

**Wilhelm Weisweber:**

**Termersetzung als Basis für eine einheitliche Architektur in der maschinellen Sprachübersetzung.** Das experimentelle MÜ-System des Berliner Projekts der EUROTRA-D-Begleitforschung (KITFAST). Sprache und Information, Bd. 28, Niemeyer, Tübingen 1994.

Mit einigem zeitlichen Abstand zum Ende des EUROTRA-Projekts und seiner "Begleitforschungsprojekte" in Bonn, Bielefeld, Stuttgart und Berlin erscheinen nunmehr Monographien, die Überblicke über die Systeme bzw. über Einzelthemen oder zusätzliche Versuche bieten. So wurde in einem der letzten LDV-Foren eine Arbeit über die Anwendung der EUROTRA-Philosophie (und -Software) auf die exotische Sprache Persisch rezensiert; heute liegt eine Studie von Wilhelm Weisweber zum KIT-FAST-Projekt an der TU Berlin vor, die im wesentlichen den Aspekt des Termersetzungsverfahrens behandelt. Dieses Verfahren sollte im experimentellen MÜ-System der Berliner Gruppe den meisten der Analyse- und Syntheseschritte zugrundegelegt werden. Obwohl der Rezensent für die formale Seite des Architekturkonzepts nur teilweise kompetent ist, kann gesagt werden, daß die im Buch geschilderte einheitliche Architektur lediglich zum Teil im System ausgeführt worden ist; aus diesem Grund hat das innovative Kernstück der Arbeit, das Kap. 6.3. (Rolle der Termersetzungssysteme in der MÜ) auch nur theoretischen Charakter.

Doch beginnen wir von vorne und systematisch: Die Kapitel 1 (Einführung) und 2 (Maschinelle Sprachübersetzung) sind eine kondensierte Einführung in das Berliner Projekt und seine Überlegungen zur MÜ-Architektur, über die sich der interessierte Leser hier rasch und vollständig informieren

kann; das Kap. 3 (Produktionssysteme) behandelt einige alternative Ansätze, deren vergleichende Wertung jedoch äußerst knapp ausfällt. Jedenfalls ist nicht direkt ersichtlich, inwieweit die dann in den Kapiteln 4-6 sehr ausführlich dargestellten Termersetzungssysteme die Schwierigkeiten der anderen Systeme überwinden können. Insbesondere für das Problem der Abhängigkeiten unter den TE-Regeln werden nur theoretische Lösungen angeboten. Warum muß z.B. dem Knuth-Bendix-Algorithmus ein ganzes Teilkapitel gewidmet werden, wenn er in der MÜ gerade nicht verwendet werden kann?

Überhaupt verliert man manchmal die Übersicht, ob der gerade behandelte (und formal immer sehr detailliert dargestellte) Teil-Algorithmus nun in der MÜ bzw. im KIT-FAST-System tatsächlich eingesetzt wird, theoretisch einsetzbar wäre oder aufgrund irgendwelcher prinzipieller Schwierigkeiten gar nicht einsetzbar ist. Viele Abschnitte sind so als Einzelkapitel interessant zu lesen, z.B. die Abschnitte 6.3. Termersetzungssysteme in der MÜ, 6.4. TE-Systeme in KIT-FAST: hier wird ein interessanter Ansatz vorgestellt, GPSG-Analyse als TE-System zu formalisieren. Aber gerade da, wo es spannend wird, bricht der Text ab und wird durch die Dokumentation der Implementierung (von Teilen!?) abgelöst.

Diese besteht aus mehreren Seiten Prolog Klauseln mit vielen sehr technischen Kommentarzeilen, die nur für den Spezialisten verständlich sind. Von größerem allgemeinem Interesse wären die Editoren und Compiler, deren Beschreibung jedoch auf 2 Seiten zusammengedrängt ist. Ebenso isoliert steht die Schlußseite über das MÜ-System.

Das Kap. 8 (Zusammenfassung und Ausblick) präsentiert relativ zusammenhanglos einige Ideen zur Weiterentwicklung, z.B. die

Einführung von Alternationen, was in allen derartigen Projekten und Systemen ein wichtiger Faktor ist. Die anderen Ideen werden nur allgemein angerissen (variable Analysetiefe, Desambiguierung, Reversibilität).

Die beiden Anhänge (Installations- und Benutzungshandbuch) - 100 von insgesamt 240 Seiten - sind nach Meinung des Rezensenten unnötigerweise auf teures Papier gebracht worden - wieso haben die Herausgeber das zugelassen? Das experimentelle MÜ-System ist (sehr lobenswert!) über ftp und Diskette verfügbar - die Anweisung, wie man sich die Dateien auf dem Server der TU Berlin holt, könnte aber die ganzen Anhänge ersetzen, die man dann als Benutzer auf den Files sowieso mit bekommt.

Die Lektoren und Herausgeber der Reihe und des Verlags haben auch eine Menge von z.T. sinnentstellenden und die Lektüre erschwerenden Tippfehlern übersehen - das müßte eigentlich im Zeitalter der automatischen Rechtschreibkorrektur nicht mehr sein:

S. 18: "Im KBMT-System wird zwar der Begriff Transfer vermeiden..." (das hätte allerdings auch das Korrekturprogramm nicht gefunden!)

S. 19: "die Übersetzung des Textes" etc. Ein weiterer Faktor, der bei der Lektüre zusätzlich ermüdet, sind die vielen Wiederholungen, sowohl stilistischer als auch argumentativer

Art:

S. 23: Angestrebte Architektur des MÜ-Systems. .

Die angestrebte Architektur ...

Das angestrebte MÜ-System

Schließlich scheint bei der Liste von Parametern zur MÜ-Architektur einiges mit der Kursivschrift nicht geklappt zu haben (S. 21/22): es ist z.B. aus der Liste nicht zu ersehen, ob KIT-FAST nun "Nur Satz-" oder "Satz- und Textrepräsentation" (oder beides?) auf einer Ebene oder auf jeder einzelnen Ebene bevorzugt, ob nun "Integration" oder "Trennung" von Satz- und Textrepräsentation gewählt worden ist. Das Ganze wird dann im folgenden Kapitel

noch einmal verbal entwickelt - ganz verstanden hat es der Rezensent jedoch immer noch nicht (kann an ihm liegen). Ähnliches passiert bei der kurzen Beschreibung von Regeln des SALAT-Systems (S. 35/36): sind die Regeln (1) und (2) nun Transfer- oder Syntheseregeln - oder beides?

An vielen wiederholten Stellen schließlich erfährt man (mit allen einschlägigen Zitaten) Details über das Berliner MÜ-System und seine Zusammenhänge mit BACK, z.B. Seite 25 und 68 und anderswo; einmal hätte genügt. Man gewinnt manchmal den Eindruck, das Buch sei aus diversen Artikeln zusammengeklebt worden, was durchaus ein gebräuchliches Vorgehen ist; man sollte nur etwas mehr Sorgfalt bei der Herstellung von Redundanzfreiheit anwenden.

Wenn man jedoch die kleinen Mängel in Kauf nimmt (von denen doch einige auch durch bessere Herausgebere tätigkeit hätten beseitigt werden können), hat man mindestens alle Informationen zu einem interessanten MÜ-Projekt in der EUROTRA-Begleitforschung in einem Band zusammen, die ja nach Meinung vieler Experten im Wissenschaftsbereich einen anerkannten Platz gefunden hat.

*Johann Haller, IAI Saarbrücken*

**Wilhelm Weisweber:**

**Termersetzung als Basis für eine einheitliche Architektur in der maschinellen Sprachübersetzung.** Zweite Rezension.

Weiswebers Arbeit ist in dem Rahmen der EUROTRA-D-Begleitforschung an der TU Berlin entstanden, also in einem Forschungsprojekt in dem die wissenschaftliche Fundierung der maschinellen Sprachübersetzung (MÜ) erweitert werden sollte. Das Projekt EUROTRA-D-Begleitforschung ist nun abgeschlossen und Ergebnisse werden vorgelegt.

Die EUROTRA-D-Begleitforschung ist insgesamt positiv zu bewerten. Es ist bei

den Gruppen, geleitet von Christian Rohrer an der Uni Stuttgart als auch Bernd Mahr an der TU Berlin gelungen das immanent anwendungsorientierte MÜ-Vorhaben mit der aktuellen linguistischen Forschung zu verbinden und mit eigenständigen Arbeiten zu fördern. Die EUROTRA-D-Begleitforschung vollbrachte vielleicht erstmalig in der Geschichte, daß MÜ nicht nur mit der aktuellen linguistischen verbunden worden ist, sondern auch für die Linguistik interessante, richtungsweisende Forschungsergebnisse vorlegte. Die Arbeiten der Berliner Gruppe über GPSG (sowie auch die parallelgeführten Arbeiten über LFG in Stuttgart) erweckten eine breite internationale Aufmerksamkeit. In diesem Rahmen ist die Arbeit von Wilhelm Weisweber eingebunden.

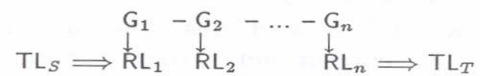
Termersetzung als mögliches linguistisches Modell ist durch Chomskys Präferenz für Produktionssysteme (rewriting systems) verdrängt worden. In der Linguistik, und insbesondere in der Syntax, sind stets axiomatische Symbolmanipulationssysteme benutzt worden, in welchen man aus einem abstrakten Ausgangssymbol  $S$  ausgehend alle Sätze einer Sprache mit Hilfe von (rekursiven) Neuschreibe-Regeln (Generierungsregeln) aufzählen wollte. Dank der Theorie der formalen Sprachen gelang es schon recht früh, den Grammatiken (Aufzählmechanismen) systematisch den leistungsgleichen Automaten (Erkennungsmechanismen) zuzuordnen. Das Beschreibungsmo- dell wurde öfters modifiziert bzw. erweitert: Merkmale sind eingeführt worden. Es entstanden attributierte Grammatiken mit deren Hilfe die syntaktische Analyse verfeinert, die semantische Interpretation modelliert und auch die Textstrukturen erfaßt wurden.

Das inhärente Aufzählungskonzept der generativen Grammatiken kollidierte jedoch mit den MÜ-Modellen (und mit Sprachverstehenssystemen schlechthin), in welchen die normale Intuition die Befolgung eines Abbildungskonzepts erwarten läßt. Die Existenz von verschiedenen Repräsentationsebenen und die Stratifikation der MÜ-Systeme sind natürlich in der Literatur informell öfters behandelt worden (Pause

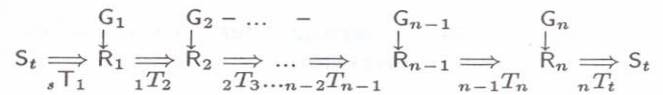
1986, Hauenschild und Umbach 1988, speziell auch in EUROTRA-Bereich durch Arnold et al. 1986:298). Die letzten Konsequenzen sind jedoch nicht gezogen worden.

Weisweber greift die EUROTRA-Modellierung auf, er verlagert aber dabei die linguistische Aufmerksamkeit von der Generierung (d.h. von der Bestimmungsverhältnis von Grammatik  $G_i$  zu Repräsentation  $RL_i$ ) auf die Termersetzung (d.h. direkter Übergang von  $G_i$  zu  $G_{i+1}$ ):

EUROTRA-Sicht:



Weiswebers-Sicht:



wobei die RL-s bzw. R-s die sprachlichen Repräsentationen die G-s die Grammatiken sind. Weisweber sieht zwischen allen Repräsentationsebenen Termersetzungen (doppel-indizierte T-s), d.h. direkte Übergänge vor.

Das Modell von Weisweber enthält zwei Verallgemeinerungen:

1. Für die Übersetzung (und implizit für die linguistische Modellierung überhaupt) ist die Abbildungsrelation zwischen den benachbarten Repräsentationen entscheidend. Die Grammatiken bestimmen lediglich die Wohlgeformtheit der Sätze auf der jeweiligen Ebene. Die Abbildung wird jedoch nicht zwischen den G-s, sondern zwischen den R-s definiert.
2. Die Regeln der Abbildungen von  $R_i$  zu  $R_{i+1}$  sind überall formal gleich, sie können als ein Termersetzungsverfahren verallgemeinert werden.

Wesentlich ist bei Weisweber, daß die linguistischen Operationen nicht als Generierung (bzw. ihre Umkehrungen) aufgefaßt

werden, sondern als Überführung (Abbildung) zwischen den Repräsentationsebenen im Sprachsystem, so zwischen der syntaktischen und der semantischen Repräsentation oder (funktional-)semantischen Repräsentation und Wissensrepräsentation usw. Die Regeln einer formalen Grammatik (die die Sprache konstituieren) sorgen lediglich für die Wohlgeformtheit der Eingabe und der Ausgabe. Die Bestimmung der Zielstrukturen RH geht aus der Eingabestruktur Ri als Ergebnis der Termersetzungen hervor.

Die Arbeit gliedert sich in acht Kapitel und Anhang mit Systemdokumentation.

In Kapitel 1 wird das Thema dargestellt und eingeordnet. Hier werden bereits die kritischen Wissensbereiche angesprochen, die in der Arbeit aufeinander bezogen werden, nämlich:

1. Die Modellierung der Maschinellen Sprachübersetzung,
2. Die Vorarbeiten an der TU Berlin und
3. Termersetzung und Termersetzungsverfahren

Kapitel 2 stellt die MÜ-Modelle dar. In dieser Darstellung handelt es sich nicht um die wissenschaftsgeschichtliche Genese der MÜ-Modelle, sondern um die Suche nach Ansätze der Vereinheitlichung der Systemarchitektur. Weisweber folgt verständlicherweise stark Christa Hauenschild und abstrahiert (über Analyse, Transfer und Synthese hinweg) ein einheitliches Grundmodell, das aus Paaren von Repräsentationen R und RH besteht, die durch zwei kontextfreien Grammatiken  $G_i$  und  $G_{i+1}$  definiert (generiert) werden (S. 17). Zwischen Rund  $R_{i+1}$  werden Termersetzungen definiert.

Kapitel 3 stellt die Produktionssysteme (im Sinne von Davis & King 1977) vor. Insbesondere werden in diesem Rahmen drei MÜ-Systeme dargestellt: SALAT, CAT und TFS. Die Auswahl ist dadurch motiviert, weil diese Systeme eine einigermaßen homogene Architektur aufweisen, gegenüber etwa SUSY, TAUM-METEO oder METAL, die beim Übersetzen in den

verschiedenen Phasen (in der Analyse, in dem Transfer, in der Synthese) unterschiedliche Techniken verfolgen.

Kapitel 4 ist der abstrakteste Teil der Arbeit, hier werden die formalen Grundlagen der Termersetzung behandelt. Das Problem ist, zu gewährleisten, daß die Termersetzungen terminieren. Weisweber greift hier den Knuth-Bendix-Algorithmus auf, der Termersetzungen automatisch durchführt. Bei konfluenten Systemen kann auch die Terminierung der Termersetzungen garantiert werden.

Termersetzungen (=TE) und Termersetzungsverfahren (=TES) sind "bestimmte Ausprägungen" der Produktionssysteme (S. 41). Zwischen den in der Syntax üblichen "formalen Grammatiken" (=FG-en) und Termersetzungsverfahren gibt es zwei Unterscheidungs Momente:

1. Die TES bilden Terme in Terme ab, d.h. sie manipulieren Ketten (oder (Strukturen), während FG-en (Typ 3 und Typ 2) Einzelsymbole umsetzen. Erst bei Typ-O-Grammatiken operieren an Ketten oder Bäumen. TES sind also Operation mit transformationeller Mächtigkeit. Oder anders gesagt: Die Wirkung der TES ist vergleichbar mit der Wirkung von "höheren" (kontextsensitiven, Typ 1 und unrestringierten Typ 0) Grammatiken. Bekanntlich macht die Unentscheidbarkeit die unrestringierten (rekursivaufzählbaren) Sprachen für praktische Verwendung ungeeignet.
2. Die TES sind differenzierter als die einfachen Symbolersetzungssysteme, die lediglich terminale und nichtterminale Symbolen unterscheiden und innerhalb dieser zwei Klassen alle Symbole gleich behandeln. Terme werden durch Signaturen im Voraus (d.h. vor den TES) erzeugt. Sie unterscheiden stets Sorten und Operationen auf beiden Seiten der Gleichungen, insbesondere ist es von vornherein möglich, Verben als Funktoren und die Nominal- und Präpositionalphrasen als ihre Argumente unterschiedlich zu behandeln. Die linke

Seite einer TE erfüllt die Aufgabe der Strukturbedingungen (S. 42), wobei die Einbeziehung der Merkmalnotation möglich ist.

Die nicht-konfluenten Systeme (Kapitel 5) können u. U. auf konfluente zurückgeführt (d.h. lösbar gemacht) werden. Dies ist deshalb bedeutsam, weil die natürlichsprachlichen Systeme nicht immer konfluent sind.

Die Anwendung der Termersetzungsverfahren auf die natürliche Sprach - folgt im Kapitel 6. Vorgestellt wird das System KIT-FAST. Für die mathematisch weniger vorgebildeten bietet wahrscheinlich dieses Kapitel den besten Einstieg. Hier wird die Mächtigkeit der TES anschaulich vorgeführt und verständlich gemacht. Weisweber zeigt, daß die TES auf allen Beschreibungsebenen (Syntaktische Analyse, Funktor- Argument-Strukturen, Transfer und Generierung) eingesetzt werden können und die Leistung erbringen, für die sonst verschiedene Formalismen benötigt worden sind. Es wird hier klar, daß der aufwendige und u.U. schwerfällig wirkende TES-Formalismus für den Linguisten eine erstaunlich einheitliche Behandlung der Sprachverarbeitung ermöglichen wird.

Der eigentliche Textteil des Buches endet mit der Beschreibung der Implementierung des MÜ-Systems (Kapitel 7) und dem Kapitel 8 "Zusammenfassung und Ausblick".

In dem Anhang steckt eine vollständige professionell erstellte Systemdokumentation (A: Installationshandbuch, B: Benutzungshandbuch und C: Beschreibung der Systemdateien). Das System steht übrigens auf dem FTP-Server der TU Berlin zur Verfügung und kann "probiert werden"

ftp.cs.tu.berlin.de

Auch weitere Beschreibungsteile können von dem Autor aus Berlin geholt, bzw. angefordert werden.

Unabhängig davon, ob sie ihm voll zugeschrieben wird, oder als Teil des Forschungsprojektes EUROTRA-D-Begleitforschung angesehen wird, ist Weiswebers Dissertation eine

einmalige Leistung, sowohl aus der Perspektive der Computerlinguistik als auch aus der der theoretischen Linguistik. Denn die TES, kontrolliert durch den Knuth-Bendix-Algorithmus, können nicht nur in Übersetzungssystemen, sondern auch generell in beliebigen linguistischen Vorhaben eingesetzt werden und sie können Sprachen (natürliche Sprachen) mit transformationeller (Typ 0) Komplexität bearbeiten. Es wäre interessant, TES als universelle linguistische Beschreibungsmodelle zu betrachten und die Leistungsgrenzen des Modells genauer zu bestimmen.

Arnold, D.J. /Krauer, S. /Rosner, M. / de Tombe, L./ Varile, G.B.: The (C,A),T Framework in EUROTRA: A theoretically committed notation for MT, in: Proceedings 11th COLING86 Bonn, 1986, S. 297-303.

Chomsky, Noam: Formal Properties of Grammars, in: Luce /Bush and Galanter (eds.) Handbook of Mathematical Psychology 11. Wiley, New York, 1963, S. 323-418.

Hauenschild, Ch. /Umbach, C.: Funktor-Argument-Struktur, Die satzsemantische Repräsentations- und Transferebene im Projekt KIT-FAST, in: Schütz, R. (Hg.) Workshop Semantik und Transfer, EUROTRA-D Working Papers No 6. IAI Universität des Saarlandes, Saarbrücken 1988, S. 16-35.

Pause, P.E.: Zur Modellierung des Übersetzungsprozesses, in: B. Hori, I. /Weber, H.J. (Hgg.) Neue Ansätze in Maschinellem Sprachübersetzung: Wissensrepräsentation und Textbezug, Niemeyer, Tübingen 1988, S. 45-74.

*Istvan Batori*, Univ. Koblenz-Landau

**Hans-Joachim Höll:**  
**Computergestützte Analysen phonologischer Systeme. Exemplarisch am Beispiel einer historisch-vergleichenden Ortsgrammatik aus dem schwäbisch-fränkischen Übergangsbereich.** Sprache und Information, Bd. 27, Niemeyer, Tübingen 1994, XVI, 319 S., DM 148,-.

Hans-Joachim Höll hat sich die schwierige Aufgabe gestellt, ein Programmsystem zur Phonemisierung laut schriftlich vorliegender

Corpora zu entwerfen, das so variabel ist, daß unterschiedliche praktische und theoretische Vorgaben gleichermaßen berücksichtigt werden können. Da gerade im Bereich der Phonologie viele konkurrierende Theorieansätze und ebenso viele unterschiedliche ‚Schulen‘ existieren, muß sich ein großer Teil der Arbeit der Darstellung und kritischen Wertung dieser z. T. stark divergierenden Theorien und Lösungsvorschläge widmen. Der Verfasser stellt schließlich mit dem Programm ‚DiaPhon‘ ein System vor, das viele der anvisierten Ziele erreicht und einem an Computerarbeit gewöhnten Phonologen eine wertvolle Hilfe sein kann!

In der Einleitung skizziert er die Schwerpunkte seiner Arbeit. Der erste besteht in der "Diskussion linguistischer und insbesondere phonologischer Fragestellungen" (S. 9), der zweite befaßt sich "mit bereits existenten phonologischen Analyseystemen" (S. 9) und der eigenen, maschinenunterstützten Analyse einer Mundart des schwäbisch-fränkischen Übergangsbereiches, die von Ottmar Sexauer<sup>2</sup> beschrieben wurde. Diese Untersuchung dient als Folie für die Vorstellung des Programmsystems und seines Leistungsumfanges.

Der erste Schwerpunkt, der im wesentlichen in Kapitel 2 abgehandelt wird, ließe sich auch als Einführung in die Phonetik und Phonologie mit Rezensionsteil bezeichnen: Der Verfasser sammelt die einschlägigen phonologischen Theorien von Trubetzkoy bis Vennemann und sichtet sie kritisch. Gleichzeitig versäumt er es nicht, durch das Gegenüberstellen sich z.T. überschneidender oder widersprechender Theorien zu zeigen, wie stark die zu erwartenden Phonemisierungsergebnisse von dem jeweiligen Theorieansatz abhängen. Eine große Zahl von Fußnoten bringt hier jeweils noch detailliertere Informationen, die manchmal den spannenderen, weil kritischeren, Teil des Textes ausmachen. Insgesamt bündelt dieses Kapitel die zentralen Fragen der Phonologie mit besonderer Berücksichti-

gung des Deutschen und seiner latenten Probleme (z.B. Wertung des Schwa, Affrikaten/Diphthonge, silbische Konsonanten, Sonorität etc.), so daß jeder Student der Phonologie diesen Abschnitt als guten Einstieg in die Examensvorbereitung verwenden könnte. Jede der diskutierten Fragen wird mit einer Bewertung abgeschlossen, inwiefern das Problem für die Ausgestaltung des Programmes zur computergestützten Phonemisierung relevant ist.

Die folgenden Kapitel, die zunächst die Datenbasis der Arbeit beschreiben und schließlich die Verfahren, die der Verfasser für die Auswertung programmiert hat, sind naturgemäß stärker auf formale Aspekte konzentriert, wenngleich auch hier immer wieder der Rückverweis auf die theoretischen Differenzen erfolgt, die sich z.B. ergeben, wenn als Bezugseinheit für die Phonemanalyse einmal die Silbe, im anderen Fall aber eine morphologisch determinierte Einheit oder eine Mischung aus beidem gewählt wird. Der Verfasser verweist hier mehrfach auf die Offenheit des von ihm erstellten Programmes. Es ist seine Absicht, die Analysemodule so zu gestalten, daß der Benutzer die Möglichkeit hat, verschiedene Transkriptionssysteme ebenso zu bearbeiten, wie er z.B. die Auswahl zwischen monophonematischer oder biphonematischer Bewertung der Affrikaten bzw. Diphthonge hat.

Zunächst beschreibt der Verfasser in Kapitel 3 die Datengrundlage seiner Untersuchung: Die oben bereits erwähnte Ortsgrammatik von Sexauer aus dem Jahr 1927 ist wahrscheinlich für den Aufbau und das Austesten des Systems so gut geeignet wie es andere Corpora auch wären (S. 129); die moderate Transkription folgt dem in der deutschen Dialektologie vorherrschenden Teuthonista-Prinzip, das keineswegs "verstärkt" durch das IPA3 abgelöst wird, wie der Verfasser S. 53 behauptet (z.B. verwenden alle Teilprojekte des neuen Gesamtbayerischen Sprachatlas eine Teuthonista-Transkription). Obwohl viele Bemerkungen Sexauers darauf schließen lassen, daß er über ein hohes Probleme-

<sup>1</sup> Leider folgt die Arbeit dem Trend, bei immer geringerer Leistung des Verlages immer teurer zu werden.

<sup>2</sup> Sexauer, Ottmar: Die Mundart von Pforzheim. Leipzig 1927

<sup>3</sup> Internationales Phonetisches Alphabet



wußtsein bezüglich der Unterschiede von Stadtmundart vs. Landmundart verfügt (vgl. Höll, S. 122), ist sein Blick doch sehr stark von der ‚Zerfallsmetaphorik‘ getrübt.<sup>4</sup> In neueren Arbeiten/Datensammlungen ist das Problem der Varietäten besser durchdrungen und daher auch das Datenmaterial so stabilisiert und gekennzeichnet, daß Rückschlüsse auf koexistierende lautliche Systeme sicherer möglich wären (S. 129ff.; vgl. Bellmann 1994). Hier hat der Verfasser bei der Analyse der Sexauerschen Daten, deren Genese unbekannt ist (S. 124), möglicherweise Probleme zu bewältigen, die ein neueres, in dieser Hinsicht problemgeschärftes Corpus nicht enthalten hätte.

Kapitel 4 stellt wieder einen Rezensionsteil dar, in dem die älteren Arbeiten vorgestellt und kritisiert werden, die sich mit der teilautomatischen phonologischen Analyse von Corpora befassen. Es ist immer das Privileg jüngerer Arbeiten, die Fehler und Unschärfen älterer Ansätze schonungslos aufdecken zu dürfen: An dieser Stelle sei jedoch daran erinnert, daß es einen himmelweiten Unterschied ausmacht, ob ein Analysesystem dieser Art - wie hier - auf einem PC mit Turbo Pascal 5.0 konstruiert wird, oder auf Maschinen, wie z.B. der vom Verfasser zitierten, legendären "Sperry Univac 1100/60" (S. 172). Die langen Papierausdrucke waren damals nicht Selbstzweck, sondern die wesentliche Kommunikationsmöglichkeit mit der Maschine überhaupt; m.E. überschätzt der Verfasser in diesem Zusammenhang (vgl. S. 158) die Fähigkeit eines Benutzers, Ergebnisdaten von größerem Umfang auf dem Bildschirm zu erfassen und zu kontrollieren, wengleich ihm natürlich zuzustimmen ist, daß das interaktive Arbeiten mehr "Kreativität und

Spontanität (sie!) eines linguistischen Bearbeiters" (S. 158) zuläßt.

Kapitel 5, das sich diesen Fragen widmet, folgt den heute üblichen Ansichten bezüglich der Mensch-Computer-Interaktion und skizziert die wesentlichen Merkmale, die das Programmsystem DiaPhon aus software-ergonomischer Sicht haben sollte.

Kapitel 6 beschreibt die Komponenten im einzelnen, aus denen DiaPhon besteht: Datenaufbereitungssystem, Datenanalysesystem und Peripheriesystem sind die Kernmodule.

Das Datenaufbereitungssystem ermöglicht das Editieren, Aufbereiten und Speichern eines Corpus. [. . .] Das Datenanalysesystem beinhaltet all die Systemkomponenten, die die wortphonologische Analyse unmittelbar unterstützen. [. . .] Das Peripheriesystem beinhaltet all die Systemkomponenten, die weder zum Datenaufbereitungssystem noch zum Datenanalysesystem gerechnet werden können. Hierzu gehören z.B. die Programmteile, die dem Benutzer die Konstruktion von Zeichensätzen oder die Anpassung der Tastatur erlauben. (S. 167)

Das Kapitel stellt diese Module vor und gibt - orientiert am Hauptmenü - die verschiedenen Analysemöglichkeiten des Programmes mit Beispielen wieder. Wichtige Menüpunkte sind die für eine phonologische Analyse grundlegenden Prozeduren wie

- > Inventaruntersuchungen, die entweder von einer phonemisierten oder nicht phonemisierten Datenbasis ausgehend die Anzahl der vorkommenden Segmente (Phone/Phoneme) auflisten;
- > Oppositionsuntersuchungen, bei denen die Wahl besteht zwischen ‚klassischen‘ Minimalpaaren und sog. subminimalen Paaren; weitere Menüunterpunkte geben u. a. die Möglichkeit, die Suche

<sup>4</sup> Im allgemeinen ist jedoch die Lage derart, daß in den Dörfern des nähere Umkreises, ja auch in noch entfernteren, der Hauptteil wenigstens der männlichen Bevölkerung irgend wie in der Industrie tätig ist. [...] Mit dieser Entwicklung geht, wie gesagt, der Zerfall der Ma. [= Mundart, B.K.] Hand in Hand." (Sexauer 1927, S. 33)

<sup>5</sup> Bellmann, Günther: Einführung in den Mittelrheinischen Sprachatlas (MRhSA). Tübingen 1994. Bellmann beschreibt dort die strengen Maßstäbe, welche an eine moderne "biserielle" Datenerhebung anzulegen sind, die zwei Varietäten (Basisdialekt vs. Regionaldialekt) untersucht.

auf bestimmte Paare zu beschränken. Der Verfasser verweist hier auf Probleme bei der Auswahl der Paare, die sich vor allem daraus ergeben, daß keine Bedeutungsüberprüfung vorgenommen werden kann und somit in den Ergebnislisten nicht nur bedeutungsverschiedene Ausdrücke kontrastiert werden, so daß der Bearbeiter kontrollierend eingreifen muß (S. 218);

- > Untersuchung der distributionellen Relationen, deren wohl wichtigste die Suche nach komplementärer Verteilung ist. Daneben werden Positionsstatistiken (für initiale, mediale und finale Position) erstellt, Listen für Vorgänger- und Nachfolgersegmente ausgegeben sowie Listen für gemeinsame Distributionen und die Silbenstruktur; den Abschluß der Abwählmöglichkeiten bildet eine phonotaktische Analyseeinheit für konsonantische und vokalische Cluster, die recht deutlich die Probleme der vom Verfasser breit diskutierten Frage zeigt, ob die Silbe und/oder morphologisch relevante Einheiten als Untersuchungsbasis gewählt werden sollen.
- > Zwei weitere Menüpunkte geben Listen für die Intensitätsdistribution (Haupt- und Nebentonverteilung) sowie die Möglichkeit zur Ausgabe von Merkmalsmatrizen und zu Suchaufträgen an das Material, die sich nach bestimmten Merkmalen ausrichten lassen.

Den Abschluß der Arbeit bilden ein alphabetisch nach einem Lemma sortiertes Wortverzeichnis der Sexauerschen Belege sowie ein ausführliches Literaturverzeichnis.

Im Resümee kommt der Verfasser nicht um die Feststellung herum, daß die "Analyseergebnisse [. . .] in starkem Maße von der Validität und Reliabilität der Transkription und der Repräsentativität der Materialbasis abhängig" sind (S. 260); dies ist die dialektologische Fassung der alten informatischen Grundweisheit *garbage in - garbage out* .

Offen bleibt die Frage, wie ein Einsatz des Programmsystems bei Daten aussehen muß, welche nicht in der Weise kodiert sind,

die der Verfasser präferiert. Hier verläßt ihn seine Methodentoleranz kurzfristig, wenn er z.B. Bürkle (1988) nachruft:

Ein hoffentlich letztes Credo für eine Codierung phonetischer Symbole als Prototypen findet sich bei BÜRKLE, der an dem traditionellen Verfahren ausdrücklich festhält, obwohl zu dieser Zeit (1988) längst graphikfähige Terminals mit frei programmierbaren Zeichensätzen im Handel waren. (S. 157)

Die geschmähte Kodierung von "Prototypen" (= Kodierung von Einzellaute durch Zeichenkombinationen, z.B. *TJ* als ‚N7‘ (S. 156)) ist kein "traditionelles Verfahren", sondern eine hardwareunabhängige, weltweit übertragbare Lösung, die es jedem Benutzer freistellt, daß - wie der Verfasser selbst sagt - "Letztlich [...] einige wenige, automatisiert durchführbare Substitutionsanweisungen [genügen], um eine Bearbeitung auf einem anderen Rechner [. . .] zu ermöglichen." (S. 157). Als kleiner Vorteil der Prototypen-Lösung ist m.E. anzusehen, daß das Material notfalls auch ohne technische Hilfsmittel lesbar bleibt, während bei der vorgeschlagenen Lösung beliebige Abfolgen von ASCII-Zeichen<sup>6</sup> entstehen, die ohne passenden Bildschirm- oder Druckertreiber unlesbarer Zeichensalat sind. Im gleichen Zusammenhang ist stark zu bezweifeln, daß in bezug auf schwere Lernbarkeit und Fehleranfälligkeit zwischen einer ‚Prototypen‘-Kodierung, der dies vorgeworfen wird (S. 156), und einer beliebig im ASCII-Kode angesiedelten große Unterschiede bestehen: Auch bei dieser müssen ja die ca. 200 Zeichen (S. 181f.), mit denen eine Transkription auskommen soll, in irgendeiner Weise von einem menschlichen Bearbeiter **gemerkt** werden, zumal der Verfasser mittels einer Tastaturanpassung bis zu 20 (!)

<sup>6</sup> Beliebige sind die Folgen deshalb, weil sich erst bei der lokalen Erstellung der Bildschirmzeichen (S. 183f.) entscheidet, welches Zeichen zu welchem ASCII-Wert gestellt wird. Bei der bekannten Erfindungsfreude der Dialektologen ist mit Sicherheit anzunehmen, daß sich jeder eine andere Zuordnung ausdenkt.



Tastaturebenen für die Eingabe der Transkriptionszeichen bereitstellt (S. 186); die ASCII-Zeichen werden durch "Umschalttaste,-Grundtaste" (S. 187) erreicht. Was nun leichter merkbar oder fehleranfälliger ist - N7 oder CTRL-@ ?

Trotz des etwas ernüchternden Eindrucks einer sehr großen Streubreite der phonologischen Theorien, die jeweils Einfluß auf die zugehörige Datenerhebung und -auswertung haben, bleibt zu erwarten, daß die vom Verfasser geleistete Problemdarbietung und Programmierung dem Linguisten eine gute Hilfe bei der Analyse großer Datenmengen sein kann. Aus der Schlußbetrachtung geht hervor, weshalb der Titel der Arbeit ("Computergestützte Analysen") so vorsichtig formuliert ist:

Eine computative phonologische Analyse ohne die aktive Kontrolle eines Linguisten ist aufgrund des Faktums, daß eine allgemein anerkannte Verfahrensweise zur Analyse phonologischer Systeme nicht explizit formuliert werden kann, nicht möglich. Es ist vor allem die *Komplexitätsbarriere* (natürliche Sprachen erlauben eine Vielzahl an Deutungsmöglichkeiten), die eine Vollautomatisierung einer phonologischen Sprachanalyse ausschließt. (S. 260)

Bernhard Kelle, Univ. Freiburg

## Neuerscheinungen:

Stefan Eherer

### Eine Software-Umgebung für die kooperative Erstellung von Hypertexten

1995. IX, 161 Seiten. Kart. DM96,-16s 749.-1 SFr 96,- ISBN 3-484-31929-1 (Sprache und Information. Band 29)

Der Einsatz von Textverarbeitungssystemen hat den Schreibprozeß stark verändert. Neben der Vereinfachung von Korrekturen ist vor allem das Einbinden der auf der Maschine möglichen Medien, wie z.B. Graphik, Bilder oder sogar Ton und Video, die entscheidende Neuerung. Hypertexte erweitern diese Möglichkeiten um maschinengestützte Verweise, die einzelne Texte oder Textfragmente miteinander verknüpfen. Gleichzeitig ist ein Trend zur Verteilung der Aufgaben auf mehrere Autoren zu beobachten, der nicht nur durch die Größe der Texte, sondern auch durch deren inhaltliche Komplexität motiviert ist.

Beide Aspekte werfen jeweils spezifische Probleme auf, die die hier vorgestellte Software-Umgebung lösen soll. Auf der einen Seite muß der Autor eines Hypertextes im Vergleich zu linearen Texten neue Organisationsformen von Texten erlernen, die den Leser bei der Erfassung des Gesamtzusammenhangs eines Hypertextes unterstützen. Auf der anderen Seite müssen die Aktivitäten der Gruppe von Autoren koordiniert und deren Ergebnisse zu einem einheitlichen Ganzen integriert werden.

Um die geforderte Unterstützung zu erreichen, wird auf der Basis von Erkenntnissen der Schreibforschung sowie des computerunterstützten kooperativen Arbeitens ein Modell des kooperativen Schreibens entwickelt und in einer deduktiven Objektbank implementiert. Das Objektmodell erweitert bestehende Modellierungsansätze um abstrakte Ideen, die die inhaltlichen Anforderungen an den zu erstellenden Text beschreiben. Das Prozeßmodell definiert die Aktivitäten der Autoren als hierarchisch geschachtelte, endliche Automaten, deren Zustände durch bestimmte Objektkonstellationen charakterisiert sind. Zustandsübergänge sind die von den Autoren ausgeführten Operationen, deren Anwendbarkeit durch Regeln und Integritätsbedingungen des Objektmodells festgelegt wird.

### Future Perspectives of Dialogue Analysis

Edited by

FRANZ HUNDSNURSCHER and EDDA WEIGAND

1995. IX, 199 Seiten. Kart. DM 124.- 1 6s 967.- 1 SFr 124.-. ISBN 3-484-75008-1 (Beiträge zur Dialogforschung. Band 8)

Contents: Preface. - K. ADAMZIK, Dialoganalyse: eine Disziplin auf der Suche nach ihrer Identität. - F. HUNDSNURSCHER, Dialogue Analysis, the Proper Way. - E. WEIGAND, Looking for the Point of the Dialogic Turn. - E. Roulet, L'analyse du dialogue dans une approche modulaire des structures du discours: L'exemple du dialogue romanesque. - A. CAPONE, Dialogue Analysis and Inferential Pragmatics. - A.H. JUCKER, Discourse Analysis and Relevance. - B. Naumann, Mailbox Chats: Dialogues in Electronic Communication. - H. WESTHEIDE, Fremdsprachenvermittlung auf der Basis dialogischer Texte. Heutige Praxis und Zukunftsperspektiven.

Max Niemeyer Verlag GmbH & Co. KG  
Postfach 2140. D-72011 Tübingen