Anwendung nutzbar zu machen oder gar linguistische Verfahren und Systeme im Hinblick auf die DV-Anwendung zu entwickeln.

Dieser Prozess der Umsetzung linguistischer Methoden in die Anwendung bei Informatik und Informationswissenschaft muss behutsam und in kleinen Schritten vor sich gehen, will man nicht - wie das Beispiel der maschinellen Sprachübersetzung gezeigt hat - durch Unerfüllbarkeit allzu großer Hoffnungen die gerade angelegten Brücken und Wege wieder zerstören.

Bei der nun folgenden Skizze verschiedener Anwendungsebenen linguistischer Verfahren ist unterschieden zwischen zwei grundsätzlichen Anwendungsbereichen: dem Bereich der Allgemeinsprache wird dabei der Bereich der Fachsprache gegenübergestellt. Dies geschieht unter der Annahme, dass verschiedene linguistische Probleme "höherer" Ebenen (etwa im Bereich der Semantik oder Pragmatik) gegenwärtig nur im eingeschränkten Bereich einer Fachsprache (u.U. sehr stark eingegrenzt auf eine "Mikrowelt") bewältigt werden können, wenn linguistische Verfahren auch ökonomisch brauchbare Ergebnisse zu liefern imstande sein sollen.

Ein breiteres Anwendungsfeld (Allgemeinsprache) + beliebige Fachsprachen) haben linguistische Verfahren auf der (flexions-)morphologischen und syntaktischen Sprachebene. So scheint es heute kein Problem mehr zu sein, bei der maschinellen Sprachanalyse für eine flektierende Sprache wie das Deutsche oder das Russische linguistische Verfahren zur Reduktion von Wortformen auf zugehörige Grundformen einzubeziehen. Dies gilt etwa auch für die formale Kompositazerlegung oder für die Ermittlung mehrwortiger fester Sequenzen (nicht notwendig für eine semantisch orientierte Kompositaanalyse). Über allgemeinsprachliche Prozeduren lässt sich weitgehend auch der syntaktische Bereich einer Sprache bewältigen (etwa im Hinblick auf eine Vereindeutigung von Wortklassenmehrdeutigkeit). Beispiele für die Leistungsfähigkeit linguistischer Verfahren in diesem Bereich bringen die Systeme zur morphologischen und syntaktischen Analyse, die am Sonderforschungsbereich Elektronische Sprachforschung in Saarbrücken bzw. im Rahmen des Projekts Condor der Firma Siemens entwickelt werden.<sup>6</sup>

Der Einsatz derartiger Methoden in der automatischen Indexierung könnte bereits heute wesentliche Leistungsverbesserungen für die automatische Dokumentation erbringen (hier zeigen gerade die derzeit eingesetzten Systeme GOLEM/PASSAT und STAIRS (trotz FAIR) noch erhebliche Schwächen). Als weitere Einsatzmöglichkeit bietet sich die Registererstellung an: dabei können linguistisch orientierte Systeme helfend eingreifen, wie überhaupt das Verlagswesen (etwa im Rahmen der automatischen Fehlererkennung auf statistisch-linguistischer Basis und bei der maschinellen Lexikonauflistung) hier wesentliche Bausteine vorfinden kann. Die kommerzielle Textverarbeitung und Büroautomatisierung lässt für entsprechende Dokumentations- und Retrievalprozesse eine ähnliche Verwendungsweise erwarten.

Die maschinelle Übersetzung bleibt offensichtlich trotz aller Rückschläge ein wichtiges Desiderat auch schon für die nahe oder nähere Zukunft. Dies zeigen nicht zuletzt die jüngsten nationalen Bemühungen (Test des SYSTRAN-Systems<sup>7</sup> bei der GMD) und die Anstrengungen im Rahmen der Europäischen Gemeinschaft (LEIBNITZ-Gruppe). Von dem Vorhandensein eines guten vollautomatischen Übersetzungssystems größeren Ausmaßes kann derzeit jedoch keine Rede sein. Auch hier wird man umdenken müssen. Sicherlich gibt das SYSTRAN-System - wie gut auch der geplante Test ausfallen mag - bei weitem nicht den Stand des derzeit Machbaren und auch ökonomisch Vertretbaren wieder. Dennoch zeigt es mit der Methode der lexikalischen Einschränkung der Semantik auf Fachgebiete durch Fachgebietsmarkierungen einen Weg zu einer breiteren Anwendungsmöglichkeit mit dem Ziel einer brauchbaren (d.h. lesbaren und für den Fachmann verständlichen) maschinellen Übersetzung auf. Hier auf besserer syntaktischsemantischer Grundlage fortzuschreiten, wird ein lohnender Weg sein. Nicht zu vernachlässigen - wenn nicht

gar in den Vordergrund zu stellen - ist dabei die Entwicklung verbesserter Verfahren zur maschinenunterstützten Übersetzung. Hier kann die maschinelle Lexikographie bereits unter Verwendung einfacher linguistischer Algorithmen zur Flexionsanalyse (evtl. auch zur Derivation und Kompositaanalyse) beträchtliche Hilfestellungen anbieten.

Die Informatik ist (notgedrungen) einen beschwerlichen Weg gegangen. Einmal musste und konnte sie zunächst davon ausgehen, dass der Gesprächspartner des Computers ein DV-Spezialist war, der sich bemühen musste, in der internen oder maschinennahen Sprache des Rechners seine Prozedur zu 'programmieren', zum anderen hatte sie mit den sog. 'problemorientierten Sprachen' (etwa ALGOL für Mathematiker, COBOL für den Kommerz, FORTRAN für die (numerisch orientierte) Wissenschaft) für einige "Fachsprachen" so etwas wie eine Mittellösung anzubieten, um den DV-Laien bei der Programmierung von technischem Ballast zu befreien. Mittlerweile hat sich der Kreis der DV-Benutzer stark erweitert, das Ausbildungsspektrum der Kommunikationspartner des Computers hat sich verbreitert und die Anwendungsbereiche - besonders in der nicht-numerischen Datenverarbeitung - haben sich ausgedehnt. Dies gilt aber auch für die Funktion des Computers selbst. Er ist nicht mehr allein ein Instrument zur Ausführung eines vorprogrammierten Prozesses, sondern immer mehr ein Hilfsmittel während des Problemlösungsprozesses selbst. Schlagworte dafür sind Begriffe wie "Informationssystem", aber auch "Strukturiertes Programmieren".<sup>8</sup> Der dazu notwendige Dialog mit dem Computer soll ökonomisch sein (Kosten, Verweilzeit), aber dort, wo es sinnvoll erscheint, benutzernah und -freundlich (d.h. er soll in der natürlichen Sprache oder einer der natürlichen Sprache angenäherten Form ablaufen können). Ohne Zweifel hat hierbei die automatische Sprachverarbeitung mittelfristig eine wichtige Aufgabe zu bewältigen. In der Linguistik und der ihr benachbarten Artificial Intelligence (AI) ist dieser Weg methodisch durchaus vorbereitet; dies zeigt eine Vielzahl von AI-Projekten zur Beschreibung von Mikrowelten<sup>9</sup>. Diese Modelle alle als "exotisch" abzutun, wäre der Sache nicht dienlich: vielmehr müssen manche als Exemplum für die grundsätzliche Machbarkeit betrachtet werden. Da ökonomische Gesichtspunkte, z.B. Fragen der Portabilität sowohl der Programme auf andere Rechner als auch der Methoden auf andere "Welten" oder "Weltausschnitte", bei diesen Experimenten nicht im Vordergrund gestanden haben, wären bei der Umsetzung für eine praktische Verwendung noch Erfahrungen zu sammeln. Es sei noch einmal betont, daß nicht alle Anwendungsbereiche der Datenverarbeitung hierfür in Frage kommen werden und dass vor allem - wegen der zu erwartenden höheren Kosten - abgewogen werden muß, inwieweit dem Benutzer eine stärker formalisierte Sprache in der M-C-Kommunikation zugemutet werden kann.

Sicherlich wird - das kann man trotz der derzeitigen finanziellen Situation behaupten - der Ausbildungsbereich (Computerunterstützter Unterricht im weitesten Sinne) für den natürlich-sprachigen Dialog Mensch-Computer besonders in Frage kommen.

Es seien noch zwei Teilbereiche erwähnt, für die sich linguistische Methoden im DV-Einsatz anbieten. Doch muss man derzeit wohl davon ausgehen, dass sich hier in naher Zukunft praktikable Lösungen nur in Einzelfällen anbieten werden:

Ein wichtiges Gebiet der Informationswissenschaft ist die Inhaltskonzentration und -kondensation von Dokumenten (Abstracting) zur rascheren Information von Fachwissenschaftlern. Ohne Zweifel ist mit jeder Abstraktion eine gewisse Informationsreduktion verbunden. Bereits das humane Abstracting ist daher prinzipiell ein problematisches Vorgehen, das zumindest genaue Kenntnisse des Sachgebiets und der Zielgruppe voraussetzt. Bei dieser gleichsam natürlichen Schwäche jeglicher Abstracts ließe sich vielleicht auch ein automatisches Verfahren verant-

worten, bei dem linguistische und statistische Methoden zusammenwirken, das manchen zelebral erstellten Produkten nicht nachstehen muß. Es ist nur zu fragen - ganz zu schweigen von der derzeitigen Machbarkeit - ob solche Systeme noch benötigt werden, wenn geeignete automatische Dokumentations- und Informationssysteme - auch Fakten-Retrieval-Systeme - auf Volltextebene zur Verfügung stehen.

Wesentlich interessanter, weil es im Ergebnis letztlich eine große Veränderung der Umwelt nahezu eines jeden Menschen in der industrialisierten Welt mit sich bringt, ist die Bewältigung der automatischen Analyse der gesprochenen Sprache (Demgegenüber ist die automatische Synthese gesprochener Sprache ein leichter lösbares Problem, doch von der Anwendungsseite her weniger interessant). Wie die jüngsten Studien zu diesem Fragenkreis zeigen, wird die Analyse gesprochener Sprache nur über das Zusammenwirken aller sprachlichen Analyseebenen zu realisieren sein (also in der Linguistik von der Phonetik über die Morphologie und Syntax zur Semantik und Pragmatik), ganz abgesehen von einerseits mentalen Problemen, die im Zusammenwirken mit der kognitiven Psychologie bewältigt werden müssen und andererseits von den technischen Fragen einer geeigneten Auswahl und Umsetzung von Prozessdaten in digitale Informationen. Bei geeigneter Restriktion (z.B. auf Einzelworte) sind vielleicht im nächsten Jahrzehnt erste Ergebnisse für den praktischen Gebrauch (telefonisches Auskunftssystem?) zu erwarten; vorerst ist dieser Bereich der Grundlagenforschung zuzurechnen.

Es bleibt zu hoffen, dass der Kommunikationsprozess, der mittlerweile verstärkt zwischen Informatikern, Informationswissenschaftlern und Linguisten eingesetzt hat, hilft, das Problem einer benutzer-, d.h. menschenfreundlichen Kommunikation Mensch-Computer zu bewältigen und damit eine wichtige, wenn auch eher technische Voraussetzung zu schaffen, die Kommunikation zwischen den Menschen zu verbessern.

### Anmerkungen

1 Sparck-Jones/Kay /1/ S.1: "natural bedfellows"

2 So wird von Wersig /2/ die Informations- und Dokumentationswissenschaft im Rahmen ihrer strukturellen Beschreibung der Kommunikationswissenschaft untergeordnet (vgl. S 21). Bauer - Goos /3/ leiten ihre Einführung in die Informatik ein über die Kommunikationsfunktion der natürlichen Sprache. Übermittlung (konkreter) Nachrichten bzw. (abstrakter) Information.

3 Chomsky /4/, /5/, /6/ gegenüber /7/

4 Kunz/Rittel /8/ S. 29 f.

5 Friedman /9/

6 Vgl. /10/ und /11/

7 Informationen zum Test sind bei der GMD, Dr. R. Müller, erhältlich.

8 Vgl. etwa den bei IBM Deutschland entwickelten Pseudo-Code in natürlicher Sprache zur Vorbereitung eines Programms in strukturierter Form in /12/ S. 51.

9 Kritische Literaturübersichten u.a. in /13/ und /14/

10 Vgl. v.a. den Ergebnisbericht einer Studiengruppe zu diesem Problemkreis /15/ und die Projektberichte zu /16/.

### Literatur

/1/ Sparck-Jones, K.; Kay, M.: Linguistics and Information Science. New York/London 1973.

/2/ Wersig, G.: Information - Kommunikation - Dokumentation. Pullach 1974.

- /3/ Bauer, F.L.; Goos, G.: Informatik. Eine einführende Übersicht. Berlin/Heidelberg/New York 1973.
- /4/ Chomsky, N.: Three models für the description of language. I.R.E. Trans. Inf. Theory 2 (1956) 113-124.
- /5/ Ders.: On certain formal properties of grammars. Inf. And Control 2 (1959) 137-167.
- /6/ Ders.: Formal properties of grammars. In: Handbook of mathematical psychology (ed: Luce, D.; Bush, E.; Galanter, E. Bd. 2 New York 1963.
- /7/ Ders.: Syntactic Structures. Den Hag 1957
- /8/ Kunz, W.; Rittel, H.: Die Informationswissenschaften. Heidelberg/Berkeley 1969.
- /10/ Arbeitsberichte des Sonderforschungsbereichs 'Elektronische Sprachforschung' Gruppe A (Automatische Lemmatisierung / maschinelle Übersetzung) Saarbrücken 1972 ff. (Vervielf.)
- /11/Arbeitsbericht des Projekts CONDOR, Siemens, München 1975 (Vervielf.)
- /12/ Melekian, N.: Neue Methoden und Techniken der Programmierung (7): Methodische Programmentwicklung (1). In: IBM-Nachrichten 26 (1976) 48-55.
- /13/ Walker, D.E.: Automated Language Processing. In: Anm. Rev. of Inf. Science and Technology 8 (1973)
- /14/ Damerau, F.J.: Automatic Language Processing. In: Anm. Rev. of Inf. Science and Technology (11) 1976 (in Verb.)
- /15/ Newall, A. et al.: Speech understanding systems. Amsterdam / London 1973
- /16/ Walker, D.E. et al.: Speech understanding research. Stanford Research Institute, Menlo Park, CA., Juni 1975