Informatik-Fachberichte 313

Herausgeber: W. Brauer
im Auftrag der Gesellschaft für Informatik (GI)

Subreihe Künstliche Intelligenz

Mitherausgeber: C. Freksa
in Zusammenarbeit mit dem Fachbereich 1
„Künstliche Intelligenz" der GI

Stephan Busemann

Generierung natürlicher Sprache mit Generalisierten Phrasenstruktur-Grammatiken

Springer-Verlag
Berlin Heidelberg New York London Paris
Tokyo Hong Kong Barcelona Budapest

Autor
Stephan Busemann
DFKI GmbH
Stuhlsatzenhausweg 3
W-6600 Saarbrücken 11

CR Subject Classification (1992): I.2.7

ISBN-13: 978-3-540-55818-7 e-ISBN-13: 978-3-642-77714-1
DOI: 10.1007/978-3-642-77714-1

Satz: Reproduktionsfertige Vorlage vom Autor / Herausgeber
33/3140-543210 – Gedruckt auf säurefreiem Papier

Vorwort

Linguistische Grammatiktheorien bilden langfristig eine unverzichtbare Grundlage für die Verarbeitung natürlicher Sprache, will man flexible und vielfältig einsetzbare natürlichsprachliche Systeme entwickeln. Aus informatischer Sicht erscheinen sehr viele linguistische Formalismen zunächst nicht verwendbar, weil sie sich nicht ausreichend effizient implementieren lassen.

Am Beispiel der linguistischen Syntaxheorie der Generalisierten Phrasenstruktur-Grammatik (GPSG) und des Problems der Generierung geschriebener natürlicher Sprache zeigt dieses Buch, wie die Brücke zwischen informatischen und linguistischen Forderungen geschlagen werden kann, so daß eine nutzbringende Symbiose entsteht.

Das Thema des Buches ist interdisziplinär: Fragen der linguistischen Theoriebildung werden in Bezug gesetzt zu Problemen der effizienten Sprachverarbeitung. Das Buch führt in die Theorie der GPSG ausführlich ein, diskutiert sie sowohl aus linguistischer wie aus informatischer Sicht und münzt die Kritik um in einen effizient implementierten Formalismus. Auf dieser linguistischen Grundlage werden Algorithmen zur Sprachgenerierung vorgestellt und bewertet. Die Abbildung von semantischen Ausgangsstrukturen auf syntaktische GPSG-Strukturen erfolgt mit Hilfe von getrennt repräsentierten Produktionsregeln, wie sie aus den Produktionensystemen der Künstlichen Intelligenz bekannt sind.

Das Buch wendet sich an Informatikerinnen[1], die an Problemen der Verarbeitung natürlicher Sprache interessiert sind, sowie an Computerlinguistinnen und Linguistinnen mit einem guten Hintergrund in Informatik. Für manche Leserinnen wird das Glossar wichtiger linguistischer Fachbegriffe nützlich sein (die darin enthaltenen Begriffe sind im Text bei ihrem ersten Vorkommen durch einen Pfeil (→) gekennzeichnet).–

Dieses Buch ist eine überarbeitete Fassung meiner Dissertation, die im Juli 1990 von der Universität des Saarlandes angenommen wurde. Die Arbeit entstand zum überwiegenden Teil innerhalb der Projektgruppe KIT (**K**ünstliche **I**ntelligenz und **T**extverstehen) am Fachbereich Informatik der TU Berlin zwischen 1985 und 1990.[2] Sie wurde am Deutschen Forschungszentrum für Künstliche Intelligenz in Saarbrücken fertiggestellt.

Meine Beschäftigung mit GPSG erfolgte im Rahmen der Projekte KIT-NASEV und KIT-FAST. Das Projekt KIT-NASEV entwickelte neue **A**nalyse- und Synthese**v**erfahren für die maschinelle Übersetzung, wobei eine zentrale Fragestellung war, wie GPSG überhaupt für Parsing und Generierung nutzbar gemacht werden kann. Im daran anschließenden Projekt KIT-FAST wurden **F**unktor-**A**rgument-**S**trukturen für den **T**ransfer hinzugefügt. Damit konnte maschinelle Übersetzung auf der Grundlage eines Transferansatzes mit Hilfe einer oberflächennahen, satzsemantischen Repräsentationssprache realisiert werden.

In diesem Buch greife ich wesentliche Ergebnisse der Forschung in KIT-NASEV und KIT-FAST auf. Viele Kolleginnen haben zum Gesamtresultat beigetragen und hier beschriebene Teile mitgestaltet. Im Rahmen des Projektes KIT-NASEV wurde eine Version des GPSG-Formalismus entworfen

[1] Gemeint sind selbstverständlich die „Informatikerinnen und Informatiker". Um solche klobigen Formulierungen (oder orthographische Purzelbäume wie in „InformatikerInnen") zu vermeiden, verwende ich durchgehend die feminine Form, wenn an beiderlei Geschlecht gedacht ist.

[2] KIT umfaßt Projekte aus verschiedenen Teildisziplinen der KI und der Computerlinguistik, was einen regelmäßigen, fruchtbaren Gedankenaustausch zwischen den Kolleginnen ermöglicht.

und implementiert, die für Sprachverarbeitung besonders gut geeignet ist. Daran haben maßgeblich Christa Hauenschild, James Kilbury, Wilhelm Weisweber, William Keller und ich mitgewirkt. Auf der Grundlage dieses Formalismus habe ich zwei Verfahren entworfen und implementiert, die zeigen, wie mit GPSG geschriebene natürliche Sprache generiert werden kann. Die benutzten Grammatik-Fragmente des Deutschen und Englischen wurden hauptsächlich von Susanne Preuß und Margaret Garry im Rahmen der genannten Projekte geschrieben und von Susanne Preuß und mir getestet. Die verwendeten Funktor-Argument-Struktur-Fragmente entwickelten federführend Christa Hauenschild und Carla Umbach.–

Ich danke meinem Doktorvater Wolfgang Wahlster für seine Unterstützung über einen langen Zeitraum hinweg und seine Aufmunterung, die besonders in den schwierigen Phasen der Arbeit wichtig war. Hans Uszkoreit, dem zweiten Gutachter meiner Dissertation, danke ich für zahlreiche Hinweise, Ideen und spannende Diskussionen im Bereich der Unifikationsgrammatiken, die mir viele neue Perspektiven erschlossen.

Von prägendem Einfluß war für mich die enge Zusammenarbeit mit Christa Hauenschild in KIT-FAST. Christa verdanke ich wichtige Zugänge zur formalen Linguistik; mit ihr über GPSG, Generierung und die wichtigen Dinge des Lebens zu diskutieren, war nicht nur außerordentlich anregend, sondern es machte auch ganz einfach Freude. Susanne Preuß, Carla Umbach und Wilhelm Weisweber bin ich für viele Anregungen und kritische Fragen dankbar, die mich immer wieder zum Überdenken meiner Grundannahmen brachten.

Allen Mitgliedern der Projektgruppe KIT an der TU Berlin und dem Leiter der KIT-Projekte Bernd Mahr danke ich für ihre freundschaftliche Unterstützung und Geduld, ohne die die Arbeit in dieser Form wohl nicht zustandegekommen wäre. Besonders wichtig waren Kai von Lucks unermüdliches Drängen und seine Antriebskraft, die bis zu seinem Weggang aus Berlin diese Arbeit mit in Gang brachten. Bei der Implementation des Berliner GPSG-Systems bildete Michael Eimermachers Tracepaket für Waterloo-Core-Prolog eine wichtige Hilfestellung. Christof Peltason erleichterte mit zahlreichen Service-Programmen die Benutzung eines KIT-weiten Netzes verschiedenster Rechner, wovon ich sehr profitiert habe. Cordula Lippke leistete wertvolle technische Hilfe bei der Erstellung von Graphiken.

Dem Deutschen Forschungszentrum für Künstliche Intelligenz in Saarbrücken und meinen Kolleginnen an meinem dortigen Arbeitsplatz bin ich dankbar für die Möglichkeit, die Arbeit abzuschließen.

Mein besonderer Dank gilt Katharina Morik für ihre freundschaftlichen Anstöße und Ideen sowohl in fachlicher als auch in ganz menschlicher Hinsicht. Schließlich danke ich meiner Frau Monika, die so oft auf gemeinsam verbrachte Zeit verzichtete und mir doch immer wieder Kraft zum Weitermachen gab.

S.B.

Saarbrücken,
Juni 1992

Inhaltsverzeichnis

Abbildungsverzeichnis