99. Computerlinguistik

1. Linguistik und Datenverarbeitung
2. Maschinelle Sprachanalyse
3. Bearbeitung großer Datenmengen
4. Mensch – Maschine – Kommunikation
5. Bibliographie (in Auswahl)

1. Linguistik und Datenverarbeitung


Versucht man, das Feld über das Verhältnis von Linguistik und Computerwissenschaft (Informatik) zu erfassen, so ergibt sich folgendes Bild. Insbesondere in den 50er und frühen 60er Jahren haben sich Linguistik und Informatik teilweise gemeinsam entwickelt. Die Entwicklung von z. T.
sehr komplexen formalen Sprachen zur Erfassung von Problemen aus den verschiedenen Wissenschaften (Programmiersprachen) und von Algorithmen zu ihrer Analyse und Synthese sowie zur Übersetzung zwischen formalen Sprachen ging Hand in Hand mit der Entfaltung der algebraischen Linguistik, als deren Abkömmlinge die verbreitetsten Grammatikmodelle der neueren Sprachwissenschaft zu gelten haben (Floyd 1964).


Eine gut lesbare Übersicht zur Computerlinguistik ist Dietrich/Klein 1974, einen Literaturbericht gibt Straßer 1977. Der Stand wichtiger Aktivitäten in der Bundesrepublik und der Schweiz ist in „Kolloquium zur Lage […]“ doku-

2. Maschinelle Sprachanalyse

Unter maschineller Sprachanalyse wird meist der an der Wissenschaftssystematik der Linguistik orientierte Zweig der CL verstanden („Linguistik mit dem Computer“). Die maschinelle Sprachanalyse bemüht sich also darum, die Ergebnisse linguistischer Arbeit „auf den Computer zu bringen“. Deshalb ist es auch nicht möglich, eine allgemeine Zielsetzung für die maschinelle Sprachanalyse anzugeben.

Im phonetisch-phonologischen Bereich sind vor allem Arbeiten zu nennen, die sich mit der Überführung von akustisch gegebenen Sprachsignalen in phonetische, phonologische (oder auch graphematische) Repräsentationen beschäftigen, die also ein kontinuierliches Signal in eine Folge von diskreten Signalen umsetzen (automatische Spracherkennung, IKP 1971, Speech Communication) und damit die Voraussetzung für die eigentliche Sprachverarbeitung auf Digitalrechnern schaffen. Obwohl das Ziel der automatischen Spracherkennung eine Analyse auf der Ebene der Lautsegmente ist, ist sie praktisch nicht ohne Information von höheren Beschreibungsebenen (lexikalisch-syntaktisch und semantisch) durchzuführen.


Im Bereich der linguistischen Theoriebildung selbst spielt der automatische Grammatiktester zwar eine wichtige, letztlich aber auf die eines technischen Hilfsmittels beschränkte Rolle. Dagegen hat z. B. die Tatsache, daß man zur automatischen syntaktischen Analyse neben der Grammatik selbst immer auch ein Verfahren für die Handhabung der Grammatik (Algorithmus) zur Verfügung haben muß, insofern Folgen für die Theorie gehabt, als dadurch die Reflexion über das Verhältnis des statischen und des dynamischen Teils eines linguistischen Verarbeitungsmechanismus mit ihren Folgen für das Verhältnis von Kompetenz und Performanz in Gang gekommen ist (vgl. z. B. Winograd 1972, Bäori 1976, 1978).


3. Bearbeitung großer Datenmengen

3.1. Dokumentationssysteme

elaborierten formalen Sprache formuliert (Indexierungssprache) (3) der Zugriff zur gespeicherten Information (Retrieval) kann über natürliche Sprache erfolgen.

3.1.1. Indexierung


Eine vergleichende Bewertung konkurrierender Analyseverfahren ist bisher nicht gelungen (Sparck Jones/Kay 1976), insbesondere ist nicht gezeigt worden, daß linguistisch fundierte Systeme besser arbeiten als statistisch orientierte (speziell zur Syntax z. B. Salton/Lesk 1968).

3.1.2. Retrieval


Ein weites Feld für die CI eröffnet sich, wenn die Interaktion zwischen Benutzer und System in natürlicher Sprache ablaufen soll. Ob ein Datensatz naturliche Sprache möglich und wie restringiert diese Sprache sein soll, richtet sich wesentlich nach dem Benutzerkreis (vgl. dazu die in Kühlen 1978 zusammengestellte Literatur über künftige Entwicklungen im Dokumentationswesen). Entscheidend ist weiter das Verhältnis von Benutzersprache und Indexierungssprache. Die Benutzersprache muß mindestens so reich sein, daß Information unter allen den Gesichtspunkten erfragt werden kann, die die Struktur der IS bestimmen. Entsprechend hat sich die Analyse der Benutzersprache an dem zu orientieren, was in der IS ausgedrückt werden kann.

Dokumentationssysteme, die einen Datenzugriff in natürlicher Sprache erlauben, sind den Frage-Antwort-Systemen zuzurechnen (Simmons 1970). Der wesentliche Unterschied zu den in der künstlichen Intelligenz entwickelten Frage-Antwort-Systemen (vgl. 4) besteht darin, daß bei den ersteren die Effektivität leitender Gesichtspunkte ist, während letztere Sprachverstehen simulieren wollen. Erstere suchen nach Dokumenten, die einen bestimmten Inhalt haben (Referenz-Retrieval), letztere direkt nach Inhalten (Fakt-Retrieval).
Dennoch wird eine zukünftige Synthese beider Typen von Dialogsystemen nicht ausgeschlossen (Smith 1976, Kühlen 1978).

3.2. Wörterbucharbeiten

Maschinell gespeicherte Wörterbücher haben gegenüber konventionellen Wörterbüchern (vgl. Art. 94) den Vorzug, daß sie schnell und systematisch gelesen, bearbeitet und ausgewertet werden können. Sie werden in der Regel für bestimmte, relativ wohldefinierte Zwecke erstellt und erfas-


4. Mensch – Maschine – Kommunikation


5. Bibliographie (in Auswahl)

S. Bátóri, On the Procedural Motivations of Natural Lan-
Anwendungsberichte


Peter Eisenberg, Hannover