

GLDV-Jahrestagung

ANALYSE UND SYNTHESE GESPROCHENER SPRACHE Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung

in Zusammenarbeit mit der Fachgruppe
Sprachanalyse der Gesellschaft für Informatik

Die Jahrestagung 1987 der Gesellschaft für Linguistische Datenverarbeitung unter der Thema *Analyse und Synthese gesprochener Sprache* fand vom 4. bis 6. März im Vortragssaal des Landesmuseums Bonn statt. Die Tagung war von Gerd Willée und Burkhard Schaefer (Organisationskomitee) sowie COLING-erprobten und bewährten Helfern vom IKP, Bonn vorbereitet worden. Das Programmkomitee bildeten H.G. Tillmann (Vorsitz), Winfried Lenders, Bernd S. Müller und Klaus G. Schweisthal. Der Tagungsband ist als Band 9 der Reihe "Linguistische Datenverarbeitung" bei Olms erschienen (Tillmann, H.G., Willée, G. (Hrsg.): *Analyse und Synthese gesprochener Sprache*, Hildesheim: Olms, 1987) und war bereits zu Tagungsbeginn den Teilnehmern verfügbar.

Folgende Beiträge wurden vorgestellt:

- B. Kaspar, Lochschmidt, H. Wolf: Spracherkennung und -synthese in der Telekommunikation.
W. Menzel, I. Starke: Wortformenanalyse und Schrifttext-gesteuerte Sprachsynthese.
S. Bauer et al.: Graphem-Phonem-Umsetzung: Lexikon versus Regelkatalog.
K. Huber: Sprachsynthese ab Text.
J.P. van Hemert, U. Adriaens-Porzig, L.M.H. Adriaens: Speech Synthesis in the SPICOS-Project
R. Sommer: Segmentierung von Sprachmaterial: Plädoyer für Koartikulation berücksichtigende Segmente.
G.F. Meier: Semantische Probleme beim Sprachverstehen.
E. Nöth, A. Batliner, R. Lang, W. Oppenrieder: Automatische Grundfrequenzanalysen und Satzmodusdifferenzierung.
R. Bannert, K.-J. Engelbert, B. Keck: Die Intonation der Spracherkennung. Das Black-Board-Model des Ikaros-Projekts.
D. Mergel, H. Ney, A. Noll, A. Paeseler: Wechselwirkungen zwischen akustischer Erkennung und linguistischen Einschränkungen im SPICOS-Projekt.
G. Niedermair: Linguistische Analyse im System SPICOS.
U. Ehrlich, H. Niemann: Semantische Verarbeitung.
A. Boner: Vier Jahre computerunterstützte Sprachübersetzung – Erfahrungsbericht.
U. Jekosch: Umsetzung von Phonemwörtern in ihre orthographische Form für unbeschränkten deutschen Wortschatz.
G. Thurman: Der Einsatz semantischer Verfahren in sprachverstehenden Systemen.
H. Gessler, K. Kotten, K.G. Schweisthal: Velare a-Realisationen im Bairischen.
I. Hertlich: Interindividuelle Variabilität in Langzeitstrukturen des Sprachsignals.
G. Dreckschmidt: The Linguistic Component in the Speech Understanding System SPICOS.
B. Möbius: Untersuchungen zu mikroprosodischen Grundfrequenzvariationen im Deutschen.
T. Goldbeck, W.F. Sendtmeier: Phonetische Manifestation der Intonation bei gleichzeitiger Realisierung unterschiedlicher linguistischer Funktionen.
E. Matschall, D. Schmidbauer: Berücksichtigung von akustisch-phonetischen Ambihuitäten bei einer bottom-up Worthythesen-Generierung.
A. Batliner: Der Einsatz der Diskriminanzanalyse zur Prädiktion des Modus.
K. Kohler: Funktionen von F_{01} -Gipfeln im Deutschen.
C. Reuter: Mikroprosodie und Synthese von Intonation.
R. Gartenberg: Zur Stilisierung von Intonationskonturen.
K. Fellbaum, L. Fliegner, M. Rusch: Erzeugung einer einfachen Wort- und Satzprosodie durch Modifikation der Sprachgrundfrequenz und der Lautdauer.
U. Heid, B. Weck, D. Roesner: Generierung deutscher Sätze aus semantischen Repräsentationen im Sinne von E. Steiner.
- Eine Panel-Diskussion "Automatische Segmentierung fließender Rede" wurde von H.G. Tillmann moderiert.

BEMERKUNGEN ZU DEN VORTRÄGEN DER JAHRESTAGUNG 1987 DER GLDV

Vom 4. bis 6. März fand in Bonn die Jahrestagung der Gesellschaft für linguistische Datenverarbeitung (GLDV) statt.

Die Gesellschaft hat für diese Tagung die Analyse und Synthese lautlicher Sprachereignisse gewählt, eine Thematik, die selbst innerhalb der LDV stiefmütterlich behandelt wird, ja lange Zeit nur im Rahmen der Akustik oder Nachrichtentechnik zuhause war. Verstehen von gesprochener Sprache ist nach wie vor ein noch wenig erforschter Bereich der Kommunikation, so daß von einer Simulation natürlicher Prozesse nicht gesprochen werden kann. Bei künstlicher Analyse durch technische Anlagen spielen letztlich auch die physiologischen und psychologischen Faktoren der menschlichen Perzeption und Lautvorstellung kaum eine Rolle, da nur die physikalischen Parameter der gesprochenen Sprache, d.h. der Vielzahl der Varianten individueller Sprechakte, verarbeitet werden müssen. Die Auffindung der für die Umkodierung in Grapheme nötigen Invarianten (bei einzelnen fälschlich als Phoneme) setzt die Segmentierung voraus, die der jeweiligen Orthographie entspricht. Das Gehirn löst dieses Problem durch zahlreiche Rückkopplungs- und Assoziationsprozesse, wie durch propriozeptive Referenzen, Summationen und Inhibitionen, so daß sprachinterne Artikulationsgewohnheiten zur selektiven, d.h. relevanten Hörgewohnheit führen, die ihrerseits zur graphematischen Zuordnung über trial-error-Prozesse die Basis ist. Da eine (Mutter-)Sprache größtenteils im Satzverband mit bedeutungstragenden Einheiten (Formativen) erlernt wird, ist der Verstehensprozeß weitgehend mit semantischen Rückkopplungen verknüpft, die ihrerseits durch situative und Wissensinformationen verstärkt, korrigiert oder präzisiert werden. Auf der Tagung wurde von mehreren Forschungsteams eine Lösungsmöglichkeit vorgelegt, die Silbe, Wort oder "Semantik" mehr oder minder einbeziehen will. Allerdings bauen die meisten Versuche auf DUDEN-Silben auf bzw. legen Silbengrenzen ad hoc fest. Silbentheoretische Erörterungen wurden – sicher auch aus Zeitgründen – nicht diskutiert, obwohl gerade die Silbengrenze eine umstrittene Frage ist. Die Silbentrennung – dies zeigen Ausführungen der SPICOS-Bearbeiter – wird von der Koartikulation überlagert, die sehr unterschiedliche Zeitstrecken bestreicht.

So wichtig und interessant die in den 60er Jahren durchgeführten Diphonuntersuchungen sind, so bleiben sie doch isolierte Kunstprodukte, da die Koartikulation oft silbengrenzenüberschreitend wirksam wird, was in einzelnen Sprachen Vokalassimilationen, Brechungen, Palatalisierungen, Umlautungen (innerhalb des Invarianzenbereiches) erzeugt. Der Zugriff zu größeren Einheiten, mindestens Disyllabeme oder -exeme bzw. Wörter in allen morphologischen Vari-

anten, paßt sich mehr an den natürlichen Verdichtungsprozeß im Gehirn an und erlaubt größere Toleranzen. Dies zeichnet sich bei mehreren Vorträgen zur Analyse und zur Synthese deutlich ab. Man könnte besonders für Dialogsysteme sogar – dies kommt bei der Forschung der Bundespost zum Ausdruck – von der Tatsache Gebrauch machen, daß sprachliche Äußerungen weitgehend auf feste Syntagmen und Redestereotypen aufgebaut sind. Erfolgreicher als die Analyseforschung ist zweifelsohne die (auf der Tagung repräsentierte) Syntheseforschung, der auch mehr Beiträge gewidmet sind. Es scheint, daß hier auch Koartikulation und damit verknüpft der Zeitfaktor noch ein gutes Stück Arbeit verlangt, daß aber "verständliche Sprache" selbst ohne Prosodik erzeugt werden kann, da als Endabnehmer ja der Mensch mit seiner relativ großen Verstehenstoleranz steht.

Über die Transkriptionssysteme wurde auf der Tagung wenig gesagt, zumindest scheint im wesentlichen etwa die Genauigkeitsstufe von API befolgt zu werden, die jedoch im Koartikulationsbereich (Pranasalisierung, Praaspiration, Diphonübergänge usw.) selbst bei europäischen Standardsprachen nicht hinreicht. Nicht ganz klar ist – so besonders bei den Züricher Arbeiten – die Rolle bzw. Notwendigkeit morphologischer Umkodierungen, wie z.B. die Deflektion bei der Lemmatisierung. Morphologische Analyse und Synthese bieten eine Menge von Fehlerquellen, sind teilweise recht aufwendig und sind bei der massenhaften Zunahme an Internationalismen und Fremdwörtern (mit Originalaus-sprache und Originalschreibung) regelgestört. Es empfiehlt sich auch in der Synthese mit Vollwortformen aus dem Speicher zu arbeiten. Einige Arbeiten haben sich mit der Satzmelodik, d.h. mit Grundfrequenzanalysen und entsprechenden Synthesen befaßt, wobei richtig die Folge der Silbenkernfrequenzen berechnet wird. Sowohl die mikroprosodischen als auch konturenorientierten Berichte zeigen gute Fortschritte gegenüber früheren Ergebnissen. Als Problem erscheint hier die Auswirkung unterschiedlicher Sprechgeschwindigkeiten auf die Intonation, wie dies im SPICOS-Projekt auch für die Findung von Silbenkernen der Fall ist. Recht interessant ist in diesem Zusammenhang die Untersuchung Kohlers, die wichtige Hinweise auf die Wechselbeziehung von Betonungsveränderungen und Verschiebungen des F_0 -Gipfels gibt. Einige Beiträge betonten die Notwendigkeit, daß auch für die akustische Analyse höhere Ebenen herangezogen werden müssen. Explizit wurde in 6 Vorträgen auf die Einbeziehung der Semantik (in unterschiedlicher Definition) bei Analyse und Synthese teils theoretisch, teils auch mit konkreten Angaben verwiesen. Teils sind damit semantische Minima-Merkmale für syntaktische Operationen gemeint, teils wird auch an Lexikonsemantik gedacht. Es ist aber keinem Vortragenden gelungen (möglicherweise durch zeitliche

Beschränkung bedingt) zu zeigen, wie in der Analyse mittels semantischer Entscheidungsoperationen Eindeutigkeit hergestellt wird.

Für die Satzsyntax ist natürlich eine inhaltliche (semantische) Ausgangsbasis erforderlich, wofür einige Vorschläge gemacht wurden, wenngleich nur an Modellen. Ob und wie weit das STEINER'sche Verfahren sich hierfür eignet, läßt sich erst sagen, wenn die entsprechenden Veröffentlichungen vorliegen. Unsere eigenen prädikatenlogischen Untersuchungen haben im Laufe von 20 Jahren mehrfacher Korrekturen diese Verfahrensweise als einzig mögliche bestätigt, wir bezweifeln aber, daß die Case-Grammar-Methode und die STEINER'schen Verbklassen dafür ausreichen, zumindest, wenn EUROTRA-D multilingual komplexe Informationen verarbeiten will. Es ist auf der Tagung mehrmals die Frage nach dem Zweck und der Anwendungsmöglichkeit der Forschungen gestellt worden. Unabhängig davon, wie man zu dem reklameträchtigen Terminus "Künstliche Intelligenz" stehen mag, ist sicher, daß Sachverhalte und Denkpro-

zesse in Computern nur verarbeitet werden können, wenn das semantische Problem ernsthaft, d.h. nicht nur in theoretischen Modellen und an einzelsprachliche Syntax gebunden, für Analyse, d.h. Inhaltsverstehen (Faktenrecherchen) und für die Synthese (Dialog-Antworten) in Angriff genommen wird.

Es ist der Verdienst der Veranstalter und Gastgeber, daß diese Problemschau für alle Teilnehmer sehr instruktiv und sicher auch zu Verbesserungen anregend war. Selbst für sprachhistorische Betrachtungen sind einige Ergebnisse aufschlußreich, wie die Münchner dialektologische Untersuchung zum bayrischen *a* gezeigt hat. Die Zukunft der LDV hängt ohne Zweifel davon ab, ob die sechste Computergeneration das hält, was die fünfte nur versprochen hat – und das hängt in erster Linie von dem linguistischen Beitrag ab, denn ohne semantische Kleinarbeit lassen sich keine Sachverhalts-Informationen verarbeiten.

Prof. Dr. Georg F. Meier, Olching

Gesellschaft für
Linguistische Datenverarbeitung e. V.
z. Hd. Frau Prof. Dr. B. Endres-Niggemeyer
FH Hannover
Fachbereich BID
Hanomagstraße 8
3000 Hannover 91

Absender

Name

Vorname

Straße

() Ort

Telef.



Aufnahmeantrag für persönliche Mitgliedschaft

Ich beantrage, mich in die Gesellschaft für Linguistische Datenverarbeitung e. V. (GLDV) als Mitglied aufzunehmen

Name Vorname

Akadem. Grad oder Berufsbezeichnung beschäftigt bei Student ja/nein

Interessengebiete

Arbeitsschwerpunkte

Datum Unterschrift