Rezensionen

Sven Naumann:
Generalisierte Phrasenstrukturgrammatik: Parsingstrategien, Regelorganisation und Unifikation. (Linguistische Arbeiten 212) Tübingen: Niemeyer, 1988. 180 Seiten; kart. 70,- DM

Naumanns Buch ist eine der wenigen deutschsprachigen Abhandlungen über GPSG und dennoch ist es kaum bekannt geworden und wird selbst in späteren verwandten Arbeiten nicht zitiert (Fisher 89, Weisweber & Preuß 92). Motivation genug fünf Jahre nach dem Erscheinen nachzufragen, wer dieses Buch lesen (oder auch gelesen haben) sollte.

Das Buch gliedert sich in drei Teile: eine Einführung in die GPSG, die Vorstellung des von Naumann entwickelten GPSG-Parsers und seine Implementation.

Die GPSG-Einführung behandelt systematisch alle Komponenten einer GPSG, die verschiedenen Regeltypen, die Instantiierungsprinzipien sowie die Merkmalprinzipien. Die Darstellung ist knapp und beschränkt sich teilweise so stark auf formale Definitionen, daß es schwer ist zu folgen. Aber wo sie ausführlicher ist, da ist sie auch kritisch (hinterfragt den Sinn der Vorschläge aus Gazdar et al. 1985 (bekannt als GKPS)), abwägend (zwischen verschiedenen GPSG Versionen) und einordnend (vergleicht GPSG mit TG). An einer Stelle überrascht Naumann damit, daß er zeigt, daß die ECPO-Eigenschaft von kontextfreien Grammatiken nicht die Voraussetzung dafür ist, daß solche Grammatiken in ID- LP Grammatiken umgewandelt werden können, wie dies in vielen anderen Publikationen behauptet wird (z.B. in GKPS
1986:49 oder in Evans 1987:38). Naumanns Argumentation basiert jedoch darauf, daß neue Konstituentennamen beliebig eingeführt werden können. Das belegt, daß er die GPSG-Implementation aus informatischer und weniger aus linguistischer Sicht betreibt.

Im Zusammenhang mit der Erläuterung seines Parsers stellt Naumann klar heraus, welche Probleme bei der Abbildung der reinen GPSG-Lehre in ein ablauffähiges System entstehen. Er wählt dann ein Drei-Phasen-Modell, das zunächst aus ID-, LP-, und Metaregeln eine kontextfreie Grammatik erzeugt, die die Eingabe für den Parser bildet. Dieser generiert für einen Eingabesatz aufgrund der Grammatik und unter Anwendung der Merkmalprinzipien (FFP, HFC, CAP) eine Menge von Strukturbeschreibungen, aus der in der dritten Phase durch Merkmaldefaults (FSD) und Kookkurenzbeschränkungen (FCR) die gültigen Strukturen ausgefiltert werden. Auffällig bei diesem Vorgehen ist vor allem der frühe Einsatz der LP-Regeln, die in ihrer Anwendung auf die Merkmale der ID- und Metaregeln beschränkt bleiben. Ein anderer, zur gleichen Zeit entstandener Ansatz der Berliner Gruppe wendet beispielsweise die LP-Regeln ganz zum Schluß an (Hauenschild & Busemann 1988:14). Dieses Vorgehen bedingt jedoch den Einsatz eines ID-LP Parsers, während Naumann sich auf einen Parser für kontextfreie Sprachen beschränkt.

Naumanns System ist in LISP geschrieben. Die Seiten 99 bis 126 des Buches enthalten den gut annotierten Code.
Das System war seinerzeit auf einem MSDos Rechner entwickelt worden und bot entsprechend unbefriedigende Antwortzeiten. Naumann hat das System getestet mit einigen englischen Beispielsätzen aus

dem Bereich der" unbounded dependencies" . Der Test basierte auf einem Lexikon mit 38 Einträgen und einer "ausmultiplizierten" Grammatik mit 169 kontextfreien Regeln. Im Kapitel 10 werden die berechneten Strukturen präsentiert.

Naumann versucht nirgendwo vorzugeben, daß sein Buch mehr darstellt als es ist, nämlich der Bericht über einen getreuen Ansatz zur Implementation des GPSG-Ansatzes, wie er in GKPS vorgeschlagen wurde. Als solcher (und als gute Einführung in die Theorie) ist er durchaus lesenswert. Es bleibt dennoch die Frage, warum das Buch so unbeachtet geblieben ist. Vielleicht liegt es daran, daß hier ein deutschsprachiges Buch (von einem deutschen Verlag) vorliegt, das jedoch nur englische Syntax behandelt. Das erschwert sicherlich die Akzeptanz in der angelsächsischen Forschungsgemeinde. Der von Uszkoreit (1986) vorgestellte GPSG-Ansatz für das Deutsche wird nur am Rande erwähnt, was wiederum das geringe Interesse im deutschsprachigen Raum erklären mag. Vielleicht liegt es aber auch daran, daß hier ein System in LISP vorgestellt wird, während Prolog-Implementationen seit Mitte der 80er Jahre mehr im Blick
punkt standen. Schließlich mag es auch daran liegen, daß dieses Buch erst erschien, als sich vielerorts das Interesse bereits auf HPSG zu verlagern begann.

Literatur

Evans, Roger: *Theoretical and Computational Interpretations of Generalized Phrase Structure Grammar*. (Cognitive Science Research Paper 085) Brighton, GB: The University of Sussex. August 1987.

Fisher, Anthony: *Practical Parsing of Generalized Phrase Structure Grammars.* Computational Linguistics 15, 3 (1989), 139-148.

Gazdar, Gerald; Klein, Ewan; Pullum, Geoffrey; Sag, I van: *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press, 1985.

Hauenschild, Christa; Busemann, Stephan: *A constructive version of GPSG for machine translation.* (KIT-Report 59) Berlin: TU-Berlin, Projektgruppe KIT. Februar 1988.

Uszkoreit, Hans: *Word order and constituent structure in German.* (CSLI Lecture Series) Chicago University Press, 1987.

Weisweber, Wilhelm; Preuß, Susanne: *Direct Parsing with Metarules.* (KIT-Report 102) Berlin: TU-Berlin, Projektgruppe KIT. Dezember 1992.

Martin Volk, Univ. Koblenz-Landau

Ulrich Schmitz: Computerlinguistik. Eine Einführung Westdeutscher 1992. Verlag, Opladen,

Es handele sich um ein Denkbuch, so der Autor. Dies bedeutet Distanz zum Inhalt und Reflexion darüber, was man tut, wenn man Sprache und Sprachforschung mit Computern zusammenbringt. Es ist eher gedacht als Standortbestimmung der Computerlinguistik als ein Lehrbuch oder eine Einführung ins Studium der Computerlinguistik im engeren Sinne. Hierzu wird ausführlich auf aktuelle Literatur verwiesen. Die Adressaten sind Studienanfänger oder Laien mit geisteswissenschaftlichem Hintergrund, die sich für die Denkweise computerlinguistischen Vorgehens interessieren. Es behandelt kritisch und in der Bewertung ambivalent eine wissenschaftliche Haltung gegenüber dem Gegenstand Sprache, in der der Zweck im Vordergrund steht. Vor allem geht es bei der Instrumentalisierung von Sprache um solche Zwecke, die wiederholbar und technisch in Form von Maschinen nachbaubar sind. Die Frage ist, wie sich diese Mensch-Maschine-Beziehung auswirkt, wenn Sprache im Spiel ist. Ein markantes Beispiel dafür ist die Annahme von Regeln und/oder Schemata, mit denen Wiederholbares im Sprachgebrauch formuliert werden kann. Um Distanz und gleichzeitig eine begründete Technikfreundlichkeit herzustellen, stellt der Autor diese

Perspektive den Möglichkeiten des "Nicht-Geregelten" gegenüber. Insofern geht es um mehr als Computerlinguistik, es geht um "einen gewachsenen Teil unserer Kultur" (S.13). Diese Sichtweise wird an den wesentlichen Themenbereichen der Computerlinguistik aufgezeigt.

Das einleitende Kapitel ("Einladung zur Computerlinguistik") beschränkt sich auf die Nennung des anvisierten Ziels des Buchs sowie auf allgemeine Angaben zum Fach Computerlinguistik. Plastisch dargestellt sind die Anwendungsbeispiele, die den Zugang zu den sprachlichen Problemen und damit zur linguistischen Herangehensweise eröffnen. Neben den traditionellen Bereichen, die Form und Struktur in den Vordergrund stellen, werden prozedurale Ansätze skizziert, die die Verarbeitung von Information thematisieren. Im weiteren werden statistische Methoden und die Arbeit an Wörterbuchdatenbanken aufgezeigt. Nicht vergessen bleibt die gegenwärtig (noch) vorherrschende Tendenz der Entwicklung von Grammatikformalismen. Die Skizze von Nachbarfächern, Forschungsaufgaben und Studiengängen rundet den Überblick ab. Die Bemerkungen zur Berufsperspektive bleiben situationsgemäß sehr vage.

Im 2. Kapitel geht es dann bereits um grundsätzliche Fragen (" Wissenschaft und Technik: das Leben im Griff"). Besprochen werden formale und maschinenorientierte Verfahren (z.B. Abbildung und Simulation oder die Verwendung von Schemata). Im weiteren gibt es eine Kurzeinführung in die Programmiersprache PROLOG, die den bereits Kundigen die Prinzipien klarer macht, zum Erlernen aber wirklich zu kurz ist. Deutlich werden soll die Eingeschränktheit solcher Ansätze im Vergleich zu Fragestellungen, die sonst noch möglich waren.

Konkret auf die traditionellen linguistischen Teilbereiche geht das 3. Kapitel ein ("Menschliche Sprache und mathematische Form"): Phonetik, Morphologie, Syntax, Semantik, Pragmatik, Textlinguistik. Zugleich handelt es sich hier um die Kerngebiete der Computerlinguistik. Sie werden unter dem Gesichtspunkt des analytischen Denkens, des Meßbaren und der Regel diskutiert. Immer wieder erscheint der Hinweis auf das Gegenteil, unterstützt durch Zitate abendländischer Denker (u. a. Derrida, Schleiermacher, Wittgenstein), daneben Zitate von Sprachwissenschaftlern wie Chomsky, Fillmore, Paul, Saussure.

Die beiden folgenden Kapitel setzen sich mit einigen der Hauptanwendungsgebiete der Computerlinguistik auseinander.

Die Mensch-Maschine-Kommunikation steht im 4. Kapitel im Vordergrund und damit grundsätzliche Unterscheidungen zwischen Menschen und Maschinen ("Menschen handeln, Maschinen operieren"), zwischen natürlicher Sprache und Programmiersprachen. Nichtsdestoweniger wird der Gesichtspunkt eingebracht, daß neue Werkzeuge das Handeln der Werkzeuggebraucher beeinflussen. Andersherum gesehen, werden natürlich nur solche Werkzeuge erfunden, die einer bestimmten Geisteshaltung ihrer Gebraucher entsprechen und vor allem: sie werden "anthropomorphisiert". Der Abschnitt" Über Computer sprechen" behandelt den geläufigen metaphorischen Sprachgebrauch. "Mit Computern sprechen" geht auf Dialogmodellierungen ein, wobei (selbstironisch) ebenfalls Metaphorisierungen eingebracht werden ("die Maschine als Lehrer" und "die Maschine als Wissens-Diener"). Schließlich aber wird ganz dezidiert Stellung bezogen: Mensch-Maschine-Kommunikation ist tot (und daher unter Umständen auch attraktiv) und etwas ganz anderes als Mensch-Mensch-Kommunikation, die lebendig und durch Gefühle geprägt ist. Gleichwohl ist eine Vielzahl kommunikativer Vorgänge zwischen Menschen auf der Ebene von Operationen anzusiedeln. In diesem Sinne ist das 5. Kapitel der automatischen Textgenerierung gewidmet (" Texte von Menschen und Maschinen"). Hier bringt der Autor einen eigenen Ansatz ein, was sich posi-

tiv durch stärkere Konkretheit des Beispiels bemerkbar macht. Es geht um die Generierung von Nachrichtentexten. Das Modell ist entwickelt auf der Grundlage von authentischen Fernsehnachrichten (Textmaschine für die Tagesschau). Es wird versucht zu zeigen, daß hier ein Bereich vorliegt, in dem Regeln und Schemata, Wiederholbares und Voraussehbares, die Produktion von Nachrichtentexten bestimmen. Ein Katalog von semantischen Stereotypen (z.B. Besuch, Treffen, Vereinbarung) wird in einer Datenbank gespeichert. Die aktuelle Version wird nach bestimmten Regeln gemischt und durch Referenzangaben (Spezifikationen von Personen, Zeit und Ort) ergänzt. Fazit der Geschichte: Menschen machen Texte in der Art, wie sie für Maschinen modelliert werden (Informieren anstelle von Argumentieren). Insofern sind Werkzeugkonstruktion und Werkzeuggebraucher nicht so weit voneinander entfernt! Zum Schluß (6. Kapitel: "Computer als Werkzeug") nochmal die Gegenüberstellung von Reduktion und Grenzen einerseits und Nützlichkeit andererseits. Und dann schließlich noch ein kleines Fragezeichen in bezug auf zukünftige Entwicklungen. Der Ausgang der Geschichte ist offen.

Generell zum Buch läßt sich folgendes sagen. Beim Lesen sind Personen- und Sachregister hilfreich. Das Personenregister skizziert überdies einen weit gespannten Rahmen, innerhalb dessen eine Vielzahl von Sprachphilosophen und Sprachwissenschaftlern zu Wort kommen. Das Sachregister gibt einen guten Überblick über die behandelten linguistischen Themen. Zum Weiterarbeiten hilft ein umfangreiches Literaturverzeichnis. In Einführungskursen ist das Buch m.E. vor allem begleitend nützlich. Gut ist, daß alle Probleme durch Beispiele klar gemacht werden. Dies betrifft sowohl die sprachliche und linguistische Ebene als auch die Implementierungsebene, wiewohl die oft sehr knappe und vereinfachende Darstellung für tatsächliche Anwendungen nicht ausreicht. Will man allerdings der komplexen Sichtweise und der argumentativ aufgebauten Darstellung tatsächlich gerecht werden, so ist etliches an Vorwissen und intensivem Mit-Denken Voraussetzung. Den größten Gewinn wird die Lektüre daher für die LeserInnen bringen, die die Ansätze, in die eingeführt wird, bereits kennen. Für sie jedoch ist es ein Buch, in das hineinzusehen sich auch zum wiederholten Male lohnt.

Annely Rothkegel, Univ. Saarbrücken

G. Leitner (Hsg.): Theorie und Praxis der Korpus-Analyse: New Directions in English Language Corpora. (Topics in English Linguistics; Bd. 9). 368 Seiten, Berlin, New York: Mouton de Gruyter, 1992

Einleitung

Die Untersuchung von Korpora hat gegenwärtig Konjunktur. Dies erklärt sich aus zwei Faktoren:

1. Der erste ist das weitgehende Mißvergnügen an bisherigen Bemühungen im Bereich der formalen Linguistik. Angetreten mit dem Anspruch, Sprache (und das hieß damals: beliebige Texte) formal beschreiben zu können, hat die Zunft sich mittlerweile eines schlechteren besonnen. Die Diskussion um die verschiedenen Forma!ismen hat deren Eleganz mit der weitgehenden Abstinenz erkauft, die Daten zu betrachten: Es konnten immer weniger Phänomene immer eleganter erklärt werden, und die Beispiele sind immer weniger nachvollziehbar geworden (in meinem Dialekt ist dieser Satz möglich). Der Versuch, Grammatiken mit größerem Abdeckungsgrad zu er-

stellen, wenn er denn überhaupt unternommen worden ist, hat zu dem Eindruck geführt, die Aufgabe sei erstens langwierig (was zutrifft), zweitens mit den gegenwärtigen Formalismen nicht zu machen (was weitestgehend ebenfalls zutrifft), und drittens prinzipiell endlos, weswegen man in Richtung auf probabilistische Grammatiken und eben Korpora ausgewichen ist. Auch bei den Sponsoren von NLP-Projekten (wie der EG) ist die Bereitschaft geschwunden, Projekte mit nicht einsetzbaren Ergebnissen zu fördern (der BMFT ist hierin eine Ausnahme). Dies hat speziell in den USA dazu geführt, über die Kriterien der Abnahme von Projekt-Ergebnissen nachzudenken und die Frage, was ein System eigentlich leisten muß, mit dem Hinweis auf eine Sammlung von Texten zu beantworten, die ein System analysieren können muß.

2. Der zweite Faktor des Interesses an Korpus-Arbeiten liegt in den überraschenden Möglichkeiten, die eine robuste und schnelle, meistens statistisch basierte Analyse großer Textmengen mit sich bringt.

   t> Eine Vorreiterrolle haben hier die Sprachmodelle in den Speech-Understanding-Projekten gespielt, die in den praktischen Einsätzen klare Dominanz über linguistische Ansätze erzielen konnten. Kaum ein Projekt arbeitet ohne solche Techniken.

   t> Eine Ausweitung der Techniken auf den Bereich der Übersetzung (über die Terminologie-Zuordnung hinaus) ist zwar bisher nicht erfolgreich, hat aber den Effekt gehabt, daß die Thematik des Alignments von bilingualen Korpora studiert und mit überraschend guten Ergebnissen angewandt worden ist.

   t> Basierend darauf konnte man sich der Identifizierung von mehrsprachiger Terminologie, der Zuordnung von Textteilen und der Erstellung entsprechender Referenz-Datenbanken zuwenden; erste Produkte dieser Art sind bereits erhältlich.

   t> Im Lexikon-Bereich wurden Korpus-Techniken angewandt bei der Identifizierung von Verb-Argumenten, einem notorisch komplexen Thema (Arbeiten von Shieber), bei der Ermittlung von semantischen Definitionen aus Lexika (Projekt Aquilex), bei der Disambiguierung von Lesarten mithilfe bilingualer Korpora, oder bei der Ermittlung semantischer Ähnlichkeiten (Arbeiten von G. Ruge).

   t> Im Syntax-Bereich sind Fragen der Frequenz bestimmter Strukturen von Interesse, die das Parsing leiten können; im Zusammenhang damit Fragen von fachsprachlichen Varianzen syntaktischer Patterns; und die Fragen nach der Repräsentation solcher Information.

All diese Arbeiten zeigen, daß die Analyse von Korpora ein Feld ist, das Erfolge zumindest im Language Engineering verspricht. Die Frage dabei ist, wie sich statistische und linguistische Techniken kombinieren lassen, um linguistisches Wissen durch die Untersuchung der praktischen Sprach-Benutzung optimal anwenden zu können.

Das vorliegende Buch stammt aus einer anderen Tradition. Es handelt sich um die Proceedings der 11. Konferenz des ICAME 1990 in Berlin, des *International Computer Archive* of *Modern English.* Diese Institution hat seit jeher starke Wurzeln in der

corpus-analytischen Tradition, die sich umgekehrt soziolinguistischer Argumente bedient. Dort hat sich ja von Anfang an eine gewisse Skepsis gegenüber dem Konzept des idealen Sprecher-Hörers gehalten, und man ist den Fragen, was ein Sprachsystem konstituiert, mithilfe der Empirie, und das heißt, mithilfe von Sprachkorpora, nähergetreten. Jetzt finden diese Arbeiten Interesse bei der NLP-Zunft; und aus diesem Gesichtspunkt, und aus dem Gesichtspunkt eines praktischen Interesses an der Entwicklung nicht-trivialer NLP-Systeme, ist die folgende Besprechung geschrieben.

1. Corpus Design and Text Encoding

Das Buch beginnt enttäuschend. Der erste Abschnitt behandelt methodische Fragen beim Design von Korpora. Es finden sich Beiträge von

t> Pieter de Haan: The optimum corpus sampie size. Er stellt fest, daß die Größe von Korpora vom Untersuchungszweck abhängt, weil die zu bearbeitenden Phänomene in entsprechender Signifikanz vorkommen müssen.

t> J. Clear: Corpus Sampling. Er gibt einige Prinzipien an, die man beim Anlegen von Korpora beachten sollte ("core" language, Texttypen-Einteilung, Unterscheidung zwischen Sample und Monitoring Korpus) und definiert Zielgruppen des Oxford Korpus.

t> G. Leitner: International Corpus of English (ICE): Corpus design - problems and suggested solutions. Er präsentiert die Taxonomie des ICE und die wesentlichen Design-Prinzipien und schlägt einige Änderungen vor: In einem Top-Down Approach sollen zunächst "large-scale socio-communicative environments that (co-)determine their own specific socio-communicative norms and the choice of language" (48) betrachtet werden, darunter die jeweiligen "major communicative events", die zeit- und ortsgebunden sind, und dann die Einteilung nach "textual and linguistic structure in text linguistic categories".

t> J. Kirk: The Northern Ireland Transcribed Corpus of Speech. Er präsentiert diese Sammlung gesprochener Sprache (240 K Wörter) und einige Probleme bei ihrer Erfassung.

t> Chr. Mair: Problems in the compilation of a corpus of Standard Caribbean English: A pilot study. Der Beitrag stellt sich die Frage, ob Caribbean English ein eigenes Sprachsystem sei.

t> L. Burnard: The Text Encoding Initiative: A progress report. Der Beitrag berichtet über den damaligen Stand der TEI, kurz nach Erscheinen der ersten Version der Guidelines. Hier hat sich mittlerweile ja einiges getan, so daß der Beitrag etwas zu spät kommt, um noch völlig aktuell zu sein.

Wenn man den Beiträgen des ersten Abschnitts nun Glauben schenken soll, so ist das Erstellen von Korpora - das ist das Enttäuschende - eine aussichtslose Sache:

t> Man weiß nicht, was man sammeln soll: "the phenomenon to be sampled is poorly defined" (Clear, 21).

t> Die Sammlung folgt bestimmten Prinzipien und theoretischen Ambitionen, die doch andererseits sich erst erweisen sollen: "The sampling problem is precisely that a corpus is inevitably biassed in some respects" (Clear, 23).

t> Man weiß nicht, wieviel man sammeln soll: "the suitability of the sample depends on the specific study that is undertaken, and there is no such thing as the best, or optimum, sample size as such". (de Haan, 4; dagegen Clear, 40: "more is definitely better").

t> Man weiß nicht, wer sammeln soll, weil das Ergebnis vom Befrager abhängt: "where the fieldworker used dialect forms, the informant cooperatively followed, but in some cases not until then" (Kirk, 68); außerdem sollte er vom Torfstechen und vom Guinness-Trinken etwas verstehen.

t> Man weiß nicht, was man mit den einmal gesammelten Texten anfangen soll: Jeder interpretiert die Daten, wie sie ihm wichtig scheinen (Mair, 78), und übersieht vielleicht das Wesentliche.

t> Wenn man schon sachlich nichts weiß, sollte man sich zumindest auf eine einheitliche Darstellung einigen (TEI mit SGML), aber auch hier gibt es berechtigte Sorgen: "The obvious risk that, in endeavoring to please all , the TEl may end by pleasing no-one" (Burnard, 106).

Das Problem an einem Beispiel:

1. Der Vorteil der Korpus-Arbeit soll sein, daß eine vieldiskutierte Problematik des Englischen, nämlich" whether it is one system or a set of overlapping, differing codes would be treated as an empirical issue" (Leitner, 33).

2. Mair führt dies am Problem des Caribbean English durch: Unterscheidet sich Caribbean English vom Englischen? Vorsicht: Gibt es das Caribbean English seinerseits überhaupt? Müßte man nicht annehmen, daß auf Jamaica, Trinidad, Tobago vielleicht ein jeweils eigenes Sprachsystem herrscht? "A wrong assessment of the situation at the time of the collection of the texts for the corpus will have fatal results" (Mair, 77). Dasselbe Argument für die verschiedenen Regionen, Schichten usw. von Trinidad ...

3. Selbst wenn man dann Unterschiede entdeckt: Begründen sie ein eigenes Sprachsystem, oder handelt es sich nur um einen Ausrutscher? "In evaluating the evidence, one faces the usual problems of having to decide whether an instance of unusual verb complementation is due to a typographical error, e.g. omission of a preposition, or whether one is dealing with a genuine linguistic datum" (Mair, 87).

4. Sodaß sich die ursprüngliche Fragestellung mit Hinweis auf Korpus-Arbeiten gerade nicht beantwortet: "a computerised corpus will show whether or not these are systematic enough to posit an independent local norm" (Mair, 85), weil man Normen als solche nicht beobachten kann, sondern allenfalls ihre Manifestationen; und weil sich immer darüber streiten läßt, ob es sich wirklich um Manifestationen der Norm handelt.
   Alles hängt eben von der Bewertung durch den Interpreten ab, sodaß sich die postulierte Objektivität der Korpus-Arbeit zuletzt wieder in den Standpunkt des Forschers verflüchtigt
   …

Tröstlich ist nun freilich, daß die meisten Autoren nicht konsequent sind und ihre eigenen Beteuerungen der Unmöglichkeit der Korpus-Arbeit nicht allzu wichtig nehmen; denn faktisch sammeln sie ja die Daten, teilen sie ein, finden auch interessante Phänomene; und das sind die Teile des Buches, die wirklich lesenswert sind.

2. Automated Text Analysis

Der zweite Teil des Buches behandelt Probleme bei der praktischen Arbeit mit Korpora, untersucht Fragen der Text-Aufbereitung, der Annotationen und des Text-Retrievals.

Der erste Beitrag (N. Belmore, Pinpointing problematic tagging decisions), diskutiert Unterschiede im Tagging zwischen Brown und LOB corpus; sie sind vor allem

auf verschiedene Arten der Tokenisierung zurückzuführen (ist can't als ein oder zwei Tokens zu behandeln?), und der Beitrag berichtet von Arbeiten, die eine Synchronisation der beiden Systeme gestatten sollen; das eigentliche Tagging-Thema (Zuweisung von Kategorien zu Tokens) wird aber noch gar nicht behandelt.

Zwei Beiträge berichten von niederländischen Projekten (DEVIL und LINKS). Der Beitrag von W. Meijs zu "Inferences and lexical relations" ist ein Irrläufer in diesem Buch; er versucht, Wort bedeutungen mit atomaren Prädikaten des Typs *baby → child (x) and recently_born (x)* zu behandeln, was nun oft genug zu nichts geführt hat. Interessant ist, daß die Bedeutungsdefinitionen in LDOCE untersucht worden sind, und Meijs stellt fest, daß sie mitnichten so hierarchisch klassifiziert sind, wie die AI-schemata suggerieren, sondern daß sie oft zirkulär sind, Lücken enthalten, und daß Verweise fehlen, die man intuitiv machen würde *(person* und *animal* z.B.). Das weist darauf hin, daß Korpus-Arbeiten immer noch der verständigen Aufbereitung bedürfen, um sinnvoll benutzbar zu sein, und die lexikalische Arbeit zwar erleichtern, aber nicht ersparen. Der zweite Beitrag (S. Janssen: Tracing cohesive relations in corpus sampies using dictionary data) ist interessant zu lesen: Sie entwickelt ein Verfahren zur Disambiguierung, in dem zunächst lexikalische Bedeutungsdefinitionen analysiert werden; daraus werden Genus-Terme gefunden, zwischen denen dann Hyponym/Hyperonym-Relationen bestimmt werden. Es gelingt der Autorin, sechs Bedeutungen von rich in LDOCE korrekt kontextuell zu disambiguieren. Die Technik ist erwägenswert, indem man nämlich entsprechende Kontext-Termini identifiziert, die in den Lexikon-Definitionen verwendet werden, und zur Laufzeit versucht, die besten Matches solcher Termini mit dem aktuellen Text zu finden.

Die Beiträge von St. Fligelstone (De veloping a scheme for annotating text to show anaphoric relations) und G. Sampson (SUSANNE - A deeply analysed corpus of American English) zeigen Schwierigkeiten, die auftreten, wenn man komplexe linguistische Phänomene mit Korpora behandeln will. Fligelstone berichtet vom Versuch, Anaphora zu taggen; dazu ist in der Lancaster Grammar factory ein eigener Editor für die entsprechenden Markups erstellt worden. Der Versuch ist nicht erfolgreich gewesen, da das Phänomen theoretisch nicht klar abgegrenzt war: Trotz eines Manuals von über 100 Seiten war es nicht möglich, das Corpus konsistent zu behandeln ("consistency could not be achieved" (161); und das Ziel, ein Phänomen theorie-neutral zu beschreiben, wird als "unattainable objective" (165) gesehen. Ähnliche Ergebnisse sind aus dem Beitrag von G. Sampson zu folgern, der das Göteborg Korpus mit syntaktischer und semantischer Tiefeninformation anreichern wollte, ebenfalls theorie-neutral. Ergebnis ist die Feststellung, daß bereits der Versuch, semantische Tiefenkasus (diesmal auf Stockwell basierend) für 150 englische Verben konsistent zu markieren, bezogen auf das gegebene Korpus, nicht gelungen ist (187), sodaß man sich wieder auf die "klassischen" Rollen (Subjekt, direktes j indirektes Objekt), und die "klassischen" Adverbialen Angaben (time, place usw.) zurückgeworfen sah (und sieht).

Diese Ergebnisse sind erwägenswert, erstens weil sich zeigt, daß konsistente Korpus-Arbeiten nur gemacht werden können, wenn die zu markierenden Phänomene zuvor bereits klar durchdrungen sind. Dann stellt sich aber die Frage, was die Korpus-Arbeit noch leisten kann, außer Häufigkeitsanalysen: Der Anspruch, daß die Korpora solche Phänomene erst erweisen, konterkariert sich, wenn diese zuvor bereits gewußt sein müssen. Zweitens sind gerade solche komplexeren Themen sicherlich nicht theorieneutral zu beschreiben; und das beeinträchtigt die Wiederverwendbar-

keit und damit den Wert der gesamten Markierungs- Arbeiten.

Der letzte Beitrag in dieser Gruppe (S.Sutton, A. McEnery, Information Retrieval and Corpora), stellt fest, daß "the field of IR is clearly relevant to the tasks of accessing and using machine readable corpora" (208). Diese Auskunft ist billig zu haben, verpaßt aber die Chance, die Vorteile der IR-Techniken in Datenorganisation, Zugriff usw. genauer zu diskutieren (wie überhaupt den Fragen der Organisation und des technischen Managements von großen Korpus-Daten kein gesteigertes Interesse zu gelten scheint: Man hat den Eindruck, die Welt-Festplatten-Kapazität ist unerschöpflich). Die Verweise auf Hypertext- und wissensbasierte Systeme sind dagegen nutzlos, weil auch im Feld der IR unklar ist, wie solche Techniken in größeren Applikationen funktionieren sollen.

Insgesamt zeigt der zweite Teil doch einige Grenzen der Korpus-Arbeiten auf und stellt klar, daß es den Königsweg für die computerlinguistischen Arbeiten (viel Preis ohne Schweiß) auch hier nicht gibt. Die automatische Extraktion von Bedeutungen aus Lexikon-Definitionen erfordert ein erhebliches Maß an Korrektur- und Nachbearbeitungs-Aufwand, wenn man wirklich gute Ergebnisse haben will; und die Hoffnung, komplexere Phänomene aus Korpus-Annotationen extrahieren zu können, wird man wohl deutlich dämpfen müssen. Vorläufig scheint es leidlich möglich, Wortklassen-Angaben und darüber errichtete syntaktische Patterns zu markieren und auch zu benutzen. Komplexere Phänomene scheinen aber vorerst außerhalb der Zugriffsmöglichkeiten der Korpuslinguistik zu liegen.

3. Corpora in Language Description

Der dritte Abschnitt versammelt eine ganze Palette von Arbeiten und Themen aus der praktischen Arbeit und dem Umgang mit Korpora. Zwei Beiträge befassen sich empirisch mit der Frage, ob English ein oder eine Menge von Sprachsystemen sei.

E. W. Schneider (Who(m)? Case marking of wh-pronouns in written British and American English) kann diesbezüglich keine signifikanten Unterschiede erkennen, lediglich textsortenabhängige (whom ist häufiger in US-religiösen und UKregierungs amtlichen Texten): "In many, in fact most respects of this fairly ,~ubtle area of grammar, British and America:Q. English behave identically" (242). Das ist zwar ein gutes Resultat, stellt den Autor aber nicht zufrieden: Vielleicht gibt es noch Nischen, Dialekte, subtilere Phänomene? Weitermachen!: "it can be stated that the diversity of the two major varieties of English appears to be further-reaching in scope but subtler in kind than might have been suspected" (243). Hier ist der Wille der Vater der Forschung.

Auch S. V. Shastri (Opaque and transparent features of Indian English) kommt zu dem Ergebnis, "that the most obvious transparent feature of IE is code-mixing" (273), opaque features "turn out to be statistically nonsignificant" (274). Die Hoffnung, im indischen Englisch ein eigenes Sprachsystem vor sich zu haben, gibt er allerdings ebenfalls nicht auf: "all this points towards the need for systematic quantitative studies based on large databases before we can say anything decisive about variety features" (274).

Zwei Beiträge bestätigen bereits anderweitig bekannte Probleme und Lösungsansätze. Dazu gehören einmal G. Barnbrook (Computer Analysis of spelling variants in Chaucers *Canterbury* Ta*les),* er untersucht Schreibvarianten im Mittelenglischen mit bekannten Techniken von Spell Checkern (add final letter / add initial letter / change two letters usw.), kommt dabei mit seinem Problem ganz gut zurecht und berichtet die bekannten Probleme (inkorrektes Matching usw.).

G. Knowles (Pitch Contours and tones in

the Lancaster /IBM spoken English corpus) berichtet über Versuche, Akzente akustisch zu ermitteln; er stellt fest, daß die FO-Analyse allein dafür nicht ausreichend ist, und weist auf Kontextabhängigkeiten hin: fallende oder steigende Konturen begründen Bedeutungsunterschiede und hängen gleichzeitig von ihnen ab. Dieses Ergebnis bestätigt entsprechende deutsche Untersuchungen.

Kay Wikberg (Discourse category and text type classification: Procedural discourse in the Brown and LOB corpora) diskutiert die Klassifikationen der Texte (Love Story, detective fiction usw.) und weist korrekt darauf hin daß »discourse topic hardly qualifies as a serious candidate for scientific classmcation" (249). Dies führt er aus an der Kategorie E von LOB, der Texte versammelt wie homecraft, food, trade, professional journals, travel, pets, hobbies usw. Der Hinweis, daß die Klassifikation ein bißchen linguistisches Wissen (z.B. im Bereich des procedural discourse) berücksichtigen sollte, ist berechtigt, einige interessante Details (daß Gebrauchstexte viele *if/when* Konstruktionen, viele Aufzählungen und Bullet-Elemente enthalten) fallen auch ab.

Der Beitrag von A. Renouf (What do you think of that: A pilot study of the phraseology of the core words of English) ist ein Beitrag für das "Handbuch des unnützen Wissens". Sie sucht im Birmingham Korpus Ketten, die nur aus high-frequency words bestehen (Birmingham corpus). Die längsten Ketten sind:

t> *he* was and *what he* was and *he* was *not* to be

t> *it* not to be *there but it* was *there* and *when the*

t> *it* was for you *it* was all for you you *said* 1.

Daß gewisse Patterns auch Sinn ergeben, läßt sich gar nicht vermeiden (all you *have to* do *is, in such* a way as to, *it* was one of *the most). What* do you *think* of *that?*

Auch der Beitrag von Chr. Geisler (Relative Infinitives in spoken and written English) tendiert in diese Richtung; er findet, daß wissenschaftliche Texte mehr Passiv-Konstruktionen enthalten, daß Face-to-Face-Texte mehr Personalpronomina enthalten (222), und daß Pronomina, die auf *thing, -one,* und –body enden, in fiktionalen Texten häufiger sind als in nichtfiktionalen. Erweitert das jetzt unseren Kenntnisstand über Infinitive?

Die letzten drei Beiträge des Buches sind fast die interessantesten vom Standpunkt der NLP-Anhängerschaft.

H. Hasselgard (Sequences of spatial and temporal adverbials in spoken and written English) untersucht Stellungen von Adverb-Phrasen. Sie unterscheidet *clusters* (freie Adverbiale) und *combinations* (wohl valenzgebundene). In ihrem Korpus (736 Sätze) überwiegen die *clusters.* Wenn dabei mehrere Adverb-Phrasen auftreten, stehen sie zusammen, und zwar Orts- vor Zeit-Angaben. Dies gilt nicht für *combinations.* Temporale Adverbiale sind dabei mobiler als lokale; speziell im gesprochenen Englisch wandern temporale Adverbien gern in die initiale Position (was zu 50% am Adverb *then* liegt, das zu Zwecken der Textkohäsion benutzt wird). Offen bleiben zwei Fragen: Wie verhalten sich die Adverbien vs. Adverbiale in dieser Hinsicht (mobil sind wohl v. a. die kurzen Adverbien, Adverbiale aus mehreren Wörtern müßten wohl eher zur Cluster-Bildung neigen)? Und, interessant für unsereinen: Wie stellt man solche In-formationen in einer formalen Grammatik dar?

Der Aufsatz von G. Kjellmer (Grammatical or nativelike?) ist deswegen interessant, weil er genau das Dilemma der gegenwärtigen NLP-Grammatik-Arbeiten bespricht: Nicht alles, was grammatisch möglich ist, ist auch akzeptabel, oder wird auch gesprochen. Kjellmer spricht von" the existence of lexically-based restrictions that operate wi-

thin the rule-defined field" (329) und gibt dafür drei Beispiele (aus dem Brown Corpus):

> 1. Verbal tenses: Manche Verben kommen fast nur im Präsens vor (amount illustrate *derive denne distinguish),* andere fast nur in Past Tense *(exdaim pause smile mutter peer).*
>
> 2. Den Infinitiv nach TO benützen manche Verben oft *(TO purchase minimize dear free recover),* manche selten *(amount differ intend deduct range).*
>
> 3. Passivity: manche Verben stehen gern im Passiv *(baptize subject situate tempt retrieve institute),* manche fast nie *(come get want wish like love talk).*

Diese Beobachtungen werfen die Frage auf, wie solche Phänomene in einer formalen NLP-Grammatik darzustellen sind: Man kann ja nicht behaupten, daß ein Verb wie *baptize* nie im Aktiv auftritt; es tritt eben nur sehr selten im Aktiv auf. Eine formale Sprachbeschreibung muß solche Phänomene ausdrücken können, d.h. sie muß Schemata von Präferenzen, Gewichtungen usw. viel mehr in ihre Überlegungen einbeziehen, als das bisher getan worden ist. Und der Ausgangspunkt dieser Gewichtungen muß im lexikalischen Bereich liegen.

Der letzte Aufsatz (J. Noel: - Collocation and bilingual text) "explores a computerized procedure for uncovering collocation data in bilingual texts" (346). Die Arbeit variiert die Church-Methode des Alignment, indem sie es auf Zeilenbasis berechnet (Fr-En mit 30:35 Zeichen pro Zeile); damit werden 2700 Zeilen eines bilingualen Korpus behandelt. Die Markierung der Kollokationen, und das ist das Enttäuschende, wird aber' per Hand vorgenommen, und die Aussage, daß man manches findet, was nicht im Lexikon steht, verliert an Gewicht, da man eben nur das findet, was man vorher im Text markiert hat. Die Frage wäre gewesen, den Versuch einer automatischen Analyse zu wagen und zu sehen, ob es relevante Patterns gibt, und ob diese im Lexikon zu finden sind oder nicht. Ansonsten hat der Ansatz seine Schranken bei Sprachpaaren, in denen Kollokationen stark im Satz (und d.h. über die Zeilengrenzen hinaus) verschoben werden können; hier scheinen die "klassischen" satzbezogenen Techniken robuster zu sein.

Insgesamt fällt im dritten Teil auf, daß einige Beiträge unter dem Fehlen einer klaren Aufgabenstellung leiden, und eben Wissenswertes und weniger Wissenswertes in einerlei Gewand versammeln.

Ein Interesse der Beiträge liegt darin, herauszufinden, ob das Englische *ein* Sprachsystem ist, oder aus vielen Subsystemen besteht (d.h. ob indisches und US-Englisch verschiedene Sprachen sind); die Beiträge im vorliegenden Band bestätigen diese Annahme eher nicht, ihre Verfechter geben aber noch nicht auf.

Fragen, die von der praktischen Seite der computerlinguistischen Anwendungen sich ergeben, sind z.B.:

> [> Häufigkeiten bestimmter syntaktischer Strukturen
>
> [> Einfluß der verschiedenen Textsorten / Fachsprachen auf diese Häufigkeiten
>
> [> praktische Relevanz und Häufigkeit bestimmter grammatischer Phänomene (syntaktische Strukturen, in Abhängigkeit vom lexikalischen Material)

Die Aufgabe ist dann, diese Informationen darzustellen und in die entsprechendene Analyse-Verfahren zu integrieren.

Einige Phänomene, die dabei beachtet werden müssen, beschreibt das vorliegende Buch. Insofern ist es eine interessante und empfehlenswerte Lektüre, auch wenn man sich manches Kopfzerbrechen, das die Autoren anregen wollen, nun wirklich nicht

machen muß.

Gregor Thurmair, Sietec München

Marc Domenig, Pius ten Hacken:
Word Manager: A System for Morphological Dictionaries. HildesheimjZürichjNew York: Olms. 211 Seiten, 39,80 DM, 1992.

Marc Domenig ist Professor für Informatik an der Universität Basel und hat sich schon in seiner Dissertation und seiner Habilitation mit computerlinguistischen Themen beschäftigt, die zu dem hier beschriebenen System hinführen. Pius ten Hacken war, bevor er nach Basel wechselte, als Computerlinguist an der Universität Utrecht u. a. in der niederländischen Eurotra-Gruppe tätig. Das System Word Manager ist in der Grundanlage Domenigs Werk; ten Hacken dürfte vor allem gründliche Kenntnisse einer weiteren Sprache (Niederländisch) und einer überaus anspruchsvollen Anwendung, der automatischen Übersetzung (Eurotra), beigetragen haben.

Word Manager ist ein aufwendiges Softwaresystem zum Aufbau und zur Pflege wortgrammatischer Daten und Regeln. Der Ausdruck morphology umfaßt hier die ganze Wortgrammatik, d.h. Morphologie und Wortbildung (vgl. Schubert 1993). Die innere Grammatik des Wortes wird in der Computerlinguistik nur zu oft auf die leichte Schulter genommen, was sicherlich nicht zuletzt daran liegt, daß viele Wissenschaftler sich an Modellen orientieren, die für das Englische entwickelt worden sind, das bekanntlich auf der Wortebene weniger formenreich ist . Versucht man wie Domenig und ten Hacken, das Problem der Wortgrammatik umfassend anzugehen, wird schnell klar, daß weder schnelle noch linguistisch anspruchslose Lösungen der Aufgabe gewachsen sind, komplexe Wörter in Stämme, Flexions- und Ableitungsmorpheme zu zerlegen und umgekehrt zu einem bestimmten Stamm alle flektierten Formen und Ableitungen zu bilden, insbesondere dann nicht, wenn mehrdeutige Analysen und Übergenerierung ausgeschlossen werden sollen. Angesichts der Komplexität des Problems versuchen Domenig und ten Hacken daher eine großangelegte Lösung, die Domenig (1990) schon früher als "computationally expensive" bezeichnet hat. Umfang und Kosten führen die Autoren zu dem Konzept eines zentralen, viele Anwendungen bedienenden Servers.

Die grundlegende Architektur ist modular. Die leere, sprachunabhängige Datenbank ist von den Autoren entwickelt worden und gilt für die darauf aufbauenden Ebenen als vorgegeben. Sie bietet Arbeitsoberflächen für Grammatiker, die für eine bestimmte Sprache ein wortgrammatisches Regelsystem, und für Lexikografen, die das dazugehörige wortgrammatische Wörterbuch aufbauen. Hinzu kommen Schnittstellen für die Clientanwendungen, die die Datenbank befragen. Word Manager ist damit eine Arbeitsumgebung für beliebige Sprachen und nicht das fertige Morphologiepaket für Sprache X, das gelegentlich Anbieter sprachtechnologischer Anwendungen schnell irgendwo zu kaufen versuchen. Domenig und ten Hacken nennen den Kern ihres Systems daher auch Conceptual Morphological Knowledge Specification environment.

Die Autoren haben ihr System gründlich getestet. Sie referieren Teile ihrer Implementierung für Deutsch, Italienisch, Niederländisch und Englisch. Dies sind Sprachen, deren Wortgrammatik sich nicht auf "einfache" Agglutination beschränkt, sondern Vor- und Nachsilben, Ablaut, Fugenlaute und Trennbuchstaben sowie viele andere Morphemveränderungen kennt. Daß Domenig und ten Hacken ihren Word Manager an diesen wortgrammatisch

komplexen Sprachen mit einem Maß an Vollständigkeit ausprobiert haben, das auch anspruchsvollen Anwendungen gerecht wird, läßt hoffen, daß sich das System als praxisrobust erweisen wird.

Die Arbeit bewegt sich zwischen einer begrifflichen Definition des gewählten Modells und der in ihm abgebildeten sprachlichen Regelmäßigkeiten einerseits und einer detaillierten technischen Beschreibung der Bedienoberflächen, Zeichensätze und Regelnotationen andererseits hin und her. Sie verzichtet jedoch nicht darauf, auch konkurrierende Ansätze zu besprechen. Insbesondere betrachten die Autoren das System der niederländischen Lexikologiedatenbank CELEX, Koskenniemis Zweiebenenmorphologie und DATR nach Evans und Gazdar.

Eine interessante Arbeit, die mit erfreulicher praktischer Untermauerung einen Be
reich darstellt, der gern vergessen wird, auf dem aber der ganze Rest des computergrammatischen Gebäudes ruht.

Literatur

Domenig, Mare (1990): *Lexeme-based morphology: a computationally expensive approach intended for a server architecture.* In: Hans Karlgren (Hg.): Papers presented to the 13th International Conference on Computational Linguistics. Helsinki: Universitas. Bd. 2: 77-82

Schubert, Klaus (1993): *Semantic compositionality: Esperanto word formation for language technology.* Linguistics 31: 311-365

Klaus Schubert} FH Flensburg

Hooshang Mehrjerdian:
Automatische Übersetzung englischer Fachtexte ins Persische. Sprache und Information Bd. 25, Niemeyer, Tübingen 1993. 171 Seiten.

Bei diesem Band der Reihe Sprache und Information handelt es sich um einen interessanten Versuch, die im EUROTRA-Projekt erarbeiteten Ergebnisse auf eine nicht-westeuropäische Sprache anzuwenden. EUROTRA war als Forschungs- und Entwicklungsprogramm der Kommission der Europäischen Gemeinschaft konzipiert, das einen Prototyp eines MÜ-Systems für alle 9 Amtssprachen zum Ziel hatte: Dieses ehrgeizige Ziel wurde innerhalb der Projektlaufzeit von 1985 bis 1992 nur teilweise erreicht; die Bewerter stimmen jedoch darin überein, daß neben dem Aufbau einer konkreten Zusammenarbeit und der Schaffung einer Infrastruktur in weniger entwickelten Ländern auch eine Reihe von interessanten linguistischen Ergebnissen erzielt worden sind. Auf diesen Ergebnissen setzen in der Folge einige nationale Forschungsprojekte sowie neue industrielle Entwicklungen wie EUROLANG auf.

Auch in der akademischen Welt, wo teilweise kontroverse Diskussionen um die von EUROTRA verfolgten Konzepte geführt und oft alternative Konzepte entworfen wurden, hat dieses große Projekt seine Ausstrahlung gezeigt; an einzelnen Stellen sind Anstrengungen unternommen worden, die in langen Diskussionen für die 9 westeuropäischen Sprachen entwickelten linguistischen Modelle an weiteren Sprachen zu erproben.

Hierzu gehört auch die vorliegende Arbeit von H. Mehrjerdian, die als Dissertation von W. Lenders (Bonn) betreut worden ist.
Sie gibt (nach einem kurzen Überblick über den Aufbau der Arbeit) zunächst einen Einblick in die speziellen Belange der persischen Sprache (Farsi), die für Phonologie, Morphologie und feste Wortklassen relativ detailliert und auch für den Nicht-Spezialisten (z.B. den Rezensenten) verständlich ist; die Angaben über die Syntax sind dagegen eher etwas knapp ausgefallen. Hier müßte ein Farsi-Kenner detailliertere Urteile anbringen können.

Im weiteren wird ein Überblick über die

Prinzipien des EUROTRA-Systems sowie die Implementierung der englischen Analyse gegeben, die der Autor als Basis für den von ihm zu erstellenden Transfer Englisch-Persisch sowie die persische Generierung verwendet.
Die dabei entstehende Kurzdokumentation der in der englischen Analyse verwendeten Attribut-Wert-Paare ist von großem Nutzen für alle, die sich mit einer automatischen Syntaxanalyse des Englischen beschäftigen müssen; EUROTRA selbst hat sich die Mühe einer solchen (leicht verständlichen) Darstellung nur selten gemacht! Erst in den jetzt erscheinenden Heften der Reihe "Studies in Machine Translation and Natural Language Processing" (zu beziehen über das Veröffentlichungsbüro der EG), die dem Verfasser noch nicht zur Verfügung standen, ist eine solche Richtung zu erkennen.

Weniger von allgemeinem Interesse ist dagegen die (sehr mit technischen Einzelheiten belastete) Beschreibung von Transfer, für die Beschreibung der Synthese sind wiederum detaillierte Farsi-Kenntnisse notwendig. Hier wäre fast zu überlegen gewesen, die ganze Dissertation in englischer Sprache abzufassen - sie hätte sicherlich mehr Publikum gefunden.

Das letzte Kapitel" Wörterbücher" bringt in etwas ungleich gewichteter Verteilung nunmehr die detaillierten Informationen zum Persischen (die für das Englische im Analysekapitel auftauchten); Transfer- und Fachwörterbücher werden nur noch kurz gestreift, was etwas im Widerspruch zur expliziten Erwähnung der Fachsprache im Titel steht. Ähnliches gilt für das Schlußkapitel "Zusammenfassung und Ausblick", das Verbesserungsideen etwas zusammenhanglos präsentiert: neben technischen Problemen der vom Autor verwendeten EUROTRA-Version steht die inhärente Problematik der Satz-für-Satz-Übersetzung sowie die Notwendigkeit eines vorgeschalteten Grammar-Checkers.
Man hat den Eindruck, als sei der Autor froh, daß er sich endlich durch den großen Brocken EUROTRA durchgearbeitet habe und nun ein bißchen erschöpft sei.

Darin besteht eigentlich das größte Verdienst dieser Arbeit: in einem verhältnismäßig frühen Stadium mit einem äußerst komplexen und noch nicht ausgereiften System in einer exotischen Sprache experimentiert zu haben; zumindest liegen damit erstmals Elemente' einer (moderneren) formalen Beschreibung des Persischen vor.

Johann Haller, Univ. Saarbrücken