American Journal of Computational Linguistics *Microfiche 77*

THE FINITE STRING

NEWSLETTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

VOLUME 15 - NUMBER DECEMBER 1978

George E Heidorn has accepted appointment as Editor of AJCL His term begins with the first issue for 1979

The experiment with a microfiche journal is over. The next issue of the journal will be on paper, with possible microfiche supplement.

The retiring Editor apologizes to the membership for the long delay in release of the present material for publication. Grave personal difficulties interfered with all of Hays's routine activities during the period, and voluntary activities necessarily suffered most

Released for publication March 25, 1979.

AMERICAN JOURNAL OF COMPUTATIONAL LINGUISTICS is published by the Association for Computational Linguistics

EDITOR David G Hays, 5048 Lake Shore Road, Hamburg, New York 14075

EDITORIAL ASSISTANT William Benzon

MANAGING EDITOR Donald E Walker, Artificial Intelligence Center, SRI International, Menlo Park, California 94025

TECHNICAL ADVISOR Martin Kay, Xerox Palo Alto Research Center

American Journal of Computational Linguistics *Microfiche 77:2*

CONTENTS

ACL: MINUTES OF THE 16TH ANNUAL BUSINESS MEETING 3

SECRETARY-TREASURER'S REPORT 7

OFFICERS FOR 1979 9

OFFICERS 1963-1979 10

NSF: SUPPORT FOR COMPUTATIONAL LINGUISTICS 12

NEWS: SHORT NOTES 14

ARIST REPRINT REQUEST 16

SUMMER LINGUISTICS AT TEXAS 18

PH D PROGRAMS IN COMPUTATIONAL LINGUISTICS 19

JOURNAL: COMPUTATIONAL LINGUISTICS AND COMPUTER LANGUAGES . 21

DISCOURSE PROCESSES 22

BOOK NOTICES 23

YALE AI PROJECT RESEARCH REPORTS AVAILABLE 26

SUMMARY OF RESEARCH ON COMPUTATIONAL ASPECTS OF EVOLVING THEORIES Raymond D. Gumb 27

TAXONOMY: INFORMATION SCIENCES Editors of Information Systems 31

MACHINE AIDS TO TRANSLATION A CONCISE STATE OF THE ART BIBLIOGRAPHY Wayne Zachary 34

REVIEWS: ON HUMAN COMMUNICATION, 3D ED, BY COLIN CHERRY William L. Benzon 41

ABHÅNGIGKEITSGRAMMATIK, BY JURGEN KUNZE Kenneth F. Ballin 47

GLANCING, REFERRING AND EXPLAINING IN THE DIALOGUE SYSTEM HAM-RPM W. Wahlster, A. Jameson, and W. Hoeppner 53

A CRITICAL LOOK AT A FORMAL MODEL FOR STRATIFICATIONAL LINGUISTICS Alexander T. Borgida 68

ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

MINUTES: 16th Annual Business Meeting
26 July 1978
University of Illinois, Urbana, Illinois
Jonathan Allen, President, presiding

MINUTES OF THE PREVIOUS MEETING

Allen noted that the minutes of the previous meeting had been published in the Finite String, Volume 14, Number 3, Microfiche 65 of the American Journal of Computational Linguistics.

FINANCIAL STATUS

Don Walker, Secretary-Treasurer, reviewed the financial status of the Association, a copy of which is attached to these Minutes. He presented income and expenses both for 1977 and for 1978 through 21 July. The balances of $6,332.19 for 1977 and $7,060.63 for 1978 constitute assets of $13,392.82. However, the major costs for AJCL for the current year are yet to be incurred, and dues for AFIPS have not yet been billed. At the same time, the $1,000 advance to cover costs of the TINLAP-2 Meeting is likely be returned, along with a portion of the excess of income over expenses, which will be shared with ACM/SIGART.

The indebtedness of the Association to the Center for Applied Linguistics, which housed the previous Secretariat, has been paid off. Of the total of $13,486.06, funds from the previous Secretariat provided $9,913.92; $3,572.14 was taken out of income for 1977.

Walker remarked that the income item from AFIPS for 1977 of $2,365, which reflected a disbursement of surplus funds from the National Computer Conference, was unlikely to be repeated again soon. AFIPS is in the process of creating a new journal, ABACUS, modeled after Scientific American, and future surpluses probably will be used to defray or at least to backup the startup costs. The journal is expected to be self-sustaining, and might eventually show a profit.

The current balance in the TINLAP-1 Account is $109.41; approximately 75 copies remain.

MEMBERSHIP

Walker reported membership figures for 1977 of 500 individual and 201 institutional, for a total of 701. The current figures for 1978, through 21 July, are 405 individual and 208 institutional, for a total of 613. A slightly more detailed breakdown is attached to these Minutes.

EDITOR'S REPORT

Dave Hays, Editor of the AJCL, announced that he was resigning, effective at the end of the year. The next issue will include the TINLAP Proceedings; the final issue of the current year will contain a complete index for the five years of its publication. The Journal was established in 1974 as an NSF-sponsored experiment in microfiche publication. Anticipating the implications of George Heidorn's survey (see below), Hays remarked that this mode of publication is likely to be replaced by microprocessor technology and might never receive a full scale trial.

Allen expressed the gratitude of the Association to Hays for his devotion, his constructiveness, and his tireless efforts in establishing and sustaining the AJCL. This tribute was affirmed by the members. Allen then announced that Heidorn, currently Associate Editor, would replace Hays as Editor in January.

A NEW FORMAT FOR THE AJCL

Heidorn presented the results of his survey of the membership regarding a new format for the AJCL. Of 513 questionnaires mailed (to both current members and to members who paid for 1977 but not yet for 1978); 212 were returned. The responses favored creation of a hard copy edition, similar in format to the Communications of the ACM, with or without an accompanying microfiche version. Most members felt that such a change would encourage a wider readership and an increase in the submission of techical articles. A more comprehensive report by Heidorn is included elsewhere in this issue.

Allen reported that the Executive Committee, after reviewing Heidorn's findings, has decided to issue the Journal in the new format with both hard copy and microfiche versions sent to each member, beginning with the first issue of 1979. The microfiche also may contain appendixes for technical articles, program listings, and other material of interest to the membership but less appropriate for inclusion in hard copy form, like the list of members.

NEXT MEETING

Allen stated that a decision about the time and place of the 1979 meeting had not yet been made. Asilomar, near Monterey in California, is being considered, possibly around the time of the next International Joint Conference on Artificial Intelligence, which will be held in Tokyo from 20 to 24 August. In the discussion, members expressed concern that the choice should not discourage attendance by graduate students.

For 1980, the Executive Committee recommended that a third TINLAP meeting be held. The President, Vice President, and Secretary-Treasurer constitute an interim committee to investigate this possibility and to negotiate with SIGART, as appropriate. Allen also remarked that an offer already had been received from Aravind Joshi to host TINLAP-3 at the University of Pennsylvania.

NOMINATIONS FOR OFFICERS FOR 1979

Aravind Joshi, reporting for the Nominating Committee (Joshi, Petrick, and Chapin), announced the following nominations for officers for 1979:

Nominating Committee: Jonathan Allen, MIT
Executive Committee: Stanley Rosenschein, RAND
Secretary-Treasurer: Donald Walker, SRI International
Vice President: Bonnie Lynn Webber, University of Pennsylvania
President: Ronald Kaplan, Xerox Palo Alto Research Center

No additional nominations were received from the floor. Allen called for a vote, which was unanimous, and the slate was declared elected.

NEW BUSINESS

Carol Lane presented a resolution supporting the ratification of the Equal Rights Amendment to the U.S. Constitution. After extensive discussion and after motions to amend and to table were defeated, the members affirmed the following substitute resolution by a vote of 20 to 13:

> WHEREAS, inclusion in the Constitution of these United States is the basic unalienable right of every citizen;
>
> WHEREAS, the Association for Computational Linguistics views as intolerable the selective exclusion of over one-half the population of this country;
>
> WHEREAS, the Equal Rights Amendment, writing women into the Constitution, must be ratified by three-fourths of the states (38) prior to its incorporation
>
> THEREFORE, BE IT RESOLVED, that all future conventions, meetings, and conferences of the Association for Computational Linguistics will, for the duration of time during which the Equal Rights Amendment is under consideration by the several states of the United States, be held only in those states that have ratified the Equal Rights Amendment.

RESOLUTIONS

Having set out in advance the syntax of his report for the Resolutions Committee, Ron Kaplan expressed the gratitude of the Association to Dave Waltz, his session organizers, and the University of Illinois for the organization and conduct of the meeting; to the National Science Foundation, and particularly to Carol Ganz Brown, for its financial support; to the current officers for their constructive efforts during the first part of their elective term (and with encouragement to continue these efforts for the rest of the year); to the retiring members of the AJCL Editorial Board--Robert Barnes, Fred Damerau, Gary Martins, John Olney, and Naomi Sager--for five years of effective service; to Dave Hays for his countless hours and fruitful endeavors in

realization of the AJCL; and to George Heidorn for his willingness to become the new editor.

Dave Hays called attention to the efforts of Martin and Iris Kay in the preparation of the AJCL Bibliography, and they were duly included in the list of resolutions.

The members affirmed these sentiments enthusiastically, and Allen directed the Secretary-Treasurer to express the appreciation of the Association formally to Dave Waltz.

The meeting adjourned.

Donald E. Walker
Secretary-Treasurer

Attachments: Financial Status, Membership Status, Officers for 1979

ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

Secretary-Treasurer's Report
(as of 21 July 1978)

FINANCIAL STATUS	1977	1978
Income:		
Dues	$13285.74	$11337.73
Back Issues	1790.00	825.00
Meetings	1395.00	
AFIPS	2365.00	
Mailing Labels	62.40	107.75
Interest	352.13	246.58
TINLAP-1 Account	435.00	52.50
	$19685.27	$12569.16
Expenses:		
AJCL Production	$ 4125.20	$ 1110.76
AJCL Bibliography	1412.50	1084.67
AJCL Editorial	208.61	385.33
Meeting Expenses	390.65	1000.00
AFIPS Dues	500.00	
Secretariat Services	1538.26	1046.00
Postage	839.60	675.62
Supplies	370.02	55.27
Printing	83.60	143.28
TINLAP-1 Account	312.50	7.50
Center for Applied Linguistics	3572.14	
	$13353.08	$ 5508.53
Balance:	$ 6332.19	$ 7060.63
Assets:		
Savings		$11207.54
Checking		2153.25
Petty Cash		32.03
		$13392.82

Center for Applied Linguistics Account

Debt (as of 3-14-77)	$13486.06
Paid (out of 1976 funds)	9913.92
(out of 1977 funds)	3572.14
	0.00

TINLAP-1 Account

Current Balance	$109.41

MEMBERSHIP STATUS	1977	1978
Individual	500	405
US	364	300
Foreign	136	105
Institutional	201	208
US	92	109
Foreign	109	110
Special	19	19
TOTALS	701	613

ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

Officers for 1979

Dr. Ronald M. Kaplan Xerox Palo Alto Research Center 3333 Coyote Hill Road Palo Alto, CA 94304	President 415:494-4416
Professor Bonnie Lynn Webber Computer and Information Science The Moore School University of Pennsylvania Philadelphia, PA. 19104	Vice President 215:243-8540
Dr. Donald E. Walker Artificial Intelligence Center SRI International Menlo Park, CA 94025	Secretary-Treasurer 415:326-6200x3071
Dr. Jerry R. Hobbs Artificial Intelligence Center SRI International Menlo Park, CA 94025	Executive Committee (1977 - 1979) 415:326-6200x2229
Dr. Bertram C. Bruce Bolt Beranek and Newman 10 Moulton Street Cambridge, MA 02138	Executive Committee (1978 - 1980) 617:491-1850
Dr. Stanley J. Rosenschein Rand Corporation 1700 Main Street Santa Monica, CA 90406	Executive Committee (1979 - 1981) 213:393-0411
Dr. Stanley R. Petrick IBM Watson Research Center P.O. Box 218 Yorktown Heights, NY 10598	Nominating Committee (1977 - 1979) 914:945-2175
Dr. Paul G. Chapin Linguistics Program National Science Foundation Washington, DC 20550	Nominating Committee (1978 - 1980) 202:254-6326
Professor Jonathan Allen Electrical Engineering MIT, 36-575 Cambridge, MA 02139	Nominating Committee (1979 - 1981) 617:253-2509
Dr. George E. Heidorn IBM Watson Research Center P.O. Box 218 Yorktown Heights, NY 10598	Editor, AJCL 914:945-2776

OFFICERS
ASSOCIATION FOR MACHINE TRANSLATION AND COMPUTATIONAL LINGUISTICS (1963-1968)
ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (1968-1979)

	1963	1964	1965	1966
President	Yngve	Hays	Lehmann	Garvin
Vice-Pres	Hays	Alt	Garvin	Oettinger
Sec-Treas	Josselson	Josselson	Josselson	Josselson
Executive	Rhodes	Sebeok	Sebeok	Sebeok
Committee	Garvin	Garvin	Hockett	Hockett
---------	Lehmann	Lehmann	Kuno	Prendergraft
Editor (FS)	Roberts	Roberts	Roberts	Roberts
Nominating	See	Yngve	Yngve	Yngve
Committee	Oettinger	Oettinger	Hays	Hays
---------	Lamb	Lamb	Lamb	Lieberman

	1967	1968	1969	1970
President	Kuno	Walker	Kay	Plath
Vice-Pres	Walker	Mersel	Plath	Friedman
Sec-Treas	Josselson	Josselson	Josselson	Josselson
Executive	Satterthwait	Satterthwait	Satterthwait	Wall
Committee	Hockett	Fromkin	Fromkin	Fromkin
---------	Pendergraft	Pendergraft	Montgomery	Montgomery
Editor (FS)	Roberts	Roberts	Roberts	Roberts
Editor (MTCL)	Yngve	Yngve	Yngve	Yngve
Nominating	Garvin	Garvin	Garvin	Kay
Committee	Hays	Kuno	Kuno	Kuno
---------	Lieberman	Lieberman	Walker	Walker

	1971	1972	1973	1974
President	Friedman	Simmons	Barnes	Woods
Vice-Pres	Simmons	Fromkin	Woods	Wall
Sec-Treas	Josselson	Roberts	Roberts	Roberts
Executive	Wall	Wall	Martins	Martins
Committee	Robinson	Robinson	Robinson	Joshi
---------	Montgomery	Chapin	Chapin	Chapin
Editor (FS)	Roberts	Roberts	Roberts	
Editor (AJCL)				Hays
Nominating	Kay	Kay	Simmons	Simmons
Committee	Plath	Plath	Plath	Barnes
---------	Walker	Friedman	Friedman	Friedman

	1975	1976	1977	1978
President	Joshi	Petrick	Chapin	Allen
Vice-Pres	Petrick	Grimes	Allen	Kaplan
Sec-Treas	Roberts	Roberts	Walker	Walker
Executive	Martins	Diller	Diller	Diller
Committee	Rieger	Rieger	Hobbs	Hobbs
---------	Nash-Webber	Nash-Webber	Nash-Webber	Bruce
Editor (AJCL)	Hays	Hays	Hays	Hays
Assoc Editor			Heidorn	Heidorn
Nominating	Simmons	Joshi	Joshi	Joshi
Committee	Barnes	Barnes	Petrick	Petrick
---------	Woods	Woods	Woods	Chapin

	1979	1980	1981	1982
President	Kaplan			
Vice-Pres	Webber			
Sec-Treas	Walker			
Executive	Rosenschein	Rosenschein	Rosenschein	
Committee	Hobbs			
---	Bruce	Bruce		
Editor (AJCL)	Heidorn			
Assoc Editor				
Nominating	Allen	Allen	Allen	
Committee	Petrick			
---	Chapin	Chapin		

FS = The Finite String
MTCL = Machine Translation and Computational Linguistics
AJCL = American Journal of Computational Linguistics

N S F SUPPORT FOR COMPUTATIONAL LINGUISTICS

Paul G. Chapin, Director of the Linguistics Program of the National Science Foundation, announced the following grants for research of obvious relevance to computational linguistics.

LANGUAGE UNIVERSALS ARCHIVING PROJECT
Charles A. Ferguson and Joseph Greenberg
Stanford University - $49,200 - 13 months

COMPUTER STUDIES IN FORMAL LINGUISTICS
Joyce Friedman
University of Michigan - $40,000 - 24 months
(The Intelligent Systems Program of NSF awarded the same amount)

COMPUTATIONAL COMPLEXITY OF GRAMMAR & NL RECOGNITION PROBLEMS
William Rounds
University of Michigan - $44,600 - 24 months

American Journal of Computational Linguistics

N S F Support for Computational Linguistics

During the Fiscal Year 1978, the Division of Science Information of NSF issued at least the following grants for support of research relevant to computational linguistics.

Correlation of Language Structure with Information
Zellig S. Harris
University of Pennsylvania - $163,329 - 29 months

Integrated Man/Machine Interface for Network Resource Utilization
Martha E. Williams
University of Illinois - $174,432 - 24 months

High-level Language Subcommittee

The Microprocessor Standards Committee of the IEEE formed a subcommittee to work on standsrds for the five high-level languages widely used in microprocessor applications: Basic, Fortran, Cobol, PL/M, and Pascal. The object foreseen was the identification and endorsement of appropriate standards. The first meeting was called in October by Bruce Ravenel of Language Resources, 1311 Lombard Street, San Francisco, CA 94109.

Manpower Survey in Linguistics

What fields need linguists or persons with some linguistic knowledge? What are the "demands and perspectives" of present and possible employing institutions? Anyone with information relevant to these issues can correspond with Dr. Walther Kindt, Fakultät für Linguistik und Literaturwissenschaft, Universitat Bielefeld, Postfach 8640, 4800 Bielefeld 1, Germany.

Computational Proofreading

The detection of orthographic errors in keyboarding of Swedish text is the topic of work undertaken by Rolf. Gavare, Department of Computational Linguistics, Göteborgs Universitet, Norra Allegatan 6, S-413 01 Göteborg, Sweden, who invites correspondence.

MT: English and Thai via Montague Grammar

The work of Joyce Friedman is being applied in work on translation. Kurt Godden, 1408 E. 27, Lawrence, KS 66044, invites contact.

Association for Women in Computing.

AWC was founded in 1978 to promote communication, professional development and advancement, and education.. Membership is open without restriction. The Correspondent of AWC is Anita Cochran, 5A137 Bell Laboratories, Murray Hill, NJ 07974; 201-582-7817.

Spelling Correction

A program to check spelling in English text has been written by Ralphe E. Gorin, AI Laboratory, Computer Science Department, Stanford University, CA 94305, with additions by William Plummer and Jerry Wolf of BBN and Richard Johnsson and Philip Karlton of Carnegie Mellon University.

Online Bibliography in Sign Language and Related Areas

The Syracuse Information Retrieval Experiment system has been adapted by Jim Bourg of the Gallaudet College Library and is being used by William C. Stokoe, Linguistics Research Lab, Gallaudet College, Kendall Green, Washington, DC 20002, for a pscyho- and sociolinguistic bibliography. Anyone with a console may inquire about access; no charge is levied at present.

Request for reprints

ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY

Volume 14 of ARIST is in preparation. Authors of chapters need help in finding relevant recent publications. They will appreciate receiving offprints from authors at their respective addresses.

If the appropriate chapter writer is not apparent, write to

Martha E. Williams
Editor, ARIST
R.R. No. 1
Monticello, Illinois 61856

CHAPTER TOPICS AND WRITERS

COMPUTER ARCHITECTURE FOR NATURAL LANGUAGE AND IR APPLICATIONS

Professor P. Bruce Berra
(& Ellen Oliver)
Syracuse University
Syracuse, New York 13210

RETRIEVAL TECHNIQUES

Professor Michael McGill
School of Information Studies
113 Euclid Avenue
Syracuse University
Syracuse, New York 13210

COST ANALYSIS OF SYSTEMS AND SERVICES

Mr. Colin Nick
Applied Communication Research
P.O. Box 5849
Stanford, California 94305

LIBRARY AUTOMATION

Ms. Mary Jane Probst Reed
Associate Director for
Research and Planning
Washington State Library
Olympia, Washington 98504

Mr. Hugh Vrooman
Illinois State Library
Centennial Building
Springfield, Illinois 62706

INFORMATION SYSTEMS IN LATIN AMERICA

Professor Tefko Saracevic
School of Library Science
Case Western Reserve University
Cleveland, Ohio 44106

(Additional authors for this and following chapters on next frame.)

INFORMATION SYSTEMS IN LATIN AMERICA (continued)

Gilda Braga
Instituto Brasileiro de informacao em Ciencia e Technologia
Av. General Justo 171; 4º
Rio de Janeiro, Brazil

Alvaro Quijano Solis
Biblioteca Daniel Cosio Villegas
El Colegio de Mexico
Camino Al Ajusco No. 20
Apartado Postal 20-671
Mexico 20, D.F.
Mexico

INFORMATION SYSTEMS FOR CONSUMER CONCERNS

Professor Vivian Sessions
School of Library Science
McGill University
3459 McTavish Street
Montreal, Quebec H3A 1Y1
Canada

COMPUTERS AND PUBLISHING

Mr. David Staiger
American Institute for Aeronautics & Astronautics
1290 Avenue of the Americas
New York, New York 10019

EDUCATION AND TRAINING FOR ONLINE SYSTEMS

Ms. Judy Wanger
1523 Sixth Street, Suite 12
Santa Monica, California 90401

DATA BASE MANAGEMENT SYSTEMS

Dr. Ronald Wigington
(& Michael A. Huffenberger)
Chemical Abstracts Service
2540 Olentangy River Road
Columbus, Ohio 43202

SYSTEMS DESIGN--PRINCIPLES AND TECHNIQUES

Dr. Ronald Wyllys
Graduate School of Library Science
University of Texas
Austin, Texas 78712

FUNDAMENTAL PRINCIPLES AND THEORIES OF INFORMATION SCIENCE

Dr. Pranas Zunde
School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332

Summer Linguistics at Texas

The University of Texas at Austin announces a special Summer Linguistics Program which will consist of a substantial offering of graduate courses given by our own faculty members and a distinguished list of visiting scholars. In addition to these courses, workshops and seminars (for credit as conference courses by arrangement with individual faculty members) will be available on topics such as syntactic universals, conditions on rule application, opacity and scope, formal vs. substantive explanation in phonology, etc. We invite applications from graduate students in linguistics and allied disciplines such as education, English, foreign languages, psychology, philosophy, anthropology, and others.

The list of courses and instructors will include the following:

LIN 380K Generative Phonology—Robert T. Harms
LIN 380L Transformational Grammar—Jorge Hankamer and Ivan Sag
LIN 381M Phonetic Theory—Peter MacNeilage
LIN 384 Outline of Turkish Syntax—Jorge Hankamer (tentative)
LIN 393 Semantics—Robert E. Wall
LIN 393 Seminar in Phonetics and Phonology—Björn Lindblom
LIN 393 Seminar in Syntax and Semantics—Emmon Bach and Barbara Partee
LIN 394 New Directions in Historical Linguistics—Robert D. King
LIN 396 Seminar in Linguistic Variation—John Baugh

In addition to the above-listed faculty members, the following scholars will be available for individual consultation: Lee Baker, Peter Cole, David DeCamp, Polly Jacobson, Lauri Karttunen, W. P. Lehmann, Fritz Newmeyer, Susan Schmerling, and others. Several intensive Oriental and European language courses will also be taught as a part of the regular UT summer session.

Low cost accommodations will be available in housing cooperatives. Classes will begin June 13 and exams will end July 20; the Program will thus not conflict with the LSA Linguistic Institute in Salzburg.

Tuition and fees: For Texas residents, the price of summer courses is $64.70 for one three-hour course, $89.90 for two three-hour courses. For out-of-state residents, it is $159.70 for one three-hour course, $304.90 for two three-hour courses. In addition, there is a $10.00 property fee, refundable at the end of the course.

For application materials, please complete the detachable section and mail it to:

Summer Linguistics Program
Department of Linguistics
University of Texas at Austin
Austin, Texas 78712

DEADLINE: May 1, 1979

PhD PROGRAMS IN COMPUTATIONAL LINGUISTICS

During the summer of 1978, Alan K. Melby wrote to many American universities, asking about the graduate work in computational linguistics. He has supplied a copy of his list of affirmative answers, presented here in the casual format of its compilation. Melby points out that MIT and Yale did not respond but would be considered by a student planning work on computers and language.

California

Ken Wexler
School of Social Sciences
University of California
Irvine, CA 92717

(cognitive science, mathema-
(tical linguistics

Electrical Engineering Dept.
University of California
Los Angeles, CA 90024

(some work in AI

Clara Bush
Department of Linguistics
Stanford University
Stanford, CA 94305

(under Prof. Terry Winograd

Connecticut

David Michaels
Room 230 H.R. Monteith Bldg.
University of Connecticut
Storrs, CT 06268

(analysis and synthesis of
(speech--Haskins Laboratory
(connection

Illinois

G. K. Krulee
Department of Linguistics
Northwestern University
Evanston, IL 60201

(with Computer Science Dept.

C. C. Cheng
Linguistics
University of Illinois
Urbana, IL 61801

Kansas

David Dinneen
Department of Linguistics
University of Kansas
Lawrence, KS 66045

Massachusetts

Emmon Bach
Department of Linguistics
University of Massachusetts
Amherst, MA 01002

Michigan

Joyce Friedman
Computer and Communication Sciences
University of Michigan
Ann Arbor, MI 48108

(with qualifications--ask JF

Minnesota

Center for Research in Human
Learning
University of Minnesota
Minneapolis, MN 55455

(psychologists with interest in AI

New York

David Hays
State University of New York
Buffalo, NY 14214

Lewis Levine
Department of Linguistics
Washington Square
New York, NY 10003

(students can work under
(Naomi Sager

Pennsylvania

Simon Belasco
Department of Linguistics
Pennsylvania State University
University Park, PA 16802

(qualified "yes"

Rhode Island

J. J. Wren
Box E
Brown University
Providence, RI

Texas

Department of Linguistics
University of Texas
Austin, TX 78712

Washington, D.C.

Department of Linguistics
Georgetown University
Washington, DC 20007

JOURNAL: C L & C L

COMPUTATIONAL LINGUISTICS AND COMPUTER LANGUAGES

EDITORS: T. FREY, T. VAMOS

PUBLISHER: COMPUTER AND AUTOMATION INSTITUTE, BUDAPEST

CONTENTS OF NUMBER 11:

I. Nemeti:	On a property of the category of partial algebras
Gy. Revesz:	A note on the relation of turing machines to phrase structure grammars
P.B. Schneck:	A new program optimization
B. Dömölke E. Santa-Toth:	Formal description of software components by structured abstract models
G. Fay:	Cellular design principles, a case study of maximum selection in codd-icra cellular space /I/
H. Heiskanen:	Semantic theory from a systematical viewpoint
T. Legendi:	Callprocessors in computer architecture
Gy. Hell:	Mechanical analysis of Hungarian sentences

One double-issue or two issues per year of ca. 350 pp., 20.5 x 28.5 cm.

1977 (numbers 12 + 13) will be published 1978/1979.

Price per issue: Hfl. 42,-- + postage.

Send orders to:

John Benjamins B. V.
Amsteldijk 44 / P. O. Box 52519
1007 HA Amsterdam / The Netherlands
Tel.: (020) 738156 / Telex 15798 jbds
Cables: BENPER / Amsterdam

* * * * *

New Journal

DISCOURSE PROCESSES

A MULTIDISCIPLINARY JOURNAL

EDITOR: ROY O. FREEDLE
Educational Testing Service
Princeton, New Jersey 08540
609-921-9000, ext. 2651

ABLEX PUBLISHING CORPORATION
355 Chestnut Street
Norwood, New Jersey 07648
201-767-8450

Personal subscription. $19.50. Institutions: $45.00.

CONTENTS: VOLUME 1, NUMBER 1, JANUARY-MARCH 1978

The role of culture-specific schemata in the comprehension and recall of stories
Walter Kintsch and Edith Greene

A code in the node: The use of a story schema in retrieval
Jean M. Mandler

An experimental investigation of contingent query schemes
Catherine Garvey and Mohamed BenDebba

Inference and coherence
Edward J. Crothers

How to catch a fish: The memory and representation of common procedures
Arthur C. Graesser

B O O K S: SHORT NOTICE

IGOR A. MEL'CUK. STUDIES IN DEPENDENCY SYNTAX. Ann Arbor: 1979
Karoma Publishers, Inc. 163 + ix pp., paper only $4.50
6 by 9 inches ISBN 0-89720-001-2

Foreward (by Paul T. Roberge, editor) v

Preface xiii

Dependency Syntax 1

The Predicative Construction in Dyirbal 23

Types of Surface-Syntactic Relations 91

Bibliography 151

Abbreviations and Symbols 162

I. MEL'ČUK, R. RAVIČ. AUTOMATIC TRANSLATION, 1964-1970.

Departement de linguistic et philologie
Universite de Montreal
Case postale 6128, Succursale "A"
Montreal, P.Q. H3C 3J7

Ce volume qui est une suite au guide bibliographique analytique de la T(raduction) A(utomatique), publie a Moscou en 1967, recense 1357 ouvrages de la T.A. et les domaines voisins, parus entre 1964 et 1970, aussi bien en U.R.S.S. qu'en Occident. Pour la plupart des ouvrages recenses, on a donné un résumé détaillé (en russe) qui se veut une description suffisante du contenu. Le guide est destiné aux linguistes, aux traducteurs et aux informaticiens, chacun y pouvant trouver des renseignements utiles. Prix approximatif du volume: $16.

This volume, a continuation of the Analytical Bibliographical Guide to A(utomatic) T(ranslation), published in Moscow in 1967, includes 1357 works in and on TA, as well as neighboring domains, which appeared between 1964 and 1970 either in the Soviet Union or in the West. For most of the publications listed, a detailed abstract is provided, which is intended to adequately represent the contents of the corresponding item. The guide is addressed to linguists, translators and information processing specialists all of whom will hopefully find in it useful data. Approximate price will be $16.

J.D. Goldstein, D.W. Lakamp, A. Pietrzyk. INFORMATION SERVICES ON RESEARCH IN PROGRESS: A WORLDWIDE INVENTORY.

National Technical Information Service, US Dept. of Commerce
5285 Port Royal Road
Springfield, VA 22161

462 pp., softcover
PB-282 025/AS
$14.50 hard copy
$3.00 microfiche

Part I: World Trends in Information on Research in Progress· An Overview

Part II: Profiles of Information Systems and Services on Research in Progress

Part III: Indexes to organization and system names, organization and system acronyms, persons respinsible and subject coverage given

Starr Roxanne Hiltz, Murray Turoff. THE NETWORK NATION: HUMAN COMMUNICATION VIA COMPUTER. Addison-Wesley: 1978

Hardbound ISBN 0-201-03140-X c. $26.50
Paperbound ISBN 0-201-03140-8 c. $14.50

The Nature of Computerized Conferencing

Potential Applications and Impacts of Computerized Conferencing

Projecting the Future

F.W. Lancaster. TOWARD PAPERLESS INFORMATION SYSTEMS. Academic 1978

179 pp., $13.50 ISBN 0-12-436050-5

In this nontechnical presentation the author first describes the paperless information/communication systems currently being developed for intelligence agencies. Then existing scientific and technological communication systems are critiqued. Finally the intelligence design is reformulated for the science and technology environment.

L. Bourrelly and E. Chouraqui. THE DOCUMENTATION SYSTEM SATIN 1.
VOL. 1 - GENERAL DESCRIPTION AND USER'S MANUAL (1974) 398 PP.
VOL. 2 - SYSTEM GENERATION AND SET-UP INSTRUCTIONS (1978) 397 PP.

SATIN 1 is implemented at the "Laboratoire d'Information pour les Sciences de l'Homme" in Marseille and is designed specifically for the documentation processes in the humanities and social sciences, archeology, geography, history, etc., and permits the representation of information about artifacts of many kinds (paintings, sculputures etc.) in addition to standard document content.

Herbert E. Bruderer. SPRACHE - TECHNIK - KYBERNETIK.

Aufsätze zur Sprachwissenschaft, maschinellen Sprachverarbeitung, künstlichen Intelligenze und Computerkunst. Verlag Linguistik, Münsingen/Bern 1978, 187 Seiten, 39 Schweizer Franken (Foreign orders must be prepaid: Surface 44 Swiss Francs, Air 50 Swiss Francs), ISBN 3-85784-000-5

Herbert E. Bruderer. HANDBOOK OF MACHINE TRANSLATION AND MACHINE-AIDED TRANSLATION. AUTOMATIC TRANSLATION OF NATURAL LANGUAGES AND MULTILINGUAL TERMINOLOGY DATA BANKS. New York 1978.

North-Holland Publishing Company 959 Pages, ISBN 0-7204-0717-6.

Foreward by Prof. K. Bauknecht, Department of Computer Science, University of Zurich. English translation by the Commission of the European Communities, Brussels. c.1600 item bibliography.

Herbert E. Bruderer (Hg./ed.). AUTOMATISCHE SPRACHÜBERSETZUNG. (in preparation)

Wissenschaftliche Buchgesellschaft, Darmstadt 1979. etwa 450 Seiten (Wege der Forschung, Band 272), 41-45 DM. ISBN 3-534-06312-0, nur für Mitglieder der Buchgesellschaft.

YALE A.I. PROJECT: RESEARCH REPORTS AVAILABLE

Send orders to: *M.S. Barham, Adm. Asst., A.I. Project, Department of Computer Science, Yale University, 10 Hillhouse Ave.-320DL New Haven, CT 06520*

< > # 78 - Riesbeck, C.K. & Schank, R.C. - Comprehension by Computer: Expectation-Based Analysis of Sentences In Context.

< > # 88 - Lehnert, W.G. - The Process of Question Answering

< > #116 - Cullingford, R.C. - Script Application: Computer Understanding of Newspaper Stories

< > #127 - Schank, R.C. & Carbonell, J.G. - Re: The Gettysburg Address

< > #128 - Schank, R.C., Wilensky, R., Carbonell, J.G., Kolodner, J.L. & Hendler, J.A. - Representing Attitudes: Some Primitive States

< > #131 - Lehnert, W.G. - Representing Physical Objects In Memory

< > #137 - Charniak, E. - On The Use Of Framed Knowledge In Language Comprehension

< > #139 - Riesbeck, C.K. & Charniak, E. - Micro-SAM and Micro-ELI: Exercises in Popular Cognitive Mechanics

< > #140 - Wilensky, R. - Understanding Goal-Based Stories Ph.D. Thesis

< > #141 - Schank, R.C. - Inference in the Conceptual Dependency Paradigm: A Personal History

< > #142 - Kolodner, J.L. - Memory Organization For Natural Language Data-Base Inquiry

< > #143 - Schank, R.C. & Birnbaum, L.A. - Real-Time Integrated Parsing

< > #144 - Schank, R.C. & Lebowitz, M. - Big Words

< > #145 - Schank, R.C. - Interestingness: Controlling Inferences

< > #146 - Carbonell, J.G. Cullingford, R.E. & Gershman, A.V. - Knowledge-Based Machine Translation

SUMMARY OF RESEARCH ON COMPUTATIONAL ASPECTS OF EVOLVING THEORIES

Raymond D. Gumb

Temple University

The concept of an evolving theory [3] is a natural extension of concepts in free tense logic with equality. In the semantics of free tense logic, an individual can have a property at one time that it does not at another, the domain of discourse can vary with the passage of time as individuals are born and die, and individual terms can refer at one time but not at another. Understanding a theory to be a set of sentences in the language of free tense logic with equality, an evolving theory is an indexed set of theories ordered in time. Intuitively, an evolving theory might represent the life work of a thinker, where the individual theories in the evolving theory correspond to chapters of the life work written at different times. The evolving theory reflects changes in the thinker's views with the passage of time.

Evolving theories have been studied from both semantic and proof theoretic perspectives, and various concepts such as the semantic concept of satisfiability and the proof theoretic concept of consistency have been extended to apply to evolving theories [3]. The semantics is given in terms of metaphor theory which stands intermediate

between Leblanc's truth-value semantics and model theory and translates readily into both. The deductive system for (a class of) evolving theories is called the <u>forest method</u> because a tree is generated for each point in time. The forest method, in effect a generalization of Kripke tableaux constructions, is mechanizable. The forest method is correct with respect to the semantics [3].

The forest method can be applied whenever the restriction on the temporal order relation has the <u>computable Kripke closure property</u>, a property of properties of relations which has been characterized model theoretically in [7]. The computable Kripke closures are a subclass of the <u>monotonic closures</u> [7], which have closure properties much like transitivity (a computable Kripke and monotonic closure). A preservation theorem giving the syntactic form of axiomatizations of the first-order monotonic closures [7] suggests a generalization of the Roy-Warshall transitive closure algorithm. The preservation theorem might also be used to determine (much as an entry in an engineering handbook) whether, given a property of relations Pr, there is a Pr-closure algorithm.

Evolving theories can be based in other intensional logics such as as modal [4] and intuitionistic logics. Some (but not all) of the logics underlying kinds of evolving theories where the forest method is applicable can be axiomatized [1, 6].

In certain (but not all) cases, the forest method (or restrictions of it) can be used to give effective proofs of the extended joint consistency theorem, a result which incorporates the Craig and Lyndon interpolation lemmas and the Robinson joint consistency theorem. Roughly, the theorem states that two theories T_1 and T_2 are mutually inconsistent just in case there is a separating sentence F such that F (not F) is a logical consequence of T_1 (T_2) and F "talks about" only individuals and relations that both T_1 and T_2 do. Effective proofs of the theorem are well-known in standard first-order logic, and similar results have been established in free logic with equality [5] and in a family of free modal logics with equality [4]. Since the proof is effective, a (depth first) algorithm can be extracted from it for constructing, given a closed forest for the union of T_1 and T_2, a separating sentence F.

Evolving theories, outfitted with additional devices to enhance their plausibility, appear to be a natural for representing temporal knowledge. When the extended joint consistency theorem holds for the representation language of a knowledge base and when an effective proof has been given, the separating sentence algorithm might be a useful tool for pinpointing inconsistencies in the knowledge base [2]. Only very restricted versions of the algorithms mentioned above have been programmed, those in LISP [2] and SNOBOL.

REFERENCES

[1] Barnes, R. and Gumb, R., "The Completeness of Presupposition-Free Tense Logic," Zeitschrift für mathematische Logik und Grundlagen der Mathematik, forthcoming. (Abstract in Journal of Symbolic Logic, 42(1977), 146-7.)

[2] Gumb, M. and Gumb, R., "Logical Techniques for Pinpointing Inconsistencies in the Knowledge Base," Proceedings of the 41st Annual Conference of the American Society for Information Science, 1978, forthcoming.

[3] Gumb, R. Evolving Theories. Flushing, N.Y.: Haven, forthcoming. (Abstracts in IJCAI5, v. 1, 567, and JSL, forthcoming.)

[4] -------, "An Extended Joint Consistency Theorem for a Family of Free Modal Logics with Equality," forthcoming.

[5] -------, "An Extended Joint Consistency Theorem for Free Logic with Equality," Notre Dame Journal of Formal Logic, forthcoming. (Abstract in JSL, 42(1977), 146.)

[6] Leblanc, H. and Gumb, R., "Soundness and Completeness Proofs for Three Brands of Intuitionistic Logic," forthcoming in a Haven anthology.

[7] Weaver, G. and Gumb, R., "First-Order Properties of Relations Having the Monotonic Closure Property," forthcoming. (Abstracts in Proceedings of the 1978 ACM Computer Science Conference, 19, and JSL, forthcoming.)

American Journal of Computational Linguistics

TAXONOMY: I N F O R M A T I O N S C I E N C E S

The following taxonomy is that used by the journal INFORMATION SYSTEMS.

1	general aspects	☐
2	analysis, modelling, description and evaluation of information systems	☐
2.1	analysis	☐
2.2	design	☐
2.3	modelling	☐
2.4	description	☐
2.5	implementation	☐
2.6	evaluation	☐
2.7	description of realized systems	☐
3	data base systems	☐
3.1	global aspects, global design	☐
3.2	system analysis for DBS, userdemands	☐
3.3	feasibility studies, evaluation of DBS, summary of experiences	☐
3.4	formal description of data base systems and data base languages	☐
3.5	data models, information models	☐
3.5.1	hierarchic DM	☐
3.5.2	network DM	☐
3.5.3	relational DM	☐
3.5.4	others	☐
3.6	data definition languages (DDL)	☐
3.7	data translation	☐
3.8	procedural data manipulation languages (DML)	☐
3.9	non-procedural (descriptive) DML, terminal languages (interactive languages), natural language interfaces	☐

- 3.10 dialog functions, computer assistance, computer guidance, dialog support, I/O-functions ☐
- 3.11 implementation aspects ☐
- 3.12 architecture of DBS, interfaces ☐
- 3.13 distributed DBS ☐
- 3.14 file management systems ☐
- 3.15 data structures, operations upon data structures ☐
  - 3.15.1 data structures ☐
  - 3.15.2 operations ☐
  - 3.15.3 pictorial data structures and operations, data bases in computer graphics and CAD ☐
- 3.16 storage structures, access path, access methods, search strategies ☐
- 3.17 reorganization, selforganising structures, optimization ☐
- 3.18 storage technology, specialized hardware for DBS (data base processors) ☐
- 3.19 security, integrity ☐
- 3.20 privacy ☐
- 3.21 description of realized data base systems ☐
  - 3.21.1 DBTG-based DBS ☐
  - 3.21.2 relational-based DBS ☐
    - 3.21.3.1 ADABAS ☐
    - 3.21.3.2 DMS-II ☐
    - 3.21.3.3 DMS-1100 ☐
    - 3.21.3.4 IDS II ☐
    - 3.21.3.5 IDMS ☐
    - 3.21.3.6 IMS ☐
    - 3.21.3.7 SYSTEM 2000 ☐
    - 3.21.3.8 TOTAL ☐
    - 3.21.3.9 others ☐

- 4 method and model base systems ☐
- 4.1 method base systems ☐
- 4.2 model base systems ☐
- 4.3 description of realized systems ☐

- 5 planning and decision support systems ☐
- 5.1 forecasting systems ☐
- 5.2 planning information systems ☐

5.3 decision support systems. ☐
5.4 control systems ☐
5.5 description of realized systems ☐

6 question answering systems, cognitive methods ☐

6.1 representation of knowledge, ☐
6.2 problem solving ☐
6.3 natural language systems ☐
6.4 pattern processing ☐
6.5 deduction and inference ☐
6.6 artificial intelligence methods, cognitive methods ☐
6.7 description of realized systems ☐

7 document retrieval systems ☐

7.1 indexing, classification, thesaurus problems ☐
7.2 evaluation measures ☐
7.3 documentation services ☐
7.4 library automation ☐
7.5 description of realized systems ☐

8 distributed systems ☐

8.1 distributed processing ☐
8.2 distributed storing ☐
8.3 distributed control ☐
8.4 data communication, protocols ☐
8.5 architecture, topology ☐
8.6 descript on of realized systems ☐

9 special application oriented information systems ☐

9.1 management information systems ☐
9.2 macro economic information systems ☐
9.3 information systems in public administration ☐
9.4 information systems in medicine ☐
9.5 technical information systems ☐
9.6 information systems in jurisprudence ☐
9.7 ecology information systems ☐
9.8 description of realized systems ☐

MACHINE AIDS TO TRANSLATION
A CONCISE STATE OF THE ART BIBLIOGRAPHY

WAYNE ZACHARY

Analytics
Willow Grove, Pennsylvania

Machine aids to translation, or MAT, is a loosely defined field whose boundaries shade gradually into such areas as computer science, information science, computational linguistics, and theoretical linguistics. It is therefore difficult to decide precisely what material belongs in a MAT bibliography. One can cast the net broadly and include a great deal of material that considers MAT only tangentially, or be more restrictive and include only the (much smaller body of) work which clearly concerns MAT. The latter course is taken here in order to avoid burying the essential literature under a mountain of peripheral references. Thus, the enormous body of work on fully automatic machine translation is not included (although Young (1978) contains an extensive survey of this literature). A few exceptions to this principle of inclusion (e.g., ALPAC, 1966) provide important contextual information concerning the history, current status, or future development of MAT.

While the history and current theoretical concerns of MAT are covered in the bibliography, the emphasis is on applications and operational

systems. This reflects the development of MAT as a more practical alternative to pure machine translation and its position as one of the only areas of computational lingustics that has progressed to the point of wide scale application.

A few other ground rules have been followed. First, since the field continues to change so rapidly and this is a state-of-the-art survey, few references over a decade old are included. Second, where a single MAT system is described almost identically in several research reports and/or publications in several languages, only a single reference is given, in English where possible. Third, works dealing strictly with hardware advances such as new graphics display technology, micropragrammable display fonts, or multilingual printers are also excluded.

This bibliography was compiled from a much larger one containing nearly 750 entries and is intended to provide a concise summary of the current work in the field with strong focus as stated above on machine aided translation systems that are presently in operation.*

---

* This research was conducted under a subcontract to Analytics from Chase, Rosen and Wallace, Inc. with funds provided by the Office of Research and Development, Gentral Intelligence Agency.

BIBLIOGRAPHY

ALPAC (Automatic Language Processing Advisory Committee)
1966 Language and Machines - Computers in Translation and Linguistics National Academy of Sciences, National Research Council: Washington, D.C.

American Mathematical Society
1968 "Research on Machine Aids to an Editor of Scientific Translations." NTIS document PB-177 775.

Berzon, V.E.
1971 "Some Techniques for Formalizing the Process of Establishing U-Relations Between Sentences in a Corrective Text," in Nauchnó-Tekhniceskaya Informatsia, Seriya 2, No. 8.

Bisby, R. and Kay, M.
1972 The MIND Translation System: A Study in Man-Machine Collaboration. NTIS Document Ad 749 000, Rand Corp.: Santa Monica, Calif.

Bruderer, H.E.
1977 "The Present State of Machine Aided Translation," Overcoming the Language Barrier, in Commission of the European Communities, pp 529-556.

Bruderer, H.E.
In Press Handbook of Machine and Machine Aided Translation, North Holland: New York (published in German in 1975).

Burge, John
1978 "The TARGET Project's Interactive Multilingual Dictionary," Project Technical Report No. 13. Depts. of Modern Languages and Computer Science; Carnegie-Mellon University.

Charniak, E. and Wilks, Y.: eds
1976 Computational Semantics, North Holland: New York.

Chevalier M., Danserau, J., and Paulin, G.
1978 TAUM-METO: Description du Systeme. University de Montreal, Montreal, Canada.

Commission of the European Communities
1977 Overcoming the Language Barrier: Third European Congress on Information Systems and Networks. Vol. 1. Munich: Verlag Dokumantations.

Daley, Cd. D.H., and Vechino, Lt. Col. R.F., USAF
1973 "The West German Federal Bureau of Languages and Machine Aided Translation in Germany." In Federal Linguist, Vol. 5, No. 3-4; pp 14-18.

Dubuc, Robert
1972 "TERMIUM: System Description," in Meta: Journal of Translation, Vol. 17, No. 4, pp 203-219.

Dubuc, Robert, and Gregoire, Jean-Francois
1974 "Banque de Terminologie et Traduction," in Meta: Journal of Translation Vol. 20, No. 4, pp 180-184, (in French).

Goetschalckx, J.
1974 "Terminology and Documentation in International Organizations," in Babel, Revue Internationale de la Traduction, Vol. 20, No. 4, pp 185-187.

Grimes, Joseph E.
1970 "Computing in Lexicography," in The Linguistic Reporter, Vol. 12, Nos. 5-6, pp 1-4.

Hann, Michael
1974 "Principles of Automatic Lemmatisation," in ITL, Review of Applied Linguistics, Vol. 23, pp 3-22.

Hirschberg, Lydia
1965 "Dictionaries Automatiques pour Traducteurs Humains," in Journal des Traducteurs, Vol. 10, No. 3, p 78, (in French).

Hlavac, T.
1973 "Dealing with the Language Barrier by Means of Computer," in Kniznue and Ved. Inf. (Czcheoslovakia), Vol. 5, No. 2, pp 70-75.

Ivanova, I.S.
1969 "Problems of Automatic Thesarus Construction," in Nauchno-Tekhniceskaya Informatsiya, Seriya 2, No. 1, pp 17-20

Joint Publication Research Service
1977 "Machine-Assisted Translation in West Germany," translation of article by various authors, NTIS document JPRS 68726.

Jordan, S.R., Brown, A., and Hutton, F.C.
1977 "Computerized Russian Translation at ORNL," in Journal of the American Society for Information Science, January, pp 26-33.

Josselson, H.
1967 "Research in Automatic Russian-English Scientific and Technical Lexicography," Final Technical Report, Wayne State University. NTIS document PB-180 930.

Kay, Martin
1976 "The Proper Place of Man and Machines in Translation," in American Journal of Computational Linguistics, Microfiche 46.

Krollman, Fredrick
1971 "Linguistic Data Banks and the Technical Translator," in Meta: Journal of Translation, Vol. 16, Nos. 1-2, pp 117-124.

Krollman, Fredrick
1974 "Data Processing at the Translators Service," in Babel, Revue Internationale des Traducteurs, Vol. 20, No. 3, pp 121-129.

Krollman, Fredrick
1977 "User Aspects of an Automatic Aid to Translation as Employed in A Large Translation Service," In Overcoming the Language Barrier, by the Commission of the European Communities, pp 243-257.

Lippman, Erhard O.
1971 "An Approacn to Computer Aided Translation," in IEEE Transactions on Engineering Writing and Speech, Vol. 14, No. 1, pp 10-33.

Lippman, Erhard O.
1975 "On-Line Generation of Terminological Digests in Language Translation," in IEEE Transactions on Professional Communications, Vol. PC-18, No. 4, pp 309-318.

Lippman, Erhard O., and Plath, W.J.
1970 "Time Sharing and Computer Aided Translation," in The Finite String, Vol. 7, No. 8, (microfiche).

Loh, Shiu-Chang
1976 "Translation of Three Chinese Scientific Texts into English by Computer," Association of Literary and Linguistic Computing Bulletin, Vol. 4, No. 2, pp 104-106.

Loh, S.C., and Kong, L.
1977 "Computer Translation of Chinese Scientific Journals," in Overcoming the Language Barrier, in Commission of the European Communities, pp 631-646.

Luther, D.A., Montgomery, C., and Case, R.
1977 "An Interactive Text-Editing System in Support of Russian Translation by Machine," in IFIPS National Computer Conference Proceedings, pp 789-790.

Lyttle, E.G., et. al.
1977 "Junction Grammar as a Base for Natural Language Processing," in American Journal of Computational Linguistics, 1975, No. 3, (microfiche).

Mathias, Jim
1973 "The Chinese-English Translation Assistance Group and Its Computerized Glossary Project," in Federal Linguist, Vol. 5., Nos. 3-4, pp 7-13.

New Scientist
1977 "Many Hands Make Chinese Translator Work" in The New Scientist, Vol. 76, No. 1073, p 88.

Price, James
1970 "Abstract of the Development of a Theoretical Basis for Machine Aids for Translation from Hebrew to English," in Hebrew Computational Linguistics Bulletin, Vol 2., 65-83.

Schmidt, R., and Vollnhals, O.
1974 "The Use of the Lexicographical Branch of a Data Bank System to Produce a Phraseological Technical Glossary," in Machine Assisted Translation in West Germany, JPRS document 69726.

Schulz, J.
1971 "Le Systems TEAM, une Aide a la Traduction," in Meta: Journal of Translator, Vol. 16, Nos. 1-2. (in French).

Schulz, Joachim
1974 "Lexicography with TEAM -- Automatic Dictionary Composition," in Machine Assisted Translation in West Germany, JPRS document 68726, pp 23-34.

Shaffer, Richard A.
1978 "California Firm to Unveil a Computer that Processes Words for Translators," in The Wall Street Journal, 24 October 1978.

Sinaiko, H.W.
1971 "Translation by Computer," in Science, Vol. 174, pp 1182-1184.

Smith, Raoul
1978 "Computational Bilingual Lexicography: A La Recherche du mot Juste," paper read at Foreign Broadcast Information Service Seminar on Computer Support to Translation.

Stallings, W.
1975 "The Morphology of Chinese Character: A Survey of Models and Applications," in Computer and the Humanities, Vol. 9, pp 13-24.

Walker, Gorden, Kuno, Susumu, Smith, Barbara, Hold, Roland
1968 "Chinese Mathematical Text Analysis," in IEEE Transactions on Engineering Writing and Speech, Vol, 11, No. 2, pp 118-128.

Warotamasikkhadit, U., Kanchanawan, N., and Londe, D.
(no date) "The Design and Construction of a System to Transliterate Thai by Computer," in 6th Australian Computer Conference Proceedings, pp 833-839.

Weaver, W.
1955 "Translation," in Machine Translation of Languages, W.N. Locke and A.D. Booth, eds., Technology Press of MIT and John Wiley and Sons: New York.

Weber, Heintz Josef
1976 "Automatiche Lemmatiseirung -- Zielsetzung und Arbeitsweisceines Linguistischen Identifikationsverfahrens," in Linguistische Berichte, Vol. 44, pp 30-47 (in German).

Wilks, Yorick
1973 "An Artificial Intelligence Approach to Machine Translation," in Computer Models of Thought and Language, Roger Schank and Kenneth Colby, eds., W.N. Freeman and Co.: San Francisco, pp 114-115.

Unknown
1976 "The Lexicography Information System (LEXIS) of the Bundeswher Language Service," in Machine Assisted Translation in West Germany, NTIS Document JPRS 68726.

Young, Mary E.
1978 "Machine Translation (A Bibliography with Abstracts)," available from National Technical Information Service, NITS document PS-78/0448.

ON HUMAN COMMUNICATION:

A Review, a Survey, and a Criticism

COLIN CHERRY

3rd Ed. Cambridge, MA
M.I.T PRESS, 1978

374 + xv pp.
ISBN 0-262-03065-9

Reviewed by

WILLIAM L. BENZON

Language, Literature and Communications
Rensselaer Polytechnic Institute
Troy, New York 12181

The comments on the dust jacket for the third edition of Colin Cherry's On Human Communication come from reviews of previous editions which appeared in such diverse places as the Canadian Journal of Psychology, Physics Today, and Romance Philology, and indicate that it is a broad ranging and fascinating book. The range is certainly broad, so broad that I have included a table-of contents as an appendix to this review rather than attempting to summarize the contents of the book.

But it is no longer fascinating. Students of human communication have thought and debated much and even learned a little between the publication of the first edition of On Human Communication in 1957 and the publication of the third edition in 1978. But very little of that material has become part of the substance of Cherry's book (although some of it is cited). Thus, while the first edition may well have been "A Review, a Survey, and

a Criticism" (the book's subtitle), the third edition is not. Too much is left unconsidered.

The most serious gap is in the consideration of natural languages. Cherry gives the impression that information theory, statistics, Fourier analysis, and perhaps a little logic are the most important formal tools for the analysis of natural language. That may have been true when Cherry wrote the first edition of the book, but it is certainly not true now. Cherry does make a few references to Chomsky, but none of the substance of the Chomskian revolution (not to mention post-Chomskian developments) has affected Cherry's treatment of natural language. The texture of theorizing and model building in linguistics, psycholinguistics, and cognitive psychology has undergone considerable change since the Fifties, but little of that has become part of the substance of On Human Communication.

Similarly, a great deal of work has been done on nonverbal communication in the last two decades. While Cherry alludes to some of this work, he makes no attempt to summarize any of the major lines of inquiry. As with modern linguistics, the material is too diverse to cover it all in the sort of survey which Cherry intends On Human Communication to be. What I find so disheartening is that so little of this material is mentioned at all, especially when one realizes that Cherry has added a new chapter ("Human Communication: Feeling, Knowing, and Understanding") to the third edition for the purpose of talking about what is specifically human about human communication. That Cherry should devote ten pages to Zipf's law while not even mentioning the work of Paul Ekman, Carroll Izard, and Manfred Clynes (to mention only the work which comes most readily to my mind) on the expression and communication of emotion is bizarre.

Finally, Cherry's treatment of semiotics is relatively insulated from most of the semiotic research of the last two decades. Semiotics is itself

such a diverse and amorphous enterprise (as diverse and amorphous as the study of human communication is) that it is perhaps unfair to criticise Cherry for shortchanging it. But Cherry introduced the topic of ritual into his final chapter and that is a subject on which semioticians have had a great deal to say (I am thinking of structural anthropologists such as Claude Levi-Strauss, Edmund Leach, and Victor Turner). Consequently I am inclined to view Cherry's neglect of semiotics perhaps more harshly than I otherwise would.

No doubt Cherry could be charged with other sins of ommission, but the three I've mentioned are serious enough. It is equally beyond doubt that an edition of On Human Communication which included this material would be a very different book, not Colin Cherry's book at all. If Cherry had attempted and achieved a synthesis of his material, then its value as a synthesis might well outweigh the dated nature of some of the elements of the synthesis. But Cherry wasn't after a synthesis; he simply wanted to see what was out there. There is now much out there which Cherry hasn't seen. Consequently On Human Communication is not a good guide. The person who wants or needs a general introduction to the subject of human communication which reflects the current state of the art(s) will have to look elsewhere.

APPENDIX: Abbreviated table of contents from On Human Communication

Chapter 1:	Communication and Organization - An Essay	1
1	The Scheme of This Book	2
2	What Is "Communication"?	3
3	What Is It That We Communicate?	9
4	Some Difficulties of Description of Human Communication	15
5	Co-operative and Non-co-operative Links	16
6	Communication and Social Pattern	19
7	Group Networks	26
8	Communication Is an Act of Sharing	30
Chapter 2:	Evolution of Communication Science - an Historical Review	31
1	Language and Codes	32
2	The Mathematical Theory of Communication	41
3	Brains - Real and Artificial	52
4	On Scientific Method	62
Chapter 3:	On Signs, Language, and Communication	68
1	Language: Science and Aesthetics	68
2	What Is a Language?	77
3	Toward a Logical Description of Language	87
4	Features as the "General Co-ordinates" of Speech	101
5	Statistical Studies of Language "Form"	102
6	Words and Meaning: Semantics	111
Chapter 4:	On Analysis of Signals, Especially Speech	124
1	The Telecommunication Engineer Comes onto the Scene	124
2	Spectral Analysis of Signals	131
3	Speech Representation on the Frequency-Time Plane	146

4 The Specification of Speech 160

Chapter 5: On the Statistical Theory of Communication 169

1 Doubt, Information, and Discrimination 169

2 Hartley's Theory: "Information" as Logical "Instructions to Select" 172

3 When the Alternative Signs Are Not Equally Likely to Occur 178

4 The Use of Prior Information: Redundancy 182

5 Messages Represented as Wave Forms: "Continuous" Information 189

6 Communication as Information, When Noise is Present 198

7 The Ultimate Capacity of a Noisy Channel 206

8 Mandelbrot's Explication of Zipf's Law - Continued 211

9 Comments on Information Interpreted as Entropy 214

Chapter 6: On the Logic of Communication (Syntactics, Semantics, and Pragmatics) 219

1 "Significs" - of Mental Hygiene 219

2 Are Different Measures of "Information" Needed? 231

3 About "Semantic Information" 233

4 Syntactic, Semantic, and Pragmatic "Information" A Relationship 243

5 Language, Logic, and Experiment 252

Chapter 7: On Cognition and Recognition 258

1 Recognition as Our Selective Faculty 258

2 Some Simple Philosophical Notes 262

3 Recognition of Universals 269

4 The Importance of Past Experience: Reality and Nightmare 271

5 The Intake of Information by the Senses: Some Quantitative Experiments 282

6 The Search for Invariants, in Pattern Recognition 291

7 On the Brain as a "Machine" 300

Chapter 8: Human Communication: Feeling, Knowing, and Understanding 305

1 Communication Is Always an Act of Sharing 306

2 Signs of Cause versus Signs of Meaning 312

3 The Importance of Ritual 314

4 Spontaneous Speech. The Extraction of Meaning 318

5 Human Language and Animal Signaling 329

6 On Human Communication 334

Appendix 339

References 344

Index 365

A B H A N G I G K E I T S G R A M M A T I K

JURGEN KUNZE

Zentralinstitut fur Sprachwissenschaft
Akademie der Wissenschaften der DDR

Akademie-Verlag
Berlin, DDR

504 pp.
1975

Reviewed by

KENNETH F. BALLIN

Department of Linguistics
SUNY at Buffalo
Amherst, New York 14261

Jürgen Kunze establishes his dependency grammar with four components. The syntactic is the most important. The three non-syntactic components are the paradigmatic component, the selectional component, and the assigning component. In the first chapter of his book ABHANGIGKEITSGRAMMATIK (Dependency Grammar) the reader gets introduced to some of the basic concepts useful in understanding the notions explicated later on. Subordination or dependency is introduced by way of a diagram, known as a tree, consisting of several connected points. A point or node that is connected to one closer to the top of the page is subordinate to it. This is called direct dependency. Indirect subordination is when two nodes are connected with one or more points in between them. These three nodes comprise a part tree.

Obviously there are several part trees which combine to make a tree. If the bottom-most node of our little part tree is not superordinate to any other point then the part tree is an end complex. Every node is an end complex with itself as its only member.

Once one decides to attach words to these nodes it changes from a connect the dots game to some sort of meaningful diagram The first step in this change is to bring order to the diagram. Since language is the object of study here and the language the book was written in procedes from left to right, the author has ordered his tree from left to right. This type of tree is known as a W tree, i.e. where each node is attached to a word. The book deals with M trees. These are trees in which the nodes are connected to signal combinations (Merkmalkombinationen). A marked tree is one in which all the connections are subordinate relations on one kind or another from a set containing all the kinds of subordinate relations possible.

In making his investigations, Kunze has limited his field of study to modern day written German. This suffices as for in any pure theoretical investigation it is acceptable to assume the observed language is a set of given sentences. The practicability of his theory depends on finding a standard of correctness. In this case tapping the knowledge of a native speaker is of no help Four ways are suggested as possibilities for this standard of correctness. The first is grammatical correctness in which all sentences are acceptable as long as they function as members of their classes, i.e. nouns as nouns, verbs as verbs. Second is a more refined grammatical correctness taking the meaning of the verb into account. Third is the suggestion of a very strict grammar

bordering on grammar and semantics including semantic categories such as ABSTRACT or CONCRETE. The fourth consideration is a semantic grammatical correctness. Though this standard of correctness is needed to make the theory work, measuring correctness is not a major factor.

There is, says Kunze, a base language and base structures that can express semantic and syntactic ambiguities. It is important when studying these structures to consider which categories and qualities are contained in it. A category is a variable with set value ranges, for example, in German, case. Qualities are restrictions imposed for ordering, appearance or non appearance of sentence fragments. Plainly not all categories and qualities are in every sentence of a language. An expansion of the base language leads to a simplification of the descriptive system but also costs quite a bit as far as analysis is concerned.

On starting into the meat of the matter tha author writes that in no way can one expect such a simple tool as dependency trees to encompass the linguistic relations within the sentence that are conditioned through language. This inadequacy is evidenced by the following situation. Every grammatical structure has an ordered dependency tree. It is however possible to have two different structures represented by the same tree. This is one of the principles for the representation of sentence structures using dependency trees. Reduction is another principle by which we get sentences like 'My friend will bring the book' from sentences like 'My friend will bring you the book tomorrow.' An additional principle removes those nodes which were dropped from the latter sentence to arrive at the former This procedure is only permis-

sible when the middle steps are somewhat acceptable.

An interesting concept is introduced by the author. He calles it the configuration criterion. It ways that one element is substitutable for another it it has all the same grammatical properties. This concept is used frequently in deciding what is dependent on what.

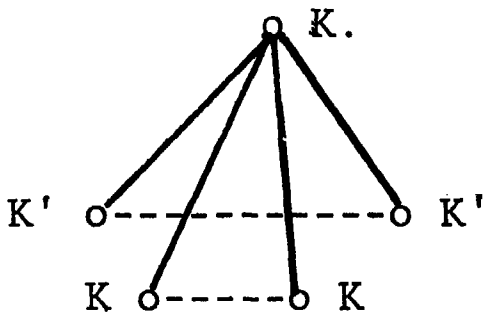
The fourth chapter deals with the non-syntactic components. Paradigmatic categories are established. The categories are Genus, commonly called gender, Person, and Number. Each of these three categories is established on the basis of data similar to the following example. I saw him. We saw him. These two sentences are syntactically equal but paradigmatically quite different. This illustrates the category of number, the first person singular changes to the first person plural. The author splits these categories again to account for the phenomenon of endings in German. It is possible to have a possessive pronoun with a masculine ending referring to a female person. Quasi categories are also established. These are tempus, modus, and case, and they are only quasi categories because they affect other parts of speech in a sentence. Kunze constructs a list which enumerates the category responsible for their relations, e.g. a noun is paradigmatically related to its apposite through case.

The separation of the paradigmatic from the selectional is due to the ease with which the former are presentable. Selectional relations are more narrowly defined in this case than in generative grammar. As with the paradigmatic relations there are nine selectional relations, five of which belong to the in-

ventory and the other four don't.

There are demands made on a system of subordinate relations. The first of these is that the marked tree should be an adequate representation of the syntactic structure of the sentence. Secondly, the subordination relations must allow all categories, qualities, and relations in the base structure to represent and differentiate the paradigmatic and selectional relations that can't be expressed through assigning.

Affectation ways (Wirkungswege) are dashed lines connecting two nodes dominated by a third node (see diagram). They represent



other relations that exist between nodes aside from subordination. That these affection ways of both the paradigmatic and selectional relations must be represented through subordination relations is another demand made on the system. The last demand made is that the conditions for the paradigmatic and selective points (Vorgaben) must also be represented.

The principle called the differentiation principle proves these last two are met. The system makes this determination by using a a knowledge of dependency trees, a fixed inventory of paradigmatic and selectional relations, and a fixed language base in a way which yields the required relations.

The last concept developed by the author is that of bundles.

There are four types of bundles - a simple bundle, an elementary bundle, a complex bundle, and a complex implication bundle. A bundle is a tree used to represent not a sentence but a set of sentences, i.e. trees. In a complex bundle the paradigmatic and selective properties need only be given once.

Chapter 8 is a discussion of some questions that were brought out as a result of this theory.

GLANCING, REFERRING AND EXPLAINING IN THE DIALOGUE SYSTEM HAM-RPM

W. WAHLSTER, A. JAMESON, W. HOEPPNER

Project: 'Simulation of Language Understanding'
Germanisches Seminar der Universitat Hamburg
von-Melle-Park 6, D-2000 Hamburg 13, West Germany

SUMMARY

This paper focusses on three components of the dialogue system HAM-RPM, which converses in natural language about visible scenes. First, it is demonstrated how the system's communicative competence is enhanced by its imitation of human visual-search processes. The approach taken to noun-phrase resolution is then described, and an algorithm for the generation of noun phrases is illustrated with a series of examples: Finally, the system's ability to explain its own reasoning is discussed, with emphasis on the novel aspects of its implementation.

1. THE TREATMENT OF VISUAL DATA

The natural language dialogue system HAM-RPM[1] converses with a human partner about scenes which either one or both are looking at directly (or have a photograph of). At present the system, which is implemented in FUZZY (LeFaivre 1977), is being tested on two domains: the interior of a living room and a traffic scene.

Since it is assumed that both partners begin the dialogue with relatively little specific knowledge about the scene, most of the specific information used by the system during the conversation must be obtained by a process more or less analogous to looking at the scene. We have found it worthwhile to make the analogy quite close, requiring the system to retrieve its visual data by doing something like casting a series of glances centered on various points in the scene.

Fig. 1 is a schematic drawing of a section of our traffic scene, showing a tree with a parking lot in front of it. How easy is it to recognize the various objects in Fig. 1 when glancing at point A? CAR9 and CAR8 will be about equally easy to recognize as cars. TREE4 will probably be recognized more easily, since it is equally close to point A, and very large, and since there are no similar types of objects. On the other hand, CAR3 will be less easily recognizable, since it is farther away. MAN4 is probably too far away to be recognizable as a man at all (he is recognizable only from the points nearest him, as is shown by the four arrows pointing away from him).

Just this information is stored in HAM-RPM in a separate associative network corresponding to point A. In all, there are about a hundred such small networks (represented by the small dots in Fig. 1), corresponding to possible glances at the scene. The statements about the nature of the various objects which are recognizable from the point in question are ordered, in a way characteristic of the FUZZY programming language, in terms of their recognizability, so that they will automatically be retrieved in that order[2].

1) The system's overall structure is described in (v. Hahn et al. 1978) as are the goals and methodological principles which guide the research within the project.

2) These networks are implemented as CONTEXTs in the sense introduced by the language CONNIVER.

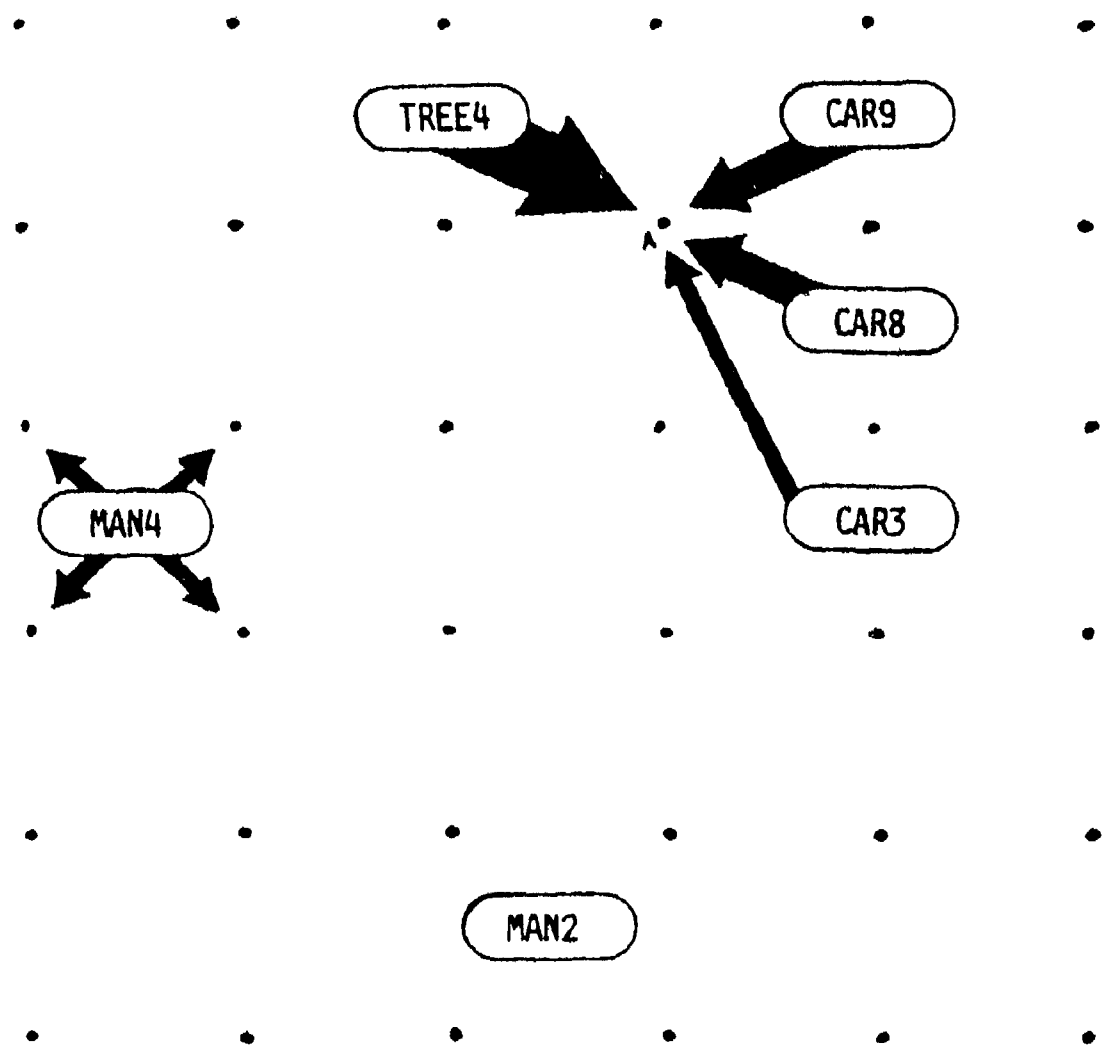


Fig. 1. "The man in front of the tree"

A simple example will show how the data stored in this way can be used by the system. When interpreting the definite description *the man in front of the tree*, assuming that TREE4 is the one meant, the system enters several CONTEXTs in front of TREE4, within each retrieving the internal names of the men recognizable from that point. It doesn't find MAN4 until it has entered the CONTEXT corresponding to point B. It then enters a couple more, and, finding no further men, assumes that it has found the referent of the definite description. Information not only about the respective types of the various objects, but also about their other attributes is stored in a similar way.

Why is it worth all this trouble to make the system sensitive to the recognizability of the various facts about a scene from the various points within it? After all, the facts themselves could be stored very straightforwardly.

Our principal justification is that, for a dialogue system which is supposed to communicate effectively with a human partner, the bare facts

about the scene are less important than the way the partner himself would be likely to perceive them. If only the facts themselves are known, information may be lacking which is essential for the production of a communicatively adequate response. For example, the definite description whose interpretation was just sketched was, strictly speaking, ambiguous, as there is a second man in front of the tree whom the system would have considered to be the referent of the description if MAN2 hadn't been there. Yet the system didn't even notice this ambiguity, since it stopped shortly after finding the first man.

To be sure, the resolution of such ambiguities could also be achieved by giving the system general information on the recognizability of objects for human beings and letting it choose on that basis which of the potential referents of the description was the one which the partner was most likely to have intended to refer to. Instead of doing this, we have made the system itself a model of its partner, so that instead of referring to a model, it only has to 'be itself' or 'act naturally' in order to communicate effectively[1].

In addition to the interpretation of ambiguous utterances, there are other situations in which this approach can be applied elegantly (Fig. 2).

SITUATION	INFORMATION
1) Interpretation of an ambiguous definite description	Which object the speaker is probably referring to
2) Generation of a definite description	Which reference points will be easy for the listener to find
3) Description of a part of a scene	Which objects the speaker might be interested in hearing about

Fig. 2

1) Two of the reports (v. Hahn 1978a, 1978b) which have been issued by the HAM-RPM group deal with the question of the nature of the relation between the dialogue partner model and the human partner in some detail.

When describing the location of an object with reference to other objects, the system will usually find a number of potential reference points; in general, it should mention those which are visually easiest for the listener to find. This is likely to happen if it itself finds these reference points particularly easily. When answering a vague question, such as a request to describe what is on the other side of the street, the system will have to select among the many visible facts those which the listener might be interested in hearing about. In many cases, these will be the visually most salient facts.

2. NOUN-PHRASE RESOLUTION

Two of the components of HAM-RPM which make use of the visual data are those responsible for noun-phrase resolution, that is, the determination of the potential referents of a noun phrase, and noun-phrase generation, that is, the construction of noun phrases to identify objects uniquely.

The procedures which resolve noun phrases work on the shallow structure of the input sentence. This is what is obtained after multiple-word phrases and idioms have been replaced with canonical expressions, the words have been looked up in the lexicon, and a simple morphological analysis has been performed.

A definite noun phrase is recognized within the shallow structure as a structure consisting of a definite article, possibly one or more attributes, a noun, and possibly a relative clause (Ritchie 1977). In a way reminiscent of Winograd's SHRDLU (Winograd 1972), processes involving semantics and pragmatics are activated in HAM-RPM as soon as possible during the analysis of the input sentence.

The noun-phrase interpreter tries to find a unique referent for each definite noun-phrase by using the knowledge stored in the conceptual and referential semantic networks and performing visual search algorithms. For example, the definite description *The picture hanging to the left of the red chair*, referring to Fig. 3, is replaced with the internal object-name PICTURE1 in the shallow structure of the sentence. This strategy can save a good deal of unnecessary processing: if no object is found which satisfies the description, there is no further parsing, but rather feed-back to

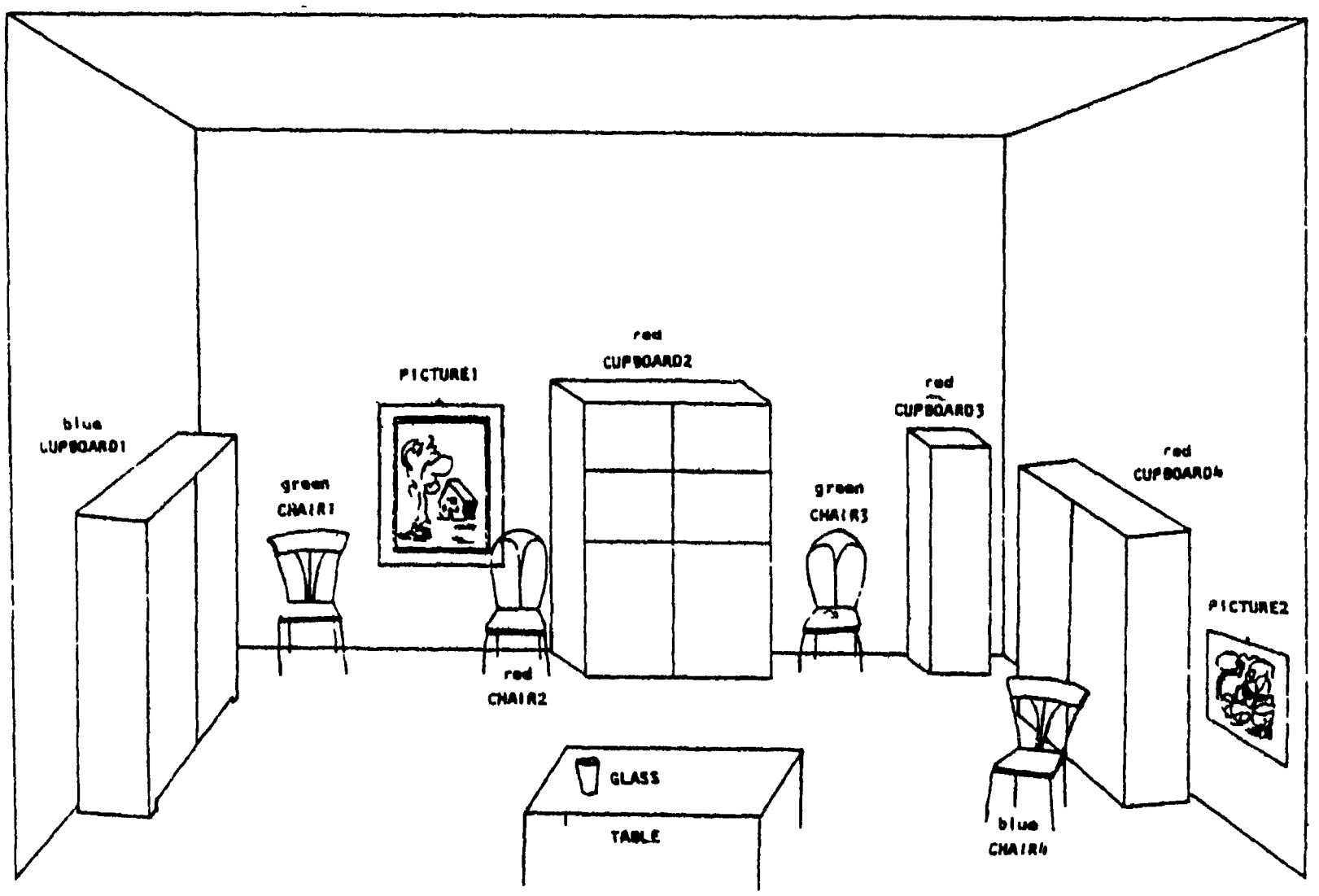


Fig. 3

the conversational partner. In the case where more than one potential referent is found, the one mentioned most recently is assumed to be the referent. If none of them has been mentioned recently, the system asks the partner for further details, assuming, as it were, that he does have some particular object in mind. These details take the form of a noun phrase, which may be either complete or elliptical. Further attributes of the intended object may be specified, it may be characterized in terms of its spatial relations to other objects, or the noun originally used in the description may be replaced with a more specific one.

Not all noun phrases, of course, can be replaced immediately with a specific referent. One such case is exemplified by the description *the chair in front of the red cupboard*. Applied to the scene in Fig. 3, the noun phrase *the red cupboard* cannot be replaced, because there is more than one red cupboard, but it cannot be ignored, either, because there is more than one chair.

The entire noun phrase can only be interpreted when it is recognized that there is only one pair of objects which stand in this relation to one another.

Another case where a definite noun phrase can't simply be replaced directly by its referent is the generic description with definite article, as in the sentence *The chair is something to sit on.* Lately we have been thinking about what formal features of a sentence might be helpful in recognizing such descriptions (see Grosz 1976).

Two clues which tend to favour a generic interpretation are the absence of any referential attribute and the presence of an adverb such as *usually* or *normally*. On the other hand, a generic interpretation becomes somewhat less plausible if the noun phrase is the object of a local preposition, as in *on the chair*, if the sentence is in the past tense; or if the verb can be generally classified as one involving visual perception or spatial relations. We assume that, no matter how many weak inference rules of this sort are incorporated into the system, there will still be some ambiguities which can only be resolved by other means, including interaction with the speaker.

A general goal in this connection is a sort of compatibility between noun-phrase resolution and noun-phrase generation, in the sense that the system should be able to understand any kind of noun-phrase that it can generate, and vice versa.

3. NOUN-PHRASE GENERATION

The method we have developed for the inverse process, noun-phrase generation, is distinguished from earlier approaches mainly in three respects.

The first is its use of what might be called a 'worst-case-first' strategy. The second is the way it takes into consideration the ease with which the listener will be able to interpret the description it generates, when more than one uniquely identifying description is possible (Herrmann & Laucht 1976). The third is its use of complex spatial relations to deal with the 'worst cases', that is, those in which several objects are indistinguishable on the basis of their properties alone.

Let's examine a few examples of the behavior of the algorithm First, two trivial cases.

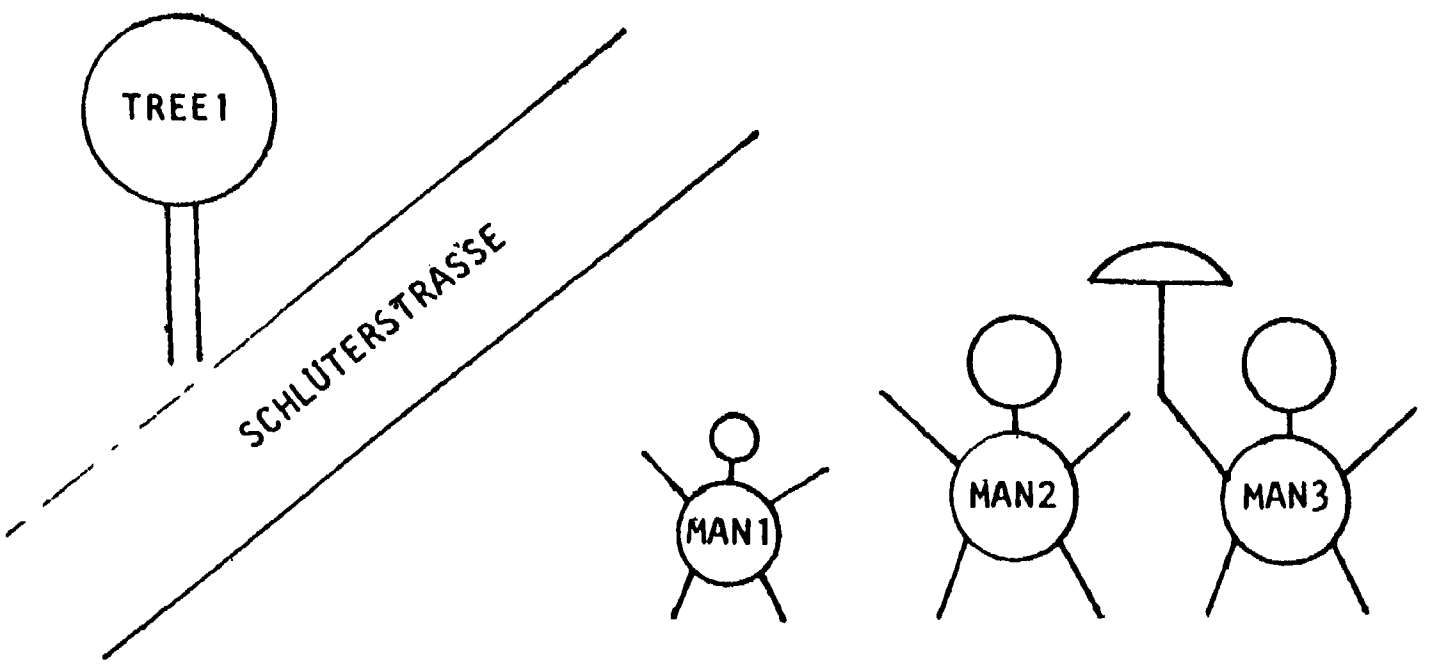


Fig. 4

The street in Fig. 4 has a proper name, and is thus referred to simply as *Schluterstrasse*. The tree is the only one in the discourse world, and hence is identified as *the tree*. The number of interesting possible strategies becomes greater when the object to be described is one of several belonging to the same conceptual class. Consider for example MAN1 in Fig. 4. The system looks among its properties for one which distinguishes it from MAN2 and MAN3, and describes it as *the small man*. A similar process underlies the generation of the noun phrase *the big man with the umbrella* to refer to MAN3.

Note that the system uses redundant labels. This is a consequence of the sequential nature of its noun-phrase generation: First, the property 'big' is found. When the system notices that there is another big man in the scene, it looks for a further distinguishing property and finds the umbrella. This property would in fact be adequate in itself, but the system doesn't attempt to find a minimal characterizing set of attributes. This sort of redundancy, which is often found in human beings, saves time both in the generation and in the interpretation of definite descriptions.

HAM-RPM frequently uses negative characterizations of various kinds, as, for example, when MAN2 is described as *the big man without an umbrella*. Now let's turn to some more complex problems of noun-phrase generation. So that the pictures don't get too cluttered, we will use examples from a simple

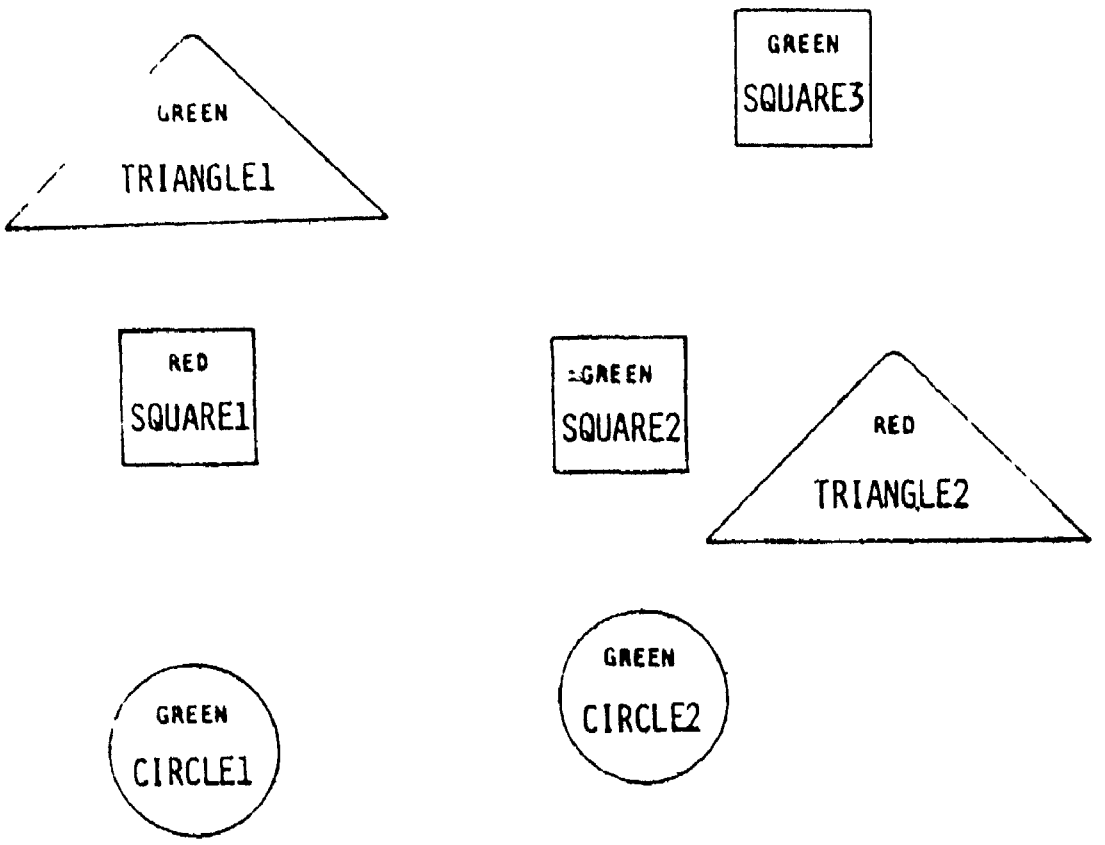


Fig. 5

domain of geometrical figures (Fig. 5). Consider CIRCLE1 in Fig. 5. Note that there are two green circles in the scene. The presence of several objects which are indistinguishable on the basis of their attributes alone is the worst case which can occur. The reason why we have spoken of a 'worst-case-first' strategy is that the system checks for this case early, rather than trying immediately to construct a simpler characterization such as those in the last few examples given.

Informal observation shows that human beings also often notice the presence of identical objects in a scene immediately. The only way to distinguish these two circles is by reference to spatial relations, for example, *the green circle in front of the red square*.

We may note in passing two ways in which the form of a description may be constrained by the form of the question which is being answered. First, properties which have been presupposed in the question should not be mentioned in a description. Consider the question *Which square is red?*. The answer *The red square* is clearly unacceptable, so instead the system answers *The square in front of the green triangle* (= SQUARE1 in Fig. 5). A second constraint of this sort is that the system should not produce circular

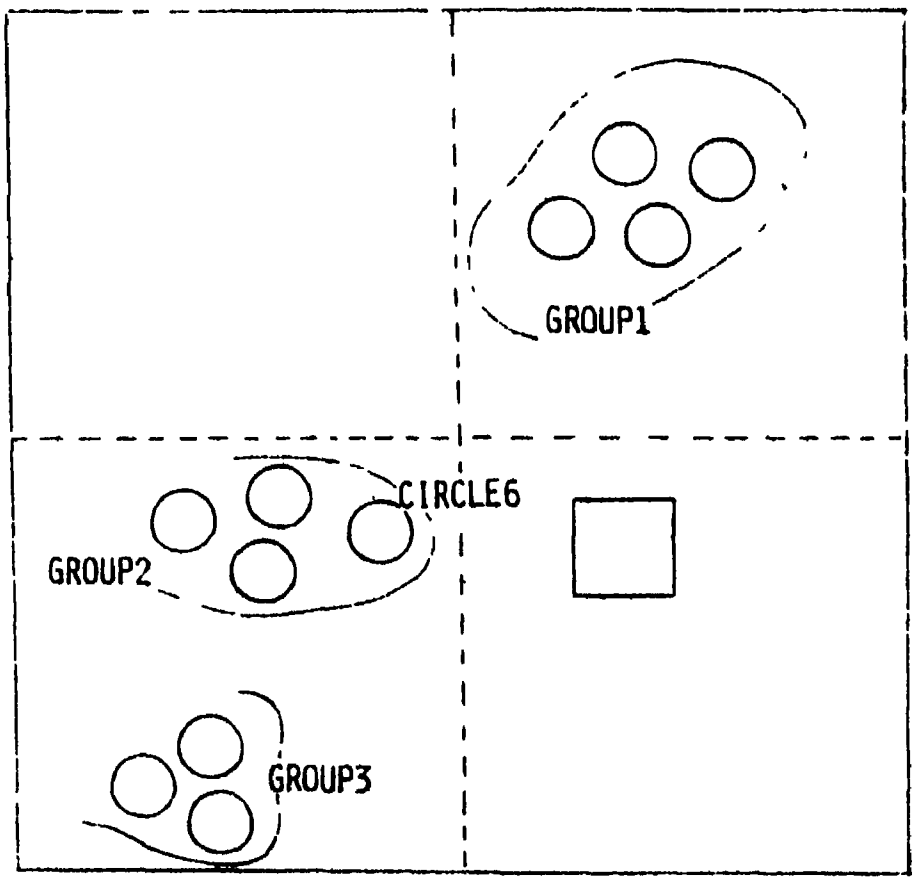


Fig. 6

descriptions. For example, when answering the question *Where is the red triangle?*, the system should not answer *To the left of the green square which is to the right of the red triangle*, although each half of this description is perfectly natural when considered in isolation.

It sometimes happens that objects chosen as spatial reference points in a description in turn have to be identified with the help of other reference points. For example, CIRCLE2 in Fig. 5 is described as *the green circle in front of the green square which is to the left of the red triangle*. As this example shows, the products of such recursive applications of the generation algorithm can soon become difficult to understand. We have made the maximum allowable depth of recursion a parameter which can be set to various values for experimental purposes.

Perhaps the most difficult problem in noun-phrase generation is the unique identification of an object when there are objects with exactly the same properties in its immediate neighbourhood (Fig. 6). This is a task which often causes difficulties even for a human speaker. To set the stage, suppose that CIRCLE6 in Fig. 6 is to be identified. The system first describes its position relative to the other circles in its group: *the right-hand circle* Then it identifies the group of which CIRCLE6 is a member with-

in the scene as a whole, distinguishing it first from GROUP2: *in front and to the left* and then from GROUP3: *which is to the left of the square.* Thus the complete description is *The right-hand circle in front and to the left which is to the left of the square.*

To put the point more generally, complex scenes sometimes have a hierarchical structure in which groups of similar objects serve as units which have to be identified in much the same way that objects themselves are. The remarks we have made about circular descriptions and recursion depth apply on the level of groups as well.

Concluding this sketch of HAM-RPM's noun-phrase interpreter and generator, we would like to stress that all these algorithms are domain-independent.

4. EXPLANATION

Although all of the examples discussed up to now have involved some sort of description of visible aspects of a scene, HAM-RPM frequently makes use of general knowledge and inference rules to draw conclusions.

For example, the system might be asked *Is the parking zone tarred?*, where the parking zone in question, though part of the scene, is hidden from view. It would then try to answer the question using approximate inferences based on fuzzy knowledge (Wahlster 1978), concluding that the parking zone might very well be tarred, because a parking zone is in a sense a part of a street, and streets, like thoroughfares in general, are usually tarred. Inferences which stand on such shaky ground as this one are of limited use to the conversational partner unless the system can describe the reasoning which underlies them.

Furthermore, not just any description will be satisfactory: the system ought to act in accordance with the following three maxims, as formulated by (Grice 1975):

1. Make your contribution as informative as is required.
2. Don't make your contribution more informative than is required.
3. Be relevant.

Thus, describing an inference chain in every detail will not in general be communicatively adequate, if some of the inferences are essentially defini-

tional, and hence conceptually trivial Only when the dialogue partner has repeatedly requested details about inferences will it be sensible to mention all of them.

Now let's look at the way we have tried to achieve these goals in HAM-RPM, using the example just given. Three processes are essential. First, while the reasoning is being performed, a sort of trace of the inference process is stored in a separate data base called INFERENCE-MEMORY. Second, after an explanation of the conclusion has been requested, this part of memory is traversed to find those of the assumptions used which are on a communicatively appropriate level of detail. Finally, these assumptions are expressed in natural language.

An essential role in the first two of these phases is played by the meta-knowledge associated with each inference rule which is available to the system. As you can see from the two inference-rule definitions in Fig. 7,

```
META-KNOWLEDGE.
```

- Apply the control knowledge coded in TRACE-PROCEDURE-DEMON7
- Don't use instantiations of premises with a degree of belief less than 0.3
- The degree of uncertainty of this rule is 0.5

```
RULE.        If you want to show (X IS Y)
                 show that       (X ISA Z)
                          and    (Z IS Y)
```

```
META-KNOWLEDGE:
```

- Apply the control knowledge coded in TRACE-PROCEDURE-DEMON7
- Don't use instantiations of premises with a degree of belief less than 0.4
- The degree of uncertainty of this rule is 0 8

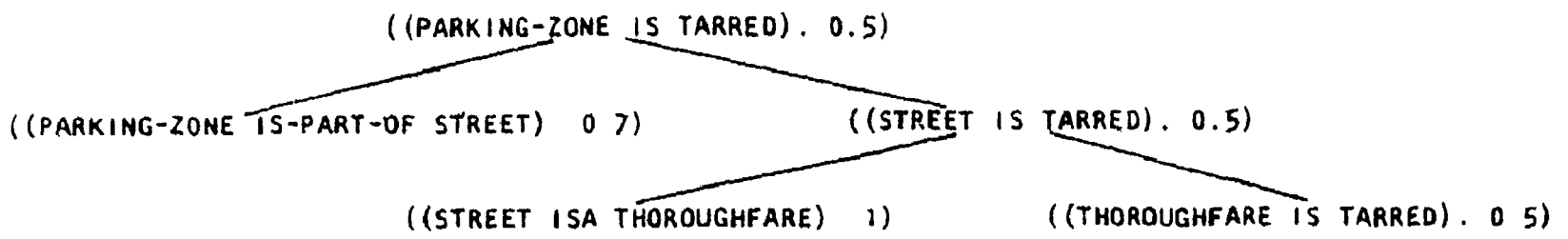
```
RULE.        If you want to show (X IS Y)
                 show that       (X IS-PART-OF Z)
                          and    (Z IS Y)
```

Fig. 7

she such piece of meta-knowledge concerns the degree of uncertainty associated with the rule. The most interesting piece of meta-knowledge in this situation is the specification of a particular FUZZY procedure demon. These demons enforce during the application of an inference rule global control regimes specified by the programmer (LeFaivre 1977). In particular, one of the things done by TRACE-PROCEDURE-DEMON7 is the storage of the reasoning steps in INFERENCE-MEMORY.

Suppose now that the assumptions at the top of Fig. 8 are represented in semantic networks. Applying the two rules in Fig. 7 to them, the system builds up the goal tree[1] in Fig. 8. The internal trace which is built up by the procedure demon is shown at the bottom of Fig. 8. Note that the

```
((PARKING-ZONE IS-PART-OF STREET)  0 7)
((STREET ISA THOROUGHFARE)  1)
((THOROUGHFARE IS TARRED)  0.7)
```




```
(((PARKING-ZONE IS TARRED) 5)((PARKING-ZONE IS-PART-OF STREET).7)((STREET IS TARRED) 5))  0.8)
(((STREET IS TARRED) 5)((STREET IS THOROUGFARE)1)((THOROUGHFARE IS TARRED) 5). 0 5)
```

Fig. 8

entries in the inference memory are ordered in terms of the degree of uncertainty of the executed inference procedures. This means that the most uncertain entries will be mentioned first in the explanation, and the most

[1] The conflict-resolution strategy which is used is one which favours more specific ones.

trivial ones probably not at all. This reflects our hypothesis that degree of uncertainty is the most important factor determining the relevance of a step in an inference chain, as far as justification of the conlusion is concerned

Our approach to explanation is distinguished from the previous efforts of Winograd in his SHRDLU system (Winograd 1972) and of the MYCIN group (Scott et al. 1977). In SHRDLU each theorem calls the functions MEMORY and MEMOREND explicitly, which manipulate the inference memory. We have improved over this by integrating the management of the inference memory into a higher process, which controls all reasoning processes. The structure of the inference rules themselves is therefore not obscured by the presence of trace commands. Our approach generalizes the corresponding features of MYCIN, in which the conceptual complexity of a rule is a logarithmic function of its certainty factor and the goal tree is traversed in steps whose size is specified by a numerical argument of the WHY command (Davis et al. 1977).

This approach is also related to recent research by Davis in his TEIRESIAS system (Davis 1977) and Sussman in his AMORD (De Kleer et al. 1977) and EL (Stallman & Sussman 1977) projects, in which general problems of an explicit control of reasoning are explored, in that it is based on an explicit representation of control and meta-knowledge, which typically is 'hidden away' in the interpreter and therefore is inaccessible to the inference system.

The explanation facility of HAM-RPM is far from being complete. Ultimately, the system must understand exactly what the dialogue partner failed to comprehend.

ACKNOWLEDGEMENT

This research is currently being supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

Davis, R. (1977): Generalized Procedure Calling and Content-Directed Invocation. *Proceedings of the Symposium on Artificial Intelligence and Programming Languages. SIGART Newsletter 64, 45-54*

Davis, R , Buchanan, B., Shortliffe, E. (1977): Production Rules as a Representation for a Knowledge-Based Consultation Program. *Artificial Intelligence 8, 15-45*

De Kleer, J., Doyle, J., Steele, G.L., Sussman, G.J. (1977): AMORD - Explicit Control of Reasoning. *Proceedings of the Symposium on Artificial Intelligence and Programming Languages. SIGART Newsletter 64, 116-125*

Grice, H.P. (1975): Logic and Conversation. *Cole, P., Morgan, J.L. (eds): Syntax and Semantics. Vol. 3: Speech Acts. N.Y.: Academic,41-58*

Grosz, B. (1976): Resolving Definite Noun Phrases. *Walker, D.E. (ed.): Speech Understanding Research. Stanford Research Institute, Chapter 9*

v. Hahn, W (1978a): Überlegungen zum kommunikativen Status und der Testbarkeit von natürlichsprachlichen Artificial-Intelligence-Systemen (Some Thoughts on the Communicative Status and the Testability of Natural Language AI-Systems). HAM-RPM Report No 4, April 1978. (also to appear in: Sprache und Datenverarbeitung)

v. Hahn, W (1978b): Probleme der Simulationstheorie und Fragepragmatik bei der Simulation natürlichsprachlicher Dialoge (Some Problems of Simulation-Theory and the Pragmatics of Questions in Connection with the Simulation of Natural Language Dialogues). HAM-RPM Report No. 6, May 1978 (also to appear in: Ueckert, H., Rhenius, D. (eds.): Komplexe menschliche Informationsverarbeitung. Beitrage zur Tagung 'Kognitive Psychologie' in Hamburg. Bern: Huber)

v. Hahn, W., Hoeppner, W., Jameson, A., Wahlster, W. (1978). HAM-RPM: Natural Dialogues with an Artificial Partner. *Proceedings of the AISB/GI Conference on Artificial Intelligence, Hamburg, 122-131*

Herrmann, T. & Laucht, M. (1976): On Multiple Verbal Codability of Objects. *Psychological Research, 38, 355-368*

LeFaivre, R.A. (1977): FUZZY Reference Manual. Rutgers University, Computer Science Dept., March 1977

Ritchie, G.D. (1977): Computer Modelling of English Grammar. University of Edinburgh, Computer Science Dept., Report CST-1-77

Scott, C.A., Clancey, A., Davis, R., Shortliffe, E.H. (1977): Explanation Capabilities of Production-Based Consultation Systems. *American Journal of Computational Linguistics, Microfiche 62*

Stallman, R.M. & Sussman, G.J. (1977): Forward Reasoning and Dependency-Directed Backtracking in a System for Computer-Aided Circuit Analysis. *Artificial Intelligence 9, 135-196*

Wahlster, W. (1978): Die Simulation vager Inferenzen auf unscharfem Wissen: Eine Anwendung der mehrwertigen Programmiersprache FUZZY (The Simulation of Approximate Reasoning Based on Fuzzy Knowledge: An Application of the Many Valued Programming Language FUZZY). HAM-RPM Report No. 5, May 1978 (also to appear in Ueckert, H., Rhenius, D. (eds.): Komplexe menschliche Informationsverarbeitung. Beiträge zur Tagung 'Kognitive Psychologie' in Hamburg. Bern: Huber)

Winograd, T. (1972): Understanding Natural Language. N.Y.: Academic

A CRITICAL LOOK AT A FORMAL MODEL FOR STRATIFICATIONAL LINGUISTICS

Alexander T. Borgida
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 1A7

SUMMARY

We present here a formalization of the stratificational model of linguistics proposed by Sampson [13] and investigate its generative power. In addition to uncovering a number of counter-intuitive properties, the results presented here bear on meta-theoretic claims found in the linguistic literature. For example, Postal [11] claimed that stratificational theory was equivalent to context-free phrase-structure grammar, and hence not worthy of further interest. We show, however, that Sampson's model, and several of its restricted versions, allow a far wider range of generative powers. In the cases where the model appears to be too powerful, we suggest possible alterations which may make it more acceptable.

1. Introduction

Linguistic theories are at least partially interested in presenting the regularities found in natural languages. Given the current dominance of the Transformational Generative (TG) school in the field of linguistics, it seems necessary for theories competing for attention to possess a formal model. In addition to the advantages normally derived from presenting results through a formalism, such as precision, succinctness and verifiability, one can also comment on the veracity of meta-theoretic claims. It was using such formal arguments that Chomsky and his collaborators demonstrated the inability of finite automata and of context-free grammars to describe all natural language constructs. Similarly, the formal work of Peters and Ritchie [8,9] was important in uncovering inadequacies of two notions of TG theory namely, the "recoverability of deletions condition" and the "universal base hypothesis".

Finally, since many generative linguists want grammatical theories which characterize natural languages, they fault any theory which is "too powerful" in the sense of being able to describe languages which clearly cannot be natural languages, such as non-recursive sets. Furthermore, computer scientists working on natural languages will have to give in the future more consideration to the work of linguists, especially on "exotic" languages, in order to be able to observe a wider range of phenomena. Such access will be facilitated if the formalisms in which the grammars

are presented lend themselves to computer implementation for purposes such as parsing, testing, etc. This entails, among other things, that linguists should avoid as much as possible features which make their grammars generate non-recursive sets, and hence it is one of the purposes of the present paper to point out such features and discuss possible ways of avoiding them.

In this paper we will discuss one model proposed for the stratificational theory of linguistics. This theory, advanced by S. Lamb, H.A. Gleason Jr. and their collaborators ([5],[6],[7]), advocates that languages be described in terms of several subsystems, known as strata. Each stratum has its own set of units and a tactics specifying the "correct" ("allowable") structures on that stratum. A specific grammar might for example have strata corresponding roughly to semantics, syntax-morphology and phonology, although this is by no means standard. Furthermore, the strata are linearly ordered as levels, and there is a realization relation which connects adjacent strata by attaching to every well-formed structure on one stratum, zero or more accompanying structures on the adjacent strata. Note therefore that a particular utterance has simultaneous expression on each stratum.

In this paper we examine the formal model for stratificational linguistics proposed by Sampson ([13]). This model uses rewrite grammars $G_1, G_2, \ldots$ to describe the tactics, while the realization relation is essentially a rewrite system R acting as a transducer between the languages of the tactics. More specifically,

realization connects adjacent tactics G_j and G_{j+1} by matching sentences u in the language generated by G_j with those sentences v in the language of G_{j+1} which can be derived from u by using rules from R. An important property of the linguistic realization relation is the fact that every structure on some stratum can have only a finite number of "realizates" on the next stratum. This means that the rewrite system R must be constrained so that it has no recursive symbols. Such a rewrite system will be called acyclic.

We investigate here the effect of acyclic rewrite systems acting as transducers on axiom sets, varying the type of the derivations and rules allowed.

We prove in this paper that regular languages are closed under transduction by acyclic rewrite systems, but that the linear context-free languages are mapped onto the recursively enumerable sets. This implies that stratificational grammars with non-selfembedding tactics would be too weak while those with even one context-free tactics would be too strong. If the realization derivation is restricted to be in some sense "left-most", then we show that the transduction can be performed by a finite state device known as an a transducer. Furthermore, if productions with null right-hand sides are not allowed in an acyclic rewrite system then all the derivations can be made left-most. This provides one possible method of restricting the generative power of acyclic rewrite systems.

By deriving a recursive characterization of the languages generated with n-strata in terms of (n-1)-stratal languages, we can show that if the realization is restricted to being leftmost, then the languages described are homomorphic images of the intersections of the languages generated by the tactics. In particular, this means that we can find natural families of stratificational grammars which generate for example the sets recognized in real time by nondeterministic multitape Turing machines. This result partially confirms a hitherto unproven claim by Sampson, and discredits Postal's [11] classiciation of stratificational grammars as just another variant of context-free phrase-structure grammars.

Finally, we investigate the use of ordered rules in linguistic grammars and prove that in several models they allow the generation of sets which are not even recursively enumerable a clearly unsatisfactory situation.

The remainder of the paper is structured as follows. In Section 2, we present the formal definitions and notation to be used, including the formal model for stratificational grammars. In Section 3, we examine the properties of "acyclic rewrite systems", which form the principal novel component in our definition of stratificational grammars. We then return in Section 4 to examine the generative power of stratificational grammars and relate the results to linguistics.

2. Definitions

We repeat here some important definitions from [12], and assume that the reader is familiar with the other basic notions of formal language theory.

A vocabulary V is a finite set of symbols, and we use V^+ to denote the set of all non-null strings consisting of symbols from V; using e to denote the null string, we also define V^* to be $V^+ \cup \{e\}$.

A rewrite system RW is a pair (V,R) where V is a vocabulary and R is a finite set of rules (productions) of the form $u \to v$, where $u \in V^+$ and $v \in V^*$; u is known as the left hand side of the production (lhs.) and v is its right hand side (rhs.).

A word $x \in V^+$ is said to directly derive or generate in R another word $y \in V^*$ (denoted by $x \Rightarrow_R y$) iff there exist words u,v,w,z such that x = wuz, y = wvz and $u \to v$ belongs to R. Let $\overset{+}{\Rightarrow}_R$ be the transitive closure of $\Rightarrow_R$, and $\overset{*}{\Rightarrow}_R$ its transitive reflexive closure. A sequence of words $w_1, w_2, \ldots, w_n$ such that $w_1 \Rightarrow_R w_2 \Rightarrow_R \ldots \Rightarrow_R w_n$ is said to be a (free) R-derivation (or simply a derivation) of w_n from w_1.

Given a rewrite system RW = (V,R) and a subset AX of V^*, the language generated by R from axiom set AX with free derivations is defined to be the set $\mathcal{L}(AX,RW) = \{w \mid u \in AX,\ u \overset{*}{\Rightarrow}_R w\}$.

Given the rewrite system RW = (V,R), define the dominance relation $<$ on $V \times V$ by: $d<b$ iff $xby \to udv$ is one of the productions in R (for some strings x,y,u,v) or if there exists some c in V such that $d<c$ and $c<b$. Then RW is defined to be

acyclic (abbreviated a.r.) iff the relation < is anti-symmetric and anti-reflexive.

If $u \to v$ is a production in a rewrite system, it will be called a null rule if v is the null string e, and it will be called context-free if $|u|$, the length of u, is 1.

A rewrite grammar G is a quadruple (N,T,S,P) where N and T are the sets of nonterminals and terminals respectively, S is a distinguished nonterminal and $\mathcal{G} = (N \cup T; P)$ is a rewrite system. In this case, if $u =*>_{\mathcal{G}} w$ then this is called a G-derivation, or a derivation in G, and the language generated by G, denoted by L(G), is defined to be the set $\{t \mid S =*> t \text{ in } G,\ t \in T^*\}$. We assume the reader is familiar with the terminology of type 0 (recursively enumerable or RE), type 1 (context sensitive), type 2 (context free) and type 3 (regular) languages, and corresponding families of grammars and automata. A type 2 grammar will be called linear if all its productions are of the form $A \to aBb$, where $A,B \in N$, $a,b \in T \cup \{e\}$, and will be called selfembedding if for some $A \in N$ there is a G-derivation $A =*> uAv$ where u and v are not null.

New languages can be obtained from old ones through such set operations as union, intersection and concatenation.

One can also define mappings over strings and then extend them to sets of strings in the obvious way. One such mapping is the substitution s which associates with every symbol b of some alphabet T, a set of words s(b) over another alphabet T'; defining s(xy) = s(x)s(y) and s(e) = e, a substitution can be extended to strings. If the sets s(b), are regular, finite or

e-free then s is said to be regular, finite or e-free respectively; if s(b) contains a single word then s is called a homomorphism, and the braces for sets are dropped. A homomorphism h can also be e-free, or it can be length-preserving, if $|h(b)| = 1$ for all symbols b. If $\mathcal{L}$ is a family of languages then we use $H(\mathcal{L})$ and $H^0(\mathcal{L})$ to represent the families of languages obtained from elements of $\mathcal{L}$ through e-free homomorphisms and homomorphisms respectively.

One final operation on strings is reversal defined by Rev(b) = b if $|b| < 2$ and Rev(xy) = Rev(y)Rev(x).

One can also use automata to perform mappings between strings. The a-transducer $M = (K, T_1, T_2, k^0, F, \tau)$ is an extension of the finite automaton, where T_1 and T_2 are the input and output alphabets, and τ is a finite subset of $K \times T_1^* \times T_2^* \times K$ (the transition set). The relation $\vdash$ is defined on $K \times T_1^* \times T_2^*$ by the rule $(k,uv,z) \vdash (k',v,zx)$ if $(k,u,v,k') \in \tau$. The output of M for input word w is one of the strings in the set $\{z \mid (k^0,w,e) \vdash^* (k,e,z),\ k \in F\}$. An a-transducer is said to be e-output free if for any (r,u,v,s) in τ, the string v cannot be null.

A collection of languages $\mathcal{A}$ is said to be closed under the operation σ if $\sigma(L) \in \mathcal{A}$ whenever $L \in \mathcal{A}$. A (full) trio is a family of languages containing at least one non-empty set, closed under e-free homomorphism (arbitrary homomorphism), inverse homomorphism, and intersection with regular languages.

Finally, omitting detailed justification (see [3]), the following formal definition captures the essential aspects of the notion of stratificational grammar, as presented by Sampson [13]:

Definition An n-stratal rewrite grammar (n-RSTRAT) is a 5-tuple RST = (n,TCT,RLZ,V_C,V_E), where V_C and V_E are the set of "content units" and "expression units" respectively, TCT = $(G_1,G_2,\ldots,G_n)$ is a vector of n rewrite grammars, and RLZ = $(R_0,R_1,\ldots,R_n)$ is a vector of $n+1$ acyclic rewrite systems. The transduction performed by such a grammar will be defined by T-RSTRAT(RST) = $\{(u,v) \mid w_0=u \in V_C^+$, $w_{n+1}=v \in V_E^*$, there exist $w_j \in L(G_j)$ such that $w_j \Rightarrow^* w_{j+1}$ via R_j-derivations for $j = 0,1,\ldots,n\}$. Its language is described by L-RSTRAT(RST) = $\{v \mid (u,v) \in$ T-RSTRAT(RST)$\}$.

In this formal model, the grammar is thought of as transducing "meaning" into "sound" in the following manner: starting with a string of "content units" (expressing the meaning of an utterance), the realization rewrite rules are repeatedly applied until a string of "expression units" is obtained. The realization derivation is constrained by the requirement that for each tactics there exists an intermediate stage in the realization derivation which conforms to the tactics specifications (i.e. belongs to the language generated by the tactics). The above formalism is based mainly on Lamb's version of stratificational linguistics; an alternate approach, closer in spirit to Gleason's model, is presented in [3].

3. Generative power of acyclic rewrite systems

To begin with, we remark that the formal definition of stratificational grammars in [13] allows in the realization system rewrite rules with null left-hand sides (i.e. rules of the form e → u). Unfortunately, such rules could be applied to some string an arbitrary number of times. In our stratificational model, this would result in any string having an infinite number of realizates. Furthermore, rules of the form e → u can also be used to establish context-free dependencies in strings generated even from singleton axiom sets. For example, if R = ({c,d},{e→cd}) then $\mathcal{L}(\{e\},R) = \{w | w \in \{c,d\}^*$, w has the same number of "c" and "d" symbols}, which is known to be a non-regular context-free language. The phenomena described above do not appear to have linguistic equivalents, and run counter to the stratificational philosophy which envisages only finitely many realizations for any structure.

As it turns out, in practice rules of the form e → u are only required to introduce in the realization derivation syntactically determined elements, such as "do" in questions. Such insertions need however be performed only once, at the end of every realization derivation between two tactics. Therefore they can be accomplished through normal acyclic rules if each e → u in R is replaced by rules v → uw and v → wu for all v → w in R. For this reason, we will continue to use the definition of rewrite systems which only allows productions with non-null left-hand sides.

We next investigate the effect of a.r. on simple types of axiom sets.

Theorem 3.1 Let AX be a regular set over alphabet T and let E be some alphabet disjoint from T. If RW = (V,R) is an a.r. then $\mathcal{L}(AX,RW) \cap E^*$ is also a regular set.

Proof Let G = (N,T,S,P) be a type 3 grammar generating AX, and without loss of generality assume that $N_G \cap V_R$ is empty. Furthermore, normalize R so that all its rules are of the form a→bc, a→e or bc→d. This can be accomplished in a 3-step process: first, replace rules of the form u → abv ($a,b \in V$, $u,v \in V^*$) by rules u → aa, $\bar{a}$ → bv where $\bar{a}$ is a new symbol; repeat this until all rules have rhs. no longer than two symbols. Next, replace rules of the form abu → v by ab → $\bar{a}$, $\bar{a}$u → v, until all lhs. of rules are at most two symbols. Finally, eliminate rules of the form a → b by adding to R a rule y → zbz' whenever y → zaz' is in R.

Our goal is to produce a type 3 grammar such that R-derivations are "precomputed" in its productions. For example, if the grammar G originally had productions X→aY and Y→bZ, while R contained the rule ab→d, then the final grammar would contain production X→dZ.

For this purpose, consider the following iterative construction:

INITIALIZATION: Let G_1 be G; let $T' = T \cup V_R$.

CONSTRUCTION 1: For every integer i, given grammar $G_i = (N_i,T_i,S_G,P_i)$, construct from it a type 3 grammar $G_{i+1} = (N_{i+1},T',S_G,P_{i+1})$ as follows:

1. for every $a \in T_i$, let P(i,a) be the set of all productions in G_i which have the symbol "a" on the rhs.;
2. to begin with, let P_{i+1} contain P_i, and N_{i+1} contain N_i;
3. IF b→cd is a production in R, THEN for every A→bB in P(i,b), ADD to N_{i+1} a nonterminal [A;B;b→cd], and ADD to P_{i+1} productions A→c[A;B;b→cd] and [A;B;b→cd]→dB;
4. IF b→e is in R, THEN for every A→bB in P(i,b) ADD production A→B to P_{i+1};
5. IF bc→d is in R, THEN for every pair of productions A→bB in P(i,b), and C→cD in P(i,c), ADD to P_{i+1} the new production A→dD <u>if $B \Rightarrow^* C$ in G_i</u>;

END;

Suppose that we were able to establish that

$$\mathcal{L}(L(G),R) = \bigcup_{i=1}^{\infty} L(G_i) \qquad (1)$$

From the construction it is easy to see that P_i is always a subset of P_{i+1} (and hence $L(G_i) \subset L(G_{i+1})$), and if

$$G_m = G_{m+1} \text{ for some index } m \qquad (2)$$

(i.e. no new productions are added to G_m in Construction 1), then G_j would be equal to G_m for every $j > m$.

But, if such an m exists then $\mathcal{L}(L(G);R) = \bigcup_{i=1} L(G_i) = L(G_m)$ and G_m is the type 3 grammar we are looking for. Therefore, it remains to establish equalities (1) and (2).

To prove (1), we first define a new type of derivation ("single, left-right pass") relation "$=o>_R$" as follows:

$u =o>_R v$ iff there exists integer n such that for j = 1,...,n $x_j \to y_j$ is a rule in R and z_j is some string with the property that $u = z_0x_1z_1\ldots x_nz_n$ and $v = z_0y_1z_1\ldots y_nz_n$ (if n = 0, then u = v).

We then claim that

$$L(G_{i+1}) = \{w \mid \exists\ v \in L(G_i) \text{ such that } v =o>_R w\} \qquad (3)$$

This equality can be demonstrated by straightforward inductions on, respectively, the lengths of derivations in G_{i+1}, and the integer n appearing in the definition of $=o>_R$. In both cases the important points are that if $A \Rightarrow bB$ in G_i ($A,B \in N_i$, $b \in T_i$) then either $A \Rightarrow bB$ in G_{i+1} (by step 2 in Construction 1) or $A \Rightarrow uB$ if $b \to u$ is in R (by steps 3 or 4); and if $A \Rightarrow bB =^*> bC \Rightarrow bcD$ in G_1 then $A \Rightarrow dD$ in G_{i+1} in case $bc \to d$ is in R (step 5).

We are now in a position to prove (1). First, suppose that w belongs to $\mathcal{L}(L(G),R)$ and w was obtained from $u \in L(G) = L(G_1)$ in an R-derivation with n steps: $u = u_1 \Rightarrow_R u_2 \Rightarrow_R \ldots \Rightarrow_R u_n \Rightarrow w$. If we note that for any strings x,y $x \Rightarrow_R y$ implies $x =o>_R y$ then by (3) we have for i = 1,...,n that $u_i \in L(G_i)$. But then $w = u_n$ must belong to $L(G_n)$, and hence to $\bigcup_{i=1}^{\infty} L(G_i)$. Conversely, if w belongs to $\bigcup_{1=1}^{\infty} L(G_i)$ then there must exist an index m such that $w \in L(G_m)$. Using (3) it is then trivial to prove by induction on m that there exists $v \in L(G)$ such that $v = v_1 =o>_R v_2 =o>_R \ldots =o>_R v_m = w$ for some $v_i \in L(G_i)$ (1 = 1,...,m). But in that case $w \in \mathcal{L}(L(G),R)$ because by definition $x =o>_R y$ implies that $x =^*>_R y$ for any strings x,y. This concludes the proof of identity (1).

To prove (2), one might try to demonstrate that the construction halts after some precomputable number of steps. This approach unfortunately runs into the following problem: the addition of a new production to G_i in step 4, allows new pairs of variables B' and C' to be connected by a derivation $B' \Rightarrow^* C'$; this may allow new production $A' \rightarrow dD'$ to be added to G_{i+1} in step 5, which in turn may eventually allow step 4 to add a new rule to G_j for some $j>i$, etc.

The above compels us to look for an alternative proof of (2): exhibiting a grammar G^o such that every G_i is a subgrammar of G^o. This would mean that the increasing sequence of grammars $G_1, G_2, \ldots$ is bounded above, and hence converges to one of its elements.

To construct G^o, remember that by definition of R there is an anti-symmetric relation $<$ on V_R. Using this, we assign to every symbol in V and every production in R a unique index number, according to the following algorithm:

INDEXING ALGORITHM:

1. I(b) := 0 for every $b \in V$ such that there is no $d \in V$ and $d>b$;
2. FOR i=0 to |V| DO WHILE not all symbols have an index;
   IF I(b)=i & b→cd is in R, THEN I(c) := I(d) := i+1 and
   I(b→cd) := i+1;
   IF I(b)=i & I(c)≤i & bc→d is in R THEN I(d) := i+1 and
   I(bc→d) := i+1;

IF I(b) = i & I(c)≤i & cb→d is in R THEN I(d) := i+1 and

I(bc→d) := i+1;

IF I(b) = i and b→e is in R THEN I(b→e) := i+1;

END

END

By the acyclicity of R, the above algorithm produces a unique value for every symbol and production. Suppose the highest index value assigned is n. Then G^O will be constructed from G by repeated modification in n passes through the following:

CONSTRUCTION 2: Let $G^O = (N^O, T', S, P^O)$ be G initially;

FOR i=1 to n DO

* in the i-th pass, add to G^O all possible productions representing derivations by index i rules */

1. For every symbol 'd' in V_R such that I(d) = i, let P(d) be the set of all productions currently in G^O, with 'd' on the rhs.;
2. IF b→dc (or b→e) is in R and has index i (iff I(b)=i), THEN alter G^O in exactly the same way as in steps 3 (or 4) of CONSTRUCTION 1; (except that P^O and N^O are used instead of P_{i+1} and N_{i+1}).
3. IF bc→d is in R and has index i, THEN for every pair of productions A→bB in P(b) and C→cD in P(c), ADD to P^O the new production A→dD (<u>whether or not B=>*C in G^O</u>).

Note that in the i-th pass the only productions added to G^o have on the rhs: a terminal symbol of index strictly greater than i. Therefore, in successive passes through the loop after the i-th one, P(d) remains unchanged for all symbols "d" with $L(d) \leq i$. Furthermore, the output G^o remains unchanged if passed through CONSTRUCTION 2 a second time.

Secondly, note that if some grammar K remains unchanged by CONSTRUCTION 2 then it does so through CONSTRUCTION 1 as well, because every rule in R is eventually considered in CONSTRUCTION 2, and in each case at least those productions which would have been added by CONSTRUCTION 1 are added by CONSTRUCTION 2.

Therefore, since G is a subgrammar of G^o, G_i will be so for every i greater than 1, and the proof is completed. □

Increasing the range of sets from which we choose the axiom sets, we obtain the following:

Theorem 3.2 Let $G = (N,T,S_G,P_G)$ be an arbitrary type 0 grammar. Then there exists a linear context-free language LIN_G, and an a.r. R_0 (which is dependent only on the set $N \cup T$), such that

$$L(G) = \mathcal{L}(LIN_G,R_0) \cap T^*$$

Proof (Notational convention: let $V = N \cup T$, and if $\epsilon\{-,\sim,\vee\}$ then use V to represent the set $\{\dot{a} | a\epsilon V\}$, and $\dot{w} = \dot{a}_1 \ldots \dot{a}_j$ if $w = a_1 \ldots a_j$.)

It is known ([1]) that there exist two linear context-free languages L_1 and L_2, as well as a homomorphism h, such that $L(G) = h(L_1 \cap L_2)$. We have constructed ([3]) pairs of new such languages:

$$L_1 = (\{\bar{\%}\bar{S}_G\bar{\%}\$\tilde{\%}\} \cup \{\bar{\%}\bar{v}_1 \ldots \bar{\%}\bar{v}_{m-1}\bar{\%}\bar{S}_G\bar{\%}\$\tilde{\%}\tilde{v}_m \ldots \tilde{\%}\tilde{v}_{2m-2}\tilde{\%}\check{z} \mid m>0,\ z\epsilon T^*,$$
$$v_i \epsilon V_G^+, \mathrm{Rev}(v_{i-1}) = v_{m+i-1} \text{ for } i = 1,\ldots,m-1\}$$

and

$$L_2 = \{\bar{\%}\bar{w}_1\bar{\%} \ldots \bar{w}_n\bar{\%}\$\tilde{\%}\tilde{w}_{n+1}\tilde{\%} \ldots \check{w}_{2n} \mid n>0,\ w_i \epsilon V^+ \text{ for } i<n,$$
$$\text{and } \mathrm{Rev}(w_{n-i}) \Rightarrow_G w_{n+i+1} \text{ for } i = 0,\ldots,n-1\}$$

In this case, the homomorphism h is defined as $h(\check{x}) = x$ if $x\epsilon T$, null otherwise. Observe that L_1 is dependent solely on the vocabulary V and it only checks whether the strings around the central '$\bar{\%}\bar{S}_G\bar{\%}\$\tilde{\%}$' are mirror images of each other. But the following rewriting system does exactly the same job, and, in addition, performs the homomorphism h:

$$R_0 = \{\bar{\%}\bar{S}_G\bar{\%}\$\tilde{\%} \to e,\ \bar{\%}\tilde{\%} \to e,\ \bar{x}\tilde{x} \to e \text{ for all } x\epsilon V_G,\ \check{a} \to a \text{ for } a\epsilon T\}$$

Then $T^* \cap \mathcal{L}(L_2, R_0) = h(L_1 \cap L_2) = L(G)$ and by observation it is clear that R^0 is acyclic. □

This result is surprising, especially from a linguistic point of view, and demonstrates the power of acyclic rewrite systems. Since it is undesirable that linguistic mechanisms be so powerful, we will attempt to put bounds on them. One way to do so is to restrict the places where steps in derivations can occur.

Essentially, in a k-leftmost derivation there is a k-symbol wide "window" on the derivational forms where rewriting can occur, and this window is only allowed to move to the right.

Definition 3.1 Let $w_0 \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_n$ be an R-derivation, where for $i = 1,\ldots,n$ productions $u_i \rightarrow v_i$ are used to obtain $w_i = x_i v_i u_i$ from $w_{i-1} \triangleq x_i u_i y_i$ $(x_i, y_i, w_i \in V^*)$. For any integer k, this is said to be a k-leftmost derivation if for all $i = 1,\ldots,n-1$ there exist s_i and t_i in V^* such that $x_i = t_i s_i$ with $|s_i| \leq k$ and $|t_i| \leq |t_{i+1}|$. This definition of R-derivations gives rise to the new language

$$\mathcal{L}(AX,RW,k\text{-ld}) = \{w \mid x \in AX,\ x \Rightarrow_R^* w \text{ in k-leftmost derivation}\}.$$

Theorem 3.3 Given an a.r. RW = (V,R), there exists an a-transducer Θ_R such that $\mathcal{L}(AX,R,k\text{-ld}) = \hat{\Theta}_R(AX)$.

Proof By the acyclic nature of the rules in R, any string of length k can be rewritten into a string of length at most kd^t where d is the length of the longest rhs. of a rule in R, and t is the number of symbols in V. Therefore, if we define a Turing machine transducer Θ_R which simulates on its working tape k-leftmost R-derivations, then it need have only a bounded, finite-length tape. But this can obviously be kept in a finite memory, and hence Θ_R can be made into an a-transducer. □

Therefore, leftmost constraints on R-derivations lead to a much more restricted version of rewrite systems because all trios (in particular LINEAR-CFL) are closed under a-transduction.

A second method of bounding the power of a.r. is to restrict the form of the productions allowed in R.

Theorem 3.4 If RW = (V,R) is an a.r. with no null productions, then for every axiom set AX, $\mathcal{L}(AX,RW) = \mathcal{L}(AX,RW,k\text{-ld})$ for some constant k.

Proof Suppose R has r rules involving t symbols and let c be the length of the longest lhs. of a production. Now observe that if $v \Rightarrow_R^* w$, then no symbol in w can have more than $k=c^t$ ancestors in v, all of which must be adjacent in v (i.e. the presence of a symbol in w can depend only on the presence of at most k adjacent symbols in v). This value of k can be obtained as follows: since the rules are acyclic, new symbols must appear after every application of a production and hence every symbol in w can be the result of applying at most t productions; since each of these uses at most c symbols as context, we get the value of c^t. The adjacency requirement comes from the condition that there be no e-rules in R. Consider now some R-derivation $tXv \Rightarrow^* t\tilde{t}$ $(t,\tilde{t},v\in V^*, X\in V)$, where no symbol in t is rewritten, but X is. We will prove by induction on the length of the derivation that there is an equivalent k-leftmost derivation.

Basis. If the derivation has 0 or 1 steps then it is clearly k-leftmost.

Induction step. Break up the derivation into steps, to see where X is rewritten:

$$tXv \underset{①}{\Rightarrow^*} tXw\tilde{v} \underset{②}{\Rightarrow} tYzv \underset{③}{\Rightarrow^*} t\tilde{t}$$

where we used rule $Xw \rightarrow Yz$ in step ②.

Now find the last production in ① which produces only non-ancestors of Y.

(a) If there is no such production, then by our opening remarks ① must be k-leftmost, and hence ③ can be made k-leftmost by induction.

(b) Otherwise, suppose that the last such rule was $\alpha \to \rho$. Then we claim that the derivation in further detail is

$$tXv \underset{④}{=^*>} tXu_1\alpha u_2 \underset{⑤}{=>} tXu_1\rho u_2 \underset{⑥}{=^*>} tXw\tilde{u}_1\rho u_2 \underset{②}{=>} tYzu_1\rho u_2 \underset{③}{=^*>} t\tilde{t}$$

The significant part of this claim is that no production in ⑥ affects the string ρu_2, and this is true by our choice of $\alpha \to \rho$ as the last production generating non-ancestors of Y, hence of Yz, and the necessary contiguity of ancestors. But now note that step ⑤ can be postponed to yield the following reordered derivation:

$$tXv \underset{④}{=^*>} tXu_1\alpha u_2 \underset{⑥}{=^*>} tXw\tilde{u}_1\alpha u_2 \underset{②}{=>} tYz\tilde{u}_1\alpha u_2 \underset{⑤}{=>} tYz\tilde{u}_1\rho u_2 \underset{③}{=^*>} t\tilde{t}$$

By repeating the construction in part (b) on ① = ④⑥ this time (instead of ④⑤⑥) we will eventually (by a second induction, if desired) achieve case (a), and thus complete the proof. #

Note that in the above proof we had only excluded the use of null rules (the symbol ρ could not be null) so that other productions with left-hand sides longer than right-hand sides are still allowed in R.

Finally, we include for completeness the following result whose proof is trivial.

Proposition 3.5 Let RW = (V,R) be an a.r. which has only context free rules. Then there exists a finite substitution s_R such that for every axiom set AX, $\mathcal{L}(AX,RW) = s_R(AX)$. #

4. Stratificational Grammars

We now return to the notion of stratificational grammar which led us originally to consider acyclic rewrite systems. To begin with, note that the original definition of n-RSTRAT grammar has no constraint on the derivations occuring on the tactics, while in practice linguists appear to view the derivations as being leftmost (i.e. the leftmost nonterminal is the one rewritten). Therefore, throughout the following discussion we will examine the differences arising out of this variation.

First, we present a recursive characterization of the n-RSTRAT languages. For this purpose, define the language generated by a 0-RSTRAT grammar $RST^0 = (0,(),(R^0),V_C,V_E)$ as L-RSTRAT(RST^0) $\mathcal{L}(V_C^*,R^0) \cap V_E^*$. Then the following theorem is an obvious consequence of the definition of L-RSTRAT:

Theorem 4.1 If $RST = (n,(G_1,\ldots,G_n),(R_0,\ldots,R_n),V_C,V_E)$ is an n-RSTRAT grammar, and TOP(RST) is the (n-1)-RSTRAT grammar $(n-1,(G_1,\ldots,G_{n-1}),(R_0,\ldots,R_{n-1}),V_C,T_n)$, then L-RSTRAT(RST) = $\mathcal{L}(\text{L-RSTRAT(TOP(RST))} \cap L(G_n),R_n) \cap V_E^*$.

Using Theorems 4.1, 3.1 and the known closure properties of the regular languages, it is easy to see that if all the tactics $G_1,\ldots,G_n$ of an n-RSTRAT grammar are non-selfembedding then the stratificational grammar can generate only a regular language.

On the other hand, as soon as one of the tactics is allowed to be of type 2 and selfembedding, then by Theorems 4.1 and 3.2

the RSTRAT grammar can generate an arbitrary RE set. Even more surprisingly, this can be accomplished using a "universal realization relation", meaning that to obtain any RE set we need only vary the tactics, not the realization rewrite system. This situation is similar to that found for TG in [9], where the transformational component can be varied while the base grammar is kept fixed.

Therefore, in this stratificational model there seems to be no alternative between the insufficient descriptive power of finite automata and the excessive power of arbitrary Turing machines. These results hold even if the derivations on the tactics are constrained to be leftmost. We must therefore search for further limitations on the realization process. In section 3 we considered several possible ways of doing this, namely eliminating null or context-dependent rules, or making the realization derivation leftmost. In linguistic grammars there is a clear need for context-dependent realization rules, hence these cannot be eliminated. Although in Sampson's model null realization rules appear to be needed (more on this below), it is possible to envisage alternative models which avoid them. By Theorem 3.4, the absence of null rules is equivalent to restricting the realization derivation to being k-leftmost. Furthermore, based on current linguistic literature there appears to be no objection to limiting the realization to being k-leftmost. Therefore, we will examine the generative power of n-RSTRAT grammars under this constraint.

Theorem 4.2 If STR = $(n,(G_1,\ldots,R_n),(R_0,\ldots,R_n),V_C,V_E)$ is an n-RSTRAT grammar with realization derivations restricted to be k-leftmost for some integer k, then there exist homomorphism h and languages $L_1,\ldots,L_n$ such that for $i = 1,\ldots,n$ $L(G_i)$ is of the same type[1] as L_i, and L-RSTRAT(STR) = $h(L_1 \cap \ldots \cap L_n)$.

Proof The proof is based on a number of results about trios, which we summarize here from [4]:

(a) For $i = 1,\ldots,n$ the families of languages of the same type as $L(G_i)$ are trios.

(b) If L is a trio then $H(L)$ is a trio and $H^o(L)$ is a full trio.

(c) If $L_1,\ldots,L_n$ are trios then $H(H(L_1 \cap \ldots \cap L_{n-1}) \cap L_n)$ is a trio and it is equal to $H(L_1 \cap \ldots \cap L_{n-1} \cap L_n)$; similarly $H^o(H(L_1 \cap \ldots \cap L_{n-1}) \cap L_n) = H^o(L_1 \cap \ldots \cap L_{n-1} \cap L_n)$ is a full trio.

(d) trios are closed under intersection with regular sets and e-output bounded a-transductions, while full trios are also closed under arbitrary a-transduction.

We now prove the theorem by induction on n.

Basis. For n=1, by Theorems 4.1 and 3.3 there exist a transducers Θ_1 and Θ_0 such that L-RSTRAT(RST) = $\Theta_L(\Theta_0(V_C^*) \cap L(G_1)) \cap V_E^*$; then our theorem holds by notes (a) and (d) above with h being the identity map.

Induction step. For the case n+1, by Theorems 4.1 and 3.3 there exists a-transducer Θ_{n+1} such that L-RSTRAT(RST) is equal to

$$\Theta_{n+1}(\text{L-RSTRAT}(\text{TOP}(\text{RST})) \cap L(G_{n+1})) \cap V_E^*. \qquad (4)$$

---

[1] Meaning type 3, type 2, type 1, type 0, linear language.

But by induction, there exist homomorphism h'' and languages $L''_1,\ldots,L''_n$ such that L-RSTRAT(TOP(RST)) = $h''(L''_1 \cap \ldots \cap L''_n)$. Substituting this in (4) and applying notes (a), (b) and (c) we find a homomorphism h and languages $L_1,\ldots,L_{n+1}$ of the same type as $L''_1,\ldots,L''_n$ and $L(G_{n+1})$ such that L RSTRAT(RST) = $h(L_1 \cap \ldots \cap L_{n+1})$. #

Remark that by Theorems 3.4 and 3.5 the same result holds in the case when the realizations do not contain null rules, and by examining the above proof it can be seen that the homomorphism h can be restricted to being e free in this case.

The following converse to Theorem 4.2 can be easily established:

Theorem 4.3 Given homomorphism h from T to T', and rewrite grammars $G_1,\ldots,G_n$ with terminal alphabets T, then for i = 0,...,n there exist context-free acyclic rewrite systems R_i such that L-RSTRAT((n,$(G_1,\ldots,G_n)$,$(R_0,\ldots,R_n)$,T,T') = $h(L(G_1)\cap\ldots\cap L(G_n))$.

Proof For j<n, define R_j to be $\{a\rightarrow a \mid a\in T\}$, by the definition of RSTRAT-derivations this will simulate the intersection of the languages generated by the tactics. Finally, define R_n to be $\{a\rightarrow h(a) \mid a\in T\}$, thus performing the homomorphism on the intersection. #

To begin with, the above theorems partially confirm Sampson's hitherto unproven claim ([13: page 11]) that stratificational languages are the result of intersecting the languages of the tactics. Note however two important qualifications to this claim:

the realization derivation must be k-leftmost and a homomorphism must be applied to the intersection of the languages.

Theorems 4.2 and 4.3 show that with k-leftmost realization derivations, the type i languages (i = 1,2) can be obtained by using a type i grammar on one of the tactics, and making the other ones non-selfembedding. If all the tactics generate context-free languages (as in the case when tactic-derivations are leftmost) then n-RSTRAT grammar can generate the homomorphic intersections of the CFLs. For n≥2, this is known to equal the RE sets if null realizations are allowed; if null realization rules are not allowed then for n≥3 the n-RSTRAT grammars generate the family QUASI of sets recognized by nondeterministic Turing machines in real or linear time ([2]). These observations demonstrate that n-RSTRAT grammars can be appropriately modified so that they generate various language families intermediate between the regular and RE sets. Unfortunately, even when the realization derivation is restricted to being k-leftmost, 1-RSTRAT grammars with context-sensitive tactics and 2-RSTRAT grammars with context-free tactics can generate the RE sets, unless null realizations are restricted. The basic problem with restricting null rules lies in the pronounced bias of this model towards the realization of terminal units from one tactics to the next. In practice, in order to describe linguistic phenomena it is necessary to have information about the entire derivation process on some tactics. Sampson accomplishes this by introducing "pseudo-

terminals" into strings; for example, if the application of production x→y is to be noted for later use, then either rule x→py or x→yp would be used in the tactics to introduce p as a marker of the occurrence of x→y. The chief drawback of this approach is that the "pseudo-terminals" such as p must eventually be deleted, making null rules necessary. One possible solution may be to discover some bound on the number of null rule applications needed, resembling the "cycling function" proposed by Peters and Ritchie ([9]). Another solution is to consider a new formal model which allows realization to access uniformly all parts of the derivations on tactics; this approach is considered in [3].

Before concluding, we take a brief look at the problems raised by one addition to the basic model discussed so far, namely ordered rules. It has often been found useful in linguistic descriptions to use rules of the form "A→u if some condition C holds, otherwise A→v"; basically, these types of rules avoid stating the negation of condition C, which may be cumbersome. In certain stratificational descriptions, this has lapsed into the use of rules of the form "A→u if this can lead to a completed derivation, otherwise A→v". This notion is formalized by Sampson through the assignment of "weights" or "preference values" to certain rules. Thus A→u may be given value 1 while A→v receives value 0, and these values are accumulated throughout the derivation. At the end, only those expression strings resulting from some content string are taken

which have derivations with maximal preference values. The fundamental problem with this use of "ordered rules" is that even in context-sensitive grammars it is in general recursively undecidable whether a certain derivation can be successfully completed or not. In fact, we show that using "ordered rules" we can generate even non-recursively enumerable sets, an obviously undsirable situation.

Theorem 4.4 There exists a context-sensitive grammar G with one "ordered rule" which generates a non-RE language.

Proof The proof rests on the well known result that there exists an RE language L^o over some alphabet T, whose complement is not RE, and that there is a type 1 grammar $G^o = (N^o, T^o, S^o, P^o)$, where $T^o = T \cup \{b,\#\}$, such that $L(G^o) = \{w\#b^{i(w)} \mid w \in L^o$, $i(w)$ is some integer, depending on w $\}$ ([12]). Consider the grammar $G' = (N', T', S', P')$ where $N' = N^o \cup T^o \cup \{Y, S', Z\}$, $T' = \tilde{T} \cup T$, P' contains P^o and additional productions as described below. The grammar G' behaves informally as follows:

(a) from S', we generate some string wS' such that $w \in T^*$, using productions from $\{S' \to aS' \mid a \in T\}$;

(b) then we apply the ordered rule " $S' \to YS^o$ with weight 1, $S' \to Z$ with weight 0"; the plan is that the new nonterminal Y can be rewritten into a terminal, #, if and only if Y appears in a string belonging to $\{wYw\#\}\{b\}^*$ (i.e. iff rules of G^o can be used to rewrite S^o into some $w\#b^j$, where w is the same as the guess made in (a)). Once some derivation from S^o is completed, it is

clear that context-sensitive rules can be used to check out the above condition for Y. In addition, the same rules can place "bars" over all the symbols thus checked, resulting, if successful, in a sentence of the form $\bar{w}\#\bar{w}\#b^j$.

(c) Z on the other hand simply travels across the string w and places "dots" on top of every symbol, using rules from $\{sZ \rightarrow Z\dot{s} | s \in T\}$

The result will be that $L(G') = \{\bar{w}\#\bar{w}\#b^j | w \in L^o\} \cup \{\dot{w} | w \notin L^o\}$. Suppose that L(G') is RE, and let h be the homomorphism which deletes all symbols not in T, and removes the dots from the others. Then h(L(G')) is also RE because the RE sets are closed under homomorphism; but h(L(G')) is the complement of L^o, and thus not in RE by our choice of L^o. Therefore by contradiction, L(G') is not RE.

A similar proof can be given for stratificational grammars with two or more context-free tactics. These results draw attention to the need to redefine the notion of "ordered rule" in stratificational usage, and point out that care must be taken whenever formalizing aspects of linguistic practice.

In conclusion, our investigation of the formal properties of the stratificational model proposed by Sampson revealed certain unintuitive properties which make it less desirable as a tool for natural language description. Thus, the use of realization rules with null left hand sides was shown to allow unbounded number of

realizations for certain strings. More significantly, we showed that n-RSTRAT grammars with even one tactics allowing self-embedding could generate all RE sets. Since there are well known problems raised by this possibility, most significant being the inability to decide grammaticality, we identified a linguistically acceptable restriction on the realization, namely k-leftmost derivations, which led to improvements in some situations. Under this additional constraint, classes of n-RSTRAT grammars were shown to variously generate the context-free languages, the Quasi-realtime languages and the context-sensitive languages. Unfortunately, even in this case n-RSTRAT grammars could generate non-recursive sets, unless null realizations were restricted, and we discussed the problems inherent in this approach. Finally, we examined the definition of "ordered rules" used in some stratificational grammars, and formalized by Sampson, showing that it allowed the generation of even non-RE sets with type 1 tactics.

The above formal results about the generative power of stratificational grammars hopefully answer the requests of critics such as Pit'ha ([10]), and demonstrate the inaccuracy of Postal's classification of stratificational grammars as simply variants of context-free phrase structure grammars. The results also indicate some of the problem areas in this formal model for stratificational linguistics. We emphasize though that the problems are specific to this particular formalism, and should not be taken as condemnations of stratificational linguistics in general, since there are other stratificational models which avoid these pitfalls.

Acknowledgements

I would like to thank Professor Ray Perrault for his much appreciated advice and help both during my graduate student career and after it. I am also grateful to Teresa Miao for typing this paper and to Peter Schneider for proof reading it.

References

[1] Baker, B. and R. Book (1974). "Reversal Bounded Multi-pushdown Machines", J. Computer Systems Science - 8, 1974, 315-332.

[2] Book, R.V. and S.A. Greibach (1970). "Quasi-realtime Languages", Math. Systems Theory - 4, 1970, 97-111.

[3] Borgida, A.T. (1977). "Formal Studies of Stratificational Grammars", Ph.D. Dissertation, University of Toronto, also Technical Report No.112.

[4] Ginsburg, S. (1975). Algebraic and Automata-Theoretic Properties of Formal Languages, North-Holland Publishing Co.

[5] Gleason, H.A. Jr. (1964). "The organization of language: a stratificational view", Monograph Series on Language and Linguistics - 17, p.75-95, Georgetown University.

[6] Lamb, S. (1966). Outline of Stratificational Grammar, Georgetown University Press, Washington.

[7] Lockwood, D.G. (1972). Introduction to Stratificational Linguistics, Harcourt Brace Jovanovich, Inc.

[8] Peters, P.S. and R.W. Ritchie (1971). "On restricting the base component of Transformational Grammars", Information and Control - 18.

[9] Peters, P.S. and R.W. Ritchie (1973). "On the generative power of Transformational Grammars" Information Sciences - 6, 49-83.

[10] Pit'ha, P. (1974). "On a new form of Lamb's stratificational grammar", Slovo a Slovesnost - 35, p.208-218; translated from the original Czech by D.G. Lockwood.

[11] Postal, P. (1964). Constituent Structure: A Study of Contemporary Models of Syntactic Description, Indiana University, Bloomington, Ind. First appeared in Int. J. Amer. Linguistics - 30.1, part 3.

[12] Salomaa, A. (1973). Formal Languages, Academic Press, New York.

[13] Sampson, G. (1970). Stratificational grammar: a Definition and an Example, Janua Linguarum, Series Minor: 88, The Hague Mouton.