CATALOGED BY ASTIA
AS AD No. 403518

403 518

NSF-63-5

CURRENT RESEARCH AND DEVELOPMENT in SCIENTIFIC DOCUMENTATION No. 11

1963

National Science Foundation
Office of Science Information Service

Use of funds for printing this publication approved by the Director of the Bureau of the Budget, February 23, 1961

For sale by the Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C. — Price $4.00.

CONTENTS

		Page
Introduction		xvii
Checklist		xxi
1.	INFORMATION NEEDS AND USES	
1.1	**Advanced Information Systems, Inc**	7
1.2	**American Chemical Society**	8
1.3	American Institute of Biological Sciences	10
1.4	American Institute of Chemical Engineers	11
1.5	American Institute of Physics	12
1.6	American Mathematical Society	13
1.7	American Meteorological Society	14
1.8	**American Museum of Natural History**	15
1.9	American Psychological Association	16
1.10	Arthur D. Little, Inc.	17
1.11	**Association of Special Libraries and Information Bureaux** (Phillips), England	17
1.12	**Association of Special Libraries and Information Bureaux** (Slater), England	18
1.13	Association of Special Libraries and Information Bureaux (Whyte), England	18
1.14	Biological Abstracts	18
1.15	Bureau of Social Science Research, Inc.	19
1.16	**Chemical Abstracts Service**	19
1.17	**Columbia University**	20
1.18	**Creedmoor Institute for Psychobiologic Studies**	20
1.19	General Electric Co.	21
1.20	General Electric Research Laboratory	22
1.21	Georgia Institute of Technology (Barker)	22
1.22	Georgia Institute of Technology (Crosland)	23
1.23	**Herner and Co.**	23
1.24	Information Dynamics Corp.	25
1.25	Institute for Advancement of Medical Communication	26
1.26	Institute for Scientific Information	27
1.27	International Business Machines Corp.	29
1.28	**International Federation for Documentation (FID),** Netherlands	29
1.29	John I. Thompson and Co., Inc.	30
1.30	**Lehigh University**	31
1.31	**Library of Congress**	32
1.32	Massachusetts Institute of Technology (Barnett)	33

Page

1. INFORMATION NEEDS AND USES—Continued

1.33 Massachusetts Institute of Technology (Galliher) . 33
1.34 National Federation of Science Abstracting and Indexing Services 34
1.35 National Institutes of Health 34
1.36 New York Botanical Garden 35
1.37 **Országos Mezögazdasagi Könyvtár És Dokumentációs Központ** (Hungary) 36
1.38 Research, Inc. 36
1.39 **Rutgers, The State University** (Clarke) 36
1.40 Rutgers, The State University (Harrar) 37
1.41 **"Shell" Research Ltd.,** England 37
1.42 Standard Oil Co. 38
1.43 Stanford Research Institute (Engelbart) 38
1.44 Stanford Research Institute (Kincaid) 39
1.45 **Surveys and Research Corp.** 39
1.46 Syracuse University 39
1.47 **Tufts University** 40
1.48 United Kingdom Atomic Energy Authority (Bell), England 40
1.49 United Kingdom Atomic Energy Authority (Sabel), England 41
1.50 U. S. Atomic Energy Commission 42
1.51 University of California, Livermore 42
1.52 University of Chicago 42
1.53 University of Illinois 43
1.54 **University of Kansas** 43
1.55 **University of Washington** 44
1.56 **Western Electric Co., Inc.** 45
1.57 Wildlife Disease Association 46
1.58 **Yale Medical Library** 46

2. INFORMATION STORAGE AND RETRIEVAL

2.1 Advanced Information Systems, Inc. 61
2.2 **American Bar Foundation** 62
2.3 American Diabetes Association, Inc. 63
2.4 American Institute for Research 63
2.5 American Legal Data Processing Association . . . 65
2.6 **American Medical Association** 66
2.7 Armed Services Technical Information Agency . . 66
2.8 **Armour Research Foundation** 69
2.9 Association of Special Libraries and Information Bureaux, England 70

Page

2. INFORMATION STORAGE AND RETRIEVAL—Continued

2.10 Badische Anilin- & Soda-Fabrik AG., Federal Republic of Germany 71
2.11 Boeing Co. 72
2.12 Bolt Beranek and Newman, Inc. 72
2.13 **Brown University** 73
2.14 **Bureau of Ships** 73
2.15 Cambridge Language Research Unit, England . . 74
2.16 **C-E-I-R, Inc.** 75
2.17 Centre d'Etudes Nucléaires de Saclay, France . . . 76
2.18 Centre National de la Recherche Scientifique, France 77
2.19 Chemical Abstracts Service 78
2.20 Commission on Professional and Hospital Activities 79
2.21 Communauté Européenne de l'Energie Atomique (Euratom), (Braffort), Italy 79
2.22 Communauté Européenne de l'Energie Atomique (Euratom), (Meyer-Uhlenried), Italy 80
2.23 **Compagnie de Saint-Gobain** and **IBM-France**, France 83
2.24 Computer Associates, Inc. 84
2.25 Cornell Aeronautical Laboratory, Inc. 85
2.26 **Creedmoor Institute for Psychobiologic Studies** . . 85
2.27 Datatrol Corp 86
2.28 David Taylor Model Basin 87
2.29 Department of Defense 88
2.30 Diebold Group, Inc. 88
2.31 Documentation Inc. 89
2.32 Douglas Aircraft Co., Inc. 90
2.33 Electro-Optical Systems, Inc. 91
2.34 **Engineers Joint Council** 91
2.35 Esso Research and Engineering Co. (Jahoda) . . . 92
2.36 Esso Research and Engineering Co. (Rupp) . . . 93
2.37 **F. Hoffman-La Roche & Co. Ltd.**, Switzerland . . 93
2.38 **Farbenfabriken Bayer AG**, Federal Republic of Germany 95
2.39 General Electric Co. (Bartlett) 97
2.40 General Electric Co. (Hubbell) 97
2.41 General Electric Co. (Thompson) 98
2.42 General Electric Co. (Wang) 98
2.43 **Georgetown University School of Medicine and Mt. Alto VA Hospital** 98
2.44 Gmelin Institute Documentation Center, Federal Republic of Germany 99
2.45 Harvard University 100

Page

2. INFORMATION STORAGE AND RETRIEVAL—Continued

2.46 **Hatfield College of Technology, England** 100
2.47 Herner and Co. 101
2.48 **Hughes Aircraft Co.** 102
2.49 Human Relations Area Files, Inc. 104
2.50 Index & Retrieval Systems Inc. 105
2.51 Indiana University 105
2.52 **Information Dynamics Corp.** 106
2.53 **Information for Industry, Inc.** 107
2.54 Institute for Advancement of Medical Communication (Schultz) 107
2.55 Institute for Advancement of Medical Communication (Welt) 108
2.56 **Institute for Bio-Medical Computer Research** . . . 109
2.57 Institute of Library Science, Hungary 109
2.58 Institute for Scientific Information 110
2.59 International Business Machines Corp. (Baxendale) 111
2.60 International Business Machines Corp. (Courtney) 112
2.61 International Business Machines Corp. (Griffin) . . 113
2.62 International Business Machines Corp. (Kochen) . 114
2.63 International Business Machines Corp. (Nienburg) . 116
2.64 **IBM-France**, France 118
2.65 Itek Corp. (Kuipers) 118
2.66 Itek Corp. (O'Brien) 120
2.67 Japan Information Center of Science and Technology, Japan 120
2.68 Johns Hopkins University 121
2.69 Jonker Business Machines, Inc. 122
2.70 Karlova Universita, Czechoslovakia 123
2.71 **Lehigh University** 123
2.72 Leningrad State University imeni A. A. Zhdanov, U.S.S.R. 124
2.73 **Library Association**, England 125
2.74 Library of Congress 125
2.75 Library Research Circle and **Documentation Research and Training Centre**, India 126
2.76 Maison des Sciences de l'Homme and **Ecole Pratique des Hautes Etudes, VIe Section**, France 127
2.77 Massachusetts Institute of Technology (Kessler) . . 128
2.78 Massachusetts Institute of Technology (Teager) . . 129
2.79 National Academy of Sciences-National Research Council 129
2.80 National Biomedical Research Foundation 130

Page

2. INFORMATION STORAGE AND RETRIEVAL—Continued

2.81 National Book League, England 131
2.82 National Bureau of Standards (Alexander) . . . 132
2.83 **National Bureau of Standards (Stern)** 134
2.84 National Bureau of Standards (Stevens) 134
2.85 **National Chiropractic Association** 135
2.86 National Library of Medicine 136
2.87 National Physical Laboratory, England 137
2.88 Octrooiraad, Netherlands 137
2.89 Pacific Southwest Forest and Range Experiment Station 138
2.90 Planning Research Corp. 138
2.91 Princeton University 138
2.92 Project Lawsearch 139
2.93 **Pulp and Paper Research Institute of Canada and The Institute of Paper Chemistry** 139
2.94 Radio Corp. of America 140
2.95 Reading Chemists' Club 141
2.96 Rensselaer Polytechnic Institute 141
2.97 Rockford Research Institute Inc. 142
2.98 **Rome Air Development Center** 142
2.99 **Rutgers, The State University** 143
2.100 Science Index Group 144
2.101 Scientific Documentation Centre Ltd., Scotland . . 145
2.102 **Société d'Économie et de Mathématique Appliquées (SEMA)**, France 146
2.103 **Southwestern Legal Foundation** 147
2.104 **Spencer Chemical Co.** 148
2.105 **Stanford Research Institute** 149
2.106 System Development Corp. (Borko) 149
2.107 System Development Corp. (Harrington) 152
2.108 **Technische Hochschule München, Federal Republic of Germany** 153
2.109 Thompson Ramo Wooldridge Inc. (Edmundson) . 155
2.110 Thompson Ramo Wooldridge Inc. (Garvin) . . . 158
2.111 U. S. Army Biological Laboratories 159
2.112 U. S. Department of Agriculture 159
2.113 **U. S. Naval Postgraduate School** 160
2.114 U. S. Patent Office 161
2.115 **UNIVAC Division of Sperry Rand Corp.** 165
2.116 **Universidad Industrial de Santander, Colombia** . . 166
2.117 Université Catholique de Louvain, Belgium . . . 166
2.118 Université de Paris, France 167

Page

2. INFORMATION STORAGE AND RETRIEVAL—Continued

2.119 University of Arizona 168
2.120 University of Chicago 169
2.121 University of Illinois Library 170
2.122 University of Michigan 170
2.123 **University of Oklahoma** 171
2.124 University of Pennsylvania (O'Connor) 171
2.125 University of Pennsylvania (Prywes) 172
2.126 University of Pittsburgh 173
2.127 University of Rhode Island 174
2.128 Western Reserve University 174
2.129 **Independent Work (Farradane), England** 179
2.130 **Independent Work (Lawlor)** 180
2.131 **Independent Work (Lipetz)** 181
2.132 Independent Work (Schneider) 182
2.133 Independent Work (Williams) 183

3. MECHANICAL TRANSLATION

3.1 Académie de la République Populaire Roumaine, Rumania 190
3.2 Birkbeck College, England 191
3.3 Cambridge Language Research Unit, England . . 192
3.4 Centre National de la Recherche Scientifique (Sestier), France 193
3.5 Centre National de la Recherche Scientifique (Vauquois), France 194
3.6 Communauté Européenne de l'Energie Atomique (Euratom), Italy 194
3.7 Computer Concepts Inc. 195
3.8 **Computer Usage Co., Inc.** 196
3.9 Electrotechnical Laboratory, Japan 196
3.10 First Research Center, Defense Agency, Japan . . 197
3.11 Georgetown University 198
3.12 Georgia Institute of Technology 198
3.13 Harvard University 199
3.14 Institut Za Eksperimentalnu Fonetiku, Yugoslavia . 200
3.15 Institute of Electronics, Automatics and Telemechanics, Academy of Sciences Georgian S.S.R., U.S.S.R. 201
3.16 Institute of Linguistics, Academy of Sciences U.S.S.R., U.S.S.R. 201
3.17 Institute of Mathematics, Siberian Branch, Academy of Sciences U.S.S.R., U.S.S.R. 202

Page

3. MECHANICAL TRANSLATION—Continued

3.18 International Business Machines Corp. 202
3.19 **IBM-France,** France 204
3.20 Karlova Universita (Novák), Czechoslovakia . . . 204
3.21 **Karlova Universita** (Sgall), Czechoslovakia 204
3.22 **Kasvatusopillinen Korkeakoulu,** Finland 205
3.23 Kyushu University, Japan 205
3.24 Leningrad State University imeni A. A. Zhdanov, U.S.S.R. 206
3.25 **Lockheed Missiles & Space Co.** 207
3.26 Machine Translation, Inc. 207
3.27 Massachusetts Institute of Technology 207
3.28 National Bureau of Standards 209
3.29 National Physical Laboratory, England 210
3.30 **Ohio State University** 210
3.31 RAND Corp. 211
3.32 Research, Inc. 212
3.33 Summer Institute of Linguistics 212
3.34 **Thompson Ramo Wooldridge Inc.** 213
3.35 Universidad Nacional Autonoma de Mexico, Mexico 214
3.36 Università degli Studi di Milano, Italy 214
3.37 Université Libre de Bruxelles, Belgium 216
3.38 University of California, Berkeley 217
3.39 University of Michigan 220
3.40 University of Texas 220
3.41 University of Washington 223
3.42 Uniwersytet Warszawski, Poland 224
3.43 Vilniaus Valstybinis V. Kapsuko Vardo Universitetas, Lithuania 225
3.44 **Výzkumný Ústav Matematických Strojů,** Czechoslovakia 225
3.45 Washington State University 226
3.46 Wayne State University 226
3.47 **Yale University** 228
3.48 Independent Work (Corbe), France 228

4. EQUIPMENT

4.1 American Brake Shoe Co. 234
4.2 **Centre National de la Recherche Scientifique,** France 235
4.3 FMA, Inc. 235
4.4 General Electric Co. 236
4.5 General Electric Research Laboratory 236
4.6 **Hallicrafters Co.** 237

Page

4. EQUIPMENT—Continued

4.7 Herner and Co. 237
4.8 Houston Fearless Corp. 238
4.9 Index & Retrieval Systems Inc. 238
4.10 Information Retrieval Corp. 238
4.11 Intectron, Inc. 239
4.12 Itek Corp. 239
4.13 Magnavox Research Laboratories 240
4.14 Massachusetts Institute of Technology (Baumann) . 240
4.15 Massachusetts Institute of Technology (Scott) . . 241
4.16 Memistor Corp. 241
4.17 Mergenthaler Linotype Co. 242
4.18 National Bureau of Standards (Bagg) 242
4.19 National Bureau of Standards (Stern) 243
4.20 National Cash Register Co. 244
4.21 Photon, Inc. 245
4.22 University of Cambridge, England 246
4.23 Walter Reed Army Medical Center 246

5. POTENTIALLY RELATED RESEARCH

5.1 *Character and Pattern Recognition*

5.1.1 Arthur D. Little, Inc. 252
5.1.2 Bell Telephone Laboratories, Inc. 253
5.1.3 Budd Electronics 253
5.1.4 Burroughs Control Corp. 254
5.1.5 Burroughs Corp. 255
5.1.6 Case Institute of Technology 256
5.1.7 Chrysler Corp. 257
5.1.8 Electrotechnical Laboratory, Japan . . . 258
5.1.9 Farrington Electronics, Inc. 259
5.1.10 General Dynamics/Electronics 259
5.1.11 Ing. C. Olivetti & C.,S.P.A., Italy 260
5.1.12 Institute of Automatics and Telemechanics, Academy of Sciences U.S.S.R., U.S.S.R. . . 261
5.1.13 Institute of Cybernetics, Academy of Sciences Ukrainian S.S.R., U.S.S.R. 261
5.1.14 International Business Machines Corp. (Greanias) 262
5.1.15 International Business Machines Corp. (Rohland) 264
5.1.16 International Business Machines Corp. (Shelton) 264
5.1.17 Kyoto University, Japan 264

Page

5. POTENTIALLY RELATED RESEARCH—Continued

5.1 *Character and Pattern Recognition*—Continued

5.1.18 Link Division of General Precision, Inc. . . . 265
5.1.19 **Litton Systems, Inc.** 265
5.1.20 **Lockheed Missiles & Space Co.** 266
5.1.21 **Massachusetts Institute of Technology** . . 267
5.1.22 Melpar, Inc. 268
5.1.23 National Bureau of Standards 268
5.1.24 National Physical Laboratory, England . . 269
5.1.25 New York University (Boni) 269
5.1.26 New York University (Freeman) 270
5.1.27 Panoramic Research, Inc. 271
5.1.28 **Philco Corp.** (Bogusz) 271
5.1.29 Philco Corp. (Chatten) 272
5.1.30 Philco Corp. (Frank) 272
5.1.31 Post Office Department 272
5.1.32 Rabinow Engineering Co., Inc. 272
5.1.33 Radio Corp. of America 272
5.1.34 **Radio Research Laboratories,** Japan . . . 273
5.1.35 **Raytheon Co.** 273
5.1.36 Recognition Equipment Inc. 274
5.1.37 SCOPE Inc. 274
5.1.38 Stanford Research Institute (Brain) 275
5.1.39 Stanford Research Institute (Rosen) . . . 276
5.1.40 Swarthmore College 276
5.1.41 Sylvania Electronic Systems 277
5.1.42 Technische Hochschule Karlsruhe, Federal Republic of Germany 277
5.1.43 University of California, Berkeley 278
5.1.44 University of Michigan 278
5.1.45 University of Michigan and System Development Corp. 279

5.2 *Speech Analysis and Synthesis*

5.2.1 Air Force Cambridge Research Laboratories (Petrick) 281
5.2.2 Air Force Cambridge Research Laboratories (Wathen-Dunn) 282
5.2.3 Autonetics 283
5.2.4 Bell Telephone Laboratories, Inc. 283
5.2.5 Bulova Research & Development Laboratories 284
5.2.6 Cognitronics Corp. 284
5.2.7 Columbia University 284

Page

5. POTENTIALLY RELATED RESEARCH—Continued

5.2 *Speech Analysis and Sythesis*—Continued

5.2.8 **Cornell Aeronautical Laboratory, Inc.** . . . 284
5.2.9 Federal Scientific Corp. 285
5.2.10 General Dynamics/Electronics 285
5.2.11 International Business Machines Corp. . . 286
5.2.12 Kyoto University, Japan 286
5.2.13 **Litton Systems, Inc.** 287
5.2.14 Massachusetts Institute of Technology (Forgie) 287
5.2.15 Massachusetts Institute of Technology (Stevens) 287
5.2.16 Motorola, Inc. 289
5.2.17 New York University 289
5.2.18 **Philco Corp.** 290
5.2.19 **Radio Corp. of America** 290
5.2.20 Radio Research Laboratories, Japan . . . 291
5.2.21 Royal Institute of Technology, Sweden . . 292
5.2.22 Sylvania Electronic Systems 292
5.2.23 Thompson Ramo Wooldridge Inc. 293
5.2.24 **Universität Bonn**, Federal Republic of Germany 293
5.2.25 University College London, England . . . 294
5.2.26 University of Edinburgh, Scotland 294
5.2.27 University of Michigan 294
5.2.28 University of Rochester 295

5.3 *Linguistic and Lexicographic Research*

5.3.1 American Bible Society 298
5.3.2 Arthur D. Little, Inc. 298
5.3.3 **Brown University** 298
5.3.4 Centro per l'Automazione dell'Analisi Letteraria, Italy 299
5.3.5 **Československá Akademie Věd**, Czechoslovakia 299
5.3.6 Charles W. Adams Associates, Inc. 300
5.3.7 **Forskningsgruppen för Kvantitativ Lingvistik (KVAL)**, Sweden 301
5.3.8 General Electric Co. 302
5.3.9 Hebrew University, Israel 302
5.3.10 **Humboldt-Universität zu Berlin, Germany** . 303
5.3.11 Indiana University (Householder) 303
5.3.12 Indiana University (Sebeok) 304

Page

5. POTENTIALLY RELATED RESEARCH—Continued

5.3 *Linguistic and Lexicographic Research*—Continued

5.3.13 Institut National des Techniques de la Documentation, France 305
5.3.14 Institute of Linguistics, Leningrad Department, Academy of Sciences U.S.S.R., U.S.S.R. 306
5.3.15 Institute of Mathematics, Siberian Branch, Academy of Sciences U.S.S.R., U.S.S.R. . 307
5.3.16 **IBM-France**, France 308
5.3.17 **Karlova Universita**, Czechoslovakia . . . 309
5.3.18 Laboratoire d'Analyse Lexicologique, France 309
5.3.19 Librascope Division of General Precision, Inc. 310
5.3.20 Manchester University, England 310
5.3.21 Massachusetts Institute of Technology (Barnett) 310
5.3.22 Massachusetts Institute of Technology (Chomsky) 311
5.3.23 **Matematikmaskinnämnden**, Sweden . . . 312
5.3.24 National Language Research Institute, Japan 313
5.3.25 Radio Corp. of America 313
5.3.26 RAND Corp. 314
5.3.27 Summer Institute of Linguistics 315
5.3.28 **Tuskegee Institute** 316
5.3.29 **Universität Bonn**, Federal Republic of Germany 316
5.3.30 University of California, Berkeley 318
5.3.31 University of Cambridge, England 318
5.3.32 University College London (Quirk), England 319
5.3.33 University College London (Smith), England 320
5.3.34 University of Edinburgh, Scotland 320
5.3.35 **University of Gothenburg**, Sweden 321
5.3.36 University of Pennsylvania (Brown) . . . 321
5.3.37 University of Pennsylvania (Harris) . . . 322
5.3.38 **University of Pittsburgh** 325
5.3.39 **University of Wisconsin** (Fowler) 326
5.3.40 **University of Wisconsin** (Roberts) 326
5.3.41 **Independent Work (Abraham)**, Rumania . 327
5.3.42 Indepnedent Work (Spiegelthal) 328
5.3.43 **Independent Work (Wenck)**, Federal Republic of Germany 329

Page

5. POTENTIALLY RELATED RESEARCH—Continued

5.4 *Artificial Intelligence*

5.4.1 Aeronutronic 330
5.4.2 Air Force Cambridge Research Laboratories 331
5.4.3 Armour Research Foundation 331
5.4.4 Astropower, Inc. 332
5.4.5 Autonetics 333
5.4.6 Cambridge Language Research Unit, England 333
5.4.7 Cornell Aeronautical Laboratory, Inc. . . 334
5.4.8 Cornell University 336
5.4.9 Institute of Cybernetics, Academy of Sciences Ukrainian S.S.R., U.S.S.R. 336
5.4.10 Librascope Division of General Precision, Inc. 337
5.4.11 Massachusetts Institute of Technology (Minsky) 339
5.4.12 Massachusetts Institute of Technology (Ross) 339
5.4.13 Mitre Corp 342
5.4.14 National Bureau of Standards 342
5.4.15 Northwestern University 343
5.4.16 Philco Corp. (Bryan) 343
5.4.17 Philco Corp. (Chatten) 344
5.4.18 Physikalisches Institut der Technischen Hochschule Aachen, Federal Republic of Germany 344
5.4.19 Princeton University 345
5.4.20 RAND Corp. 345
5.4.21 RAND Corp. and Carnegie Institute of Technology 346
5.4.22 Rockford Research Institute Inc. 347
5.4.23 Stichting Studiecentrum Voor Administratieve Automatisering, Netherlands 347
5.4.24 System Development Corp. 348
5.4.25 Technische Hochschule Karlsruhe, Federal Republic of Germany 351
5.4.26 Università degli Studi di Milano, Italy . . 352
5.4.27 University of California, Berkeley 354
5.4.28 University of California and System Development Corp. 356
5.4.29 University of Chicago 356
5.4.30 University of Illinois 357
5.4.31 University of Michigan 359
5.4.32 University of Pennsylvania 361

Page

5. POTENTIALLY RELATED RESEARCH—Continued

5.4 *Artificial Intelligence*—Continued

5.4.33 University of Southampton, England . . . 363
5.4.34 University of Texas 363
5.4.35 Independent Work (Roy), Scotland 363

5.5 *Psychological Studies*

5.5.1 Advanced Information Systems, Inc. . . . 364
5.5.2 Harvard University (Bruner) 365
5.5.3 Harvard University (Stone) 365
5.5.4 Medical Research Council, England 368
5.5.5 Montana State University 368
5.5.6 University of California, Los Angeles . . . 369
5.5.7 University of Connecticut 370
5.5.8 University of Maryland 370
5.5.9 University of Michigan 371
5.5.10 William Alanson White Institute 372

Glossary 374

List of Acronyms and Abbreviations 379

Index of Individuals and Organizations 383

Index of Sponsors 415

Equipment Index 418

Subject Index 423

INTRODUCTION

This is the eleventh in a series of descriptive reports on current research and development in scientific documentation. These reports appear semiannually. Additional copies of issue Nos. 9 and 10 are available from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. Additional copies of this report are available from the Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C.

These reports are compiled by the Office of Science Information Service of the National Science Foundation as a service to individuals and organizations interested in scientific documentation. All pertinent activities in the United States and abroad that have come to the attention of the Foundation staff are included. The descriptive statements were prepared by the research workers themselves, with only minor editing by the National Science Foundation. In those cases where no significant change was noted in the description of a project, a brief statement indicating the project's continuation has been included along with pertinent references. The statements that are used do not necessarily represent the views of the National Science Foundation. It is suggested that in quoting material from the statements describing specific projects, the source be referred to as "Statement (s) provided for the National Science Foundation report, *Current Research and Development in Scientific Documentation.*"

The statements are grouped under the five major headings below. To facilitate reference, the statements have been assigned consecutive numbers within the various sections. The table of contents indicates these numbers, and the indexes at the close of the report use these numbers to identify the section and statement to which the index entry refers.

1. *Information Needs and Uses*—studies and analyses of the information needs of scientists, of the uses made of scientific and technical information, and of communication problems in science and technology, including publication studies and experiments with new publishing formats and techniques.

2. *Information Storage and Retrieval*—studies of methods, systems, and procedures for analyzing, organizing, encoding, storing, and searching subject matter, including theoretical studies of information storage and retrieval.

3. *Mechanical Translation*—research on problems of automatic translation from one natural language to another.

4. *Equipment*—development of devices for the processing of scientific information including devices for reading, storing, searching, transmitting, and translating.

5. *Potentially Related Research*—work on problems not immediately connected with scientific documentation but whose solution is likely to have an impact on the future of documentation, including such fields as character and pattern recognition, speech analysis and synthesis, linguistic and lexicographic research, artificial intelligence, and certain psychological and sociological studies.

Each section is preceded by an introductory summary which calls attention to the work of organizations being reported for the first time, to projects and studies which have been completed or discontinued, and to closely related research projects in other sections of the report. The nature of many of the research projects is such that it is difficult to categorize them according to research without scattering some work that is related. For example, some research groups with different ultimate objectives, such as mechanical translation procedures versus information retrieval procedures, may be engaged in similar research on automatic methods for parsing the sentences of one or more natural languages. Such relationships are pointed out in the introductory summaries. The summaries are based on the information contained in the contributors' descriptive statements and are not intended to be evaluative.

Some new features introduced in the previous issue are intended to make the report a more useful and convenient reference aid. These features include a glossary of specialized terms used in some of the project descriptions, a list of acronyms and abbreviations, a separate index for equipment, and a sample checklist of the kind of information to be included in the descriptive statements. In this issue the organizations reporting on new work for the first time are indicated in boldface type in the table of contents and in the introductory summaries. Comments of readers on the usefulness of these features and the introductory summaries are welcomed. Suggestions and comments for improving or extending the indexes and lists are also welcomed.

This issue includes 423 statements describing over 500 research projects, studies, and experiments in 264 organizations, an increase of 60 statements, approximately 100 projects, and 42 organizations over issue No. 10. The research efforts of 84 organizations in 23 foreign countries are described in a total of 112 statements.

The National Science Foundation is preparing a cumulative bibliography of all publications and reports listed in the *CRDSD* series. The Office of Technical Services, U. S. Department of Commerce, with the support of the National Science Foundation, is organizing a service

which will ensure the continuing availability of these publications and reports. The principal purpose of such service will be to make it possible for interested persons to obtain copies of any report listed in the series when the original source no longer has copies for distribution. OTS will provide Xerox or microfilm copies. The beginning of this service will be announced in the near future.

There are undoubtedly numerous projects, studies, and experiments which are not included in this issue simply because they have failed to come to the attention of the Foundation staff. The National Science Foundation would appreciate receiving information on other pertinent programs both in the United States and abroad for possible inclusion in future reports. To assist possible contributors in the preparation of statements describing their work, a sample checklist of information required may be found on page —. Correspondence should be addressed to the Program for Documentation Research, Office of Science Information Service, National Science Foundation, Washington 25, D. C.

Scientists and other individuals or organizations concerned with scientific documentation will be placed on the mailing list for these reports upon request. Additional copies are available from The Office of Technical Services or the U. S. Government Printing Office, as described above.

This report was compiled by Mr. Eugene Pronko, Mrs. Nancy R. Donald, and Miss Mary G. Lewis of the Program for Documentation Research, Office of Science Information Service, National Science Foundation.

CHECKLIST

Form Approved
Budget Bureau No. 99-R192

CURRENT RESEARCH AND DEVELOPMENT IN SCIENTIFIC DOCUMENTATION

Material for *Current Research and Development in Scientific Documentation* should be typewritten and double spaced, in duplicate. Statements in English are preferred. Foreign-language statements, however, are acceptable and should be forwarded immediately so that a translation can be made and returned to the sender for review. The following information should be included, whenever applicable, in each descriptive statement.

1. Name and address of organization, complete enough to be used as a mailing address. Use the language of the country in which organization is located; please suggest English translation for non-English organization title.
2. Name of principal investigator or project leader.
3. Concise description of project, presenting a complete summary of the research underway. The text of this description should cover the following project characteristics in the exact order listed below:
   a. Statement of the purpose of the project.
   b. Scope, methodology, and approach including the following specific facts where applicable:
      Subject matter used in research or experiments (e.g., metallurgical literature, Russian physics texts, etc.)
      In mechanical translation research, the language (s) worked with.
      Pertinent quantitative information (e.g., number of people interviewed, of documents indexed, or words in mechanized glossary, etc.).
      Equipment used in course of research. Please give identifying numbers for computers.
      In equipment development, the stage of development and the type of documentation applications for which the equipment is being designed or may be used.
   c. Statement of progress made during previous 6-12 months. Mention specific results obtained whether they be preliminary, intermediate, final, positive, negative, etc.
   d. Next phase of the project to be undertaken or planned.
   e. Source of financial support, if work is being done on a grant or contract basis, and duration of the project.

   If more than one project is to be described, a separate text should be submitted for each such project.
4. Complete bibliographical citations to recent publications or essential background papers directly related to the work described. Authors, titles, and sources should be cited in the language in which the document is written. If copies of these publications have not already been sent to us, please include *two copies each* with your statement so that we may retain them for our reference.
5. Definitions of terminology used in a unique or specialized way in project descriptions should be provided on a separate sheet of paper. These will be considered for inclusion in a glossary appended to the report, intended to clarify terms which may be unfamiliar to users of the report.

1
INFORMATION NEEDS AND USES

1. INFORMATION NEEDS AND USES

The following summary calls attention in particular to new projects reported here for the first time, to progress made in certain previously reported work, to developments or projects which have been completed, and to closely related work which appears in other sections of this report.

COMMUNICATION PROBLEMS IN SCIENCE

The scientist is both a generator and user of written and nonwritten information. His reporting of research by means of the printed word is only one aspect of the larger problem of communication among scientists. But, by and large, it is this one aspect which has received the most attention in recent years in a wide range of attempts to solve the myriad problems created by the rapid growth of the scientific literature. Nonwritten scientific communication is also of great importance. The scientist attends conferences, talks with associates in his particular field of interest, and interacts with members of his own group. Information exchange is an essential part and consequence of all these activities. As a generator-user of information the scientist engages in human communication processes about which much is yet to be learned. And as if oral and written communication were not enough, today's scientist must face the problem of communication with machines that will hopefully enable him to search and select the information that he needs when he needs it.

Some studies are concerned with different aspects of the communication process, including technical writing effectiveness and the problems involved in man's use of machines as an extension of his intellectual capabilities.

Tufts University (1.47) is making a critical review of the literature and past research on the effectiveness of technical writing as a means of communication. An anticipated result of this study is the indication of areas where future research will be required.

Advanced Information Systems Company, Inc. (1.1) has a study underway to determine the effects of behavioral factors upon the success of information systems.

Research underway at the **University of Washington** (1.55) concerns (a) human attitudes toward computer use and results, (b) machine learning processes, and (c) memory organization for meaning association and retrieval.

STUDIES OF NEEDS AND USES

Requirements of institutions and of individual categories of scientists, the uses made of available information systems and services, and the problem of determining the best kinds of mechanisms to serve the needs of the scientific community continue to receive increased attention.

Lehigh University (1.30) has undertaken a study to develop and evaluate research methodology for scientists' patterns of information uses and requirements.

Jonker Business Machines, Inc. (2.69) is investigating the requirements of information retrieval networks with respect to the merging of collections of data and information, to methods of making index vocabularies compatible, and to the utilization of various types of disseminable search devices and centralized or disseminable microform document storage systems.

The Creedmoor Institute for Psychobiologic Studies (1.18) is conducting a study of literature needs of neuropsychiatric research workers using an unstructured interview technique.

An analysis of the information requirements of chemical engineers is being continued by the **American Institute of Chemical Engineers** (1.4).

A preliminary survey of the inquiries received by the Scientific Information Office, United Kingdom Atomic Energy Authority, Winfrith (1.48) during the first 6 months of 1962 has been completed.

Information for Industry, Inc. (2.53) is making a quantitative analysis of patent information and patterns of use in chemistry and related fields as criteria for design of a magnetic tape index system.

Herner and Company (1.23) has a study underway to determine the feasibility, structure, and possible mode of organization of an abstracting and indexing service for the mental health literature.

Columbia University (1.17) is engaged in a project to identify, describe, and evaluate specialized social science information services in the United States.

Országos Mezögazdasági Könyvtár És Dokumentációs Központ (1.37) is surveying the possibilities for an information service for agricultural specialists.

Western Electric (1.56) is investigating the technical information needs of its engineers.

The National Bureau of Standards (2.82) has completed its preliminary investigation of the information processing needs of the Food and Drug Administration.

"Shell" Research Ltd. (1.41) is investigating the pattern of use by individual scientists of the different services provided by an information organization.

A project at Aslib (1.11) is attempting to determine the adequacy

of literature searching by scientists prior to starting experimental work for new projects, the difficulties they encounter, and the extent to which assistance from librarians and information officers is available and used. Another study at **Aslib** (1.12) concerns the use made of library and information services in industrial and academic organizations.

Herner and Company (1.23) is conducting an evaluative study of the character and degree of use of the *Chemical Engineering Thesaurus.*

Yale Medical Library (1.58) has completed an investigation of the citation use of medical doctoral theses during the period 1850-1960 and is currently investigating the citation use of medical periodicals covering the period 1951-1960, which will yield data on such journal characteristics as frequency, language, country of origin, and field or specialty. Studies are also being conducted on the recorded use of books and journals.

The **American Museum of Natural History** (1,8) is studying the usability of microfilm copies of museum catalogs.

PUBLICATION STUDIES AND EXPERIMENTS

Primary and secondary publications represent the major portion of available recorded scientific information and the general means by which new scientific knowledge is communicated to the scientific community as a whole. Consequently, major efforts have been directed toward improving these vehicles of scientific communication.

Publication studies are concerned primarily with availability of publications, the pattern of their dissemination among scientists, and their use and value.

Studies of the Characteristics of Scientific Publication

Knowledge of the characteristics of scientific publication in the various fields of science is essential in discerning trends and patterns in publication activity, growth rate of the literature, sources of support for scientific research, and the publication habits of scientists, and possibly in identifying the institutions in which the more important work is concentrated.

In a recently completed study of *Physics Abstracts,* the **American Institute of Physics** (1.5) studied the abstracting of three types of journals for timelag, rate of completion, and depth of coverage of each journal issue.

A study of the content, influence, availability, and value of scientific conference papers and proceedings has been completed by **FID** (1.28).

The publication trend of papers presented at national meetings of the **American Chemical Society** (1.2) was examined for the purpose of planning more effective dissemination channels for meeting papers.

Herner and Company (1.23) has completed the compilation and analysis of science news media, with particular emphasis on the field of medicine.

Rutgers State University (1.39) is conducting a study of the impact of photographic copying on the publication of scholarly materials in the United States and will attempt to derive a formula showing the point at which it becomes more economical to subscribe to a journal than to copy from it.

Readership Studies and Publications Use Evaluation

Library of Congress (1.31) is having a contract study (see The Diebold Group, 2.30) made of user reactions to the *Monthly Index of Russian Accessions.*

The Institute for Advancement of Medical Communication (1.25) recently completed a study designed to evaluate the effectiveness of the U. S. Public Health Service's 5-year-old Russian Scientific Translation Program.

The **American Chemical Society** (1.2) is conducting a study of readers' reactions to the ACS basic journals program and is also planning to publish an experimental journal to elicit readers' reactions to proposed modifications.

Surveys and Research Corporation (1.45) is making a study of *U. S. Government Research Reports* and *Keywords Index to U. S. Government Technical Reports* to determine the comparative effectiveness of the two publications as announcement mechanisms and to determine user reaction to them.

Computer-Produced Indexes and Digests

The **University of Kansas** (1.54) is preparing for publication the first issue of a new permuted title index of Slavic journal articles in the social sciences and humanities. The original wording in the titles will be permuted in transliterated form.

Chemical Abstracts Service (1.16) has developed a computer-produced digest of information in the field of biological action of organic compounds. *Chemical-Biological Activities* provides a KWIC index to compound names and conceptual information, together with molecular formula, notation, and author indexes.

Euratom (2.22) has produced a cumulative index to the 12 volumes (1950-1961) of *Nachrichten für Dokumentation* by means of an IBM 1401 computer program, and intends to add entries from other representative documentation journals to produce a comprehensive cumulative index to the periodical literature in the field of documentation.

Format Studies and Experiments

The **American Institute of Physics** (1.5) has in progress the experi-

mental development of a "contents" journal and a combined index for all AIP journals. They also have under study by an independent contractor (see 2.131) the applicability of citation indexing techniques to physics.

To demonstrate the TABLEDEX index, the National Biomedical Research Foundation (2.80) has produced by computer a bibliography of International Geophysical Year literature illustrating two different TABLEDEX formats.

New Printing Techniques

To meet the requirements imposed by utilization of computing and other equipment in index production and bibliography compilation and to handle the sometimes unconventional symbols peculiar to certain scientific publications, new printing techniques continue to be explored.

Itek Corporation (2.65) is carrying on research and development tasks on (a) computer composition and the generation of typesetting control data and (b) computer programming with a flicker-free display capability for textual and graphic processing, indexing, and editorial processing.

Computer programs are being developed by **Matematikmaskinnämnden** (5.3.23) to insert corrections, additions, and deletions in a punched text, yielding a tape with updated text that can be fed into a teletypesetter reading unit. They are also developing an automatic hyphenation program.

Some of the equipment development and studies described in Section 4 (Equipment) are pertinent to the problem of applying new printing techniques, especially the work of Mergenthaler Linotype Company (4.17) and Photon, Inc. (4.21). [See also the introduction to Section 4 (Equipment), p. 233].

COMPLETED PROJECTS AND STUDIES

The following projects and studies have been completed:

Bureau of Social Science Research, Inc. (1.15)
Georgia Institute of Technology (1.22)
John I. Thompson and Company, Inc. (1.29)
Rutgers State University (1.40)
Stanford Research Institute (1.44)

Association of Special Libraries and Information Bureaux (1.13) [discontinued]

ADVANCED INFORMATION SYSTEMS, INC. 1.1

3002 Midvale Avenue, Los Angeles 34, Calif.

JOHN A. POSTLEY

A study underway seeks to determine the effects of behavioral factors upon the attainment of the objectives of information systems. Tech-

niques are being sought which will enable the system designer to anticipate problem areas likely to be encountered in the implementation of a particular system. The study postulates that behavioral and organizational considerations can be assessed in terms of a limited number of factors, and that the degree of success in the change to a particular information system in a given organizational structure can be predicted on the basis of these factors.

In the first phase of the study, an attempt is being made (a) to identify the psychological, sociological, and organizational factors relevant to the goals of a library information system; and (b) to determine the qualitative, and in some cases, the quantitative, patterns which reflect the relationships between attitudes and organizational structures, on the one hand, and the achievement of information system goals on the other. In the last phase of the project, trial simulations of particular organizational structures and their information systems in the environment of a management training laboratory will be undertaken, utilizing the results of the first phase.

To date, initial identification has been made of the behavioral factors which might tend to affect the value, in terms of the acceptance or non-acceptance, of an information system. Four tests (questionnaires) have been selected to measure individual attitudes with respect to behavioral factors. Techniques have been developed to relate the measurement of attitudes obtained by the tests to the objectives and goals of the information system.

The project was initiated in May 1962 with support from the National Science Foundation.

1.2 AMERICAN CHEMICAL SOCIETY

1155 16th Street, NW., Washington 6, D. C.
RICHARD H. BELKNAP, *Director of Planning*

The first phase of a study of the publication habits of American chemists has been completed. An attempt was made to evaluate the publication outlets available to chemists and biochemists doing research in American academic institutions. Work is continuing on extending this study to chemists working in industrial and Government laboratories. It is hoped that specific areas will be identified which appear to have unsatisfactory coverage. Special attention will also be given to the needs of the chemist as author and generator of information.

I. Publication Trend of Conference Papers

To determine the publication trend of papers presented at national meetings of the ACS, an analysis was made of the final publication status of all papers presented before the major chemistry divisions during the 131st ACS National Meeting (1957). Although it was not

possible within the scope of this study to determine exactly the total number of meeting papers eventually published, a good insight into the final fate of these papers was acquired. Valuable information, such as a comparison of publication trends among the various fields of chemistry, a breakdown of papers by journals, timelag involved between presentation and publication of papers, was brought out during this analysis. The results will be useful in the planning of more effective dissemination channels for meeting papers. A more intensive study involving both presentation and publication of these papers is being planned.

II. Basic Research Sponsorship

A survey has been initiated to determine the amount of basic research sponsorship as reflected in chemistry. This investigation is based on the premise that a quantitative analysis of the articles in the major basic journals of chemistry in the United States would provide a good yardstick for evaluating both the amount of basic chemical research and the relative contributions of the sponsor groups. Publication trends as established in this survey show a general pattern in the financing of research in chemistry. As more data on funding are released, a better evaluation of research costs will be possible. Policies concerning the management of research must be clearly formulated, and an accurate method for determining the operating cost structure of basic research must be established. Then, an exact relation between the sponsor-provided funds and the research results can be determined.

III. User Needs for Basic Journals

Foremost among the considerations continuously facing the ACS in its endeavor to formulate short- and long-range plans for its publication activities are readers' reactions to its basic journals program. Several fundamental questions must be frequently answered: (a) What are the current and future needs of chemists for technical periodical literature? (b) How well do present journals fill existing and future needs? and (c) What opportunities exist for the ACS to modify its approach in the fulfillment of the current and future needs of the profession? A study seeking the answers to these questions is being started. The first phase will consist of personal interviews to explore in depth the thinking of a number of opinion makers in the chemical profession and the opinions of readers in each of the major areas of chemistry.

IV. Publication Format Experimentation

A plan to test the reaction of authors and readers to new ideas on ways to produce a more readable, efficient, and inexpensive journal is in its final stage. It is hoped to publish, for a restricted time, a very limited edition of an ACS journal which will incorporate all the ideas over which control can be exerted: inexpensive format, rewritten titles,

special design for the order of presentation of material, highly edited and condensed copy, etc. Copies of this "experimental" edition will be distributed to selected subscribers who also will receive the regular version of the journal. In this way reaction can be tested to see if these ideas can be moved from the discussion to the action stage.

1.3 AMERICAN INSTITUTE OF BIOLOGICAL SCIENCES

Biological Sciences Communication Project (BSCP)
2000 P Street, NW., Washington 6, D. C.
CHARLES W. SHILLING, *Project Director*

The basic research project is responsible for experimentation on new or nonconventional methods, techniques, and systems for improving the technical information aspects of scientific information problems of the life scientists.

A preliminary study of communication channels in biology has been developed to determine the manner and means by which biologists locate and transmit unconventional information. This study will be conducted by interviewing and providing questionnaires to selected laboratories.

A preliminary study has been made to determine the use made of the translated (cover-to-cover) Russian journals. Twenty-five U.S. biological journals were selected as being representative of or related subject- and content-wise to the Russian journals. All references or citations in each article were tabulated and categorized by national origin (4).

In cooperation with the American University, a seminar was held to identify and review the problems associated with the flow of biological sciences information between the scientist who produces it to the scientist who uses it. A report of the seminar is available in two parts (2)(5).

In an effort to evaluate a system using the same principles as are involved in machine techniques, an indexing method has been established and a permuted title index has been produced manually for the *AIBS Bulletin* from January 1951 to December 1961. The procedure was developed with several basic principles in mind: that it would be adaptable and useful to the publications staff of a single scientific journal, that it would provide for a minimum amount of professional staff time, that it should not include other than conventional office equipment, and that it should provide a cumulative feature (3).

The above studies are sponsored by the National Science Foundation.

A grant from the National Institutes of Health provides for a 3-year study to identify the serial literature in the life sciences on a worldwide basis. The major emphasis in this work will be on the collection of

bibliographic data relating to the serials and the location of unique holdings in the United States (1).

References:

(1) Dopkowski, Philip L., and Joanna E. Shields. "Science Information Training Program," *BSCP Communique* 6-62, May 1962.

(2) Hattery, Lowell H. *Information and Communication in Biological Science.* Washington, D. C.: Center for Technology and Administration, School of Government and Public Administration, The American University, December 1961. (Available from BSCP)

(3) Janaske, Paul C. "Manual Preparation of a Permuted-Title Index," *BSCP Communique* 7-62, June 1962.

(4) "Citation of Russian Literature in 25 Selected U.S. Biological Sciences Journals For the Years 1959, 1960, and 1961," *BSCP Communique* 5-62, July 1962.

(5) *Information Handling & Science Information,* ed. by Paul C. Janaske. Washington, D. C.: American Institute of Biological Sciences, 1962.

AMERICAN INSTITUTE OF CHEMICAL ENGINEERS 1.4

345 West 47th Street, New York 17, N. Y.

ROLLIN D. MORSE, *Chairman,*

Standards Subcommittee for Information Retrieval

The experiment with publishing indexing information and abstracts in the American Institute of Chemical Engineers' publications, and also in *Petroleum Refiner,* is continuing, with no change reported since the previous statement [*Ed.*]. A continuing analysis of the information requirements of chemical engineers is being made, and needed extensions to the program are being developed.

References:

(1) American Institute of Chemical Engineers. *Chemical Engineering Thesaurus,* 1961.

(2) Engineers Joint Council. *Proceedings of the Engineering Information Symposium.* New York: The Engineers Joint Council, January 17, 1962. (Price: $2.00)

(3) Holm, B. E. "Information Retrieval—A Solution," *Chemical Engineering Progress,* vol. 57, no. 6, June 1961, pp. 73-78.

(4) Morse, Rollin. "Information Retrieval," *Chemical Engineering Progress,* vol. 57, no. 5, May 1961, pp. 55-58.

(5) Morse, Rollin, and Eugene Wall. "New Information Retrieval Standards," *Petroleum Refiner,* May 1961.

1.5 AMERICAN INSTITUTE OF PHYSICS

335 East 45th Street, New York 17, N. Y.

HUGH C. WOLFE, *Director of Publications and Director, Documentation Research Project, and* PAULINE ATHERTON, *Assistant Director, Documentation Research Project*

The research phase of the Reference Retrieval System Project (1) is in progress. Approximately 1,000 research physicists have supplied descriptions of their research fields and requests for lists of references. These data are being analyzed. Such data from a group of nuclear physicists (2) have been analyzed and made available to physicists, journal indexers, and abstract journal editors. Some checking of existing reference retrieval tools against the indicated requirements has been done. Future plans include the improvement of present indexes to physics research literature based on analysis of the data collected.

Three limited experiments with citation indexing and bibliographic coupling have been made by the project staff (3). Following up on these, an experimental project on the applicability of citation index techniques in physics has been undertaken through subcontract with a specialist in this field (see 2.131).

A study of *Physics Abstracts* has been made (4). The abstracting of three types of journals [(a) in English with no author abstracts, (b) in English with author abstracts of articles, (c) in German] was studied for timelag, rate of completion, and depth of coverage of each journal issue.

Experimental development of a "contents" journal and a combined index for all AIP journals is in progress. Several guides to the literature of physics have been published or are in preparation (5). These publications are widely distributed to graduate students, teachers, librarians, and research physicists.

These projects are supported by the National Science Foundation.

References:

(1) American Institute of Physics. *A Project for the Development of a Reference Retrieval System for Physicists,* April 1962.

(2) Atherton, Pauline. *A Preliminary Report on Phase I of the Reference Retrieval System Development Project.* New York: American Institute of Physics, April 1962.

(3) Atherton, P., and J. C. Yovich. *Three Experiments With Citation Indexing and Bibliographic Coupling of Physics Literature.* New York: American Institute of Physics, April 1962.

(4) American Institute of Physics. *Study of Physics Abstracts—Abstracting Three Types of Journals,* August 1962.

(5) American Institute of Physics. *Annual Physics Book List,* 1960 and 1961 editions; *Check List for an Undergraduate Physics*

Library (in preparation); *The Periodical Literature of Physics* (out of print; revised edition, wider in scope, is in preparation).

AMERICAN MATHEMATICAL SOCIETY 1.6

190 Hope Street, Providence 6, R. I.
GORDON L. WALKER

A study is being made of the use of the Photon machine for the composition of *Mathematical Reviews*, a journal which is most expensive to compose by conventional techniques because of the inclusion of a number of foreign languages and the many symbols used by mathematicians.

Project efforts are concerned with (a) selecting the characters most frequently used in composition of the journal for inclusion on the custom 1,440-character Photon disc, (b) determining the most convenient arrangement of these on the disc, and (c) selecting special-character designs which fit in with the basic alphabets. It is hoped to develop the Photon's facility for accepting automatic input from any perforated tape or card system to the point at which it can handle large portions of mathematical composition. It is planned to investigate the ways in which the Photon may be adapted to various methods of data storage and retrieval, with particular reference to the possibilities for automatic preparation of author and subject indexes.

The Photon has been installed. Two special offset lenses, which permit the setting of all characters on the disc in superior and inferior positions, have been satisfactorily tested. These eliminate the necessity for putting superscripts and subscripts as special characters on the disc. Preliminary studies have shown that the semicircle that contains the italic letters should also contain all of the Roman numbers and the most frequently used symbols, such as the "equals" sign. Italic numbers are likely candidates for omission from the disc and if included will be put on an infrequently used semicircle of the disc. The Century Schoolbook face for the Roman, italic, and boldface alphabets has been chosen. Experiments are being made to determine the appearance of other symbols when used with this type and to determine adjustments needed in the size of Greek characters and symbols. A tape reader and a code converter have been installed on the Photon which make possible the direct transcription of the tape from Flexowriters and Justowriters to Photon film. By means of two small conversion boards, readily interchanged in the machine by hand, the Photon can read either 6-channel or 8-channel Flexowriter tape.

It is planned to prepare a type-specimen book containing specimens of the various type faces set in different sizes, including a complete list of the mathematical symbols on the disc. A 100-page volume in the Memoir

series, published by AMS, will be set on the Photon. A six-page index for the book *Hydrodynamic Instability,* Proceedings of Symposia in Applied Mathematics, volume 13, will be prepared on a Justowriter, and the tape used as input for the Photon. Further experimentation will be made on the adaptation of punched cards or tapes made for other uses to Photon composition. In particular, it will be determined whether the Flexowriter punched cards, which now contain the entries for the Combined Membership List, can be used as an input for the Photon, either by means of a tape converter or additional relay circuits in the machine. A study is being made of the use of Flexowriter tapes to make up the necessary control records used in the editorial process.

The project is supported by the National Science Foundation for a 2-year period which began in January 1962.

Reference:

(1) American Mathematical Society. *Development of the Photon for Efficient Mathematical Composition,* Report No. 1 to the National Science Foundation, June 13, 1962.

1.7 AMERICAN METEOROLOGICAL SOCIETY

Meteorological and Geoastrophysical Abstracts,
P. O. Box 1736, Washington 13, D. C.
MALCOLM RIGBY, *Editor and Supervisor*

Continuous efforts are underway to mechanize fully all indexing, retrieval, and printing operations for abstracts in meteorology, oceanography, hydrology, etc., which are compatible with present tabulator or computer development, in order to (a) speed up regular indexing operations (especially cumulative indexes); (b) speed up bibliographic searches, organization, indexing, and publication thereof; (c) expand capability for producing a larger volume of bibliographies and a larger variety of indexes; (d) systematize, speed up, update, and make more versatile operational tools such as subject heading lists, classification schemes, geographic place authority lists, and serial title listings; and (e) produce monthly and cumulative title listings according to the UNIDEK system.

The number of titles indexed each year has risen from 8,000 to 12,000 during the past 2 years making mechanization imperative, and the number of bibliographies has increased to 15 or 20 per year with a potential demand for 50 to 100 per year—possible only with machine selection and printout. Equipment used is the IBM 870 Document Writer System, IBM 082 sorter, IBM 407 tabulator, IBM 704 computer, and IBM 1401 computer.

Progress in the last 6 months has included: (a) issuance of a second

experimental issue of *Meteorological and Geoastrophysical Titles* with a systematic UDC (UNIDEK) index as well as a modified KWIC index and a bibliography arranged by serial title. Preparation is being made for 11 subsequent issues this year and a 6-, 12-, and 24-month cumulative index; (b) issuance of a Subject Heading list, with equivalent UDC numbers, of 20,000 headings prepared from a deck of punched cards; and (c) issuance of a UDC-Subject Headings list arranged by UDC numbers, for systematic checking, elimination of redundant headings, cross-referencing, and use in assignment of subject headings.

The next step, just begun, involves preparation of models of punched-card or tape-storage presentation of entire bibliographic citations and abstracts in order to expedite punching and taping Volumes I–XIII (90,000 abstracts) of *Meteorological and Geoastrophysical Abstracts* and current punching of future issues for indexing and retrieval.

These projects are partially supported by the U. S. Air Force, U. S. Weather Bureau, U. S. Army, the National Aeronautics and Space Administration, and the National Science Foundation. It is expected that by 1964 most of the above program, including machine preparation of cumulative indexes, will be operative.

References:

(1) American Meteorological Society. *Meteorological and Geoastrophysical Titles,* ed. by Malcolm Rigby, vol. 1, no. 1, April 1961; vol. 1, no. 2, October 1961; vol. 2, no. 1, January 1962.

(2) American Meteorological Society. *Alphabetical List of Subject Headings with Corresponding Universal Decimal Classification Numbers for Meteorology (551.5) and Related Subjects.* Washington, D. C., October 1960, 2 vols., 480 p.

AMERICAN MUSEUM OF NATURAL HISTORY 1.8

Central Park West at 79th Street, New York 24, N. Y.

SYDNEY ANDERSON, *Assistant Curator, Department of Mammalogy*

Problems in the retrieval of information from natural-history museums and potential methods of aiding researchers in obtaining information are being studied. A recent survey of collections in North America reveals about 1,500,000 specimens of mammals in more than 300 collections. Valuable material is often not used because of difficulties in obtaining information. The usability of microfilm copies of museum catalogs is being studied. A copy of films of 160,000 catalog entries in the Department of Mammalogy has been made and is being loaned to investigators at other institutions. Feasibility and costs of a punched-card system are being studied.

1.9 AMERICAN PSYCHOLOGICAL ASSOCIATION

1333 16th Street, NW., Washington 6, D.C.
WILLIAM D. GARVEY *and* BELVER C. GRIFFITH

The purposes of this project are: (a) to provide general information about the conduct of scientific work in psychology and the behavior of the psychologist, especially as these relate to the generation of information and needs for information services, and (b) to provide a framework within which the contribution of specific information institutions may be evaluated and recommendations for revisions or additions to them may be made.

During the past year, the objective has been to develop a natural history of scientific information exchange in psychology. A series of studies has been conducted, and others are planned, in an effort to obtain quantitative descriptions of the component parts in the following conceptualization of this process.

Several general functions can be identified in any successful exchange of scientific information: the origination of information, its transmission, storage, and use. The first problems have been to identify the persons and institutions fulfilling these functions and to determine the scope of their activity.

The most difficult areas under investigation include the operation of feedback systems that exist in the activity and their effect, if any, on the output of the source. In general, there appear to be two sources of feedback: the editorial filtering and transformation of the information, i.e., editors and program committees, and the user. Presumably, these sources of feedback help shape the output of the originators of research.

Finally, there is the description of the user's behavior: his access to the information, his decision as to the relevance of information, the type of information he uses, and the manner in which he uses it.

It would appear from the results of the studies to date that within this scheme for information exchange there is the possibility of obtaining data that will enable the science and profession of psychology to adjust its production and dissemination of material for more effective scientific communication.

The project is supported by the National Science Foundation.

Reference:

(1) Garvey, W. D., and B. C. Griffith. "The NSF–APA Project on Scientific Information Exchange," presented at the 1962 Annual Meeting of the American Psychological Association, St. Louis, Mo., September 1962.

ARTHUR D. LITTLE, INC. 1.10

Acorn Park, Cambridge 40, Mass.
VINCENT E. GIULIANO, *Principal Investigator*

A project is underway to investigate the degree of centralization or decentralization which should be planned in designing distribution, indexing, retrieval, abstracting, and other literature processing mechanisms best suited to the needs of the scientific community. An objective of this study is to provide means for identifying needs for new or revised information center facilities; a closely related objective is to provide means for identifying duplications in existing facilities.

The first phase of the study has involved (a) operational analyses of some existing document processing systems; (b) development of mathematical models relating to the effectiveness of document searching procedures and indexing systems, the cost of documentation systems, and informational network interrelationships among scientists, libraries, and document centers; and (c) a review of the art of information processing.

The second phase of the study will involve continued effort along all three lines, with major emphasis on the further development of mathematical models to provide a framework for evaluating various concepts of literature indexing, abstracting, storage, dissemination, and retrieval. Models will be tested with representative data.

A promising line of investigation appears to be the development of information network models which regard individuals, information centers, documents, etc., to be associated portions of an overall network configuration. A very simple model of this type has been proposed for identification of information centers according to their technical orientation. A small electrical analog network has been built to test this model; it deals with libraries and topics drawn from the area of political science. The behavior of this network appears to be promising, and larger scale tests are planned.

The project is being carried out under contract to the National Science Foundation.

Reference:

(1) Giuliano, V. E., and P. E. Jones. "Linear Associative Information Retrieval." (To be published)

ASSOCIATION OF SPECIAL LIBRARIES AND INFORMATION BUREAUX 1.11

3, Belgrave Square, London, SW. 1, England
MOIRA PHILLIPS, *Principal Investigator*

A project has been undertaken with the aim of examining the adequacy of literature searching by scientists prior to starting experi-

mental work for new projects, the difficulties they encounter, and the extent to which assistance from librarians and information officers is available and used. The investigation will comprise a postal questionnaire survey concerning literature searching habits, followed by a diary survey of noncurrent literature read by scientists who returned questionnaires. Approximately 1,000 scientists from industrial concerns, universities, and government establishments will be approached.

1.12 ASSOCIATION OF SPECIAL LIBRARIES AND INFORMATION BUREAUX

3, Belgrave Square, London, SW. 1, England
M. SLATER, *Principal Investigator*

A study is underway of the use made of library/information services in industrial and academic organizations. The broad aim of the program is to investigate the information requirements of users of such services, the practical use to which the information is put, and the amount of library time and skill involved. An additional aim is to test a hypothesis that library users can be classified into groups possessing recognizable and predictable needs and behavior patterns.

The method employed is a questionnaire survey, supplemented by observation of each library/information service. The sample consists of 12 to 20 organizations in or near London. Users of library/information services are asked to fill in a questionnaire relating to a specific visit or request made to the service. It is hoped that the results of this study will have direct applicability to the more efficient and realistic organization of new services and the reorganization of existing services.

The survey began on June 27, 1962. By early August, 220 questionnaires had been obtained from 8 organizations. Analysis of these questionnaires is already in progress. It is anticipated that a report will appear early in 1963.

1.13 ASSOCIATION OF SPECIAL LIBRARIES AND INFORMATION BUREAUX

3, Belgrave Square, London, SW. 1, England
I. D. WHYTE, *Principal Investigator*

The investigation of the transmission of self-published United Kingdom research results has been discontinued.

1.14 BIOLOGICAL ABSTRACTS

University of Pennsylvania, 3815 Walnut Street, Philadelphia 4, Pa.
G. MILES CONRAD, *Director*

Projects which are continuing, with no change reported since the previous statement, include (a) work on improving the utility of

B.A.S.I.C., a permuted title index; (b) studies of the problems of establishing specialized information services; and (c) research into the vocabulary of biologists. [*Ed.*].

BUREAU OF SOCIAL SCIENCE RESEARCH, INC. 1.15

1424 Sixteenth Street, NW., Washington 6, D. C.
IVOR WAYNE, *Project Director*

A survey of users of the metallurgical information searching service at Western Reserve University's Center for Documentation and Communication Research (see 2.128) has recently been completed. Several dimensions of user satisfaction-dissatisfaction, such as speed, coverage, relevance, and cost, were investigated and the findings presented in the final report (1).

The project was sponsored by the National Science Foundation.

Reference:

(1) Bureau of Social Science Research. *A Survey of Users of the American Society for Metals–Western Reserve University Searching Service,* BSSR Report No. 352, July 1962, 42 p.

CHEMICAL ABSTRACTS SERVICE 1.16

The Ohio State University, Columbus 10, Ohio
G. MALCOLM DYSON, *Director of Research and Development*

Recent research has resulted in a computer-produced digest of information pertinent to the field of biological action of organic compounds. A sample issue of the digest, entitled *Chemical-Biological Activities* (CBAC), has been issued (1). Designed to speed the flow of information from primary journal to working scientist, CBAC includes a digest of research results within 14 days after publication of an original paper and provides a Keyword-in-Context index to the compound names and conceptual information, together with molecular formula, notation, and author indexes.

Relevant information is extracted from selected primary journals and stored in the system in terms of a machine language. Strict thesauric control is maintained during codification. Using a dictionary tape, the computer translates the units of machine language into a form approximating a normal English sentence, and produces a printout in which a series of such sentences, accompanied by the several descriptions of each compound cited, is listed under the title of the original paper. A modified IBM 1401 computer, with a 1403 printout in 120 characters, is used.

After further research and market studies, it is planned to publish CBAC biweekly, with cumulative indexes issued on a semiannual or annual basis.

Research leading to this publication was partially supported by a grant from the National Institutes of Health.

References:

(1) Chemical Abstracts Service. *Chemical-Biological Activities,* Sample Issue, September 1962.

(2) Dyson, G. Malcolm, and Michael F. Lynch. "Chemical-Biological Activities, a Computer-Produced Express Digest," presented at the meeting of the American Chemical Society, Atlantic City, N. J., September 1962. (To be published in *Journal of Chemical Documentation)*

1.17 COLUMBIA UNIVERSITY

Bureau of Applied Social Research, New York 27, N. Y.
PAUL F. LAZARSFELD

A project has recently been initiated to identify, describe, and evaluate specialized social science information services in the United States. One objective of the study is to obtain a better understanding of the structure of the U. S. network of social science information services and the way in which researchers and others use the services.

The survey will be conducted by a questionnaire to all services identified as being within the scope of the project. Inquiries will include questions on subject coverage, size and type of collection, information services, user restriction, and publications. A directory of such services, based on the survey, will be prepared and published at the completion of the project.

The study is supported by the National Science Foundation and is expected to be completed in late 1963, with publication of the directory in 1964.

1.18 CREEDMOOR INSTITUTE FOR PSYCHOBIOLOGIC STUDIES

Station 60, Queens Village 27, N. Y.
JOHN R. WHITTIER *and* HENRY BLACK,
Co-principal Investigators

A study is being made of the literature needs of neuropsychiatric research workers.

Information is gathered by unstructured interviews, averaging an hour in length, with 40 to 50 full-time research workers and research directors. To provide some comparison or contrast, 10 or 12 people working in fields other than neuropsychiatry will also be interviewed.

As of August 1962, field work was about 75 percent completed. It is anticipated that the final report will be ready for publication be-

fore the end of 1962. The main point that has emerged is the very great need for additional secretarial help and skilled personal bibliographical assistance. Electronic machines and/or complex organizations based on them appear to be largely irrelevant. Some peripheral information on the distribution and use of reprints by scientists is being collected and will probably be reported separately.

The project is supported by the National Institutes of Health under a 1-year grant which expires in October 1962.

GENERAL ELECTRIC COMPANY 1.19

Applications Section, Computer Department, Phoenix, Ariz.
MARSHALL SPANGLER, *Principal Investigator*

As part of a continuing effort in the investigation and development of information storage and retrieval systems, there is a current project to study indexing problems with particular emphasis on mechanized, or automatic, indexing.

In conjunction with the preparation of bibliographies, a program has been written for the GE-225 computer to process text lines either from titles or running text, as in abstracts, in order to produce an alphabetical list of significant words in context with adjacent words of the text from which they were derived. This is commonly known as a permuted title index.

Each word in the alphabetical list is tagged with its document number or identification number if titles are being processed, or with page, paragraph, and/or line of a particular document if running text is being indexed.

If descriptors or keywords are desired for preparation of coordinate indexing systems, the program will generate punched cards containing keywords only. Each card will contain one keyword, together with identification or document numbers for each reference in which that keyword appears.

A further purpose of the program which is being perfected is to arrange bibliographical data into a given format and provide printout on the computer's high-speed printer. This printout can be made directly on reproducible media. Further processing can yield listings of document numbers by year, by authors, and by sources.

A contemplated extension of this project is to generate indexes by means of word analysis, taking into consideration syntactic and semantic aspects of text lines.

The computerized indexing and bibliographic system is currently undergoing tests.

Reference:

(1) Spangler, Marshall. *General Bibliography on Information Storage and Retrieval,* General Electric Company Technical Information Series, Computer Department, Phoenix, Ariz., R62CD2, March 1962; revised October 1, 1962.

1.20 GENERAL ELECTRIC RESEARCH LABORATORY

Project Analysis Section, Research Application Department, Schenectady, N. Y.

W. ADAIR MORRISON *and* MARY T. WRIGHT

An analysis of scientific papers published in selected research journals in 1961 has been completed, and a report is available (1).

Twenty-three journals, selected to give good coverage of publications in physics, chemistry, and metallurgy and ceramics, were examined, and the papers published in them in 1961 were classified by scientific area of subject matter, by the organization from which they originated, by the type of institution—university, governmental, industrial, or nonprofit laboratory, or individual—in which they originated, by the sources of support acknowledged, and by the State of origin. For the combined areas of science and for the three individual areas, the number of publications from each organization is given, numbers and proportions from the different kinds of institutions and from the different States are compared, and numbers and proportions of papers acknowledging Government support are noted. The results of this analysis are compared with those of the previous surveys of the same journals—1957, 1958, and 1960. The major results are presented graphically; additional data and details are given in extensive appendixes.

Trends seen in previous surveys persist: An increasing proportion of publications acknowledge Government support; an increasing proportion of the papers come from industrial organizations.

References:

(1) Wright, Mary T., and W. Adair Morrison. *An Analysis of Selected Scientific Journals for 1961,* Report No. 62-RL-3078A.

(2) Morrison, W. Adair, and Mary T. Wright. "Origin and Support of Scientific Publications," *Journal of Metals,* vol. 14, May 1962, pp. 334-335.

1.21 GEORGIA INSTITUTE OF TECHNOLOGY

Engineering Experiment Station, Technical Information Section, Atlanta 13, Ga.

DALE L. BARKER, *Project Director*

Literature citations from Russian chemical journals are being ex-

amined statistically to identify and characterize the literature cited. Ten thousand citations have been collected and are being analyzed. The study will indicate Soviet chemists' use of non-Soviet literature, the relative importance of various Russian journals, and geographical and institutional characteristics of Soviet chemical research. The work is about 80 percent complete.

GEORGIA INSTITUTE OF TECHNOLOGY 1.22

Price Gilbert Memorial Library, Atlanta 13, Ga.
DOROTHY M. CROSLAND, *Project Director*

A study to evaluate various possibilities for improvement in the training of individuals to staff various technical libraries and information centers has been completed.

The project was supported by the National Science Foundation.

Reference:

(1) *Proceedings of the Conferences on Training Science Information Specialists,* October 12-13, 1961, and April 12-13, 1962. Atlanta: Georgia Institute of Technology, 1962.

HERNER AND COMPANY 1.23

1401 K Street, NW., Washington 5, D. C.
SAUL HERNER, *President*

I. Study of Abstracting and Indexing Service for Mental Health Literature (William A. Creager, Saul Herner, Mary Herner, Gregory Abdian, and Willard Fazar, *Principal Investigators*)

A study has been undertaken to determine the feasibility, structure, and possible mode of organization of an abstracting and indexing service encompassing the literature of mental health and subjects related to it. The study is concerned with the scope, character, and sources of materials that would be covered; style, method, and speed of abstracting and indexing; format, production, and frequency of publication; acquisition, maintenance, and availability of abstracted and indexed materials; user attitudes, needs, and preferences; and costs. The first two areas of consideration are presently being studied.

The project is sponsored by the National Institute of Mental Health.

II. Science News Media (Mary Herner, Gregory Abdian, and Saul Herner, *Principal Investigators*)

A compilation and an analysis of science news media in the English language have been made in order to help guide policy regarding the broad dissemination of information on current scientific events (1). The study turned up 242 news publications, 37 of which were news-

papers, 123 of which were newsletters, 80 of which were magazines, and 2 of which were in other formats. The primary subject of the media was medicine.

The project was conducted under contract to the National Science Foundation.

III. THESAURUS USE EVALUATION (Mary Herner and Walter Johanningsmeier, *Principal Investigators*)

An evaluative study is being made of the character and degree of use of *Chemical Engineering Thesaurus* (published by the American Institute of Chemical Engineers) and the index data, in the form of catalog cards, utilizing roles and links, which are being provided for articles published in *Chemical Engineering Progress*. The study will be aimed at both corporate and individual users of the two tools. In addition to furnishing an initial evaluation of the tools, the study will serve as a basis for evaluating roles and links under conditions of actual use.

The project is being carried out under contract to the National Science Foundation.

IV. EVALUATION OF TAPE TYPEWRITER USE (Mary Herner and Walter Johanningsmeier, *Principal Investigators*)

An attempt has been made to evaluate the relative merits of the use of tape typewriters as a means of speeding up the preparation of manuscripts and the diminution of proofreading. The preparation of the body and indexes to the publication *Basic Research Resumés* was used as the occasion of the study (see 2.47). After detailed time studies, cost evaluations, and overall examinations of the resulting products, it was decided that the use of tape typewriters for manuscript typing and duplication is wholly inefficient, and this procedure was discarded in favor of an ordinary electric typewriter with a carbon ribbon. This ordinary typewriter has proved far superior to the tape typewriter in practically every respect.

The study was performed for the U. S. Air Force Office of Scientific Research.

V. STUDY OF GOVERNMENT RESEARCH REPORT SERIES SYMBOLS AND CATEGORIES (Mary Herner, Saul Herner, Gregory Abdian, and Paul J. Mahany, *Principal Investigators*)

The study of the identifying names and symbols used to differentiate Government research report series is in an advanced stage of completion; the ultimate objective is a standardization and, hopefully, a simplification of report symbols and categories. This study involves the definition of the meanings of the most frequently used symbols, the conversion of the meanings into report types, and the description of the uses and circumstances of use of the various types. Symbols, report types, and

definitions of report types are each being compiled by rank order based on the numbers of reports represented. It is hoped that this defining, subdefining, and ranking will help furnish realistic guides for eventual standardization of report types.

The study is sponsored by the National Science Foundation.

Reference:

(1) Herner, M., G. Abdian, and S. Herner. *Scientific and Technical News Media in the English Language.* Washington, D. C.: Herner and Company, August 1962.

INFORMATION DYNAMICS CORPORATION [1] 1.24

2 Lakeside Park, Route 128, Wakefield, Mass.
DAVID P. WAITE, *President*

A recently completed study (1) concerned the selection of a system for producing higher quality typography in the announcement journals issued by the National Aeronautics and Space Administration (NASA) Scientific and Technical Information Facility, Bethesda, Maryland.

An installation and operations cost analysis was made of 13 alternative systems selected from 38 production work-places studied. The appearances of the end product obtainable from each system were examined by studying a wide variety of samples of existing publications and a series of comparable samples furnished by competing equipment manufacturers. One of the surprising results of this study was that the highest appearance-quality publication does not cost appreciably more than the lower quality alternatives when a printing run of approximately 6,000 or more copies is required. After the decision was made to proceed with a system design for mechanized typographic composition, selection was made of the necessary equipment, keyboard layouts, code configuration, and other engineering considerations to enable early start-up of a manual system intended to be meshed later into fully automatic operation from keyboard input, through computer assembly, to phototypesetter page composition.

The work was done under a subcontract of a National Aeronautics and Space Administration contract.

The analysis of specific Government agency user needs for mechanized printing continues, accompanied by study of practical production systems for producing reprinted directories, catalog listings, announcement journals, special bibliographic lists, and other such publications. A final engineering report now in preparation is to be submitted by the end of 1962 to the project sponsor, the National Bureau of Standards.

The transparent 5- by 8-inch microform originally developed for

[1] Formerly Forbes & Waite, Inc.

handling full-text copy of NASA technical reports is being studied to determine its applicability to handling other scientific collections and the possibility of the format's being adopted by several other Government agencies.

Reference:

(1) Berul, L. H. *Selecting a System for Producing Higher Quality Announcement Journals,* prepared under subcontract to No. NASW-3714.

1.25 INSTITUTE FOR ADVANCEMENT OF MEDICAL COMMUNICATION

9650 Wisconsin Avenue, Bethesda 14, Md.
RICHARD H. ORR, *Director*

I. TRANSLATION OF SOVIET BIOMEDICAL LITERATURE (Richard H. Orr, *Principal Investigator)*

A study designed to evaluate the effectiveness of the U.S. Public Health Service's 5-year-old Russian Scientific Translation Program has recently been completed. A major part of this program was the support of cover-to-cover translation of nine Soviet biomedical journals. Three major techniques were used in the study: (a) A sample of 404 American scientists and 103 medical librarians were interviewed; (b) References to Soviet work in American journals and to American work in Soviet journals were analyzed; and (c) Articles in the translated Soviet journals were submitted to the editorial boards of similar American journals for an assessment of scientific merit. The interviews and citation counts were performed by Herner and Company.

A detailed report of the findings is being prepared for publication. The major conclusions can be summarized as follows: Although the translation program has succeeded in helping to make American workers more aware of relevant Soviet biomedical literature, there is little evidence that this increased awareness has been accompanied by more frequent use of Soviet material or by a higher regard for Soviet research in the biomedical field. American scientists apparently continue to rely on their usual journals and on the long-established abstracting and indexing services as their major sources of information about Soviet work, rather than on the special English-language editions of Soviet journals supported by this program. Finally, the editorial standards of the translated Soviet journals are, on the whole, not comparable to those of similar American journals.

The project was conducted under a contract with the National Library of Medicine.

II. PUBLICATION EXPERIMENT (Richard H. Orr, *Principal Investigator*)

In the study of an experimental bibliographic tool for workers engaged in cerebrovascular research (2), the current focus is on determining the effect of step-wise progression from word to concept indexing. The test audience's use of this periodical for current awareness and retrospective searching is being assessed by questionnaires and diaries, which will be supplemented later by interviews.

The work is supported by the National Institutes of Health.

III. INFORMATION FLOW TO MEDICAL PRACTITIONERS (Walter E. Boek, *Research Associate*)

Ninety-four published and unpublished reports of studies on the flow of information to medical practitioners have been collected (1). The information-gathering habits of no other group of basic or applied scientists have been investigated as intensively; however, despite the effort and money expended, analysis of these reports indicates that surprisingly little is known with any certainty about how the practicing physician uses the information channels available to him, and that the methodology required to answer satisfactorily many of the important questions has not yet been developed. Critical reviews of both the findings and the methodology previously reported are being prepared for publication. The work has now entered the phase of designing studies to develop and test new study methods and to obtain data that will fill important gaps in the present understanding of the physician as a user of information.

The project is sponsored by the National Institutes of Health.

References:

(1) Boek, Walter E. "An Annotated Bibliography of Studies on the Flow of Medical Information to Practitioners, Part I," Institute for Advancement of Medical Communication, New York, September 1961; Part II, September 1962.

(2) Orr, Richard H., and Eleanor Crouse. "Secondary Publication in Cardiovascular, Endocrine, and Psychopharmacologic Research," *American Documentation*, vol. 13, no. 2, April 1962, pp. 197-203.

INSTITUTE FOR SCIENTIFIC INFORMATION 1.26

33 South 17th Street, Philadelphia 3, Pa.
EUGENE GARFIELD, *Director, and*
IRVING H. SHER, *Director of Research*

I. CITATION INDEX PROJECT

Over 1 million reference citations have been processed from approximately 75,000 source (referant) articles, taken from 600 source journals

published in 1961. Articles published in 1962 will be processed using new card format and techniques developed from earlier experiments. 1961 material will continue to be processed in order to obtain a complete file for a single target year, thereby creating a research tool which can be used for comparative studies.

Computer programs have been written for various citation editing procedures including standardization of journal abbreviations and transliterated authors' names. Several large-scale statistical analyses will be programmed. This information will be used to conduct research on the "taxonomy" of subject disciplines through SR (Source-Reference) Cycling. The specific fields of medical genetics, molecular chemistry, and psychopharmacology were selected for experiments on defining new fields of science through comprehensive citation indexing. Preliminary studies indicate that statistical information already compiled by the project will be invaluable for various valuations including the selection of journals by small or specialized libraries.

A developmental program for a unified science index, including a comprehensive citation index, has been outlined for processing 10 million citations over a 3-year period. Research on the use of citation indexes for selective dissemination of information is well advanced. This includes studies on a daily citation index service, published in a newspaper format, and a weekly personalized service based on lists of "question citations" provided by the user.

Citation index research is sponsored by the National Science Foundation and the National Institutes of Health.

II. Abstracting Experiments

Readers of the *Index Chemicus* responded very favorably to experimental use of authors' summaries in English, French, or German. The experiments have now become operational. ISI abstracters are now experimenting on preparing summaries of Soviet articles within stringent time schedules. Cooperation from Soviet editors has been solicited.

Miniprint reproductions of authors' abstracts in *Current Contents* are under investigation. In addition, sale of authors' abstracts through OATS (original article tear sheet) service is being tested with Russian and Japanese articles.

III. Miniprint

Comparative studies on the use of full-size text, microcards, microprint, microfilm, and Miniprint have been extended partly because of citation index requirements. Miniprint experiments have been conducted for author indexes and authors' summaries in *Current Contents*. Similar experiments have been tried for reducing costs of authors' reprints.

IV. SCIENTIFIC MEETINGS

Preliminary plans are underway for publication of a new service named *AGENDA, Advance Programs of Scientific Meetings,* to begin early in 1963. It will bridge the gap between original presentation of scientific papers at meetings and the time required for their ultimate publication. It has been estimated that 40 percent of papers presented at meetings are not published even after 2 years (3). Fifty percent of conference proceedings are neither indexed, abstracted, nor listed (1) (2).

References:

(1) Hanson, C. W. and Marion Janes. "Lack of Indexes in Reports of Conferences," *Journal of Documentation,* vol. 16, no. 2, 1960, pp. 65-70.

(2) Hanson, C. W., and Marion Janes. "Coverage by Abstracting Journals of Conference Papers," *Journal of Documentation,* vol. 17, no. 3, 1961, pp. 143-149.

(3) Orr, Richard H. "The Metabolism of New Scientific Information: A Preliminary Report," *American Documentation,* vol. 12, no. 1, 1961, pp. 15-19.

INTERNATIONAL BUSINESS MACHINES CORPORATION 1.27

Advanced Systems Development Division, Monterey & Cottle Roads, San Jose, Calif.

L. D. STEVENS, *Manager, Information Storage and Retrieval Systems, and* D. R. ROYSE, *Manager, Market Department*

Work on the broad market analysis program in information processing continues, with no change reported since the previous statement [*Ed.*].

INTERNATIONAL FEDERATION FOR DOCUMENTATION (FID) 1.28

7 Hofweg, The Hague, The Netherlands

PAUL POINDRON,[1] *Investigator*

A study on the content, influence, availability, and value of scientific conference papers has been completed (1) (2). The study was carried out by FID with the collaboration of the Union of International Associations and the International Council of Scientific Unions.

Conferences held by international organizations concerned with the basic sciences (including applied sciences) during the years 1954 through 1958 were the subject of study.

Among the resulting suggestions for ways to improve the quality

[1] Conservateur en Chef, Direction des Bibliothèques de France

and availability of conference publications were the following. In order to make conference proceedings useful to research workers, particularly those unable to attend the conference, arrangements should be made to (a) publish, if possible, all papers before the conference begins; the text of unpublished papers should be available on request; (b) publish proceedings as soon as possible after the conference, within one year at the latest; and (c) send all conference publications, immediately after they appear, to the secretariats of the international bibliographies concerned so that each individual paper may be indicated or abstracted (using the author's summary). A single calendar should give all particulars of all conferences, and some appropriate means should be found for giving information about publications before they are issued. Union catalogs should show which libraries in a particular country have copies of listed conference papers. Standardization of terminology, style (particularly the layout of the title page), or rules for cataloging conference publications would be helpful.

The study was carried out under a Unesco contract.

References:

(1) International Federation for Documentation. "The Content, Influence and Value of Scientific Conference Papers and Proceedings," *Unesco Bulletin for Libraries,* vol. 16, no. 3, May-June 1962, pp. 113-126.

(2) International Federation for Documentation. "Availability of Scientific Conference Papers and Proceedings," *Unesco Bulletin for Libraries,* vol. 16, no. 4, July-August 1962, pp. 165-176.

1.29 JOHN I. THOMPSON AND COMPANY, INC.

1118 22d Street, NW., Washington 7, D. C.

THEODORE R. O'DONNELL

The study to determine the present degree of effectiveness and possible improvements in disseminating, within the entire Government R&D structure, newly generated technical reports of Government-sponsored research has been completed (1). Primary consideration was given to the expeditious furnishing of new reports to Government R&D activities and their contractors and grantees who have a need for the information at the time it is documented.

The study involved a field investigation of 31 representative Government information activities; an operations research appraisal of information distribution systems, including the development of an analytical model for effective dissemination; an analysis of currently produced technical information; and research involving environmental factors affecting the distribution process.

Conclusions reveal that the present degree of effectiveness is unsatisfactory, and that improvements in the Federal technical information distribution system are possible in several areas. First, primary distribution as it now exists is ineffective and wasteful and should be replaced by an alternate method for providing early awareness of new technical information. Second, announcements of technical information are too bulky, too infrequent, and too thinly distributed; they vary too much in content to be effective. Third, description patterns applied to technical information are, in general, not oriented toward ready recognition of application by the user. Fourth, no apparent effort is made to avoid redundancy in content of different reports within a collection. Finally, there is no coordinated Government-wide policy for the documentation and dissemination of technical reports.

The study was sponsored by the National Science Foundation.

Reference:

(1) O'Donnell, T. R., J. L. Lewis, and J. I. Glendinning, *Federal Government's System for Distributing its Unclassified R&D Reports,* report submitted to National Science Foundation under Contract No. NSF C-244, John I. Thompson and Company, June 20, 1962.

LEHIGH UNIVERSITY 1.30

Department of Psychology, Bethlehem, Pa.
FRANCIS J. WUEST, *Principal Investigator*

A study is underway to develop and validate three methods of analyzing the information requirements and use patterns of scientists.

The methods to be employed are the interview technique, the diary work sample, and the problem solving situation. The Departments of Chemistry and Metallurgy at Lehigh will provide the population for the study. The work is being conducted in three successive phases; one method will be tested in each phase.

In Phase I, all members of the Chemistry and Metallurgy Departments will be interviewed to develop biographical data, a complete description of other professional activities in which each subject engages, an estimate of the proportion of time spent in each of these areas of endeavor, and a description of available information collections. On the basis of this information, further interviewing will be conducted, using the "critical incident" technique (a method for identifying requirements that are of particular importance to the success of a task), in order to develop a description of the information requirements and the information sources used by each member of both departments. Each source of information will be rated on a utility

scale for each of the specific purposes for which it could be used, thus providing a quantified, subjective estimate of the utility of the information sources described.

In Phase II, the diary-work-sample method will be utilized. Using the results of Phase I, a period of time will be selected which is likely to contain activity representative of all types in which the subjects engage. During this period, a sample of the activities of each subject will be monitored to determine the actual frequency and pattern of information use. The subject will note each occasion on which he seeks information in any source other than his memory and supply complete data for each occasion.

In Phase III, a simulated work situation will be used to allow observation of the critical aspects of the subject's behavior under controlled conditions. Each subject will collect the information necessary to solve three common problems in a simulated work environment, e.g., preparation of a lecture, development of a research technique, and answering a question of the type which might be posed to a consultant. Each subject will have access to a collection of information related to the problem and will verbalize a running description of his activities. The problem will be terminated when the subject believes he has an adequate solution. Data to be recorded will include time to reach solution, number and kinds of information sources used, search sequence, percent of total relevant information sources used, and percent of all sources inspected which were relevant.

In all phases, relative frequency distributions will be used to describe the frequency with which the available information sources were used and the general pattern of information use. Comparisons will be made between the results obtained and the interrelationship between the results derived from each method, and the advantages and disadvantages of each will be examined.

The project was begun in September 1962 for a 2-year period with support from the National Science Foundation.

1.31 THE LIBRARY OF CONGRESS

Washington 25, D. C.

HENRY J. DUBESTER, *Chief, General Reference and Bibliography Division*

In order to determine future support levels and the possible need for changes or improvements, a study of user reactions to the *Monthly Index of Russian Accessions (MIRA)* is underway. A limited number of field interviews with known *MIRA* subscribers, selected to reflect the diversity of users and uses, were pursued in order to develop a questionnaire designed to elicit present use and value judgments as well as

estimates of consequences if *MIRA* ceased being published. Questionnaires were then addressed to known recipients and other groups. The second phase of the study will involve the analysis of the returned questionnaires. This phase will be followed by a more intensive sampling of user reactions through a telephone survey. The final report, which will include results from all three phases of the study, is due in October 1962.

Support for the study is provided by the National Science Foundation. The study itself is being carried out by The Diebold Group, Inc. (see 2.30), under contract with the Library of Congress.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1.32

Cooperative Computing Laboratory, Cambridge 39, Mass.
MICHAEL P. BARNETT

Work on computer-controlled printing continues. A new photocomposing program (PC4) has been coded for the IBM 709 computer. It translates source material punched on cards or Flexowriter tape into the code of the Photon machine. It has a more powerful set of control codes than the PC2 system reported previously (1). The photocomposing programs are being used in several experimental studies concerned with the photocomposition of numerical, algebraic, and verbal material produced by a computer.

The work is supported by a grant from the National Institutes of Health.

Reference:

(1) Barnett, M. P., K. L. Kelley, and M. J. Bailey. "Computer Generation of Photocomposing Control Tapes: Part I," *American Documentation*, vol. 13, no. 1, January 1962, pp. 58-65.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1.33

International Abstracts in Operations Research, Cambridge 39, Mass.
HERBERT P. GALLIHER, *Editor*

The experiment with a special type of index, termed the "Digest," in *International Abstracts in Operations Research* continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the National Science Foundation.

Reference:

(1) *International Abstracts in Operations Research*, vol. 1, no. 4, Part I, September 1962.

1.34 NATIONAL FEDERATION OF SCIENCE ABSTRACTING AND INDEXING SERVICES

301 East Capitol Street, Washington 3, D. C.
RAYMOND A. JENSEN, *Executive Secretary*

A study underway has as its objective the development of a national plan for science abstracting, indexing, and reviewing activity in the United States.

After a phase of general orientation with particular emphasis on members of the Federation, work is proceeding in two broad areas. Present methods of distribution and use of abstracting, indexing, and reviewing services by scientists and engineers are being analyzed in detail. Extensive study is also being made of all phases of producer operations, including acquisition, selection, and processing. Special attention is being paid to opportunities for developing more useful products for technical personnel and for effecting more economical methods of production by the services, through increased coordination and cooperation.

The project is supported by the National Science Foundation, and initial phases of the project are being carried out under subcontract by Robert Heller & Associates, Inc. It is anticipated that conclusions and recommendations with respect to a national plan will be completed by early 1963.

1.35 NATIONAL INSTITUTES OF HEALTH

Division of Research Grants, Bethesda 14, Md.
LYNDA L. CAHOON, *Chief, Research Documentation Section*

The first volume of *Research Grants Index,* an annual scientific subject index to current research projects supported by the Public Health Service, was issued in May 1962 (1). In this volume, which was prepared by the photocopy technique, a modification was made of the conventional keyword arrangement of subject headings in order to group together as subheadings certain categories of related information, such as anatomical sites.

Further modifications are being made for the 1962 *Research Grants Index,* now in preparation. Descriptors-in-context will be used instead of project titles. All entries will be reviewed by the authors of the indexed documents before publication. The indexing terms are typed on camera-ready cards from punched paper tape, which is also used in updating procedures. An additional tape is punched simultaneously for experimental input to the Minneapolis-Honeywell 800 computer.

In conjunction with the preparation of the publication, emphasis is placed on the development and current revision of a thesaurus for use as an indexing guide.

Reference:

(1) U. S. Department of Health, Education, and Welfare. *Research Grants Index, Fiscal Year 1961,* Public Health Service Publication No. 925. Washington, D. C.: U.S. Government Printing Office, May 1962.

NEW YORK BOTANICAL GARDEN 1.36

Bronx Park, New York 58, N. Y.
SYDNEY W. GOULD,[1] *Director*

The first volume of the International Plant Index, *Family Names of the Plant Kingdom,* has been published (1). It was prepared by the use of punched cards and data processing machines as a test volume. It is hoped that after the genera are published, the Index proper will consist of the Latin binomials of plants with citations in alphabetical sequence by genera, divided into commonly worked groups, namely the flowering plants, pines, ferns, mosses, algae, fungi and lichens, and bacteria and viruses. Later the names and usages of the economic plants may be indexed and vernacular names in the languages using Western alphabets added. The first volume, known as IPIx-1, lists the names of the families and higher taxa of the Plant Kingdom—perhaps the first such comprehensive listing of the entire kingdom in modern times.

It is anticipated that punched cards of Latin plant names may be used as catalogs of the contents of the major herbaria of the world. In addition, groups of cards can be furnished to experimental taxonomists, or punched cards of the indexes may be of use in the preparation of a new flora, such as is now being planned for Western Europe. Many other applications of these cards are possible, some of which are suggested in the first volume now published.

The project is sponsored by the National Science Foundation.

Reference:

(1) Gould, Sydney W. *Family Names of The Plant Kingdom,* Vol. I of The International Plant Index, New York Botanical Garden and Connecticut Agricultural Experiment Station, 1962. (Available from IPIx Office, 123 Huntington Street, P. O. Box 1106, New Haven 4, Conn. Price: $2:50 for U. S. and Canada; $3.00, foreign)

[1] Address: Connecticut Agricultural Experiment Station, P. O. Box 1106, New Haven 4, Conn.

1.37 ORSZAGOS MEZÖGAZDASAGI KÖNYVTÁR ES DOKUMENTÁCIÓS KÖZPONT [NATIONAL AGRICULTURAL LIBRARY AND CENTRE FOR DOCUMENTATION]

53 Attila ut, Budapest, I., Hungary

A. MIKÓ, *Director*

Plans have been made to pursue a survey of the possibilities for a documentation service of specialists working in large socialist agricultural enterprises. The emphasis of the survey will be on the investigation of current documentation needs of these qualified experts, and on the storage, retrieval, and quick availability of relevant information required in scattered production centers.

This program is supported by the Ministry of Agriculture.

1.38 RESEARCH, INC.

1111 Union Central Tower, Cincinnati 2, Ohio

JOHN SCOTT DAVENPORT

The program to define the communication process in mathematical model form in both closed system (face-to-face) and open system (mass) states continues. The program is still in its primary phase and is expected to be so for some time to come because of theoretical problems which must be resolved. Contact has been established with communications research programs in several universities and with certain experimental psychologists who are pioneering in this area. The groups involved are agreed that program fruition is 3 to 5 years away.

1.39 RUTGERS, THE STATE UNIVERSITY

Graduate School of Library Service, New Brunswick, N. J.

ROBERT F. CLARKE

A study is being made of the impact of photographic copying on the publication of scholarly materials in the United States, with special interest in scholarly materials of limited sales potential.

Actual records of copying and publication for a recent period are being used to learn if photocopying is influencing sales potential of scholarly materials. An effort is being made to determine whether a formula can be established showing the point at which it becomes more economical to subscribe to a journal than to copy from it (excluding binding, storage, and cataloging costs).

The project is supported by the U. S. Air Force Office of Scientific Research.

RUTGERS, THE STATE UNIVERSITY 1.40

Graduate School of Library Service, New Brunswick, N. J.
HELEN JOANNE HARRAR

A study designed to examine and evaluate the philosophy, programs, and operations of each of the three types of cooperative storage warehouses in operation and to recommend optimum measures for achieving their objectives has been completed.

It was found that (a) savings effected by the warehouses through compact storage and simplified cataloging could be realized by individual libraries without the central warehouse if the same storage modes and quality of cataloging were used, and (b) there appears to be a growing trend away from nonlocal cooperative storage, even among participants in these warehouses. Nonstorage gains, such as liberalized lending policies or reduction in duplication among the libraries, do not depend on centralized cooperative storage, compact storage, or simplified cataloging. There appears to be no reason why such gains could not be continued and extended without the cooperative storage warehouse.

The project was supported by the U. S. Air Force Office of Scientific Research.

Reference:

(1) Harrar, Helen Joanne. *Cooperative Storage Warehouses.* Doctoral Thesis. New Brunswick, N. J.: Rutgers, The State University, Graduate School of Library Service, May 1962. 203 p. (mimeographed)

"SHELL" RESEARCH LIMITED 1.41

Thornton Research Centre, P. O. Box No. 1, Chester, England
L. J. B. MOTE, *Head of Technical Information Division*

The operation of a Technical Information Division is being studied with the object of achieving a greater understanding of the various information processes involved. The study also includes the use made of the service by the research staff. The Division, which consists of a Technical Inquiry and Abstracting Service and a Reports Office and Library, serves both a research center concerned with the application of petroleum products and an adjacent refinery. Resources include records of approximately 5,000 inquiries (each taking more than 30 minutes to answer) dealt with over the last 12 years. A faceted classification is being developed for selected aspects of petroleum technology, e.g., performance, properties, processing, composition, etc., and will be used with mechanical selection equipment.

An investigation into the requirements of research scientists for

assistance in information retrieval is continuing. Initially, records of the less simple technical inquiries received over the past 10 years have been analyzed in terms of source, age, and number of documents providing answer, as well as status and profession of scientist concerned. Differences in the needs of research scientists for assistance in information retrieval are being studied. An attempt has been made to relate these differences to the work undertaken in terms of number of subject fields involved, degree of organization of relevant literature, etc.

The pattern of use by individual scientists of the different services provided by an information organization is also being studied.

References:

(1) Mote, L. J. B. "New Technical Information Division Building at Thornton Research Centre," *Aslib Proceedings,* vol. 14, no. 1, January 1962, pp. 11-18.

(2) Mote, L. J. B., and N. L. Angel. "Survey of Technical Inquiry Records at Thornton Research Centre, 'Shell' Research Limited," *Journal of Documentation,* vol. 18, no. 1, March 1962, pp. 6-19.

(3) Mote, L. J. B. "Reasons for the Variations in the Information Needs of Scientists." (To be published)

1.42 THE STANDARD OIL COMPANY

Management Methods Unit, Midland Building, Cleveland 15, Ohio

D. T. H. CAMPBELL, *Project Leader*

Research concerned with the problem of user needs for technical information within the heterogeneous technical operations of a large corporation and how best to meet these needs continues, with no change reported since the previous statement [*Ed*].

1.43 STANFORD RESEARCH INSTITUTE

Menlo Park, Calif.

DOUGLAS C. ENGELBART, *Project Leader*

Work on the project entitled "Augmented Human Intellect Study" continues, with no change reported since the previous statement [*Ed.*]. The research receives partial support from the U.S. Air Force Office of Scientific Research.

Reference:

(1) Engelbart, D. C. *Augmenting Human Intellect: A Conceptual Framework.* Summary Report, SRI Project 3578, Contract AF 49 (638)-1024, Stanford Research Institute, Menlo Park, Calif., October 1962.

STANFORD RESEARCH INSTITUTE 1.44

Behavioral Sciences Research Program, Menlo Park, Calif.
HARRY V. KINCAID, *Principal Investigator*

Phase I of a basic research project on the information requirements of practicing physicians has been completed.

Phase I was strictly an exploratory and feasibility study which included a review of the pertinent literature, interviews with people prominent in the field of information flow to practicing physicians, and a limited number of interviews with practicing physicians.

A number of hypotheses and possible methods of approach were developed. No publications are planned concerning this phase of the project.

It is anticipated that further research will be undertaken in the future.

Work was supported by a contract with *Modern Medicine,* a controlled circulation medical journal.

SURVEYS AND RESEARCH CORPORATION 1.45

1010 Vermont Avenue, NW., Washington 5, D. C.
LIBERT EHRMAN, *Vice President*

A study of *U. S. Government Research Reports* and *Keywords Index to U. S. Government Technical Reports,* a new permuted title index published by the Office of Technical Services, has been initiated to determine the comparative effectiveness of the two publications as announcment mechanisms and to determine user reaction to them. The project will include a study of the economics of production processes.

The project is supported by the National Science Foundation and is scheduled for completion in May 1963.

SYRACUSE UNIVERSITY 1.46

Project IPECAC, Syracuse University Research Institute, B-6 Collendale, Syracuse 10, N. Y.
WARREN B. WALSH, *Project Director, and*
ROBERT R. EASTER, *Principal Investigator*

The study of the use by American scientists of English-language translations of Soviet scientific journals is nearing completion. The cooperation of the users of these translated journals has made possible the gathering of significant data and opinions. Over 3,000 individuals responded to the letters of inquiry. Both the inquiries and the returns correspond percentage-wise with the distribution of U. S. scientists as shown by the National Register of Scientific and Technical Personnel.

More than 6,500 usable comments were made by the respondents concerning their uses of the translated journals, lag-time, costs, and journal contents.

It is anticipated that the final report will be ready for publication before the end of 1962. The finished report tabulates, summarizes, and, to a limited extent, interprets these comments. Display is both verbal and graphic, and the report includes a discussion of the methodology of the study.

The project is supported by the National Science Foundation.

1.47 TUFTS UNIVERSITY

Institute for Psychological Research, North Hall, Medford 55, Mass.

PAUL RONCO, *Principal Investigator*

A critical review is being made of the literature and past research on the effectiveness of technical writing as a means of communication. Primarily oriented to the technical report, the study is concentrated on the variables in the writing and graphic process which have some measurable communication effect on the reader. It is reasoned that significant improvements in writing and graphic processes could in turn improve the effectiveness of a technical message.

It is anticipated that this project should provide results of value to two different groups. First, it will give researchers a critical and comprehensive review of past work, with possible recommendations on areas not covered but worthy of investigation. Second, it should provide to current practitioners in technical writing, editing, and report production a collection of significant past research findings which can be applied at the present time.

Under contract to the National Science Foundation, the project was initiated in July 1962 for a 1-year period.

1.48 UNITED KINGDOM ATOMIC ENERGY AUTHORITY

Information Office, Atomic Energy Establishment, Winfrith, Dorchester, Dorset, England

G. A. BELL, *Leader of Project*

A preliminary survey of the inquiries received by the Scientific Information Office during the first 6 months of 1962 has been completed, and the results will be published. This survey was carried out as a close parallel to the survey conducted by the Harwell Information Office (see 1.49).

The purpose of the investigation was to discover the usage of the information service by the staff of the Establishment, the variation of

use with the technical status of the inquirer, the deployment of the Information Office effort in answering inquiries relative to the technical status of the inquirer, and an estimate of the relative worth of the major tools used by the Information Office in answering inquiries.

Reference:

(1) Blick, A. R., and A. G. Ashby-Pickford. *Some Deductions From Winfrith Information Office Statistics,* AEEW—M200, December 1961 (unclassified).

UNITED KINGDOM ATOMIC ENERGY AUTHORITY 1.49

Atomic Energy Research Establishment, Harwell, Didcot, Berkshire, England

C. S. SABEL *and* J. E. TERRY

Details of inquiries addressed to the Scientific and Technical Information Office at the Atomic Energy Research Establishment, Harwell, are recorded on edge-punched cards. Analyses can be made of the kind of inquiries received, the sources used in answering the inquiries, and the pattern of use of the information service by the scientific staff of a large research establishment.

A new, slightly modified punched card has been in use since January 1962. This card enables the number of successful and unsuccessful uses of a source of information (e.g., a given abstract journal) to be recorded. It is hoped to relate the percentage success of given sources to types of inquiry, length of search time, and other parameters. The analysis of results from these cards will be undertaken when all the punched cards for 1962 are available.

Beginning in January 1963, the Information Offices at Harwell, Culham, and Winfrith (see 1.48) will record inquiries on a card designed jointly.

References:

(1) Sabel, C. S. *An Investigation of Information Office Retrieval Patterns at A.E.R.E. Harwell, Using Edge-Punched Cards.* AERE.-M 617, January 1961. 22 p.

(2) Sabel, C. S., J. E. Terry, and J. H. Moss. "Edge-Punched Card Examination of Retrieval Patterns in Information Offices, and Related Investigations," *Journal of Documentation,* vol. 18, no. 3, September 1962, pp. 111-132.

1.50 U.S. ATOMIC ENERGY COMMISSION

Washington 25, D. C.
EDWARD J. BR..NENKANT, *Director, Division of Technical Information*

Studies continuing, with no change reported since the previous statement, include (a) investigation of various methods of preparing and organizing information for *Nuclear Science Abstracts* and other publications, and (b) development of a pilot system to inventory and describe current research and development projects sponsored by the Atomic Energy Commission [*Ed.*].

1.51 UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory, Livermore, Calif.
FRANK GIALLANZA *and* JAMES KENNEDY

A working program has been written for the IBM 1401 computer to produce a list of citations and a corresponding Keyword-in-Context (KWIC) index. The program is completely flexible. Any citation format may be used with an insignificant word table of infinite size.

Also completed are IBM 1401 programs to produce a Keyword-Plus-Title index, 3- by 5-inch catalog cards for the library, and 3- by 5-inch cross-reference cards and subject authority cards for nine laboratories that belong to the cooperative weapon indexing committee.

Under development is a program which will incorporate the KWIC index feature with the logical operators AND, OR, and IF to enable users to retrieve information using normal English language as input. A limited 1401 version of this program is now in use. A larger, more flexible program is contemplated for the IBM 7090 computer.

1.52 UNIVERSITY OF CHICAGO

Graduate Library School, Chicago 37, Ill.
VIRGINIA R. REED

A study of duplicate indexing and lack of indexing in the literature of metal cutting will be completed in December 1962. The study has consisted of examination of the complete bibliography of metal cutting from January 1, 1943, to December 31, 1956, produced by the Research Information Service of John Crerar Library for the American Society of Tool Engineers (now the American Society of Tool and Manufacturing Engineers). The bibliography was developed through examination of over 30 indexing services. Bibliographies, footnotes, and other references found in the indexed articles were examined for other pertinent articles, and searching and reporting of the current literature revealed still more articles.

Year of publication, language, and type of publication for indexed and unindexed articles were compared. The number of times each article was indexed, the distribution of articles in indexes, the combination of indexes for articles indexed more than once, the ratio of uniquely indexed articles to total articles in each index, and the method by which unindexed articles were found were all examined.

Reference:

(1) Reed, Virginia R. *A Study of Duplicate Indexing and Lack of Indexing in the Literature of Metal Cutting.* Master's thesis. University of Chicago Graduate Library School. (Copies will be available from the Photoduplication Department of the University at a small cost.)

UNIVERSITY OF ILLINOIS 1.53

Institute of Communications Research, Urbana, Ill.
CHARLES E. OSGOOD, *Director*

Work continues on (a) investigating both the structure of the network of psychological journals in terms of reciprocal bibliographical citing and the characteristics of each journal in the network as a source and recipient of such "information," and (b) determining in quantitative terms the connotations that members of the psychological profession have for their own journals, with no change reported since the previous statement [*Ed.*]. The National Science Foundation, through the American Psychological Association (See 1.9), has provided funds for these two projects.

UNIVERSITY OF KANSAS 1.54

University Library, Lawrence, Kans.
EARL FARLEY, *Project Director*

The first issue of a new permuted title index of Slavic journal articles in the social sciences and humanities is being prepared for publication. Such an index will provide rapid and specific access by research specialists in the field of Slavic studies to the pertinent material being received by the University of Kansas.

The index will contain in transliterated form the original wording of titles from about 3,000 articles in current issues of the Library's collection of over 200 Slavic journals devoted to the social sciences and humanities (1). These titles will be processed by a changed and reassembled version of the existing *Chemical Titles* program for the IBM 1401 Data Processing System. Chief among the lists produced is one placing each significant word of all the titles in alphabetical order, surrounded by its original context of up to 60 characters and spaces. This

Keyword-in-Context system, developed by H. P. Luhn (2) and now used in scores of different applications, is adaptable to any subject or language. In this case, Russian in transliteration, but not in translation, is the intended form of output. Because of this specific format and the provision of author, title, and table of contents listings, the index will tend to supplement the other approaches available to users of the *Monthly Index of Russian Accessions.*

Preliminary steps of planning, including arrangements for the program, funds, and staff, have been accomplished since May 1962. A similar index was begun in June 1962 in cooperation with the University of Oklahoma Law Library for its Space Law Collection (see 2.123).

Further work in 1962-63 will complete adaptation of the program, printout, and distribution of the initial issue. Requests to be included in this distribution will be considered. A logical extension in the scope of this system may be the inclusion of titles received by other libraries.

References:

(1) Backus, Oswald P., III. "Recent Experiences with Soviet Libraries and Archives: Uncommon Resources and Potential for Exchange," *College and Research Libraries,* vol. 20, November 1959, pp. 469-473, 499.

(2) Luhn, H. P. *Keyword-in-Context Index for Technical Literature (KWIC Index).* Yorktown Heights, N.Y.; IBM Advanced Systems Development Division, RC 127, 1959.

(3) Farley, Earl. "Future Scholarly Uses in Libraries of Information Retrieval Equipment," *Books and Libraries at the University of Kansas,* vol. 2, no. 2, May 1962, pp. 1-6.

1.55 UNIVERSITY OF WASHINGTON

Department of Electrical Engineering, Seattle 5, Wash.

DAVID L. JOHNSON, *Professor, Electrical Engineering*

The purpose of research underway is an investigation and analysis of those factors fixing the interface existent in man-computer systems. Emphasis is being placed upon three specific areas: (a) human attitudes toward computer use and results, (b) machine learning processes, and (c) memory organization for meaning association and retrieval. All three areas are being examined with respect to their potential in moving or fixing the man-machine interface for various types of problems. The work is interdisciplinary in nature and is being carried out jointly by engineers and a clinical psychologist.

Approximately 200 responses from an attitude survey have been received from people with diverse computer backgrounds. The results

of this survey are currently being statistically evaluated by the IBM 709 computer. This work will be continued by means of personal interviews and more complete studies of the limitations on the effectiveness of computers imposed by human reaction to the devices, and their results in different problem types.

A program for the IBM 709 is being developed for the general solutions of problems in algebra and trigonometry. The purpose of this study is not the development of the learning program per se, but rather the evaluation of various types of prelearning and probabilistic learning techniques in both limited and general machine learning devices.

Preliminary work is underway in a study of meaning retrieval and association in memory. The purpose of this work is not directly a study of human memory processes, but a study of the characteristics required of a computer to efficiently simulate human memory responses. This study is currently considering a two-stage memory, the association within each, and the association and processing between the couple. In the initial investigations, a limited field of information is being considered. Techniques of quantum mechanics are being investigated, along with the spatial and dimensional association approaches.

All of the work described is either in the preliminary or intermediate phase; the research has been underway for approximately 1 year.

The research is being carried out under the auspices of the U.S. Air Force Office of Scientific Research.

WESTERN ELECTRIC COMPANY, INC. **1.56**
Engineering Research Center, Box 900, Princeton, N.J.
D. W. MERRY, *Project Leader*

An investigation is underway to determine the technical information needs of Western Electric and the means by which these needs can be met.

The information requirements of company engineers are being investigated through a sample survey of enginers at five locations. Analysis of the results has just bcgun. A state-of-the-art study is underway to determine if extant information retrieval systems can be adapted to the needs of the company.

Comparative evaluation of mechnical and manual retrieval systems will begin soon, using a limited number of company technical reports as input to the systems.

The efficacy of a coordinate-indexed company apparatus card system is being determined. More than 7,200 different relays are being used in a pilot study which has now reached the program preparation stage.

1.57 WILDLIFE DISEASE ASSOCIATION

2000 P Street, NW., Washington D.C.
CARLTON M. HERMAN *and* DAVID E. DAVIS

The American Institute of Biological Sciences' experiment in publishing *Wildlife Disease* on 3- by 5-inch Microcards continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the Council on Library Resources, Inc., and the National Science Foundation.

1.58 YALE MEDICAL LIBRARY

333 Cedar Street, New Haven 11, Conn.
FREDERICK G. KILGOUR, *Medical Librarian*

I. Studies on Recorded Use of Books and Journals (Frederick G. Kilgour)

Studies are being conducted on the recorded use of books and journals in the Yale Medical Library. Investigations of journal usage are designed to identify heavily used current titles. Studies of book usage have revealed categories of users, and use by date of publication, while further studies are determining subject matter and types of books used.

II. Citation Use Studies (L. Miles Raisig, *Principal Investigator*)

An investigation of the citation use of medical doctoral theses during the period 1850-1960 was completed in April 1962 as a part of the Selective Book Retirement Program, Yale University Library, 1959-62, and will be reported in 1962 and 1963. Results include charts and tables of national and university distribution of 2,452 theses, and of distribution of French and German university theses by the years of publication.

An investigation is in progress on the citation use of medical periodicals covering the period 1951-60. Such journal characteristics as frequency, language, country of origin, and field or specialty will be sought in the estimated 40,000 citations to be analyzed. Journals will be ranked in field or specialty by newly devised indexes of use. This investigation is supported by the National Institues of Health.

References:

(1) Kilgour, Frederick G. "Recorded Use of Books in the Yale Medical Library," *American Documentation,* vol. 12, no. 4, October 1961, pp. 266-269.
(2) Kilgour, Frederick G. "Use of Medical and Biological Journals in the Yale Medical Library," *Bulletin of the Medical Library Association,* vol. 50, no. 3, July 1962.
(3) Raisig, L. Miles, and Frederick G. Kilgour. "The Use of

Medical Theses as Demonstrated by Journal Citations (1850-1960)," to be presented at the Second International Congress on Medical Librarianship, Washington, D.C., June 16-22, 1963.

(4) Raisig, L. Miles. "Mathematical Evaluation of the Scientific Serial," *Science,* vol. 131, May 13, 1960, pp. 1417-1419.

(5) Raisig, L. Miles. "Statistical Bibliography in the Health Sciences," *Bulletin of the Medical Library Association,* vol. 50, no. 3, July 1962.

(6) Raisig, L. Miles. "Citation Use of Medical Periodicals, 1951-1960," to be published in *Bulletin of the Medical Library Association,* 1963-64.

2
INFORMATION STORAGE AND RETRIEVAL

2. INFORMATION STORAGE AND RETRIEVAL

The research and development activities reported in this section are concerned with the design of new information systems, the comparative testing or evaluation of proposed or operating systems, the determination of the theory underlying the design and operation of information systems, and the problems involved in organizing information for storage and searching, including the development of mechanized methods for processing the natural language of documents to provide input for information storage and retrieval systems.

About one-fourth of the organizations in this section appear for the first time reporting new programs underway; however, most organizations report progress on long-range programs, the initiation of new phases of such programs, or the start of entirely new projects. Thirty organizations from 15 foreign countries are reporting on 43 different projects or studies.

Much of the effort is concentrated on systems design, but there is a constantly increasing interest in the testing, evaluation, and comparison of new systems and techniques; in more theoretical studies of systems design; in mechanized processing or searching of full natural-language text; and in problems of indexing vocabulary development and control.

The following summary calls attention only to new work reported in this section for the first time, to projects which have been completed, and to closely related projects which appear in other sections of this report.

DESIGN OF INFORMATION SYSTEMS

In the area of systems design, much of the work being done anticipates the mechanization of storage and searching operations. Much of the design activity, therefore, is focused on problems involved in organization and maintenance of file or machine storage, and the usual feasibility studies and systems analyses which precede the designing of any suitable information system. Library problems and the mechanization of library operations are also included in this section.

File Organization and Maintenance

UNIVAC Division of Sperry Rand Corporation (2.115) recently completed several studies and analyses concerned with the storage and

maintenance of a master "index" and the processing of search requests against it.

Stanford Research Institute (2.105) is designing and constructing an experimental model of the multiple instantaneous response file (MIRF).

Feasibility Studies and Systems Analyses

The National Bureau of Standards (2.82) has completed its preliminary investigation of the information processing needs of the Food and Drug Administration, including the procedures and systems involved in the collection, creation, evaluation, application, storage, and retrieval of chemical and biological information.

The **National Chiropractic Association** (2.85) has a study underway to determine the feasibility of establishing a system for abstracting, classifying, coding, and retrieving information pertaining to all aspects of the human spine and related structures.

Information Dynamics Corporation (2.52) is engaged in an operations research study on the comparative economic analysis of two hypothetical systems: a subject-specialized information services network and a geographically regionalized information services network.

The **Creedmoor Institute** (2.26) is performing a detailed anaylsis of procedures involved in and time required for manual literature searches and will also compare these manual methods with mechanized methods of information handling.

Société d'Économie et de Mathématique Appliquées (2.102) is conducting methodological investigations of information retrieval systems with a view to applying the results to the construction of an information retrieval language.

Library Problems and Mechanization

As a follow-up to its recently completed systems study, the **University of Illinois Library** (2.121) has initiated a program of testing and evaluation of computer-based library technical operations.

Rutgers State University (2.99) is studying the comparative costs of preparing and using book and card catalogs.

The **Bureau of Ships** (2.14) has an information retrieval project which includes the problem of library operations and their automation.

Work on machine-aided production of library records is being conducted independently by **Ben-Ami Lipetz** (2.131).

DESIGN AND DEVELOPMENT OF SPECIALIZED SYSTEMS BASED ON MECHANIZED TECHNIQUES

A number of specialized systems being developed on the basis of mechanized techniques for storing and retrieving information cover a wide range of document types and subject matter in such fields as patents, medicine, law, and engineering. The systems described here,

for the most part, are being designed to store and retrieve documents, bibliographic references, or data. A few systems are concerned with the analysis by machine of stored information and data.

The Bureau of Ships (2.14) of the U. S. Department of the Navy is engaged in a cooperative project with the David Taylor Model Basin (2.28) to design and test an information storage and retrieval system for Buships' reports collection.

The U. S. Naval Postgraduate School (2.113) has developed a semi-automatic bibliographic information retrieval system for technical documents in the applied physical sciences.

In developing a retrieval system for the literature of interest to the pulp and paper industry, the aim of **The Institute of Paper Chemistry** (2.93) and the **Pulp and Paper Research Institute of Canada** (2.93) is to establish compatibility between manual indexes and large machine-oriented systems.

A project at the **Universidad Industrial de Santander, Colombia** (2.116) has as its goal the development of a system for automatic retrieval of bibliographic references relating to the fields of chemical, electrical, industrial, mechanical, and metallurgical engineering.

In a joint project on mechanized information retrieval **F. Hoffmann-La Roche & Company, Ltd.**, and **Sandoz Ltd.** (2.37) of Switzerland have developed a system based on the principle of "codeless scanning."

Hughes Aircraft Company (2.48) has a program underway to collect, index, and abstract the literature on the electrical and electronic properties of materials and to evaluate and compile the experimental data from this literature; a modified coordinate index employing faceting (properties of materials) will be subject to machine manipulation.

The Armour Research Foundation (2.8) is designing a coordinate indexing and retrieval system for a file containing published information on an experimental project related to the properties of fused salts.

Development of a literature storage and retrieval center for the aquatic sciences has been completed at the University of Rhode Island (2.127).

General Electric Company's Computer Department (2.42) has completed the development of a system for automatically extracting intelligence from a large mass of data; the system is now operational on an IBM 7090 computer.

Hatfield College of Technology (2.46), England, is engaged in an extension of the work of P. R. P. Claridge on mechanized indexing which was reported previously by Warren Spring Laboratory, England (see *CRDSD,* No. 9, Statement No. 2.66).

Patent Information Systems

Information for Industry, Inc. (2.53) is making a quantitative analysis

of patent information and patterns of use in chemistry and related fields as criteria for design of a magnetic tape index system.

Euratom (2.22) is preparing to process patents using the keyword method in cooperation with BREVATOME, Paris.

The **U. S. Patent Office** (2.114) has just initiated a project on mechanical searching of patent information on alloys which will employ a hierarchical random-access file for use on an IBM 1401 computer with a disc file. The project to develop a system for storing and retrieving information about resuscitators has been completed and is now in use.

Medical Data and Literature

Brown University (2.13) has undertaken a project to plan for the development of a program for application of modern technology to medical information handling.

The **American Medical Association** (2.6) has a program underway to build a deep index to the literature on drugs and therapy; an IBM 1401 computer will be used in the initial phases of the project.

The National Library of Medicine (2.86) MEDLARS project, previously reported in Section 1, now appears in this section. The project is scheduled for completion in the fall of 1963.

Systems for Legal Literature

In the project of the **American Bar Foundation** (2.2), whose goal is to devise a computer system for organizing and searching legal literature, a pilot experiment is being conducted to produce by computer a thesaurus of significant words based on a corpus of current cases reported in the legal literature.

The purpose of the **Southwestern Legal Foundation** (2.103) project is to develop and test a system for identifying and retrieving documents or portions of documents pertinent to a specific research need embodied in a search request composed of keywords; full text storage and search techniques will be employed.

The **University of Oklahoma** (2.123) is studying a KWIC program to determine its adaptability to retrieval of space law materials in its Space Law Collection.

In independent work being conducted by **Reed C. Lawlor** (2.130) methods are being developed for the application of modern mathematical techniques and computers to the analysis of problems of law.

Automatic Dissemination of Information

The various problems involved in automatic dissemination of information are being studied at **Technische Hochschule München** (2.108).

STUDIES AND EXPERIMENTS ON INDEXES AND INDEXING

Most of the work on problems of indexes and indexing reported in the previous issue is continuing at a steady rate. In this issue, a slight

increase in new projects and studies concerned with indexing vocabulary problems, terminology control, and the relationship of concepts and terms is noted.

Permutation Indexing

Byproduct tape from Flexowriter composition of the **Spencer Chemical Company** (2.104) technical abstract bulletin is being processed by computer to form a printed, permuted, and truncated keyword index for manual access, together with a coordinate card file for machine retrieval.

Citation Index

Ben-Ami Lipetz (2.131), under contract to the American Institute of Physics (see 1.5), is preparing an experimental citation index by recording instances in which authors in a selected small body of foreign-language physics journals cited earlier papers in familiar English-language physics journals; an evaluation of the usefulness of this type of citation index as a reference aid will be made.

Vocabulary Problems

Rutgers State University (2.99) is making a study of the compatibility of three technical indexing schemes which are used by the National Aeronautics and Space Administration, the Armed Services Technical Information Agency, and the U. S. Atomic Energy Commission.

Brown University (2.13) is engaged in work on specifying medical terms and concepts and the proper organization and display of the numerous complex interrelationships between medical terms and concepts.

J. Farradane (2.129) is conducting independent studies on grouping of descriptors and the expression of relations between descriptors.

The **Engineers Joint Council** (2.34) is developing a vocabulary of engineering terminology by collecting, collating, and converting to punchcard form numerous engineering glossaries, dictionaries, subject heading lists, classifications, and thesauri which will eventually form the basis for an engineering thesaurus.

The possibilities of deep indexing in the field of punchcard machines and electronic computers are being studied at **Octrooiraad** (2.88).

Thesaurus Building

The Institute of Paper Chemistry (2.93) and the **Pulp and Paper Research Institute of Canada** (2.93) are developing thesauri for pulp and paper terminology and forestry terminology.

The **American Bar Foundation** (2.2) is using legal case reports as the experimental material to produce a computer-constructed thesaurus by statistical analysis methods.

In connection with the development of a retrieval system for published

information on fused salts, the **Armour Research Foundation** (2.8) is using mechanized means to compile a thesaurus.

The Armed Services Technical Information Agency (2.7) will publish the revised second edition of the *Thesaurus of ASTIA Descriptors* by December 1962.

Human Factors Studies

The National Bureau of Standards (2.83) experimentally tried the Aslib (See 2.9) findings with respect to a fixed limit for indexing time on a 10,000-document sample with the resulting conclusion that a maximum limit of 5 minutes per document is not only satisfactory but also desirable.

The **Institute for Advancement of Medical Communication** (2.54) is conducting a study of the effect of subject-matter background and indexing experience on the choice of descriptors for documents.

A test of inter-indexer consistency conducted by the **National Book League** (2.81) resulted in 70-percent consistency being achieved and only 18 percent of the resulting 30-percent inconsistency being of importance.

MECHANIZED METHODS FOR PROCESSING LANGUAGE AND GRAPHIC MATERIALS

The development of mechanized methods and techniques for processing the language of documents to provide input for information systems is being pursued from a number of different aspects: conversion of the document's terminology to the specialized language of the system, use of natural language of documents as input to the system, linguistic processing, and semantic analysis.

Information that is either in a system or to be put into a system must be analyzed to determine its content, and appropriate identifying marks must be chosen to represent this information content in a language form that is susceptible to machine manipulation. Today there exists no "standard" information retrieval language, and it is quite unlikely that one will be available in the near future. There are, however, several programming languages for general-purpose symbol manipulation which are designed for facilitating the writing of programs for processing natural-language data. One of these languages, COMIT, was developed at Massachusetts Institute of Technology as a user-oriented language based on a notation familiar to linguists which makes it much easier to use. Other symbol manipulating languages receiving widespread use are IPL-V and LISP.

Conversion of Document's Terminology

David Taylor Model Basin (2.28) in cooperation with the **Bureau of Ships** (see 2.14) is developing a formal, retrieval-oriented language

for indexing which employs a simple, stylized notation suitable for mechanical processing.

Société d'Économie et de Mathématique Appliquées (2.102) is constructing a documentation language which is designed to allow the storage and retrieval of very general information and which, though simple, will be more elaborate than keywords.

A **Euratom** (2.29) study of the composition of machinable abstracts has led to identification of three types of abstracts—retrieval, structural, and positions.

Natural Language of Documents As Input

Research on machine search of natural-language text at Thompson Ramo Wooldridge Inc. (2.109) resulted in machine processing of natural-language questions leading to a more effective retrieval than did a similar process in which a human being translated the original question into a search instruction.

Mechanized Searching of Chemical Formulas

Work at **Massachusetts Institute of Technology** (5.3.21) is continuing on the mechanized searching of chemical formulas, both by the use of the Shadow program for linearized formulas and by other methods for formulas that are treated as two-dimensional structures.

AUTOMATIC CLASSIFICATION, INDEXING, AND ABSTRACTING

Three projects on automatic indexing at Thompson Ramo Wooldridge Inc. (2.109) have yielded strong evidence that automatic indexing based on thesaurus techniques leads to acceptably high quality, compared to human processes, and is economically feasible.

Spencer Chemical Company (2.104) has a study in progress on the feasibility of a computer selection of keywords from the abstracted text.

A study of automatic book indexing being conducted by **Rutgers State University** (2.99) is aimed at developing a system for machine indexing and at determining the efficiency of this procedure as compared with manually prepared indexes.

IBM-France (2.64) has planned a series of experiments in automatic indexing of technical texts and technical vocabulary analysis.

A project at **Western Reserve University** (2.128) is aimed at achieving a fully automatic processing of previously generated abstracts.

An automatic linguistic analysis project at **Universität Bonn** (5.3.29) will include experiments in automatic abstracting.

The possibility of mechanical abstracting is being studied by **Humboldt-Universtät zu Berlin** (5.3.10) which is using a bibliographic collection of 300,000 titles in the field of linguistics as experimental material.

STUDY AND DEVELOPMENT OF SUBJECT CLASSIFICATION SYSTEMS, CODES, AND NOTATIONS

The Library Association (2.73) of England is conducting a study of a new general scheme of classification and the production of a pilot scheme.

Technische Hochschule München (2.108) has worked out a classification scheme for information retrieval systems which is based on an actual study of about 100 existing and proposed information retrieval systems. The redundancy of various documentation codes, including generic codes and prime-number codes, is also under study.

In the newly organized **Documentation Research and Training Centre** (2.75) in Bangalore, India, a depth classification for pharmacology is being developed.

Georgetown University School of Medicine (2.43) is developing a machine coding system for historical and physical data on cardiovascular diseases.

Institute for Bio-Medical Computer Research (2.56) is engaged in the development of a case history code for data from hospital charts and patient interviews and in the design of forms for recording this data.

Information for Industry, Inc. (2.53) is studying notation schemes for the chemical and biological dichotomy of U. S. and foreign pharmaceutical patent information.

The work on developing a decimal notation for biochemical information, previously reported by Vanderbilt University (see *CRDSD,* No. 10, Statement No. 2.90), is now being carried out independently by John H. Schneider (2.132).

"Shell" Research Ltd. (1.41) is developing a faceted classification for selected aspects of petroleum technology.

An English edition of a 248-page Unesco report by Eric de Grolier on *A Study of General Categories Applicable to Classification and Coding In Documentation* is now available from Unesco.

TESTING AND EVALUATION OF SYSTEMS AND TECHNIQUES

Rome Air Development Center (2.98) is establishing a controlled information test facility to conduct controlled tests in the area of information storage and retrieval, including comparison of manual versus automatic methods of storing and retrieving textual and graphic information, evaluation of automatically produced abstracts, and the preparation and evaluation of a Semantic Screening Model.

Rutgers State University (2.99) is conducting a study of the comparative efficiency of information retrieval modes.

An evaluative study is being made by **Herner and Company** (1.47) of the character and degree of use of *Chemical Engineering Thesaurus* and

the index data, in the form of catalog cards, utilizing roles and links, which are being provided for articles published in *Chemical Engineering Progress.*

Testing and evaluation of a catalog of compounds and their roles from about 800 polyethylene and polypropylene patents have been completed, and the system is now in use in an examining operation at the U. S. Patent Office (2.114). Data gathered from an experiment that compared the quality and quantity of manual versus machine searching of 800 patents are also being evaluated.

In a joint experiment conducted by **Compagnie de Saint-Gobain** and **IBM-France** (2.23) to find a method for automatic document selection by computer, 92 percent of the documents sought were found and the number of irrelevant documents did not exceed 25 percent.

Farbenfabriken Bayer AG (2.38) is comparing the efficiency of peek-a-boo and edge-punched card systems; determining the time required for obtaining information from a peek-a-boo card system; and determining and comparing the costs involved in the usage of three systems—normal card register, edge-punched cards, and peek-a-boo cards. The studies thus far have revealed that mechanical methods of documentation do not show any advantage over the peek-a-boo system with respect to time and cost.

In a preliminary, informal examination of various legal information problems by the **American Bar Foundation** (2.2), an experiment was conducted in which KWIC indexing of titles was compared with indexing by a subject heading classification system; results showed that 64.4 percent of the title entries contained as keywords one or more of the subject heading words under which they had been indexed and 25.1 percent contained logical equivalents.

THEORETICAL AND MATHEMATICAL STUDIES

Research on a theoretical study of methods and techniques of information storage and retrieval in order to provide a detailed analysis of the theories underlying the design and operation of information systems and to establish a basis for future research and development is underway at **Lehigh University** (2.71).

Phase I of a research program to seek out new and unique solutions to the fact correlation problem has been completed by Electro-Optical Systems Inc. (2.33).

RELATED RESEARCH, EQUIPMENT, AND NOTABLE CHANGES

Much of the research in Section 3 (Mechanical Translation) is related to the projects in this section that are concerned with automatic

analysis of language. The mechanical translation projects are concentrated, of course, on the automatic analysis, or parsing, of the foreign languages under study as the first step in the translation process, whereas the information retrieval projects are concerned principally with the automatic analysis of English. In both cases valuable insight into the nature of language is being gained that will promote further developments.

The linguistic research described in Section 5 (Potentially Related Research) also is related to work in this section that deals with language analysis. Some projects, such as those of **Humboldt-Universität zu Berlin** (5.3.10), Indiana University (5.3.11), the National Bureau of Standards (5.4.14), Radio Corporation of America (5.3.25), **Universität Bonn** (5.3.29), the University of Pennsylvania (5.3.37), and the **University of Pittsburgh** (5.3.38), are being conducted with the needs of information retrieval specifically in view. In other cases, the more immediate aims of the research efforts might differ. However, the knowledge and understanding that are being gained through these programs will undoubtedly prove applicable to many tasks in information retrieval.

The equipment reported as being in use in the projects covered by this section ranges from simple hand-manipulated cards through punch-card tabulating equipment to high-speed computers with large-capacity storage. The IBM 7090 and 1401 computers are most frequently reported as being used in searching and selecting operations. The IBM 1401 computer is the most frequently reported equipment used in index production activities. In some instances the IBM 1401 is used in combination with the larger IBM 7070 or 7090 computers.

A research group at the **U. S. Patent Office** (2.114) has recently been established to explore problems in the areas of linguistics, mathematics, and self-organizing systems.

As a corollary effort to international cooperation in information retrieval among examining patent offices, the **U. S. Patent Office** (2.114), with the assistance of a National Science Foundation grant, is sponsoring a research associates program in which six candidates have been selected from foreign Patent Offices to conduct research for a 2-year period at the U. S. Patent Office.

The Indian Statistical Institute has organized the **Documentation Research and Training Centre** (2.75) in Bangalore, India, "to train documentalists for service within India, and also to extend such facilities to documentalists needed for other countries particularly in the fast-developing Afro-Asian countries." The second function of the Centre is to organize a continuing research program on documentation techniques.

A change in organizational names and the transfer of a statement to another section of this report should be noted. The **Bureau d'Etudes sur**

le Traitement Automatique de l'Information dans les Sciences Humaines has changed its name to **Service d'Etudes Sémiologiques et Documentaires** (Centre for Semiological and Documentation Studies) and now comes under the Maison des Sciences de l'Homme and the Ecole Pratique des Hautes Etudes (VIe Section) (2.76). Forbes & Waite Inc. has changed its name to **Information Dynamics Corporation** (2.52). The work of Harvard University (5.5.3) on the application of computers to content analysis, previously reported in this section, now appears in Section 5.5 (Psychological Studies).

The following organizations have reported the completion of projects, studies, experiments, or research in this issue:

Electro-Optical Systems, Inc. (2.33)
General Electric Company (2.42)
National Biomedical Research Foundation (2.80)
National Bureau of Standards (2.82)
University of Rhode Island (2.127)

ADVANCED INFORMATION SYSTEMS, INC. 2.1

3002 Midvale Avenue, Los Angeles 34, Calif.
R. M. HAYES

Work continues on the study of the organization of large files with self-organizing capability. The work at present is in three areas: (a) The first is intended to tie the analytical results of the previous study (1) to actual data. In particular, statistical data of activity in a large biomedical library are being gathered and analyzed. These data will be used to clarify the several distribution functions defining file activity. (b) The study of search strategies is continuing. Several protocols of actual search behavior have been obtained from reference librarians. The protocols are being analyzed for heuristic processes that could be used in a computer program, or for training human search behavior. (c) Analysis and testing of the two models developed as test beds in the previous work are continuing. A computer program has been written in FORTRAN and run on an IBM 7090 computer which simulates file activity and then tests several features of the word-model reorganizing processes. The bit-processor reorganizing program has been completed in IPL-IV and is being tested.

The project is supported by the National Science Foundation.

Work is continuing on the programming and testing of the model for evaluation of total system effectiveness which was developed under contract to the U. S. Air Force (Rome Air Development Center).

Reference:

(1) Advanced Information Systems Co. *Report on the Organization of Large Files with Self-Organizing Capability* (interim report), National Science Foundation Contract C-162.

2.2 AMERICAN BAR FOUNDATION

1155 East 60th Street, Chicago 37, Ill.
WILLIAM B. ELDRIDGE [1] *and* SALLY F. DENNIS,[2]
Principal Investigators

The aim of the project is to devise a computer system for organizing and searching legal literature.

A pilot experiment is being conducted using as experimental material 5,800 current cases from the *Northeastern Reporter.* Half of the substrate cases will be used to produce a computer-constructed thesaurus of significant words, using statistical methods of analysis. The thesaurus will be represented for the computer as "index-word space," in which each word will have a "definition" expressed as its geometrical coordinates. The remaining cases will be used in an indexing experiment, in which frequencies of significant words will be analyzed with respect to "index-word space," generating the coordinates of "concepts" as indexing terms. Documents will be filed in a modified file under concepts.

The work will be carried out using an IBM 7090 computer.

Experimental questions will be solicited from lawyers and scholars both in practice and in research. They will be asked to search their own questions by conventional means, keeping accurate time records. Half of the conventional searches will be limited to the 5,800 cases in the experimental system, and the other half will embrace the entire *Northeastern Reporter;* this division is intended to produce an estimate of the extent to which the volume of a full system could be cut by deleting older material.

In a preliminary, informal examination of various legal information problems, an experiment was directed at comparing the results of automatic computer indexing of titles by the KWIC system with human indexing using a subject heading system. One source of data consisted of 803 legal research projects and theses titles indexed under a modified form of the Index to Legal Periodicals (ILP) system. The other source of data consisted of 2,625 legal articles indexed under the ILP system. Interpretation of data revealed, among other things, that 64.4 percent of the title entries contained as keywords one or more of the ILP subject heading words under which they were indexed, and 25.1 percent contained logical equivalents. The remaining 10.5 percent of the title entries had non-descriptive titles (3).

References:

(1) Eldridge, W. B., and S. F. Dennis. *Report of Status of the Joint American Bar Foundation–IBM Study of Electronic Methods Applied to Legal Information Retrieval,* ABF Publication, August 1962.

[1] American Bar Foundation
[2] International Business Machines Corporation

(2) Eldridge, W. B., and S. F. Dennis. "The Computer as a Tool for Legal Research," to be published in *Law And Contemporary Problems.*

(3) Kraft, Donald H. "Comparison of Keyword-in-Context (KWIC) Indexing of Titles with a Subject Heading Classification System," presented at the annual meeting of the American Documentation Institute, Hollywood-by-the-Sea, Fla., December 11-14, 1962.

AMERICAN DIABETES ASSOCIATION, INC. 2.3

1 East 45th Street, New York 17, N. Y.

ARNOLD LAZAROW,[1] VAUN NEWILL,[2] *and* JOSEPH IZZO,[3]

Principal Investigators

The feasibility of using modern computer techniques for retrieving diabetes-related literature is being studied.

The world's medical literature is being scanned for material pertinent to diabetes research. A comparative study relating depth of subject description to retrieval efficiency is being implemented. The identical subject material is being indexed at different levels in order to determine efficacy of retrieval versus cost. The following levels are included: National Library of Medicine subject headings, keywords in the title, keywords in the author summary, keywords in the available published abstract, Western Reserve University's telegraphic abstract, and keywords in the total text.

A preliminary consultants' meeting was held in order to develop suitable hierarchial classifications for use with machine retrieval systems. Included were research scientists who are interested in diabetes and have special detailed knowledge in various fields, i.e., anatomy, biochemistry, medicine, ophthalmology pharmacology, and physiology.

The project is supported by a 5-year grant from the National Institutes of Health.

Reference:

(1) Newill Vaun A. "Report of Research and Development of Medical Information Retrieval Projects," presented at the Rochester Conference on Data Acquisition and Processing in Biology and Medicine, July 1962.

AMERICAN INSTITUTE FOR RESEARCH 2.4

410 Amberson Avenue, Pittsburgh 32, Pa.

JAMES W. ALTMAN, *Program Director, and* DAN PAYNE,

Project Director

Guidelines or rules for text reduction were developed and evaluated on a project to advance to its fullest the art of manually preparing

[1] Department of Anatomy, University of Minnesota, Minneapolis 14, Minn.
[2] Department of Preventive Medicine, Western Reserve University, Cleveland, Ohio
[3] Department of Medicine, University of Rochester, Rochester, N.Y.

informative abstracts. In terms of objectives, the study was intended to (a) develop guidelines that result in abstracts which provide maximal support to abstract-users, and (b) develop these guidelines so that they result in reliable, i.e., consistent, abstracts of scientific-technical material. Achieving these objectives would reduce the time required of scientists and technicians in the performance of text-mediated duties without degrading their performance. The amount of textual material to be stored, retrieved, and handled in support of such duties could also be reduced.

The literature and 276 abstracting organizations and services were surveyed for information concerning the preparation of information abstracts. From the results of these surveys, various principles for abstracting scientific-technical material were collated and organized. These principles were broken down into specific abstracting tasks and their related subsections. These tasks were organized into a standard Abstract Form, and detailed instructions were prepared for its use.

The evaluation of the abstracts prepared in accordance with the Form required two separate tests: (a) a test for consistency, and (b) a test to determine the degree to which abstracts support performance, relative to original text. The test of consistency was based upon expert judgment of similarity of amount of information contained in the various versions of an abstract. The judgments were made concerning each of 13 subsections of the Abstract Form which were relevant for the papers abstracted for this study. Each subsection was compared with the same subsections of the other abstracts until all combinations were exhausted. The judgments were concerned with only three possibilities: (a) less information contained in one than the other, (b) information content essentially the same, and (c) more information contained in one than the other. The measure of consistency was the percentage of ratings indicating essentially the same information, with all ratings of all abstracts taken as the base. The test of performance support involved comparing performance using abstracts with performance using full text, under two conditions of time, on a text-mediated task. The task used in the study was general in nature: to comprehend and organize information contained in textual material. The tests employed were prepared by an expert in the subject field, and were designed to test for this comprehension and organization of information. Two general types of items were used: those which required the identification-recognition of information, and those which required the manipulation, or application, of information. The subjects were 20 senior electrical engineering students, divided into 2 groups of 10 each such that mean ability of the groups, as indicated by grade-point average, was similar. The textual material consisted of the six technical papers and the related abstracts. The groups received both text and test material in such

a way that each group read, and was tested on, three abstracts and three original papers, in counterbalanced order. The test material for the original paper and related abstract was identical. Three measures of performance were obtained: accuracy under time limits, accuracy with no imposed time limit, and total test time.

The conclusions were: (a) The abstracting procedure, an Abstracting Form and associated instructions, produces reasonably consistent abstracts. An expert judge rated 13 subsections of 6 technical papers prepared by 3 different abstractors as 88 percent consistent, i.e., contained identical information; (b) The abstracts prepared were a substantial reduction of the original text. Considering the six abstracts used in a performance test (judged as containing the most information, but not necessarily the longest), mean percentage reductions obtained were: 47 percent reduction of words, 28 percent reduction of figures, and 27 percent reduction of equations; and (c) Level of performance, as measured by accuracy on use-tests, supported by abstracts was equivalent to that supported by original text, regardless of test time restriction. However, less total test time was required with abstracts than with full text.

Financial support for the project is provided by the U. S. Air Force (Rome Air Development Center).

References:

(1) Payne, D., Sara J. Munger, and J. W. Altman. *A Textual Abstracting Technique: A Preliminary Development and Evaluation Support.* Pittsburgh: American Institute for Research, August 1962.

(2) Payne, Dan, Sara J. Munger, and James W. Altman. *A Textual Abstracting Technique: A Preliminary Development and Evaluation Support, A Supplemental Volume.* Pittsburgh: American Institute for Research, August 1962.

(3) Strasel, H. C., M. P. Willis, Sara J. Munger, and D. Payne. *Automatic Abstracting Evaluation Support: Study Plan.* Pittsburgh: American Institute for Research, February 1962.

AMERICAN LEGAL DATA PROCESSING ASSOCIATION 2.5

Alcala Park, San Diego 10, Calif.
DANIEL SANTUCCI, *Project Leader*

The project to determine the feasibility of a data processing service for lawyers, and also one for physicians, continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Heard, Leroy, Ronald McEuen, and William Benz. "The Courts and Electronic Data Processing," presented at Round Table

meeting of the Research Projects Committee of the American Legal Data Processing Association, December 15, 1961, San Diego. 18 p. (multilithed)

(2) Santucci, Daniel A., and George Yahn. "Electronic Highways for the National Data Processing Service for Lawyers. Mistakes and Pitfalls in Microwave System Selection. Private Systems versus Common Carrier Systems," presented at the IRE Convention, San Diego, 1961. 22 p. (multilithed)

(3) Santucci, Daniel A., George Yahn, and Leroy Heard. "The Data Processing Services for Lawyers and Physicians. Microfilm and Tele-Facsimile Transmission From Computer Center to Law Offices and Libraries," presented at the IRE Convention, San Diego, 1961. 19 p. (multilithed)

2.6 AMERICAN MEDICAL ASSOCIATION

535 North Dearborn Street, Chicago 10, Ill.

FRED R. WHALEY, *Manager, Drug Documentation*

A program is underway to build a deep index to the literature on drugs and drug therapy. The early stages of this work will be of a research nature. Initial work on the analytical phase will include the organization of terminology, taking into account the depth requirements of the drug evaluation section of the Department of Drugs. The IBM 1401 computer will be used for the initial work on the manipulative phase. Several display techniques for answers to questions will be investigated.

2.7 ARMED SERVICES TECHNICAL INFORMATION AGENCY

Air Force Systems Command, U.S. Air Force, Arlington Hall Station, Arlington 12, Va.

Col. JAMES O. VANN, *Commander*

During the past year, ASTIA has concentrated on increasing the effectiveness of the interchange of scientific information throughout the Department of Defense. Plans have been made and work started on a 19-point program which is aimed directly at meeting the challenge of increasing masses of scientific and technical information. The 19 points are itemized briefly as follows.

1. Revision of the *Thesaurus of ASTIA Descriptors.* This involved a cooperative effort by hundreds of librarians, engineers, scientists, and several related activities both in and outside Government. Publication date for the Thesaurus is December 15, 1962.

2. Multiple-Word Index. Abbreviated but informative entries are being developed as computer printouts for all indexes, bibliographies, and the *Technical Abstract Bulletin (TAB).* This program also includes

the addition of a permuted title index (KWIC) as a fast announcement service to supplement *TAB*.

3. Abstract Standardization. A uniform abstract covering all bibliographic elements for technical reports is being devised for use by originators and authors, thus paving the way for automation on common grounds and large savings in time, expense, and repetitive bibliographic efforts.

4. Cooperative Priority Program. DOD contractor and laboratory engineers, scientists, and specialized librarians, as well as science graduate students, will be given opportunities to enhance their areas of interest by on-the-site efforts in ASTIA.

5. Implementing Field Operations. Direct transmission networks between ASTIA Headquarters and its Field Divisions are being established, together with more timely and complete technical information services at these locations.

6. Classified *Technical Abstract Bulletin (TAB)*. Complete abstracts referring to classified reports will be carried in this edition of the *Technical Abstract Bulletin*. Previously only an unclassified version has been prepared. This program is now operational.

7. Additional Field Divisions. Previously, only four Field Divisions for ASTIA services existed: New York, Dayton, San Francisco, and Los Angeles. Several additional ones are planned under this program. Huntsville, Alabama, has now been added; others contemplated are Boston, Seattle, Houston, Denver, Dallas, St. Louis, Chicago, Orlando, Philadelphia, and a possible office in Europe for DOD ASTIA activities there.

8. Interservice Data Exchange Program (IDEP). This is a program for promoting the interchange of data between all three military services resulting from testing of ballistic missile components and parts in the pre-operational phases. This program is now operational.

9. Contractor and Subcontractor Reports. This is an active acquisitions program aimed at placing in ASTIA all reports generated by 10,000 prime contractors of DOD and their 300,000 subcontractors. The total is over 300,000 technical reports annually.

10. Service to Potential Bidders. ASTIA will make an analysis of proposed projects and provide potential bidders with the latest state-of-the-art knowledge. The purpose is to obtain the most efficient proposals possible in replies to DOD Requests for Proposals.

11. Interchange of Magnetic Tapes. The prerequisite here is an essentially complete automatic data processing system in ASTIA and uniform processing of information. Exchange of compatible tapes will then be feasible and permit rapid interchange of information with least duplication of effort.

12. Contractor Pre-Reports. Through brief statements or abstracts

to ASTIA, prior to report publication, engineers and scientists will make available information otherwise long delayed by the timelags incurred in preparation, editorial, review, and publication functions given the formal report.

13. Technical Analysis and Synthesis. Technical reviews and studies prepared by scientific analysts will be made and published. These will be compiled from current reports, pre-reporting information and DD613s (see No. 15 below) to provide state-of-the-art summaries in high priority R&D programs.

14. *TAB* and Microfilm for Latin America. Unclassified and unlimited information, such as is now provided NATO and SEATO countries, will be made available in Latin America in support of current State Department policies.

15. Directed Expansion of RDT&E Program. Bibliographic control will be established for all DOD project cards (DD613). This will enhance management control of research, development, test, and evaluation programs and provide profiles on these tasks to promote active dissemination of supporting information.

16. Microfilm to NSF-OTS Regional Libraries. Twelve libraries strategically located throughout the country are being supplied complete-text microfilm of all unclassified and unlimited DOD reports for public consumption. This is an addition to the cooperative programs already underway between the Office of Technical Services, the National Science Foundation, and ASTIA. This program is now operational.

17. Research and Experimentation. It is planned to explore the latest techniques and mechanical aids as they may be adaptable to scientific progress. Research within ASTIA itself is high on the order for this program.

18. In-House Printing of *TAB*. In order to make information available sooner, *Technical Abstract Bulletin,* as well as bibliographies, technical reviews, indexes, and other reference aids will be printed at ASTIA.

19. Multiple-Copy Requests. Because of limited resources, ASTIA has previously provided only one copy of a report to its users. Experience shows that engineers and scientists require, on occasion, many copies when studying the same problem. This program is designed to meet that need. Before doing so, it is planned to fully automate the reproduction facilities at ASTIA.

There are over 400 identifiable events on the schedule in order to implement this 19-point program. Activity lines have been layed out using the Program Evaluation Review Technique (PERT) process. Full implementation is planned within the next 5 years. In order to accommodate the program, a large-scale computer system is planned for installation in the near future.

References:

(1) Gillum, T. L., P. H. Klingbiel, C. N. Mooers, and E. Wall. *Philosophy of and Guidelines for Revision of the ASTIA Thesaurus,* November 1, 1961.

(2) Klingbiel, Paul H. *Language Oriented Retrieval Systems,* February 1962.

(3) *ASTIA Guidelines for Cataloging & Abstracting,* June 1962 (Revision).

ARMOUR RESEARCH FOUNDATION 2.8

Technical Information Research Section, 10 West 35th Street, Chicago 16, Ill.

C. R. McCULLY *and* A. P. WENNERBERG, *Project Leaders*

A system of coordinate indexing and retrieval is being designed for a file containing published information on an experimental project related to the properties of fused salts. In addition to current information which is added regularly, the file is based on some 1,500 references selected in a survey of the literature back to 1950. In the survey, *Technical Abstract Bulletin, U. S. Government Research Reports,* ASTIA files, and *Chemical Abstracts* were examined. Current issues of these source materials as well as more than 50 primary journals are surveyed regularly for current information. A second survey will extend the coverage back to 1940.

The future interests and information problems of the laboratory personnel will continue to broaden the scope of the file within the ultimate limits defined earlier as the properties of fused salts. Inasmuch as retrieval terms cannot be defined or even anticipated for the ultimate file, a system of coordinate indexing was judged to be most suitable for the purpose of accommodating an ever-expanding and unpredictable number of terms.

Of the 1,500 references selected in the original survey and filed as annotated abstract cards, copies of 200 articles, reports, books, symposia, tables, etc., have been acquired and indexed to date.

The Termatrex system has been initiated for this file. In addition the IBM keypunch, sorter, and printer are being used in the thesaurus compilation.

During the next year, it is anticipated that the 1,500 documents will be obtained, indexed, and filed. At the end of this period it is anticipated that editing will be complete and the first thesaurus will be ready for use.

The work accomplished thus far has been done in conjunction with the experimental project on fused salts supported by the U. S. Navy, Bureau of Naval Weapons.

2.9 ASSOCIATION OF SPECIAL LIBRARIES AND INFORMATION BUREAUX

3, Belgrave Square, London, SW.1, England
CYRIL CLEVERDON, *Director of Project*

An investigation into the comparative efficiency of indexing systems has been conducted at the College of Aeronautics, Cranfield, and the final report on the testing and analysis has now been published (2). The first part of the work (1) entailed the indexing of 18,000 documents by four different systems, while the test program (2) involved more than 7,000 searches in the resulting catalogs.

The analysis has shown most clearly the interrelationship between the ratio of the recall of relevant documents and the ratio of relevant documents among the total number of documents retrieved. As one figure rises, so the other is bound to fall, and indications are that information retrieval systems generally are operating at a recall ratio of 70 to 90 percent, with relevancy in the range of 8 to 20 percent.

The time for indexing varied from 2 to 16 minutes, but no significant improvement was obtained at indexing times higher than 4 minutes. This timing represented actual time spent on a document, and in a normal working situation may be taken to represent an indexing rate of 8 to 10 documents an hour. Uniterm gave better results than the other systems, but the difference was not sufficiently large enough to justify recommending its use in all circumstances. The major factor militating against a higher efficiency was shown to be human error, largely in the decisions as to the concepts in a document which should be included in the indexing.

In addition to the indexing done by the project staff, some 4,000 documents were indexed by librarians and scientists in 49 organizations, and the analysis of this indexing showed similar trends.

A technique for testing existing indexes with comparatively little effort has been developed and has been shown to be capable of providing valuable information on the efficiency of operation. The results of two such tests have been reported (2).

The main conclusion drawn from the project is that the two most difficult tasks in the operation of an information retrieval system are (a) the decision as to which concepts in a document should be included in the index, and (b) the formulation of the search program. The translation of the concept indexing into the descriptor language is a purely clerical task. Potentially all descriptor languages are capable of retrieving the same information, and the main difference will lie in their ability to improve the relevance ratio by not retrieving irrelevant references. An investigation is now being made of the various devices which (a) can be used with different types of descriptor languages, and (b) will improve the relevance ratio.

This work is supported by the National Science Foundation.

References:

(1) Cleverdon, C. W. *Report on the First Stage of an Investigation into the Comparative Efficiency of Indexing Systems,* September 1960. (Available from the College of Aeronautics, Cranfield, Bletchley, Bucks. Price: $1.40)

(2) Cleverdon, C. W. *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems.* October 1962. (Available from the College of Aeronautics, Cranfield, Bletchley, Bucks. Price: $5.00)

BADISCHE ANILIN- & SODA-FABRIK AG. 2.10

Hauptlaboratorium, Ludwigshafen am Rhein, Federal Republic of Germany

ERNST MEYER

For the coding of chemical structural formulas, a machine was built which makes it possible to scan, with photocells, the formulas drawn on a grid printed on paper and to punch the bonds as holes in punched cards or punched tapes. The positions of hetero-atoms, double bonds, etc., are punched by hand.

From the punched cards produced in this manner, †topological memory units are compiled with an IBM 1401 computer which afford the possibility of a search for any desired structural units. Machine programs must still be worked out for conversion into other codes.

References:

(1) Mooers, Calvin N. *Ciphering Structural Formulas—The Zatopleg System,* Zator Technical Bulletin 59, 1951.

(2) Meyer, Ernst, and Klaus Wenke. "Ein System zur topologischen Verschlüsselung organisch-chemischer Strukturformeln für die mechanisierte Dokumentation" ["A System of Topological Coding of Organic-Chemical Structural Formulas for Mechanical Documentation"], *Nachrichten für Dokumentation,* vol. 13, no. 1, 1962, pp. 13-19.

(3) Meyer, Ernst. "Eine Maschine zur Verschlüsselung chemischer Strukturformeln für die Dokumentation" ["A Machine for Coding of Chemical Structural Formulas for Documentation"], *Nachrichten für Dokumentation,* vol. 13, no. 3, 1962.

† *See* Glossary.

BROWN UNIVERSITY 2.13

Institute for Health Sciences, Providence 12, R. I.

LAWRENCE C. KINGSLAND, JR., *and* GLIDDEN L. BROOKS, *Principal Investigators*

A project has been undertaken to plan for the development of a program for application of modern technology to medical information handling.

Current clinical, medical, and research literature (reports, journal articles, monographs, and textbooks) are being studied in an effort to construct a junior version of a "giant textbook" or "fact book" focusing on current research and clinical medicine at specialist or expert level. This preliminary work, which will furnish concrete current examples of medical information handling problems, involves: (a) preparation of suitable keyword and other depth indexes, (b) careful specification of thousands of important current medical terms and concepts, and (c) proper organization and display of the numerous complex interrelationships between medical terms and concepts. More thorough word study and analysis will become practicable as additional machine-readable medical text and depth indexing tapes are made available.

Long-term goals may include development of still more effective bodies of information which will help (a) bridge the gap between the rapidly developing basic sciences and clinical medicine, (b) develop new insights into the mechanisms of disease (by facilitating examination of the logical framework and revealing now hidden second- and third-order associations), (c) program teaching machines at more advanced levels, and (d) develop effective programs for computer-assisted diagnosis.

The project is supported by the National Institutes of Health.

BUREAU OF SHIPS 2.14

Department of the Navy, Washington 25, D. C.

MARION E. BONNIWELL, *Librarian*

Project SHARP (Ships Analysis and Retrieval Project), a cooperative project between the Bureau of Ships, Technical Information Branch (Code 33), and the David Taylor Model Basin (see 2.28), Applied Mathematics Laboratory, was set up to design and test an automated information storage and retrieval system for the Bureau of Ships' Technical Information Branch's reports collection, utilizing existing computer facilities at the David Taylor Model Basin. The system envisions the retrieval of information through the use of syntactically related descriptive terms and role indicators. The system will have the capability of searching a bibliographic file by source; subject; document security classification and/or distribution limitations; contract, project, and

economy numbers; Bureau of Ships' program number; personal author; title; date; source's report number; ASTIA and/or OTS document number; and combinations of the foregoing. Complete subject search is contingent upon further development of FROLIC (Formal, Retrieval-Oriented Language for Indexing Content) and a companion thesaurus of descriptive terms. In addition, other library and inhouse operations, such as the automated production of *Acquisitions Bulletin,* charge-out-records file, pending requests file, user interest register, printed catalog cards, and statistical records of requests and distribution, are to be included. It is intended to retain parts of the card catalog for searching documents where the use of the computer would not be required.

The entire program is to be carried out in five phases. To date, Phase I, Preliminary Design of the System, and portions of Phase II, Checking out the Preliminary Design, have been completed.

The project was undertaken in June 1961 for a period of 37 months.

2.15 CAMBRIDGE LANGUAGE RESEARCH UNIT

20 Millington Road, Cambridge, England
R. M. NEEDHAM, *Principal Investigator*

Research continues on information retrieval systems; theoretical and practical work has been concentrated on the Theory of Clumps (1) (2). In this work the grouping of points in a similarity space is studied, seeking (a) computational techniques for setting up the similarity space; (b) appropriate definitions of closely associated groups, or "clumps"; and (c) computational techniques for discovering clumps in a given set.

Programs are available for setting up matrices of similarity coefficients for data in various forms, the matrices being up to 1,000 by 1,000. The programs have been improved to take advantage of recent large extensions to the EDSAC 2 computer, and these improved programs have been used for an experiment in anthropological classification (3).

Work is continuing on clump-finding among keywords from the C.L.R.U. information retrieval experiment; this is at present concerned with indexing documents by the clump-membership of their keywords. The resulting tentative index codes are being studied to determine the resolving power of the automatic classification from which they are derived.

Work is also continuing on the construction of a thesaurus-type classification of words, based on the application of clumping procedures to small groups of words which are "near-synonyms" in that they can replace each other in some, but not all, contexts (4). Some 2,000 of these groups, or "rows," have been prepared from the *Oxford English Dictionary,* and these are now being investigated with the same programs as are being used for the information retrieval classification experiments.

These studies are supported by the National Science Foundation.

References:

(1) Needham, R. M. *Research on Information Retrieval Classification and Grouping 1957-61.* Doctoral Thesis. Cambridge, England: C.L.R.U., ML.149, 1961.

(2) Needham, R. M. "A Method for Using Computers in Information Classification," presented at the International Federation for Information Processing Societies conference, Munich, August 1962.

(3) Work described at the Wenner-Gren Foundation Symposium on "The Use of Computers in Anthropology," 1962. (To be published)

(4) Sparck Jones, K. "Mechanised Semantic Classification," in *Proceedings of The First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

C-E-I-R, INC. 2.16

1200 Jefferson Davis Highway, Arlington 2, Va.
HERBERT ROBINSON, *President*

I. MECHANIZED INDEX PRODUCTION (Frank Piovia and Wayne B. Swift)

A project is underway to produce a Uniterm index to chemical patents. Approximately 12,500 chemical and chemically related patents issued in the United States in 1962 will be indexed. The index is updated every 2 months to include such patents issued during the period. Patents are reviewed by qualified chemical indexers, and selected terms are evaluated, keypunched, and converted to tape from punchcards. Term classification, processing, and camera-ready copy preparation are accomplished by an IBM 1401-7090 computer complex. The index preparation constitutes a significant advance over previously available methods for machine preparation of coordinate indexes. Published volumes also include *Patent Gazette* abstracts for each patent indexed.

This work is being done under a contract with Information for Industry, Inc., Washington D. C., and will continue through 1963.

II. COMPATIBILITY OF MATERIALS INFORMATION RETRIEVAL SYSTEMS (Wayne B. Swift)

A project has been initiated to formulate a method for compatibility of materials information retrieval systems. Materials information embodied in several systems at the Materials Central of Wright-Patterson Air Force Base is being studied.

This work is sponsored by the U. S. Air Force, Aeronautical Systems Division, Wright-Patterson Air Force Base.

III. Design Study For An Information System On Information Sources and Services (Donald A. Melnick)

Specific objectives of a recently initiated study are (a) to determine the technical and economic feasibility and practicability of establishing and maintaining a central file of current information and data relating to the world's scientific information sources and services; and (b) to design an operational system by which such a central file can be effectively and economically established, maintained, and exploited. The term "scientific information sources and services," as used here, includes any activity, organization, or mechanism established for the purpose of providing scientists and engineers with information or data for use in the conduct of research and development. Included are primary journals, abstracting and indexing services, special libraries, specialized information and data centers, etc.

Among the several phases of the study, the following are included: (a) determination of the scope of coverage, types of users, and output products required by users of the proposed central file; (b) development of a tentative subject classification and/or indexing system for input items; (c) development of a system for coding input data; and (d) determination of present and future hardware requirements for an effective and economical mechanized system for the file.

The project is being carried out under a 6-month contract with the National Science Foundation.

2.17 CENTRE D'ETUDES NUCLÉAIRES DE SACLAY

Service de Documentation, Boîte postale no. 2, Gif-sur-Yvette (S.et O.), France

R. GUILLOUX, *Chef du Service de Documentation*

The research program conducted by the French Commissariat à l'Energie Atomique (C.E.A.) in the field of information storage and retrieval is aimed primarily at the mechanized preparation of various kinds of conventional and nonconventional indexes.

The first step of this program was to prepare conventional subject and author indexes to the *Bibliographie scientifique hebdomadaire* issued by the Service de Documentation (2,000 titles a week). The system developed is now fully operational. It uses punched tape typewriters (Flexowriter) and punched-card machines with small-size, 40-column cards (ICT-SAMAS) (1). The printed cumulative indexes are intended to be the main tools for information retrieval. Searches, however, can be conducted directly with the punched-card files until the annual indexes are completed. These files will also serve to semi-automatically prepare current bibliographies on subjects in which C.E.A. scientists are particularly interested. Photographic techniques apt to improve the system in this respect are being investigated.

The next step in the program is to develop and test automatic indexing procedures. From the results of an experiment carried out in 1961 with an IBM 1401 computer (2), a set of semi-empirical rules has been devised to process English titles. Within each title, the computer singles out sequences of words which are relatively short but self-sufficing and sorts them in alphabetical order from each of their respective components. Lastly, it condenses the resulting index by pruning away the occurrences with little informative content. Such an index stands halfway between a KWIC index and conventional indexes. A sample index dealing with about 700 recent papers pertaining to controlled thermonuclear fusion and plasma physics was printed and circulated among 100 scientists working in this field to test their response.

A citation index is being prepared in order to test, for information retrieval, its usefulness as a complementary device to the above-mentioned index.

With regard to applied linguistics, the structure of English compounds relative to plasma physics is being investigated with a view to automatic indexing and translation (3).

References:

(1) Chonez, A. "La mécanisation partielle des tâches bibliographiques au C.E.A.," to be published in *Revue Internationale de la Documentation.*

(2) Vandeputte, N. *Traitement sur ordinateur IBM 1404 de textes scientifiques anglais en vue d'études linguistiques et statistiques,* Note C.E.A. No. 369, 1961. 40 p.

(3) Balestic, Ch.-J. *Etude de syntagmes anglais relatifs à la physique des plasmas (Première partie),* note interne AFD 32 (1), December 1961. 38 p.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE 2.18

Section d'Automatique Documentaire (SAD), 23 rue du Maroc, Paris 19e, France

J.-C. GARDIN

Progress has been made in the development of SYNTOL [a set of logico-linguistic rules for expressing and retrieving the subject matter of scientific abstracts; a description of the grammatical and lexical features of the language is incorporated in a report to Euratom (see 2.76, refs. 1 and 3)]. SAD has been concerned with two aspects of studies on SYNTOL: (a) designing an experimental program for the IBM 7090 computer (1) (3), and (b) devising an algorithm for the automatic translation of scientific abstracts (in physiology) from French into SYNTOL (2).

Both studies have been partially supported by contracts with Euratom, ending June 1962; research will continue along the same lines in 1963, with the collaboration of the Service d'Etudes Sémiologiques et Documentaires (see 2.76).

The project on automatic processing of sociological data in "content analysis" has been completed. About 1,000 propositions have been processed, each composed of about 30 words, with some syntactical indications, derived from 150 questionnaires on intergroup relationships in Ghana (4). The program has been written by Mme. Renaut for the Gamma 60 (Cie des Machines Bull); results will be published in 1963.

References:

(1) *Le SYNTOL, étude d'un système général de documentation automatique,* Rapport Final adressé par l'Association Marc Bloch à l'Euratom; Partie 2: *Programmation du SYNTOL sur IBM 7090,* by R.-C. Cros.

(2) *Ibid.,* Partie 4: *Analyse automatique,* by M. Coyaud.

(3) Gardin, J.-C., and F. Levy. "Quelques opérations automatiques fondées sur la grammaire du SYNTOL," presented at the International Congress on Symbolic Languages, Rome, April 1962.

(4) Piault, Colette. *Une analyse de contenu au Ghana, avec exploitation automatique des données sur Gamma 60* (mimeographed report to the Centre National de la Recherche Scientifique).

2.19 CHEMICAL ABSTRACTS SERVICE

The Ohio State University, Columbus 10, Ohio

G. MALCOLM DYSON

Director of Research and Development

For the project on storage and machine retrieval of chemical information, a modified IBM 1401 computer which has a 1403 printout in 120 characters has been installed and brought into satisfactory operation. It uses the CAS Character Set (Model 1) in which uppercase and lowercase letters; normal, subscript, and superscript numerals; and some Greek and italic letters are available, together with a wide range of conventional scientific signs such as brackets, parentheses, plus, degree, and equals. Any character or group of characters can be underlined. This computer has been programmed with the algorithms for detecting any given substructures in a molecule.

The algorithm for universal computation of molform from ciphers has been prepared, and work has started on a method of mechanical ciphering from randomly labeled structures. This has as its objective the utilization of labor at a lower level of sophistication for the entry of notations from the super-collective index to the computer tapes. The

reorganization of the super-collective file has been put in hand and is progressing satisfactorily. It represents a "pre-editing" phase used prior to transfer to tape in order to decrease machine time used in searching.

This project is supported by the National Science Foundation.

COMMISSION ON PROFESSIONAL AND HOSPITAL ACTIVITIES 2.20

First National Building, Ann Arbor, Mich.

WILLIAM H. KINCAID, *Assistant Director,* VERGIL N. SLEE, *Director, and* JOHN H. GRIFFITH, *Special Research Assistant*

Research on the potentialities of computers in the development of methodology useful in handling medical data accumulated under differing classification systems, particularly those systems used in the classification of diseases, operations, and other therapy, continues, with no change reported since the previous statement [*Ed.*]. The research is supported by the National Institutes of Health.

COMMUNAUTÉ EUROPÉENNE DE L'ENERGIE ATOMIQUE (EURATOM), CENTRE COMMUN DE RECHERCHE 2.21

Centre de Traitement de l'Information Scientifique (CETIS), Casella Postale No. 1, Ispra (Varese), Italy

P. BRAFFORT *and* A. LEROY

Research on automatic documentation continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Detant, M., and Leroy. "Elaboration d'un programme d'analyse de la signification," presented at the Troisième Congrès International de Cybernétique, Namur, September 11-15, 1961.

(2) Lustig, G., and K. Meyer-Uhlenried. "Application de l'ordinateur IBM 1401 dans le domaine de la documenation automatique," presented at the Deuxième Congrès de l'Association Française de Calcul et de Traitement de l'Information, Paris, October 17-20, 1961.

2.22 COMMUNAUTÉ EUROPÉENNE DE L'ENERGIE ATOMIQUE (EURATOM), CENTRE COMMUN DE RECHERCHE

Centre de Traitement de l'Information Scientifique (CETIS), Casella Postale No. 1, Ispra (Varese), Italy
K. H. MEYER-UHLENRIED

Within CETIS (P. Braffort, Head), research has been carried on in the automatic documentation field for information storage and retrieval, especially by using the keyword method. An IBM 1401/IBM 7090 computer program has been written for the storage of complete bibliographical units and for the retrieval of documents. Questions are put to the machine as combinations of keywords. The program has already been tested and used (1). The data are stored according to the methods developed by CETIS (2) (3), and at the moment are retrieved according to the coordinate index method. By modifying the same program, cumulative indexes may be produced in an economical way. As far as the possibilities of logical connections of keywords are concerned, the question program will be extended in close cooperation with the Euratom Centre d'Information et de Documentation (CID), Brussels. Processing of patents is being prepared in cooperation with BREVATOME, Paris (4).

I. Indexing and Classification of Abstracts by Machine

Automatic assigning of keywords to given abstracts is being examined as well as the automatic classification of these abstracts. Corresponding IBM 1401 and IBM 7090 programs are in preparation. For test purposes, selected texts from *Nuclear Science Abstracts (NSA)* and from the field of documentation are used.

II. Studies on Preparation of Compatible Input for Mechanized Systems

Another main object is the preparation of scientific information for input to computers, especially by taking into consideration decentralized data collection and centralized data processing on a supranational level. For this purpose a † data processing chain has been developed which consists of the following links: manual methods, application of punched tape equipment, conventional punched-card equipment, and computers (5) (6). In this process the widest possible use is made of the keyword method. On the other hand, methods have been developed for automatic conversion of punched tapes and punched cards of any origin in the IBM 1401 FORTRAN code or in any other code (7-15). The solution of this problem was of some importance as far as the processing of the punched tapes generated on machines with

† *See* Glossary.

English, French, German, or Italian keyboards was concerned. Under a research contract with the Gmelin Institute, Frankfurt (Prof. Dr. E. Pietsch, Head), the use of tape-controlled typewriters and converters in documentation is being studied. The following machines are included in this research program: FRIDEN Flexowriter, BIMA "P," SUPERTYPER 8500, IBM 870 Document Writing System, IBM 047 tape-to-card converter and Systematics C 750 converter.

As a result of these tests CETIS and the Institut für Dokumentationswesen, Frankfurt (Dr. Cremer, Head), suggested equipping the tape-controlled typewriter BIMA "P" with a wiring board allowing any coding in the 8-channel punched tape code (16).

III. Mechanization of Library Operations

In order to rationalize library work, conventional programs for tape-controlled typewriters in connection with punched-card equipment have been developed for the registration of accessions (especially of report literature) and the journal orders and journal accounting.

IV. Studies on Composition of Machine-Usable Abstracts

The composition of machinable abstracts is being studied within several research contracts. A collection of experiment descriptions containing a great deal of quantitative data has led to the distinction of the so-called † "Retrieval Abstract" and the so-called † "Structural Abstract," contrary to the conventional abstracts, within a research contract with the University of Freiburg i.Brsg.—Forstbotanisches Institut (Head, Prof. Dr. Marquardt) (17). Moreover, studies are being made of how the automatically retrieved information should be most usefully treated by the scientist. The basic ideas of the so-called † "Positions Abstract" have been conceived within a research contract with Beratungsstelle für Dokumentation und Information, Bonn (Graf Rothkirch-Trach, Head). The "Positions Abstract" is composed in such a way that the logical relations between its different parts are evident and automatic processing is possible (18).

V. Index and Glossary Compilation by Machine

In the "Documentation of Documentation" field, a cumulative index has been produced of the 12 volumes (1950-1961) of *Nachrichten für Dokumentation,* first as a model, by means of the IBM 1401 program described above (19). Other representative documentation journals are being collected, and one common supercumulative index will be produced. For analyzing the literature on documentation, a glossary of German and English terms plus German definitions has been set up (20). It is planned to publish French, Dutch, and eventually Italian

† *See* Glossary.

versions of the glossary. Another IBM 1401 program had been used for the production of such glossaries in several languages.

References:

(1) Meyer-Uhlenried, K. H., G. Lustig, M. Micarelli, E. Peterhansl, and G. Wächter. *IBM 1401 Program for Automatic Documentation,* July 16, 1962.

(2) Meyer-Uhlenried, K. H. "Richtlinien für die Auswertung von Fachliteratur als Grundlage für eine Automatische Dokumentation," *Rapport CETIS No. 13,* pp. 23-35.

(3) Bernstein, H. H. "Richtlinien zum Aufbau von Literaturkarteien im Hinblick auf ihre Eingabe in elektronische Datenverarbeitungssysteme," *Rapport CETIS No. 13,* pp. 36-45.

(4) Lenoir, M. *Etude de la Mecanographie d'une Documentation Brevets, Compte Rendu de la Réunion CID–DOCA–BREVATOME des 15 et 16 Mars 1962. 15 p.*

(5) Meyer-Uhlenried, K. H. "Das System DOCA—ein Weg zur automatischen Dokumentation," *Rapport CETIS No. 37,* pp. 1-12.

(6) Bernstein, H. H. "Der Lochstreifen als Bindeglied zwischen Steilkartei und Computer in der Dokumentation," *Rapport CETIS No. 37,* pp. 33-42. (To be published in *Nachrichten für Dokumentation)*

(7) Behringer, F. A. "Programmierung des streifengesteuerten Kartenlochers IBM 047 für ein codegerechtes Zusammenspiel mit dem FRIDEN Flexowriter und der Tabelliermaschine IBM 407," *Rapport CETIS No. 21,* pp. 30-60. (To be published in *Nachrichten für Dokumentation)*

(8) Behringer, F. A. "Bau einer Zusatzschaltplatte zur Erweiterung der Programmkapazität der IBM 047," *elektronische datenverarbeitung,* no. 2, 1962; to be published also in *Nachrichten für Dokumentation.*

(9) Behringer, F. A. "Ausbau der IBM 047 zu einem universellen Lochstreifen-Lochkarten-Übersetzer," *elektronische datenverarbeitung,* no. 3, 1962; to be published also in *Nachrichten für Dokumentation.*

(10) Behringer, F. A. "Über das 'Umlautproblem' bei Eingabe deutscher Flexowriter-Lochstreifen in den Streifen-Karten-Umwandler IBM 047," *Rapport CETIS No. 41,* pp. 3-14. (To be published in *Nachrichten für Dokumentation)*

(11) Behringer, F. A. "Einige Trickschaltungen zum Porgrammieren des Lochstreifen-Lochkarten-Umwandlers IBM 047," *Rapport CETIS No. 41,* pp. 15-45. (To be published in *Nachrichten für Dokumentation* and *IBM-Nachrichten*)

(12) Behringer, F. A. "Umcodierungsschaltungen für die IBM 047 zur Übersetzung von Flexowriter Lochstreifen in FORTRAN-

Lochkarten," to be published in *Nachrichten für Dokumentation.*
(13) Vardy, P. "Flexowriter–IBM 026 FORTRAN Code-Umwandler," to be published in *Nachrichten für Dokumentation.*
(14) Mühlen, W. "Lochstreifenverabeitung mit der Maschinenkombination IBM 047-IBM 884. Schaltung zur Umcodierung von Lochstreifen," to be published in *Nachrichten für Dokumentation.*
(15) Conze, K. "Programmbeschreibungen für Lochstreifenschreibmaschinen zur Erfassung bibliographischer Angaben." (In preparation)
(16) "Lochstreifen-Schreibmaschine BIMA mit Code-Umwandler," *Nachrichten für Dokumentation,* no. 2, 1962, pp. 97-98.
(17) Geinitz, D. "Retrieval Abstract und Strukturreferat," to be published in *Nachrichten für Dokumentation.*
(18) Rothkirch-Trach, Graf. "Positionsreferat," to be published in *Nachrichten für Dokumentation.*
(19) Buntrock, H. "Ein kumulativer Index für die 'Nachrichten für Dokumentation,' Jahrgang 1950-1961," to be published in *Nachrichten für Dokumentation.*
(20) Buntrock, H. "Dokumentationslexikon," to be published in *Nachrichten für Dokumentation.*

COMPAGNIE DE SAINT-GOBAIN 2.23

Centre de la Villette, 52, bld de la Villette, Paris 19e, France
and
IBM-FRANCE
5, Place Vendôme, Paris 1, France
M. FOURNIER,[1] M. L. DERIBERE-DESGARDES,[1]
G. PICOT,[1] *and* F. LEVÉRY[2]

An experiment was carried out to find a method for automatic document selection by computer. The documents used were abstracts of periodical articles and patents prepared during a 4-month period by the Chemical Documentation Service of the Compagnie de Saint-Gobain.

The method consisted of representing the documents on a magnetic tape index. An unlimited number of keywords, with no indication of their hierarchical relationship, were assigned to the documents; 10,000 documents were analyzed with an average of 8 keywords per document. The set of keywords used constituted a lexicon wherein the synonymous keywords, or ones of very close meaning, were grouped under the same number. At the end of the experiment, the lexicon included 12,316 keywords grouped under 5,957 numbers.

[1] Compagnie de Saint-Gobain.
[2] IBM-France.

The requests for selection of documents were composed of a series of keywords connected by the conjunctions AND, OR, and NOT. Searches were made on the IBM 705 computer which compared the document numbers assigned to each of the keywords and performed the logical operations indicated by the conjunctions. By introducing an increasing number of keywords in the course of machine operation, more refined answers to one request could be obtained. Thirty requests could be handled simultaneously during the experiment. About 100 requests were processed. A conventional manual search has revealed that 92 percent of the documents sought were found; the number of irrelevant documents did not exceed 25 percent. These results verified the validity of the method and indicated improvements which could be introduced.

References:

(1) Levéry, F., et al., IBM-France, and F. Biet, et al., Compagnie de Saint-Gobain. *Une Expérience de Sélection Automatique,* IBM-France, December 1960. The text of this pamphlet has been reprinted in "Economie et Comptabilité," *Gestion des Entreprises,* vol. 15, no. 54, June 1961, pp. 2641-2663, and in *La Documentation en France,* vol. 29, no. 3, May-June 1961, pp. 3-35.

(2) Picot, G., M. L. Deribere-Desgardes, and F. Levéry. "Une Expérience de Selection Automatique de Documentation," *Revue Internationale de Documentation,* vol. 29, no. 1, February 1962, pp. 8-13, and *Bulletin des Bibliothèques de France,* vol. 7, no. 4, April 1962, pp. 185-205.

(3) IBM. *General Information Manual. An Information Retrieval Experiment by Saint-Gobain Company Paris, France.* IBM Data Processing Division, 1962.

2.24 COMPUTER ASSOCIATES, INC.

44 Winn Street, Woburn, Mass.

T. E. CHEATHAM, JR., *and* S. WARSHALL

Research on "adaptive retrieval," i.e., the design of files capable of reorganization on the basis of the monitoring of retrieval requests received, continues.

Most of the work to date has been oriented toward a retrieval system in which elements of the corpus of allowable data are small simple facts for which the abstracting problem is relatively trivial (although optimal file organization is not). The difficulty in these systems is, specifically, the development of a †query translator so powerful that,

† *See* Glossary.

even if the file is fairly complex, many of the customary linguistic constraints in the structure of queries may be significantly relaxed.

In recent months, effort has been concentrated primarily on the development of a general-purpose compiling facility which will permit inexpensive experimentation with both variation of the basic source language and techniques of response to minor violations of the formal syntax of the language.

References:

(1) Cheatham, T. E., Jr., and S. Warshall. *Translation of Retrieval Requests Couched in a "Semi-Formal" English-Like Language,* CA-61-9, October 1961.

(2) Cheatham, T. E., Jr., and G. O. Collins, Jr. *Design Concepts for a Programming System to Support a Large Scale Operations Center,* CA-61-1, May 1961.

(3) Cheatham, T. E., Jr., and Gene F. Leonard. *Exercise and Evaluation of Command and Control Systems,* CA-61-2, May 1961.

(4) Warshall, S. *A Syntax Directed Generator,* CA-61-7, December 1961.

CORNELL AERONAUTICAL LABORATORY, INC. 2.25

P.O. Box 235, Buffalo, N.Y.

PAUL G. ROETLING, *Project Engineer*

The graphic format data handling study continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the U. S. Air Force (Rome Air Development Center).

CREEDMOOR INSTITUTE FOR PSYCHOBIOLOGIC STUDIES 2.26

Station 60, Queens Village 27, N. Y.

HENRY BLACK, *Library Consultant*

The limitations of bibliographical machines are being investigated.

The first part of the project is a detailed analysis of the procedures involved in and time required for handling 12 or 14 fairly extensive literature searches that arose in connection with the routine work of the Institute. Special attention is given to the questions of whether it would have been technically feasible to handle these searches by means of computer-centered bibliographic machines and whether, in the light of time actually required by conventional methods, it might have been practical to do so. The report on this is substantially completed and will be ready for publication in September 1962.

The second part consists of a critical examination of some of the policy, organizational, accounting, and financial problems that would

arise if an attempt were made to supply bibliographic services by a computer-center organization to any major field. This is necessarily largely a "think piece," but reflects an effort to outline difficulties which, insofar as can be determined, have not been seriously discussed anywhere. Formal work on this paper was begun in August 1962; it is anticipated that it will be ready for publication in late 1962 or early 1963.

2.27 DATATROL CORPORATION

8115 Fenton Street, Silver Spring, Md.

HUGH P. DONAGHUE, WILLIAM HAMMOND, *and* STAFFAN ROSENBORG

Investigations are directed toward optimum exploitation of electronic data processing techniques for information storage and retrieval. Two major projects are currently underway.

I. Automated Legal Retrieval

In cooperation with the George Washington University, Graduate School of Public Law, a pilot system has been developed for automated legal retrieval. Input to the system consisted of statutes, case histories, briefs, and other legal documentation in the fields of merger and monopoly.

The novel search strategy was demonstrated at the annual meeting of the American Bar Association in San Francisco in August 1962, using an IBM 1401 computer (1). Prior to the demonstration, the legal vocabulary employed in indexing was analyzed using an IBM 7090 computer to determine the relationship of statistically associated terms.

Further studies are contemplated, using factor analysis and other statistical methods. While the association factor, used in the present strategy to expand the search, gives the relationships of one term to another, a factor analysis study will permit determining the relationship of a single term to a group of terms. From this, it will be possible to learn how terms cluster when related to the same concept.

II. Convertibility of Indexing Vocabularies

The immedate objective of the second project is to generate sufficient data to determine if it is feasible and practical to develop a table of equivalents that will permit automatic conversion from one technical indexing terminology to another.

An experimental study just completed (2) included the compilation of a table of indexing equivalents for the descriptors used by the Armed Services Technical Information Agency (ASTIA) in terms of the subject headings used by the U. S. Atomic Energy Commission (AEC). Further studies in this area are continuing.

This work was supported by the National Science Foundation.

References:

(1) "A Search Strategy for Legal Retrieval—A Joint Project by Datatrol Corporation and the Graduate School of Public Law, The George Washington University," presented at the Annual Meeting of the American Bar Association, San Francisco, Calif., August 1962.

(2) Datatrol Corporation. *Experimental Study of Convertibility Between Large Technical Indexing Vocabularies—with Table of Equivalents,* August 31, 1962. Final Report on Contract NSF C-259.

DAVID TAYLOR MODEL BASIN 2.28

Applied Mathematics Laboratory, Washington 7, D.C.

THOMAS S. WALTON *and* MILTON SIEGEL

The objective of the project is to investigate effective techniques for indexing and searching a large collection of technical documents, and to develop an automated information storage and retrieval system for the Technical Information Branch of the Bureau of Ships (see 2.14).

The principal effort has been directed toward the construction of a general indexing language for expressing the subject content of scientific and technical documents in unambiguous terms. The grammatical structure for such a language has now been developed (1). This language, called FROLIC (Formal, Retrieval-Oriented Language for Indexing Content), employs a simple, stylized notation suitable for mechanical processing.

Effort is currently concentrated on the compilation of a basic vocabulary of about 800 generic and specific terms. These are assigned to a small number of general categories (2) and are arranged in hierarchies to form the core of a thesaurus for the system. This will guide indexers and searchers in the proper choice of terms for descriptions and queries. Furthermore, the complete set of word relationships will be accessible to the searching program to permit mechanical simulation of some of the intellectual activities of human searchers.

The next phase of the project will be the indexing of a sample of approximately 1,000 documents in order to evaluate the overall effectiveness of the procedures. It is expected that programs for maintaining and searching files of document descriptions will be developed for the IBM 1401 and 7090 computers.

The Bureau of Ships, U. S. Department of the Navy, provides financial support, and the project is scheduled for completion during fiscal year 1964.

References:

(1) Walton, Thomas S. *A Formal, Retrieval-Oriented Language for Indexing the Subject Content of Documentary Material* (a forthcoming David Taylor Model Basin report).

(2) Vickery, B. C. "Subject Analysis for Information Retrieval," in *Proceedings of the International Conference on Scientific Information,* Area 5, pp. 41-51. Washington, D. C.: National Academy of Sciences–National Research Council, 1959.

2.29 DEPARTMENT OF DEFENSE

Washington 25, D. C.

H. EDMUND STILES, *Principal Investigator*

Work on the design of an all-computer document retrieval system continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Stiles, H. E. "The Association Factor in Information Retrieval," *Journal of the Association for Computing Machinery,* vol. 8, no. 2, April 1961, pp. 271-279.

(2) Stiles, H. E. "Machine Retrieval Using the Association Factor," in *Machine Indexing: Progress and Problems.* Washington, D. C.: American University, 1962.

2.30 THE DIEBOLD GROUP, INC.[1]

430 Park Avenue, New York 22, N. Y.

DONALD F. BLUMBERG, *Associate Director, Planning Division*

A series of research projects underway are primarily directed to the study of the interaction of information systems and planning-decision processes. A basic question now being considered is the quantitative effect of information flow, rate of acquisition of data, and content of information on the decision to plan, develop, and implement alternative plans and programs. The planning processes under consideration are primarily examples drawn from industrial, military, or scientific environments in which planning decisions (as opposed to action decisions) are of critical importance.

The following projects are now underway:

(a) An evaluation is being made of the scope, interaction, and utilization of the technical data base of a large and complex military agency. The primary mission of this agency is to coordinate and direct a significant amount of systems and operations analysis, and programming, of large military man-machine systems for users.

(b) Mechanized techniques are being developed for the generation

[1] Formerly reported as John Diebold and Associates, Inc.

of planning alternatives in the data processing-programming field (1).

(c) An analysis is being made of the nature and extent of utilization of a monthly accession index of Russian scientific textual information, as produced by the Library of Congress (3). This project is supported by funds from a National Science Foundation grant to the Library of Congress (see 1.31).

(d) Development and evaluation of a plans-obsolescence index for storage and retrieval continue. Particular attention has been given to the technical and economic anaylsis of image storage versus digital storage for browsing, search, and heuristic linking in complex planning and decision information systems.

(e) A major project, recently initiated, will examine the information flow and editorial decision processes in large newspaper and publishing organizations. A specific objective of this study will be to determine the overall impact upon the newspaper-publishing field of the present and future state of the art in information technology.

References:

(1) Blumberg Donald F. "Information Systems and the Planning Process," presented at the First International TIMS-CORS meeting, Toronto, Canada, May 1962.

(2) Blumberg, Donald F. "Information Bureaus—New Opportunities for the Service Bureau," presented at the Association of Data Processing Service Organizations, Philadelphia Symposium, October 1962.

(3) The Diebold Group, Inc. *Extent and Character of the Utilization of the Monthly Index of Russian Accessions,* report to the Library of Congress, November 1962.

DOCUMENTATION INCORPORATED 2.31

7900 Norfolk Avenue, Bethesda, Md.

EUGENE MILLER, *President*

Research and development in information technology continues, with no change reported since the previous statement [*Ed.*]. Support for the several projects is provided by the National Aeronautics and Space Administration, U.S. Air Force (Rome Air Development Center), U.S. Air Force Office of Scientific Research, and the National Institutes of Health.

2.32 DOUGLAS AIRCRAFT COMPANY, INC.

Missile & Space Systems, 3000 Ocean Park Boulevard, Santa Monica, Calif.

L. R. BUNNOW, *Group Engineer, Computing Engineering Section, and* GRETCHEN W. KORIAGIN, *Head Librarian, Missile and Space Systems Engineering*

The purpose of the project is to develop a system for the mechanization of many of the activities of a technical engineering library. This system is being developed in four subsystems: (a) the bibliographic data subsystem which supplies index cards, accession lists, and subject indexes; (b) the dictionary subsystem which coordinates the vocabulary used in indexing and prepares a printed dictionary; (c) a system of selective dissemination of information; and (d) a system for information retrieval in the form of literature searches.

The mechanized system deals with report literature both internally and externally generated. All Douglas technical libraries are providing input to the system. A total of 40,000 documents have been indexed into the system. The vocabulary contains 9,000 words. Approximately 10,000 technical employees will be served by the system. A Friden Flexowriter is used for input preparation; an IBM 7090 computer, for updating tape files and for selective dissemination of information and literature searches; and an IBM 1401 computer, for printing index cards, accession lists, and literature searches.

The bibliographic data and dictionary subsystems are operational and have been thoroughly checked out with good results. Selective dissemination is operational on an experimental basis. The literature searching subsystem is operational on the IBM 7090 and under analysis on the IBM 1401.

It is planned to complete all previously mentioned phases and then make a detailed evaluation of the entire system.

References:

(1) Brandenberg, W., H. C. Fallon, C. B. Hensley, T. R. Savage, and A. J. Sowarby. *Selective Dissemination of Information, SDI 2 System,* International Business Machines Corp., Advanced Systems Development Division, 1961.

(2) Bunnow, L. R. *Study of and Proposal for a Mechanized Information Retrieval System for the Missile and Space Systems Engineering Library,* Douglas Aircraft Co., Inc., Report SM 37418, May 1960.

(3) Koriagin, G. W. and L. R. Bunnow. *Mechanized Information Retrieval System for Douglas Aircraft Company, Inc. Status Report,* Douglas Aircraft Co., Inc., Report 39167, January 1962.

(4) Koriagin, G. W. *Library Information Retrieval Program.* Doug-

las Aircraft Company, Inc., Engineering Paper No. 1269, January 1962.

ELECTRO-OPTICAL SYSTEMS, INC. 2.33

Advanced Electronics and Information Systems Division, Pasadena, Calif.
ROBERT E. WALL, JR.,[1] *Project Supervisor*

Phase I of the research program to seek out new and unique solutions to the fact correlation problem has been completed (1). The work was sponsored by the U. S. Air Force (Rome Air Development Center).

Reference:

(1) Wall, Robert E., Jr. *Study of Fact Correlation for Intelligence Analysis,* Electro-Optical Systems, Inc., EOS Report 1590, June 14, 1962. Final report prepared for U. S. Air Force under Contract AF 30 (602)-2383. RADC-TDR-62-384 Vol. I.

ENGINEERS JOINT COUNCIL 2.34

345 East 47th Street, New York 17, N. Y.
EUGENE WALL

In order to create the basis for effective storage and retrieval of engineering information, a vocabulary of engineering terminology, with its interterm relationships, is being developed. The end result of the effort will be an *Engineering Thesaurus,* which will provide the indexer of engineering literature with additional indexing terms beyond those he may already have selected from his own knowledge and from the original document, thus adding additional access points to the material as it is stored. Similarly, it will facilitate searches by providing a list of possible terms by which pertinent documents might have been indexed because of generic or semantic association.

Numerous engineering glossaries, dictionaries, subject heading lists, classifications, and thesauri are being collected and their terminologies collated and converted to punched-card form. As of August 15, 1962, approximately 45,000 terms had been converted to punched cards; approximately 43,000 others are being processed. Terminological contributions are still being prepared by a number of other engineering societies and organizations. It is thus expected that about 100,000 terms will be treated in subsequent steps of the program, although most of these will be duplicates obtained from the different sources.

In the next phase of the project, the vocabulary terms collected from the various sources will be collated, and the degree of terminological overlap among sources will be determined. Those terms which are

[1] Present address: Applications Research Corp., 4241 Redwood Ave., Los Angeles, Calif.

employed more commonly—perhaps 10,000 to 20,000 terms—will be chosen for use in the *Engineering Thesaurus.* Interterm relationships (including synonymous, generic, and variable relationships) will be developed by subject-oriented terminology panels for the chosen terms. The remainder of the terms, e.g., those used by only one source, will be listed in the thesaurus according to source.

The collection of engineering terminology, its collation and analysis, and the development of interterm relationships are being supported by the National Science Foundation. It is expected that this phase of the work can be completed in approximately 1 year and that the Engineers Joint Council will publish the *Engineering Thesaurus.*

References:

(1) American Institute of Chemical Engineers. *Chemical Engineering Thesaurus,* 1961. ($10.00)

(2) Engineers Joint Council. *Engineering Information Symposium,* January 17, 1962. 95 p. ($2.00)

(3) Armed Services Technical Information Agency. *Thesaurus of ASTIA Descriptors—Edition II,* 1962 (scheduled for publication December 15, 1962).

2.35 ESSO RESEARCH AND ENGINEERING COMPANY

Technical Information Division, P.O. Box 51, Linden, N. J.

G. JAHODA, *Principal Investigator*

Work on the development of a technique for determining index requirements by analyzing representative search questions is being continued. Search headings have been characterized in terms of specificity of access point and number of component parts. The three basic types of indexes—the alphabetic-subject index, the hierarchically classified index, and the coordinate index—are being evaluated in terms of these search heading characteristics and by one additional characteristic: the number of steps required to get at document identifications. The technique is being tested on a representative sample of questions asked of the company's internal reports collection.

Reference:

(1) Jahoda, G. "A Technique for Determining Index Requirements," presented at the annual meeting of the American Documentation Institute, Hollywood-by-the-Sea, Fla., December 13, 1962.

ESSO RESEARCH AND ENGINEERING COMPANY 2.36

Engineering Information Center, P. O. Box 209, Madison, N. J.
W. H. RUPP[1] *and* G. JAHODA,[2] *Principal Investigators*

A coordinate index on IBM cards and on magnetic tape and a machine-prepared printed permuted terms index are being constructed for the company's engineering divisions. The annual rate of growth of the index is about 10,000 documents of internal basic data of an engineering nature. Programs for the IBM 1401 and/or IBM 7090 computers have been developed for the following operations: (a) translation of indexable information into an encoded standardized index language, (b) preparation of a printed list of terms used in the index with corresponding codes, (c) preparation of punched cards for searching on the IBM 108 sorter, (d) preparation of printed permuted terms index, and (e) storage of index on magnetic tapes for computer searching.

Most documents will be microfilmed and mounted, 48 images per jacket. Documents selected from the index searches can be retrieved either in the form of Xerox copy from microfilm or hard copy from central engineering files.

Test searches of the coordinate index with an IBM 108 sorter are now underway.

F. HOFFMAN-LA ROCHE & CO. LTD. 2.37

Basle, Switzerland
and
SANDOZ LTD.
Basle, Switzerland
F. WEGMÜLLER,[3] R. BECHER,[3] *and* H. R. SCHENK[4]

Since 1958, Roche and Sandoz have cooperated in the field of documentation concerning the published literature on chemistry, pharmacology, medicine, and allied sciences of interest to the pharmaceutical industry.

A project for automatic and mechanized information retrieval has been set up on the basis of an internal literature review and abstracting service in use since 1920 at Roche. The requirements of the project were (a) abstracting of about 400 scientific periodicals, (b) distribution of the printed abstracts to the scientific staff of both firms as an information medium, (c) mechanized and automatic documentation by means of electronic computers, and (d) use of one and the same type in clear language for both information and documentation.

[1] Head, Engineering Information Center, P. O. Box 209, Madison, N. J.
[2] Head, Systems Research and Library, Technical Information Division, P. O. Box 51, Linden, N. J.
[3] La Roche & Co. Ltd.
[4] Sandoz Ltd.

After an exhaustive investigation a method was worked out, called "Codeless Scanning," which is now in successful and satisfactory operation. The characteristic features of the method are as follows: The abstracts contain a field indication (e.g., organic chemistry, microbiology, etc.); a set of keywords and auxiliary terms in plain (i.e., uncoded) language and arranged in a logical sequence; authors; place of publication; plain bibliographical citation; title of the paper in original language; and, when necessary, a short summary possibly including graphs, tables, formulas, etc. Field indications, sets of keywords (including the auxiliary terms), authors, and citations of the abstracts are typed on the Flexowriter, then completed with titles and texts on ordinary typewriters. Where appropriate, Xerox copies of parts of the original are used without alternation. The completed sheets are transferred onto offset plates by xerography and then printed, assembled in individual fascicules corresponding to the particular interests of the subscribers by means of the field indication, and distributed. The Flexowriter machines are programmed; the resulting tapes contain automatically all codes necessary for further processing by means of an IBM tape-to-card converter at Roche or the corresponding Remington machine at Sandoz. The punched cards are fed into the computer (IBM 1401 at Roche and Remington UCT at Sandoz) and converted into magnetic tape. The search programs of the computers enable differential selections of the material, e.g., positive, negative, and alternative selection of keywords, auxiliary terms and/or parts thereof, authors, localities, and periodicals. Several searches may be conducted at the same time. The result of the machine search is a list printed in plain language, containing exactly the same full set of keywords and auxiliary terms, authors, and citation as in the original abstracts. Thus the inquirer obtains not only a list of document numbers but also an interpreted selection of the desired literature.

The aim was to create a fast, sure, labor- and time-saving method which does not necessitate the tedious and practically uncheckable translation into a code. A code covering the whole field from theoretical chemistry through the biological sciences would be impossibly large and cumbersome and would never be complete. As an alternative, a limited number of standardized keywords (at present about 1,300) are used, selected on the basis of previous experience. In addition, free terms are used, the choice depending exclusively upon the necessity of providing a more detailed description than can be given by the appropriate keyword alone. This enables subdivision of all material into a reasonable number of standardized items on one level; at the same time it permits the indication of minor, but important, facts, e.g., trademarks, laboratory numbers, and scientific nomenclature, for which a complete code system is quite impracticable. In the case of organic chemistry,

even the indication of the numbering or stereochemical characteristics may be given in terms of the usual nomenclature.

By means of the computer, a cumulative index is being printed in the form of a card file which is frequently used by individuals for minor search problems. Also, these index cards contain the full text in plain language and allow a high working speed, thus forming a complete set of cross-references. Work is continuing with the aim of reducing still further the labor involved in accelerating information and documentation.

References:

(1) Wegmüller, F., R. Becher, B. Hoffmann, and H. R. Schenk. " 'Codeless Scanning,' ein neues Verfahren der automatischen Dokumentation," *Separatum Experientia* (Basel, Birkhauser Verlag), vol. 16, no. 8, 1960, p. 383.

(2) Hoffmann, Bernhard A. "Die Voraussetzungen für den Einsatz von Elektronischen Datenverarbeitungs-Maschinen in der Dokumentation," *Nachrichten für Dokumentation,* vol. 12, no. 1, 1961, pp. 27-28.

(3) Wegmüller, Fritz. "Dokumentation in der pharmazeutischen Industrie mit Hilfe der Codeless Scanning-Methode," *Nachrichten für Dokumentation,* vol. 12, no. 1, March 1961, pp. 18-20.

(4) Schenk, Hans-Rudolf. "Der Einsatz von Lochkartenmaschinen und Computers in der Codeless Scanning-Methode," *Nachrichten für Dokumentation,* vol. 12, no. 1, 1961, pp. 22-26.

(5) Wegmüller, F., and H. R. Schenk. "Codeless Scanning, ein neuartiges Verfahren der Dokumentation," *Information* (Zürich), January-April 1961.

(6) "Codeless Scanning," *Revue de la Mécanographie,* no. 167, 1961, pp. 413-416.

(7) Schenk, H. R. "Der Einsatz eines UNIVAC UCT-Systems in der Dokumentation," *Lochkarte* (Frankfurt), vol. 25, no. 19, 1962, pp. 19-27.

FARBENFABRIKEN BAYER AG 2.38

Literaturgruppe der Pflanzenschutz Wissenschaftlichen Abteilung
[*Literature Group of the Scientific Plant Protection Department*]
Leverkusen, Federal Republic of Germany
WOLFGANG BARTELS

The Literature Group surveys the scientific literature and publishes a journal for the employees of the Plant Protection Section, containing abstracts of various publications in the fields of plant pathology and

plant protection. As a result of these literature surveys, methods of evaluating literature abstracts are constantly being improved. The following major points are being investigated: (a) The file contains 3,500 peek-a-boo cards for one particular subject. The time necessary for including a literature abstract in this system is being determined. Also, the time necessary for obtaining information from the cards is being ascertained. (b) The effectiveness of the peek-a-boo and edge-punched card systems is being compared, using as test material the files concerned with plant protection. The limits of usage of the two systems are being determined. (c) Costs involved in the usage of three filing systems (normal card register, edge-punched cards, and peek-a-boo) are being calculated. In particular, costs involved in answering various questions are being compared for the three systems.

To date, the studies have revealed that mechanical methods of documentation do not show any advantage over the peek-a-boo system. This was tested in regard to working time and costs needed for dealing with the literature of plant protection. When mechanical methods are used, the answering of questions takes most of the time.

Investigations have been started to find out how far back from the present time literature files would have to be checked under normal usage. For this purpose, literature references contained in publications in the field of plant pathology and plant protection during the last 10 years have been checked to determine the dates of the various citations.

References:

(1) Bartels, W. "Die Anwendung der Cordonniermethode bei der Auswertung von Literatur des Gebietes Pflanzenpathologie und Pflanzenschutz," *Zeitschrift für Pflanzenkrankheiten (Pflanzenpathologie und Pflanzenschutz,* vol. 66, 1959, pp. 627-636.

(2) Bartels, W. "Die Sichtlochkartei und ihre Folgeeinrichtungen —Ein Beitrag zur Methodik der Erfassung und Auswertung von Pflanzenschutzliteratur," *Nachrichten für Dokumentation,* vol. 12, 1961, pp. 77-85.

(3) Bartels, W. "Sichtlochkarten—Kerblochkarten, Vorteile und Grenzen beider Methoden bei der Karteibefragung, dargestellt am Beispiel einer Pflanzenschutz-Literaturkartei," *Nachrichten für Dokumentation,* vol. 12, 1961, pp. 137-146.

(4) Bartels, W. "Vergleiche der Wirtschaftlichkeit von Sichtloch-, Kerbloch- und Steilkarteien," Zusammenfassung und Diagramme in *Kurzberichte und Sonderdrucke zur 8. Öffentlichen Arbeitssitzung des Arbeitsausschusses für Kostengrundlagen der Dokumentation der Deutschen Gesellschaft für Dokumentation e.V. am 9. Oktober 1961 in Bad Dürkheim an der Weinstrasse,* Frankfurt am Main, 1961, pp. 12-16.

GENERAL ELECTRIC COMPANY 2.39

Information Systems Operations, 4901 Fairmont Avenue, Washington 14, D. C.
WAYNE D. BARTLETT, *Manager*

Research and development activity pertinent to automatic indexing and abstracting of full text is underway. It is recognized that the automatic selection of keywords to create an index, or the automatic selection and assembly of keywords, key sentences, and parts of sentences to create an abstract is a difficult problem which will not be solved by any single, simple selection criterion. As a result, a comprehensive, systematic approach to these problems has been adopted in the form of a project entitled META (Methods of Extracting Text Automatically).

META is a project based upon a broad engineering approach directed to the development of algorithms for machine indexing and abstracting of textual information. A secondary objective is the development of a methodology for designing particular algorithms which are a function of the nature of the textual information to be processed.

META involves the hypothesis, empirical testing, and evaluation of many criteria for analysis and selection of the most significant words and sentences of the textual information. These criteria may be broadly classed as statistical, linguistic, and semantic. The testing and evaluation of the various criteria are being done with the aid of a digital computer.

META consists, in part, of a flexible set of computer programs which have been written for the GE 225 computer. The programs provide a modular system which has been designed to test the various hypotheses and methods of analysis and selection without extensive programming cost or time. The system currently consists of approximately 10,000 operational instructions. Various glossaries have been compiled, and approximately 100,000 words of full text have been keypunched from technical articles. Experimentation is continuing, and results will be reported subsequently.

GENERAL ELECTRIC COMPANY 2.40

Cincinnati 15, Ohio
J. R. HUBBELL, *Principal Investigator, and*
J. M. BENBOW, *Co-principal Investigator*

Research on a classification system using various map grids as a graphic illustration for word hierarchical order and arrangement continues, with no change reported since the previous statement [*Ed.*].

2.41 GENERAL ELECTRIC COMPANY

Technical Military Planning Operations, Santa Barbara, Calif.
FREDERICK B. THOMPSON, *Principal Investigator*

Development of a general mathematical theory of information systems continues, with no change reported since the previous statement [*Ed.*].

2.42 GENERAL ELECTRIC COMPANY

Computer Department, Phoenix, Ariz.
T. L. WANG, *Principal Investigator*

Development of a system for automatically extracting intelligence from a large mass of data has been completed, and the system is now operational on an IBM 7090 computer.

The study was sponsored by the Minuteman High Reliability Components Program of Autonetics, a division of North American Aviation, Inc.

References:

(1) Wang, T. L. "An Information System With the Ability to Extract Intelligence From Data," *Communications of the ACM*, vol. 5, no. 1, January 1962, pp. 16-18.
(2) Wang, T. L., T. Vickers, and A. Fox. *Reliability Study Information System User's Manual*, March 1961.

2.43 GEORGETOWN UNIVERSITY SCHOOL OF MEDICINE

Department of Medicine, Washington 7, D. C.
and
MT. ALTO VA HOSPITAL
Washington 7, D. C.
HUBERT V. PIPBERGER[1]

Attempts are being made to develop a digital machine code for historical and physical data from patients with cardiovascular diseases. This code will be applied in large-scale statistical and epidemiological studies of heart disease by means of digital electronic computers.

A large checklist has been prepared using data from the case histories, physical examinations, and laboratory findings for patients presenting the chief complaint of chest pain. Data are being collected only from patients where the final diagnosis could be verified by objective means. Attempts are being made to apply data reduction methods to this original information. Redundant information is being eliminated. Weight factors in terms of the differential diagnosis of chest pain are being

[1] Georgetown University School of Medicine

determined. Subjective factors for attributing weight factors to signs and symptoms are not being used.

The project has been in progress since June 1, 1960, and several hundred case histories have been collected. The data are being transcribed on checklists of the "yes" or "no" type using the FOSDIC system of the National Bureau of the Census. Microphotographs of these checklists are being scanned optically, and the data are transcribed directly onto digital magnetic tape. In this form they are fed into an IBM 7090 computer for evaluation.

The project is supported by the National Heart Institute.

GMELIN INSTITUTE DOCUMENTATION CENTER 2.44

Varrentrappstrasse 40-42, Frankfurt am Main 13,
Federal Republic of Germany
E. H. E. PIETSCH, *Director*

Work continues on the statistical analysis of word frequency in the literature (current reports, conference papers, dissertations, patents, industrial papers) processed by the Zentralstelle fuer Atomkernenergie-Dokumentation beim Gmelin-Institute (Center for Atomic Energy Documentation in the Gmelin Institute). A descriptor list was compiled from this literature which, starting in 1962, is published in the series AB *Bibliographic Reviews* (formerly series A and series B publications). The subject matter of the bibliographic series B-1961 (12 monthly publications) was analyzed statistically with respect to the nature of titles and to word frequency for further mechanical treatment.

It is intended to continue this statistical analysis of word frequency. As a framework the descriptors are divided into two groups (fixed and more or less varying descriptors), and only the first group seems to be useful for direct application for storage, thesauri, etc. The group with rigid semantic content is subdivided into subgroups, such as A (apparatus, instruments, devices); P (processes, reactions, methods); S (substances, classes of materials, especially correlated with chemical elements); and O (general concepts, scientific topics and disciplines, properties).

The second group consists of G (geographical names, locations); I (institutions, corporations, universities); and N (names, which are loosely connected with a second concept). For this group a suitable method for processing indexes, especially for mechanical documentation systems with limited storage volume, must be found and worked out.

2.45 HARVARD UNIVERSITY

Computation Laboratory, Cambridge 38, Mass.
GERARD SALTON, *Assistant Professor of Applied Mathematics*

Research is continuing in automatic content analysis and in the automatic identification and classification of information. The proposed methods emphasize on the one hand the use of linguistic criteria and of citations available with natural-language texts, and on the other the design of abstract models to represent the structure of retrieval systems.

The following efforts are representative of work performed within the past 6 months: (a) a program for automatic syntactic analysis and the automatic generation of phrases and word groups, using a small function word dictionary and contextual analyses to replace the complete word dictionary normally required for this purpose (1) (2); (b) an experiment to determine whether citations available with documents could be used as an aid in the determination of document content (3); and (c) the construction of a model for representing retrieval operations by structure matching processes (4).

This research is sponsored by the U. S. Air Force Cambridge Research Laboratories.

References:

(1) Lemmon, A. "Report on a Syntactic Analysis Program for Information Retrieval," in *Information Storage and Retrieval,* Scientific Report No. ISR-2 to the Air Force Cambridge Research Laboratories, September 1962.

(2) Salton, G. *Some Experiments in the Generation of Word and Document Associations.* (To be published)

(3) Salton, G. "The Use of Citations as an Aid to Automatic Content Analysis," in *Information Storage and Retrieval,* Scientific Report No. ISR-2 to the Air Force Cambridge Research Laboratories, September 1962.

(4) Salton, G. "The Manipulation of Trees in Information Retrieval," *Communications of the ACM,* vol. 5, no. 2, February 1962. Also in *Information Storage and Retrieval,* Scientific Report No. ISR-1 to the Air Force Cambridge Research Laboratories, November 1961; ASTIA No. AD–274 816.

2.46 HATFIELD COLLEGE OF TECHNOLOGY

Roe Green, Hatfield, Herts., England
G. H. WRIGHT, *County Technical Librarian*

Research underway is an extension of work on mechanized indexing previously reported at Warren Spring Laboratory (*CRDSD,* No. 9, Statement No. 2.66).

The new program of work envisages (a) erection of an improved

model of the tape punching typewriter equipment used for input to the system; (b) alterations in the camera to increase to a total of 900 the number of bits of code which the fiche can hold; (c) construction at minimum cost of a new selector, including an electronic unit to control its logical operation, which will not have the restraints of Filmorex coding and which can use the 900 bits of code in search; (d) development of a machine language which will take advantage of the improvements in the coding capacity and logical power of the equipment; and (e) trial of the system, at first in sociology and later in other subjects as the language develops, in answering inquiries received by the County Technical Information Service; the language will be evaluated and improved in the light of the results obtained.

It is hoped that this new approach will result in a retrieval percentage above the 80 percent so far achieved with other systems and in the work at Aslib (see 2.9), and that the logical controller will enable queries containing a number of alternatives to be satisfactorily answered. The system, which is of the random-access type, is based on discrete fiches, and multiple copies will be made for the main entry points. Since only the relevant portion of the collection will have to be searched, it should be possible to provide answers to inquiries within 30 minutes even at the comparatively slow searching speed (100 fiches per minute) proposed.

References:

(1) Claridge, P.R.P. "Information Handling in a Large Information System," in *Proceedings of the International Conference on Scientific Information,* pp. 1203-1220. Washington, D.C.: National Academy of Sciences-National Research Council, 1959.

(2) Claridge, P.R.P. "Mechanized Indexing of Information on Chemical Compounds in Plants," *Indexer,* vol. 2, no. 1, 1960, pp. 4-19.

HERNER AND COMPANY 2.47

1401 K Street, NW., Washington 5, D.C.

MARY HERNER *and* WALTER JOHANNINGSMEIER

Work is continuing on a study of deep indexing techniques and methods of presenting correlative indexes in alphabetical page form. The indexing phase of the work, which is being done under a contract with the U. S. Air Force Office of Scientific Research, is concerned with a method of indexing which is essentially a cross between permuted indexes such as KWIC and the "links" developed at E. I. duPont de Nemours & Co. The contents of each document are described in a series of highly stylized telegraphic phrases in which unessential connectives are dropped and in which only terms from a carefully screened

vocabulary are used. Each telegraphic phrase is treated as a set or entity, and the terms in each set are permuted by means of a fixed procedure in which each term is used once as the "lead" term for the set. The permuted phrases are arranged alphabetically by "lead" terms and printed out in a form closely akin to a typical book index with indentation of subordinate terms, and a listing of all subordinates occurring under each "lead" term.

In keeping with the spirit of the study, both the intellectual and mechanical phases of the index preparation have undergone a series of refinements. The study began with a professional indexer making up the telegraphic phrases and permuting by the significant terms within them. This was followed by the writing of each descriptive phrase once and having a clerk do the permutations. This was further streamlined by having the phrases dictated rather than written. The procedure has now been taken a step farther, and the indexing language and the method of presentation have been made even more highly standardized and stylized in anticipation of permutation and printout by punchcard and magnetic tape devices. At present, a detailed study is being made of the comparative advantages of the various mechanical methods of preparing and presenting permutation indexes.

In another project, which is being performed on behalf of the Joint Council Subcommittee on Cerebrovascular Disease of the National Institutes of Health, an experimental keyword index is prepared from the titles of articles in selected sections of *Cerebrovascular Bibliography*. The purpose of the experimental index is to determine the degree to which workers in the field of cerebrovascular disease prefer to search subject terms on a pre- or post-coordinated basis.

2.48 HUGHES AIRCRAFT COMPANY

Library Services and Documentation Research Section,
Culver City, Calif.
H. THAYNE JOHNSON, *Section Head*

Work is continuing slowly on machine applications of retrieval terms and basic English grammar (see *CRDSD*, No. 10, Statement No. 2.34).

In addition, a program is underway to collect, index, and abstract the literature on the electrical and electronic properties of materials and to evaluate and compile the experimental data from this literature. It is estimated that about 30,000 items extending back to 1940 will be documented. These will constitute a highly selective technical information center complete with a specialized modified coordinate index.

Nine major categories of materials are being covered: semiconductors, insulators, ferroelectric dielectrics, metals, ferrites, ferromagnetics, electroluminescent materials, thermionic emitters, and superconductors.

Extended studies are proceeding on subcategories and properties under these major headings for use in both the indexing system and the issuance of data sheets covering the experimental data. These studies are now completed for semiconductors and insulators.

Additional studies are being accomplished to provide maximum utilization of the indexing system. A punched-card program has been developed for monthly updating of descriptor cards, with a tape updating program now being considered. However, these are based strictly upon manipulation of the index primarily by material, while it would be desirable to provide rapid use of the system through either material or property, or both. Accordingly, access facets are being explored which will include several levels of information beginning with the article containing the data and running through groups of data associated with one material or with one material and property (Indium Arsenide–Absorption) down to the individual set point established by the original experiment. Complete cross-referencing between facets would provide optimum use of the data.

Although the program is slanted toward materials applications, the use of data processing techniques makes it valuable for other uses such as: (a) Master tab runs of materials and properties can be produced; (b) valuable statistical counts can be made, such as average number of descriptors per article or average number of articles in the system per material and property; (c) state-of-the-art surveys can be easily accomplished, for the system will indicate not only areas where research has been done but also areas where gaps in the literature (and therefore to some extent gaps in research) occur; and (d) a measure of the growth of research activities in the field can be developed.

The latter project is supported by the U. S. Air Force, Directorate of Materials and Processes, Aeronautical Systems Division.

References:

(1) Johnson, H. Thayne, Emil Schafer, and Everett M. Wallace. *Electrical and Electronic Properties of Materials, Information Retrieval Program,* Technical Documentary Report No. ASD-TDR-62-539, Directorate of Materials and Processes, Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, June 1962.

(2) Johnson, H. Thayne. *A Hierarchic Approach to Automated Information Retrieval,* Report No. SRS 491, Hughes Aircraft Co., Culver City, Calif., March 1962.

(3) Johnson, H. Thayne. "A Polydimensional Scheme for Information Retrieval," presented at the Annual Convention of the American Documentation Institute, Boston, November 8, 1961; published in *American Documentation,* vol. 13, no. 1, January 1962, pp. 90-92.

(4) Johnson, H. Thayne, and Miles C. Pine. *A Punched Card Program for Posting Descriptor Cards in Information Retrieval Systems,* Report No. SRS 576, Hughes Aircraft Co., Culver City, Calif., September 1962.
(5) Johnson, H. Thayne, and Lou E. Vaughn. *A Theoretical Information Retrieval System for Retrieving Highly Specific Materials Data,* Report No. SRS 577, Hughes Aircraft Co., Culver, City, Calif., September 1962.
(6) Schafer, Emil. *Insulation Materials Descriptors Used in the Electrical and Electronic Properties of Materials Information Retrieval Program,* Special Report No. 1 (Contact AF 33 (616)-8438), Hughes Aircraft Co., Culver City, Calif., July 1962.
(7) Schafer, Emil. *Semiconductor Materials Descriptors Used in the Electrical and Electronic Properties of Materials Information Retrieval Program,* Special Report No. 2 (Contract AF 33 (616)-8438), Hughes Aircraft Co., Culver City, Calif., September 1962.
(8) Wallace, Everett M. *Information Retrieval Program, Electronic/Electrical Properties of Materials,* First, Second, and Third Quarterly Reports (Contract AF 33 (616)-8438), Hughes Aircraft Co., Culver City, Calif., October 1961, January 1962, April 1962.
(9) Johnson, H. Thayne. "An Information Retrieval Program on the Electrical and Electronic Properties of Materials," presented at the Materials and Information Symposium, Dayton, Ohio, November 28-29, 1962.

2.49 HUMAN RELATIONS AREA FILES, INC.

P. O. Box 2054, Yale Station, New Haven, Conn.
FRANK W. MOORE, *Director of File Research*

The project underway seeks to facilitate comparative research in human behavior by providing in paper or microfilm form a special rapid-access library collection of social science documents, which have been analyzed and organized according to codes presented in (2) and (3), for distribution to 50 subscribing institutions. Natural-language documents containing data of significance to the social sciences pertaining to a selected sample (over 200) of the world's cultures are collected, reproduced, analyzed, coded, distributed, and stored. To date, some 500,000 pages from 4,000 documents have been processed for the files, of which about 100,000 pages from 800 documents have been added in the last 12 months; microfilming of previously processed data has continued at the rate of 100,000 pages every 12 months. Methodology, codes, and equipment used are described in the references cited below, and have not been substantially altered recently. During the forthcoming 12 months, a similar quantity of documents is scheduled for

processing, with increased emphasis on the addition of new materials to established files; a revised edition of (3) is to be completed for publication.

Financial support on a continuing basis is received from 21 member institutions; current analysis and research are supported by a grant from the National Institute of Mental Health.

References:

(1) Moore, Frank W. "Cross-Cultural Documentation," in *Readings in Cross-Cultural Methodology*, ed. by F. W. Moore, pp. 277-282. New Haven: Human Relations Area Files, 1961.

(2) Murdock, George P. *Outline of World Cultures*, 2d ed., revised. New Haven: Human Relations Area Files, 1958.

(3) Murdock, George P., et al. *Outline of Cultural Materials*, 4th. ed., revised. New Haven: Human Relations Area Files, 1961.

INDEX & RETRIEVAL SYSTEMS INC. 2.50

Woodstock, Vt.

A. B. KYLE, *Systems Manager*

Study of methods for evaluating the optimum form, content, and usefulness of an interdisciplinary science index continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Bennett, Frank G., IV. "Current Index to Abstracting Services," *Sci-Tech News*, Fall 1961, p. 129.

INDIANA UNIVERSITY 2.51

Research Center in Anthropology, Folklore, and Linguistics, Bloomington, Ind.

THOMAS A. SEBEOK, *Principal Investigator, and*
ELAINE K. RISTINEN, *Research Associate*

An attempt is being made to devise a useful information storage and retrieval system for linguistic documents, using the facilities of the Indiana University Research Computing Center and about 10,000 documents stored at the Research Center in Anthropology, Folklore, and Linguistics as a first sample. The investigation is directed toward the discipline of linguistics and designed first of all for specialists in that field. An IBM 709 computer will be programmed to develop the system as well as to process the organized data. It is intended that the linguists working on this project will have sufficient experience in computer programming to be able to supervise the entire programming effort.

Since automatic indexing is in an early stage of development, items

will be stored in condensed form and indexed manually. Variations of keyword schemes will be tried. Subject classifications are being discussed in relation to storage and retrieval. Inasmuch as there are so many variables involved in the whole problem, empirical procedures and an experimental attitude prevail. Although a more or less homogeneous group of users is considered at the outset, possible expansion to include users with special needs is contemplated for the future. For instance, retrieval questions relating linguistics to other fields are of some importance.

The characteristics of linguistic documents are being given special consideration. There are three kinds of descriptors: (a) names of persons, such as authors and cited authors; (b) names of languages and their varieties; and (c) technical terms, including dates. Multilingual access is essential. Every effort will be made to make indexing simple, so that it can be done by clerical workers; special effort will go into nonrecurring jobs, such as programming.

The work can be divided into several phases. The first phase involves devising the system, primarily the indexing system. The next phase is pretesting the indexing system; then comes the programming; and finally retrieval testing, which, when completely satisfactory, is the end of the procedure. The project is in the first phase now—the devising of the indexing system. Preliminary indexing has been completed on a sample of 100 documents. Some plans have been made for later phases also. For instance, when retrieval testing is done, a block lookup system will probably be tried for linguistic terminology (1).

It is hoped that by mid-1963 a sample of from 100 to 1,000 documents will be fully indexed, pretested, programmed, and ready for retrieval testing, and that multilingual indexes sufficient to index the entire sample library of 10,000 items will be completed. These multilingual indexes, since they are likely to be of value to others, will then be published separately.

This investigation is sponsored by the U. S. Air Force Office of Scientific Research of the Office of Aerospace Research.

Reference:

(1) Te Nuyl, Th. W. "The 'L'unité' mechanized documentation system," *Revue de la Documentation,* vol. 28, no. 4, 1961, pp. 140-147.

2.52 INFORMATION DYNAMICS CORPORATION [1]

2 Lakeside Park, Route 128, Wakefield, Mass.
DAVID P. WAITE, *President*

An operations research study has been initiated to undertake a comparative economic analysis of two hypothetical systems: one, a subject-

[1] Formerly Forbes & Waite, Inc.

specialized information service network; the other, a geographically regionalized information service network. It is expected that the study will provide greater insight into the fundamental characteristics of two major systems' concepts applicable to national dissemination of scientific information. This will involve the identification and selection of performance requirements which each system must be designed to satisfy; identification of variations in the user community which each system must accommodate; and study and selection of various bases of comparison against which two systems can be compared. Various mathematical expressions of the two systems will be developed as a means for making a comparative economic analysis.

The study is sponsored by the National Science Foundation.

INFORMATION FOR INDUSTRY, INC. 2.53

1000 Connecticut Avenue, NW., Washington 6, D. C.
LYNN J. BARTLETT, *President*

Nearing completion in the autumn of 1962 is a 2-year study of notation schemes for the chemical and biological dichotomy of U. S. and foreign pharmaceutical patent information. The present phase is centered on testing information handling techniques for systems with a relatively small corpus of documents which are characterized by a large number of variables yet to be controlled.

A quantitative analysis of patent information and patterns of utilization in chemistry and related fields has been applied to the *Uniterm Index to U. S. Chemical Patents,* published by Information for Industry, Inc. The results have been employed as design criteria for the preparation of a magnetic tape system. A total of 3,580,000 references comprise the index to a corpus of 102,343 patents issued between 1950 and 1961. The study of use patterns indicated a requirement for post-coordinated searching in the magnetic tape index and also for assembling skeletal abstracts to patents. This product will be marketed in November 1962.

INSTITUTE FOR ADVANCEMENT OF MEDICAL COMMUNICATION 2.54

9650 Wisconsin Avenue, Bethesda 14, Md.
CLAIRE K. SCHULTZ,[1] *Research Associate*

Work continues on further automating the preparation and printing of the index to the papers presented at the annual meeting of the Federation of American Societies for Experimental Biology (1). For the 1962 index, which covered almost 3,000 papers, a UNIVAC I computer program was written for high-speed printout in a format

[1] Address: Line Lexington, Pa.

closely resembling traditional subject indexes and therefore requiring little adaptation on the part of users. Currently a thesaurus is being developed to enable the computer to perform some of the editing of author-generated input that previously required human intervention. Since papers presented at Federation meetings cover a very broad area of biomedical research, it is expected that printed editions of this thesaurus will be useful to information services in this field. In addition, the feasibility of using a Photon for computer printout is being explored in an effort to improve the readability and typographic sophistication of the printed index. The work is sponsored by the National Institutes of Health.

A related study concerns the effect of subject-matter background and indexing experience on the choice of descriptors for documents. A random sample comprising 300 papers given at the April 1962 Federation meeting is being indexed independently by a physiologist, a biochemist, a pharmacologist, a pathologist, an immunologist, and a specialist in nutrition, none of whom have had previous indexing experience. The indexing terms generated by these scientists will be analyzed for differences attributable to disciplinary background and will be compared with terms chosen for these same papers by authors, program chairmen, and experienced indexers with relatively little subject-matter knowledge.

References:

(1) Schultz, Claire K., and Clayton A. Shepherd. "The 1960 Federation Meeting: Scheduling a Meeting and Preparing an Index By Computer," *Federation Proceedings,* vol. 19, no. 2, July 1960, pp. 682-699. Reprinted in *Medical Documentation,* vol. 5, no. 4, October 1961, pp. 95-105.

(2) Schultz, Claire K. "Some Characteristics of an Efficient Information Retrieval System," *Journal of Chemical Documentation,* vol. 2, no. 2, April 1962, pp. 103-105.

2.55 INSTITUTE FOR ADVANCEMENT OF MEDICAL COMMUNICATION

Washington Office, 1028 Connecticut Avenue NW., Washington 6, D. C.
ISAAC D. WELT, *Associate Director, and Chief, Washington Office*

Research continues on the "combined index-abstract" approach to information storage and retrieval (1), utilizing comprehensive collections of cardiovascular and psychopharmacology literature. Current emphasis is on improving the efficiency of indexing and decreasing the cost of preparing a printed index by using such mechanical aids as the step-camera, tape typewriter, and computer. The advantages and disadvantages of the tape typewriter for transcribing index entries, for

exchanging material with cooperating activities, and for converting text to machine-readable form are being studied.

A thesaurus of some 200 terms has been developed for the field of psychopharmacology and is now being tested by a number of indexers. The world literature in psychopharmacology has been carefully studied. Although the 10,000 papers examined represent many hundreds of journals, approximately one-half were published in less than 100 journals. The scattering of references complies with Bradford's Law. Whereas some 23 languages are represented in the collection, 62 percent of the papers are published in English, French, German, Italian, Russian, and Spanish, in descending order of frequency. A complete report on the characteristics of psychopharmacology literature is being prepared for publication.

The research is supported by the National Heart Institute and the National Institute of Mental Health.

Reference:

(1) Welt, Isaac D. "A Combined Indexing-Abstracting System," in *Proceedings of the International Conference on Scientific Information,* pp. 449-459. Washington, D. C.: National Academy of Sciences-National Research Council, 1959.

INSTITUTE FOR BIO-MEDICAL COMPUTER RESEARCH 2.56

11 East 44th Street, New York 17, N. Y.
LIDA G. GOTTSCH, *Principal Investigator*

A comprehensive Case History Code has been developed which enables a large amount of data to be coded from hospital charts and patient interviews. The code further assigns the information directly to spaces on punchcards and makes possible the transfer of data to paper or magnetic tape for computer analyses of the information.

Currently, forms are being designed for recording the case history, examination, and analysis in the code. Drafts have been submitted for criticism to researchers in allied fields. When criticisms and suggestions have been evaluated, the forms will be revised, printed, and disseminated to interested medical researchers.

The form-design project is supported by the National Institutes of Health.

INSTITUTE OF LIBRARY SCIENCE 2.57

Muzeum utca 3, Budapest, VIII, Hungary
G. OROSZ, *Investigator*

Investigation is continuing on the applicability of machine-punched cards as storage media and of the standard punched-card machines as searching devices for information retrieval.

Further research is underway on code theory of double-row marginal-punched cards and on the elaboration of some new types of coding systems.

References:

(1) Orosz, G. *Zwei Verfahren des Informationsauffindens vermittels Hollerith-Maschinen,* Budapest, 1960. (In Hungarian with summary in German)

(2) Orosz, G. "Mechanical Information Retrieval Based on the Collation Principle," in *Yearbook of the National Szechenyi Library 1959* [*Az Orszagos Szechenyi Konyvtar Evkonyve 1959*], Budapest, 1961. (In Hungarian with summary in English)

(3) Orosz, G. "Theory of the Subordinated Coding Systems of the Double Row Marginal Punched Cards," in *Yearbook of the National Szechenyi Library 1960* [*Az Orszagos Szechenyi Konyvtar Evkonyve 1960*]. (In press) (In Hungarian with summary in English)

(4) Orosz, G., and E. Pataky. "Verfahren zur Anwendung von Lochkartenmaschinen beim Auffinden von Informationen," to be published in *Dokumentation,* vol. 9, no. 5, 1962.

2.58 INSTITUTE FOR SCIENTIFIC INFORMATION

33 South 17th Street, Philadelphia 3, Pa.

EUGENE GARFIELD, *Director, and* IRVING H. SHER, *Director of Research*

I. Chemical Linguistics

CHEMTRAN is ISI's designation for a complete grammar of organic chemical nomenclature. Once completed, CHEMTRAN would include all the necessary algorithms for the interconversion of chemical nomenclature, structural diagrams, notations, codes, descriptors, etc. Under the surveillance of outside user study groups, it is planned to conduct comparative studies on various indexes that would result from applying such programs to the chemicals indexed in the *Index Chemicus.* Thus, beginning with a systematic chemical name, CHEMTRAN would generate a series of "chemonyms" including molecular formula, structural diagram, line notation, CBCC code, etc. As the state of the art advances in pattern recognition machines, use of the structural diagram as input becomes practical. CHEMTRAN then would also include recognition as well as generation algorithms for computing "chemonyms" from diagrams.

II. Rotated Molecular Formula Indexes

An experimental cumulative rotated molecular formula index, the RotaForm Index,® was prepared by an IBM 7090-1401 computer complex covering over 180,000 new chemical compounds indexed during the

period 1960-62. Preliminary tests indicate that it will be extremely useful in searching fields like metallo-organics.

Experimentation is underway with a new index called "Rotadex" which will significantly speed up use of the conventional molecular formula index and the RotaForm Index.® Rotadex gives a true and complete rotation of the molecular elements, includes columnar format presentation, and includes a descriptor code of 4 to 5 alphanumeric symbols which describe (a) the fundamental chemical features or structures of a compound, and (b) the kind of compound. It will permit extremely sophisticated structural generic searches.

References:

(1) Garfield, E. "An Algorithm for Translating Chemical Names to Molecular Formulas," *Journal of Chemical Documentation,* vol. 2, no. 3, 1962, pp. 177-179.

(2) Garfield, E. "Generic Searching by Use of Rotated Formula Indexes," to be published in *Journal of Chemical Documentation.*

INTERNATIONAL BUSINESS MACHINES CORPORATION 2.59

Research Laboratory, Monterey and Cottle Roads, San Jose, Calif.

PHYLLIS BAXENDALE

Research is directed to the identification and classification of information as it obtains in printed text. Two types of investigation are underway: One concerns reference retrieval systems; the other, information retrieval systems.

The reference retrieval investigations are presently focused on the automatic indexing of technical documents and the search strategies appropriate to the mode of indexing. The approach is two-fold: syntactic and statistical. In each instance an empirical model of language is formulated and operating instructions specified by which "valid index terms" may be selected from various units of text. Thus far the syntactic model has been based on the syntactic and some semantic features of titles, paragraph captions, etc. This model was tested and evaluated as it operated on 500 titles of physics theses. To provide greater depth of indexing, it is now planned to expand the concept to more complex statements of English.

Using approximately 4,000 theses titles as the data base, statistical investigations are underway on a methodology by which index terms may be meaningfully related. The initial experiment will test a concept of semantic association which can be quantified and for which statistical measures of association can be designed.

The information retrieval investigations exploit the concept of logical

syntax for application in a "question-answering" system. It is the task of descriptive logical syntax to delineate both the formal-language sentences and the natural-language sentences and the interconnection between them, so that the rules of inference of the formal language can be exploited indirectly in the natural language.

Each of these experimental programs is implemented with the aid of an IBM 7090 computer. The syntactic models are programmed in COMIT language. Source material may be entered alternatively by punched cards or by Flexowriter tape through the medium of a comprehensive code based upon that designed by S. Newman, et al.

References:

(1) Newman, S. M., R. W. Swanson, and K. C. Knowlton. "A Notation System for Transliterating Technical and Scientific Text for Use in Data Processing Systems," in *Advances in Documentation and Library Science,* vol. 3, part 1, pp. 345-376. New York: Interscience Publishers, Inc., 1960.

(2) Baxendale, P. "Machine-Made Index for Technical Literature—An Experiment," *IBM Journal of Research and Development,* vol. 2, no. 4, October 1958.

(3) Baxendale, P. "An Empirical Model for Machine Indexing," in *Machine Indexing: Progress and Problems.* Washington, D. C.: American University, 1962.

(4) Bennett, J., P. Lewis, and P. Baxendale. *A Prospectus for a Computer Program to Investigate the Application of Statistical Methodology to Information Retrieval Problems,* Research Note, IBM Research Laboratory, San Jose, February 1962.

2.60 INTERNATIONAL BUSINESS MACHINES CORPORATION

Federal Systems Division, Bethesda, Md.

ROBERT H. COURTNEY, JR., *Manager, Systems Development Department*

Several research projects involving the application of information technology to intelligence are under development.

A project is underway to test experimentally the feasibility of automatically classifying documents based on the statistical distribution of words in the documents. Basic techniques for performing automatic classification have been developed and programmed for the IBM 709 computer.

A new technique for the indexing and retrieval of photographs has been programmed and demonstrated on the IBM 709 computer. Research is now in process to develop advanced techniques of image analysis and correlation.

Other projects currently under development include a detailed analysis of variations in document indexing among individual indexers, a study of retrieval patterns generated as a function of requests, and the development of a formatted file design for the IBM 1410 computer.

References:

(1) Garland, J. L., and K. W. Webb. *Information Retrieval System Evaluation Technique* (Summary Report), IBM, Systems Development Department, 1962. 26 p.

(2) Meadow, H. R. *Statistical Analysis and Classification of Documents,* IBM, Systems Development Department, 1962. 116 p.

(3) Mumbower, L. E., T. W. Richards, et al. *Photo and Graphic Indexing and Retrieval System for Photo-Intelligence Operations,* IBM, Systems Development Department, 1962. 77 p.

(4) Muth, E. P. *Document Indexing and Retrieval Study,* IBM, Systems Development Department, 1962. 84 p.

(5) Winter, R. E., J. L. Garland, and K. W. Webb. *Research Report Registry Feasibility Study* (Final Report), IBM, Systems Development Department, 1962. 32 p.

INTERNATIONAL BUSINESS MACHINES CORPORATION 2.61

Advanced Systems Development Division, San Jose, Calif.

MARJORIE GRIFFIN, *Manager, Library*

The IBM ASDD and Research Library provide conventional services and are a proving ground for the application of library procedures to IBM machines. Special IBM punched cards have been designed for ordering books, pamphlets, periodicals, and reports, and for interlibrary loans. Preprinted routing slips are used to route periodicals. Master shelf-list punched cards are used to produce catalog and circulation cards. From theses catalog cards three separate alphabetic book catalogs—by author, title, and subject—are automatically prepared, listing the volume holdings of two IBM San Jose libraries. These catalogs, which contain author, title, subject, call number, date, and library location information, are updated every 2 months. The circulation cards are convenient for overdue notices and termination listings; they are also an efficient medium for collecting statistical data. Processing has been streamlined by conversion from IBM EAM equipment to the IBM 1401 computer.

Bibliographies from the descriptors of 29,000 documents are compiled from an IBM 7090 computer within 4 hours. The program admits the logical connectives AND, OR, and NOT. The endings of the keywords can be masked if a general rather than a specific search is required.

The library is currently preparing a total system with emphasis on speed and reduction of errors in input and real-time information.

References:

(1) Griffin, Marjorie. *The Library of Tomorrow,* IBM ASD Technical Memorandum ASDJ-M-040, San Jose, April 1961.

(2) Marckworth, Lois. *Dissertations in Physics, An Indexed Bibliography of All Doctoral Theses Accepted by American Universities, 1861-1959.* Stanford, Calif.: Stanford University Press, 1961.

2.62 INTERNATIONAL BUSINESS MACHINES CORPORATION

Experimental Systems Department, Thomas J. Watson Research Center, P. O. Box 218, Yorktown Heights, N. Y.

MANFRED KOCHEN, *Project Leader*

The broad aim of this project is to specify general characteristics of useful nonarithmetic processing systems. This involves analysis of new concepts and methods for representing, storing, and processing large quantities of non-numeric information. The use, rather than development, of new hardware (e.g., IBM photostore-7090-console complex), programming techniques (e.g., IPL V), and methods of analysis (probabilistic graph theory) is stressed.

The immediate aim is to demonstrate the possibility of an expandable Adaptive Man-machine Non-arithmetic Information Processing (AMNIP) system that could help a scientist (and perhaps other decision-makers also) in his technical judgments. This is being accomplished by the construction of (a) a sample data base of facts recorded by humans who translate from a variety of bibliographic and technical sources into a special language for tabular storage; (b) a system of IPL-V 7090 programs for updating, searching, finding clusters of related items (e.g., groups of articles with similar citation patterns), presenting users' interests by successive approximations, and computing elementary logical consequences; (c) a mathematical model for an AMNIP system to calculate capacity for long-range adaptation by reorganization of recorded data, through training by a population of users, to a useful degree; (d) formal, English-like languages for improved man-machine communication, and perhaps for translation of simple English text into a form suitable for automatic logical processing; and (e) demonstrations and experiments, using the above, to test hypotheses about information growth rates, man-machine conversation, hypothesis selection, search by sampling, use of citation patterns, and logical-syntactic transformations.

Concerning (a), about 5,000 sentences have been recorded, and this

exhausts the capacity of the 7090 IPL-V systems currently used. This data base will be converted and greatly augmented for storage on a photodisc, using existing machine-form biographic records and citations among patents. The system of programs referred to in (b) is about half completed. A new FORTRAN 7090 program for finding clusters in a large citation net has been completed. This entire system of programs is being transferred to a photostore-IBM-7090-console integrated complex to provide an on-demand experimental system.

The clustering algorithm applied to finding groups of people with similar reading interests produced two overlapping clusters. The acceptance rate of notices submitted by members of the same cluster was about 6 to 10 times that of notices submitted by nonmembers. The measure of similarity between the reading interests of two people (but applicable to estimating similarity between two rows of any incidence matrix) was shown to be statistically optimal. It was also used for inferring similarities in texts based on observed similarities in samples of text.

An important relation between the least mean search time *before* examination of a file location and the least mean search time *thereafter* was derived. This "search equation" applies to the design of new random-access storage and search systems. An important new technique for solving it in integers was also developed, with explicit solutions when the probability distribution of items over file locations is trapezoidal.

With regard to (d), the construction of a formal, English-like language (LOGOS) was extended by reconstruction of relative pronouns. A system for replacing parentheses occurring in sentences of this language by English groupers (e.g., either-or) and connectives was developed. Algorithms for sentence recognition and for transforming sentences of this formal language into propositions resembling those of an applied predicate calculus have been developed and are being programmed.

A system of IBM 7090 computer programs for automatic abstracting and indexing is nearly operational. It will permit experimentation with three techniques for clustering and ranking similar sentences or words in a text, based on co-occurrence statistics.

The U. S. Air Force Systems Command, U. S. Air Force Cambridge Research Laboratories, and U. S. Air Force Office of Scientific Research respectively, are providing partial support for various aspects of this project dealing with practical systems construction, AMNIP system analysis, and logical analysis of grammar (H. Bohnert, principal investigator of latter).

References:

(1) Abraham, C. T. "Sampling for Co-Occurrence," *Quarterly Report IV*, vol. 1, May 1962. Contract AF 19 (626)-10.

(2) Abraham, C. T. *Multi-stage Sampling for Estimation of the Number of Items Common to Two or More Lists of Items,* August 1962. Contract AF 19 (626) -10.
(3) Abraham, C. T. "A Note on a Measure of Similarity Used in the DICO Experiment," *Quarterly Report III,* vol. 1, February 1962. Contract AF 19 (626) -10.
(4) Bohnert, H. G. *A System of Grouping for English-Like Languages,* May 1962. (To be published)
(5) Bohnert, H. G. *The Logic of the Relative Pronoun "That."* (To be published)
(6) Grems, M. *Reference Manual—Glossary for Information Processing,* DPD Publication 8089, April 1962.
(7) Kochen, M. *High-Speed Document Perusal,* Final Technical Report, AFOSR-2871, April 1962. Contract AF 49 (638) -1062.
(8) Kochen, M., and E. Wong. "Concerning the Possibility of a Cooperative Information Exchange," *IBM Journal of Research and Development,* vol. 6, no. 2, April 1962, pp. 270-271.
(9) Kochen, M. "Adaptive Information Retrieval," *IBM Research News,* vol. 5, no. 5, June 1962.
(10) Kochen, M. *Adaptive Man-Machine Non-Arithmetic Information Processing,* Final Report, AFCRL TR 62-01, June 1962. Contract AF 19 (604) -8446.
(11) Reisner, P. "An Experimental Investigation of Clustering," *Quarterly Report IV,* vol. 1, May 1962. Contract AF 19 (626) -10.
(12) Wong, E. *A One-Dimensional Model of Search,* IBM Technical Note, August 1962.

2.63 INTERNATIONAL BUSINESS MACHINES CORPORATION

Advanced Systems Development Division, P. O. Box 344, Yorktown Heights, N. Y.
R. E. NIENBURG *and* T. R. SAVAGE

Development work is continuing on the SDI (Selective Dissemination of Information) Systems. Research activities in coding, human factors, and programming languages are also included.

The aim of the efforts is to provide an integrated automatic system which permits dissemination, retrieval, indexing, classification, and hard-copy distribution, all under computer control and on the basis of programmed (rather than human) decisions. Interrogation of the system should be available by means of ordinary English questions.

The SDI 2 System has been implemented on the IBM Tape 650 and through FORTRAN II. The SDI 3 System is implemented on the IBM 1401 computer. The SDI 4 System has been programmed for the IBM 7090 computer, and the SDI 5 System has been partially designed.

Experiments comparing the use of titles as opposed to titles and abstracts as notification devices, and comparing the effect of educational background on encoding performance, have been completed and the results published. Experiments comparing the effect of varying the number of index terms and selecting index terms by machine have been completed.

An experiment to determine the use of hard copies is in progress. Experiments on the effect of word truncation on indexing and the value of dissemination notification have been completed. An Automatic Record Analysis Language (ARAL) has been programmed to permit easy alphabetic string processing using the FORTRAN system.

References:

(For a complete list of earlier publications, see *CRDSD, No. 9,* Statement No. 2.30. Copies of these publications are available from Librarian, IBM, ASDD, Yorktown Heights, N. Y.)

(1) Resnick, A., and C. B. Hensley. *The Use of Diary and Interview Techniques in Evaluating a System for Disseminating Technical Information,* Report No. 17-055, IBM, ASDD, Yorktown Heights, N. Y., 1961.

(2) Rath, G. J., A. Resnick, and T. R. Savage. "The Formation of Abstracts by the Selection of Sentences," *American Documentation,* vol. 12, no. 2, April 1961, pp. 139-143.

(3) Rath, G. J., A. Resnick, and T. R. Savage. "Comparisons of Four Types of Lexical Indicators of Content," *American Documentation,* vol. 12, no. 2, April 1961, pp. 126-130.

(4) Resnick, A., and T. R. Savage. "A Re-evaluation of Machine-Generated Abstracts," *Human Factors,* August 1960, pp. 141-146.

(5) Brandenberg, W., et al. *Selective Dissemination of Information—SDI 2 System,* Report No. 17-031, IBM, ASDD, Yorktown Heights, N. Y., 1961.

(6) Hensley, C. B., et al. *Selective Dissemination, Report on a Pilot Study—SDI 1 System,* Report No. 17-039, IBM, ASDD, Yorktown Heights, N. Y., 1961.

(7) Resnick, A. "The Relative Effectiveness of Titles and Abstracts for Notification in a Selective Dissemination System," *Science,* vol. 134, October 6, 1961, pp. 1004-1006.

(8) Resnick, A. "Comparative Effect of Different Education Levels on Indexing in a Selective Dissemination System." (In preparation)

(9) Sowarby, A. J. *The SDI 3 System for the IBM 1401 Data Processing System,* IBM 1401 General Program Library, No. 10.3.004, June 25, 1962.

(10) Stiassny, S. *Coefficients of Similarity Between Documents,* Report No. 17.01.012.008, IBM, ASDD, Yorktown Heights, N. Y., September 1959.

2.64 IBM-FRANCE

5, place Vendôme, Paris 1, France
F. LEVÉRY

A series of experiments is planned in the fields of automatic indexing of technical texts and technical vocabulary analysis.

A statistical method will be tested to determine the degree of closeness in meaning of words. The method will consist of studying the pairs of words which appear together in the majority of texts and calculating a coefficient of correlation from the frequencies. Such work will result in a standard list of notions frequencies for a particular kind of information.

Starting from this list, new experiments will be made so as to obtain a list of keywords representing each text. The method will use statistical comparison between the distribution of frequencies of notions contained in a text and the standard distribution obtained for the entire corpus.

2.65 ITEK CORPORATION

Information Sciences Laboratory, Lexington 73, Mass.
J. W. KUIPERS, *Manager,* R.F. BARNES, *and* R. S. GLANTZ

A research program for the design and development of mechanized retrieval systems emphasizes grammatic normalization as a basic method for dealing systematically with linguistic expressions which may be presented for input to a system. In its major features the basic grammar is related both to graph theory and to functional logic. Major progress has been made in the development of the basic grammar and in procedures for its application in the normalized representation (L-representation) of natural-language expressions. Major progress has also been made on a prescriptive method, based on L-representation, for selecting and formulating subject index data for documents (L-indexing). A sample of some 3,000 scientific and technical papers dealing with aspects of nuclear energy has served as case material for development of methods and testing of results. Although procedures are not complete or fully tested, results at this stage in the development program show great promise for further extension and use. Reports on the present status of this work have been issued (1-3). The research program is being furthered in the directions outlined in the original program proposal (1). Investigations are also being made of the

mathematical properties of the sets of document descriptions generated by the application of L-representation methods. Another task group is concentrating on the specification of procedures for an operational system which employs normalized input data. Documents in the field of optics are being used for processing and for the development of auxiliary tools required for an integrated system. A Digital Equipment Corporation (DEC) PDP-1 computer, augmented with a 500,000-word disc memory and additional peripheral equipment, is being used to perform essential machine functions to make possible further studies on file organization, searching tactics, and the design of searching equipment.

Another active program now in progress involves setting up and operating a document-handling system with relatively uncontrolled input. At present, "document derivative data" to be input to the system are recorded in a machine-interpretable punched-tape format. A general-purpose computer is used to generate a multiplicity of data forms as required for various system functions, for searching files, for announcement lists, etc. "Document derivative data" for 10,000 documents have been recorded and are being used in experimental processing. To carry out the machine processing for this experimental system, the DEC PDP-1 computer with augmented disc memory is being used. Computer programs for major system functions have been completed. Programs for some machining functions have also been prepared for an IBM 1401 computer, which is being used for the several development projects.

Research and development tasks related to the programs described above are also being carried on in the following areas: computer composition and the generation of typesetting control data; time-shared computer programming; and computer programming with a flicker-free display capability for textual and graphic processing, dynamic factor analysis, indexing, decision processing, and editorial processing. A large number of computer programs and subroutines have been completed and tested and are now being used in a variety of applications.

A study and development project, carried on for the U. S. Air Force Cambridge Research Laboratories, is concerned with machine processing methods in a specialized document and library center (5). As part of this project, a special-purpose logic control device has been constructed which can permute, abridge, or repeat series of punched paper tape records automatically only on the basis of data contained in each record.

References:

(1) Kuipers, J. W. *Summary of Project Activities, Contract NSF-C88, IL-4000-19,* February 1961.

(2) Lipetz, B. A. "Compilation of an Experimental Citation Index from Scientific Literature," *American Documentation,* vol. 13, no. 3, July 1962, pp. 251-266.
(3) Williams, T. M., R. F. Barnes, and J. W. Kuipers. *Discussion of Major Features of a Restricted Logistic Grammar for Topic Representation,* IL-5206-26, February 1962.
(4) Barnes, R. F. "Language Problems Posed by Heavily Structured Data," *Communications of the ACM,* vol. 5, no. 1, January 1962, pp. 28-34.
(5) Reports submitted to U. S. Air Force Cambridge Research Laboratories, under Contract AF 19(604) 8438.

2.66 ITEK CORPORATION

Applied Technology Department, Lexington 73, Mass.
B. O'BRIEN, *Manager*

A program is continuing for the development of new automatic techniques for the high-speed transport and handling of unitized film records. Emphasis is on simple and flexible methods of controlling the routing of unit records in accordance with specific instructions or data which are descriptive of the information contained in the record.

Development of processing and viewing equipment is also continuing in related project tasks. Further improvement is also being made in the design of a unitizing printer used to produce sheet film unit records from 16 mm. or 35 mm. roll film.

2.67 日本科學技術情報センター [JAPAN INFORMATION CENTER OF SCIENCE AND TECHNOLOGY]

15 Itiban-tyo, Tiyoda-ku, C.P.O. Box 1478, Tokyo, Japan
丹羽保次郎 [YASUJIRO NIWA], *President, and*
三輪大作 [DAISAKU MIWA], *Vice President*

In the JICST (Japan Information Center of Science and Technology) development program, emphasis is placed on reinforcement of information processing, storage, and retrieval systems.

The JEIPAC, a transistorized information processing machine and a modification of the TOSBAC 4134 model manufactured by the Tokyo Shibaura Electric Company, has been installed. The experiment for retrieval of metals literature is continuing, using the modified classification codes of ASM-SLA system. Approximately 10,000 documents will be coded and stored tentatively on magnetic tape during 1962. The coding system for machine retrieval is also being developed in some parts of the chemical field and electrical engineering field.

The JEIPAC has also been programmed for automatic indexing designed after the IBM KWIC indexing system. An author index to

227,000 documents for the annual index (1961) to *Current Bibliography on Science and Technology* and a patentee index to 150,000 Japanese patents (1948-1961) were prepared by conventional EAM equipment, including the IBM Cardatype 858 Accounting Machine. The *Current Bibliography on Science and Technology* is the abstract journal which makes world literature available in the Japanese language; it covers 3,000 scientific periodicals of 42 countries.

Reference:

(1) Abe, Koji, and Hiroshi Mitsuhashi. "Experiments on Machine Retrieval of Metallurgical Literature by JICST Electronic Information Processing Automatic Computer (I), (II)," *JICST Monthly,* vol. 5, no. 6, 1962, p. 7. (In Japanese)

THE JOHNS HOPKINS UNIVERSITY 2.68

Baltimore 18, Md.
ROBERT H. ROY, *Principal Investigator*

The purpose of the investigation is the optimization of the operation of a university research library. Corollaries to this broad general purpose will be (a) formulation of a cost model, to include user costs as well as costs of operation; (b) development of decision rules for shelving, storage, etc.; (c) analysis and optimization of operational procedures; and (d) assessment of the applicability of computer procedures to circulation control, storage policy, and information retrieval.

The methodologies associated with operations research and systems engineering are being used in analyzing the various facets of library operation. Among these, cost accounting, various statistical techniques, and work sampling are being employed, and it appears very likely that inventory theory and possibly queueing theory will play a part. An IBM 1620 computer has been used and current developments envisage regular utilization of an IBM 1401 computer and possibly an IBM 7090 computer.

During the summer and fall of 1961 most of the research effort was directed toward organization of the team and familiarization with the overall problem. Within the past 6 months progress has been made as follows: A questionnaire directed toward assessment of library use has been devised, tested, and revised. The final document has been used throughout the entire library for three 1-week sample periods. Tabulation of data on the IBM 1620 computer has been completed, and results will be analyzed. Storage costs, not presently charged in current accounting procedures, have been measured by a decision, for use in cost models, to charge for space at cost of construction (measured in 1962 dollars) capitalized at the rate of return of the University Consolidated Investment Account. To this, library maintenance costs will

be added. This decision recognizes that these are sunk costs but, in a situation where available space is constantly less than needed, the decision seemed sound. These data have been obtained. Intensive studies of rates of acquisition are currently being made. Historical acquisition data, which are available for the entire library for the past 10 years, will be punched into IBM cards for analytical purposes. Plans are nearly completed for recording and storing on tape information on all current acquisitions. Germane to this, some field studies of shelf organization have been made and further sampling studies of this kind are planned. Anticipated objectives of these various steps will be to yield decision rules on: (a) shelving plans for a new library now under construction (in over-simplified terms, how much empty shelf space to leave between different classifications when books are moved into the new library); (b) when to "thin out" in order to avoid excessive reshelving as acquisitions continue. This point in time cannot yet be stated in quantitative terms, but it is known to be well in advance of saturation. The concept of "load factor" appears useful in this regard, and (c) how to "thin out," i.e., which items to dispose of or to remove to a more remote location.

With respect to computer utilization for circulation control purposes, all flow diagrams have been completed both for a trial system to be used in one branch library and for a prospective complete system to be used in the library as a whole. The trial system will be manual insofar as users and librarians are concerned, but it is anticipated that a fully developed method may utilize "charge-a-plate" cards for both books and patrons, cards which will produce information which can be "read" by the computer.

Phases of the project to be carried forward in the immediate future are (a) installation, operation, and evaluation of circulation control procedures in one branch library; (b) data processing and analysis of acquisitions and installation of routines for storing current acquisitions data on tape; (c) analysis of questionnaire responses; and (d) application of these steps to further development of the projected cost model.

The project is supported by the National Science Foundation.

2.69 JONKER BUSINESS MACHINES, INC.

404 N. Frederick Avenue, Gaithersburg, Md.
FREDERICK JONKER, *Principal Investigator*

One task concerns the establishment of a generalized treatment of the theory of information retrieval. The theory will be in three parts: index terminology, indexing methods, and retrieval methods and devices. The theory of index terminology postulates the existence of a

continuum of terminology. The theory of indexing methods postulates the existence of a continuum of methods. Both continua form a two-dimensional array. All practical indexing systems can be described by a point in this two-dimensional array. The third part describes the basic search approaches and discusses and compares the physical implementation of these.

The second task concerns an investigation of the requirements of information retrieval (I.R.) networks. I.R. networks comprise a center and a large number of remote sources of information and users of information. The most general type of I.R. network is a Planetary Network. The Exchange Network and the Dissemination Network are special cases of the Planetary Network. Planetary Networks require the merging of collections of data and information and methods of making their index vocabularies compatible. The centers of networks can utilize computers and other data processing equipment. Networks will also utilize various types of disseminable search devices and centralized or disseminable microform document storage systems.

This 2-year effort is supported by a U. S. Air Force Office of Scientific Research contract. Reports on the work are expected to become available toward the end of 1962.

KARLOVA UNIVERSITA [CHARLES UNIVERSITY] 2.70

Institute of Numerical Mathematics, Faculty of Mathematics and Physics, Linguistic Group, Krasnoarmějcu 2, Prague 1, Czechoslovakia
PETR SGALL, *Leader*

The results of the testing of an elementary model of language for information storage and retrieval on a short electrotechnical text are being evaluated. Several methods of automatic abstracting, i.e., statistical and semantical, have been considered (1).

Reference:

(1) Čulík, Karel, and Bohumil Palek. "Automatické referování" ["Automatic Abstracting"], *Metodika a technika informací*, no. 3, 1962.

LEHIGH UNIVERSITY 2.71

Department of Philosophy, Christmas-Saucon Hall, Bethlehem, Pa.
DONALD J. HILLMAN, *Assistant Professor of Philosophy*

Research is underway on a theoretical study of methods and techniques of information storage and retrieval. The purpose of the study is to provide a detailed analysis of the theories underlying the design

and operation of information systems and to set up a basis for future research and development.

The role played by models in theory construction is first described, and an investigation is conducted to determine the viability of certain models selected from the fields of symbolic logic, mathematics, and mathematical linguistics. Among the more important of these models are those based on an atomistic system of Boolean Algebra, the Calculus of Relations, *n*-dimensional Euclidean Geometry, and the Theory of Trees, respectively. On the basis of these models, and others related to them, it is expected that new theories will be developed and made subject to experimental corroboration.

A general study of information systems, their requirements, and the methodology of theory construction has been completed, and criteria established for the evaluation of theories. The scope of Boolean Algebra and related models based on the propositional calculus has been studied and theoretical limitations imposed. The next phase of the project will be concerned with the applicability of models based on qualificational logic.

The research is sponsored by the National Science Foundation and is of 2 years' duration.

References:

(1) Hillman, Donald J. "Problems, Systems, and Methods." in *Study of Theories and Models of Information Storage and Retrieval,* Report No. 1, August 3, 1962.

(2) Hillman, Donald J. "The Boolean Algebra Model," in *Study of Theories and Models of Information Storage and Retrieval,* Report No. 2, August 3, 1962.

(3) Hillman, Donald J. "A Positive Model for Systems of Special Classification," in *Study of Theories and Models of Information Storage and Retrieval,* Report No. 3, August 29, 1962.

2.72 ЛЕНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ А. А. ЖДАНОВА

[LENINGRAD STATE UNIVERSITY IMENI A. A. ZHDANOV]

Experimental Laboratory of Machine Translation, Leningrad, U.S.S.R.

Н. Д. Андреев [N. D. ANDREEV], *Laboratory Director*

Research is in progress on information retrieval in the field of radioelectronics. The first object of work has been the construction of a special vocabulary which connects the †*Intermediary Language* (IL) with the †*Retrieval Language* (RL). The RL vocabulary now includes more than 3,000 compound units called †*koinoglyphs* (*each* koinoglyph consists of several †*semoglyphs*) and is growing at an average rate of 15 koinoglyphs per month.

† *See* Glossary.

Information retrieval in the field of juridical literature is being studied, and a symbolic system for writing simple combinations of juridical notions has been developed.

Another study in information retrieval is concerned with automating the dispatcher service of the merchant marine. Work on vocabulary and semantic analysis of ships' correspondence is in progress. Here too the RL is connected with the IL.

References:

(1) Андреев, Н. Д., Д. А. Керимов [Andreev, N. D. and D. A. Kerimov]. "Возможности использования кибернетической техники при решении правовых проблем" ["The Possibilities of Using Cybernetic Technique in Solving Legal Problems"], in Кибернетика на службу коммунизму [*Cybernetics in the Service of Communism*], vol. 1, 1961, pp. 234-241.

(2) Андреев, Н. Д. [Andreev, N. D.]. Машинные языки как элемент универсального кода науки [*Machine Languages As An Element of a Universal Code of Science*], Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста [Reports at the Conference on Information Processing, Machine Translation and Automatic Text Reading], Moscow, 1961. Translation in JPRS 13172, Foreign Developments in Machine Translation and Information Processing, available from OTS.

LIBRARY ASSOCIATION 2.73

Chaucer House, Malet Place, London, W.C. 1, England
D. J. FOSKETT

Work underway involves study of a new general scheme of classification and production of a pilot scheme.

The project is supported by a grant from the Scientific Committee of NATO.

LIBRARY OF CONGRESS 2.74

Washington 25, D.C.
GILBERT W. KING, *Survey Director*

A six-member team survey of the possibilities for automating large research libraries, with the Library of Congress as primary focus of the survey, is in process. The emphasis of the survey is on the

information organization, storage, and retrieval functions of libraries whose collections number in the millions and which serve research through the availability of both current and retrospective literature.

The project is supported by the Council on Library Resources, Inc. The survey team expects to complete its task and submit a final report in September or October of 1962. Plans have been made for publication of the report.

Reference:

(1) Dubester, Henry J. "Mechanization of Subject Headings," *Library Resources and Technical Services,* vol. 6, no. 3, Summer 1962, pp. 230-234.

2.75 LIBRARY RESEARCH CIRCLE *and* DOCUMENTATION RESEARCH AND TRAINING CENTRE

696 Cross Road 11, Bangalore 3, India
S. R. RANGANATHAN, *Chairman*

Two reports on the investigation of mnemonics in classificatory language, conducted by Abdul Rahman and T. Ranganathan, have been published. In nonseminal mnemonics (1), a comparative study is made of the incidence of systematic mnemonics and of four kinds of scheduled mnemonics in CC and DC. In seminal mnemonics (2), the occurance of seminal mnemonics in CC is exhaustively tabulated. After a description of the potency of such mnemonics, an ideal scheme of classification with not more than a few seminal schedules is indicated as a field for research.

In the Documentation Research and Training Centre (3), which was recently founded at Bangalore by The Indian Statistical Institute, S. R. Ranganathan and A. Neelameghan are designing a depth classification of pharmacology.

References:

(1) Rahman, Abdul, and T. Ranganathan. "Nonseminal Mnemonics," *Annals of Library Science,* vol. 9, 1962, pp. 1-14.

(2) Rahman, Abdul, and T. Ranganathan. "Seminal Mnemonics," *Annals of Library Science,* vol. 9, 1962, pp. 53-67.

(3) Indian Statistical Institute, Documentation Research and Training Centre, *Course of Training in Documentation, Prospectus and Syllabus,* 1962.

MAISON DES SCIENCES DE L'HOMME **2.76**
and
ECOLE PRATIQUE DES HAUTES ETUDES, VIe Section
Service d'Etudes Sémiologiques et Documentaires (SESD),
23 rue du Maroc, Paris 19e, France
J.-C. GARDIN

The provisional Bureau d'Etudes sur le Traitement Automatique de l'Information dans les Sciences Humaines has become a permanent department of the Maison des Sciences de l'Homme and the Ecole Pratique des Hautes Etudes, VIe Section, under a new name, as indicated above. In addition to the functions of the former Bureau d'Etudes (see *CRDSD,* No. 9, Statement No. 2.38), the Service d'Etudes Sémiologiques et Documentaires [Centre for Semiological and Documentation Studies] has taken up a program for training graduate students in the study of "systems of signs" (natural or artificial languages, codes, etc.), for practical applications in the field of automatic documentation. Courses will be given, beginning October 1962, at the Ecole Pratique des Hautes Etudes; part of them will be devoted to the mathematical aspects of such studies.

The main activity of the group, however, is devoted to completing a series of experiments with SYNTOL (Syntagmatic Organization Language), a general system for automatic documentation devised under a contract with Euratom. The logico-linguistic rules of the system are described in the final report to Euratom, part 1 (1). Around 3,000 abstracts (mainly in English and French) have been "manually" translated into SYNTOL and submitted to an experimental retrieval program for the IBM 7090 computer (2). The fields covered are physiology, psychology, sociology, and cultural anthropology. Two of the specialized dictionaries which have been prepared for each field have been mimeographed for illustration purposes (3). The dictionary for sociology and cultural anthropology, prepared in collaboration with the Centre d'Analyse Documentaire pour l'Afrique Noire (4), will be ready for distribution in 1963.

The experiment has consisted of processing three sets of about 15 questions, related to the different fields of the above corpus, and comparing the answers obtained to a pre-established list of pertinent documents, empirically determined for each question (5). A critical evaluation of the results has been undertaken with the collaboration of the Centre d'Analyse Documentaire pour l'Afrique Noire; results will be published in 1963.

Part of the experiment is supported by funds from the Ford Foundation.

References:

(1) *Le SYNTOL (Syntagmatic Organization Language), étude d'un système général de documentation automatique,* Rapport Final adressé par l'Association Marc Bloch à l'Euratom; Partie 1: *Aspects théoriques,* by J.-C. Gardin.

(2) *Ibid.,* Partie 2: *Programmation sur IBM 7090,* by R.-C. Cros.

(3) *Ibid.,* Partie 3: *Exemples de Lexiques,* by F. Lévy, Natacha Gardin, and Radmila Zygouris; fasc. A, *Introduction et Champ commun;* fasc. B, *Physiologie;* fasc. *C., Psychologie.*

(4) *Ibid.,* fasc. D, *Sociologie et Ethnographie,* by Françoise Izard. (In preparation)

(5) Gardin, J.-C., and F. Lévy. *Le SYNTOL, ses propriétés et ses applications en documentation automatique,* presented at the International Federation of Information Processing Societies conference, Münich, August 1962.

2.77 MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge, Mass.

M. M. KESSLER *and* W. N. LOCKE

A project underway seeks to design a model for a technical information system.

The first phase of the work is concerned with the preparation of a body of factual information that will give guidance to system design. The working medium consists of some 10,000 articles published in various physics journals.

The results so far have been published (1-5).

The second phase of the work, starting around November 1962, will be concerned with several possible system configurations.

The 1-year project is sponsored by the National Science Foundation.

References:

(1) Kessler, M. M. *An Experimental Study of Bibliographic Coupling between Technical Papers,* R-1, M.I.T., November 21, 1961; revised June 15, 1962.

(2) Kessler, M. M. *Bibliographic Coupling Between Scientific Papers.* R-2, M.I.T., July 9, 1962.

(3) Kessler, M. M. and F. E. Heart. *Analysis of Bibliographic Sources in the Physical Review (Vol. 77, 1950 to Vol. 112, 1958),* R-3, M.I.T., July 13, 1962.

(4) Kessler, M. M. *Analysis of Bibliographic Sources in a Group of Physics-Related Journals,* R-4, M.I.T., August 6, 1962.

(5) Kessler, M. M. *Biblographic Coupling Extended in Time: Ten Case Histories,* R-5, M.I.T., August 20, 1962.

(6) Kessler, M. M., and F. E. Heart. *Concerning the Probability That a Given Paper will be Cited,* R-6, M.I.T., November 5, 1962.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY 2.78

Computation Center and Research Laboratory of Electronics, Cambridge, Mass.
HERBERT M. TEAGER, *Principal Investigator*

Research is continuing into on-line, real-time devices, languages, and systems. The systems problems associated with time-sharing have largely been solved; and development is continuing in producing a basic algebraic compiler, scheduling and monitor routines, and a basic filing and retrieval system for programs, data, and output. IO hardware configuration problems have largely been resolved with a choice of typewritten, plotter, and graphical input capability at each console. The feasibility of personalized, hand-drawn input has been explored, and operational software is being developed for this purpose. Stress has accordingly been shifted to applications for which the real-time aspect of computation is the key issue, and for which graphical languages can provide a substantial improvement. The areas of electrical field and network problems are being investigated with a secondary emphasis on the development of general subroutines for some common information processing requirements, such as inversion of partially symbolic matrices and conversion of a boundary value problem into a set of difference equations.

The research is supported by the U. S. Office of Naval Research and the National Science Foundation.

References:

(1) Teager, H. *Real-Time, Time-Shared Computer Project,* Sixth and Seventh Quarterly Progress Reports, October 8, 1962.
(2) Wilde, D. *Electrical Network Analysis by Digital Computer.* Master's Thesis. Cambridge, Mass.: Massachusetts Institute of Technology, September 1962.

NATIONAL ACADEMY OF SCIENCES-NATIONAL RESEARCH COUNCIL 2.79

Division of Chemistry and Chemical Technology, Washington 25, D.C.
I. MOYER HUNSBERGER, *Principal Investigator*

The Chemical Notations Systems Project, in which a thorough analysis is being made of the characteristics of all codes and notations as well as of other nonconventional systems presently used to designate chemical structures, is drawing to a close.

The final report on the study is being prepared and will be submitted to the project sponsor, the National Science Foundation.

2.80 NATIONAL BIOMEDICAL RESEARCH FOUNDATION

8600 16th Street, Silver Spring, Md.
R. S. LEDLEY, F. S. ZUSMAN, L. S. ROTOLO, *and* J. B. WILSON

Computer programs have been prepared to compile a TABLEDEX index, a coordinate index that retains the convenience and familiarity of a bound-book form and that can also be automatically organized on a digital computer. A TABLEDEX index basically consists of three parts: The first is a retrieval-word list of index terms used; the second is the TABLEDEX tables, or index proper; the third is the bibliography proper. A conventional index of authors and special terms is also often included.

In order to demonstrate the appearance and use of a TABLEDEX index, the computer programs were applied to a bibliography of scientific articles prepared for the International Geophysical Year by the Library of Congress. The input to the computer programs consists of bibliographic reference cards giving the article citation and the associated retrieval words and other annotations. These cards are first transcribed onto IBM punchcards from which they are recorded onto magnetic tape. The computer programming system, which is written for the IBM 7090 computer with IBM 1401 computer input-output, processes the information recorded on the magnetic tapes and then generates automatically the various parts of the TABLEDEX format from this tape. The computer programming system assigns accession numbers to the documents, extracts the retrieval words from the article abstract and annotations (as previously marked by the Librarian), prepares an alphabetical list of all words utilized as retrieval words, prepares the TABLEDEX tables, develops the author and special term index, determines document frequency for each retrieval word and assigns frequency numbers to these terms, organizes the number TABLEDEX tables, and prints out all this information as requested for any particular bibliography.

The *International Geophysical Year Bibliography* index (2) presents two different TABLEDEX formats as printed out by the computer, in order to permit comparison of these two different formats of a TABLEDEX index in actual use. Thus, this IGY volume consists of the following six basic parts: Part I, which is the list of retrieval words used in the bibliography, includes authors' names and other special terms which are distinguished by asterisks; Part II consists of the alphabetical TABLEDEX tables by means of which the searches are made; Part III is the index of authors and special terms; Part IV is a list of retrieval words with frequency code numbers for use with Part V; Part V is the number TABLEDEX tables by means of which searches can be

made in terms of code numbers for the retrieval words; and Part VI is the bibliography proper or list of article citations.

The project was supported by the National Science Foundation. The bibliography material was compiled and the retrieval words were determined by Mendall S. Thompson and Thomas D. Gomery, of the Library of Congress.

References:

(1) Ledley, Robert S. "TABLEDEX: A New Coordinate Indexing Method for Bound Book Form Bibliographies," in *Proceedings of the International Conference on Scientific Information,* pp. 1221-1243. Washington, D. C.: National Academy of Sciences-National Research Council, 1959.

(2) Zusman, Fred S., Mendall S. Thompson, James B. Wilson, Louis S. Rotolo, and Thomas D. Gomery. *The International Geophysical Year Bibliography: An Example of Tabledex Formats,* a joint project by National Biomedical Research Foundation and the Library of Congress, 1962.

NATIONAL BOOK LEAGUE 2.81

7 Albemarle Street, London, W. 1, England

BARBARA KYLE,[1] *Social Sciences Documentation*

Four international bibliographies covering the fields of economics, political science, sociology, and social anthropology, published by the International Committee on Social Sciences Documentation of Unesco, are being classified by a new "faceted" classification for literature of the social sciences, in preparation for tests of the new system with systems currently in use.

In February 1962 a preliminary report on the work to date was issued (1). In March 1962 a short test was made to determine how far two indexers, both knowing the system of classification and indexing to be used, could achieve consistency. A total of 246 titles (those ending in digit 5) from the *International Political Science Bibliography* were indexed and classified. Seventy-percent consistency was achieved. The causes for the 30-percent inconsistency (18 percent being of some importance) have been analyzed, and modifications in the system have been made in an attempt to eradicate them in the future.

The following work is planned before the completion of the project: (a) further consistency tests in fields other than political science, (b) testing of efficiency of indexing in social anthropology and political science, and (c) preparation of a final report.

The project is supported by the National Science Foundation.

[1] Present address: Aslib, 3, Belgrave Square London, SW.1, England

References:

(1) Kyle, Barbara. *First Report on the Preparation for Tests of the Kyle Classification for the Social Sciences,* National Book League, February 1962.

(2) Kyle, Barbara. *Consistency Analysis of Two Indexers in Using K. C. for Political Science Material,* National Book League, May 1962.

2.82 NATIONAL BUREAU OF STANDARDS

Data Processing Systems Division, Washington 25, D. C.
S. N. ALEXANDER

I. Office of Technical Services Study (Ethel Marden and Patrick E. Doyle)

The purpose of a cooperative study with the Office of Technical Services (OTS), Department of Commerce, is to determine the contributions which automatic information handling techniques might make to OTS services and efficiency. The first phase of the study has progressed to the experimental automatic production of indexes to reports announced by OTS.

A related task has been the production of a permuted title index to the reports from the OTS collection normally announced in *U. S. Government Research Reports* (1). The efforts devoted to the production of the permuted title index have contributed to the work on more detailed and comprehensive indexes. This task is being supported by the National Science Foundation.

Acquisition of Systematics' Universal Code Tape-to-Card Converter (punch paper tape) and its companion Card-to-Tape Converter has rendered more feasible the utilization of data available from other agencies in machine form. Such equipment may make possible the economic utilization of automatic composing and printing devices in the production of OTS publications.

In a second phase of the study, present and potential functions of the massive 3- by 5-inch card catalog maintained by OTS are being analyzed in order to design a machine file to fulfill those functions. Concentrated attention is being given to the functions of the file related to the preparation of the announcement journals and indexes and to the satisfaction of reference interrogations.

II. Food and Drug Administration Study (Ethel Marden)

An investigation was made of the information processing requirements of the Food and Drug Administration (FDA), with particular emphasis on the feasibility of applying machine techniques to all or part of FDA's information processing activities. The study included FDA's needs with respect to the recording and storing of technical in-

formation, its need for an integrated file system which would incorporate data from several organizational units within FDA, and the potential demands on a mechanized file for information retrieval. Some consideration was also given to the kind of equipment which would be necessary to process FDA's scientific and technical information. A report containing the findings of the feasibility study has been submitted to FDA (2).

III. PATENT RETRIEVAL SYSTEM (Ethel Marden)

The HAYSTAQ project, a part of the cooperative program with the U. S. Patent Office for mechanizing patent search operations (see 2.114), is a series of computer programs for processing and searching chemical information. Data for the chemical structure search program are now being processed by several data preparation routines in preparation for making extensive machine searches. The result is a steadily accumulating magnetic tape file of coded data which represent the chemical compounds making up the file.

Experimental computer programs were also written for performing screening operations to segment the file as a preliminary step to making detailed searches; there have not yet been sufficient machine runs to evaluate their utility. Research is continuing into investigations of new methods for searching chemical structure information, as well as along two ancillary lines of investigation: (a) increased machine preparation of chemical structure data for mechanized searching, and (b) the development of grammars or languages for the expression of technical information which is to be manipulated by automatic equipment.

References:

(1) U. S. Department of Commerce, Office of Technical Services, *Keywords Index to U. S. Government Technical Reports (Permuted Title Index)*, vol. 1, no. 1, June 15, 1962. (Issued semimonthly)

(2) National Bureau of Standards, Data Processing Systems Division. *Findings, Conclusions and Recommendations of a Cooperative Study of the Work Processes, Proceedings, and Systems Involved in the Collection, Creation, Evaluation and Application of Chemical and Biological Information and Data in the Bureau of Medicine, Bureau of Biological and Physical Sciences, and FDA District Offices,* report to the Food and Drug Administration, Department of Health, Education, and Welfare, May 1962.

(3) Koller, H., E. Marden, and H. Pfeffer. "The HAYSTAQ System: Past, Present, and Future," in *Proceedings of the International Conference on Scientific Information,* pp. 1143-1179.

Washington, D. C.: National Academy of Sciences-National Research Council, 1959.

(4) Moore, R. T. "A Screening Method for Large Information Retrieval Systems," in *Proceedings of the Western Joint Computer Conference,* Los Angeles, May 1961, vol. 19, pp. 259-271.

2.83 NATIONAL BUREAU OF STANDARDS

Basic Instrumentation Section, Washington 25, D. C.
JOSHUA STERN, *Chief*

The Aslib findings with respect to a fixed limit for indexing time to be spent per document (see 2.9) were tried experimentally on a 10,000-document sample. The majority of documents indexed were either journal articles in the physical and engineering sciences or abstracts from journals such as *Nuclear Science Abstracts.*

For the document sample involved, it was found that a maximum limit of 5 minutes per document is not only satisfactory but also desirable; the time spent beyond 5 minutes does not result in any better indexing. (At present, 10 to 12 terms and 7 to 8 concepts are selected per document.) Indexing time as used here includes all operations carried out by the indexer/analyst/coder except the preparation of abstracts and citations. Specifically, this involves scanning the document, selecting the most important concepts, translating the selected concepts into index terms through the use of the coding vocabulary, and writing down the set of index terms on a form. Two full-time indexers (professional physical scientists) were involved in the trial, and will continue to index the remainder of the collection within set time limits of 5 minutes or less.

Time records show that abstracts take about one-half the time of journal articles; "Either you can do them or you can't, and you know which right away." Time-consuming difficulties most often arise with journal articles when the emphasis of the title differs from that of the abstract or of the document itself or when the emphasis of the abstract differs from that of the document. The latter case is the most severe.

2.84 NATIONAL BUREAU OF STANDARDS

Data Processing Systems Division, Washington 25, D. C.
M. E. STEVENS

The Research Information Center and Advisory Service on Information Processing (RICASIP) is engaged in a continuing program of collection of information and maintenance of current awareness on research and development activities in the field of information processing and literature retrieval and the preparation of state-of-the-art reviews in various areas of this broad field. The most recent state-of-the-

art review is concerned with information selection systems which retrieve replica copies. A survey concerned with automata theory is in progress.

Current planning is directed toward indexing selected subsets of items from the collection for storage and retrieval, using equipment such as the Rapid Selector, Lodestar, and Microcite.

RICASIP is jointly sponsored by the National Science Foundation and the National Bureau of Standards.

References:

(1) Marden, Ethel C., and Herbert R. Koller. *A Survey of Computer Programs for Chemical Information Searching,* NBS Technical Note 85, Washington, D. C., February 1961. Available from Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C., PB-161586.

(2) Stevens, Mary E. *Automatic Character Recognition: A State-of-the-Art-Report,* NBS Technical Note 112. Available from Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C., PB-161613.

(3) Bagg, T. C., and M. E. Stevens. *Information Selection Systems Retrieving Replica Copies: A State-of-the-Art Report,* NBS Technical Note 157, December 31, 1961. Available from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. Price: $1.25.

NATIONAL CHIROPRACTIC ASSOCIATION 2.85

Department of Research and Statistics, 201 Palatine Drive, Alhambra, Calif.

HENRY G. HIGLEY, *Director*

A study is underway to determine the feasibility of setting up a system for abstracting, classifying, coding, and retrieving information pertaining to all aspects of the human spine and related structures. A relatively small, highly specialized system is needed. The system must be able to retrieve information as well as data. It must be flexible, capable of future expansion, and economical. It would be advantageous if the material could be retrieved on at least two levels: (a) material directly related to the question, and (b) material indirectly or partially related to the question.

The first part of the program consists of determining the function that the information must perform, areas of inquiry, and the frequency with which the various types of inquiry occur. The data are being obtained by reviewing the requirements of the Department of Research and Statistics in the past 3 years and requesting similar information

from the other departments and agencies of the National Chiropratic Association.

Secondly, a study is being made to locate the principal sources of information. It is estimated that some 1,600 documents directly related to the subject area are published yearly; about one-half of them are listed in *Index Medicus.*

When the project is completed, the information obtained will be used for the design of a system of classification and coding and the preparation of a thesaurus.

It is hoped that the project, which was begun in the summer of 1962, will be completed within a year.

2.86 NATIONAL LIBRARY OF MEDICINE

Bethesda 14, Md.

SEYMOUR I. TAINE, *Chief, Biblographic Services Division*

The project for the design, development, and installation of an electronic information storage and retrieval system known as MEDLARS (Medical Literature Analysis and Retrieval System) continues. The following objectives have been defined for this novel, computer-based indexing and bibliographic service in the field of medicine: (a) compilation on a recurring, scheduled basis of bibliographies in relatively broad subject fields; (b) compilation, on a demand basis, of bibliographies in more restricted subject areas; and (c) assembly and high-speed printout of camera-ready copy for *Index Medicus* and other formal bibliographic publications.

MEDLARS inputs per year will consist of approximately 250,000 English and foreign journal articles and 13,000 monographs which will be indexed in depth by indexers using the *Medical Subject Headings* list, presently being revised and expanded. The inputs will be converted into punched paper tape by input typists, proofed by editorial personnel, and read into the computer sequentially for storing in digital form on magnetic tape. Major outputs from the system may be either from the computer printer or a Graphic Arts Composer. In order to provide high-quality, compact, and fast output, a magnetic tape driven Graphic Arts Composer is being developed to compose a total of 221 different characters in font sizes varying between 6 and 14 points at speeds of 440 characters per second. The film output from this composer will be used for preparing lithographic plates for printing *Index Medicus, Cumulated Index Medicus Bibliography of Medical Reviews,* and other formal bibliographic publications. Complex retrieval will be conducted by using the logical conjunction, disjunction, and negation with logical conjunction of non-negated elements of subject headings and other terms, such as language, journal title, author, place and year of publication, form, and entry date into the system.

The General Electric Company has successfully completed the system design, and in February 1962 began a 1-year effort on the detailed development and engineering of the system. The final phase of the project, involving the ordering of equipment, installation, and implementation, will begin in September 1962 and is scheduled for completion in the fall of 1963.

The project was begun in 1961 for a 2½-year period under contract with the Information Systems Operation, Defense Systems Department, General Electric Company.

NATIONAL PHYSICAL LABORATORY 2.87

Autonomics Division, Teddington, Middlesex, England
A. R. MEETHAM

The study of a machine-generated descriptor language for indexing and retrieval of the propositions of Euclid in English is being prepared for publication.

Methods of evaluating retrieval systems are being investigated in more detail.

OCTROOIRAAD (Netherlands Patent Office) 2.88

The Hague, Netherlands
G. J. KOELEWIJN *and* J. DEKKER

Research is continuing on the possibilities of far-reaching mechanization of novelty search in the patent literature. Through the cooperation of the U. S. Patent Office (see 2.114), a test will be run on the ILAS. The test will include about 3,000 patent specifications in the file on carburetors. It is hoped that the test will provide some insight into the use of interfixes in practice.

The possibilities of deep indexing in the field of punched-card machines and electronic computers are being studied. As far as functional components of these machines (e.g., printing mechanisms, magnetic storage devices, adders, and analog-digital converters) are concerned, the research group has become convinced that fairly deep indexing is possible without using the interfix. A test system for printing mechanisms has been completed and was tested during the summer of 1962. A system for analog-digital converters is expected to be completed for test and operational use in autumn of 1962. Both systems use direct coding on IBM cards. At present, a simple sorting machine is used for searching. A start has been made on studying the indexing of patent specifications describing computers; these specifications are often very voluminous and time consuming in manual searching. Due to lack of time and manpower, this latter study has been temporarily suspended.

Reference:

(1) Koelewijn, G. J. "Mechanical Selection on the Basis of Concepts and Their Relationships," *ADIA Proceedings,* Beiheft zu den Nachrichten für Dokumentation, Frankfurt am Main, 1961, pp. 208-216.

2.89 PACIFIC SOUTHWEST FOREST AND RANGE EXPERIMENT STATION

U. S. Forest Service, Box 245, Berkeley 1, Calif.
THEODOR B. YERKE, *Librarian*

The study to determine the usefulness of a library card catalog in forestry that is based on the Oxford classification system and supplemented by an alphabetical subject index continues, with no change reported since the previous statement [*Ed.*].

2.90 PLANNING RESEARCH CORPORATION

1333 Westwood Boulevard, Los Angeles 24, Calif.
V. A. OSWALD, JR., *and* W. B. MITCHELL, *Senior Associates*

Work on the development of techniques for exploiting unformatted textual data continues, with no change reported since the previous statement [*Ed.*]. The project is sponsored by International Business Machines Corporation.

2.91 PRINCETON UNIVERSITY

Statistical Techniques Research Group, 167 Nassau Street, Princeton, N. J.
JOHN W. TUKEY

Experimentation with a citation index for the journal literature of statistical methodology continues. Project purposes include (a) development of convenient and effective compressed codes and coding systems, (b) experimentation with the usefulness of various formats of citation indexes, and (c) preparation of an effective citation index for statistics. The work is supported by the National Science Foundation.

Reference:

(1) Tukey, John W. "Keeping Research in Contact With the Literature: Citation-Indices and Beyond," *Journal of Chemical Documentation,* vol. 2, no. 1, January 1962, pp. 34-37.

PROJECT LAWSEARCH 2.92

1028 Connecticut Avenue, NW., Washington 6, D. C.
WILLIAM H. B. THOMAS

Project Lawsearch, which is directed toward the development of a mechanical search system, is nearing completion. Coordinate indexing is used. The search technique is that of optical coincidence, using superimposable or peek-a-boo cards. Approximately 2,000 judicial decisions and their legal materials have been indexed in the areas of the liability of motor carrier for personal injury to passengers and for loss of or injury to goods, and in a selected area under the Federal Motor Carrier Act. A list of about 800 index terms or subject headings has been derived. The terms may be combined or coordinated to effectuate a search based on facts or law, or both.

From the work of the project, information will be available relating to the nature and type of material suitable for a system of the kind proposed, the feasibility of various levels of indexing depth, and the various means of establishing index terminology and the importance of operational procedure in the processing of raw index data.

The project is sponsored by the Council on Library Resources, Inc.

PULP AND PAPER RESEARCH INSTITUTE OF CANADA 2.93

3420 University Street, Montreal 2, Quebec, Canada
and
THE INSTITUTE OF PAPER CHEMISTRY
Appleton, Wisc.
J. E. TASMAN,[1] P. M. NOBBS,[1] and EDWARD F. THODE[2]

A retrieval system is being developed for literature of interest to the pulp and paper industry. The project has as its goal the establishment of a framework within which many systems and subsystems may operate in full compatibility, from the manual index in a small mill laboratory to a large machine-oriented system in a central documentation center. A subsidiary goal is the development of methods appropriate for machine retrieval at the documentation centers of the cooperating institutions.

In order to insure compatibility with systems in other process industries, as much use as possible is being made of the information retrieval plan adopted by the American Institute of Chemical Engineers and outlined in the *Chemical Engineering Thesaurus* published by that society.

Work on the project, which is still in the early stages of development,

[1] Pulp and Paper Research Institute of Canada
[2] Institute of Paper Chemistry

has been largely directed toward the compilation of two thesauri of keywords: one covering pulp and paper terminology; the other, forestry terminology. Preliminary drafts of these thesauri, which are compatible with the *Chemical Engineering Thesaurus,* have been completed. Other work thus far accomplished has been (a) the machine generation of an index of all possible "four-letter words" for use in possible arbitrary assignment of mnemonic codes, and (b) the examination of various logical combinations for machine searching. One indication is that exclusion logic will be essential in retrieval from large collections.

In the next phase of the development, it is intended to carry out experimental concept-coordination indexing, with the aid of the draft thesauri, of some 3,000 to 5,000 abstracts taken from the *Abstract Bulletin of The Institute of Paper Chemistry.* In conjunction with this, a keyword file will be set up which will allow manual or machine retrieval of abstracts by identification number according to various search arguments. Statistical analysis of the results will be used to compare the usefulness of various combinations of logic patterns in improving the thesauri, and to assess the quality of the retrieval being obtained.

2.94 RADIO CORPORATION OF AMERICA

Data Systems Center, Bethesda, Md.
JACK MINKER, *Manager, Information Technology*

The ACSI-MATIC system, a large-scale information processing system, is approaching an initial operational status in the Pentagon where it will serve as an important adjunct to the task of intelligence analysis.

The initial system will employ a Sylvania 9400 computer as the central processor operating with a Telex Mass Memory Module disc file system and conventional peripheral equipment. Several pieces of equipment have been developed especially for integrating the standard items in the system. One of these is a special analyst's console which will enable the intelligence analyst to communicate his choices, wishes, or requests directly to the computer in simple English and standard military terminology. A prototype of this has been built and is undergoing evaluation.

The ACSI-MATIC programming system is particularly designed for the problems which large-scale nonscientific processing entails. It consists of two parts: (a) an executive program that provides a centralized input-output and program control feature, and (b) an assembly program which produces allocatable machine code for the purpose of providing dynamic storage allocation.

The files which contain the intelligence data are provided with extensive glossaries and hierarchical listings which indicate what terms are included in the system and just how all terms are interrelated.

The initial executive program, the assembly program, and the thesaurus routines have undergone extensive checking and are fully operational. Routines required to load the thesaurus data base are debugged, and loading of this data base has commenced. Routines required to process information into records are being debugged and integrated into the system. The retrieval strategy has been worked out fully and is undergoing debugging. Once the information data base is loaded, the generalized retrieval routine will be tested.

Several scientific intelligence problems utilizing information from the primary files of the ACSI-MATIC system are also under study.

The project is sponsored by the Office of the Assistant Chief of Staff for Intelligence (OACSI), Headquarters, Department of the Army.

References:

(1) Climenson, W. D., N. H. Hardwick, and S. N. Jacobson. "Automatic Syntax Analysis in Machine Indexing and Abstracting," *American Documentation,* vol. 12, no. 3, July 1961, pp. 178-183.

(2) Colilla, R. A., and B. H. Sams. "Information Structures for Processing and Retrieving," *Communications of the ACM,* vol. 5, no. 1, January 1962, pp. 11-16.

(3) Gurk, H. M., and J. Minker. "The Design and Simulation of an Information Processing System," *Journal of the Association for Computing Machinery,* vol. 8, no. 2, April 1961.

(4) Minker, J. "Implementation of Large Information-Retrieval Problems," presented at the Gordon Research Conference, New Hampton, N. H., July 1961.

(5) Sams, Burnett. "Dynamic Storage Allocation for an Information Retrieval System," *Communications of the ACM,* vol. 4, no. 10, October 1961, pp. 431-435.

READING CHEMISTS' CLUB 2.95

3103 River Road, Reading, Pa.
WILLIAM J. WISWESSER

The study of the ability of chemists to learn to decode the line-formula notation continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the National Science Foundation.

RENSSELAER POLYTECHNIC INSTITUTE 2.96

Department of Management Engineering, Troy, N.Y.
NORMAN F. BARNES

Studies focused on the development of information storage and retrieval criteria, including user requirements, and their utilization to evaluate specific systems continue, with no change reported since the previous statement [*Ed.*].

2.97 ROCKFORD RESEARCH INSTITUTE INCORPORATED

140½ Mount Auburn Street, Cambridge 38, Mass.

CALVIN N. MOOERS

Study of the theory and practice of information technology continues with support from the U. S. Air Force Office of Scientific Research, the National Institutes of Health, and the Gmelin Institute. The theoretical research supported through Zator Company is being transferred to Rockford Research, an associated nonprofit entity.

The TRAC procedure-describing language is undergoing revision and trial by example. Offers for production of an interpretive computer program for the implementation of TRAC have been received.

The "reactive typewriter" project now has two Siemens teleprinters, and arrangements are being made for various forms of connection to nearby computers over commercial wire circuits (e.g., Western Union Telex and Bell System). E. S. Fergusson is assisting in this project. A report describing the reactive typewriter concept is in press (1).

Work continues on the "Mooers Revision" of ASTIA descriptors. Scope notes and scientifically based definitions for each of the 1,600 descriptors and the 95 groups have been produced. After final editorial revision, the schedule of descriptors with the scope notes will be published.

Assistance is being rendered in the compilation of two books. The book by M. Minsky on Turing machines is nearing completion. A book by Thomas Kailath on statistical detection theory in communication is underway.

Reference:

(1) Mooers, C. N. *Wanted–A Reactive Typewriter.* (In press)

2.98 ROME AIR DEVELOPMENT CENTER

Information Processing Laboratory, Griffiss Air Force Base, N. Y.

ROBERT E. IRISH, *Principal Investigator*

A Controlled Information Test Facility, which will be managed by personnel of the Information Storage and Retrieval Section, is being established to conduct controlled tests in the area of information storage and retrieval. Specific areas to be investigated include: (a) the preparation of a bibliography for the field of information processing, (b) comparison of manual versus automatic methods of storing and retrieving textual and graphic information, (c) evaluation of automatically produced abstracts (machine-produced and human-produced abstracts will be compared by using them in problem solving tasks), and (d) the preparation and evaluation of a Semantic Screening Model for the field of information storage and retrieval.

Under a U. S. Air Force contract, a portion of this work is being conducted by Utica College of Syracuse University, Utica, N. Y.

RUTGERS, THE STATE UNIVERSITY 2.99

Graduate School of Library Service, New Brunswick, N. J.
NEAL HARLOW, *Dean*

I. A Study of the Compatibility of Technical Indexing Schemes (Ann F. Painter)

Under study are the indexing systems of three major Government agencies which contribute to the Office of Technical Services—National Aeronautics and Space Administration, Armed Services Technical Information Agency, and the U. S. Atomic Energy Commission (see also 2.27). The project is divided into three phases: (a) a check of the duplication of report materials coming to OTS, (b) a comparison of duplicate indexing efforts by the three agencies and OTS, and (c) an investigation of the consistency with which subject headings are assigned in conventional (manual) versus nonconventional (machine) systems. The project will help to determine the feasibility of automating the Office of Technical Services' processes and dissemination of scientific and technological information.

The study is supported by an Office of Technical Services' contract, administered through the National Bureau of Standards, Data Processing Systems Division (see 2.82).

II. A Study of the Comparative Efficiency of Information Retrieval Modes (Richard M. Dougherty)

The services of the former Chemical-Biological Coordination Center of the National Research Council and similar bibliographical services, such as provided by the *Index-Handbook of Cardiovascular Agents,* are being investigated in depth to determine the intellectual and mechanical attributes of this approach to organization and retrieval of knowledge and to develop recommendations for optimum achievement of the intellectual objectives.

The project is supported by the U. S. Air Force Office of Scientific Research.

III. A Study of Automatic Book Indexing (Susan Artandi)

The objective of the study is to develop a system for machine indexing and to determine the efficiency of this procedure as compared with manual preparation of indexes.

Two approaches are being used: (a) development of a subject list, including synonyms, for a narrow subject area, and design of a computer program to apply it; and (b) development of a method to select indexable terms other than those which could be anticipated by a subject list primarily to index proper nouns.

The project is supported by the U. S. Air Force Office of Scientific Research.

IV. A Study of the Comparative Costs of Preparing and Using Book and Card Catalogs (Fred J. Heinritz)

The relative costs of production, maintenance, and use of book and card catalogs will be determined. An attempt will be made to ascertain parameters to aid libraries in deciding which of the two catalogs is more economical for their particular service requirements.

The project is supported by the U. S. Air Force Office of Scientific Research.

2.100 SCIENCE INDEX GROUP

906 Dashiell Road, Falls Church, Va.
ROBERT L. BIRCH, *Coordinator*

Title-wording studies are underway to pinpoint characteristic patterns in the wording of journal titles and names of organizations where such wording adversely affects retrieval from files or catalogs. Title changes listed in such sources as *New Serial Titles* are analyzed to determine common characteristics of titles deemed unsatisfactory. Interviews are conducted to determine what factors are taken into consideration in the choice of new titles or names of organizations. Various draft checklists have been prepared, indicating the factors which often frustrate publication efforts by limiting the retrievability of publications, insofar as such retrieval depends on recovery from files, library catalogs, and bibliographies. The checklists are designed for those who choose or approve titles for new journals or organizations. When satisfactory formulation of these checklists is completed, it is intended that they be made available for use in editorial offices and management situations to permit more efficient choices at the publication source, in order to insure emphasis on distinctive filing handles in the launching of new publications.

Classification and terminology studies underway are designed to serve as an idea-bank consisting of one-page summaries of ideas which are not applicable at time of filing, but which may be expected to become applicable with the development of new technologies and equipment. Particular emphasis is being placed on the problems of recovery by subject headings which will remain effective despite the unsettled terminology of breakthrough areas.

SCIENTIFIC DOCUMENTATION CENTRE LTD. 2.101

Halbeath House, Dunfermline, Scotland, United Kingdom
P. S. DAVISON

Further work has been carried out on the possibilities of photoelectric masking procedures for handling spectra. In particular, the originally proposed method (1) of treating all spectra as continuous lines and recognizing their shape, though it should be suitable for ultraviolet data, has difficulty with infrared spectra in coping with the wide variations arising from differences in particle size, concentration, and temperature which can grossly alter the position of the background line and also change the shapes of bands. Because of the very large amount of information contained in infrared (and mass) spectra, these variations can, however, be handled on a particulate basis and retain sufficient information to allow discrimination between spectra of different compounds in large collections. Photoelectric masking techniques for this have been outlined in a publication by the Centre (2); these methods would overcome the above difficulties.

With the help of members of the Infra-Red Discussion Group, a cooperative experiment is in progress to assess the magnitude of variations in infrared spectra of the same compounds from a large number of sources; it is hoped this will give some indication of the proportion of these variations arising from instrumental causes and sample preparation. These data will be needed to determine the levels of similarity required to select identical spectra.

Preliminary study has also been given to the availability of spectra (3), though it is now thought that the number estimated to be readily available, some 200,000, is too conservative, and this strengthens the need for a central Spectra Index. After discussion with the Infra-Red Discussion Group, it was clear that collection of this data in a form accessible by compound name was the first requirement, and this work has been initiated. Some 1,500 spectra or equivalent sets of figures have already been collected and experimental spectra services are being offered. Data have been sent or offered by a large number of laboratories and are also being collected from literature sources. Plans are being made to expand the activities of the Centre to cover data outside spectroscopy.

This program has been supported by Fife County Council Trust Fund, Beckman Instruments Ltd., Smith Kline and French Research Institute (England), British Drug Houses Ltd., and Bellingham and Stanley Ltd.

References:

(1) Davison, P. S. "Improving Access to Spectroscopic Data," *British Bulletin of Spectroscopy*, Supplement to Bulletin No. 45, March 1962, pp. 205-206.

(2) Davison, P. S. "Records of Research and Discovery, Neglected Assets of Science, Part II, Methods of Handling Chemical Data," *Research Applied in Industry*, vol. 14, no. 5, 1961, pp. 207-212.
(3) Davison, P. S. *Spectra Index*, Scientific Documentation Centre Ltd., June 1962.

2.102 SOCIÉTÉ D'ÉCONOMIE ET DE MATHÉMATIQUE APPLIQUÉES (SEMA)

87 Rue La Boétie, Paris 8ème, France
ROBERT LATTES *and* BERNARD ROY

Methodological investigations of information retrieval systems are underway. The most important documentation languages were studied and developed according to a common model (2), and an identical survey was undertaken for data processing systems (3). Based on these analyses, and after a bibliographic investigation, an overall model has been developed (4) which leads, through a systematic approach, to the definition and elaboration of an information retrieval system, and thus to the deduction of the schedule of decisions which must be made. This model (a graph) is supported by a double analysis: (a) fundamental characteristic elements, i. e., points to which attention must be drawn, and (b) interactions, i. e., links arising between certain pairs of characteristic elements. A tentative algebraic approach of pertinency is developed (5). The actual studies are relative to the effective operation of documentation systems.

These methodological results will be applied to the construction of a language called PERIOR, designed to allow the storage and retrieval of very general information which is only constrained to be "organized" (scheduled). This language should not duplicate an existing one. Of a technical basis more elaborate than simple keywords, it will nevertheless remain quite simple and will be set up and checked against a corpus of documents in applied mathematics, or possibly in nuclear physics.

The studies are sponsored by Euratom.

References:

(1) "Les languages documentaires. Modèle descriptif et problèmes fondamentaux," presented at the Symposium on Symbolic Languages in Data Processing, Rome, March 26-31, 1962.
(2) SEMA. *Études de Documentation Automatique, Rapport No. 1: Application d'un Modèle de Langage Formel à des Langages Documentaires,* September 1961.
(3) SEMA. *Études de Documentation Automatique, Rapport No. 2: Application d'un Modèle Descriptif des Calculateurs Électroniques a des Équipements Existants,* December 1961.

(4) SEMA. *Études de Documentation Automatique, Rapport No. 3: Conception d'un Systeme Documentaire,* March 1962.

(5) SEMA. *Études de Documentation Automatique, Rapport No. 4: Approche Algébrique de la Pertinence,* June 1962.

SOUTHWESTERN LEGAL FOUNDATION 2.103

Hillcrest at Daniels, Dallas 5, Tex.
ROBERT A. WILSON, *Principal Investigator*

The purpose of the project is to develop and test a system for identifying and retrieving documents or portions of documents pertinent to a specific research need embodied in a search request composed of keywords.

The system is being tested on a test library of 180 appellate court decisions in the field of labor and commercial arbitration law. At present, 60 such published court decisions have been punched into 80-column punchcards and stored on magnetic tape. The IBM 1401 system is used, with 12K memory and 4 tape units. Special programs have been written to provide a machine-sorted wordlist and a condensed search file utilizing root numbers to characterize all the various forms of search words and to permit human participation at the proper stages of index preparation.

The system is designed for a man-machine partnership in the handling of full natural-text input and the development of a thesaurus of search terms by the process of eliminating nonessential and nonsignificant words from the search vocabulary. During the past 6 to 12 months, approximately 100,000 words have been stored, comprising the text of 60 court decisions. When the machine eliminated duplications and nonsignificant words, only 3,200 root words or index terms remained.

The next phase of the project is the completion of the storage of all arbitration decisions from Arkansas, Louisiana, New Mexico, Oklahoma, and Texas in order to provide a complete body of law for experimental search purposes. When this has been done, further experimentation in the preparation of search requests, comparison studies utilizing conventional printed index legal research procedures, and comparative cost studies need to be undertaken. Also planned is the storage of a collection of nonlegal technical documents and the application of the dual indexing system to nonlegal materials. Further experimentation with methods for retrieving portions of lengthy documents applicable to the question at hand is also needed.

References:

(1) Wilson, Robert A. "Progress in Case Law Retrieval," in *Pro-*

ceedings, National Conference on Law and Electronics, Arrowhead Lake, Calif., May 1962.

(2) Wilson, Robert A. "Computer Retrieval of Case Law," Southwestern Law Journal, vol. 16, no. 3, September 1962, pp. 409-438.

2.104 SPENCER CHEMICAL COMPANY

Technical Information Department, Merriam, Kans.
R. H. WHEATER, *Supervisor*

An information system for both manual and machine retrieval has been coupled to an existing technical abstract dissemination program. The technical abstract bulletin is prepared on a specially designed Flexowriter which permits fully machinable treatment of scientific text, including chemical formulas. Selective programming of the Flexowriter produces a byproduct tape from the composition of the abstract bulletin. This tape is fed to a computer (current use is being made of the IBM 1620 for logic processing) to form a printed, permuted, and truncated keyword index for manual access together with a coordinate card file for machine retrieval (search coordination is likewise performed on the 1620). The system also produces a lexicon and an author/affiliation card file which are used for bibliographic control and additional search functions.

The volume of documents totals about 10,000 to 12,000 patents and technical publications per year. The system may later be extended to include internal reports, as well as a computer-produced "alerting service."

For the present, the indexing instructions are given to the Flexowriter operator by an underlining of the keywords which are either contained in or suggested by the abstract. A modified "rolelink" relationship is employed in the keyword indexing, primarily to lend clarity to the permuted index. A study is also in progress on the feasibility of a computer selection of keywords from the abstracted text for indexing purposes.

This system has been designed to handle the company's immediate needs with a truncated term format, using the systems and data processing resources which are now available within the company. However, the input, in "whole data" form, which has served to provide the truncated format is being temporarily stored in order to allow future conversion to an expanded format and processing with magnetic tape equipment.

STANFORD RESEARCH INSTITUTE 2.105

Menlo Park, Calif.

E. LEROY YOUNKER, *Project Leader*

An experimental model of a data retrieval file in which all data are interrogated simultaneously is being designed and constructed. The purpose of this work is to demonstrate the feasibility of conclusions reached in an earlier study (1).

The experimental model will contain 1,000 documents that are described by an accession number and eight English words (descriptors), selected from a 3,000-word dictionary. A search question, consisting of two or more descriptors, will be entered by means of an electric typewriter. The machine will indicate immediately whether or not any file document includes the descriptors of the search question, and if so, how many documents respond. The machine will then resolve multiple responses and type out the accession number and full set of descriptors of each responding document.

The project is sponsored by the U. S. Air Force (Rome Air Development Center).

Reference:

(1) Goldberg, J., et al. *Multiple Instantaneous Response File,* Final Report, SRI Project 3101, Contract AF 30(602)-2142, Stanford Research Institute, Menlo Park, Calif., August 1961. RADC-TR-61-233 (ASTRIA No. AD-266 169).

SYSTEM DEVELOPMENT CORPORATION 2.106

Center for Research in Systems Development,
2500 Colorado Avenue, Santa Monica, Calif.

HAROLD BORKO, *Project Leader*

The activities of the Information Retrieval and Linguistics Project may be divided into two areas of work: documentation research, and linguistics and communication analysis. Documentation studies are concentrating on the derivation of automatic and semiautomatic procedures for indexing, classifying, and abstracting documents. The studies in linguistics and communication have as their objectives the explication of linguistic information, both structural and semantic, to assist in machine processing of natural-language text, and the identification of psychological factors that facilitate man-machine communication using a natural-language vocabulary.

I. Documentation Research

Work has continued on a mathematically derived classification system. The most frequent content words are used to form a document-word correlation matrix, and this matrix is factor analyzed. In addi-

tion to the results obtained using *Psychological Abstracts,* the technique has been applied to 405 abstracts of current computer literature published in *IRE Transactions of Electronic Computers.* Since these abstracts had been classified using a logically derived classification system both by the IRE and by Maron of RAND, comparisons could be made between these classification systems and the automatically derived classification system. The initial results indicate that mathematically derived classification systems can be applied to abstracts of documents in the computer field (1) (2). The programs for this technique are written for the IBM 7090 computer. Details of a descriptor word index have also been published (9).

In associative indexing, the most strongly co-occurring word pairs can be represented in the form of a map, which can be used either for analysis or for retrieval of documents from which the keywords were derived. The association map provides a procedure whereby index terms can be grouped by concept and not simply by alphabetic arrangement. The making of association maps has been improved in several ways which afford greater flexibility to the user of such maps. Ways of making use of association maps, both on-line and off-line mode, have been described (4). These programs are written for the Philco 2000 computer.

Work is also continuing on determining criteria for an acceptable abstract and on deriving procedures for automatic abstracting (3) (10) (11).

II. Linguistics and Communication Analysis

In content analysis via specification of word senses, analysis of the specialized sense occurrences in a 5,000-word corpus is being made. Attempts will be made to develop a set of rules for recognizing the specialized sense occurrences that are indicated syntactically in the corpus, and to generate just enough semantic and factual information in machine-usable form to permit automatic recognition of specialized sense occurrences (not indicated syntactically) to obtain approximate specifications of all specialized sense occurrences therein (8).

In the evaluation of report summaries, an experiment was conducted to determine the effects of format restriction on the selection behavior of humans in preparing report summaries, as well as on the information characteristics and usability of the summaries. Restricted format summaries, designated as term diagrams, consisted of a spatial arrangement of grammatical subjects and objects together with appropriate modifiers interconnected by verb word groups. The results included the following: First, mean summary preparation time was significantly greater for the term diagram procedure than for natural language; second, although the term diagram format was a more compact medium for conveying information, subjects conveyed no more information using term

diagram format than using natural language; third, decision makers using term diagram format summaries gave significantly more correct response to information requests; and fourth, when interviewed following the experiment, subjects were almost unanimous in preferring the term diagram format. Further work on the effects of formatting on man-machine communication is planned (5) (6).

In the study of semantic structure and communication reliability, the question posted is: How do the semantic structural relations of a natural-language expression affect the reliability with which it communicates an intended meaning? Sixteen expressions, familiar yet highly varied in their communication reliability, were selected for experimental stimuli. Using pairs of communicators (one subject who encoded the original stimulus item into a different word and the other who attempted to decode the message into its original form), a measure of average communication reliability was obtained. For each stimulus item the following statistical indices were obtained: associative fluency, average number of substitutions given for the expression by a subject in 2½ minutes, association diversity, and total number of different types of substitutions given for the expression by 100 subjects. The results indicated that the best predictor of communication reliability was the ratio of associative fluency to associative diversity. The product moment correlation between this ratio and communication reliability was 93 percent. Further empirical studies will be conducted to test the generality of this finding (7).

References:

(1) Borko, H. "The Construction of an Empirically Based, Mathematically Derived Classification System," in *Proceedings of the 1962 Spring Joint Computer Conference,* vol. 21, pp. 279-289.

(2) Borko, H., and M. D. Bernick. *Automatic Document Classification,* TM-771, System Development Corporation, November 15, 1962.

(3) Borko, Harold, and Seymour Chatman. *Criteria For Acceptable Abstracts: A Survey of Abstractors' Instructions,* TM-759, System Development Corporation, November 1, 1962.

(4) Doyle, L. B. "Indexing and Abstracting by Association, Part I," to be published in *American Documentation.* SP-718/001/00, System Development Corporation, April 9, 1962.

(5) Ford, J. D., Jr. *Evaluation of Report Summaries: Quantitative Results and Discussion,* TM-662/001/00, System Development Corporation, May 23, 1962. 15 p.

(6) Ford, J. D., jr. *The Effects of Formatting Restrictions on the Quality of Report Summaries Produced by Humans,* SP-707, System Development Corporation, July 11, 1962. 10 p.

(7) Katter, Robert V. *A Predictor of Semantic Communication Effect,* TM-663/000/00, System Development Corporation, August 1, 1962. 22 p.
(8) Olney, John. *Building a Concept Network for Retrieving Information from Large Libraries: Part I,* TM-634/001/00, System Development Corporation, January 26, 1962. 41 p.
(9) Stone, Eileen. *The Descriptor Word Index Program,* FN-6599, System Development Corporation, June 1, 1962. 13 p.
(10) Wyllys, R. E. *Automatic Analysis of the Contents of Documents, Part I: Historical Review,* FN-6089, System Development Corporation, December 7, 1961. 16 p.
(11) Wyllys, R. E. *Automatic Analysis of the Contents of Documents, Part II: Document Searches and Condensed Representations,* FN-6170, January 10, 1962. 26 p.

2.107 SYSTEM DEVELOPMENT CORPORATION

Biomedical Systems Department,
2500 Colorado Avenue, Santa Monica, Calif.
R. W. HARRINGTON

A three-part research and development program is being conducted jointly with the Veterans' Administration.

Part I, to plan the introduction of automated information processing in the VA's Department of Medicine and Surgery, is concerned with system requirements and with methods and techniques for standardizing medical data for use in medical care, research, education, and administration.

Part II is directed toward the development of a prototype support center which can provide technical assistance to medical researchers.

Part III, research on patient data automation, is concerned with investigating and developing means for relieving the clinician and other professional personnel of many of the routine tasks associated with handling patient data. Attention is directed toward data involving the medical record. Two major problem areas have been identified: (a) the storing of medical information so that it can be retrieved, on a real-time basis, for the manifold purposes of patient care; and (b) the storing of clinical records of patients after they are discharged so that the data can be retrieved for use by medical researchers; for the purpose of accrediting hospitals; for legal, medical, and administrative purposes; and for future patient use.

A system description and a functional analysis of a large Veterans' Administration general medical and surgical hospital have been completed. Both the system description and the system analysis are being used as a basis for iterative design and development of experiments on

hospital automation. Particular emphasis is being placed on the storing and retrieving of the medical record.

Progress has been made in all three phases of the work. Descriptive models are being prepared for functional and organizational processes within the operational system of the present Department of Medicine and Surgery; the prototype center is actively providing support to medical researchers; and a computer model that simulates a real-time hospital information system is developed. This hypothetical hospital patient data system is being used experimentally and includes representative portions of admissions and dispositions, medication records and inventory, laboratory and pharmacy reports and records, vital signs, doctors' orders, discharge summaries, and other special management data. Computer programs, based upon analysis and synthesis work, have been written for these functions. Current efforts are centered on the introduction of on-line experiments within the live data environment of the hospital.

References:

(1) Harrington, Robert W. "Patient Data—A Computer-Based System," *Hospital Progress,* May 1962.

(2) Vallance, John A. *Development of a Simulated Ward Operation,* SP-838, System Development Corporation, May 24, 1962.

(3) Wilson, Harold H. *Automated Data Processing for a Modern Hospital,* SP-812, System Development Corporation, May 4, 1962.

TECHNISCHE HOCHSCHULE MÜNCHEN 2.108

Institut für Elektrische Nachrichtentechnik und Messtechnik, Munich 2, Federal Republic of Germany

HORST G. KÖRNER, *Principal Investigator*

Work on retrieval system theory has been carried out, some of which has been published (2). A study was made of about 100 existing and proposed information retrieval systems, involving both specialized equipment and general-purpose computers. In addition to studies of the literature, visits were made to organizations in the United States, France, and Great Britain.

A classification scheme was worked out which accommodates all the studied systems. The idea underlying the classification is that the stored bit pattern in all systems can be considered a transformed matrix of the bit pattern in the "original descriptor-document matrix." This original matrix usually has a low †density of 1-bits and, except when using direct coding (e.g., peek-a-boo cards), a code is employed to compress the matrix, thereby bringing the density of the 1-bits closer

† *See* Glossary.

to the optimum of .5. Four types of codes are used: direct, †exclusive (local), superimposable, and †free (nonlocal). Because of certain common properties, the first three are called †"position codes"; the last three, †"combination codes." If the descriptors are coded (e.g., Zatocoding, Alpha-Matrex cards, WRU tapes), the matrix compression is vertical; if the documents are coded (e.g., Uniterm cards, IBM 9900), the compression is horizontal. For storage in the file the matrix is either partitioned by columns (normal file system), partitioned by rows (inverted file system), or not partitioned. Finally, file interrogation can be serial (e.g., tapes, discs) or parallel (e.g., slotted cards, peek-a-boo cards, associative memories, Lernmatrix).

The classification scheme provides for all combinations of the above-mentioned facets, i.e., 4 types of codes, 2 types of compression, 3 types of matrix partitioning, and 2 types of file interrogation. Provision for systems using multiple filing (e.g., Minicard system, rotational indexes) is also made. The classification, especially in a two-dimensional form (3), is quite useful when giving a survey of information retrieval systems, when discussing the pros and cons of various types of systems, etc. As far as could be determined, some of the possible systems have not yet been described in the literature. Their usefulness is currently being investigated.

Further work is concerned with the redundancy of various documentation codes, including generic and prime-number codes. Studies on the problems of automatic dissemination of information are also being conducted.

The work is being supported by the Deutsche Forschungsgemeinschaft.

References:

(1) Körner, H. G. "Die Matrixdarstellung und die Efficiency von Codes und Geräten beim Wiederauffinden von Information" ["Matrix Representation and the Efficiency of Codes and Devices in Information Retrieval"], presented at the annual meeting of the Deutsche Gesellschaft für Dokumentation, Würzburg, October 1960.

(2) Körner, H. G. "Geräte und Verfahren für das Wiederauffinden von Information" ["Devices and Methods for the Retrieval of Information"], in *Taschenbuch der Nachrichtenverarbeitung*, pp. 1282-1313. Berlin: Springer Verlag, 1962.

(3) Körner, H. G. "Classification Chart of Information Retrieval Systems," Karlsruhe, 1961. (Available from author)

(4) Körner, H. G. "Type of Codes, Methods, and Devices in Information Retrieval," presented at the conference of the In-

† *See* Glossary.

ternational Federation for Documentation, London, September 1961.
(5) Körner, H. G. "List of Books, Bibliographies, and Journals in the Field of Information Retrieval," Munich, 1962. (mimeographed)

THOMPSON RAMO WOOLDRIDGE INC. 2.109

RW Division, Canoga Park, Calif.
H. P. EDMUNDSON, PAUL L. GARVIN, J. L. KUHNS, C. A. MONTGOMERY, *and* DON R. SWANSON

I. Automatic Abstracting

Research on automatic abstracting of English-language text is in progress (14). The approach being used is that of studying the attributes of carefully controlled human-produced abstracts in order to deduce a mechanizable technique for creating similar products. A cycle of experiments is in progress in which machine-produced abstracts of a small collection of documents are analyzed and the results of this analysis used to improve the initial mechanized technique. A theoretical study of the content pattern of documents has been conducted. This project is sponsored by the U. S. Air Force (Rome Air Development Center).

II. Automatic Indexing and Retrieval

Research is being conducted on the problems of fully automatic indexing and retrieval based on machine search of natural-language text. The first phase of the project, intended to measure the effectiveness with which material responsive to specific questions can be selected through machine searching of the full natural-language text, has been completed (1). In the second phase of this work, further experiments with the searching of full text and the evaluation of retrieval effectiveness have continued; however, in this phase a program has been written to accept questions in natural-language form. The computer takes this input question on a word-by-word and phrase-by-phrase basis and performs a thesaurus lookup. The thesaurus carries preassigned weights for each word to indicate an estimate of its importance for retrieval purposes. Words not appearing in the thesaurus are taken as having zero weight; i.e., they are unimportant so far as retrieval is concerned. The words which have non-zero importance weights are then used, together with their synonyms and near-synonyms from the thesaurus, to construct a search request. This automatically constructed search request is then used to search the text of the experimental library. Those documents containing the largest number of words and phrases in the search request are assigned the greatest "relevance weight." The proximity of words and phrases is also taken into ac-

count in constructing the relevance weight. The output consists of a list of document numbers ordered according to the computer-calculated relevance weight. The same list shows the "true relevance weight" estimated beforehand by human examination. By inspection, one can then determine how near the top of the list the documents relevant to any particular question fall. One can derive from this output a curve showing the percent of relevant information retrieved as a function of the number of irrelevant documents also retrieved. This curve taken as an average over all questions is given in (5). The conclusion is reached that machine processing of natural-language questions led to more effective retrieval than did a similar process in which a human being translated the original question into a search instruction. This result came about simply because it is more important for retrieval purposes to be thorough (with the formulation of synonyms and near-synonyms) than it is to be clever. This research is sponsored by the Council on Library Resources, Inc.

In a separate investigation (7), a study of several thousand entries in a classified bibliography of article titles (the *Index Medicus*) revealed that a large proportion of the title entries contained words identical to or synonymous with words of the corresponding subject heading. It is inferred that a major part of the bibliography studied could have been compiled by a machine procedure operating on titles alone, provided the machine were supplied with a suitable synonym dictionary.

A third project in the field of automatic indexing has also been completed. A computer program has been written which will automatically index teletype news dispatches. Keypunching of the input is avoided since the machine procedure begins with the punched paper tape produced by teletype. Redundancy in the procedure permits one to circumvent the problem of garbles. An evaluation of the indexing of 1,200 items indicated that the machine missed less than 7 percent of the relevant information in the indexing process and assigned a number of irrelevant subject tags equal to less than 17 percent of the total. Comparison with human indexing indicated a wide margin of superiority for the machine procedure. Again, this turned out to be true only because diligence counted far more than intellect. The cost of machine indexing of these news dispatches (about 20 cents per item) is less than that of human indexing.

Thus, three projects have yielded strong evidence that automatic indexing based on thesaurus techniques, as compared to human processes, leads to acceptably high quality and is economically feasible.

References:

(1) *Word Correlation and Automatic Indexing,* Phase I Final Report—"An Experiment in Automatic Text Searching," April 30, 1960.

(2) *Word Correlation and Automatic Indexing,* Progress Report No. 3, January 1961.
(3) Swanson, Don R. "Searching Natural Language Text by Computer," *Science,* vol. 132, no. 3434, October 21, 1960, pp. 1099-1104.
(4) *Word Correlation and Automatic Indexing,* Phase IIA Final Report, January 1962.
(5) Swanson, Don R. "Interrogating a Computer in Natural Langauge," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.
(6) Swanson, Don R. "Research Procedures for Automatic Indexing," in *Machine Indexing: Progress and Problems.* Washington, D.C.: American University, 1962, pp. 281-304.
(7) Swanson, Don R., and C. A. Montgomery. "Machine-Like Indexing by People," *American Documentation,* vol. 13, no. 4, 1962, pp. 359-366.
(8) Garvin, Paul L. "Linguistic Aspects of Information Retrieval," in *Machine Indexing: Progress and Problems.* Washington, D.C.: American University, 1962, pp. 134-143.
(9) Edmundson, H. P. "A Statistician's View of Linguistic Models and Language Data Processing," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill book Co., Inc. (In press)
(10) Garvin, Paul L. "A Linguist's View of Language Data Processing," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press)
(11) Ray, L. C. "Keypunching Instructions for Total Text Input," in *Machine Indexing: Progress and Problems,* pp. 50-57. Washington, D.C.: American University, 1962.
(12) Ray, L. C. "Programming for Natural Language," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press)
(13) Swanson, Don R. "The Formulation of the Retrieval Problem," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press)
(14) *A Study for Automatic Abstracting,* Final Report, prepared under Contract AF 30(602)–2223 for Rome Air Development Center, September 1961.

2.110 THOMPSON RAMO WOOLDRIDGE INC.

RW Division, Canoga Park, Calif.
PAUL L. GARVIN, *Principal Investigator*

The purpose of a basic research project underway is to develop formal inductive methods for semantic analysis in order to achieve reliability in the compilation of semantic codes for use in information retrieval, automatic abstracting, and related activities. This is to be achieved by the adaptation of linguistic methodology to the study of semantic systems.

The first phase of the project is concerned with the further development of linguistic methodology, as based on (1) and illustrated by (2), and with the compilation of a semantic data file based on a collection of documents under analysis by a linguistic technique. In the second phase of the project, the methods developed during the first phase will be applied to the semantic data file. It is hoped that the research will yield basic insights into the nature of the semantic structure of natural languages and will allow the formulation of criteria for an inductively derived thesaurus. The latter is intended for use in future information retrieval and automatic abstracting systems.

The research is sponsored by the U. S. Air Force Office of Scientific Research.

References:

(1) Garvin, Paul L. "The Definitional Model of Language," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Co., Inc. (In press)

(2) Garvin, Paul L. "A Study of Inductive Method in Syntax," *Word,* Journal of the Linguistic Circle of New York, vol. 18, Nos. 1-2, 1962, pp. 107-120. (Technical Note No. 1 under Contract AF49 (638)-1128, AFOSR–1976)

(3) Garvin, Paul L. "Computer Participation in Linguistic Research," to be published in *Language,* Journal of the Linguistic Society of America. (Technical Note No. 2 under Contract AF 49 (638)-1128, AFOSR–2480)

(4) Garvin, Paul L. "The Impact of Language Data Processing on Linguistic Analysis," to be published in *Proceedings of the Ninth International Congress of Linguists.* (Technical Note No. 3 under Contract AF 49 (638)-1128, AFOSR–2593)

(5) Garvin, Paul L., and W. Karush. "Linguistics, Data Processing and Mathematics," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press) (Technical Note No. 4 under Contract AF 49 (638)-1128, AFOSR–2866; System Development Corporation Document SP-771).

U.S. ARMY BIOLOGICAL LABORATORIES 2.111

Fort Detrick, Frederick, Md.
HAROLD W. BATCHELOR

Trial operation of an information retrieval system adequate for the deep indexing of a collection of almost 100,000 active unpublished documents on a wide variety of subjects continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Batchelor, Harold W., and Clifford J. Maloney. "A Classified Deep Index," presented at the American Chemical Society meeting, Division of Chemical Literature, Washington, D.C., March 24, 1962.

(2) Maloney, Clifford J., and Harold W. Batchelor. "Tentative Principles of Information Retrieval," presented at the American Chemical Society meeting, Division of Chemical Literature, Washington, D.C., March 24, 1962.

U. S. DEPARTMENT OF AGRICULTURE 2.112

New Crops Research Branch, Crops Research Division, Agricultural Research Service, Plant Industry Station, Beltsville, Md.
ROBERT E. PERDUE, JR., *Botanist*

Study of problems involved in the establishment of a system to retrieve information relevant to the utilization of plants continues. The peek-a-boo retrieval system appears to best satisfy the requirements of organizational objectives. This system will allow maximum flexibility in the development of a complex subject classification and will permit the complete information program to remain directly under the immediate control of the botanists who will contribute to and withdraw from the information store.

In the initial stages, it appears that it will be desirable to index information indirectly to plant species through genera. Answers to questions will be provided in the form of generic numbers, and determination of the species involved will require lookup in the master file. This approach will keep the number of card decks at a minimum and will permit retrieval of important types of information that cannot be retrieved with an index to species. Indexing to genera will further avoid some of the complications resulting from instability of species names and synonomy.

The most serious problem that can now be foreseen is the potentially large number of answers that may be returned for other than the most complex questions. The significance of this problem can be determined only by experiment.

It is hoped that a system can be initiated on an experimental basis in early 1963.

Reference:

(1) Perdue, Robert E., Jr. "Information Retrieval with a Coordinate Index," *American Journal of Botany,* vol. 49, no. 6, part 2, July 1962, p. 677. (Abstract of paper presented before meeting of American Society of Plant Taxonomists, Oregon State University, Corvallis, Oreg., August 26-31, 1962)

2.113 U. S. NAVAL POSTGRADUATE SCHOOL

The Library, Monterey, Calif.
GEORGE R. LUCKETT, *Librarian,*
and PAUL SPINKS, *Associate Librarian*

The SABIRS (Semi-Automatic Bibliographic Information Retrieval System) program has been developed to effect a rapid and accurate system of machine information storage and retrieval.

Technical documents relating to all major branches of the applied physical sciences are included in SABIRS. Included also is a heavy concentration of materials in the fields of nuclear engineering, naval science, naval engineering, aeronautical engineering (including missile technology), intelligence, and economics. Data for 6,000 documents have been coded, processed, and made available for literature searches on a magnetic tape file. The mechanized glossary now includes 4,748 descriptors and uniterms.

The Control Data Corporation 1604 computer is used for the search and retrieval process, and the IBM 1401 computer is utilized for printing search and file updating data.

One year of experimentation with SABIRS has indicated its superiority over a manual (coordinate index) system of searching. At present a small number of students avail themselves of the mechanized service now being offered and submit requests for current literature searches. Experience over the past year has confirmed that the number of requests submitted will grow commensurately as more coded data are available for searching. Complete utilization of the system is anticipated when a minimum of 20,000 documents are retrievable through SABIRS.

At present a search output reveals only the accession numbers of items cited. However, the development of the SABIRS program to make available a readable output is anticipated in the near future.

Reference:

(1) Wildberger, August Martin. *Information Retrieval.* Master's Thesis. Monterey, Calif.: U. S. Naval Postgraduate School, 1961.

Office of Research and Development, Washington 25, D. C.
R. A. SPENCER, *assisted by seven principal investigators*

Objectives of the several research projects which collectively constitute a many-faceted approach to problems of organizing, storing, and searching technical information include: (a) design of systems of broad application; (b) language studies; (c) investigation of data analysis and file preparation problems, and creation of methods for their solution; (d) utilization of available knowledge, techniques, and equipment for construction of limited search systems; (e) consideration of the problems of classification of technical subject matter and their relationship to mechanized searching; and (f) testing and evaluation of manual and mechanized search procedures.

I. Information System for Alloy Patents

A project has been started in the area of alloys. The specific and generic components and percentage ranges are being extracted from the text of approximately 8,500 alloy patents. This information will be arranged in a hierarchical random-access file for use on an IBM 1401 computer with a disc file.

II. Semiautomatic Encoding of Chemistry for Information Retrieval

The SECIR (Semiautomatic Encoding of Chemistry for Information Retrieval) project, previously identified as PACIR, is nearing completion in its initial application to approximately 3,500 pesticide patents. The original IBM 305 RAMAC semiautomatic encoding program has been translated into a program for use on the IBM 1401 computer.

III. Coding of Patents in the Organometallic Art

Coding is complete on the 3,500 patents in the organometallic art, using a general chemical compound codesheet format. Processing of the data is underway and tests on the file will be carried out in the near future. The patents were coded in three separate sections (claims, examples, and disclosure) in an attempt to find a procedure for diminishing the deleterious effects of composite coding.

IV. Catalog of Compounds

The testing and evaluation of a catalog of compounds and their roles from about 800 polyethylene and polypropylene patents are completed, and the system is now in use in an examining operation. The general chemical compound codesheet has been applied to the catalog of compounds extracted from the polyethylene patents. Each different term (over 10,000) was coded on a separate codesheet. An automatic machine procedure assigns appropriate codes to duplicate terms in the

catalog, and a generic search file is created using the descriptors from the general codesheet. The file will be tested on random-access hardware in the near future. An experiment has been carried out that compares the quality and quantity of manual versus machine searching on the 800 patents in the polyethylene file. The data have been gathered and are being evaluated. A further experiment on this file is being carried out to determine the value of links.

V. Coding of Chemical Compound Structures

The recently described system (1) for the enumeration and ciphering of organic compounds has undergone further development. The system has now been extended to uniquely represent (a) structural formulas which have alternatively variable substructures (Markush formulas), and (b) stoichiometric chemical reactions.

A manually prepared card file of structural formulas organized on the basis of cipher properties is being maintained and augmented. The file is ordered by a 14-decimal digit index number which represents a count of structural features constituting a geometric and functional profile.

In an effort to establish the linear notation as a basis for an information retrieval system which would be independent of, but complementary to, the HAYSTAQ system, algorithms are being developed to perform the following automatic operations based on mechanical cipher analysis: (a) computation of the numerical index described above; (b) determination of the enumeration pattern of the ciphered structure; (c) interconversion among three equivalent cipher forms: a line formula, a hierarchical form, and a skeletal segment form; and (d) isolation of cipher components into independent symbolic representations of certain structural fragments.

VI. HAYSTAQ Project

The HAYSTAQ project, which is a joint effort with the National Bureau of Standards (see 2.82), seeks to create, test, and evaluate experimental search systems utilizing general-purpose computers. Testing of the structure searching program, which centers about a topological network tracing, is underway.

A corollary project, conducted during the summer of 1962, resulted in the development of a computer routine for preparation of detailed encoded descriptions of structures of chemical compounds in the HAYSTAQ format, employing as input a curtailed form of manually prepared data. Another summer project, conducted at the National Bureau of Standards, developed machine methods for more powerful screening of structures based on rapid inspection of a large number of criteria represented in the form of a binary matrix.

VII. Coordinate Index File for Transistor Circuits

A coordinate index file for miscellaneous transistor circuits, taken from 2,600 U. S. patents, is being established; about two-thirds of these have already been incorporated into the file and are in use for searching patent applications. The file preparation procedures emphasize the depth of analysis (an average of over 100 descriptors per document) required for patent searching and the importance of recording the logical relationship between terms. An extensive list of topological terms is employed. Scope notes are included in the term list and on the glossary used in searches. IBM cards are being punched for internal and logical checks using an IBM 305 RAMAC, and a search could be made from this deck. However, this deck is used on the RAMAC to punch out Batten cards from which the actual searches are made. A Lodestar Microfilm Reader, with microfilm photos of the full text of the patents, is installed in the examining division for display. Plans are being made for an evaluation study of the file; the examiners are documenting the results of searches made with this file and comparing them with prior manual searches.

The full text of 100 patents from the transistor file is being punched on IBM cards and transferred to magnetic tape, to be used in several mechanized information retrieval and automatic indexing studies using an electronic computer.

VIII. Electronic and Electrical Schematic Diagrams

A study is being made of means for effective storage and retrieval of complete electronic and electrical schematic diagrams, with three basic approaches being studied: linear, matrix, and lattice-type notation schemes. These approaches are being compared and selection will be made of the most satisfactory notation system. Search strategies will be developed for retrieval of the stored-circuit information from electronic computers.

IX. Information System for Resuscitators

A project leading to the development of a system for storing and retrieving information about resuscitators has been completed and is now in use. Key terms were applied to approximately 700 documents dealing with this type of mechanical breathing apparatus. IBM cards are being used as Batten cards, but any random-access or inverted file device could be employed.

X. Basic Research in Linguistics, Mathematics, and Self-Organizing Systems

A recently established basic research group will explore problems in the areas of linguistics, mathematics, and self-organizing systems. The primary objective in the area of linguistics involves the resolution of in-

transigence in natural language so that unambiguously stored facts and concepts disclosed in a document which correspond to queries about such facts and concepts may be retrieved. Existing files in the chemical art are being evaluated for experiments designed to test the efficacy of methods for resolving semantic problems.

Principles of inductive inference and self-organizing systems are being studied and explored with a view toward preparing a large file having self-organizing capabilities. The logical analysis and mathematical model for the self-organizing system are now at an intermediate stage in preparation for testing on the IBM 1401 computer.

XI. Information System for Automatic Transmissions

In the project relating to the search of documents dealing with automatic transmissions, a new approach has been adopted which views these devices as complex hydraulic power plants with specific load relationships. As a preliminary phase for system test purposes, simpler hydraulic power plants are being used as the subject matter. It is contemplated that the search will begin with general questions and then proceed through greater specification to the isolation of the available art most nearly approximating that sought. To further assure that the power to recall will be at a maximum, a system of generating the search topics from the documents involved is being employed with a view to the possible later extraction of the search topics from the text by mechanical means. A model of the system has been implemented successfully, using edge-notched cards. It is contemplated that as the file grows a computer or a search machine of special design will be employed.

XII. Mechanical Arts Search Project

In the mechanical arts search project being conducted in cooperation with the Netherlands Patent Office (see 2.88), the punchcard search file for the ILAS card sorter has been completed, and a set of test searches has been made. The significance of the results of the tests will be reported by the Netherlands Patent Office.

XIII. Research Exchange Program

As a result of the International Patent Office Workshop on Information Retrieval, October 1961 (2), the Ad Hoc Committee on International Cooperation in Information Retrieval among Examining Patent Offices (ICIREPAT), comprising representatives of 10 Patent Offices, was established for the purpose of formulating a cooperative research effort. The Secretariat for this committee is in the U. S. Patent Office. A second meeting of the committee took place at the German Patent Office in Munich in September 1962 for the purpose of organizing and implementing the cooperative research program.

As a corollary effort, and with the assistance of a National Science Foundation grant, the U. S. Patent Office is sponsoring a research associates program. Six candidates have been selected from the Patent Offices of England, Germany, Japan, Netherlands, and Sweden, and from the Institut International des Brevets, to conduct research during a 2-year period at the U. S. Patent Office on problems in information retrieval related to patent activity.

References:

(1) Hayward, H. Winston. *A New Sequential Enumeration and Line Formula Notation System For Organic Compounds,* U. S. Patent Office Research and Development Report No. 21, November 1961.

(2) U. S. Department of Commerce, Patent Office. *The 125th Anniversary of the United States Patent Act of 1836, Part II: Proceedings of a Conference Entitled the International Patent Office Workshop on Information Retrieval* (Washington, D. C., October 23-25, 1961). (In press)

UNIVAC DIVISION OF SPERRY RAND CORPORATION 2.115

P. O. Box 500, Blue Bell, Pa.

GILBERT KASKEY, *Principal Investigator*

A recently completed series of studies and analyses pertained to that part of the overall information (document) retrieval function concerned with the storage and maintenance of the master "index" file and the processing of search requests against it.

The project was carried out in three concurrent phases: (a) analysis of file characteristics and their data processing implications, using as test material an extensive sample of the ASTIA automated document file; (b) design of the Information File Generator (IFG), a computer program to construct a document file from a stipulated set of parameters; IFG has successfully generated a sample file based upon the frequency of usage of ASTIA descriptors and the distribution of documents by number of descriptors; and (c) mathematical formulation of information retrieval factors in an attempt to formalize and unify the concepts of file organization and search techniques.

The project was sponsored by the U. S. Air Force Office of Scientific Research.

Reference:

(1) Fossum, Earl G., et al. *Optimization and Standardization of Information Retrieval Language and Systems,* final report under Contract AF 49 (638) -835, prepared for U. S. Air Force Office

of Scientific Research, AFOSR-3216. Blue Bell, Pa.: UNIVAC Division of Sperry Rand Corp., July 1962.

2.116 UNIVERSIDAD INDUSTRIAL DE SANTANDER

Centro de Documentación e Información, c/o División de Investigaciones Científicas, Bucaramanga, Colombia
CLARE DE SILVA

A project is underway for the development of a system for automatic retrieval of bibliographic references relating to the fields of chemical, electrical, industrial, mechanical, and metallurgical engineering.

The file includes scientific and technical magazines arriving at the Centro (around 700 titles at present), plus a growing collection of microfilm, photocopies, pamphlets, bulletins, reprints, etc. Titles of articles are translated into Spanish, and Spanish abstracts and †micro-abstracts are prepared. To date, 3,000 articles have been indexed. A contract is still pending for equipment.

Present work, pending arrival and installation of equipment, is concentrated on preparation of micro-abstracts, selection of descriptors for micro-abstracted articles, and the development of descriptor dictionaries.

After installation of equipment, the system will be set up and analyzed in operation.

References:

(1) de Silva, Clare. "La búsqueda automática de información," *Revista de la Universidad Industrial de Santander,* vol. 4, no. 2, 1962, p. 167.

(2) de Silva, Clare. "El Centro de Documentación e Información de la Universidad Industrial de Santander," *Boletín de Información Extranjera,* vol. 4, no. 260, 1961, p. 924. (In Spanish)

(3) de Silva, Clare. "Documentación e Información Científica—Su Necesidad en la Moderna Investigación," presented at the Curso Inter-Americano de Fotometría de Llama, sponsored by the University and the Organization of American States, August 13, 1962.

2.117 UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Department of Mathematics, Seminar of Professor H. Florin, Louvain, Belgium
MARCEL F. NEUTS, *Principal Investigator*

Work on formulating and studying mathematical models which may be of use in the optimization of information retrieval processes continues, with no change reported since the previous statement [*Ed.*]. The project is supported by Euratom.

† *See* Glossary.

References:

(1) Neuts, Marcel F. *A Limit Theorem for Nested Binomial Variables,* Rapport Technique No. 2, Katholieke Universiteit te Leuven, Seminarium van Professor H. Florin, Leuven, Belgium, May 1961. 15 p.

(2) Neuts, Marcel F. *A Distribution-Problem in Sampling without Replacement,* Rapport Technique No. 4, Katholieke Universiteit te Leuven, Seminarium van Professor H. Florin, Leuven, Belgium, June 1961. 34 p.

(3) Neuts, Marcel F. *An Application of Decision Theory to the Coding of Keywords,* Rapport Technique No. 5, Katholieke Universiteit te Leuven, Seminarium van Professor H. Florin, Leuven, Belgium, August 1961. 9 p.

(4) Neuts, Marcel F. "A Multi-Stage Search Game," presented at the Princeton University Conference on Game Theory, October 1961.

UNIVERSITÉ DE PARIS 2.118

Laboratoire de Psychologie Sociale, Documentation Service, Paris, France

ROBERT PAGÈS, *Director of Research*

Current projects include research on (a) the processes of communication in the †"information interview"; (b) a formalized abstracting schema and other documentary forms useful in analyzing materials in the social sciences; the structure of document analyses, from the point of view of coding for mechanized handling; and spontaneous abstracting behavior among researchers and non-researchers; (c) redundancy and synonymy in communication; (d) the combination of constraints inherent in retrieval and constraints of ordered arrangement in a single linguistic and formal system for the general organization of information in a research laboratory; and (e) relationships between aspects of the scientific production structure (i.e., what kind of work is being carried on by scientists at a particular time and what information can be found related to that topic at that time) and aspects of requests made by scientific researchers to a documentation service, particularly studies of the characteristics of requests; these studies imply the search for adjustment between production and requests. Other current work includes (a) an experiment using a laboratory miniature of main documentation processes and concerning definition criteria in the expression of requests, and (b) a survey of information behavior, including customs and habits of information-taking, and resistance to change. The survey will focus particularly on the spontaneous behavior of clients.

† *See* Glossary.

References:

(1) Pagès, Robert. "L'analyse codée, technique documentaire en psychologie sociale et en sciences humaines: Présentation et résumé de la grammaire," *Chiffres,* no. 2, 1959, pp. 103-122.

(2) Pagès, Robert, Jean Bouillut, and Gérard Lemaine. *Une grille d'analyse des documents en psychologie sociale,* 1960. 22 p. (mimeo)

(3) Bouillut, Jean. "Analyse de contenu et traitement de l'information, Les Parasitages," *Chiffres,* no. 3, 1961, pp. 151-171.

(4) Pagès, Robert, Jean Bouillut, Marie-Thérèse Duflos, Danièle Vataire, and Marie-Claude Gardelle. *Problèmes et caractéristiques d'une langue d'information pour la mise en mémoire et le rappel.* (In preparation)

2.119 UNIVERSITY OF ARIZONA

Systems Engineering Department, Tucson, Ariz.
JAMES W. PERRY, *Principal Investigator*

Investigation underway has the twofold objective of (a) developing well-formulated principles for the design, development, and operation of documentation methods and systems, and (b) defining possibilities for pioneering new methods of correlating and evaluating factual information by applying automation techniques. The general methodology has been as follows: (a) analysis of previously developed document selection methods, both traditional and recently developed; (b) definition of features of similarity and difference as observed during analysis; (c) formulation of parameters to provide measures for characteristic features of various types of document selection methods; and (d) development of a mathematical model that relates structural and operational parameters of various methods to their performance characteristics.

The mathematical model is being worked out so as to permit (a) computer simulation of individual searches to select a subset of documents by various methods when characterized by different values of parameters, and (b) computation, from statistical distributions of parameters for different systems, of other parameters that characterize response to different ranges or spectra of queries. The programs are being worked out for the IBM 1401 computer at the University of Arizona and the IBM 7090 computer at the Western Data Processing Center, UCLA.

During the past 6 months, attention has centered on (a) developing the mathematical model, and (b) evaluating ranges of values for various parameters. Trial runs of the computer simulation program are being planned for the near future.

This investigation is being sponsored by the U. S. Air Force Office of Scientific Research.

Reference:

(1) Perry, J. W. "Communication Systems, Design and Theory. Part I. Documentation Selection—Analysis and Design of Methods," in *Proceedings of the 5th Annual Institute in Technical and Industrial Communications,* Colorado State University, Fort Collins, Colo. (In press)

UNIVERSITY OF CHICAGO 2.120

Graduate Library School, Chicago 37, Ill.
WILLIAM J. KURMEY[1]

An evaluation is being made of automatically prepared abstracts (autoabstracts) and indexes. The study consists of an evaluation of (a) the content and quality of autoabstracts based on word frequency selection as compared with conventional abstracts, and (b) the automatic preparation of keyword-in-context and subject-in-context indexes to the autoabstracts as compared with conventional indexes to the conventional abstracts. The investigation is based on the premise that the standards of any valuable machine abstracting and indexing agency must equal those of present, well-established, and reputable abstracting and indexing agencies.

Journal articles for which abstracts are available from reputable abstracting agencies are being converted into machine-readable form. Present plans limit the subject fields of investigation to chemistry and biology; however, extension of the project to the fields of medicine and mathematics for comparative purposes may be undertaken. The autoabstracts created from the machine-readable texts are generated by five different methods based on word-frequency distribution. These methods involve the computation of relative significance factors for each sentence using variable length chaining and weighting of word clusters. Fifty articles selected at random from each subject field will be used with the five autoabstracts compared to each corresponding abstract in the leading abstract publication of that field. Programming is underway utilizing IBM 1401, IBM 1410, and IBM 7090 systems. Major efforts have been data preparation and programming. Objectivity in evaluation may be possible by employing computer comparison programs for both autoabstracts and human abstracts, thus eliminating subjective judgement. Documents selected from the field of chemistry will be completely in abstracts and human abstracts, thus eliminating subjective judgment. machine-readable form by December 1962. Subject indexes, as well as author and title indexes, will be generated by a Keyword and Subject-

[1] Present Address: Data Systems Division, IBM Corp., Poughkeepsie, N. Y.

Word-Out-of-Context program. These indexes will be compared with index entries assigned to the abstracts in the abstract publications.

Project results will be published as a master's thesis for the University of Chicago Graduate Library School. Copies will be available from the Photoduplication Department of the University.

2.121 UNIVERSITY OF ILLINOIS LIBRARY

Chicago Undergraduate Division, Navy Pier, Chicago, Ill.
DON S. CULBERTSON, LOUIS A. SCHULTHEISS, *and* EDWARD M. HEILIGER

The final report of the University Library Information Systems Project, a feasibility study carried out under grants from the Council on Library Resources, Inc., and the University of Illinois Research Board, has been published (1). The proposed system for library technical operations—acquisitions, catalog preparation, circulation procedures, and serials control—is based on a series of electronic computer-produced documents.

A program for testing and evaluation of the system has been initiated. Block diagramming and final organization preparatory to computer programming are underway. Present plans are to write all programs in COBOL since it is felt that a general, nonmachine-oriented language would increase the utility of the programs.

Implementation of the findings of the study will be completed as rapidly as possible. Publication of results will take place after complete system testing.

The testing and evaluation program is supported by the National Science Foundation.

Reference:

(1) Schultheiss, Louis A., Don S. Culbertson, and Edward M. Heiliger. *Advanced Data Processing in the University Library.* New York: Scarecrow Press, 1962. (Cloth, $10.00)

2.122 UNIVERSITY OF MICHIGAN

Department of Library Science, Ann Arbor, Mich.
R. N. PAUL[1]

The comparative study of the classification theories of Bliss and Ranganathan continues, with no change reported since the previous statement [*Ed.*].

[1] Present address: Booth Library, Eastern Illinois University, Charleston, Ill.

UNIVERSITY OF OKLAHOMA 2.123

Norman, Okla.
MORTIMER D. SCHWARTZ, *Professor of Law,* [1]
and EARL FARLEY, *Project Director* [2]

A "Keyword-in-Context" (KWIC) program is being studied for adaptation and use in the retrieval of space law materials in the Space Law Collection at the University of Oklahoma. The Space Law Collection consists of some 1,200 individually cataloged items, including books, periodicals, manuscripts, Government publications, tape recordings, TV films, photographs, and microfilms, published in English and foreign languages. Each item is represented in a card catalog in traditional card catalog form. The KWIC program will also be utilized for a subject classification analysis of space law. The University of Oklahoma Computer Laboratory, which has an IPM 1410 computer, is assisting in this project.

The benefit of experience derived from this project and a retrieval project at the University of Kansas, utilizing the KWIC program and the IBM 1401 computer for Slavic materials (see 1.54), will be exchanged.

UNIVERSITY OF PENNSYLVANIA 2.124

Institute for Cooperative Research, Philadelphia 4, Pa.
JOHN O'CONNOR

The possibilities of mechanized indexing are being studied.

The punched-card retrieval system at the Merck Sharp & Dohme Research Laboratories (West Point, Pa.) is the subject of investigation, with particular concentration on the three indexing terms *penicillin, toxicity,* and *mode of action.* A random few documents assigned one of these terms, e.g., *penicillin,* by Merck indexers and a random few documents not assigned that term are examined. An indexing rule is sought which a computer could use and which would reproduce these assignments (with perhaps a bit of overassigning). Any such rule is then tested on a few further documents; necessary revisions are made and then tested on still further documents. The intent of the study is to get some empirical evidence for hypotheses about the kinds of computer indexing rules which might and might not work.

About 100 documents were selectively examined "by hand." Concentration was mostly on trying frequent-word indexing rules (1). Computer processing of text is underway in order to permit more thorough study. About 200 documents had been keypunched as of July 1962, and a word-frequency list for each document has been gen-

[1] University of Oklahoma, Norman, Okla.
[2] Slavic Index, University of Kansas Libraries, Lawrence, Kans.

erated by a UNIVAC I computer. These word-frequency lists are intended to permit a fairly intensive exploration of possible indexing rules involving frequent "related" words. Where frequent-word approaches fail, more subtle rules will be looked for.

Some simple ways have been described in which people operating retrieval systems can carry out, at odd moments, small-scale empirical investigations of the possibilities of mechanized indexing (2).

Copies of the machined text are available at cost. Ultimately, a random sample of about 400 documents indexed at Merck in 1955 will have been keypunched. Preliminary information about the machined text is available from the author.

The research is sponsored by the U. S. Office of Naval Research and the U. S. Air Force Office of Scientific Research.

References:

(1) O'Connor, John. *Mechanized Indexing: Some General Remarks and Some Small-Scale Empirical Results.* (Copies available from author.) (ASTIA No. AD-250 209). An abridged account appears in *Machine Indexing: Progress and Problems,* pp. 266-279. Washington, D.C.: American University, 1962.

(2) O'Connor, John. "Some Suggested Mechanized Indexing Investigations Which Require No Machines," *American Documentation,* vol. 12, no. 3, July 1962, pp. 198-203.

2.125 UNIVERSITY OF PENNSYLVANIA

The Moore School of Electrical Engineering, Philadelphia 4, Pa.
NOAH S. PRYWES *and* HARRY J. GRAY

A study is being conducted on the Multi-List system for real-time information storage and retrieval, under the sponsorship of the U. S. Office of Naval Research. The purpose is the development of techniques for mechanization of storage and retrieval from large files such as those in a business or a technical library.

The study was initially directed toward obtaining efficiency in processing by utilizing list techniques. To this end a processor and an associative memory were designed with application studies serving as the basis. Later, a system consisting of a disc file (IBM 1405 computer) to handle the U. S. Navy Supply files was devised and will shortly go into pilot operation. This latter effort is co-sponsored by the U. S. Navy, Bureau of Supplies and Accounts.

Currently, the problem of automatic stratification of a technical library is being studied. Experiments in a stratification program are being done with the use of an IBM 7090 computer.

Another effort consisted of mechanization of a growing tree using

initially the IBM 1620 computer, and then later, for larger trees, the Philco 2000 computer (Model 210) and UNIVAC Larc.

References:

(1) Moore School of Electrical Engineering, University of Pennsylvania. *The Multi-List System Technical Report No. 1,* prepared for the Office of Naval Research, Information Systems Branch (three parts, two volumes), November 30, 1961.

(2) Prywes, N. S., and H. J. Gray. "The Multi-List Type Associative Memory" (AIEE Publications S-136), in *Proceedings of the Session on Gigacycle Computing Systems at the AIEE General Winter Meeting,* January 1962, pp. 87-107.

(3) Landauer, W. I., and N. S. Prywes. "A Growing Tree for Descriptor Language Translation," in *Proceedings of the Symposium on Symbolic Languages in Data Processing,* March 1962.

(4) Reed, W. G. *Computer Simulated Construction of a Balanced Decoding Tree.* Master's Thesis. Philadelphia: University of Pennsylvania, Moore School of Electrical Engineering, July 1962.

(5) Prywes, N. S., and H. J. Gray. "The Multi-List System for Real Time Information Storage and Retrieval," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.

UNIVERSITY OF PITTSBURGH 2.126

Pittsburgh 13, Pa.

JOHN F. HORTY, *Director, Health Law Center, and* WILLIAM B. KEHL, *Director, Computation and Data Processing Center*

Work on applying information retrieval techniques developed by the University of Pittsburgh to legal materials continues, with no change reported since the previous statement [*Ed.*]. The work is supported by grants from the Council on Library Resources, Inc., the National Institutes of Health, and the Ford Foundation.

References:

(1) Bacon, C. R. T. *Technical Report No. 1, An Information Retrieval System for the IBM 650,* Computation and Data Processing Center, University of Pittsburgh, 1960.

(2) Horty, J. F. *Application of Information Retrieval Techniques to Legal Research,* University of Pittsburgh, 1960.

(3) Kehl, W. B., J. F. Horty, C. R. T. Bacon, and D. S. Mitchell. "An Information Retrieval Language for Legal Studies," *Communications of the ACM,* vol. 4, no. 9, September 1961, pp. 380-389.

2.127 UNIVERSITY OF RHODE ISLAND

Aquatic Sciences Information Retrieval Center, Taft Hall, Kingston, R. I.

SAUL B. SAILA, *Assistant Professor of Marine Biology*

Development of a literature storage and retrieval center for the aquatic sciences has been completed under grants from the U. S. Fish & Wildlife Service (Bureau of Commercial Fisheries) and the Sport Fishing Institute.

References:

(1) O'Connor, J. S., S. B. Saila, and T. A. Gaucher. "Steps Toward Retrieval of Literature," in *Current Bibliography for Aquatic Sciences and Fisheries,* vol. 3, part 3: N1-N3. London: Taylor and Francis, Ltd., 1961.

(2) O'Connor, J. S., and S. B. Saila. "A Developing Aquatic Sciences Information Retrieval System," to be published in *Transactions of the American Fisheries Society.*

2.128 WESTERN RESERVE UNIVERSITY

Center for Documentation and Communication Research, Cleveland, Ohio

J. H. SHERA, *Director, and* ALLEN KENT, *Associate Director*

I. Automatic Processing of Abstracts

A project begun in July 1962 aims at achieving fully automatic processing of previously generated abstracts for information retrieval. Metallurgical abstracts having a high degree of consistency in both style and format have been selected as the texts for investigation. Investigation is divided into five phases: (a) investigation of automatic computer input (automatic conversion to computer magnetic tape of Flexowriter punched paper tape, monotype tape, or any automatically convertible medium); (b) design of automatic programs for obtaining a computer printout of the full text in a format suitable for further processing; (c) production of a lexicon—separating those words that are metallurgically descriptive from those that are not, to permit automatic processing of words in running text; (d) design of programs for processing text for information retrieval purposes; and (e) testing for information retrieval effectiveness using questions previously posed to the files of the American Society for Metals Documentation Service. The study is sponsored by the National Science Foundation.

II. Information System and Service for Educational Research Materials

Work was completed in April 1962 on the first year of a pilot study of an information service for educational research materials sponsored

by the U. S. Office of Education. More than 200 test questions were searched over a file of 4,000 documents, 2,000 of which relate to media research and 2,000 to educational research of a general nature (1).

Three current projects relate to the literature of educational research. The first provides for the analysis, coding, and storage of approximately 4,500 additional documents relating to media research, covering the period 1920-1952.

The second of the projects, "Identification of Sources of Educational Research Literature," provides for an exhaustive study not only of the sources but also of the volume of educational research literature, procedures for acquiring that literature, the development of criteria for the selection of material, and an estimation of the costs involved in its acquisition.

The third project, "Effectiveness of a Pilot Information Service of Educational Research Materials," has as its aim an investigation in detail of the unit operations of the pilot information service. The investigation is directed toward: (a) determining the optimum depth of analysis of subject content; (b) refinement of terminology control and development of codes for new terminology; (c) further experimentation with search strategies; and (d) full exploitation of the feedback loop, based on questioner evaluation, to refine and fully exploit the capabilities of the system.

These three projects are being conducted under contracts with the U. S. Office of Education (2).

III. Study of Search Strategy

Fundamental research directed toward search strategy for information retrieval assumes that "information storage and retrieval" as a research problem has too often been oriented to storage or input codes and techniques to the neglect of retrieval aspects. Storage and retrieval are separate problems. There is a need to evaluate and establish strategies for questioning, searching, and manipulating information stores. This effort is study of the parameters for effective and efficient search strategies for information represented in two forms: metalanguage and "telegraphic" abstracts. The design of questions, statistical analysis of answers with respect to pertinence (i.e., customer satisfaction), and probability distributions of data value with respect to time will be explored through use of an existing data file, nonsynthetic questions, and customer feedback. The applicability of alternative strategies and refinements in mathematical models to accommodate varying traffic densities will be considered. Results are expected to suggest improved approaches to information retrieval. This project is supported under a U.S. Air Force Office of Scientific Research grant (3-5).

IV. Information System for Medical Literature and Data

In cooperation with the School of Medicine of Western Reserve Uni-

versity, the Center is developing and extending methods for analysis, coding, and computer exploitation of published and unpublished material in several subject areas of medicine. This involves effort directed toward the establishment of various input forms and techniques, broad systems for encoding medical terminology, and investigations of basic strategies of searching by computer. In addition, mechanized information retrieval systems are being developed which will be capable of handling and retrieving medical data. Specifically under investigation at present are methods of storing and retrieving pathological data accumulated by the School of Medicine. These projects are part of a larger project administered by the Western Reserve University School of Medicine and sponsored by the National Institutes of Health (6).

A cooperative project with the Communicable Disease Center at Atlanta, Georgia, provides for the appraisal of the value of machine searching by the establishment and evaluation of a pilot system in the literature of disease vector control and vector-borne diseases. Approximately 8,000 abstracts of articles relating to disease vector control and vector-borne diseases have been analyzed, coded, and stored on magnetic tape. Seventy-seven pilot questions have been searched and have been evaluated by the Documentation Committee of the Communicable Disease Center. Future work calls for an expanded acquisitions and analysis policy and a study of user requirements. The project is sponsored by the National Institutes of Health (7-9).

A program is underway for the formulation and application of principles, techniques, and procedures for abstracting, encoding, and searching literature relevant to diabetes, including evaluation and testing and the preparation of manuals. A comprehensive titling service is being developed to identify literature relating to diabetes, and plans for publication of the titles and their dissemination to research workers are being made. This project is a segment of a larger investigation of the problems, methods, and applications of modern computer techniques in the machine retrieval of diabetes-related literature on various levels of penetration into the information context. The American Diabetes Association subcommittee (see 2.3) is monitoring this study which is being carried out jointly by the University of Minnesota, the Center for Documentation and Communication Research, and the University of Rochester under a grant from the National Institutes of Health (10) (11).

V. Test and Evaluation of WRU-ASM System

Preliminary results of the Center's participation in the Aslib-Cranfield research project (see 2.9) have been received and evaluated. This study provides for the investigation of the comparative efficiency of various indexing systems including that of Western Reserve University.

These initial investigations have been sponsored by the National Science Foundation (12) (14).

A 2½-year program to test and evaluate procedures for the exploitation, by machine methods, of the literature of interest to metallurgists has just been completed. Reports published as a result of this program included information and data on development of an operational system, acquisition of documents, analysis and quality control, cost analysis of input and output operations, and coding. The program was extended in 1960 to provide detailed case histories on actual questions searched for subscribers to the American Society for Metals Documentation Service to provide data on search performance, relevance, and the effect of relaxation of various levels of detail in the Western Reserve University System. These case histories have produced quantitative and objective data to be applied to the development of testing techniques to measure the relevancy of search output. The program was supported by the National Science Foundation (13) (14).

VI. Retrieval System for Legal Research Materials

Faculty members of the Western Reserve University School of Law and the Center for Documentation and Communication Research are collaborating in an investigation of mechanized techniques for processing and exploiting the various materials of legal research. During the course of the next year, current systems will be studied. Preliminary investigation will be directed toward the formulation of appropriate techniques for analyzing and storing information to satisfy research requirements of varying complexity (15).

VII. Formulation of Systems' Design Criteria

Preliminary systems design criteria are being formulated for the use of an associative memory in information retrieval systems, particularly for investigating thesaural and semantic relationships among words used as index entries, in exploring the effects of alternative strategies of searching, and in normalizing natural language for purposes of searching.

References:

(1) Reeves, Pamela W., A. J. Goldwyn, Jessica S. Melton, and Allen Kent. *The Library of Tomorrow—Today;* An information service of educational research materials: U. S. Office of Education Cooperative Research Project No. 1298 and Title VII Project B-170. Cleveland: Center for Documentation and Communication Research, Western Reserve University, 1962.
Schick, Frank L. "Automated Educational Research Information," *Theory into Practice,* vol, 1, no. 2, April 1962, pp. 94-97.

(2) Reeves, Pamela. "Analysis and Searching of Education Research Literature," presented at "Information Retrieval in Action" Conference, Cleveland, April 19, 1962.

(3) Center for Documentation and Communication Research, Western Reserve University (Staff Report). *Technical Documentary Report on Theory of Documentation and Searching Strategy: Final Report,* AFOSR 2818, 1962. This report summarizes previous investigations and includes a complete bibliography (30 ref.) of all reports and papers prepared under contract No. AF 49 (638) -357.

(4) Goffman, William. *Optimal Probing by an Information Retrieval System,* August 1962. This research was supported in whole or in part under grant number AF AFOSR 62-35.

(5) Kent, Allen. *Textbook on Mechanized Informatiton Retrieval.* New York: John Wiley and Sons, Inc., 1962.

(6) Goldwyn, Alvin J. "The Doctor and the Document—A New Tool for Biomedical Research," *Wilson Library Bulletin,* June 1962, pp. 829-832, 847.

(7) Kent, Allen, A. J. Goldwyn, and Jessica S. Melton. *Fourth Progress Report on Documentation of Disease Vector Control Literature and Literature on Vector-Borne Diseases, Zoonoses and Mycoses,* PHS Grant E-2913, report to Communicable Disease Center, Atlanta, Ga., covering October 1960 to February 1962.

(8) Ferguson, Malcolm. "The Communicable Disease Literature," presented at "Information Retrieval in Action" Conference, Cleveland, April 19, 1962.

(9) Goldwyn, Alvin J. "Abstracting and Indexing in Various Subject Fields," presented at "Information Retrieval in Action" Conference, Cleveland, April 20, 1962.

(10) Lazarow, Arnold. "The Diabetes Literature Program," presented at "Information Retrieval in Action" Conference, Cleveland, April 19, 1962.

(11) Newill, Vaun. "Report of Research and Development of Medical Information Retrieval Projects," presented at the Conference on Data Acquisitions and Processing in Biology and Medicine, University of Rochester, July 17, 1962.

(12) Papers presented at "Information Retrieval in Action" Conference, April 19, 1962:

Kent, Allen. "Purpose of the Cleverdon-WRU Experiment."
Melton, Jessica S. "Search Strategies."
Rees, Alan. "Search Results."
Cleverdon, Cyril. "Conclusions."
Fairthorne, Robert A. "Implications of Test Procedures."

(13) Overmyer, LaVahn. *Test Program for Evaluating Procedures for the Exploitation of Literature of Interest to Metallurgists VI, An Analysis of Output Costs and Procedures for an Opera-*

tional Searching Service, sixth report on program initiated under National Science Foundation grant G-10338.

(14) Smith, Jean, Thomas G. Morris, Alan Rees, and Jessica S. Melton. *Case Histories: Phase Report,* report on program initiated under National Science Foundation grant G-10338.

Melton, John L. *Alternate Search Strategies: Note on Progress,* report on program initiated under National Science Foundation grant G-10338.

(15) Melton, Jessica S., and Robert C. Bensing. "Searching Legal Literature Electronically: Results of a Test Program," *Minnesota Law Review,* vol. 45, no. 2, December 1960, pp. 229-248.

Melton, Jessica S. "The Semantic Coded Approach to Indexing Literature," presented before the Special Committee on Electronic Data Retrieval, American Bar Association, St. Louis, Mo., August 8, 1961; published in *MULL,* March 1962, pp. 48-54.

(16) Papers presented at the "Information Retrieval in Action" Conference, April 18, 1962:

Rees, Alan. "Information Needs and Patterns of Usage."

Belzer, Jack. "Application of the GE 225 to Information Retrieval."

Valvoda, Mary Alice. "A Comparison of Manual and Mechanized Searching Techniques."

Buscher, William. "Programming Techniques and Alternate Searching Strategies."

(17) Rees, Alan M. "Relevancy and Pertinency in Indexing," *American Documentation,* vol. 13, no. 1, January 1962, pp. 93-94.

(18) Melton, Jessica S. "Machine Information Retrieval—Boon or Bust in solving the Communication Problem," *Journal of Engineering Education,* vol. 52, April 1962, pp. 485-491.

(19) Kent, Allen. "Resolution of the Literature Crisis in the Decade 1961-1970," *Research Management,* vol. 5, no. 1, January 1962, pp. 49-58.

(INDEPENDENT WORK) 2.129

"Torran," Crofton Road, Orpington, Kent, England
J. FARRADANE

The overt expression of relations between descriptors is considered essential for accurate storage and retrieval of information. A factual basis for such relational terms has been found in experimental psychological work on the mechanism of thinking. More recent data, which confirm the system of relations, extend possibilities by providing a basis for grouping (i.e., classifying in a new way) descriptors in a system which links accurately with the relational principles. Tests in

the construction and use of such groupings will be made. Theoretical studies for principles of "browsing," combination of indexed records, etc., are also being pursued.

Reference:

(1) Farradane, J. "Relational Indexing and Classification in the Light of Recent Experimental Work in Psychology," to be published in *Information Storage and Retrieval,* vol. 1, no. 1, 1963, pp. 3-12.

2.130 (INDEPENDENT WORK)

412 West Sixth Street, Los Angeles 14, Calif.
REED C. LAWLOR

Methods are under development for the application of modern mathematical techniques and computers to the analysis of law problems.

A mathematical theory of patent claims has been developed and tested and a FORTRAN program has been written (1).

During the past year a mathematical theory of stare decisis, i.e., the law of precedent, has also been developed. A FORTRAN program based upon this theory has also been developed and tested (2) (3).

At the present time, active work is being done on the following: (a) Improved techniques are being tested for developing Boolean equations that express decisions of courts and individual judges as functions of the facts presented in cases in selected fields of law. Standards are being established for the formulation of liberal, moderate, and conservative rules for describing past decisions and, hence, for predicting future decisions; (b) methods and programs are being developed for deriving Boolean equations automatically from a table of facts versus cases and a table of votes of various judges versus cases; (c) a computer program is being developed for writing legal opinions which not only predict the outcome to be expected on the basis of precedent but also set forth a detailed comparison of a case under investigation with various prior cases in the same field of law; and (d) numerous other methods of correlation and prediction of decisions are under investigation.

The programs and work are being tested on the IBM 7090 computer under the sponsorship of Edgar A. Jones, Chairman of the Interdisciplinary Committee of Law and the Administration of Justice of the University of California at Los Angeles.

References:

(1) Lawlor, Reed C. "Analysis of Patent Claims By Mathematical Logic," in *Proceedings 1961 American Bar Association-Patent, Trademark, and Copyright Law Section Symposium on Information Retrieval, St. Louis,* pp. 201-228. Chicago: American Bar Association, 1961. Price: $1.50.

(2) Lawlor, Reed C. "Prediction of Court Decisions," in *Proceedings 2nd National Law and Electronics Conference, May 27-29, 1962 at Arrowhead Lake, California.* Albany, N. Y.: Matthew Bender and Co., Inc., 1962.

(3) Lawlor, Reed C. "Computer Aids to Legal Decision Making," in *Proceedings 1962 American Bar Association-Symposium of Special Committee on Electronic Data Retrieval, San Francisco.* Also to be published in *MULL.*

2.131

(INDEPENDENT WORK)
Partridge Lane, Carlisle, Mass.
BEN-AMI LIPETZ

Work on machine-aided production of library records (1) (2) and on the preparation of citation indexes (3) is being continued independently (see *CRDSD,* No. 10, Statement No. 2.45).

An evaluation is being made of the usefulness of a citation index as a reference aid in the field of physics. An experimental citation index is being prepared by recording instances in which authors in a selected small body of foreign-language physics journals cited earlier papers in familiar English-language physics journals. In the resultant citation index, one will be able to look up a given cited reference from the English-language literature to learn whether and where it was cited in the foreign literature. The evaluation of the usefulness of the index will be based on the hypothesis that the foreign journals included in the sample are not currently being used as much in the United States as their technical content would justify. Statistics will be gathered and analyzed to determine whether the controlled distribution of the experimental citation index leads to increased usage of the foreign-language literature included in the index relative to comparable little-used foreign-language literature not included in the index. The usefulness and the susceptibility to mechanized processing of different degrees of completeness of references and different formats for graphic presentation of data will also be considered in designing and evaluating the citation index.

The citation index evaluation project is supported by funds from a National Science Foundation grant to the American Institute of Physics.

References:

(1) Lipetz, B. A., D. E. Sparks, and L. F. Buckland, *Study of Information Processing Operations and Problems at Air Force Cambridge Research Laboratories (AFCRL) Scientific Library,* Itek Corporation Report IL-9028.01, July 1961. Special Report No. 1 on Contract AF 19(604) 8438. 38 p. plus charts and appendixes.

(2) Lipetz, B. A., D. E. Sparks, and P. J. Fasana. *Techniques for Machine-Assisted Cataloging of Books,* Itek Corporation Report IL-9028-08. Special Report No. 3 on Contract AF 19 (604) 8438, October 1962. 65 p.
(3) Lipetz, B. A. "Compilation of an Experimental Citation Index from Scientific Literature," *American Documentation,* vol. 13, no. 3, July 1962, pp. 251-266.

2.132 (INDEPENDENT WORK)

Biological Abstracts, 3815 Walnut Street, Philadelphia 4, Pa.
JOHN H. SCHNEIDER

A project is underway to develop a decimal classification with extensive cross-references which will permit single-entry indexing and filing of documents dealing with selected scientific fields.

The scientific fields include biochemistry, biology (other than systematics), organic chemistry and chemicals, the chemical industry, physics, utilization of nuclear energy, and subjects pertaining to scientists and research in general. Approximately 22,000 documents have been indexed to varying degrees. A backlog of several thousand more is available for indexing. The decimal classification itself consists of more than 450 typed pages.

During the past 12 months efforts have been directed toward revision of basic biochemistry sections of the classification, extension of classifications dealing with special types of organic compounds, and studies of surfaces. Several thousand documents for indexing have been collected.

The next phase of the project will involve extension of natural science classifications other than biology, classification of the backlog of several thousand documents which involves simultaneous revision and extension of selected sections, and detailed cross-referencing of new topics.

Research underway is a continuation of work on the development of a decimal system previously reported at Vanderbilt University (*CRDSD,* No. 10, Statement No. 2.90).

References:

(The following mimeographed publications were printed at the American University of Beirut, Lebanon, and are available on loan from the author.)

(1) Schneider, J. H. *Organic Chemistry, the Chemical Industry, and Physics: A Decimal Classification,* June 1960. 155 p.
(2) Schneider, J. H. *Biochemistry: A Decimal Classification,* vol. I, December 1960. 78 p.

(3) Schneider, J. H. *Biochemistry: A Decimal Classification,* vol. II, June 1961. 198 p.

(INDEPENDENT WORK) 2.133

5 Colonial Village Drive, Arlington, Mass.
T. M. WILLIAMS

Work on the completion and reporting of studies on a paraphrase-subsuming (normalizing) grammar continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Williams, T. M. *A Graph-Theoretic Logistic Grammar,* IL-4000-20. (In preparation)

(2) Williams, T. M., R. F. Barnes, and J. W. Kuipers. *Discussion of Major Features of a Restricted Logistic Grammar for Topic Representation,* IL-5206-26, February 1962.

3

MECHANICAL TRANSLATION

3. MECHANICAL TRANSLATION

Interest in mechanical translation (MT) research continues to grow. Eight new MT projects are reported. Six of these projects are in organizations reporting MT activity here for the first time: **Computer Usage Company** (3.8), **IBM-France** (3.19), **Kasvatusopillinen Korkeakoulu** (3.22) in Finland, **Lockheed Missiles & Space Company** (3.25). **Výzkumný Ústav Matematických Strojů** (3.44) in Czechoslovakia, and **Yale University** (3.47). Three of the eight projects are focusing on Chinese: Ohio State University (3.30), Thompson Ramo Wooldridge Inc. (3.34), and **Yale University** (3.47). Among the foreign countries reporting new projects are Czechoslovakia, France, and Finland; the latter is reporting MT activity here for the first time.

Almost one-half (22) of the organizations reported in this issue are located in the United States. The remaining 26 are distributed among 13 foreign countries as follows: Belgium (1), Czechoslovakia (3), England (3), Finland (1), France (4), Italy (2), Japan (3), Lithuania (1), Mexico (1), Poland (1), Rumania (1), U.S.S.R. (4), Yugoslavia (1).

Machine Translation, Inc. (3.26), Research, Inc. (3.32), and the University of Michigan (3.39) have discontinued their MT projects. The University of Washington (3.41) reports the completion of its project on Chinese MT.

The work on MT at Karlova Universita, which was previously reported in this section by one statement, now appears as two separate statements. Some of the members of the Division for Machine Translation in the Czech Language Department (3.20) have transferred to the **Institute of Numerical Mathematics** (3.21) in the University where they have formed a new MT group. These groups within the University are cooperating with each other.

Mechanical translation research up to the present has been largely concentrated on the problems of lexicon and the parsing of the source and target languages, principally Russian and English, with relatively little attention being devoted to the problem of synthesis of the output language. Various procedures for the automatic parsing of Russian and English have been partially completed, and there appears to be no reason why fairly reliable computer programs for this purpose cannot be worked out. Such procedures would potentially be applicable to many problems in addition to mechanical translation. Formerly, dic-

tionary sizes of a few thousand words (in canonical form) were frequently proposed. Now, several MT glossaries are over 10,000 canonical forms in size, and a few MT dictionaries providing translation for 100,000 words or more are in operation (IBM, 3.18; University of California, 3.38). Generative grammars, the necessity for which in the translation process stems from the fact that the grammatical rules of any one language are not inherent in those of another and hence cannot be derived from them, are being developed for various languages (see, for example, Leningrad State University, 3.24; Massachusetts Institute of Technology, 3.27; University of Texas, 3.40). Transformational grammars, of importance in permitting the selection of superior target renderings, are also being developed (see Ohio State University, 3.30; University of Pennsylvania, 5.3.37). As the complexity and size of programs increase, attention is being devoted to the development of systematic, and, in some cases, mechanized, procedures for accumulating linguistic data and for the revision of computer programs for MT (University of California, 3.38; University of Texas, 3.40).

There are two trends with regard to generality of application of the MT procedures being developed. On the one hand, several groups are striving to achieve separation of the data from the procedure so that only new data need be provided to change the source or target language in a translation program. On the other hand, other groups are developing more and more detailed programs with a view to achieving the best possible program for specific pairs of source and target languages.

On the more theoretical side, more attention is being turned to problems of polysemy, or the multiple-meaning problem. Little real progress in this area appears to have been made. Some attention has also been devoted to the problem of intermediate languages for MT, but there is little agreement thus far on the role they will play in the future.

Perhaps the most noticeable trends in the field at present are the increasing number of languages under study and the increasing breadth and generality of the theoretical work. In general it should be noted that most projects are concerned with translation into the language or languages of the country in which the research is being conducted.

Work in other areas is sometimes closely related to mechanical translation research. In this issue, the portion of Section 5 devoted to linguistics (5.3) is relevant and should be consulted by those interested in MT. See, for example, the statements describing the work being done by the Československá Akademie Věd (5.3.5), Humboldt-Univer-

sität zu Berlin (5.3.10), Indiana University (5.3.11), Forskningsgruppen för Kvantitativ Lingvistik (KVAL) (5.3.7), Tuskegee Institute (5.3.28), Universität Bonn (5.3.29), University of Pennsylvania (5.3.37), and some independent work in Germany (5.3.43).

A series of translations of foreign-language works, predominantly Russian, on mechanical translation and related topics, produced by the U.S. Joint Publications Research Service (JPRS) under the title "Foreign Developments in Machine Translation and Information Processing," is available from the Office of Technical Services (OTS) of the U.S. Department of Commerce. This series, begun in 1958, was previously available only in limited quantities from various sources. Translations in this series are announced in the *Monthly Catalog of United States Government Publications,* issued by the Superintendent of Documents, U. S. Government Printing Office, and in *Technical Translations,* published bimonthly by OTS.

The Research Information Center and Advisory Service on Information Processing, National Bureau of Standards, has compiled a 519-item bibliography to the aforementioned JPRS translation series. Copies of a preliminary version of the bibliography, entitled *A Bibliography of Foreign Developments in Machine Translation and Information Processing,* are available from the Documentation Research Program, Office of Science Information Service, National Science Foundation, while the limited supply lasts. A final version of the bibliography will be available from OTS in the near future.

The Institute of Linguistics of the Academy of Sciences U. S. S. R. has compiled a 313-item annotated and classified *Bibliography of Foreign Works on Machine Translation* (R. D. Ravich, БИБЛИОГРАФИЯ ЗАРУБЕЖНЫХ РАБОТ ПО МАШИННОМУ ПЕРЕВОДУ, U.S.S.R., 1962). The works cited are, of course, exclusively non-Russian and cover the period 1960-1961. According to the introductory remarks in this bibliography, the Institute of Precision Mechanics and Computing Technique of the Academy of Sciences U.S.S.R. is preparing for publication an annotated bibliography of more than 1,000 Soviet and foreign works on MT and mathematical linguistics for the period prior to 1960.

Meetings related to mechanical translation are announced in the bimonthly publication of the National Science Foundation, *Scientific Information Notes,* available through the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C., for $1.25 per year domestic ($1.75 foreign).

3.1 ACADÉMIE DE LA RÉPUBLIQUE POPULAIRE ROUMAINE

Commission de Linguistique Mathématique,
Institut de Linguistique, Bucharest, Rumania
EM. PETROVICI, *président de la Commission;*
GR. MOISIL *and* AL. ROSETTI, *vice-présidents*

A complete algorithm has been elaborated for English-Rumanian machine translation. The work has been directed by Acad. Grigore C. Moisil; the algorithm was constructed by Erika Domonkos. Recent experiments were satisfactorily conducted on the MECIPT—1 electronic computer of the Timisoara Polytechnical Institute.

The distinct parts of the algorithms are the "input" and the "output." The "input" represents an abstracting process from the source language into numerical symbols. The "output" represents a conversion process from numerical symbols into the target language. A row of abstract symbols is obtained between these parts which includes three kinds of information: (a) the location of the word in the English glossary, showing which dictionary entry is involved; (b) the coding of the English word, extracted from the dictionary showing the word-class; and (c) the collection of all grammatical information obtained in connection with this word in the particular case in the sentence that is being translated.

The "input" identifies this information in the source language and reproduces it in the abstract machine language. The "output" recognizes this information in the machine language and applies it to the Rumanian word.

Between the "input" and "output," the words of a sentence are recorded in storage from full stop to full stop in a sequence called the "matrix of search-locations."

The successive stages of a sentence during the translation process are the following: After the words of an English sentence have been located in the dictionary and recorded in the matrix of search-locations, the abstract row of English words in English word-order results. After the syntactic rearrangements have been executed, the matrix will show the abstract row of English words in Rumanian word-order. After the actual translation, an abstract row of Rumanian words in Rumanian word-order results. Finally, the printed text is a Rumanian sentence.

The dictionary has no separate stem and ending glossary. The method of stripping typical endings has been adopted. If present, these endings are stripped, the appropriate information recorded, and the remaining part of the word is submitted to a further dictionary search.

The introduction of a single word in the dictionary requires seven

storage-locations: the longest English word has 12 letters spreading over two locations (a location has 30 bits plus a sign bit, and a letter has 5 bits); the longest Rumanian word has 18 letters, which consequently requires three locations; to each English word and each Rumanian word, an address is added.

There is a special program to arrange the word sequence in accordance with Rumanian syntax. The first part of this program recognizes the expressions and where the word-order is to be inverted; and the second part actually changes the places of these words.

The algorithm is also able to solve certain cases of ambiguity, such as: (a) some prepositions ("of" to be translated with the Rumanian "din" or "de," or with the possessive case; "to" to be translated with the Rumanian "la" or in determining the infinitive of verbs or the dative of nouns, etc.); (b) grammatical ambiguities (search is made to determine the English preterit: whether it plays the part of a verb and is supposed to be translated with the perfect, or is used as an adjective and is supposed to be translated with a past-participle, which is to be dealt with as an adjective); and (c) ambiguities solvable by grammatical methods (ambiguity of noun-verb, adjective-verb, or noun-adjective).

References:

(1) Moisil, Gr. C. "Préliminaires de la traduction automatique," *Limba Romînă*, vol. 9, no. 1, 1960, pp. 3-10.

(2) Moisil, Gr. C. "Problèmes posés par la traduction automatique. Conjugaison des verbes en roumain," *Studii si Cercetări lingvistice*, vol. 11, no. 1, 1960, pp. 7-24.

(3) Marcus, S. "Traduction d'une langue à une autre, à l'aide de la machine électronique de calcul," *Gazeta Matematică si fizică*, série B, no. 3, 1961.

BIRKBECK COLLEGE (University of London) **3.2**

Department of Numerical Automation, London, England

ANDREW D. BOOTH, *Principal Investigator*

Research activities in the field of mechanized linguistics and mechanical translation, concentrated on text preparation and analysis, continue, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Booth, D. C. "Some Applications of Digital Computers in Medicine," *Physics in Medicine and Biology*, vol. 6, no. 3, January 1962, pp. 377-388.

3.3 CAMBRIDGE LANGUAGE RESEARCH UNIT

20 Millington Road, Cambridge, England
MARGARET MASTERMAN *and* D. S. LINNEY

With support from the U. S. Office of Naval Research, work continues on the development of a general interlingual machine translation procedure in which semantic information, which has already been subjected to syntactic analysis, is classified and transferred by means of a thesaurus. Research has been concentrated on the nature of a thesaurus classification and in particular on the refinement of Roget's thesaurus structure through the use of a small number of very general classifiers. A new thesaurus model has been constructed with a totally different structure from Roget's, and experimental work has now started using a 10,000-word MT Lingua.

Work is being carried out on the syntactic analysis, or bracketing, procedure mentioned above. A technique for using punched cards as a programming tool for syntax bracketing is being experimented with. Although being tested on English, the procedure in general is intended to be interlingual. Use is made of the EDSAC 2 computer, and a small quantity of output to the program has been obtained.

With support from the U. S. Air Force Office of Scientific Research, work is continuing on the development of a program to test the interlinguality of the theory of syntactic function (4). The form of the program has been extended to include governor-dependent, word-order, and concord information.

References:

(1) Cambridge Language Research Unit. *Semantic Problems in Language* (Report on the Colloquium on Semantic Problems in Language held by the CLRU at King's College, Cambridge, September 9-10, 1961), CLRU, M.L. 163, 1962. 229 p. (mimeo)

(2) Masterman, M. "The Semantic Basis of Human Communication," lecture given at the University of Leeds, 1961. (Available in mimeo from CLRU)

(3) Masterman, M. "Semantic Message Detection for Machine Translation," CLRU, 1962; to appear in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

(4) Parker-Rhodes, A. F. "Is There An Interlingual Element in Syntax?" presented at the Ninth International Congress of Linguists, Harvard, Mass., August 27-31, 1962.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

3.4

Centre d'Études pour la Traduction Automatique, Section de Paris (CETAP), Boîte Postale No. 16, Arcueil (Seine), France
A. SESTIER, *Director*

New evidence, based on numerous examples of various kinds, has been obtained to support the theory that conventional formal binary marks of agreement or government are not sufficient for the recognition of the syntactic structure, even when supplemented with the eventual "semantic" clues. The principal reasons for this fact are that a grammatical (dependent) function can be filled by very different types of dependent phrases, and that some functions cannot be filled beside a given governor by more than one dependent phrase (namely, the strongly governed functions). Taking such facts into account in building a grammar is felt quite necessary, especially since they can be formalized rather simply and it is very difficult to add them afterwards.

The definition of the various functions in the above-mentioned sense is being carried out for Russian, French, and German. At the same time, government and dependency codes are established for limited vocabularies.

The major part of the group uses transformations only for pronouns, conjunctions, and various kinds of coordination; for restoring continuity of constituents; and also for avoiding omission of some types of constructions. Deeper and more extensive use of transformations is also being investigated in French.

A systematic lookup of the CETAP Russian stem dictionary has been carried out for the RAND form glossary, in order to compare glossary coverage (63 percent of the glossary is covered by the dictionary) and to use the glossary for checking the form (for the common part). A checked comparative edition of the common part, on any material, will be available very soon.

Research on French is for the most part supported by Euratom.

References:

(1) Tailleur, O. Note 16. *Les paradigmes du Français.*
(2) Dupuis, L. *A Morphological System as a Compromise for Ease of Compilation, Syntactic Research and Adaptability for Further Real A.L.T. Programs.*
(3) Gross, M. *On the Equivalence of Models of Language Used in the Fields of Mechanical Translation and Information Retrieval.*
(4) Sestier, A. *A System of Grammatical Description of the Natural Languages Using the Concept of Grammatical Function.*
(5) Perriault, J. Note 17. *Renseignements statistiques sur le corpus russe in CETAP.*

3.5 CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Centre d'Études pour la Traduction Automatique, Section de Grenoble (CETAG), Faculté des Sciences, Grenoble, France
BERNARD VAUQUOIS, *Director*

Research in mechanical translation involving Russian-French, German-French, and Japanese-French continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Veillon, G. "Programme d'Analyse Morphologique du Russe," in *Annales du 2ème Congres L'AFCALTI* (Assn. Française de Calcul et de Traitement de l'Information), September 1961.
(2) Auroux, A. "Méthode de reconnaissance des structures syntaxiques," *ibid.*
(3) Veyrunes, J. "Etude de la redondance dans les procédés d'indexages morphologique," *ibid.*

3.6 COMMUNAUTÉ EUROPÉENNE DE L'ENERGIE ATOMIQUE (EURATOM), CENTRE COMMUN DE RECHERCHE

Centre de Traitement de l'Information Scientifique (CETIS), Casella Postale No. 1, Ispra (Varese), Italy
P. BRAFFORT, A. LEROY, *and* Y. LECERF

Research on automatic translation involving the four languages of the European Community (Dutch, French, German, Italian) continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Verheyden, Jean. *Les constituants des constituants intermédiaires. Application à l'analyse automatique,* Rapport CETIS No. 28, Euratom, August 1961.
(2) Scheffer, Emilie. *Recueil de stemmas (Tome 1),* Rapport CETIS No. 29, Euratom, August 1961.
(3) Lecerf, Yves. "Intrinsic Machine Addressing in Automatic Translation," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

COMPUTER CONCEPTS INC. 3.7

610 South Broadway, Suite 803, Los Angeles, Calif.
PETER TOMA, *Director of Machine Translations*

The translation project can be divided into one major and two minor endeavors with correspondingly different characteristics.

AUTOTRAN—The system is being developed to produce good quality, high-speed translation of scientific texts from Russian into English. It uses idiomatic and high-frequency dictionaries in the core memory as well as micro-glossaries on tapes. During the main dictionary lookup, morphology and automatic case determiner changes are carried out. The analysis is executed by the Reciprocal Syntax technique which introduces a generalized analysis applicable also to other source languages. The first synthesis is carried out only into English, but provisions are made to attack other target languages later.

The first phase of the system, idiomatic and high-frequency lookups, and the second phase, micro-glossary search and morphology, have been successfully checked out. The Reciprocal Syntax is presently coded for the IBM 7090 computer. The following dictionaries have been prepared and are available on keypunched cards: (a) medical, 35,000 stem entries; (b) atomic energy, 12,000 stem entries; and (c) general, 55,000 stem entries.

Additional dictionary material has been accumulated in 20 subject fields. The coding of subsequent dictionaries will take place in accordance with the translation requirements. A large, trained staff of native-speaking Russian and English personnel is available to code the above-mentioned dictionaries.

MULTITRAN—To determine to what extent Reciprocal Syntax can be applied to other languages, its analysis has been experimentally simulated on English. The article "Exploding Stars" by Robert Kraft, *Scientific American,* April 1962, has been selected as the corpus. Having achieved positive results on many sentences, it has been decided to add a Unified Synthesis technique. In this way, all phases of the translation have been simulated in the following target languages: French, German, Italian, Portuguese, and Spanish. In developing this technique, it was realized that much greater advantage could be taken of the computer in the future if it were used for simultaneous translation into several languages rather than from one language into another. The experiments showed that the Unified Synthesis technique makes it feasible to use the same program with particular language switches when translating into related languages (French, Italian, Portuguese, and Spanish). The next experiment will analyze whether the Germanic Synthesis can be similarly used for Danish, Dutch, Norwegian, and Swedish.

SEMANTRAN—Preliminary research is being carried out to solve the semantic transfer problem on a generalized level. Although the use of micro-glossaries decreases, or in many instances eliminates, this problem whenever scientific texts are translated, no large-scale translation of nonscientific texts will be possible without an adjoining, almost independent semantic meaning transfer. SEMANTRAN envisages a computer-oriented solution using a 12-digit octal number for each semantic concept. Studies are being carried out to determine how a generalized program can use these concepts with control words.

Sentences containing highly ambiguous words are analyzed in various languages. The preliminary results support the feasibility of this approach.

3.8 COMPUTER USAGE COMPANY, INC.

655 Madison Avenue, New York 21, N. Y.
ASCHER OPLER, *Director of Programming Systems*

Work is in progress on development of a system for translating French to English based on the Sakai method (1) (see also 3.10).

The procedure deviates slightly from that described by Sakai in that it has proven unnecessary to go through an intermediate "machine language part of speech." The input product table currently contains 10,000 entries describing syntactic relations between adjacent parts of speech. While the development work is proceeding on an IBM 1401 computer programmed to simulate the Air Force Lexical Processing Machine AN/GSQ-16, the final system will be operational on the latter equipment.

Certain characteristics of the AN/GSQ-16 have simplified the handling of idioms and irregular verbs.

The research is sponsored by the International Business Machines Corporation.

References:

(1) Sakai, I. "Syntax in Universal Translation," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

3.9 電氣試驗所 [ELECTROTECHNICAL LABORATORY]

2-1 Nagato-cho, Chiyoda-ku, Tokyo, Japan
夢沼良一 [R. TADENUMA]

Research on English-Japanese machine translation and machine abstracting continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Tadenuma, R. *English-Japanese Machine Translation,* Research Report of the Electrotechnical Laboratory.

(2) Kita, J. *English-Japanese Machine Translation* (3), Research Report of the Electrotechnical Laboratory. (To be published)

(3) Tadenuma, R., and S. Sugiura. *Machine Abstract,* The Committee of Automation, Institute of Electrical Communication Engineers of Japan, 1961.

防衛庁第1研究所 3.10

[FIRST RESEARCH CENTER, DEFENSE AGENCY]
Mita 13, Meguro, Tokyo, Japan

坂井一郎 [ITIROO SAKAI]

A generalized translation procedure is being studied in an effort to prove the possibility of translating any language into any desired language making use of the linguistic data prepared in a given form.

No specific texts or languages have been selected except that some attempts have been made to re-encode the linguistic data prepared by other authors.

Because of the comparatively small capacity and rather low clock frequency, the NEAC-1103 machine is used for the purpose of testing the procedures only.

In synthesizing an output text, one must arrange the word order according to the rules of the output language, which are not always logical or the same as those of the intermediate language. Also, in some languages the grammatical information brought forward as a group of sememes is represented by a string of a few lexes which should be arranged in a certain order. It has been decided to prepare a set of short patterns of word order and to synthesize an output text that satisfies as many patterns as possible. The principle of programming has been established but the program has not yet been completed.

Preparation of some linguistic data is being considered in order to test the possible extent of generalization.

Reference:

(1) Sakai, I. "Syntax in Universal Translation," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

3.11 GEORGETOWN UNIVERSITY

Machine Translation Research and Language Projects,
1330 New Hampshire Ave., NW., Washington 6, D.C.
LÉON DOSTERT, *Research Director*

The following developments in machine translation research have occurred since May 1962.

The total number of keypunched words in Russian texts, covering five major scientific disciplines (chemistry, physics, biological sciences, and social sciences) has reached the target of 10 million words. The number of entries in the dictionary covering various individual disciplines in the above fields totals about 45,000. A total of 50,000 is expected by spring of 1963.

The GAT (Georgetown Automatic Translation) technique has been tested on the IBM 7090 computer, with fairly voluminous runs in chemistry, physics, and cybernetics. Preliminary evaluations by experts and outside organizations on parts of these runs have been positive with respect to the reliability of the translation.

The Polytechnic Dictionary (Russian-English), to be published by McGraw-Hill, should appear by the summer of 1963.

A test on the Georgetown Russian program for transfer into French will be conducted before the end of 1962 under the auspices of Euratom. A first test in English-Turkish will be conducted in the fall of 1962.

The monograph reviewing the last decade of research in machine translation at Georgetown University will appear by the end of 1962, with supplementary technical reports in the spring of 1963.

3.12 GEORGIA INSTITUTE OF TECHNOLOGY

Engineering Experiment Station, Atlanta, Ga.
B. J. DASHER, *Project Director*

Research on linguistic problems related to machine translation is continuing. The first phase of the investigation of German adverbs is nearing completion, and a report on this work is being written.

The investigation of the Spanish *ser-estar* problem is also continuing. Currently, Spanish adjectives are being cataloged to show the manner in which their meaning is influenced by the choice of *ser* versus *estar*, when these adjectives are used in the predicate position.

The work on machine-oriented structural grammar for English has been dormant during recent months; it will probably be renewed in the fall of 1962.

HARVARD UNIVERSITY **3.13**
Computation Laboratory, Cambridge, Mass.
ANTHONY G. OETTINGER, *Associate Professor of Applied Mathematics and of Linguistics*

A program of research in automatic language translation and mathematical linguistics is being continued.

During the past 6 months, substantial progress has been made in the development and refinement of methods for multiple-path syntactic analysis of English and Russian sentences. Since these methods will produce multiple analyses for syntactically ambiguous sentences, they represent a powerful generalization of the predictive approach to syntactic analysis in its earlier form (2, NSF-5 and NSF-7), which has, to date, yielded at best a single most probable analysis for each sentence.

Multiple-path systems for both Russian and English are now in operation on an IBM 7090 computer. The system for English has been extended in scope from its first operational version (October 1961) to include tests for number agreement between subject and predicate, with a corresponding increase in operating efficiency and in the precision of results. In addition, a sentence diagram subroutine has been added to the main program in order to provide an automatic means of representing the grammatical structures identified in a form isomorphic to a tree diagram. The current English grammar contains nearly 3,500 rules which cover most of the structures which appear or may appear in scientific papers. The system is currently being applied to the automatic analysis of some sizable samples of running scientific text.

Over the past few months, the Russian multiple-path system has been applied to a variety of test sentences with the aim of checking out both the program and the rules of the grammar table. This testing process has led to the elimination of errors in the program, and to a considerable increase in the precision and coverage of the grammar table. The first runs on a scientific text which has also been processed by the earlier predictive system tend to confirm the opinion that the new multiple-path system is both more powerful and more precise than its predecessor.

Work is now underway on a promising technique for preventing both analysis programs from retracing the identical portions of paths which are partially equivalent. Implementation of the technique should not only result in much shorter analysis times for long sentences, where the problem of redundant path-tracing has been particularly acute, but should also lead to a more compact output format with a clearer indication of the location and nature of any syntactic ambiguities in a sentence.

The Harvard Automatic Dictionary file, now stored on IBM magnetic tape, contains over 30,000 stem entries representing approximately 15,000 Russian words or over 150,000 distinct inflected forms, including words of general currency and specialized vocabulary in mathematics and electronics. Each entry is provided with grammar codes necessary for syntactic analysis and English inflection. The dictionary file is currently being used by the automatic translation group at Euratom (see 3.6), as well as by the group at the National Physical Laboratory, Teddington, England (see 3.29). Both the Russian-English dictionary and the grammar table for multiple-path analysis of English can be made available to any qualified research workers.

The research program is sponsored by the National Science Foundation.

References:

(1) Kuno, S., and A. G. Oettinger. "Multiple-path Syntactic Analyzer," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.

(2) *Mathematical Linguistics and Automatic Translation,* series of reports of which the latest is no. NSF-7, November 1961.

(3) *Papers Presented at the Seminar in Mathematical Linguistics,* mimeographed series, of which Volume 8, 1962, is the latest. (On deposit at Widener Library, Harvard University)

(4) Oettinger, A. G. *Automatic Language Translation.* Cambridge, Mass.: Harvard University Press, 1960.

3.14 INSTITUT ZA EKSPERIMENTALNU FONETIKU [INSTITUTE FOR EXPERIMENTAL PHONETICS]

Knez Mihajlova 35, Belgrade, Yugoslavia
DJORDJE KOSTIC, *Director of Research*

About 2 million words taken from the Yugoslav press and contemporary Yugoslav poetry have been classified grammatically and the distribution of grammatical forms determined. The frequencies of each meaning of each form have also been determined, and the results for the press and poetry material compared. The significance of these data for machine translation is also being considered.

A book describing the above work is expected to be completed by the end of 1962.

ИНСТИТУТ ЭЛЕКТРОНИКИ, АВТОМАТИКИ И ТЕЛЕМЕХАНИКИ АКАДЕМИИ НАУК ГРУЗИНСКОЙ ССР [INSTITUTE OF ELECTRONICS, AUTOMATICS AND TELEMECHANICS, ACADEMY OF SCIENCES GEORGIAN S.S.R.] 3.15

Ruchadze Street N 1, Tiflis, Georgian S.S.R., U.S.S.R.

А. И. Элиашвили [A. I. ELIASHVILI] *Project Leader*

Work on an algorithm for Georgian analysis which will be suitable for any language continues, with no change reported since the previous statement [*Ed.*].

References:

The following articles appear in Труды Инст. ЭАТ АН ГССР [*Transactions of the Institute of Electronics, Automatics, and Telemechanics, Academy of Sciences Georgian S.S.R.*], vol. 2, 1961.

(1) Дамения, М. Е. [Dameniya, M. E.] "О методе составления синтетических схем машинного перевода" ["A Method of Compiling MT Synthetic Schemes"], pp. 43-49. (In Georgian)

(2) Дамения, М. Е. [Dameniya, M. E.] "О способе различения омонимных форм в синтезе машинного перевода" ["A Method of Distinguishing Homonymic Forms of MT Synthesis"], pp. 51-54.

(3) Дамения, М. Е., Л. Маргвелани, Л. Тамарашвили [Dameniya, M. E., L. Margvelani, L. Tamarashvili] "A Dictionary of Georgian Verb Endings for Machine Translation," pp. 55-112. (In Georgian)

(4) Пирцхалава, М. Д. [Pirtskhalava, M. D.] "О синтетических схемах машинного перевода союза и частиц в грузинском языке" ["Synthetic Schemes for Machine Translation of the Conjunction and Particles in Georgian"], pp. 113-115. (In Georgian)

ИНСТИТУТ ЯЗЫКОЗНАНИЯ АН СССР [INSTITUTE OF LINGUISTICS, ACADEMY OF SCIENCES U.S.S.R.] 3.16

Komn. 197, Kitaiskii Proezd, 7, Moscow, U.S.S.R.

И. А. Мельчук [I. A. MEL'CHUK]

Work continues on the construction of an algorithm for the automatic analysis of Russian scientific text [*Ed.*].

Reference:

(1) Мельчук, И. А. [Mel'chuk, I. A.]. О стандартных операторах для алгоритма автоматического анализа русского научного текста ["Standard Operators for the Algorithm for Automatic Analysis of Russian Scientific Text"], Машинный перевод, Труды Инст. ТМиВТ [*Machine Translation, Transactions of the Institute of Precision Mechanics and Computing Technique*], no. 2, 1961, pp. 85-134.

3.17

ИНСТИТУТ МАТЕМАТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ АКАДЕМИИ НАУК СССР
[INSTITUTE OF MATHEMATICS, SIBERIAN BRANCH, ACADEMY OF SCIENCES U.S.S.R.]

Novosibirsk 72, U.S.S.R.

А. В. Гладкий, М. В. Рыбакова, Т. И. Шедько
[A. V. GLADKIĬ, M. V. RYBAKOVA, *and* T. I. SHED'KO]

Development of translation algorithms from German to an intermediate language and from the intermediate language to Russian continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Гладкий, А. В., М. В. Рыбакова, Т. И., Шедько [Gladkiĭ, A. V., M. V. Rybakova, and T. I. Shed'ko]. "Схема семантического языка для записи математических текстов" ["A Semantic Language Scheme for Recording Mathematical Texts"], Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста [*Reports at the Conference on Information Processing, Machine Translation, and Automatic Text Reading*], Moscow, January 24-31, 1961. Translation in JPRS 13254, 30 March 1962, available from Office of Technical Services, Department of Commerce, Washington 25, D. C.

3.18

INTERNATIONAL BUSINESS MACHINES CORPORATION

Thomas J. Watson Research Center, P. O. Box 218, Yorktown Heights, N. Y.

GILBERT W. KING, HSIEN W. CHANG, *and* FANG Y. WANG[1]

The development of an entire machine translation complex is continuing within the context of an evolutionary program for research and production. Intensive investigations are continuing in the areas of both hardware and linguistics. The performance of the hardware complex including the Photostore and the Mark II computer has undergone improvement in the light of advances made in the control of linguistic data. A number of studies of both an applied and a basic nature are being pursued within the linguistic area. The extensive Russian-English dictionary has been considerably modified and improved, and the compilation of additional scientific and technical entries continues. This dictionary, together with the proper control entries, is currently being used for experimental automatic translation on a production basis by means of a bi-directional, single-pass translation system. The transformation of this dictionary to a stage of development at which it can be employed with any more sophisticated

[1] Mr. Wang is a consultant to IBM Research Center.

translation system is just about completed. A program for optimum control and manipulation of sentence elements during automatic analysis in terms of syntactic structure and powerful linguistic theory is being elaborated, and a program for the study and definition of search strategies has been initiated. An extensive project for linguistic data gathering and recording is being contemplated for Russian. This work is designed to accompany the intensification of grammatical research in that language. In four languages—Russian, French, Chinese, and English—study programs have been instituted to produce rigorous formulations of grammatical rules for the constant improvement of automatic analysis in the case of Russian, French, and Chinese and automatic synthesis in the case of English.

A program recently initiated to conduct research in machine translation of Chinese-English aims to study the input encoding problems of Chinese characters for processing and to investigate the syntactic and semantic problems in the source and target languages. The effort has resulted in the development of a Chinese input coding device and in the compilation of an experimental Chinese-English MT lexicon based on grammatical analysis and employing the table lookup techniques of the AN/GSQ-16 (XW-1) Language Translation Complex. Continued emphasis will be placed on the study of syntactic and semantic problems of the language pair, and on the enlargement and improvement of the existing dictionary.

The machine translation program is partially supported by the U. S. Air Force (Rome Air Development Center).

References:

(1) Craft, J. L., E. H. Goldman, and W. B. Strohm. "A Table Look-up Machine for Processing of Natural Languages," *IBM Journal of Research and Development,* vol. 5, no. 3, July 1961. (Available from IBM Research)

(2) Galli, E. J. "The Stenowriter—A System for the Lexical Processing of Stenotypy," *IRE Transactions on Electronic Computers,* vol. EC-11, no. 2, April 1962. (Available from IBM Research or IRE)

(3) King, Gilbert W. "Addressing of Very Large Memories," *IBM Journal of Research and Development,* vol. 5, no. 2, April 1961. (Available from IBM Research)

(4) International Business Machines Corporation. *Final Report on Computer Set AN/GSQ-16 (XW-1),* vols. 1-6, June 20, 1959; vols. 7-9, February 9, 1960. Available upon request to Information Processing Laboratory, Rome Air Development Center (AFSC), Air Force Contract AF 30 (602)-1823; or qualified persons may obtain from Armed Services Technical Information Agency (ASTIA).

(5) International Business Machines Corporation. *Final Report on Word Analyzer,* April 1, 1962. Available upon request to Information Processing Laboratory, Rome Air Development Center (AFSC), Air Force Contract AF 30(602)-2072; or qualified persons may obtain from Armed Services Technical Information Agency (ASTIA).

3.19 IBM-FRANCE

5, place Vendôme, Paris 1, France
R. TABORY

Formalized grammars useful for the mechanical treatment of English and French are being investigated. Related auxiliary problems, such as the theory and practice of programming languages, are also dealt with. A growing interest is demonstrated for an eventual linguistic heuristics.

3.20 KARLOVA UNIVERSITA [CHARLES UNIVERSITY]

Faculty of Philosophy, Department of Czech Language, General Linguistics and Phonetics, Division for Machine Translation, Krasnoarmějcu 2, Prague 1, Czechoslovakia
PAVEL NOVÁK, *Leader*

The aim of the project is to set up a translation algorithm from English to Czech for specialized electrotechnical texts. The preparation of the analysis of English is consciously "eclectical" in order to make use of all the experience available. In the synthesis of Czech, the morphological part has been principally completed. The Division cooperates closely with the Institute of Numerical Mathematics attached to the Faculty of Mathematics and Physics (see 3.21).

Reference:

(1) Konečná, Dana, and Helena Novozamská. "Strojový překlad a informační systémy" ["Machine Translation and Information Systems"], *Metodika a technika informací,* no. 7, 1962, pp. 46-53.

3.21 KARLOVA UNIVERSITA [CHARLES UNIVERSITY]

Institute of Numerical Mathematics, Faculty of Mathematics and Physics, Linguistic Group, Krasnoarmějcu 2, Prague 1, Czechoslovakia
PETR SGALL, *Leader*

In cooperation with the Division for the Theory of Machine Translation (see 3.20), an experiment in English-to-Czech machine translation of electrotechnical texts is being prepared. It is planned to

translate a sample of text by using some elements of an intermediate language. Punchcard machines are being used to procure statistical data on the Czech language (1).

Reference:

(1) Panevová, Jarmila. "Využití strojů při zkoumání češtiny" ["The Use of Machines for the Investigation of Czech"], *Slovo a slovesnost*, vol. 23, 1962, pp. 202 ff.

KASVATUSOPILLINEN KORKEAKOULU 3.22

Jyväskylä, Finland
LENNART SCHÄRING *and* TEUVO KUIKKA

Preliminary studies are underway for a program of machine translation from Finnish to Swedish. The project thus far has included extensive preliminary analysis and simulated machine translation, but no actual machine processing.

The group is associated with the Nordiska Sommaruniversitetet.

九州大學 3.23

[KYUSHU UNIVERSITY]

Department of Communication Engineering, Fukuoka, Japan
田町常夫, 栗原俊彦, 大野克郎
[T. TAMATI, T. KURIHARA, *and* Y. OONO], *Principal Investigators*

The reciprocal translation between English, German, and Japanese is being studied, using the KT-1 experimental language-translating machine. Since the structure of Japanese is essentially different from that of other languages, emphasis has been put on syntactical processing. The special-purpose features of this machine enable basic study of syntactical and multiple-meaning problems. The translations between the three languages are performed by a common program, and the directions of the translation are specified by the combination of letters *E, G,* and *J.* An automatic method of syntactical analysis and of extracting the translation rules (for word and syntax) by a learning process has been devised. For this and other advanced studies, a new machine, KT-2, with a larger memory capacity than that of the KT-1 and with some improved functions, is now under construction. Ordinary Japanese grammar has been found to be unsuitable for mechanical translation and a new grammar will be formed for better translation.

This work is sponsored by the Japanese Ministry of Education.

References:

(1) Tamati, T., T. Kurihara, and A. Yosimura. "A Machine for Experiment in Translation," to be published in *Problems in Information Storage and Retrieval.*

(2) Tamati, T., and K. Inadomi. "Kikai-zisyo no Zidô-sakusei no Iti-hôhô" ["A Method of Automatic Formation of Mechanical Dictionary"], presented at the meeting of the Professional Committee of Automata and Control, Institute of Electrical Communication Engineers of Japan, Tokyo, July 1962.

3.24 ЛЕНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ А. А. ЖДАНОВА
[LENINGRAD STATE UNIVERSITY IMENI A. A. ZHDANOV]

Experimental Laboratory of Machine Translation, Leningrad, U.S.S.R.
Н. Д. Андреев [N. D. ANDREEV], *Laboratory Director*

Work is in progress on machine translation from many languages (Russian, English, French, German, Italian, Spanish, Czech, Polish, Rumanian, Hindi, Swedish, Indonesian, Vietnamese, Burmese, Japanese, Chinese, Turkish, Mongol, Arabic, Swahili, Hausa, Korean, Bulgarian, Kirghiz, Bengali, Lettish, Estonian, Tamil, and Finnish) into the artificially constructed †*Intermediary Language* (IL), and from the IL into Russian, German, and Indonesian. Electronics texts provide the main subject matter for the translation research. The vocabulary of the IL now includes nearly 6,500 units called †*semoglyphs*, and is growing at an average rate of 100 semoglyphs per month. The grammar of the IL was completed 2 years ago. Both the grammar and the vocabulary are being tested in a long series of experiments, the first of which was completed July 1960 and consisted of a translation of one page of a scientific text from Indonesian into IL (and further into Russian) through the use of the URAL computer. All of the algorithms are written in a symbolic language, specially developed by the group, called ortho-language (as compared with the descriptive *metalanguage* of machine translation).

References:

(1) Андреев, Н. Д. [Andreev, N. D.]. "Linguistic Aspects of Translation," presented at the Ninth International Congress of Linguists, Cambridge, Massachusetts, August 27-31, 1962.

(2) Андреев, Н. Д., В. Л. Замбржицкий, Р. В. Пазухин, Г. А. Пак [Andreev, N. D., V. L. Zambrzhitskii, R. V. Pazukhin, and G. A. Pak]. Построение и функционирование словаря языка-посредника в поле входных языков [*Construction and Functioning of an Intermediary Language Dictionary in the Field of Input Languages*], Moscow, 1961.

(3) Андреев, Н. Д., Н. Д. Кремнева, Н. Н. Волков [Andreev, N. D., N. D. Kremneva, and N. N. Volkov]. Влияние типа грамматики языка-посредника на схему алгоритмов

† *See* Glossary.

машинного перевода [*The Influence of a Type of Intermediary Language Grammar on a Scheme of MT Algorithms*], Moscow, 1961.

(4) Андреев, Н. Д. [Andreev, N. D.]. "О некоторых предварительных требованиях к переводческой машине"["On Some Preliminary Requirements for a Translation Machine"], Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста [*Reports at the Conference on Information Processing, Machine Translation and Automatic Text Reading*], Moscow, 1961.

LOCKHEED MISSILES & SPACE COMPANY 3.25

Research and Engineering Laboratory, Palo Alto, Calif.
B. D. RUDIN, *Principal Investigator*

Research is aimed at reciprocal English-Russian machine translation. The syntax and morphology of each of the languages are under separate study, and machine programs are being written for automating the parsing of both languages. The method of study consists of formulating relational rules, programming them on an IBM 7090 computer, and then subjecting them to randomly selected text for evaluation. A program has also been written which reconstructs a sentence from its parsed form (a "prosing" program).

Programs for translating parsed Russian into English are now being planned.

The study began in July 1960.

MACHINE TRANSLATION, INC. 3.26

821 Fifteenth Street, NW., Washington 5, D.C.
ARIADNE LUKJANOW LOEWENTHAL, *President*

Work on the production translation system previously reported as under development has been discontinued [*Ed.*].

MASSACHUSETTS INSTITUTE OF TECHNOLOGY 3.27

Research Laboratory of Electronics, Cambridge 39, Mass.
VICTOR H. YNGVE, *Principal Investigator*

The primary objective of the research program is to find out how languages can be translated by machine. Secondary objectives are concerned with evaluating the fidelity that can be achieved with different approaches to mechanical translation, the usefulness of the translations for various purposes, and their costs. A further objective is to add to the general knowledge of noncomputational uses of digital computing machinery and to a basic understanding of human communication.

Considerable work has been done on grammar and syntax. Understanding of the sentence structures of English, German, French, Arabic,

and Russian has advanced to the point where running computer programs are available for producing grammatical sentences at random in four of these languages. The programs represent, of course, only a modest first step in achieving the degree of understanding and in collecting the wealth of detail that will ultimately be required. The syntactic work has not been confined to the writing of these programs, however. There have been separate studies of such topics as the impersonal construction in German, the order of adjectives in German and English, the German inseparable prefixes, and various topics on word order.

In the theory of translation, there have been several investigations into the possibility of setting up structural correspondences between two languages on a formal syntactic basis at the lexical and transformational levels. In addition, there is a running computer program that will translate from a narrowly restricted set of Arabic sentences into English.

In the area of semantics, interesting results have been obtained in the analysis of those features of natural language that function as the analog of the free-variable in logical and mathematical systems. The analysis includes an investigation of some of the cases where time-order and tense-structure are involved. Work is also progressing on the problem of generating meaningful or significant sentences, as opposed to sentences that are legitimate merely in the sense of satisfying a set of syntactical criteria. This work has led to the need for a mechanical proof procedure. For this purpose, the Davis-Putnam method has been programmed. Extensive investigations are also being carried out on procedures of proof in implicationally ordered formal systems. In these investigations, the computer is being used in an essential way as a research tool.

Another investigation is concerned with methods of efficient recognition of the grammatical structure of a sentence. Efficient recognition is being sought by appropriate utilization of information abstracted from actually observed sentences.

The rapid development of computer programs in these various areas has been made possible by the completion of the COMIT system, which makes available a high-level programming language that is convenient to use and that minimizes the time required for writing and checking out the program. The COMIT system has been distributed through the SHARE organization and is available to all users of IBM 709 or 7090 computers.

The program is sponsored by the National Science Foundation.

References:

(1) Bross, John S. "Problems of Equivalence in Some German and

English Constructions," *Mechanical Translation,* vol. 7, no. 1, July 1962, pp. 8-16.
(2) Charney, Elinor K. "On the Semantical Interpretation of Linguistic Entities that Function Structurally," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)
(3) Darlington, Jared. "Interlingua and MT, a Discussion," *Mechanical Translation,* vol. 7, no. 1, July 1962, pp. 2-7.
(4) Klima, E. S. "Structure at the Lexical Level and Its Implications for Transfer Grammar," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)
(5) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 64,* January 15, 1962, pp. 207-229.
(6) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 65,* April 15, 1962, pp. 175-186.
(7) M.I.T., Research Laboratory of Electronics, *Quarterly Progress Report No. 66,* July 15, 1962, pp. 289-293.
(8) M.I.T., Research Laboratory of Electronics, *Quarterly Progress Report No. 67,* October 15, 1962, pp. 167-176.
(9) Quillian, Ross. "A Revised Design for an Understanding Machine," *Mechanical Translation,* vol. 7, no. 1, July 1962, pp. 17-29.
(10) Yngve, V. H. *An Introduction to COMIT Programming* and *COMIT Programmers' Reference Manual.* M.I.T., November 1961. (Available through SHARE distribution 1198)

NATIONAL BUREAU OF STANDARDS 3.28

Applied Mathematics Division, Washington 25, D. C.
IDA RHODES, *Project Leader*

Work is continuing on a long-range program for practical mechanical translation from Russian into English. Computer codes written earlier in this project have been adapted to the IBM 7090 computer and successfully tested on a few sample sentences. In these tests, the process of †"profiling" has been simulated; the preparation of a machine program for this phase is in process. The machine codes for "parsing" and for finding and rearranging English equivalents are largely complete but lack portions for dealing with infrequently occurring grammatical constructions. Work on the 5000-stem sample dictionary has progressed to the point where morphological information has been recorded for more than one-third of the stems.

† *See* Glossary.

The research is sponsored jointly by the U. S. Army Signal Corps and the U. S. Army Research Office.

References:

(1) Alt, Franz L. "Recognition of Clauses and Phrases in Machine Translation of Languages," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

(2) Alt, Franz L. "The Outlook for Machine Translation," in *Proceedings of the Western Joint Computer Conference,* Vol. 17, San Francisco, Calif., May 1960.

(3) Alt, Franz L., and Ida Rhodes. "The Hindsight Technique in Machine Translation of Natural Languages," *Journal of Research of the National Bureau of Standards,* Section B. vol. 3, April-June, 1962, pp. 47-51.

(4) Rhodes, Ida. "A New Approach to the Mechanical Syntactic Analysis of Russian," *Mechanical Translation,* vol. 6, 1961, pp. 33-50.

3.29 NATIONAL PHYSICAL LABORATORY

Autonomics Division, Teddington, Middlesex, England

A. M. DAY, J. McDANIEL, W. L. PRICE, A. J. SZANSER, S. WHELAN, *and* D. M. YATES

Programming for the operation of the Harvard Russian-English dictionary on the ACE computer is nearing completion. Much study still needs to be done, however, for the selection of preferred English correspondents for the Russian entries. Coding for the reflexive—ся forms of verbs in the dictionary is underway.

A second version of the NPL elementary syntactic analysis scheme for Russian has been programmed and is ready for testing. Included in the scheme are "noun-blocking" and some rearrangement of prepositional and participial phrases and also of inverted sentences.

3.30 OHIO STATE UNIVERSITY

Division of Linguistics, 216 N. Oval Drive, Columbus 10, Ohio

WILLIAM S-Y. WANG, *Principal Investigator*

Linguistic analysis is being done on Chinese and English. A theoretical paper has been prepared which examines the "traffic rules" in English syntax in the hope of better defining the form of a grammar. A preliminary formulation of the constituent structure of Mandarin syntax and a detailed analysis of the "potential verb" have been completed. A study is being made on a class of four-syllable constructions with respect to their internal composition and external distribution.

Correspondences between the grammars of the two languages are continually being sought.

In the interest of promoting research efficiency on the two languages, some Chinese grammatical studies are being translated into English, Chomsky's *Syntactic Structures* is being edited and translated into Chinese, and POSA Report No. 2 (3) is being expanded and will include a glossary of Chinese-English grammatical terms.

Controlled recordings of the complete set of citation syllables in Mandarin have been made. Detailed measurements are being taken on the major acoustical parameters of the vowel in these syllables. Attention is being paid to the interaction of these parameters on the phonetic characteristics of the accompanying consonants and tones.

Work on the syntactic component of the grammar, supported by the National Science Foundation, began in July 1961. Work on the phonological component, supported by the U. S. Office of Naval Research, was initiated in July 1962.

References:

(1) Ching, Eugene. *Four Syllable Constructions in Mandarin.* (In preparation)

(2) Fillmore, Charles J. *The Position of Embedding Transformations in a Grammar,* POSA Report No. 4, November 1962.

(3) Wang, William S-Y. *Synchronic Studies in Mandarin Grammar: A Selected Bibliography,* POSA Report No. 2, February 1962.

(4) Wang, William S-Y. *Some Syntactic Rules for Mandarin Chinese,* POSA Report No. 3, October 1962.

THE RAND CORPORATION 3.31

1700 Main Street, Santa Monica, Calif.

DAVID G. HAYS *and* THEODORE W. ZIEHE, *Project Leaders*

Although no work directed solely to machine translation has been carried out in nearly 2 years, much of RAND's linguistic research (see 5.3.26) has at least tangential relevance to MT.

References:

(1) RAND Research Memoranda (RM) are placed on deposit at several university libraries in the United States and abroad. All of RAND's research products, including text and glossaries on punched cards or magnetic tape, can be obtained by other research workers under appropriate circumstances.

RM-2066-1. *Studies in Machine Translation—6: Manual for Coding Russian Grammar,* by K. E. Harper, D. G. Hays, and D. V. Mohr, March 3, 1958; rev. December 29, 1960.

Studies in Machine Translation—7: Manual for Assigning Word Numbers and English Equivalents to Russian Forms, by

H. P. Edmundson, K. E. Harper and D. G. Hays. (In preparation)

RM-2068. *Studies in Machine Translation—8: Manual for Postediting Russian Scientific Text*, by K. E. Harper, D. G. Hays, and B. J. Scott, July 15, 1960.

RM-2538. *Studies in Machine Translation—10: Russian Sentence-structure Determination*, by D. G. Hays and T. W. Ziehe, April 1, 1960. (ASTIA No. AD—238 096)

Studies in Machine Translation—11: Guide to Tables of Russian Dependency Types and Syntactic Equivalences, by K. E. Harper and J. H. Pustula. (In preparation)

RM-2655. *Studies in Machine Translation—12: A Glossary of Russian Physics*, by A. S. Kozak, C. H. Smith, and members of the RAND ALDP Group, October 14, 1960.

(2) Hays, D. G. *Research Procedures in Machine Translation*, RAND Memorandum RM-2916-PR, December 1961.

3.32 RESEARCH, INC.

1111 Union Central Tower, Cincinnati 2, Ohio

JOHN SCOTT DAVENPORT

Work concerning evaluation of the quality of mechanical translation in terms of the communication of true (or intended) meaning has been suspended.

3.33 SUMMER INSTITUTE OF LINGUISTICS

Box 1960, Santa Ana, Calif.

JOSEPH E. GRIMES, *Principal Investigator*

Work on the possibility of using machines to resynthesize a human translator's behavior by matching word strings in material to be translated with the translation of identical strings in a large body of text has been temporarily suspended.

Reference:

(1) Grimes, Joseph E., and Manuel Alvarez. "The SIL Concordance Program," presented at the meeting of the Linguistic Society of America, Austin, Tex., July 1961.

THOMPSON RAMO WOOLDRIDGE INC. **3.34**
RW Division, Canoga Park, Calif.
H. P. EDMUNDSON, PAUL L. GARVIN, JULES MERSEL, C. A. MONTGOMERY, GERHARD REITZ, STEVEN B. SMITH, *and* DON R. SWANSON

Experimental translation of Russian physics and other literature is being carried out using general-purpose computing equipment. The principal goal of this work is to produce a machine program for automatic translation of Russian into English of usefully high quality and, within the framework of that goal, to focus particularly on problems of semantics and multiple meaning. Considerable emphasis is placed on the design and implementation of a systematic research methodology which permits efficient use to be made of a computer as a research tool for organizing and analyzing language data. Problems of syntax have also received attention. The resulting syntactic program has been designed to serve two major purposes: (a) sentence analysis for machine translation per se, and (b) recognition and "packaging" of syntactically related structures within the sentence for the purpose of studying multiple-meaning problems that are influenced by those syntactic relations.

Presently the RW Division is engaged in (a) developing computer techniques which will automatically detect errors in the machine translation output by a computer comparison of the machine translation and the human translation, (b) applying the fulcrum technique to semantic research, and (c) exploratory studies in Chinese-English translation.

The research is carried out with partial support from the National Science Foundation and the U. S. Air Force (Rome Air Development Center).

References:

(1) Swanson, Don R. "The Nature of Multiple Meaning," in *Proceedings of the National Symposium on Machine Translation,* ed. by H. P. Edmundson, pp. 386-393. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961.

(2) *Machine Translation Studies of Semantic Techniques,* Final Report, prepared under Contract AF 30 (602)-2036 for Rome Air Development Center, February 22, 1961.

(3) Swanson, Don R. " 'Word Block Model' for Russian-English Syntax," in *Information Processing.* London: Butterworths Scientific Publications, 1960.

(4) Mersel, Jules. "Research in Machine Translation at Ramo-Wooldridge," in *Proceedings of the National Symposium on*

Machine Translation, ed. by H. P. Edmundson, pp. 26-38. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961.

(5) Garvin, Paul L. "Syntactic Retrieval," in *Proceedings of the National Symposium on Machine Translation,* ed. by H. P. Edmundson, pp. 286-292. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961.

(6) Garvin, Paul L. "Syntax in Machine Translation," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press)

(7) Mersel, Jules. "Programming Problems in Machine Translation," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press)

(8) *Progress Report No. 1 to National Science Foundation on Computer-Aided Research in Machine Translation,* prepared under Contract NSF-C233, July 19, 1962.

(9) *Survey of Problem Areas in Chinese-English Machine Translation,* Technical Note No. 1 under Contract NSF-C233 with National Science Foundation, August 21, 1962.

3.35 UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Centro de Cálculo Electrónico,
Mexico City 20, Mexico
SERGIO F. BELTRAN, *Director*

The adaptation of the University of California RUSDIC to Spanish has been completed, and the perforation of cards has been started. It is planned to employ the Spanish RUSDIC in collaboration with the University of California (see 3.38). Thus, any text prepared for translation can be processed for both English and Spanish, thereby avoiding much duplicated effort.

Plans have been made for using the English and Spanish RUSDICs for constructing English-Spanish and Spanish-English dictionaries. The procedure will involve joining the translations of each Russian element and then thoroughly editing the result. Since the choice of elements for the dictionary has proved to be one of the most time-consuming phases of the preparation of the RUSDIC, it is believed that the proposed procedure will save much time in this regard.

3.36 UNIVERSITÀ DEGLI STUDI DI MILANO

Centro di Cibernetica e di Attività Linguistiche, Milan, Italy
SILVIO CECCA'TO, *Principal Investigator*

Research on automatic language translation is being carried out. Russian-English translation has been worked on primarily; German

and Italian output are being added to the procedure. An English input procedure is being developed, and an experiment of English-Latin translation of random sentences made with a limited vocabulary has been perfected.

On the basis of past experience a tabular method of classification has been adopted, in which the †correlators (200 to 300 for each language) are picked out individually, while the †correlata, i.e., the words correlated by them, are arranged in classes which are gradually being broken up into groups with a view to treating them eventually as individuals as well (when semantic analysis covers the whole field).

Correlator tables of the first degree have been prepared for five languages. The semantic categories and the transformations of †correlational nets relating to some 500 Russian verbs have been worked out for English, German, and Italian output, within the limits of a vocabulary of 2,500 head-words.

The programs previously elaborated for the IBM 650 and IBM 704 computers have been adapted for use on the IBM 7090 computer.

Research is being carried out on the problems of punctuation marks and the insertion of articles in translation from Russian. In order to reduce the vast number of combinatory operations, a hierarchy of correlational possibilities is being studied which will reduce machine work on the basis of more or less probable occurrences.

The research is carried out under contract with the U.S. Air Force (Rome Air Development Center) through its European Office of Aerospace Research, and under a collateral contract with Euratom.

References:

(1) Centro di Cibernetica e di Attività Linguistiche. *Linguistic Analysis and Programming for Mechanical Translation,* Technical Report prepared for European Office, Air Research and Development Command, U. S. Air Force, under contract No. AF 61 (052)-212, Milan, 1960. Milan: Feltrinelli Editore; New York: Gordon & Breach Publishers, Inc., 1961.

(2) Ceccato, Silvio. "Natural Languages and Artificial Languages: Analysis and Transformation," *Nonnumerical Data Processing Symposium,* Blaricum (Holland), 1961.

(3) Ceccato, Silvio. "La Storia di un Modello Meccanico dell'Uomo che Traduce," in *Almanacco Letterario 1962,* pp. 122-134. Milan: Bompiani, 1961.

(4) Ceccato, Silvio. "La Traduzione nell'uomo e nella macchine," *Civietà delle Macchine,* vol. 9, no. 5, Rome, 1961.

(5) Ceccato, Silvio. "Four Lessons in Mechanical Translation," in

† *See* "Correlation" and "Correlational Net" in Glossary.

NATO Advanced Study Institute, Automatic Translation of Languages, Venice, July 15-31, 1962. (In press)

(6) Ceccato, Silvio. "Cibernetica e Linguistica," in *Cultura e Scuola.* (In press)

(7) Glasersfeld, Ernst v. "Notes Concerning Output Matrices," in *Linguistic Analysis and Programming for Mechanical Translation,* pp. 170-174. Milan: Feltrinelli Editore; New York: Gordon & Breach Publishers, Inc., 1961.

3.37 UNIVERSITÉ LIBRE DE BRUXELLES

Groupe de Traduction et de Documentation Automatiques, Institut de Statistique, 87, avenue Adophe Buyl, Brussels 5, Belgium

P. P. GILLIS, JM. FAVERGE, *and* L. S. HIRSCHBERG

Studies in automatic linguistics are being conducted with a view toward application to mechanical translation and information retrieval.

The immediate practical objective is mechanical translation of scientific texts from Russian into French, especially high energy physics, but general questions of methodology will also be investigated. The theoretical research is based mainly on French and is conducted in close cooperation with the CETIS group of Euratom (see 3.6). An attempt is being made to make explicit the †categorical and †heuristic rules used for the purpose of syntactic analysis by the various MT groups and to distinguish between these two kinds of rules. As far as the experimental part of the research and the construction of an MT program from Russian into French are concerned, the method of the GAT group (see 3.11) is being applied, with the direct assistance of the Georgetown University and CETIS groups. The programs are written for the IBM 7090 computer of CETIS Euratom (Ispra), but some practical experiments are performed on the IBM 1620 of the University.

In a first stage, the syntactic analysis program for Russian developed at Georgetown University is being used without modification. Their Russian-English dictionary is being translated into a Russian-French one, using the Georgetown concordances of scientific texts. This permits assessment, on the basis of experience, of the extent to which the same analysis of a source language (Russian) can be used for different target languages (English and French) and what the characteristics of such an analysis must be in order to achieve this type of invariance. A French synthesis program is being constructed on the basis of the data which can be expected from typical analysis programs in the present state of the art. This synthesis program is an independent

† *See* Glossary.

entity which can be adapted to the present analysis program or, if necessary, to a new more refined version. The study of French morphology for automatic synthesis has been completed; the programs have been written and debugged (1).

Currently, the remaining French synthesis programs are being written and will be published as soon as tested on pre-analyzed sentences. Simultaneously, the Russian-French dictionary is being constructed (2).

In the area of information retrieval, an experiment is being conducted in the medical field (cancerology); results will be published later. The theoretical problems originated by the experimental work are also being studied.

Cooperation with groups of human translators is planned in order to construct specialized dictionaries based on concordances of legal and administrative texts. Information retrieval experiments are planned in order to attempt the introduction of some semantic rules in the MT programs. Theoretical research will be continued.

This research is being performed under contract from Euratom.

References:

(1) Blois, J., and E. Morlet. *Morphologie du Français pour la traduction automatique,* Rapport CETIS No. 44, Euratom, April 1962.

(2) Blois, J., and L. Hirschberg. *Manuel de codage du français,* Université de Bruxelles, Contract Euratom no. 081-61-5-Cet-B.

(3) Braffort, P., L. Hirschberg, and J. Mommens. *La programmation des problèmes non numériques,* Rapport CETIS No. 31, Euratom, September 1961.

(4) Buydens, J. *Liens de dépendance grammaticale et classification sémantique,* Rapport CETIS No. 38, Euratom, November 15, 1961.

(5) Hirschberg, L. *Discussions sur l'hypothèse de projectivité,* Rapport CETIS No. 35, Euratom, September 1961.

(6) Hirschberg, L. *Ponctuations et Grammaires de Dependance,* Université de Bruxelles, July 1962, Contract Euratom no. 018-61-5-Cet-B.

UNIVERSITY OF CALIFORNIA 3.38

Machine Translation Project, Berkeley, Calif.
SYDNEY M. LAMB, *Principal Investigator*

Research is in progress on machine translation of technical literature, with special emphasis on translation from Russian and Chinese to English. In addition, the project is cooperating with the Universidad Nacional Autonoma de Mexico (see 3.35) in extension of its systems

and techniques to translation from Russian, Chinese, and English into Spanish. The group has followed the approach of designing a general translation procedure intended to be applicable to translation from any language to any language, when supplied with properly organized information for specific language pairs, e.g., Russian-English, Chinese-English. The design of the translation procedure, along with other aspects of the research, is based upon the †stratificational theory of linguistic structure (5).

The work underway falls into four areas: (a) theoretical work on the nature of linguistic structure and on the design of MT and other †mechanolinguistic systems; (b) the design and writing of portions of a general MT program; (c) the accumulation, analysis, and organization of linguistic information on Russian, Chinese, and English; and (d) the design and writing of computer programs to facilitate future accumulation and analysis of linguistic information.

I. Theory

Work is continuing on various details of the stratificational approach to language structure (5). This work is currently concentrating on †semology and syntax.

II. Programming for MT

Until recently, programming was done for the IBM 704 computer, and much of the current programming effort is devoted to preparing versions for the newly installed IBM 7090 computer. Programs for the following stages of the decoding process have been coded but have not yet been checked out on the 7090: (a) segmentation (into †lexes) and dictionary lookup (1), (b) conversion from lexes to †lexemes, and (c) determination of syntactic relationships. The dictionary adapter (DICADAP), which reorganizes human-oriented dictionaries for efficient use by the computer (2), is operational on the 704 but not on the 7090. The dictionary lookup and segmentation program can accommodate dictionaries with large vocabulary coverage and has a speed of 800 words per second on the 7090 (1). The parsing program, which operates by grouping peripheral members of syntactic constructions with the partners presupposed by them, consists of 125 FAP instructions.

III. Accumulation of linguistic information

Work in this area is being conducted intensively for Russian and Chinese and on a small scale for English. For Russian, research is concentrated on: (a) syntax, (b) the addition of further information to the project's Russian-English dictionary (2) (6), and (c) chemical nomenclature (3). The Russian-English dictionary is intended for

† *See* Glossary.

use mainly with chemical and biological literature, and has a vocabulary coverage of about 2 million assumed current words (or around 200,000 stems) plus about 5 million stems of chemical names. For Chinese, work is in progress on dictionary compilation and on syntax. A five-volume set of cross-referenced indexes for 10,000 Chinese characters is to be published in 1963 (7). Some work is also being done on English syntax and †sememics.

IV. AUTOMATION OF LINGUISTIC ANALYSIS

Work in this area includes (a) a program for automatic †tactic analysis, which has thus far been completed only for the initial stages (4), and (b) RUSTAN, an automatic Russian text analyzer, which is based upon the decoding stages of the Russian-English translation system. (About 200,000 words of Russian text from the journal *Biokhimija* are in machinable form ready for analysis.) In addition, the project plans to take advantage of various programs designed to aid linguistic analysis which are being written as part of a separate mechanolinguistics project being conducted at the university (see 5.3.30).

The project, which has been in existence since October 1958, is supported by the National Science Foundation.

References:

(1) Lamb, Sydney M., and William H. Jacobsen, Jr. "A High-Speed, Large-Capacity Dictionary System," *Mechanical Translation,* vol. 6, 1961, pp. 76-107.

(2) Lamb, Sydney M., William H. Jacobsen, Jr., Russell K. Gardiner, and John H. Wahlgren. *On the Organization of information in an MT Dictionary,* Research Memorandum, 1961.

(3) Wahlgren, John H. "Linguistic Analysis of Russian Chemical Terminology," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

(4) Lamb, Sydney M. "On the Mechanization of Syntactic Analysis," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

(5) Lamb, Sydney M. *Outline of Stratificational Grammar,* University of California, Berkeley, 1962.

(6) Johnson, C. Douglas, Luis G. Ireland, Sydney M. Lamb, and John H. Wahlgren. *The Refined Grammar Codes of Russian*

† *See* Glossary.

Nonchemical Nonsuffixal Alphabetic Lexes, Research Memorandum, 1962.

(7) Dougherty, Ching-Yi, Sydney M. Lamb, and Samuel E. Martin. *Chinese Character Indexes.* Berkeley, Calif.: University of California Press. (In press)

3.39 UNIVERSITY OF MICHIGAN

Institute of Science and Technology, Ann Arbor, Mich.
ANDREAS KOUTSOUDAS,[1] *Project Director*

Research on the development of a theory of translation which would form the basis for the construction of a "learning" or "self-modifying" program to translate any one natural language into another has been discontinued [*Ed.*].

References:

(1) Koutsoudas, A. "Machine Translation at The University of Michigan," in *Information Retrieval and Machine Translation,* Part II, ed. by Allen Kent, pp. 765-770. New York-London: Interscience Publishers, Inc., 1961.

(2) Smoke, W., and E. Dubinsky. "A Program for the Machine Translation of Natural Languages," *Mechanical Translation,* vol. 6, November 1961, pp. 2-10.

3.40 UNIVERSITY OF TEXAS

Linguistics Research Center, P. O. Box 7980, University Station, Austin 12, Tex.
W. P. LEHMANN, *Director and*
E. D. PENDERGRAFT, *Associate Director*

I. Machine Translation Study

The program of mechanical translation research has a long-range but primarily practical purpose: to implement automatic translation of artificial and natural languages by means of a generalized computer system.

The *USASC Language Translation System* is now being programmed. The system is designed for the IBM 709 computer with two input-output channels and on-line printer. All testing is done at the U. S. Army Electronic Proving Ground, Fort Huachuca, Ariz.

The research was begun in May 1959 under sponsorship of the U. S. Army Electronics Research and Development Laboratory, Fort Monmouth, N. J.

[1] Present address: Department of Linguistics, Indiana University, Bloomington, Ind.

II. Development of a Linguistic Computer System

During the course of MT research, it became apparent that automatic programming techniques could be applied to linguistic processes other than translation. New research was initiated for the broad purpose of studying a wide range of linguistic processes. Of particular interest will be processes which are potentially applicable to scientific documentation. In addition to mechanical translation, these include inferential and self-organizing processes which may be relevant to automatic abstracting and to information storage and retrieval.

The *Linguistic Research System* is being programmed. The system is designed for the IBM 7090 computer with two input-output channels and on-line printer. Testing is carried out at C-E-I-R, Houston, Tex. The project was initiated in September 1961 under National Science Foundation sponsorship.

The research on the two above-mentioned projects may be organized into the following subtasks which are appropriate for different technical specialists. At present, three research groups are participating in the two projects; these are comprised of linguists, computer programmers, and mathematicians.

Mathematical Research. Natural language is widely regarded as an organized system or structure composed of various interrelated components. The task of the mathematical research group is to improve linguistic theory through formalization and analysis of these structural assumptions.

One major component of natural language is the structure of expression; a second is the structure of content. A mathematical hypothesis which formalizes these two structures, and the intricate relationships between them, was completed during the past year. The system of language described by the hypothesis is a hierarchy of two levels. The first, or syntactic, level makes up the structure of expression. The structure of content consists of the second, or semantic, level in the hierarchy.

Programming Research. Generalized algorithms for linguistic analysis and synthesis, both at the syntactic and semantic levels, have been deduced from the structural hypothesis. Other operational interpretations of the hypothesis have made possible the specification of a series of generalized algorithms for language translation. The algorithms contain interlingual transfer processes which operate at progressively higher levels in the structural hierarchy. As interlingual transfer is accomplished at successively higher levels, the algorithms will have the capacity to produce translations of better quality.

These linguistic algorithms are quite complex and may be realized

in the computer in many different ways. The programming research group is concerned with finding suitable computer adaptations for the linguistic algorithms. This search for programming solutions parallels, but should not be confused with, the quest for improved linguistic theory.

Satisfactory program designs have thus far been devised for the syntactic analysis algorithm and for all of the utility functions which support this process. The latter include the control function of the system, the functions which store and revise corpora and compile and revise syntactic data, the automatic programming functions which organize and compact syntactic data, and the functions which display syntactic analysis output. These basic programs, which will be the product of approximately 10 man-years of programming time, are expected to be completed before the end of 1962.

Linguistic Research. Investigations of German and English syntax have been underway for 3 years. These studies by the linguistic research group have a twofold purpose. They are, first of all, intended to verify the hypothesis by specializing it to the structures of particular languages. Secondly, they are the means by which syntactic data are prepared for the system.

Approximately 120,000 German and English construction rules have been coded as syntactic data. Of these, roughly 70,000 are dictionary rules based largely upon *Webster's New Collegiate Dictionary* and its counterpart, *Der Sprach Brockhaus.* The remaining rules are chiefly from constructions found in a corpus of Edward Ruechardt's book *Sichtbares und Unsichtbares Licht,* and its English translation *Light, Visible and Invisible.* The rules have been keypunched on IBM cards and are being recorded on magnetic tape.

Accumulating corpora is a relatively undemanding but necessary part of the work. About 1 million words of English text, and approximately the same amount of German, have been recorded on magnetic tape in preparation for experiments with automatic syntactic analysis.

References:

(1) Linguistics Research Center, University of Texas. *Eleventh Quarterly Progress Report,* for period 1 November 1961—31 January 1962.

(2) Linguistics Research Center, University of Texas. *Twelfth Quarterly Progress Report,* for period 1 February 1962—30 April 1962.

(3) Linguistics Research Center, University of Texas. *Thirteenth Quarterly Progress Report,* for period 1 May 1962—31 July 1962.

(4) Jonas, Ronald W. *System Control,* programming reference manual, Linguistics Research Center, University of Texas, 1962.

UNIVERSITY OF WASHINGTON 3.41

Department of Far Eastern and Slavic Languages and Literature, Seattle, Wash.
ERWIN REIFLER, *Project Director*

A Chinese-English Machine Translation Project was carried out during the period June 1960 to October 1961, and a final report of this work has been prepared (1).

Altogether, about 40,000 running characters of modern scientific texts from 23 fields of knowledge have been translated and studied with the purpose of determining the linguistic data of both the source and the target languages required for the compilation of an MT-operational bilingual lexicon. This involves the analysis of the formal and semantic structure of the two languages in order to ascertain (a) those semantic units which should be given membership in the bilingual lexicon, and (b) those logical procedures which should be given membership in the program of the automatic system to enable it to resolve problems of ambiguity and divergent structure.

Because of the inadequacy of Chinese and Sino-foreign dictionaries and because of the disagreement between the boundaries of the spoken and written forms in Chinese (2), much time had to be given to the determination of the Chinese semantic units and their English equivalents. As a consequence, of the 3,123 Chinese semantic units (containing 2,467 expressions belonging to the general language) extracted from the project's corpus and the 10,289 entries selected from Y. R. Chao's *Concise Dictionary of Spoken Chinese,* only 1,880 were processed before the termination of the project. These are included in the bilingual glossary of the report (1).

In lexicographical work, the principles and the approach developed in the previous Russian-English MT Project and outlined in (3) were applied. The aim has been on optimum of lexicography which permits a maximum of solutions of grammatical and nongrammatical problems by purely lexicographical means.

The final report includes: (a) a description of the work and the problems; (b) two samples of simulated machine translations elaborated on the basis of the bilingual glossary; (c) a 162-page bilingual glossary containing 1,880 Chinese semantic units: their frequencies, romanizations, and telecode numbers; their English equivalents in a specially developed system of abbreviations; and information about the area of knowledge to which they belong, and other information of importance for MT lexicography; and (d) a bibliography.

The re-indexing of all characters in the body of the Chung[1]-Hua[2]

Ta⁴-Tzu⁴-Tien³ (中華大字典) has been completed. Thus only the supplement section remains to be done; and for the purposes of publication, the new index will have to be shuffled into a number of sequences in consideration of different aspects of arrangement enhancing the speed of access, and it will have to be edited.

The project was supported by the National Science Foundation.

References:

(1) University of Washington. *The Chinese-English Machine Translation Project,* Vol. I and II, final report prepared for National Science Foundation under NSF G-13579, September 1962.

(2) Lu, C. W. "The Status of the Word in Chinese Linguistics," in *Beiträge zum Problem des Wortes im Chinesischen,* ed. by Paul Ratchnevsky, pp. 34-37 (Ostasiatische Forschungen, No. 1). Berlin: Akademie-Verlag, 1960.

(3) University of Washington, Department of Far Eastern and Slavic Languages and Literature and Department of Electrical Engineering. *Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English, USAF,* Phase I, Report No. RADC-TN-58-321 (ASTIA No. AD—148 992). Seattle: University of Washington Press, 1958.
Phase II, Report No. RADC-TR-60-11, Contract AF 30 (602) —1827. Seattle: University of Washington Press, 1960.

3.42 UNIWERSYTET WARSZAWSKI [WARSAW UNIVERSITY]

Research Center for Applied Linguistics, Department of General Linguistics, Warsaw, Poland
OLGIERD WOJTASIEWICZ, *Project Leader*

The immediate objective of the Center is to acquaint linguists and philologists of the University with recent advances in the fields of mathematical and applied linguistics through research and training. The next objective is to initiate research work which will, in all probability, concentrate on language models and formalization of syntax, mainly of Polish, Russian, English, and French.

The Center cooperates with the Department of Logic, Warsaw University (headed by Prof. Roman Suszko). Seminars held at the Center are attended by workers from various departments of Warsaw University and also some other institutions.

Reference:

(1) Wojtasiewicz, Olgierd. "Towards a General Theory of Sign Systems. I," *Studia Logica,* vol. 13, 1962, pp. 81-101; Part II to be published in *Studia Logica,* vol. 15.

VILNIAUS VALSTYBINIS V. KAPSUKO VARDO UNIVERSITETAS 3.43
[VILNIUS V. KAPSUKAS STATE UNIVERSITY]

Vilnius, Lithuania
J. KUBILIUS, V. STATULEVIČIUS, *and* M. ZACHARJAN

The compilation of the initial variant of the algorithm for machine translation of Russian into Lithuanian is completed. The text of an algebra manual served as a basis for investigation. The algorithm includes a 900-word (stem) Russian-Lithuanian dictionary, rules for Russian text analysis, and rules for Lithuanian text synthesis.

Work on the programming of the algorithm has begun. At present the choice of words in the dictionary has been programmed. Simultaneously, attempts are made to improve separate parts of the algorithm.

Statistical and informational characteristics of Lithuanian and Russian are still under investigation. The distribution of different letter combinations, the distribution of word-length, and the number of syllables in the words have been determined.

Members of the Institute of Physics and Mathematics of the Academy of Sciences are participating in this work.

VÝZKUMNÝ ÚSTAV MATEMATICKÝCH STROJŮ 3.44
[RESEARCH INSTITUTE FOR MATHEMATICAL MACHINES]

Loretánské nám. 3, Prague 1, Czechoslovakia
KVĚTA KORVASOVÁ, *Project Leader*

Research is concerned with the development of a machine program for automatic translation of English technical (especially electronics) texts into Czech. Some statistical studies of the Czech dictionary, which contains about 10,000 words, were performed for the purpose of formulating a method for fast searching of the grammatical information of the text word (4) (5). Some studies of an English dictionary of about 500 words were also made.

The algorithm for automatic translation, using a mechanical dictionary of about 500 words, was elaborated for short texts (about 50 clauses). Some of the results of analyses and investigations achieved at Charles University in Prague (see 3.20 and 3.21) were used for this project. The function of this algorithm will be tested on the EPOS 1 computer (1-3) (6), which is especially designed for data processing and which will begin operation in 1962.

The algorithm for automatic translation is designed in such a way that the extension of the algorithm for general texts does not require any substantial changes in the program. This extension will be made methodically, using the EPOS 1 computer.

References:

(1) Chlouba, V., J. Oblonský, and A. Svoboda, *Záv. zpráva VÚMS, No. 136: Magnetická bubnová paměť EPOS [Magnetic Drum Memory of EPOS]*.

(2) Chlouba, V., J. Oblonsky, and A. Svoboda. *Záv. zpráva VÚMS, No. 150: Magnetická pásková paměť EPOS [Magnetic Tape Memory of EPOS]*.

(3) Korvas, Z., J. Oblonský, and A. Svoboda. *Záv. zpráva VÚMS, No. 97: Operační jednotka a logická výstavba počítače EPOS [Arithmetic Unit and Logical Design of the Computer EPOS]*.

(4) Korvasová, K., and B. Palek. *Záv. zpráva VÚMS, No. 179: Automatické kódování českého slovníku [The Automatic Coding of the Czech Dictionary]*.

(5) Korvasová, K., and B. Palek. "The Problem of Czech Dictionary Searching Information Processing Machines." (In press)

(6) Oblonský, J., and A. Svoboda. *Záv. zpráva VÚMS, No. 122: Řadič a kód EPOS [Control Unit and Code of EPOS]*.

3.45 WASHINGTON STATE UNIVERSITY

Department of Sociology and Anthropology, Pullman, Wash.
ALLAN H. SMITH

The Japanese-English MT program continues, with no change reported since the previous statement [*Ed.*].

3.46 WAYNE STATE UNIVERSITY

Department of Slavic Languages, Detroit 2, Mich.
HARRY H. JOSSELSON, *Principal Investigator*

The main efforts of the machine translation group are directed toward several aspects of machine translation: (a) forming, correcting, improving, and expanding the Russian-English glossary based on mathematical text; (b) writing, checking out, and running syntactic routines (nominal, prepositional, and governing modifier blocking) on the IBM 7090 computer, using the interpretive system which has been developed to facilitate coding; (c) writing, checking out, and running several other programs on the IBM 1401 and IBM 650 computers with the RAMAC in order to mechanize stages preceding and following the syntactic routines; (d) analyzing the dative case in Russian in order to determine its functions and the related translations of the case endings; and (e) cooperating with other MT groups in the United States.

A Russian-English mathematical glossary on punched cards, containing coded information necessary for the automatic translation process

and English equivalents, has been constructed and checked out. The three syntactic blocking routines mentioned above, designed to group automatically elements of a sentence belonging together, are operational. A fourth, predicative blocking, is being coded and will be included into the automatic system for sentence analysis.

In order to mechanize as many stages of the translating process as possible, several machine programs have been developed. These include card-to-tape conversion programs and dictionary comparison, updating, and lookup routines, as well as input programs preliminary to syntactic routines. Investigations are also being carried out concerning the most convenient way of storing the dictionary on a disk file.

In the area of MT-oriented syntactic analysis, work in analyzing the Russian dative case is proceeding along computer-directed lines, similar to the study of the instrumental case carried out last year. In the immediate future, it is intended to investigate the functions of other classes of the Russian inflectional system. The findings dealing with the instrumental and dative cases will be mechanized as much as possible in terms of a series of "rules" or algorithms which could lend themselves to computer programming.

Work is continuing in assisting the MT group of the National Bureau of Standards (see 3.28) in coding the Russian-English Dictionary of the American Mathematical Society according to the coding format of this group. Additional information useful for MT, developed by the Wayne State University group, is also being supplied to the NBS group.

The project is supported by the U.S. Office of Naval Research.

References:

(1) Josselson, Harry H., and Arvid W. Jacobson. *Research in Machine Translation. Russian into English: Mathematical Text,* Annual Progress Report to ONR, July 31, 1959, Project No. NONR-2562(00). (ASTIA No. AD-226 834)

(2) Josselson, Harry H. "Report on Mechanical Translation Research of Wayne State University," in *Proceedings of the National Symposium on Machine Translation,* ed. by H. P. Edmundson. Englewood Cliffs, N. J.: Prentice-Hall Book Co., Inc., 1961.

(3) *Annual Progress Report to the Office of Naval Research,* Information Systems Branch, August 31, 1960.

(4) Janiotis, Amelia, and Harry H. Josselson. "Multiple Meaning in Machine Translation," in *Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis.* London: Her Majesty's Stationery Office. (In press)

(5) *Third Annual Progress Report to the Office of Naval Research,* Information Systems Branch, August 31, 1961.

(6) *Fourth Annual Progress Report to the Office of Naval Research,* Information Systems Branch, August 31, 1962.

3.47 YALE UNIVERSITY

425 Norton Parkway, New Haven, Conn.
FANG YU WANG

A project was initiated in 1960 for the study of grammatical signals in Chinese-English machine translation. Since the latter part of 1961 the grammatical phenomena in modern written Chinese have been observed using five articles on scientific and technological subjects. The following points have been emphasized: (a) the grammatical information which is carried by the "functors" of modern written Chinese; (b) the problems of classification of the parts of speech and the sentence structures in Chinese-English machine translation; (c) the problems of different kinds of ambiguity in modern written Chinese, especially the ambiguous structures; (d) the problems of segmentation of words; and (e) the possible solution of the problems of ambiguity and segmentation through grammatical signals.

This project is supported by International Business Machines Corporation.

3.48 (INDEPENDENT WORK)

32, rue d'Albon, Athis-Mons (S & O), France
MICHAEL CORBE

Work leading toward a paragraph-by-paragraph English-to-French mechanical syntax transfer continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Corbe, Michael. *Eléments d'une méthode de traduction automatique d'anglais en français.* Typewritten document distributed to a restricted group of French MT specialists (late summer 1960).

(2) Corbe, Michael, and R. Tabory. *Introduction à une syntaxe automatique de l'anglais (Méthode des fragments),* ETAAF/Sy. 1 Paper, Paris, February 15, 1961.

(3) Corbe, Michael, and R. Tabory. "Introduction to an Automatic English Syntax (by Fragmentation)," ETAAF/Sy. 1 Rev./Engl., [a more elaborate version of (2)], in *Proceedings of the First International Conference on Machine Translation of*

Languages and Applied Language Analysis. London: Her Majesty's Stationery Office. (In press)

(4) Corbe, M. Footnotes to the French translation of (3), published in *La Traduction Automatique,* vol. 2, no. 4, 1961, and vol. 3, no. 1, 1962. The Hague: Mouton and Co.

(5) Corbe, M. "La machine à traduire française aura bientôt trente ans," *Automatisme,* vol. 5, no. 3, 1960, ed. by Dunod, Paris.

(6) Corbe, M. "La linguistique automatique dans le monde en 1960," *Automatisme,* vol. 6, no. 6, 1961, ed. by Dunod, Paris.

(7) Corbe, M. "Traduction automatique. Analyse des langues appliquée Réflexions sur une conférence internationale," *Automatisme,* vol. 7, no. 3, 1962.

(8) Corbe, M. (m.c.) "La TA dans le monde," in *La Traduction Automatique (Bulletin de l'Association pour l'étude et le développement de la traduction automatique et de la linguistique appliquée),* vol. 1, 1959; vol. 2, 1961; and vol. 3, 1962; passim.

(9) Corbe, M. *Types of Separators To Be Used for Segmentation Purposes in English, and Certain Types of Resulting Segments.* (In preparation)

(10) Corbe, M. *Part-of-Speech and/or Functional Coding Manual for Syntactic Analysis of English Text.* (In preparation)

4
EQUIPMENT

4. EQUIPMENT

Together with the systems design work described in Section 2, equipment specially adapted for information handling is under development as described in this section. The two developments are parallel, the one affecting the other. Most projects in this section are concerned with the design of special-purpose equipment and integrated information systems. Particular emphasis is being placed on the development of storage devices, especially those using photographic media. These provide either hard-copy output or viewing of search results on a screen. General-purpose computers are being used in many information handling systems, but their development is not covered here.

The following summary calls attention in particular to new projects reported here for the first time, to certain progress made in previously reported work, to developments or projects which have been completed, and to closely related work which appears in other sections of this report.

Very little change has occurred in the photographic storage developments previously reported. The prototype model of the photochromic microimage Camera-Recorder has been completed and is reported to be operating satisfactorily by The National Cash Register Company (4.20). In the Massachusetts Institute of Technology's (4.14) research and development on rotating mirror photographic storage techniques, a high-speed core loader has been built for the IBM 1620 computer and a loader for the IBM 7090 computer is being planned. A small prototype reader, referred to as a computer input device, was completed, connected to an IBM 1620 computer, and successfully tried by the American Brake Shoe Company (4.1).

Development of a new document storage and retrieval system based on photographic storage techniques is reported here for the first time. RADIR (Random Access Document Indexing and Retrieval), an automated system for coding, indexing, storing, and retrieving information on 35 mm. microfilm, is in the final stage of development at The Hallicrafters Company (4.6).

The development of several document storage and retrieval systems is reported as being completed. These include the FileSearch system of FMA, Inc. (4.3), the Magnacard System of the Magnavox Research

Laboratories (4.13), and the Command Retrieval Information System of Information Retrieval Corporation (4.10).

For the MEDLARS system of the National Library of Medicine (see 2.86), Photon, Inc. (4.21) is engaged in the design, development, and fabrication of a computer printer that will serve as the output device for this system and will produce high-quality photographic composition on film at a rate of approximately 440 characters per second.

Two other new developments are noted. **Itek Corporation** (4.12) has constructed a special-purpose logic control device, called "Cross-filer," which controls the permutation of catalog card information on punched tape for the production of complete sets of catalog cards. At the **Centre National de la Recherche Scientifique** (4.2), the design of a peek-a-boo card reproducing machine intended for use in the publication of peek-a-boo card indexes has resulted in the building of a prototype model.

A state-of-the-art report on *Information Selection Systems Retrieving Replica Copies,* prepared by the Research Information Center and Advisory Service on Information Processing as National Bureau of Standards Technical Note 157, is now available from the U.S. Government Printing Office, Washington 25, D.C. (Price: $1.25).

All work on character and pattern recognition devices is included in Section 5 (Potentially Related Research) along with the fundamental studies of techniques in that area.

4.1 AMERICAN BRAKE SHOE COMPANY

Raymond Atchley Division, 2231 S. Barrington Avenue, Los Angeles 64, Calif.
CARL G. BLANYER

Development of the Photomemory, a high-density, sequential access, rapid readout photographic storage system, continues.

A small prototype reader with an 80,000-bit capacity, referred to as the computer input device, has been completed and connected to an IBM 1620 computer. Eight-bit words are read out in parallel at a rate permitting the data plate to be scanned completely in 3 seconds. A trial with the IBM 1620 was valid and generally successful. The photomemory operated normally and placed the contents of the data plate into the memory.

Partial support for the project is provided by the U.S. Office of Naval Research.

References:

(1) St. George, Emery, Jr., Carl E. Nielsen, Jr., and Eugene P. Johnson, Jr. *The Development of a Photomemory.* Annual Sum-

mary Report No. 1, prepared for ONR under Contract Nonr-2668 (00). Hydel, Inc., Waltham 54, Mass.

(2) Nielsen, Carl E., Jr. *The Development of a Photomemory.* Annual Summary Report No. 2, prepared for ONR under Contract Nonr-2668 (00). Hydel, Inc., Waltham 54, Mass.

(3) Nielsen, Carl E., Jr., and Carl G. Blanyer. *The Development of a Photomemory.* Annual Summary Report No. 3, prepared for ONR under Contract Nonr-2668 (00). Raymond Atchley Division, American Brake Shoe Co., Los Angeles 64, Calif.

4.2

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Centre d'Analyse Documentaire pour l'Archéologie (CADA), 23 rue du Maroc, Paris 19e, France

J.-C. GARDIN

A reproducing machine for peek-a-boo cards has been designed for publication purposes. A prototype is now available, built by the Société Phototechnique, Paris, for the duplication of Selecto cards (IBM card format, capacity of 5,000 punched positions); it is used by the Centre for the reproduction of two different card indexes completed in 1961 (1)(2). More effort needs to be devoted to improving the design of such a machine, which has three main defects: (a) The composition of the punch matrix is manual; different methods of automatic composition have been experimented with and have proved impractical; (b) though the speed is satisfactory (around 50 cards per minute), much time is wasted in controlling visually the obtained duplicates because of frequent defects in the punching; and (c) the printing block (for reproducing the title of each card while it is being punched) only allows for rubber characters, which provide readable but not elegant typography.

References:

(1) Christophe, J. *Catalogue des Outils orientaux de l'Age du Bronze, sur cartes perforées.* (In press)

(2) Allard, M., *S.J.*, M. Elziere, J.-C. Gardin, and F. Hours, *S.J. Analyse Conceptuelle due Coran, sur cartes perforées.* (In press)

FMA, INC. 4.3

142 Nevada Street, El Segundo, Calif.

WILLIAM R. ARSENAULT

Development of the FileSearch system has been completed. The system is now fully operational at several installations throughout the country. The system consists of (a) an electric typewriter which

punches index information about documents on edge-punched cards in binary form; (b) a recording unit which photographs simultaneously each edge-punched card and the document it describes, and produces a reel of 35 mm. coded film; and (c) a retrieval unit which electronically scans the coded film at the rate of more than 100 pages per second and locates the requested document or documents. The document can be retrieved as an image on a screen (for browsing), a hard-copy print, or a duplicate film copy (for file expansion).

References:

(1) "FMA FileSearch," *Reproduction Methods,* July 1962, p. 38.

(2) *FMA FileSearch.* FMA brochure.

(3) "FileSearch Retrieves Information," *Management and Business Automation,* April 1961, p. 40.

(4) "Automated File Storage and Retrieval," *The Office,* May 1961, p. 140.

4.4 GENERAL ELECTRIC COMPANY

Computer Department, Phoenix, Ariz.

L. W. GOOSTREE, JR., *Manager-Marketing*

Work on Project SITE (Search Information Tape Equipment), which is especially adaptable to manipulation of data recorded in natural-language representation, continues, with no change reported since the previous statement [*Ed.*].

4.5 GENERAL ELECTRIC RESEARCH LABORATORY

P. O. Box 1088, Schenectady, N.Y.

W. E. GLENN

The thermoplastic recording technique is being developed for use as a storage medium for recording documents, television, digital information, analog signals, etc. In the field of documentation the film can record at a high rate and produce an optically visible image of very high resolution. The film requires no chemical development and can be erased and reused. This process is of particular interest for high-capacity document files.

Recent progress includes improved document recordings, successful operation of a portable recorder using ¼-inch tape, successful recording of signals well in excess of 100 mc, improvement of electron guns, and improvement of tape quality.

Reference:

(1) Glenn, William E., and J. Edmond Wolfe. "Thermoplastic

Recording," *International Science and Technology,* June 1962, pp. 28-35.

THE HALLICRAFTERS COMPANY 4.6

4401 W. Fifth Avenue, Chicago 24, Illinois
ROBERT F. HALLIGAN, *President*

RADIR (Random Access Document Indexing and Retrieval), an automated system for rapidly coding, indexing, storing, and retrieving information on 35 mm. microfilm, is in the final stage of development.

RADIR is composed of (a) a coder keyboard used in conjunction with a standard microfilm camera, (b) an index unit which receives the document information at random and dispenses the information on request, and (c) a retrieval unit which searches the microfilm at 100 frames per second and displays the requested document. If a copy of the document is desired, the retrieval unit develops in less than 10 seconds a 35 mm. reproduction which will print out on an electrostatic photocopier. RADIR is a compact, simple, efficient, low-cost system of information storage and retrieval, which minimizes storage space problems, eliminates the need for alphabetical or numerical filing, and significantly reduces document retrieval time.

HERNER AND COMPANY 4.7

1401 K Street, NW., Washington 5, D. C.
HENRY T. HEATWOLE *and* SAUL HERNER

Final testing and troubleshooting of a small electronic information data searching and correlating device, called the H-44, are underway prior to putting a production prototype into routine operation for further tests under actual operating conditions. The device, designed for clerical operation, handles searches of index records in which subjects may be represented by 3-, 4-, and 5-digit numerical codes. It will search and correlate up to four such codes at a time, analyzing for logical sums, logical products, or negations, or combinations of these.

References:

(1) Herner, S. "Heatwole Associates Magnetic Tape Searcher and Correlator," presented at the panel on Information Retrieval Systems for Small and Medium-Sized Libraries, Special Libraries Association meeting, San Francisco, May 28-June 1, 1961 (mimeo.).

(2) Heatwole Associates. *Specialized Information and Data Processing Equipment by Heatwole Associates,* 1961 (pamphlet).

(3) Heatwole Associates. *H-44 Operator's Manual,* 1961 (pamphlet).

4.8 HOUSTON FEARLESS CORPORATION

Westwood Division, 11801 West Olympic Boulevard, Los Angeles 64, Calif.
H. R. LUXENBERG *and* R. J. LINTELL

Fabrication of an experimental model of an Automatic Unit Record Storage and Retrieval Device continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the U. S. Air Force (Rome Air Development Center).

4.9 INDEX & RETRIEVAL SYSTEMS INC.

Woodstock, Vt.
A. B. KYLE, *Systems Manager*

The evaluation of new techniques for modifying addressing machine equipment to be used in index production continues, with no change reported since the previous statement [*Ed.*].

4.10 INFORMATION RETRIEVAL CORPORATION

One Farragut Square South, Washington 6, D. C.
P. W. LARSEN, *Principal Investigator*

The Command Retrieval Information System (CRIS), a microimage scroll storage system with automatic document-image retrieval and reproduction, is currently in production.

Prototypes were constructed and tested, resulting in many improvements to the system. The basic single scroll capacity, 500,000 pages, remains unchanged as does the 20-second-average random retrieval time. All logic and control storage is now done using solid state components to improve reliability and reduce costs. The entry keyboard was changed to a parallel entry system to provide easier operation and allow complete control of scanning in any direction. The viewing screen was enlarged and the optical system improved. The film card output system was improved to provide use of standard and special aperture cards for image printout.

Various operational systems were derived for scientific and general research libraries, newspaper applications, and commercial information services. These systems provide a total information system capability, recognizing, first, the limited ability of the user to adequately define his search question, and second, the economic factors involved in supplying a user with a mass of information which might be pertinent to his problem. Other factors, such as the total time required, the reliability of information support, and the physical location of the user from the information system, are recognized in these systems.

The adaptation of various indexing methods and machines and of

various reproduction equipment to the CRIS input and output is continuing.

References:

(1) Larsen, Paul W. "CRIS . . . The Command Retrieval Information System," in *Advances in Microfilm Technology* (proceedings of the National Microfilm Association Eleventh Annual Meeting, April 1962), ed. by Vernon Tate. Washington, D. C.: Spartan Books, October 1962.

(2) Bushor, William E. "Information Storage/Retrieval," *Electronics,* vol. 35, no. 26, June 29, 1962, pp. 50-56.

INTECTRON, INCORPORATED 4.11

2300 Washington Street, Newton Lower Falls 62, Mass.

J. KENT BOWKER

The investigation of high-reduction microphotography continues. The data on the inherent resolution of photographic materials, which deal with optical diffusion and the aperture response of diffusing media, are being prepared for publication. Further work on the physical aspects of development and optical elements, as far as these affect image quality, is planned.

The study is sponsored by the Council on Library Resources, Inc.

ITEK CORPORATION 4.12

Information Sciences Laboratory, Lexington 73, Mass.

J. W. KUIPERS, *Manager*

A low-cost, special-purpose logic control device has been constructed. This device, called the "Crossfiler," is now undergoing operational tests. The Crossfiler reduces the clerical effort required to produce sets of catalog cards for libraries or document centers. The device accepts input in the form of punched paper tape and has two alternative output modes: (a) direct activation of an electric typewriter through cable connection, or (b) high-speed production of secondary punched paper tape for off-line activation of punched-tape typewriters. The input punched tape contains descriptive information on a series of books or documents. The input format must be designed so that coded boundaries between input data categories (call number, author, title, series, subject terms, etc.) may be recognized by rules of sequence. The device recognizes and then permutes the input data to produce typing output (or the punched tape equivalent) consisting of a full set of catalog cards for each book; in a set of cards the descriptive text is identical, but each card contains a different heading or tracing derived automatically from the input text. After completing

the permuting operation on one book or document, the device proceeds automatically to the next book or document. Operation is fully automatic. No splicing of loops from the input punched tape is involved.

Although not specifically provided for in the model now undergoing test, the device is capable of other data-manipulating functions, such as automatic abridgment or automatic rearrangement of records on the basis of categories of data contained in the records.

The device was constructed as part of a research and development project supported by the U. S. Air Force.

4.13 MAGNAVOX RESEARCH LABORATORIES

2829 Maricopa Street, Torrance, Calif.

ROLF E. WESTGARD, *Data Systems Product Manager*

The Magnacard System is now being offered for sale or lease. Application studies have been made in various areas such as law enforcement, vehicle and driver registration, inventory management, title insurance, and other billion-or-more-character storage problems. Complete programming packages are available for use with a Magnavox-provided computer or can be easily made available for almost any general-purpose digital computer. The system offers unusual EDP speed, flexibility, and storage capacity in the low-to-medium price range.

A large rapid-access file provides from 680,400,000 alphabetic characters to 1,021,000,000 numeric characters. Two such files can be associated with one Magnacard Transport Unit associated with one input-output channel of a computer. Data are transferred to and from these files at 90,000 alphanumeric characters per second.

The Magnavue card stores microfilmed images along with magnetic recorded information and can be used on the same card-handling Magnacard equipment. A display station is added for viewing or hard copy. Breadboard equipment has proven the feasibility of this means of graphic storage.

The Magnascriber prototype has been built. This device is the size of an office typewriter. It utilizes a standard card-punch keyboard and records statically on the Magnacard or Magnavue card. Verification is made by means of a punched paper tape.

4.14 MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Engineering Projects Laboratory, Data Processing Systems Division, Cambridge, Mass.

D. M. BAUMANN, *Project Leader*

Research and development continues on rotating mirror photographic storage techniques. Feasibility of the Photomemory has been demonstrated. A high-speed core loader has been built for the IBM 1620

computer, and a loader for the IBM 7090 computer is being planned. These loaders operate at a speed limited only by the storage cycle time of the computer.

The project is sponsored by the U. S. Office of Naval Research.

References:

(1) Baumann, D. M. "Rotating Mirror Photographic Storage Systems," in *Large-Capacity Memory Techniques for Computing Systems,* ed. by M. C. Yovits, pp. 373-383. New York: The Macmillan Company, 1962.

(2) Arndt, Joseph H., Jr. *Prototype Development of a Semi-permanent Photographic Data Storage and Retrieval System.* Master's Thesis. Cambridge, Mass: Massachusetts Institute of Technology, May 1962.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY 4.15

Microreproduction Laboratory, Libraries, Cambridge 39, Mass.

PETER SCOTT

Work on the development and construction of a model of a relatively inexpensive microfilm finder-reader system which is compatible with the general design of commercially available microfilm readers continues, with no change reported since the previous statement [*Ed.*]. The project, which is being carried out jointly with Edgerton, Germeshausen & Grier, Inc., of Boston, is sponsored by the Council on Library Resources, Inc.

MEMISTOR CORPORATION 4.16

270 Polaris Avenue, Mountain View, Calif.

BERNARD WIDROW, *President*

Construction of a research machine containing 1,500 memistorized adaptive weights continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Widrow, B., and M. E. Hoff. "Adaptive Switching Circuits," in *1960 IRE WESCON Convention Record, Part 4.*

(2) Widrow, B. *An Adaptive "Adaline" Neuron Using Chemical "Memistors,"* Stanford Solid-State Electronics Laboratory Technical Report No. 1553-2, Stanford University, October 1960.

4.17 MERGENTHALER LINOTYPE COMPANY

Research and Engineering Division, Graphic Systems Engineering Department, 29 Ryerson Street, Brooklyn 5, N. Y.
J. KONIGSBERG, *Project Leader*

Study and investigation are concerned with the design of a high-speed Lexical-Graphical Composer-Printer System which will automatically produce as its output a high-quality master that can be used for either direct printing or for transfer to a printing plate. The output of the system will contain both high-quality automatically justified lexical and graphical matter, such as line art and halftone cuts, with proper position correlation maintained in the output between the two kinds of matter. A printout speed of 600 characters per second is a design objective. The end item of the study will be a system design specification embodying the results of the study with emphasis toward employing proven principles, methods, and existing proven equipments such that a working system can be produced in a reasonable amount of time.

Past activities have led to a determination of mediums for the system input and output information and definition of a group of formats designed to meet the requirements for the presentation of a wide variety of printed matter, such as straight text, index, and tabular matter. Current and future activities involve an investigation of the problem of automatic justification and automatic hyphenation, an inquiry into the requirements for the automated setting of mathematical text, definition of the logic implied by the chosen output formats, and the definition of system input information in accordance with the defined logic.

The work is sponsored by the U. S. Air Force (Rome Air Development Center). Duration of the study program is 1 year, beginning January 1, 1962.

Reference:

(1) Buck, C. P., R. F. Pray, and G. W. Walsh. *Investigation and Study of Graphic Semantic Composing Techniques,* RADC TR 61-58, Final Report, Syracuse University Research Institute, College of Engineering, June 1961. (ASTIA No. AD-258 974)

4.18 NATIONAL BUREAU OF STANDARDS

Data Processing Systems Division, Washington, D. C.
T. C. BAGG *and* J. L. PIKE

A continuing project has as its objective the development of an improved design of the Rapid Selector, a device for storing and retrieving selected document images from reels of coded microfilm at high speed. A new model is being completed for use by the Bureau of Ships, while

the prototype equipment at NBS is being set up for demonstrating other potential applications. Improvements in the input and output facilities and means for greater manipulation of the coded data are underway. A series of experiments which will make use of a variety of document abstracting and encoding schemes has been planned, and preliminary work has begun. These comparative investigations are being conducted by the Research Information Center and Advisory Service on Information Processing (see 2.84), using the Rapid Selector system.

A detailed technical report on the present state of development of the system is in final review prior to publication.

The project is sponsored by the Bureau of Ships, Department of the Navy.

Reference:

(1) Bagg, T. C., and J. L. Pike. "The Rapid Selector and Other NBS Document Retrieval Studies," in *Advances in Microfilm Technology* (proceedings of the National Microfilm Association Eleventh Annual Meeting, April 1962), ed. by Vernon Tate. Washington, D. C.: Spartan Books, October 1962.

NATIONAL BUREAU OF STANDARDS 4.19

Basic Instrumentation Section, Washington 25, D. C.
JOSHUA STERN, *Chief*

Work continues on the Microcite, a document (abstract) retrieval machine for which the input consists of a stack of peek-a-boo cards selected by the searcher in accordance with both his request and the coding vocabulary. The output to the searcher is either a projected full-size image of one of three formats (3 by 5, 5 by 8, or 8 by 10 inches) or hard-copy readout of the same material. It is planned to incorporate a mechanism which will allow the machine to generate its own microfilm image arrays (matrices) in a two-step reduction. Final reduction will be approximately 64:1. Following a period of experimentation, a new matrix magazine (storage bin plus loading, unloading, and selected means) has been designed and is being constructed. It is anticipated that work underway will extend the capacity of the Microcite to accommodate up to approximately 1 million documents.

An experimental version of a new punch has been constructed and is in use for generating the third peek-a-boo primary index term card deck. The most significant improvement is the incorporation of a shock-sensing element and associated transistor circuitry to drive the punch solenoid in place of the displacement-actuated microswitch formerly used. The table of the new punch is made of chrome-plated steel instead of aluminum. The steel table does not dirty the cards

as they are punched and retains a flat surface with greater stability. Some experiments which have been carried out indicate that punching without a true die may be feasible. If this proves to be the case, it would be possible to move part of the somewhat awkward superstructure mounted above the punch table to a position under the table.

4.20 THE NATIONAL CASH REGISTER COMPANY

Electronics Division, 1401 East El Segundo Boulevard, Hawthorne, Calif.
WILBUR C. MYERS, *Director of Research*

Research and development continues on the NCR †photochromic microimage (PCMI) process which provides a capability for very high-density document storage. Linear reductions of 200:1, representing an area reduction of 40,000:1 are entirely feasible for normal business documents.

The most important characteristics which make PCMI technology unique in the field of microform storage and dissemination are (a) Photochromic films provide very high resolution with no grain; (b) photochromic films permit storage of images containing a wide contrast of gray scale because they are inherently low-gamma and grain-free; (c) photochromic films provide immediate visibility of the image upon exposure. No development process is required; (d) photochromic films provide both erasing and rewriting functions. This permits the processes of editing, updating, inspection, and error correction to be incorporated into systems; (e) the PCMI process incorporates the ability to effect a bulk-transfer readout of microimages at the 200:1 reduction level by contact printing; (f) use of high-resolution silver halide films provides both permanency for the storage of microimages and economical dissemination of duplicates; and (g) the very high density of 200:1 microimages offers the possibility of using some form of "manual retrieval" techniques for many applications. This eliminates the normal requirement in systems of this size for expensive and complex random-access hardware.

A prototype model of the PCMI Camera-Recorder has been completed and is operating satisfactorily. This unit was designed to produce 3- by 5-inch photographic master negative film transparencies containing 2,625 images at a linear reduction ratio of 200:1. This master microimage negative is then used to reproduce (i.e., publish) any number of desired 3- by 5-inch positive film transparencies. A laboratory model of a microimage viewer has been completed.

Among the applications areas where PCMI techniques are expected

† *See* Glossary.

to be important are information storage and retrieval, microform publishing, on-demand printing, and libraries.

References:

(1) Tauber, A. S., and W. C. Myers. "Photochromic Microimages—A Key to Practical Microdocument Storage and Dissemination," in *Advances in Microfilm Technology* (proceedings of the National Microfilm Association Eleventh Annual Meeting, April 1962), ed. by Vernon Tate. Washington, D. C.: Spartan Books, October 1962.

(2) *Photochromic Micro-Images,* NCR brochure.

PHOTON, INC. 4.21

355 Middlesex Avenue, Wilmington, Mass.

ROBERT G. CROCKETT, *Project Manager*

Current efforts involve the design, development, and fabrication of a computer printer for the MEDLARS system of the National Library of Medicine (see 2.86). The computer printer will serve as the output device for this system and produce high-quality photographic composition on film at a printing rate of approximately 440 characters per second.

The computer printer will operate from the magnetic tapes which form the output of the computer storage and retrieval system, via a magnetic tape handler. The coded information, which is sent from the tape handler to the computer printer, consists of character, space, and control codes and completely defines the format for each line of composition. This information is received and accumulated in a solid state control section of the computer printer. At the end of each line the tape reading stops and the stored information is processed and indexed for use. Width assignments are made for each character or space in storage, and the particular location of each letter within a line is placed in temporary storage. Information from the memory circuits is then fed into the photographic section of the computer printer, and the codes which are read out are used to control the flashing of characters.

The photographic section consists of glass matrices which contain several hundred characters (uppercase, lowercase, numbers, punctuation, symbols, etc.). The matrices are located directly in front of a multiple light source, and any single character or group of characters can thereby be illuminated simultaneously. An optical system is used to project each character onto film on a common line so that as characters are selected and successively illuminated at high speeds, they are photographed onto sensitive film or paper. The film is contained in a magazine and consists of a long roll which is indexed as soon as

the photography of each line is complete. The output film may then be developed by a rapid continuous process, and is available for offset use through any of the conventional methods.

This development project is being performed under contract for the General Electric Company, Defense Systems Operation, Bethesda, Md., which is prime contractor for the MEDLARS project sponsored by the National Library of Medicine.

4.22 UNIVERSITY OF CAMBRIDGE

Mathematical Laboratory, Corn Exchange Street, Cambridge, England
M. V. WILKES, *Director*

Work on a monotype reader designed for on-line connection to a digital computer continues, with no change reported since the previous statement [*Ed.*].

4.23 WALTER REED ARMY MEDICAL CENTER

Department of Radiobiology, Walter Reed Army Institute of Research, Washington 12, D. C.
ALFRED FELDMAN *and* DAVID P. JACOBUS

A typewriter, by means of which chemical structures can be typed and simultaneously encoded, has been developed.

A computer program will be written to search the output from this typewriter for any structure or any desired fragment of structures (generic searching). The biological portions of the information are handled conventionally and are programmed for the RPC 4000 computer.

References:

(1) Feldman, A., D. B. Holland, and D. P. Jacobus. "The Automatic Encoding of Chemical Structures," presented before the 141st American Chemical Society meeting, Washington, D. C., March 28, 1962.

(2) Feldman, Alfred, and Harold L. Manceaux. "An Improvement in The Printing of Chemical Structures, Which Results in Their Complete Computer Codes," presented at the annual meeting of the American Documentation Institute, Hollywood-by-the-Sea, Fla., December 14, 1962.

(3) Meyer, E., and K. Wenke. "Ein System zur topologischen Verschlüsselung organisch-chemischer Strukturformeln für die mechanisierte Dokumentation" ["A System of Topological Coding of Organic-Chemical Structural Formulas for Mechanical Documentation"], *Nachrichten für Dokumentation*, vol. 13, 1962, p. 13.

(4) Waldo, W. H., and M. DeBacker. "Printing Chemical Structures Electronically: Encoded Compounds Searched Generically with IBM-702," in *Proceedings of the International Conference on Scientific Information.* Washington, D. C.: National Academy of Sciences-National Research Council, 1959.

5
POTENTIALLY RELATED RESEARCH

5. POTENTIALLY RELATED RESEARCH

Research reported here for the most part is not directly connected with scientific documentation but is included because its results may influence scientific documentation work in the future. The purpose of the section is to foster communication and exchange of ideas between workers in documentation research and those in potentially related areas of research.

The coverage in this issue is broader and more extensive than in previous issues. The statements in this section are grouped under five headings: character and pattern recognition, speech analysis and synthesis, linguistic and lexicographic research, artificial intelligence, and psychological studies.

5.1 *CHARACTER AND PATTERN RECOGNITION*

The development of automatic techniques for direct transcription of data from typed or printed form to machine-usable form holds considerable promise for application in mechanized literature search and in mechanical translation. In addition, availability of natural-language texts in machine-usable form is becoming increasingly important for further progress in research in linguistics which, in turn, should contribute to improvements in scientific documentation.

A distinction may be made between "character recognition" and "pattern recognition." Character recognition here is restricted to letters and conventional symbols, while pattern recognition applies to arbitrary patterns, some of which can be visually perceived, such as photographs, charts, and real-world objects, and others, which require nonvisual perception, such as speech. Characters, of course, are specific examples of patterns. The work on recognition of speech as a type of nonvisible pattern is considered in the separate subsection 5.2 (Speech Analysis and Synthesis). The research in these fields is interrelated, as is the research described in subsection 5.4 (Artificial Intelligence). See, for example, Cornell Aeronautical Laboratory, Inc. (5.4.7), National Bureau of Standards (5.4.14), and Philco Corporation (5.4.16 and 5.5.17).

The following summary calls attention in particular to new work reported here for the first time and to projects and studies that have been completed

CHARACTER RECOGNITION

A character recognition system based on weighted area scanning is being investigated at **Massachusetts Institute of Technology** (5.1.21).

In Japan the **Radio Research Laboratories** (5.1.34) has research in progress on printed or handwritten characters.

Burroughs Control Corporation (5.1.4) has developed and made delivery on a Typed Page Reader which reads standard elite typed documents and converts the information to teletype code.

The New York University (5.1.25) study of Russian printing styles, completed in 1959, is mentioned here only because the availability of the final report on the study had not been cited in previous issues of *Current Research and Development in Scientific Documentation.*

PATTERN RECOGNITION

In the approach being taken at **Litton Systems** (5.1.19), pattern recognition is treated as a problem in statistical hypothesis testing in which each member of a class of things exhibiting a common pattern is represented by a vector in a multidimensional vector space.

Lockheed Missiles & Space Company (5.1.20) research involves investigation of the decision-making portion of the pattern recognition problem.

A **Philco Corporation** (5.1.28) research program to seek optimum prenormalization techniques to facilitate the recognition of gray-scale aerial photographic imagery utilizes adaptive memories with current pattern recognition schemes.

Raytheon Company (5.1.35) is conducting research on the generation and analysis of signals resulting from scans of geometrical solids.

The work of the Institute of Automatics and Telemechanics, Academy of Sciences U.S.S.R. (5.1.12), previously reported in subsection 5.4, is now included in this subsection.

5.1.1 ARTHUR D. LITTLE, INC.

35 Acorn Park, Cambridge 40, Mass.
VINCENT E. GIULIANO, PAUL E. JONES, *and*
RICHARD F. MEYER

Investigations concerned with the use of statistical correlation techniques for the recognition of printed characters continue, with no change reported since the previous statement [*Ed.*].

References:

(1) Giuliano, V. E., et al. "Automatic Pattern Recognition by a Gestalt Method," *Information and Control,* vol. 4, no. 4, December 1961.

(2) Meyer, R. F., V. E. Giuliano, and P. E. Jones. "Analytic

Approximation and Translational Invariance in Character Recognition," in *Optical Character Recognition* (proceedings of the ONR-NBS Symposium on Optical Character Recognition, Washington, D. C., January 1962), ed. by George L. Fischer, Jr., et al., pp. 181-195. Washington, D. C.: Spartan Books, 1962.

(3) Giuliano, V. E., et al. "A Gestalt Method for Pattern Recognition," presented at the Third International Congress on Cybernetics, Namur, Belgium, September 11-15, 1961; published in *Revista Methodos*, Milan, Italy, no. 49-50, vol. 13, 1961.

(4) Giuliano, V. E. "A Gestalt Method of Character Recognition," *ADL Review*, June 1962.

BELL TELEPHONE LABORATORIES, INC. 5.1.2

Murray Hill, N. J.
L. D. HARMON

Work on computer reading of real-time cursive script continues, with no change reported since the previous statement [*Ed.*].

BUDD ELECTRONICS 5.1.3

A Division of the Budd Company, Inc., 43-22 Queens Street, Long Island City 1, N. Y.
AZRIEL ROSENFELD, *Manager of Research*

Studies of concepts and techniques for the analysis of pictorial data are aimed at developing rules for organizing (or "structuring") complex pictures based on the recognition of similarities and dissimilarities of visual texture and on the identification of "texture contours" (boundaries at which textural changes occur). Automatic techniques for texture measurement and boundary location are being developed. In conjunction with this program, studies in texture scaling and picture structuring are being conducted by Dr. Jacob Beck of the Department of Psychology, Harvard University.

The studies are supported in part by the U.S. Air Force Office of Scientific Research.

References:

(1) Beck, Jacob, and Azriel Rosenfeld. "Conceptual Model for Pictorial Data Analysis," Paper FB–17, presented at the 1962 spring meeting of the Optical Society of America.

(2) Rosenfeld, Azriel. "An Approach tc Automatic Photointerpretation," in *1962 Conference Proceedings, National Convention on Military Electronics*, pp. 97-102.

(3) Rosenfeld, Azriel. "Automatic Recognition Techniques Ap-

plicable to High-Information Pictorial Inputs," to appear in *1962 International Convention Record, Institute of Radio Engineers.*

(4) Technical and Status Reports on Contract AF 49(638)-1143, *Research on Pictorial Data Analysis Concepts and Techniques.* (To appear during 1962-63)

5.1.4 BURROUGHS CONTROL CORPORATION

Control Instrument Division, 67 35th Street, Brooklyn 32, N. Y.
ROBERT J. BIBBERO, *Manager of Development Engineering*

A long-range program is underway to apply the results of the corporation's theoretical studies and research on character recognition techniques to military problems of pattern recognition.

To date, this work has been concerned primarily with the development of a feasibility model of a Typed Page to Teletype Code reader for the Signal Corps, Fort Monmouth, N. J. Among the various features included in this equipment are automatic document handling, automatic line registration, line skew following, recognition of uppercase standard elite type, automatic insertion of teletype operational symbols, keyboard insertion of "can't-read" characters, reading rate of 75 characters per second, and output rates of 0-75 characters per second.

The Typed Page Reader reads standard elite typed documents in which the typed material is double spaced. The machine is capable of reading uppercase letters, numerals, and punctuation marks; a certain degree of mutilation and misalignment of characters may be tolerated. Typed characters may be tilted as much as 10° due to typewriter misalignment without causing degradation of reading accuracy. Material is read at a rate of 75 characters per second. The typed characters are converted into teletype code for transmission over a teletype line or to a tape perforator.

The Typed Page Reader has been delivered to the U. S. Army Signal Research and Development Agency, Fort Monmouth, N. J., and the feasibility of this type of equipment has since been demonstrated.

The Burroughs Corporation and the Control Instrument Division are engaged in research programs for the future development of character recognition equipment. These programs are primarily aimed at increasing the reliability of character recognition systems.

Currently, work in more difficult character recognition problems (e.g., handwritten characters) is being pursued by other Corporate activities on a research and theoretical study basis utilizing advanced statistical decision concepts and methods (see 5.1.5). Results of this work promise to improve performance of hardware similar to the Typed Page Reader by one or more orders of magnitude.

Plans are being prepared to improve the Typed Page Reader by application of the newer theoretical concepts. In addition it is planned to apply these concepts to pattern recognition problems other than those of printed characters.

The Typed Page Reader was developed under contract to the U. S. Army Signal Corps.

References:

(1) Mintz, L., and K. Brooks. "Burroughs Control Corporation Typed Page Reader," in *Optical Character Recognition* (proceedings of the ONR-NBS Symposium on Optical Character Recognition, Washington, D. C., January 1962), ed. by George L. Fischer, Jr., et al., pp. 85-92. Washington, D. C.: Spartan Books, 1962.

BURROUGHS CORPORATION 5.1.5

Research Center, Burroughs Laboratories, Box 843, Paoli, Pa.
C. K. CHOW *and* R. P. COLEMAN

The purpose of research underway is to study and to develop techniques in character and pattern recognition.

General areas under investigation are (a) quantitative representation of patterns, (b) recognition systems and their representation, and (c) methods of simplification and approximation. Recognition is considered as a statistical decision problem, and optimum systems are based on *a priori* and conditional probabilities, or their equivalents, of patterns. The structures of recognition networks and formulas for recognition weights are derived systematically from the functional form of probability distributions. Both linear and nonlinear networks are under investigation; nonlinearity is essential in order to accommodate interdependence among signals. Simulation of recognition methods is performed on a Burroughs 220 digital computer. The program consists of two main operations: estimation of statistical parameters from sample characters and recognition using these estimated values. Alphanumeric characters, both machine-printed and handprinted, are used as input.

A special-purpose scanner has been constructed to supply input data to the computer simulation program. Printing samples are scanned by a flying spot with possible maximum of 60 quantization levels and then are encoded into paper tape. The data therefore contain a reasonably faithful picture of the original samples; gray information is retained. Conversion to binary form and other signal preprocessing take place in the computer as part of the simulation program.

References:

(1) Chow, C. K. "A Recognition Method Using Neighbor Depend-

ence," to be published in *IRE Transactions on Electronic Computers.*

(2) Coleman, R. P. "Orthogonal Functions for Logical Design of Switching Circuits," *IRE Transactions on Electronic Computers,* vol. EC–10, no. 3, September 1961, pp. 379–383.

5.1.6 CASE INSTITUTE OF TECHNOLOGY

Systems Research Center, University Circle, Cleveland 6, Ohio

R. B. BANERJI, *Project Leader*

The project's goal is to develop a theory of object recognition. The method for describing a predefined object and for using that description in a computer algorithm for the recognition of a predefined object has been developed (2). There have been some recent developments in methods for manipulating the descriptions in the computer memory for generating the description of new objects (1). The intuitive ideas involved in the terms "concept," "property," "relevance," and "generalization" have been formalized.

In this approach, all objects are considered as subsets of a universe of objects, and all descriptions are obtained from the descriptions of objects having the simplest possible predefined descriptions. These "simplest objects" at present define "properties" of more complex objects. However, if the object to be described is extremely complex, it may be useful to define properties in terms of other complex objects rather than in terms of the simplest of objects. The objects which should be chosen to constitute new properties depend on the class of objects to be recognized by means of these. The development of a method for isolating such new properties forms the primary goal of the project at present. An important secondary goal of the project is to develop ways for making the descriptions more concise. The presently known methods for doing this are inadequate. They are set theoretical and hence have analogs in the propositional calculus. At present, ways and means are being considered for building the predicate calculus into the system.

The programs that are being developed will be simulated and tested in IPL–V language on the Burroughs 220 computer. An interpreter for this purpose is being developed.

References:

(1) Banerji, Ranan B. "The Description List of Concepts," *Communications of the ACM,* vol. 5, no. 8, August 1962, pp. 426-432.

(2) Banerji, Ranan B. "An Information Processing Language for Object Recognition," *General Systems,* vol. 5, 1960, pp. 117-128.

CHRYSLER CORPORATION 5.1.7

Advanced Projects Organization, P. O. Box 1827, Detroit 31, Mich.
D. N. BUELL

A concept for an optical pattern recognition device has been developed, and a simplified experimental system is under test. An IBM 7090 computer is used to simulate the computer logic. The concept provides a lens focusing an image on an array of photosensors and a rotating mirror which causes image motion over the sensors. The computing elements provide a digital †transform of the image in which disjoint sets of bits are identified with image size, orientation, and location in the field of view.

The output of the machine in the form of digital impulses is then ordered and used as input to an IBM 7090 computer program which counts, sorts, and prints the impulses (i.e., changes of state). The output is then graphed and forms base line curves which will be stored in the system memory. Ideally, when unidentified objects are processed by the pattern recognition system, the curves generated will be compared with those of known objects stored in memory and a table search will establish recognition. To enable the system to utilize stochastic principles, curves of transforms not previously included in the system memory and subsequently determined to be of proven value will be automatically included in the system memory to provide an increased base line capacity and recognition capability.

There appears to be kinship between the machine logic and circuitry and some of the behavior postulated for the human visual system.

Current studies indicate the feasibility of obtaining a transform which is invariant with respect to object rotation as well as image rotation, i.e., the transform of a rectangle is not altered when viewed from an angle such that its image is a parallelogram. The principal objective of this project is the discovery of techniques for the identification of real objects in a real environment as distinguished from character recognition. However, the techniques have an applicability to recognition of graphical data as well as textual material.

Reference:

(1) Buell, D. N. "Chrysler Optical Processing Scanner (COPS), A Character Recognition System Which is Independent of Character, Translation, Size or Orientation," in *Computers: Key to Total Systems Control.* New York: The Macmillan Company, 1961.

† *See* Glossary.

5.1.8 電気試験所
[ELECTROTECHNICAL LABORATORY]

1, 2-Chome, Nagata-Cho, Chiyoda-Ku, Tokyo, Japan

飯島泰蔵

[TAIZO IIJIMA], *Chief of Automata Section, Electronic Computer Division*]

Research on character recognition has been underway since 1958. The first phase of the research was concerned with the development of a reading machine capable of reading about 100 kinds of typewritten or printed characters. The first experimental reading machine was completed in 1959 (1), and the second at the end of 1961. Since some basic difficulties in character reading were pointed out as a result of this experimental work, the second phase of the project has been undertaken.

The purpose of this phase is to construct a basic theory of pattern recognition. The problem of the normalization of patterns has been taken up, and it is expected that the solution of this problem will be useful in increasing the efficiency of the present reading machines. It has been found that the method for normalization of patterns can be theoretically derived from the viewpoint of observation. First of all, the theory of the typical one-dimensional pattern was rigorously developed (2). The theoretical study of visual patterns (e.g., the two-dimensional pattern) is now in progress. As a byproduct of this study, the reason for the lens-system construction of animals' visual organs has been clarified.

The establishment of practical methodology with a view to applying the basic results obtained here to the reading machine will be an important problem for future consideration. Preparations are being made for the development of basic theories on the perfect method for coding patterns and the generalized correlation method for pattern recognition.

References:

(1) Wada, H., et al. "An Electronic Reading Machine," in *Information Processing*, pp. 227-232. London: Butterworths Scientific Publications, 1960.

(2) 飯島泰蔵 [Iijima, T.] パターンの正規化に関する基礎理論 典型的な1次元パターンの場合 ["Basic Theory on Normalization of Pattern (In Case of Typical One-Dimensional Pattern)"], 電気試験所彙報 [*Bulletin of the Electrotechnical Laboratory*], vol. 26, no. 5, 1962, pp. 368-388. (In Japanese)

FARRINGTON ELECTRONICS, INC. 5.1.9

7019 Edsall Road, Alexandria, Va.
ABRAHAM I. TERSOFF, *Manager, Research and Development*

The research and development program is directed toward the development of techniques for (a) accurate reading of deteriorated material, (b) reading of intermixed fonts in uncontrolled material, (c) providing optical scanners with useful self-organizing capabilities, and (d) high-speed scanning techniques which will permit the reading of thousands of characters per second.

A new character recognition technique scheduled for application to the reading of multifont numeric characters is based on following information streams within a character. Criteria relating to beginning, continuation, ending, divergence, convergence, etc., of the streams have been developed. A systematic sequential storage of these criteria is expected to produce character-identifying arrays which are invariant with respect to character size, configurational distortion, and reasonable angular misorientation.

Work on the following two complete systems incorporating such techniques is underway.

I. Development of an Automatic Address Reader for the U. S. Post Office Department (see 5.1.31) continues, with emphasis currently being placed on (a) completion and subsequent live-mail testing of the uppercase-address program, (b) computer analysis of the address-separating adequacy of the existing lowercase-address program, and (c) development of the zone number program.

II. Development of a sophisticated Selected Data Page Scanner system is nearing completion. Program capabilities designed into this scanner include specification through instruction selectors of such items as desired scan lines, scan fields, word lengths, storage modes, storage addresses, and required auxiliary operations. Reading capability includes a 62-character set and permits character stroke thickness variations in excess of 3 to 1.

GENERAL DYNAMICS/ELECTRONICS 5.1.10

1400 N. Goodman Street, Rochester 1, N. Y.
LAVEEN KANAL

The purpose of research underway is to study, both theoretically and experimentally, the application of adaptive pattern recognition techniques to various problems which require that input data be classified into groups.

The basic problem in the design of a pattern recognition system is the determination of system parameters from a small number of samples of known classification. The problem has been approached through the

use of well-founded statistical procedures, in particular, the use of iterative procedures to solve the chosen classification function. Several preliminary experiments in character recognition and speech recognition (see 5.2.10) have pointed out that the choice of a good set of "observables" from the data is necessary for success in all but the most trivial classification problems. A recent paper (1) reports on some of the basic principles used in pattern recognition and also discusses the results of several preliminary experiments in character and speech recognition.

The construction of an adaptive pattern recognizer has been completed. The machine is a special-purpose computer which will accept up to 1,024 dichotomous pattern observables from a flying spot scanner. The observables are weighted and summed by eight independent threshold logic units. The weightings are determined by an error-correcting reinforcement procedure which is done automatically by the machine.

Work in the near future will consist mainly of applying the adaptive pattern recognizer to classification problems.

Reference:

(1) Kanal, L., F. Slaymaker, D. Smith, and W. Walker. "Basic Principles of Some Pattern Recognition Systems," to be published in *Proceedings of the National Electronics Conference 1962.*

5.1.11 ING. C. OLIVETTI & C., S.P.A.

Laboratorio di Ricerche Elettroniche, Via del Parlamento, 33, Borgolombardo, Milan, Italy
GIORGIO PEROTTO, *Project Leader*

Systems have been developed for automatic pattern recognition in which the pattern is transformed into an electrical signal using a flying spot scanner.

A trial model or prototype of a character reader has been built to investigate in general the problems of printing quality and character style. A feature of this model is the reading of numerical characters of different style, including practically all types in use in typewriters and calculating machines.

Methods for overcoming difficulties occurring because of noise present in real characters and in papers are now being tested.

It is foreseen that the rejection and error rate will be very low, thus overcoming the necessity of a strict tolerance to character quality.

5.1.12

ИНСТИТУТ АВТОМАТИКИ И ТЕЛЕМЕХАНИКИ АН СССР [INSTITUTE OF AUTOMATICS AND TELEMECHANICS, ACADEMY OF SCIENCES U.S.S.R.]

Kalanchevskaya ulitsa, 15a, Moscow I-53, U.S.S.R.

Э. М. Браверман [E. M. BRAVERMAN]

Work is continuing on various aspects of pattern recognition involving a machine learning process [*Ed.*].

References:

(1) Браверман, Э. М. [Braverman, E. M.]. "Некоторые вопросы построения машин, классифицирующих обьекты по незаданному заранее признаку" ["Certain Problems in the Design of Machines Which Classify Objects According to an Identifying Feature Which Is Not Specified A Priori"], Автоматика и телемеханика [*Automatics and Telemechanics*], vol. 21, no. 10, October 1960, pp. 971 978.

(2) Браверман, Э. М. [Braverman, E. M.]. "Опыты по обучению машины распознаванию зрительних образов" ["Experiments in Teaching a Machine To Recognize Visual Forms"], Автоматика и телемеханика [*Automatics and Telemechanics*], vol. 23, no. 3, March 1962, pp. 349-364. Translation in JPRS 13694, Foreign Developments in Machine Translation and Information Processing, available from OTS.

5.1.13

ИНСТИТУТ КИБЕРНЕТИКИ АКАДЕМИИ НАУК УССР [INSTITUTE OF CYBERNETICS, ACADEMY OF SCIENCES UKRAINIAN S.S.R.][1]

4 Lysogorskaya, Kiev 28, U.S.S.R.

В. А. Ковалевский [V. A. KOVALEVSKIĬ], *Principal Investigator*

The purpose of research on pattern recognition is input of information to digital computers directly from printed, typed, and written documents. Current experiments are being conducted with typed characters primarily by means of the "Kiev" digital computer equipped with a flying spot scanner.

A program simulating the recognition of the full alphabet of a common typewriter has been prepared. Recognition, as in the case of previous work, is based on the calculation of correlation coefficients. Improvements have been made in order to expand the acceptable range of character displacement and to reduce the recognition time and the necessary memory. The reliability of recognition is very high in spite

[1] Formerly Computing Center of Academy of Sciences Ukrainian S.S.R.

of the low quality of typing. The recognition time is about 5 seconds in the case of legible characters and up to 40 seconds for badly typed ones.

Further investigations will be aimed at recognition of handprinted characters.

An experimental model of a reading device using optical correlation has been successfully tested. The device is designed for recognition of typed numerals.

In addition to these investigations, an output device for observation of graphs, characters, and various symbols on a tube screen has been designed. The device is controlled by computer. By means of a special subroutine, the computer calculates the coordinates of the dots which make up the desired pattern on the screen.

References:

(1) Ковалевский, В. А. [Kovalevskiĭ, V. A.]. "Корреляционный метод распознавания изображений" ("Correlative Method for Pattern Recognition"), Журнал вычислительной математики и математической физики [*Journal of Computational Mathematics and Mathematical Physics*], vol. 2, no. 4, 1962, pp. 684-694.

(2) Ковалевский, В. А., В. К. Елисеев [Kovalevskiĭ, V. A., and V. K. Eliseyev]. "Исследование алгоритма распознавания машинописных знаков" ["Investigation of an Algorithm for Recognition of Typed Characters"], Журнал вычислительной математики и математической физики [*Journal of Computational Mathematics and Mathematical Physics*], vol. 2, no. 5, 1962, pp. 902-911.

(3) Ковалевский, В. А., А. Г. Семеновский [Kovalevskiĭ, V. A., and A. G. Semenovskiĭ]. "Читающий автомат, основанный на анализе направлений" ["Reading Machine Based on the Analysis of Directions of Outline Scanning"], Читающие Устройства [*Reading Devices*], Institute of Scientific Information, Academy of Sciences U.S.S.R., Moscow, 1962, pp. 47-55.

5.1.14 INTERNATIONAL BUSINESS MACHINES CORPORATION

Advanced Systems Development Division, East Coast Systems Laboratory, P. O. Box 344, Yorktown Heights, N. Y.

EVON C. GREANIAS, *Senior Engineer*

An experimental machine that recognizes numbers in a broad variety of handwriting and can enter them directly into a computer system for

processing has operated successfully in preliminary tests. There are no plans to market this experimental reader commercially. It is being evaluated in tests conducted jointly at Tufts University by IBM and the university's Institute for Psychological Research.

The experimental numeric reader's design departs from previous techniques used in other character recognition machines. Previous devices scanned optically with one fixed "routine." This scanner explores the contours of a character with a moving beam of light. Its scheme employs serial scanning with a high-resolution flying spot scanner. As the scanning beam moves over its path following the contours of the numbers, analog voltages corresponding to the shape of the characters are generated and analyzed, using both analog and digital techniques. Special circuits detect the presence of important shape features of the handwritten numbers. The technique is flexible enough to allow recognition regardless of wide variations in size, registration, orientation, or shape of the numbers, so long as they are written within a 2- by 3-inch space on an ordinary IBM card. The logic (circuitry) of the experimental reader was developed on the basis of a sample selected from 3,000 numbers handwritten by a group of 30 people with no constraints placed on their style. The experimental reader system consists of the scanning and recognition unit, a buffer, and an IBM 514 Reproducing Punch. When a numeral is identified, this is recorded by means of punched holes in another IBM card which can be processed by conventional computers or accounting machines.

About 150 subjects—students, housewives, sales clerks, and others—have participated in the Tufts' experiments. Using a regular lead pencil, and allowed to make erasures, they have written more than 146,000 numbers in a variety of ways: copying random numbers, adding columns of figures, writing from a sitting and from a standing position, etc. Purpose of the experiments is to determine how well the reader performs when people have been trained to write "correctly," and how much tolerance for variations in character shapes should be provided in the next version of logic for the machine, which will be based on the much larger sample of numerals.

At the present level of logic development, 98.5 percent of the numerals written by trained subjects have been read correctly. Training consisted of approximately 30 minutes of instruction and practice on such things as avoiding excessive flourishes and gross distortions of numeral shapes. One-half of the numerals that were not read correctly were written by only 10 percent of the subjects. Of the numerals written by the other 90 percent, the reader correctly identified 99 percent, rejecting 1 percent as "unreadable" and mistaking .07 percent for the wrong numerals. The test results to date indicate that the experi-

mental reader has the kind of flexibility that would be required for practical application in handling large volumes of handwritten numerals.

5.1.15 INTERNATIONAL BUSINESS MACHINES CORPORATION

General Products Division, Rochester, Minn.
WILLIAM S. ROHLAND, *Program Manager*

Efforts are directed toward the investigation of techniques adaptable to multifont alphanumeric character recognition systems in uncontrolled applications. Computer simulation techniques are extensively used to optimize designs in conjunction with large volumes of real-life printing. Emphasis on scanner development is directed toward the general field of pattern recognition, as well as to specific character recognition applications.

Further investigation has been done in characterizing papers, inks, carbons, and pencil leads as they relate to transducers commonly used in scanning equipment.

5.1.16 INTERNATIONAL BUSINESS MACHINES CORPORATION

Thomas J. Watson Research Center, Yorktown Heights, N.Y.
G. L. SHELTON, JR., *Project Manager*

Work directed toward the automatic recognition of uncontrolled multifont print continues, with no change reported since the previous statement [*Ed.*].

5.1.17 京都大學

[KYOTO UNIVERSITY]

Department of Electrical Engineering, Kyoto, Japan
坂井利之
[TOSHIYUKI SAKAI]

Experiments have been carried out on a simple and compact recognition device for characters written on paper tape, which limits the height of the characters. The principle is to detect the sequence of occurrence of black and white when characters are passed row-by-row before an array of nine photocells aligned vertically.

The design for numerals, capital alphabets, and some other symbols has been completed, and the construction of the device will begin in the autumn of 1962.

Reference:

(1) 坂井利之 [Toshiyuki Sakai].
音声タイプライタ ["Phonetic Typewriter"], 科学 [*Kagaku*], May 1962.

LINK DIVISION OF GENERAL PRECISION, INC. 5.1.18

Binghamton, N.Y.
FRANK P. LEWANDOWSKI, *Program Manager, and*
ROBERT B. GREENLY, *Senior Engineer*

Work on the optical page scanning system which is capable of reading a variety of type styles continues, with no change reported since the previous statement [*Ed.*].

LITTON SYSTEMS, INC. 5.1.19

Communication Sciences Laboratory, 221 Crescent Street, Waltham, Mass.
GEORGE SEBESTYEN *and* ALICE K. HARTLEY

The problem of pattern recognition is treated as a problem in statistical hypothesis testing. According to this viewpoint each member of a class of things exhibiting a common pattern is reprcsented by a vector in a multidimensional vector space. Different dimensions represent different attributes while different vectors represent different examples of the pattern. The pattern can be characterized by the joint probability density of its attributes, and discrimination between examples of different patterns is based on likelihood ratio computations.

During the research program, different techniques of machine learning, i.e., characterization of the joint probabiiity densities of attributes of a pattern from a finite number of its examples, have been developed. With these techniques, multimodal densities can be approximated from a small number of examples by procedures that can be carried out rapidly. The computation techniques are demonstrated on a number of practical problems.

The program is sponsored by the U. S. Air Force Cambridge Research Laboratories.

References:

(1) Sebestyen, G. "Recognition of Membership in Classes," *Proceedings of the IRE,* Professional Group on Information Theory, vol. IT-7, no. 1, January 1961, pp. 44 ff.

(2) Sebestyen, G. "Pattern Recognition by an Adaptive Process of Sample Set Construction," *Proceedings of the IRE,* Professional

Group on Information Theory, vol. IT-8, no. 5, September 1962, pp. S82-S91.

(3) Sebestyen, George S., and Alice K. Hartley, *Study Program of Pattern Recognition Research,* final report on Contract No. AF 19 (604)-8024 for period January 1, 1961, through December 31, 1961, Litton Systems, Inc., December 31, 1961. (ASTIA No. AD—273 235)

5.1.20 LOCKHEED MISSILES & SPACE COMPANY

3251 Hanover Street, Palo Alto, Calif.

O. FIRSCHEIN, M. FISCHLER, *and* R. L. MATTSON

Research involves investigation of the decision-making portion of the pattern recognition problem. Threshold networks are used as the decision-makers, and a threshold network simulator, which uses binary weights and thresholds stored on a drum memory, performs the data processing.

Threshold network synthesis algorithms have been developed, and studies of the effect of †feature word construction and organizing set size have been carried out. Several forms of synthesis algorithms have been programmed for the IBM 7090 computer; up to 2,000 350-bit feature words can be used to determine the weights and thresholds for the decision network.

Recent work involves the automatic partitioning of classes to obtain subclasses, and the utilization of these subclasses in decision-making.

References:

(1) Firschein, O., and M. Fischler. *Automatic Subclasses Determination for Pattern Recognition Applications,* LMSC Internal Report, Pattern Recognition Note No. 4, November 1962.

(2) Fischler, M. *Hyperplane Techniques in Pattern Recognition,* LMSC Report No. 6-90-62-59, August 1962.

(3) Fischler, M., R. L. Mattson, O. Firschein, and L. D. Healy. "An Approach to General Pattern Recognition," *Proceedings of the IRE,* Professional Group on Information Theory, vol. IT-8, September 1962, pp. S64-S73. LMSC Report No. 6-90-62-2, April 1962.

(4) Mattson, R. L. *The Application of the Unger-Paull Minimization Method to Subclass Determination,* LMSC Internal Report, Pattern Recognition Note No. 1, September 1962.

(5) Mattson, R. L., and O. Firschein. *Feature Word Construction for Use With Pattern Recognition Algorithms,* LMSC Report No. 6-90-62-58.

† *See* Glossary.

(6) Mattson, R. L., O. Firschein, and M. Fischler. *Methods of Increasing Redundancy in a Class of Synthesis Algorithms,* LMSC Report No. 6-90-62-83, September 1962.

5.1.21

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Engineering Projects Laboratory, Data Processing Systems Division, Cambridge, Mass.
D. M. BAUMANN, *Project Leader*

A character recognition system based on weighted area scanning techniques is being investigated. An IBM 7090 computer program is being utilized to find the optimum weighting function for any given alphabet. Thereafter, a masked photocell can be idealized to divide the alphabet into nonexclusive subsets. This process is repeated until all the characters have been recognized.

The system has been simulated on the 7090, and a simple demonstration prototype recognition machine has been constructed and tested.

The investigation is sponsored by the U. S. Office of Naval Research.

References:

(1) Baumann, D. M. "Weighted Area Scanning Techniques for Character Recognition," in *Optical Character Recognition* (proceedings of the ONR-NBS Symposium on Optical Character Recognition, Washington, D. C., January 1962), ed. by G. L. Fischer, Jr., et al., pp. 197-208. Washington, D. C.: Spartan Books, 1962.

(2) Baumann, D. M. *Feasibility of a Weighted-Area Scanning Technique for Character Recognition.* Sc. D. Thesis. Cambridge, Mass.: Massachusetts Institute of Technology, Department of Mechanical Engineering, June 1960.

(3) Lemke, G. A. *Computer Design of Character Recognition Masks.* Master's Thesis. Cambridge, Mass.: Massachusetts Institute of Technology, Department of Mechanical Engineering, March 1962.

(4) Ohlenbusch, Cord W. *Consecutive Character Display of Printed Material by Video Scanning.* S. B. Thesis. Cambridge, Mass.: Massachusetts Institute of Technology, Department of Electrical Engineering, January 1962.

(5) Trantanella, F. M. *Design of a High Speed Page Reader.* Master's Thesis. Cambridge, Mass.: Massachusetts Institute of Technology, Department of Electrical Engineering, May 1962.

5.1.22 MELPAR, INC.

Applied Science Division,[1] *Watertown 72, Mass.*
PAUL W. COOPER[2]

Viewing pattern recognition as a problem in statistical classification wherein an n-space is partitioned into category regions with decision boundaries, a research program has been directed toward developing a useful mathematical theory of pattern recognition. Attention has been directed to boundary forms which are simple to implement and yet are optimum for wide ranges of distributions. Among the questions examined have been: characterization of the classes of distributions for which the boundaries are optimum, techniques for determining the optimum boundaries from known samples, methods of ordering the coordinate directions (attributes) according to their importance in contributing to the decision, and methods of implementation.

References:

(1) Cooper, P. W. *Classification by Statistical Methods (Pattern Recognition),* Melpar Technical Note 61/2, April 1961.

(2) Cooper, P. W. *The Hyperplane in Pattern Recognition,* Melpar Technical Note 61/6, December 1961.

(3) Cooper, P. W. *The Hypersphere in Pattern Recognition,* Melpar Technical Note 62/1, February 1962.

(4) Cooper, P. W. *Statistical Pattern Recognition with Quadratic Forms,* Melpar Technical Note 62/4, June 1962.

5.1.23 NATIONAL BUREAU OF STANDARDS

Data Processing Systems Division, Washington 25, D. C.
M. E. STEVENS *and* B. K. RANKIN, III

A current project involves a survey of progress in automatic character recognition equipment development, study of related trends in pattern recognition research, and consideration of techniques which may be applicable to recognition of Chinese characters.

Emphasis is being directed toward the possibilities of recognizing Chinese ideographs and toward the related questions of dictionary lookup on the basis of scanner output as a step toward mechanized translation of Chinese text. An approach involving both generative and recognition grammars for the construction and stroke analysis of ideographs is being pursued.

The project is sponsored by the U. S. Army Signal Corps.

---

[1] The Applied Science Division was transferred to Melpar, Inc., Arlington, Va., in September 1962.

[2] Present address: Sylvania Electronic Systems, Waltham 54, Mass.

In addition, under the sponsorship of the Bureau and other agencies, specific research tasks are undertaken from time to time such as the development of a machine model for recognition of a set of geometric shapes with size invariance.

References:

(1) Stevens, M. E. "Abstract Shape Recognition by Machine," in *Computers: Key to Total Systems Control,* pp. 332-351. New York: The Macmillan Company, 1961.

(2) Stevens, M. E. *Fourth Quarterly Progress Report, Delivery Order No. 47802PW-61-72-72, Modification No. 1, Task e,* September 25, 1962. 15 p. (mimeo)

NATIONAL PHYSICAL LABORATORY 5.1.24

Autonomics Division, Teddington, Middlesex, England

D. O. CLAYDEN, M. B. CLOWES, *and* J. R. PARKS

Work continues on a multifont reading system based upon a feature description of a character, with no change reported since the previous statement [*Ed.*].

References:

(1) Clowes, M. B., and J. R. Parks. "A New Technique in Automatic Character Recognition," *The Computer Journal,* vol. 4, no. 2, 1961, p. 121.

(2) Clowes, M. B. "The Use of Multiple Auto-Correlation in Character Recognition," in *Optical Character Recognition* (proceedings of the ONR-NBS Symposium on Optical Character Recognition, Washington, D. C., January 1962), ed. by George L. Fischer, Jr., et al., pp. 305-318. Washington, D. C.: Spartan Books, 1962.

NEW YORK UNIVERSITY 5.1.25

Division of General Education, New York 3, N.Y.

CHARLES BONI, *Chief Investigator*

The study of Russian printing styles to obtain certain types of graphic data which can be used as source-data in the design of an automatic print reader, last reported in *CRDSD, No. 5,* p. 89, was completed in 1959 (1).

The project was supported by the U. S. Air Force (Rome Air Development Center).

Reference:

(1) Boni, Charles. *Russian Type Study.* Final report on contract AF30(602)-1824, New York University, Division of General

Education, April 28, 1959. Report No. RADC-TR-59-35. (ASTIA No. AD—211 938)

5.1.26 NEW YORK UNIVERSITY

Department of Electrical Engineering, New York 53, N. Y.
HERBERT FREEMAN, *Principal Investigator*

A study is being carried out to develop systematic procedures for the analysis and manipulation of geometric line drawings in a digital computer. The approaches used are based on the method of chain encoding (2). This encoding method facilitates the development of algorithms for determining curve properties as well as for synthesizing curves to satisfy specified requirements.

Present studies involve an investigation of the quantization errors associated with the particular encoding method, the development of generalized pattern classification criteria, an investigation of possible extensions of the techniques to three dimensions, and the development of additional algorithms for curve property determination. During the past 6 months, a number of previously derived algorithms were successfully tested on an IBM 1620 computer. A pattern plotting program for this computer has been written which permits the 1620 computer directly to plot large patterns (extending over many pages of the line printer).

As an example in pattern classification, a study has been undertaken to determine the feasibility of programming a computer to solve a jigsaw puzzle. Initial results are promising, and it is expected that the study will provide valuable insight into the pattern classification problem.

A semiautomatic chain encoding device (SACE), which will facilitate the encoding of line patterns for either a PB 250 or an IBM 1620 computer, has been designed. The device should be ready for testing before the end of 1962.

The research is sponsored by the U. S. Air Force Office of Scientific Research.

References:

(1) Bodin, L. *A Computer Program for Chain-Encoded Plane Curves,* Tech. Note 400-8, Department of Electrical Engineering, New York University, October 1961.

(2) Freeman, H. "Techniques for the Digital Computer Analysis of Chain-Encoded Arbitrary Plane Curves," in *Proceedings of the National Electronics Conference,* Chicago, Ill., Oct. 9-11, 1961, vol. 17, pp. 421-432.

(3) Freeman, H. "A Technique for the Classification and Recognition of Geometric Patterns," in *Proceedings of the 3rd Inter-*

national Congress on Cybernetics, Namur, Belgium, September 11-15, 1961.

(4) Freeman, H. "On the Digital Computer Classification of Geometric Line Drawings," to be presented at the National Electronics Conference, Chicago, Ill., October 8-10, 1962.

(5) Gardner, L. *On the Recognition of Hand-Printed Characters,* Tech. Note 400-6, Department of Electrical Engineering, New York University, September 1961.

(6) Glass, J. *Analysis of Quantized Two-Dimensional Geometric Patterns,* Tech. Note 400-7, Department of Electrical Engineering, New York University, September 1961.

PANORAMIC RESEARCH, INC. 5.1.27

3946 Fabian Way, Palo Alto, Calif.
W. W. BLEDSOE

Work to extend and increase the effectiveness of the n-tuple pattern recognition method continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Bledsoe, W. W. *Lethally Dependent Genes Using Instant Selection,* Informal Report, PRI: 1, Panoramic Research, Inc., December 1961.

(2) Bledsoe, W. W. "The Use of Biological Concepts in the Analytical Study of Systems," presented to the Operations Research Society of America/The Institute of Management Sciences National Meeting, San Francisco, Calif., November 10, 1961; Issued as PRI:2, Panoramic Research, Inc.

PHILCO CORPORATION 5.1.28

Scientific Laboratory, Blue Bell, Pa.
J. F. BOGUSZ *and* E. N. POWERS

A research program of study and experiment is underway to seek optimum prenormalization techniques to facilitate the recognition of gray-scale aerial photographic imagery with current pattern recognition schemes utilizing adaptive memories. Experiments are being performed, using training and testing sequences, to relate the parameters of the input data to the capabilities of the recognition system. Image input parameters are modified by photographic adjustment of scale factor and geometric rectification, and by electro-optical spatial filtering techniques which enhance object contours, delete "clutter," and accentuate significant recognition characteristics. In addition, study and

experimentation are being directed to the problem of optimizing the number and kind of "random" criteria used for recognition.

This program is sponsored by the U. S. Air Force (Rome Air Development Center) and is expected to be completed during the second quarter of 1963.

5.1.29 PHILCO CORPORATION

Scientific Laboratory, Blue Bell, Pa.

J. B. CHATTEN *and* C. F. TEACHER

Research continues on the recognition of machine and handprinted alphanumeric symbols, with no change reported since the previous statement [*Ed.*].

5.1.30 PHILCO CORPORATION

Data Recognition Engineering Laboratory, Communications and Electronics Division, 4700 Wissahickon Avenue, Philadelphia, Pa.

A. I. FRANK, *Manager of Engineering*

Work continues on an all-electronic machine that reads and recognizes printed and typewritten postal addresses on envelopes (see 5.1.31), with no change reported since the previous statement [*Ed.*].

5.1.31 POST OFFICE DEPARTMENT

Office of Research and Engineering, Washington 25, D.C.

RICHARD W. HESSINGER, *Electronic Engineer*

Work continues on the laboratory model address-reading machine employing the electronic flying spot scanning technique, with no change reported since the previous statement [*Ed.*].

5.1.32 RABINOW ENGINEERING CO., INC.

1025 Research Boulevard, Rockville, Md.

JACOB RABINOW, *President*

Testing of a special-purpose Rabinow reader, which will be capable of reading six intermixed fonts, continues, with no change reported since the previous statement [*Ed.*].

5.1.33 RADIO CORPORATION OF AMERICA

Applied Research, Camden, N. J.

W. J. HANNAN

Work continues on a reading machine which employs optical correlation or a "mask matching" technique. Under a contract with the U. S. Air Force (Rome Air Development Center), the machine's ability to recognize complete English and Russian fonts has been established. A final report is in preparation.

電波研究所 5.1.34

[RADIO RESEARCH LABORATORIES]

Research Section of Information Processing,
Koganei-shi, Tokyo, Japan
尾方義春
[YOSHIHARU OGATA]

Research on character recognition is in progress with the aid of the NEAC 2203 computer and its input data translator (1) for printed or handwritten characters.

Projects presently underway involve recognition of (a) printed characters, including English alphabet, Japanese phonetic alphabet, and numerals (2), and (b) handwritten characters, especially numerals (3). The former is based on the correlation method.

Additional improvement was gained against rotation and inclination of the character and in the economy of memory and operation time by the introduction of new integration (1) (2).

Several methods of feature detection were applied to the recognition of handwritten numerals, and a new "shifted pattern method" is now being programmed and tested.

References:

(1) 尾方義春, 小泉深吉, 佐藤允克, 米山一彦
[Ogata, Yoshiharu, Shinkichi Koizumi, Nobukatsu Sato, and Kazuhiko Yoneyama]. 文字識別の一方法
["A Method of Character Recognition"] (in Japanese), *Quarterly Report of R.R.L.*, vol. 8, November 1962. (In press)

(2) 佐藤允克, 米山一彦, 尾方義春
[Sato, Nobukatsu, Kazuhiko Yoneyama, and Yoshiharu Ogata]. "Printed Character Recognition" (in Japanese), *Quarterly Report of R.R.L.*, vol. 7, November 1961, pp. 489-493.

(3) 米山一彦, 尾方義春
[Yoneyama, Kazuhiko, and Yoshiharu Ogata].
数字パターンの識別 ["Numerical Pattern Recognition"] (in Japanese), *Quarterly Report of R.R.L.*, vol. 7, November 1961, pp. 494-498.

RAYTHEON COMPANY 5.1.35

Missile and Space Division, Bedford, Mass.
P. W. CHENEY, T. B. KNAPP, O. LOWENSCHUSS, *and*
T. TANIMOTO[1]

Research on the generation and analysis of signals resulting from scans of geometrical solids is being continued. Various mathematical

[1] Formerly of Melpar, Inc., Applied Science Division, Watertown 72, Mass.

techniques, as well as storage and retrieval methods, have been and are being considered. Computer programs of the methods have been applied to data produced by various real solids.

The project is being conducted under contract to the Advanced Research Projects Agency.

Reference:

(1) Raytheon Company. *Radar Echo Characteristics. Semi-Annual Technical Summary Report,* BR-1769, August 1962.

5.1.36 RECOGNITION EQUIPMENT INCORPORATED

Ross at Prairie Avenue, Dallas 4, Tex.

E. G. PERRY, T. Q. LEBRUN, *and* I. SHEINBERG

A high-resolution analog Electronic Retina Character Reading System is in the final stages of development. A prototype page carrier has successfully handled pages fast enough to permit a viewing rate of 950 printed lines per minute. Maximum page handling rate is 30 pages per minute. A compatible unit-document handling machine has been designed to process documents at the rate of 1,200 per minute. A compatible adding machine tape handling machine has also been designed. A "one-font-at-a-time" alphanumeric recognition unit is being debugged.

The completely parallel operation provided by a two-dimensional retinal array permits very high character reading rates with inexpensive, low-speed electronic components. The parallel system eliminates the requirement for storage within the recognition area of the machine and the requirement to "chop" each unknown character into a series of artificial vertical slices to accomplish recognition.

The analog characteristic of the system permits retention of all the information inherent in the original character. No arbitrary quantizing with attendant loss of character information is effected in the viewing portion of the system.

The extension of the system's capabilities to intermixed, multiple alphanumerics will be completed early in 1963.

5.1.37 SCOPE INCORPORATED

121 Fallfax Drive, Falls Church, Va.

E. C. DRIESE, *Project Engineer*

Current research in data processing is centered on image analysis techniques suitable for use with aerial reconnaissance photography.

The specific goal of the work is to provide a suitable display for an adaptive memory type of cognitive system, such as CONFLEX I or the Perceptron. The display will characterize "objects of interest" on a

retinal field of the cognitive system independent of size, translation, orientation, and noise (background contrast). Possible implementations under analysis include both photoreceptive retinas and scanning systems. The image processing methods being studied are special sensory field wiring, unusual optical transforms, and extraction through video signatures. All efforts in this area of data reduction exploit the known properties of "objects of interest" to be encountered.

This project is sponsored by the U. S. Air Force (Rome Air Development Center).

Reference:

(1) Uffelman, M. R. "CONFLEX I," *1962 IRE International Convention Record*, vol. 10, Part IV, Electronic Computers and Information Theory, March 1962.

STANFORD RESEARCH INSTITUTE 5.1.38

Menlo Park, Calif.
ALFRED E. BRAIN, *Project Leader*

A research study and experimental investigation of techniques suitable for practical application to graphical data processing are underway.

The particular problem is the transfer of data from their original form on maps, charts, and aerial photographs to a storage medium from which items of detail can be retrieved at short notice. The program includes the study and development of organizations of combined fixed and adaptive networks that will permit real-time recognition of patterns, independent of size, displacement, and rotation, in the presence of interfering signals and noise; the development of components and subsystems suitable for implementing the schemes devised; and, finally, the design and construction of an experimental data processing machine.

The problem of finding a method of transforming the graphical input data into a set of simultaneous electrical signals appropriate for the pattern recognition system has been satisfactorily answered by the use of multiple optical replications. Of the several methods examined for generating numerous identical images lying in the same plane, the "fly's eye" array of lenses was selected on the basis of uniformity of illumination and resolving power. A test array has been constructed for 100 images; a checkerboard pattern of 100 by 100 elements on the retina is reproduced with at least 90-percent modulation in each image. The "wiring" of the retina consists of 100 masks, one per image, all located on the same 10- by 8-inch photographic plate. Cadmium selenide photocells were selected for converting the light transmitted by the masks into electrical inputs for the machine.

Two modules, each containing many quasi-analog storage elements

having nondestructive readout, have been developed. Both modules are in the form of magnetostrictively driven delay lines.

Several alternative system designs have been simulated in order to identify and measure the relative merits of projected methods.

The project is supported by the U.S. Army Electronics Research and Development Laboratory, Fort Monmouth, N.J.

References:

(Requests for project reports should be directed to the Data Transducer Branch, Communications Department, U.S. Army Electronics Research & Development Laboratory, Fort Monmouth, N. J.)

(1) Brain, A. E., and G. E. Forsen. *Graphical Data Processing Research Study and Experimental Investigation,* Quarterly Progress Report No. 7, SRI Project 3192, Contract DA 36-039 SC-78343, Stanford Research Institute, Menlo Park, Calif., March 1962.

(2) Brain, A. E., and N. J. Nilsson. *Graphical Data Processing Research Study and Experimental Investigation,* Quarterly Progress Report No. 8, SRI Project 3192, Contract DA 36-039 SC-78343, Stanford Research Institute, Menlo Park, Calif., June 1962.

5.1.39 STANFORD RESEARCH INSTITUTE

Menlo Park, Calif.

CHARLES ROSEN, *Project Supervisor*

Work on applying methods of integral geometry to the problem of obtaining parameters of patterns which are invariant to translation, rotation, and/or size continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Novikoff, Albert B. "Integral Geometry as a Tool in Pattern Processes," in *Bionics Symposium. Living Prototypes—the Key to New Technology,* WADD Technical Report 60-600. Wright-Patterson Air Force Base, Ohio: Directorate of Advanced Systems Technology, Wright Air Development Division, A.R.D.C., U.S. Air Force Base, December 1960.

5.1.40 SWARTHMORE COLLEGE

Electrical Engineering Department, Swarthmore, Pa.

CARL BARUS

A method is being studied for machine recognition of patterns taken from a class of patterns unknown to the designer but meeting certain restrictions. It is assumed that many specimens of each pattern of the

class to be learned are available in sequence. No assumption is made as to how the specimens of a pattern are distributed. Several specimens of each pattern are stored, with their identification. These subsets are used in identifying an unknown by suitably weighted comparisons. A learning process is employed to select subsets for storage that are representative of the statistical distributions of the respective patterns. Current work is largely concerned with this learning process. Experiments using a two-symbol, mixed-font, hand-drawn alphabet are in progress.

The work is supported by a National Science Foundation grant.

Reference:

(1) Barus, Carl. "A Scheme for Recognizing Patterns from an Unspecified Class," in *Optical Character Recognition* (proceedings of the ONR-NBS Symposium on Optical Character Recognition, Washington, D. C., January 1962), ed. by George L. Fischer, Jr., et al., pp. 227-247. Washington, D. C.: Spartan Books, 1962.

SYLVANIA ELECTRONIC SYSTEMS 5.1.41

A division of Sylvania Electric Products, Inc.,
Applied Research Laboratory,
40 Sylvan Road, Waltham 54, Mass.

STEPHEN B. GRAY, GEORGE G. PICK, WALTER WEISSBLUM, *and* DONALD B. BRICK

Work on synthesis of optimum linear weighting functions for a pattern recognizer, which operates by digital matching, continues, with no change reported since the previous statement [*Ed.*].

TECHNISCHE HOCHSCHULE KARLSRUHE 5.1.42

Institut für Nachrichtenverarbeitung und Nachrichtenübertragung,
Karlsruhe, Federal Republic of Germany

K. STEINBUCH, Director, *and* H. KAZMIERCZAK, *Project Leader*

The research program on automatic character recognition continues. A reading automaton for printed, typewritten, and handprinted material (10 numerals) based on the potential field method has been constructed and is now in operation. The reading rate is 3,000 characters per second. The recognition of alphanumeric characters is planned. Learning processes will be introduced later by using an adaptive translator (learning matrix) for simple self-programming and permanent learning.

A method of optical machine reading has been designed for specially printed numerals and location marks. The determination of characters is accomplished by the stripe angle of hatched characters printed in

their normal shape. A simple device based on this method has been constructed.

The research is supported by the Deutsche Forschungsgemeinschaft.

References:

(1) Kazmierczak, H. "The Potential Field as an Aid for Character Recognition," in *Information Processing,* pp. 244-247. London: Butterworths Scientific Publications, 1960.

(2) Steinbuch, K. "Die Lernmatrix" ["The Learning Matrix"], *Kybernetik,* vol. 1, no. 1, January 1961, pp. 36-45.

(3) Kazmierczak, H., and P. Reuschlen. "Maschinelle Erkennung von Schraffurzeichen" ["Machine Reading of Hatched Characters"], *Nachrichtentechn. Zeitschr.,* vol. 14, no. 10, October 1961, pp. 496-501.

5.1.43 UNIVERSITY OF CALIFORNIA

Department of Electrical Engineering, Berkeley, Calif.

JEROME R. SINGER, *Associate Professor and Chief Investigator*

Design of an electronic machine to "read" in a manner analogous to humans continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Singer, J. R. "Model for a Size Invariant Pattern Recognition System," in *Bionics Symposium. Living Prototypes—The Key to New Technology.* Wright Patterson Air Force Base, Ohio: Directorate of Advanced Systems Technology, Wright Air Development Division, A.R.D.C., U.S. Air Force, December 1960.

(2) Singer, J. R. "An Electronic Analog of the Human Recognition System," *Journal of the Optical Society of America,* vol. 51, no. 1, January 1961, pp. 61-70.

(3) Singer, J. R. "A Self Organizing Recognition System," in *Proceedings of the Western Joint Computer Conference,* May 1961.

(4) Singer, J. R. "Recognition with Size and Rotation Invariance —an Electronic Analog of Human Recognition Processes," presented at the International Biophysics Congress, Stockholm, Sweden, July 31-August 4, 1961.

5.1.44 UNIVERSITY OF MICHIGAN

Office of Research Administration, Cooley Electronics Laboratory, Ann Arbor, Mich.

WILSON P. TANNER, JR., *Principal Investigator*

The pattern recognition problem is being investigated in its general form without regard to particular applications. A mathematical

characterization has been formulated which encompasses a broad range of recognition schemes. A general machine for these schemes has been described and is being analyzed in terms of its components and their effect on the potential performance of the machine. Attention is also being directed toward learning situations, primarily those which do not involve feedback. A mathematical approach to context is being formulated.

The project is being carried out under contract with the U. S. Air Force Office of Scientific Research.

UNIVERSITY OF MICHIGAN **5.1.45**
Mental Health Research Institute, Ann Arbor, Mich.
and
SYSTEM DEVELOPMENT CORPORATION
Santa Monica, Calif.
LEONARD UHR,[1] J. PHILIP BENKARD,[1] *and*
REBECCA PRATHER[2]

Methods for efficient learning and induction by means of computer simulations are being studied. Several versions of a sensory pattern recognition program have been coded, and also a first attempt has been made to generalize pattern recognition to handle patterns of symbols.

The pattern recognition program has been tested with a wide range of different input patterns, including degraded photographs of faces, cartoons, and spoken words. Several recent tests have been made on its abilities to recognize patterns correctly as a function of certain aspects of its own internal structure, for example the number of operator-measures it has at its disposal, and its methods for accumulating and evaluating its past experience.

A pre-processing subroutine, designed to perform the functions of the on-off nets found in the mammalian eye, is now being used to transform a shape into its contour, compress and summarize this contour, and list it as a one-dimensional string. This subroutine has not yet been fully integrated into the pattern recognition program.

A simple language translation program that makes use of many of the principles of discovery and learning used by the pattern recognition program has been coded and tested. This program attempts to translate between language pairs presented to it in the form of successive sentence pairs. In other words, the program is presented examples of sentences in one language along with their translations in a second lan-

[1] University of Michigan
[2] System Development Corp.

guage, and must attempt to build up its vocabulary and translation rules as a function of such experiences.

When combined with the subroutine that transforms shapes into their contours, the language translation program can do pattern recognition. The contour string is given to it as its input sentence, and the pattern's name as its output sentence; it must learn to translate the many different contours possible for a single pattern into the single name. This program is inferior to the pattern recognition program in learning to recognize degraded patterns and does not have as good ability to generalize over variant examples of a pattern, except by a brute-force rote-learning accumulation of information. But the preprocessing subroutine performs a number of useful transformations; e.g., it maps all linear transformations of a pattern into a single contour string.

An attempt is being made to strengthen and generalize the language-learning program so that it will be able to work with a wider variety of pattern recognition and symbol transformation problems. The general approach is to ask the program to discover important information in its experiences, and to continually organize and reorganize this information into useful inductions that then guide its future behavior. For example, for language translation that program should be able to build up inductively (a) its own vocabulary of words and idioms, along with (b) something similar to a phrase-structure tree for each of the two languages, and (c) transformations from one language to the other, both at the level of the individual words and the higher-order classes making up the phrase-structure tree.

This research is supported by the University of Michigan, the System Development Corporation, and the U. S. Public Health Service.

References:

(1) Uhr, L., and C. Vossler. "The Search to Recognize," in *Optical Character Recognition* (proceedings of the Symposium on Optical Character Recognition, Washington, D.C., January 1962), ed. by George L. Fischer, Jr., et al., pp. 319-329. Washington, D.C.: Spartan Books, 1962.

(2) Vossler, C., and L. Uhr. "Computer Simulations of a Perceptual Learning Model for Sensory Pattern Recognition, Concept Formation, and Symbol Transformation," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.

(3) Uhr, L. "The Development of Perception and Language: Simulated Models," in *Computer Simulations of Personality*, ed. by S. Messick and S. Tomkins. New York: John Wiley and Sons, Inc. (In press)

5.2 *SPEECH ANALYSIS AND SYNTHESIS*

Research on speech analysis and synthesis can be considered to be a special form of pattern recognition which has potential impact similar to that described for character and pattern recognition. As in the case of the latter, the ultimate goal of the research and development activities in this area is the construction of a device or devices that will recognize spoken words, identify the speaker, and even perform the inverse process of synthesizing or imitating spoken words. At present, there exist only in prototype form a few systems which can recognize a small number of spoken words or digits.

The following summary calls attention in particular to new work reported here for the first time and to projects and studies that have been completed.

A program on automatic speech recognition at **Cornell Aeronautical Laboratory, Inc.** (5.2.8) has the objective of discovering means of measuring or detecting in the spoken sound certain factors which together are sufficient for identifying English consonant phonemes.

Techniques are being investigated by **Litton Systems, Inc.** (5.2.13) for producing a phonetic transcription of speech.

Philco Corporation (5.2.18) is conducting a basic study of the factors involved in the human perception of speech in order to establish the specifications of the invariant features of spoken words.

As part of a program to develop a voice-to-teletype code converter, **RCA** (5.2.19) has completed a study to determine criteria for automatic recognition of speech sounds.

At **Universität Bonn** (5.2.24) a program of morphophonemic analysis and the construction of electronic apparatus for segmentation of speech signals and recognition of audio-linguistic units is being pursued.

At Sylvania Electronic Systems (5.2.22) the research study to determine the validity of the coefficients of an appropriate set of orthonormal functions as a simplified representation of speech sounds for use in machine speech recognition has been completed.

The work previously reported by Bulova Research & Development Laboratories (5.2.5) has been transferred to **New York University** (5.2.17).

5.2.1
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

CRRBA, L. G. Hanscom Field, Bedford, Mass.
STANLEY R. PETRICK

Work on the speech pattern recognition procedure for replacing a sequence of binary categorization decisions by a single computation

which characterizes membership in one of a number of possible classes continues, with no change reported since the previous statement [*Ed.*].

5.2.2

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

L. G. Hanscom Field, Bedford, Mass.

WEIANT WATHEN-DUNN

The Speech Research Branch (CRRSV) of the Communication Sciences Laboratory investigates pertinent aspects of speech communication in order to learn more about the basic processes involved in generating and perceiving natural acoustical speech signals, analyzing and synthesizing such signals and their electrical counterparts, and processing the electrical versions in various ways. The purpose of this endeavor is to gain fundamental knowledge which will be useful in developing techniques for improving the reliability of voice communications, especially under adverse or unusual circumstances, and for performing new voice-communication functions. Effort continues on problems associated with extracting the voicing excitation function from real speech and on problems concerned with pitch perception. Further investigations toward determining the best short-time measures of raw speech signals have been made.

References:

(1) Lieberman, Philip. "Perturbations in the Vocal Pitch of Normal and Diseased Larynxes," Paper G37, presented at the Fourth International Congress on Acoustics, Copenhagen, Denmark, August 21-28, 1962; also Paper C6, presented at the Speech Communication Seminar, Stockholm, Sweden, August 29-September 1, 1962.

(2) Lieberman, Philip, and Sheldon B. Michaels. "Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech," *Journal of the Acoustical Society of America,* vol. 34, no. 7, July 1962, pp. 922-927.

(3) Lieberman, Philip, and Sheldon B. Michaels. "On the Discrimination of Missing Pitch Pulses," Paper C7, presented at the Speech Communication Seminar, Stockholm, Sweden, August 29-September 1, 1962.

(4) Michaels, Sheldon B., and Philip Lieberman. "Discrimination of Missing Pitch Pulses," Paper N5, presented at the 63rd Meeting of the Acoustical Society of America, New York, May 23-26, 1962; abstract in *Journal of the Acoustical Society of America,* vol. 34, no. 5, May 1962, p. 729.

(5) Wathen-Dunn, Weiant. "On Peak Factors and Asymmetries of Acoustical Speech Signals," Paper G44, presented at the Fourth International Congress on Acoustics, Copenhagen, Denmark, August 21-28, 1962.

AUTONETICS 5.2.3

A Division of North American Aviation, Inc., Research and Development, P. O. Box R-3, Anaheim, Calif.

J. D. BLEDSOE

An investigation of pattern recognition is being continued, with primary emphasis on acoustic data. Speech is used as a source of modulated data, whereas machinery-generated sources are used as examples of steady-state data.

Experiments are in progress to determine parameters applicable to the recognition process, and to determine methods of correlation and decision, where a large number of parameters are to be considered. These experiments utilize the IBM 7090 computer and the Autonetics RECOMP II digital computer, as well as bandpass filters, digitizing, and spectral display equipment. Both logical and adaptive correlation and decision systems are under consideration.

References:

(1) Bledsoe, J. D., and C. C. McConkle. *Evaluation and Simulation of Methods for Speech Recognition,* Autonetics Report EM-7615, May 1962.

(2) Bledsoe, J. D., and C. C. McConkle. *A System for the Data Reduction and Correlation of Spoken Words,* Autonetics Report EM-1162-115, May 1962.

(3) McConkle, C. C. *The Verdan Speech Recognition Demonstration,* Autonetics Report EM-1162-120, June 1962.

(4) Bledsoe, J. D. *An Underwater Acoustic Analysis and Recognition Study,* Autonetics Report EM-1162-130, July 1962.

BELL TELEPHONE LABORATORIES, INC. 5.2.4

Murray Hill, N. J.

L. D. HARMON

Studies leading to automatic recognition of speech continue, with no change reported since the previous statement [*Ed.*].

References:

(1) Julesz, B. "Visual Pattern Discrimination," *IRE Transactions on Information Theory,* vol. IT-8, no. 2, February 1962.

(2) Julesz, B. "Towards the Automation of Binocular Depth Perception," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.

5.2.5 BULOVA RESEARCH & DEVELOPMENT LABORATORIES

Division of Bulova Watch Company, Inc.,
62-10 Woodside Avenue, Woodside 77, N. Y.

The project described under this heading in *CRDSD*, No. 10, Statement No. 5.2.5, has been transferred to New York University (see 5.2.17).

5.2.6 COGNITRONICS CORPORATION

Treasure Hill, Pleasantville Road, Briarcliff Manor, N.Y.
DAVID H. SHEPARD

Development of speech recognition equipment continues, with no change reported since the previous statement [*Ed.*].

5.2.7 COLUMBIA UNIVERSITY

Department of Electrical Engineering, New York 27, N. Y.
CYRIL M. HARRIS, *Principal Investigator*

Work continues on the detailed analysis of plosive speech sounds, with no change reported since the previous statement [*Ed.*]. The project is supported by the U. S. Air Force Office of Scientific Research.

5.2.8 CORNELL AERONAUTICAL LABORATORY, INC.

Computer Research Department,
P. O. Box 235, Buffalo, N. Y.
H. R. LELAND, *Head, Cognitive Systems Section*

A program in automatic speech recognition has as its objective the discovery of means of measuring or detecting in the spoken sound certain factors which are together sufficient for identifying English consonant phonemes. The factors sought are (a) "position" of articulation (lips, teeth, hard palate, soft palate), (b) "manner" (stop, fricative, nasal, liquid), and (c) "voicing" (voiced, unvoiced).

The experimental procedure in current work is to make Sonograms and evaluate them visually. Analysis of Sonograms of the 33 phonemes (spoken by one speaker, as "ibby," "iddy," "iffy," etc.) has shown char-

acteristics of the second factor, manner, which confirm results by earlier workers. Equipment available for this work includes a Sonograph, a sound booth, and standard audio recording and amplification equipment.

FEDERAL SCIENTIFIC CORPORATION 5.2.9

615 West 131st St., New York 27, N.Y.
MARK WEISS, *Principal Investigator*

Research directed toward the use of significant parameters of speech in experiments on machine recognition of spoken English continues, with no change reported since the previous statement [*Ed.*]. The work is supported by the U. S. Air Force (Rome Air Development Center).

GENERAL DYNAMICS/ELECTRONICS 5.2.10

1400 North Goodman Street, Rochester 1, N.Y.
F. H. SLAYMAKER, R. A. HOUDE, *and* W. D. LARKIN

The ultimate goal of speech research underway is to develop the tools necessary to convert conversational speech to a printed English output. To this end, the following projects have been undertaken, or completed.

The Formant Vocoder study resulted in the construction of a 55-channel spectrum analyzer, multiplex, analog-to-digital conversion and recording equipment necessary for computer processing of speech data. From this computer, the processing of a first approximation model for a formant vocoder was advanced. A current sequel to this work is the construction of an experimental model formant vocoder speech compression system.

Under a recognition study, rules were developed which permitted the identification of consonant-vowel segments from data provided by the formant tracks and the gross shape of the unvoiced spectrum. These rules were applied to a limited vocabulary word recognition study. Current efforts in speech recognition center on isolating the word from continuous speech. This comprises segmentation of the speech wave into manageable units that can be compared to a stored vocabulary of sounds. These units can then be compared to a higher order dictionary to arrive at the word boundaries by linguistic analysis. It is also anticipated that the segmented units will lend themselves to a speech analysis-synthesis compression system requiring an information rate on the order of 200 bits per second.

In conjunction with the effort towards automatic recognition of speech, an effort is being devoted to the investigation of speech parameters which reflect the state and identity of the speaker.

This work has been partially supported by the U. S. Army Electronics Research and Development Laboratory, Fort Monmouth, N. J.

5.2.11 INTERNATIONAL BUSINESS MACHINES CORPORATION

Advanced Systems Development Division, San Jose, Calif.
W. E. DICKINSON, *Project Engineer*

Work on speech recognition continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Dersch, W. C. "A Decision Logic for Speech Recognition," in *Bionics Symposium. Living Prototypes—The Key to New Technology*. WADD Technical Report 60-600. Wright-Patterson Air Force Base, Ohio: Directorate of Advanced Systems Technology, Wright Air Development Division, A.R.D.C., U.S. Air Force, December 1960. Also published as IBM ASDD Technical Report 16.01.106.018, San Jose, September 1960.

(2) Dersch, W. C. "SHOEBOX, A Voice Responsive Machine," *Datamation*, June 1962, pp. 47-50.

5.2.12 京都大學 [KYOTO UNIVERSITY]

Department of Electrical Engineering, Kyoto, Japan
坂井利之
[TOSHIYUKI SAKAI]

The phonetic typewriter which was first designed for Japanese monosyllables was extended to conversational speech sound by adding a control system for conversational speech (1) (2). For the next step, the application of phonetic context to the system is intended. To get data for the design of the system, statistics of a three-phoneme sequence (trigram) of Japanese were obtained by the KDC-1 computer, and the distribution of these sequences was examined. On these bases work continues on the recognition system for conversational speech in which the influence of phonetic context on acoustical pattern is taken as the main principle, and decomposition to phoneme units (segmentation) and recognition of these units are jointly performed. The design based on this principle is finished, and the system is being constructed by the Nippon Electric Co. Ltd., Tokyo.

The project is supported by the Japanese Ministry of Education.

References:

(1) 坂井利之 [Toshiyuki Sakai]. 音声タイプライタ ["Phonetic Typewriter"], 科学 [*Kagaku*], May 1962.

(2) Sakai, Toshiyuki, and Shuji Doshita. "Phonetic Typewriter," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.

LITTON SYSTEMS, INC. 5.2.13

Communication Sciences Laboratory, 221 Crescent Street, Waltham, Mass.
GEORGE SEBESTYEN *and* WILLIAM FLOYD

Techniques are being investigated for producing a phonetic transcription of speech. Speech is represented parametrically and the parametric representation is compared with each of several hundred stored representations. An output symbol corresponding to the best match is produced, resulting in a coarse phonetically transcribed text. This is further processed to improve the transcription.

This effort is partially supported by the U. S. Air Force (Rome Air Development Center).

5.2.14

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Lincoln Laboratory, Lexington 73, Mass.
JAMES W. FORGIE *and* CARMA D. FORGIE

Work continues on the development of speech recognition programs for the TX-2 computer. The goal of the project is speech communication from man to machine. A program to classify stop consonants is now being tested. This program combines information from plosive burst spectra and formant transitions. Also under investigation is the use of formant transition cues in combinations with the fricative spectral cues used in the previously developed fricative recognition program. Other studies include a program to evaluate the effect on word recognition scores of certain errors at the phoneme recognition level.

5.2.15

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Research Laboratory of Electronics, Cambridge 39, Mass.
KENNETH N. STEVENS, *Principal Investigator*

Research on the generation, analysis, and perception of speech is

being continued. The work on speech generation includes theoretical investigations of the acoustics of speech production (11), studies of the motions of the vocal tract during speech production, and experimental studies with electrical analog speech synthesizers (4) (5) that have been developed as part of the research program. The speech analysis work includes (a) examination of methods for automatically extracting signals with low information rates from the speech wave (3)(6)(10), and (b) the accumulation of data on the properties of these signals for speech sounds corresponding to various allophones of American English and the relating of these data to articulatory events (2)(7-8). The analysis and some synthesis studies are carried out on the TX-O digital computer (1). Studies dealing with the processing of speech-like signals by human listeners are also in progress (9).

The work is supported by the U. S. Air Force and the National Science Foundation.

References:

(1) Dennis, J. B. "Computer Control of an Analog Vocal Tract," in *Proceedings of The Fourth International Congress on Acoustics,* Copenhagen, August 21-28, 1962. (In press)

(2) Fujimura, O. "Formant-antiformant Structure of Nasal Murmurs," in *Proceedings of The Fourth International Congress on Acoustics,* Copenhagen, August 21-28, 1962. (In press)

(3) Halle, M., and K. N. Stevens. "Speech Recognition: A Model and a Program for Research," *IRE Transactions on Information Theory,* vol. IT-8, 1962, pp. 155-159.

(4) Hecker, M. H. L. "Studies of Nasal Consonants with an Articulatory Speech Synthesizer," *Journal of the Acoustical Society of America,* vol. 34, 1962, pp. 179-188.

(5) Hecker, M. H. L., A. S. House, and K. N. Stevens. "Performance of the Articulatory Analog of the Speech Mechanism: A Status Report," *Quarterly Progress Report No. 64,* Research Laboratory of Electronics, M.I.T., January 15, 1962.

(6) Heinz, J. M. "An Analysis of Speech Spectra in Terms of a Model of Articulation," in *Proceedings of The Fourth International Congress on Acoustics,* Copenhagen, August 21-28, 1962. (In press)

(7) House, A. S., A. P. Paul, K. N. Stevens, and Jane B. Arnold. "Acoustical Description of Syllabic Nuclei: Data Derived by Automatic Analysis Procedures," in *Proceedings of The Fourth International Congress on Acoustics,* Copenhagen, August 21-28, 1962. (In press)

(8) House, A. S., K. N. Stevens, and A. P. Paul. "Acoustical Description of Syllabic Nuclei: An Interpretation in Terms of a Dynamic Model of Articulation," in *Proceedings of The Fourth International Congress on Acoustics,* Copenhagen, August 21-28, 1962. (In press)

(9) House, A. S., K. N. Stevens, T. T. Sandel, and Jane B. Arnold. "On the Learning of Speechlike Vocabularies," to be published in *Journal of Verbal Learning and Verbal Behavior,* vol. 1, 1962.

(10) Inomata, S. "Program for Active Segmentation and Reduction of Phonetic Parameters," in *Proceedings of The Fourth International Congress on Acoustics,* Copenhagen, August 21-28, 1962. (In press)

(11) Stevens, Kenneth N., and Arthur S. House. "An Acoustical Theory of Vowel Production and Some of its Implications," *Journal of Speech and Hearing Research,* vol. 4, 1961, pp. 303-320.

MOTOROLA, INC. 5.2.16

Military Electronics Division, Chicago, Ill.
EVAN L. RAGLAND

The feasibility study of a new technique for conversion of digital words into vocal words continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Ragland, Evan L. "Digital-to-Voice Conversion," in *Computers: Key to Total Systems Control,* pp. 135-146. New York: The Macmillan Co., 1961.

NEW YORK UNIVERSITY 5.2.17

Data Processing and Computation Laboratory,
College of Engineering, New York 53, N. Y.
LOUIS J. GERSTMAN, *Principal Investigator*

A project is underway to demonstrate the feasibility of real-time machine recognition of finite vocabularies. The approach seeks not to identify individual phonemes, but rather to extract from an †"unknown" word only as much information as is required to assign it uniquely to a previously stored classification scheme.

† *See* Glossary.

The method is quasi-spectral, deriving from the statistics of filtered, digitized waveforms indices of formant movement as a function of grossly quantized time. These indices, rather than the original waveforms, are economically stored in the ALWAC III-E computer, and form the comparison set against which the statistics for an incoming "unknown" word are tested. The computer prints out the name of the closest matching stored word.

The indices for the stored words may be examined with a view to replacing words that are too close to other words by words more distant in the measurement space. In this connection, preliminary data suggest that the words of the ICAO alphabet (Alpha, Bravo, etc.) require no such retailoring. Since the method treats speech in units as gross as the obligatory articulatory gestures which produce speech, there is good reason to hope that the scheme will not be talker-specific.

The project was recently transferred from the Bulova Research & Development Laboratories under whose heading the work had been previously reported in *CRDSD,* No. 10, Statement No. 5.2.5.

5.2.18 PHILCO CORPORATION

Scientific Laboratory, Blue Bell, Pa.
C. F. TEACHER *and* L. R. FOCHT

A basic study of the factors involved in human perception of speech is being conducted to establish the specifications of the invariant features of spoken words. These invariant features will be used to effect recognition of phonemes by multi-stage correlation techniques. This work is progressing toward the general goal of word recognition by means of recognized phoneme sequences. The analysis and extraction of the invariant speech components are expected to be completed shortly.

5.2.19 RADIO CORPORATION OF AMERICA

Surface Communications Division, Camden, N. J.
WILLARD F. MEEKER

A study has been made to determine criteria for automatic recognition of speech sounds.

Input speech was analyzed with an 18-channel filter. The rectified filter outputs were smoothed, sampled 50 times per second, and quantized to 8 levels in 3.5 db steps. The resulting digital data were recorded on magnetic tape and served as input to an RCA 501 computer used for display and analysis of data and evaluation of recognition criteria. Programs for automatic segmentation and phoneme recogni-

tion were developed and operated with moderate accuracy. The study dealt only with monosyllables formed by all combinations of 15 initial consonants. 10 vowels, and the final consonant /d/. For 10 male talkers and with initial consonants limited to 10, the average accuracy for vowels was 45 percent; for consonants, 60 percent.

The study was carried out under a U. S. Air Force (Aeronautical Systems Division, Wright-Patterson Air Force Base) contract.

References:

(1) Meeker, W. F., and L. S. Green. *Voice to Teletype Code Converter Research Program, Part I. Experimental Verification of a Method to Recognize Phonetic Sounds,* ASD Technical Report 61-666, Part I, December 1961. (ASTIA No. AD—272 029)

(2) Meeker, W. F., A. L. Nelson, and P. B. Scott. *Voice to Teletype Code Converter Research Program, Part II. Experimental Verification of a Method to Recognize Phonetic Sounds.* (In preparation)

電波研究所 5.2.20

[RADIO RESEARCH LABORATORIES]

Research Section of Information Processing,
Koganei-shi, Tokyo, Japan
YOSHIHARU OGATA, *Head of Section, and* KAZUO NAKATA,
Principal Investigator

Research on the analysis and recognition of speech is underway. The goal is to develop an efficient and natural communication link between man and computer.

The method of formant frequency extraction from the digitalized spectrum pattern was recently established, and its accuracy and source of errors were checked by a synthetic spectrum and applied to natural speech.

A preliminary approach to the automatic recognition of Japanese monosyllables is now being planned and tested.

This project is carried on with the cooperation of the Nippon Electric Company Ltd., Tokyo. The NEAC 1103 computer and its speech data translator are being used in this research.

Reference:

(1) Nakata, Kazuo, Jouji Suzuki, and Ken-ich-i Maesono. "Speech Analysis by Computer," *Journal of the Radio Research Laboratories,* vol. 9, no. 43, May 1962, pp. 277-291. (In English)

5.2.21 ROYAL INSTITUTE OF TECHNOLOGY

Stockholm 70, Sweden

GUNNAR FANT, *Head, Speech Transmission Laboratory*

Methods of acoustic specification of speech with application to efficient coding of speech data are being investigated to gain useful information for the design of analysis-synthesis telephony and speech recognition systems.

A new spectrograph and associated equipment for speech data processing are under construction (1) (2). Systematic factors affecting formant amplitude measurements have been investigated (3). The perceptual significance of physical sound segment boundaries has been investigated (4). A special study of the relative role of noise spectrum, formant transitions, and noise intensity as cues for fricative identification has been undertaken (3). In the forthcoming work, special attention will be given to the further development of synthesis strategies.

References:

(1) Fant, G. *Speech Analysis and Synthesis,* Royal Institute of Technology, Division of Telegraphy-Telephony, Speech Transmission Laboratory, Report No. 26, June 1962.

(2) Speech Transmission Laboratory. *Quarterly Progress and Status Report,* No. 1/1962 (January-March), April 15, 1962.

(3) Speech Transmission Laboratory. *Quarterly Progress and Status Report,* No. 2/1962 (April-June), July 15, 1962.

(4) Öhman, Sven. *On the Perception of Swedish Consonants in Intervocalic Position,* Thesis work for fil.lic. at the University of Uppsala. Also publ. as Report No. 25, Royal Institute of Technology, Division of Telegraphy-Telephony, Speech Transmission Laboratory, March 1962.

5.2.22 SYLVANIA ELECTRONIC SYSTEMS

A Division of Sylvania Electric Products Inc., Applied Research Laboratory,
40 Sylvan Road, Waltham 54, Mass.

HAROLD J. MANLEY, *Principal Investigator*

The research study to determine the validity of the coefficients of an appropriate set of orthonormal functions as a simplified representation of speech sounds for use in machine speech recognition has been completed. A digital computer simulation was conducted in which a sample of connected speech was analyzed and resynthesized in terms of a series of orthogonalized exponentially damped sinusoids. It was

found to be possible to synthesize each pitch period from a function set having 16 fixed frequencies with fixed damping at each frequency. The processing was done pitch synchronously on an IBM 7090 computer using 10-bit accuracy speech which had been digitized at a rate of 12,000 samples per second. The results indicate that both the phonetic content and the quality of the speaker's voice are retained in the analysis-synthesis process.

The study was supported by the U. S. Air Force (Rome Air Development Center).

References:

(1) Sylvania Electronic Systems, Applied Research Laboratory. *Optimum Speech Signal Mapping Techniques,* Final Report, RADC-TR-62-3, Contract AF 30 (602)-2446, January 10, 1962. (ASTIA No. AD–273 443)

(2) Manley, H. J., and D. B. Klein. "Analysis-Synthesis of Continuous Speech in Terms of Orthogonalized Exponentially Damped Sinusoids," *Journal of The Acoustical Society of America,* vol. 34, no. 5, May 1962, p. 724. (abstract)

THOMPSON RAMO WOOLDRIDGE INC. 5.2.23

RW Division, Canoga Park, Calif.
S. BERTRAM *and* PAUL L. GARVIN

Work on the problems of speech recognition continues, with no change reported since the previous statement [*Ed.*].

UNIVERSITÄT BONN 5.2.24

Institut für Phonetik und Kommunikationsforschung, Koblenzer Strasse 98a, Bonn, Federal Republic of Germany
G. HEIKE, H. SCHNELLE, *and* Hw. SENDHOFF

Two opposite approaches to segmentation of speech signals and recognition of audio-linguistic units are being tried: (a) formal morpho-phonemic analysis. The descriptive methods of formal syntax will be used for morpho-phonemic analysis. The combinatory restrictions on combinations of German phonemes within semantic and syntactic classes of words will be described. It is believed that automatic speech recognition of sounds independent of linguistic context cannot be successful for normally spoken language unless this language is clearly pronounced in an unnatural manner; and (b) construction of electronic apparatus. Construction of electronic apparatus will at first be mainly concerned with extraction of statistical information, e.g., the distribution of signal amplitudes or, in particular, frequency of zero crossings, etc.

5.2.25 UNIVERSITY COLLEGE LONDON

Department of Phonetics, Gower Street, London W. C. 1, England
D. B. FRY, *Head of Department*

Work continues on mechanical speech recognition, with no change reported since the previous statement [*Ed.*]. Support is provided by the Ministry of Aviation and the U. S. Air Force.

References:

(1) Denes, P., and M. V. Mathews. "Spoken Digit Recognition Using Time-Frequency Pattern Matching," *Journal of the Acoustical Society of America,* vol. 32, 1960, pp. 1450-1455.
(2) Denes, P., and J. Milton-Williams. "Further Studies in Intonation," *Language and Speech,* vol. 5, 1962, pp. 1-14.

5.2.26 UNIVERSITY OF EDINBURGH

Department of Phonetics, Minto House, Chambers Street, Edinburgh, Scotland
DAVID ABERCROMBIE, *Head of Department*

Research on the acoustic, physiological, and auditory analysis of speech continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Uldall, E. T. "Ambiguity: Question or Statement," in *Proceedings of the Fourth International Congress of Phonetic Sciences.* Helsinki: Mouton and Co., 1961.

5.2.27 THE UNIVERSITY OF MICHIGAN

Communication Sciences Laboratory, Ann Arbor, Mich.
GORDON E. PETERSON, *Director*

Fundamental research is conducted on the structure of spoken language, particularly English. The research involves the development of basic speech analyzing instrumentation, including special methods of signal recording, a sound spectrograph for speech analysis, a fundamental voice frequency analyzer, and speech formant analyzing circuits. Basic studies are conducted on the physiology of speech production, the acoustical analysis and synthesis of speech, automatic speech recognition, and grammatical analysis.

Support is provided by the U. S. Office of Naval Research, the U. S. Air Force Office of Scientific Research, and the National Institutes of Health

References:

(1) Lehiste, I., and G. E. Peterson. "Some Basic Considerations in the Analysis of Intonation," *Journal of the Acoustical Society of America,* vol. 33, no. 4, April 1961, pp. 419-425.

(2) Lehiste, I., and G. E. Peterson. "Transitions, Glides, and Diphthongs," *Journal of the Acoustical Society of America,* vol. 33, no. 3, March 1961, pp. 268-277.

(3) Peterson, G. E. "Automatic Speech Recognition Procedures," *Language and Speech,* vol. 4, part 4, October-December 1961, pp. 200-219.

(4) Peterson, G. E. "Parameters of Vowel Quality," *Journal of Speech and Hearing Research,* vol. 4, no. 1, March 1961, pp. 10-29.

(5) Peterson, G. E., and C. J. Fillmore. "The Theory of Phonemic Analysis," in *Proceedings of the Fourth International Congress of Phonetic Sciences,* pp. 112-125. Helsinki: Mouton and Co., 1961.

(6) Peterson, G. E., and F. Harary. "Foundations of Phonemic Theory," in *Proceedings of Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects,* ed. by Roman Jakobson, pp. 139-165. Providence, R. I.: American Mathematical Society, 1961.

(7) Peterson, G. E., and I. Lehiste. "Revised CNC Lists for Auditory Tests," *Journal of Speech and Hearing Disorders,* vol. 27, no. 1, February 1962, pp. 62-70.

(8) Peterson, G. E., and Norris P. McKinney. "The Measurement of Speech Power," *Phönetica,* vol. 7, no. 2/3, 1961, pp. 65-84.

(9) Shoup, J. E. "Phoneme Selection for Studies in Automatic Speech Recognition," *Journal of the Acoustical Society of America,* vol. 34, no. 4, April 1962, pp. 397-403.

(10) Sivertsen, E. "Segment Inventories for Speech Synthesis," *Language and Speech,* vol. 4, part 1, January-March 1961, pp. 27-89.

(11) Wang, W. S-Y., and C. J. Fillmore. "Intrinsic Clues and Consonant Perception," *Journal of Speech and Hearing Research,* vol. 4, no. 2, June 1961, pp. 130-136.

THE UNIVERSITY OF ROCHESTER **5.2.28**

Department of Languages and Linguistics, Rochester 20, N. Y.

DEAN H. OBRECHT, *Principal Investigator*

Investigation of perception cues embedded in the speech spectrum of Arabic continues, with no change reported since the previous statement [*Ed.*]. The study is supported by the National Science Foundation.

5.3 *LINGUISTIC AND LEXICOGRAPHIC RESEARCH*

The linguistic and lexicographic research reported here will contribute to general knowledge of language and therefore perhaps aid in research on information systems and mechanical translation. The handling of natural language offers an interesting possibility for computer applications of the future. Computers now cannot process ordinary technical prose as effectively as humans; they can receive and process only carefully prepared data. The introduction of information into a computer is difficult; the difficulty is accepted for some purposes only because of the gains that come from mass and rapid processing of the information. Research efforts are directed to acquiring large quantities of machine-usable text with which to develop and test new techniques; to developing such techniques for recognizing and manipulating the structures of natural language, e.g., parsing sentences (research in mechanical translation described in Section 3 shows progress in achieving this goal; see also the development and uses of computer languages for symbol manipulation described in Subsection 5.4); and to extracting the meaning from natural-language expressions to produce a formalized version of the input expressions.

The following summary calls attention to new work reported in this subsection for the first time, to two projects which have been completed, and to closely related projects which appear in other sections of the report.

The amount of new work being reported has significantly increased over the previous issue. Fifteen new projects from six different countries are reported here for the first time. In this new work, the computer as an aid to linguistic analysis and research is much in evidence. Some of the projects are being conducted with the ultimate ends of information retrieval and mechanical translation in mind.

Phonological analysis of several languages is being carried out at **Brown University** (5.3.3), utilizing an IBM 7070 computer, with a view to developing a method for an exact typological classification and comparison of languages in their phonological aspect.

At **Tuskegee Institute** (5.3.28) a study is being conducted to determine the statistical moments of words and their correlations between a source and target language.

The **University of Pittsburgh** (5.3.38) has an investigation underway into the feasibility of developing a programming language compatible with the needs of psychologists and educators in retrieval and analysis of natural-language material.

A CDC 1604 computer is being used in a project at the **University of Wisconsin** (5.3.39) to statistically analyze the Etruscan language.

As part of a program aimed at the automation and mechanization

of auxiliary work in linguistic research, the **Československá Akademie Věd** (5.3.5) has initiated a project to automate the files used for compiling the Dictionary of Contemporary Standard Czech.

The problems of the relationship between form and function in natural languages are being studied by means of the axiomatic method at **Karlova Universita** (5.3.17).

Universität Bonn (5.3.29) is engaged in several projects on automatic linguistic analysis, semantic analysis, word segmentation, mathematical linguistics, and theory of programming.

Humboldt-Universität zu Berlin (5.3.10) is conducting statistical research on syntax and lexical items in German, is developing an intermediate language for multilingual MT, and is studying German syllable structure for possible speech automation and acoustical MT input.

Forskningsgruppen för Kvantitativ Lingvistik (KVAL) (5.3.7) is studying vocabulary and syntactic structure of Swedish with a view toward future application to MT research.

Matematikmaskinnämnden (5.3.23) has a project underway to develop computer programs to insert corrections, additions, and deletions in a punched text, yielding a tape with updated text that can be fed into a teletypesetter reading unit.

Work is in progress at the **University of Gothenburg** (5.3.35) on mechanical linguistic analysis of English text using teletypesetter tapes and an ALWAC computer.

IBM-France (5.3.16) is doing a statistical analysis of written French in order to determine an automatic encoding scheme which will reduce overloading of communication channels.

Independent work (5.3.43) in Germany on syntax analysis of Japanese chemical texts is aimed at eventual Japanese-English MT.

In Rumania, **independent research** (5.3.41) is being conducted on the possibility of processing a scientific text with the aid of a digital computer for (a) revealing the synonymy of expressions, and (b) recognizing the logical deducibility of certain expressions from others within the limits of the same text.

Two projects, one being reported here for the first time, have been completed: the Indiana University (5.3.12) project involving the exploration of a number of basic psycholinguistic hypotheses and the **University of Wisconsin** (5.3.40) investigation, completed in 1961, into the etymological composition of English and the frequency of occurrence of phonemes and phoneme sequences in American English with the aid of a computer.

The reader's attention is directed to the work appearing in other sections on simulating human language behavior (System Development Corporation, 5.4.24, and the University of Texas, 5.4.34), on recogniz-

ing and distinguishing language patterns and their meaning (University of Michigan and System Development Corporation, 5.1.44, and the Institute of Cybernetics [formerly Computing Center] of the Academy of Sciences Ukrainian S.S.R., 5.4.9), and on psycholinguistics (University of Michigan, 5.5.9).

5.3.1 AMERICAN BIBLE SOCIETY

450 Park Avenue, New York 22, N.Y.
EUGENE A. NIDA

Investigation of all types of foreign-language translation, with emphasis principally upon translations of the Bible, continues, with no change reported since the previous statement [*Ed.*].

5.3.2 ARTHUR D. LITTLE, INC.

35 Acorn Park, Cambridge 40, Mass.
VINCENT E. GIULIANO *and* PAUL E. JONES, *Investigators*

Work continues on a realizable natural-language system for the retrieval of command and control data (see also 5.3.6), with no change reported since the previous statement [*Ed.*]. The investigation is sponsored by the U. S. Air Force Systems Command.

References:

(1) Arthur D. Little, Inc. *Studies for the Design of an English Command and Control Language System,* Report ESD-TR-62-45; also CACL-1.

(2) Giuliano, V. E., and P. E. Jones. "Linear Associative Information Retrieval." (To be published)

5.3.3 BROWN UNIVERSITY

Department of Linguistics, Providence, R. I.
HENRY KUCERA, *Principal Investigator*

Phonological analysis of several languages is being carried out by means of certain procedures of communication theory and other mathematical techniques with the view of developing a method for an exact typological classification and comparison of languages in their phonological aspect and, ultimately, for investigating the dynamism of historical linguistic change.

The languages which have been under investigation during the past year are Russian, Czech, and German. A complete statistical analysis of representative texts consisting of approximately 100,000 phonemes of each language has been completed. Syllabic structure, entropy and redundancy on the syllabic level, a so-called isotopy index, and functional load of the relevant phonological oppositions have been determined for each language.

The IBM 7070 data processing system is used in this research. Data processing equipment is used in all stages of the project, including phonemic transcription of the literary texts, syllabic analysis, and the various statistical counts.

A similar analysis of other languages is planned.

Reference:

(1) Kucera, Henry. "Statistical Determination of Isotopy," in *Preprints of Papers for the Ninth International Congress of Linguists,* Cambridge, Mass., 1962, pp. 338-346.

CENTRO PER L'AUTOMAZIONE DELL'ANALISI LETTERARIA 5.3.4

[Literary Data Processing Center]

Pont. Facolta di Filosofia, Aloisianum, Gallarate (Varese), Italy

REV. ROBERTO BUSA, *S.J., Director*

Work continues on linguistic analysis by automation, with no change reported since the previous statement [*Ed.*].

References:

(For a complete list of earlier publications by Rev. Roberto Busa, *S.J.*, concerning the automation of linguistic analysis, see *CRDSD,* No. 9, Statement No. 5.3.2.)

ČESKOSLOVENSKÁ AKADEMIE VĚD 5.3.5

[CZECHOSLOVAK ACADEMY OF SCIENCES]

Institute of the Czech Language, Department for Research in Contemporary Czech, Letenská 4, Malá Strana, Prague 1, Czechoslovakia

JITKA ŠTINDLOVÁ

As part of a program aimed at the automation and mechanization of auxiliary work in linguistic research, a project is being initiated to automate, by mechanographical indexing, the files used for the compilation of the Dictionary of Contemporary Standard Czech (the published installments of the dictionary comprise the letters A to P). The realization of the project will provide the basis for the editing work, e.g., unification and supplementing of items, cross-references, etc. It will also be of help in supplementing the first edition of the dictionary, and especially in preparing the second edition. At the same time it will furnish a broad material basis for lexical, grammatical, and stylistic research, a basis necessary for both practical purposes and theoretical study. The mechanical processing of the dictionary will also be useful in performing some newly emerging tasks, e.g., a precise, mechanically established classification of items according to word categories and morphological features will provide for the essential

prerequisites for using the vocabulary of the dictionary (about 250,000 items) for machine translation.

An evaluation is being made of an experiment in excerpting words and terminological word combinations (this is regarded as the most topical task to be served by the materials collected in the dictionary).

During the next year an attempt will be made at complex indexing of the dictionary items.

At the same time, work is being done on the technical problem of designing mechanized equipment which will adequately handle the Czech alphabet. Altogether 64 symbols are needed, or, if the Slovak alphabet is to be also considered, 67.

One of the first tasks of the effort directed toward the mechanization and automation of auxiliary work in linguistics will be the checking of the mechanographic processing of the materials contained in the archives of specialized terminology. (The archives serve mainly the lexicographic and consulting practice of the Institute of the Czech Language.) The use of the punched cards will make possible the regrouping of the items registered on the cards according to notional, alphabetic, or linguistic viewpoints.

By way of experiment, specialized terms included in the Czechoslovak State Norms are being processed.

At the same time, the materials contained in the terminological archives are being compared with the terminological excerpts drawn from the Dictionary of Standard Czech, parts of which have already been published.

The questions concerning this project were dealt with at the Besançon colloque international sur la mécanisation des recherches lexicologiques.

References:

(1) Štindlová, Jitka. "Stroje na zpracování informací a jejich význam pro jazykovédu" ["Information Processing Machines and Their Importance for Linguistics"], *Slovo a slovesnost,* 1961, pp. 208-215.

(2) Štindlová, Jitka. "Využití technických prostředků mechanizace a automatizace při organizaci archívních sbírek odborného názvosloví" ["The Utilization of the Technique of Mechanization and Automatization in Organizing the Archival Collection of Specialized Terminology"], *Naše řeč* (Praha), 1961, pp. 28-32.

5.3.6 CHARLES W. ADAMS ASSOCIATES, INC.

142 The Great Road, Bedford, Mass.
MURRAY E. SHERRY

The investigation of a realizable natural-language system for the

retrieval of command and control data continues, with no change reported since the previous statement [*Ed.*]. The project is being conducted under contract with Arthur D. Little, Inc. (see 5.3.2), with funds provided by the U. S. Air Force Systems Command.

5.3.7
FORSKNINGSGRUPPEN FÖR KVANTITATIV LINGVISTIK (KVAL)

c/o *Språkförlaget Skriptor, Stockholm 40, Sweden*
HANS KARLGREN

In linguistic research underway, a study is being made of vocabulary and syntactic structure on an empirical, statistical basis, with a view to future application to mechanical translation research.

The material used thus far has been conversational Swedish, transcribed for this purpose from tape recordings. The method in itself, however, can and will be applied to written material as well. The method is based on segmentation of the text into morphemes, the smallest units of syntactical structures. These morphemes have been arranged in distributional classes. The string of class morpheme class symbols is treated as a simplified text, containing the relevant syntactic information. The analysis of these strings is carried out, to a considerable extent, with the aid of computers. The processing also gives frequency counts for morpheme classes and individual morphemes, word stems, and inflected forms of words.

When the method has been sufficiently refined, these vocabulary and syntax counts and analyses will be carried out on a large scale.

The EDP part of the job has been done on the "Besk" computer (the MNA prototype computer, which has been slightly modified and manufactured for sale as "Facit" by the Swedish "Atvidabergs" industries).

In the last year, considerable work has been done to lay down segmentation and classification principles and manually prepare the first set of texts. The main machine programs have been written, but because of the need for many extensions of the original plans, programming is presently in an intermediate stage. Theoretical mathematical treatment of the methods has been given much time, but has not been completed because conclusive empirical data are not yet available. The method, as a whole, has shown good prospects for development in several directions.

The theoretical discussion has made more advanced computer analysis of the sequences desirable. Also, more written literature will be treated, as well as spoken language.

The KVAL group operates independently, with no financial support,

except for computer time granted by MNA, and some minor aid from the Stockholm and Upsala Universities. Costs have been defrayed by the members of the working group themselves.

In the near future two publications entitled "Talspraksstudier" (Studies of Conversational Swedish) will appear. The statistical methodology will be treated in the journal *Statistical Methods in Linguistics* to be published by Språkförlaget Skriptor, Stockholm.

5.3.8 GENERAL ELECTRIC COMPANY

Computer Department, Phoenix, Ariz.
L. W. GOOSTREE, JR., *Manager-Marketing*

Development of various analytic routines for investigating the linguistic structure of scientific and technical literature continues, with no change reported since the previous statement [*Ed.*].

5.3.9 HEBREW UNIVERSITY

Applied Logic Branch, Jerusalem, Israel
YEHOSHUA BAR-HILLEL, *Head of Group*

Investigation of the formal properties of †sequential languages and of various formal notions of syntactic complexity continues, with no change reported since the previous statement [*Ed.*]. The work is supported by the U.S. Office of Naval Research.

References:

(1) Bar-Hillel, Y., M. Perles, and E. Shamir. "On Formal Properties of Simple Phrase Structure Grammars," *Zeitschrift fuer Phonetik, Sprachwissenschaft und Kommunikationsforschung,* vol. 2, no. 1, January 1962, pp. 143-172.

(2) Shamir, E. *On Sequential Grammars,* Technical Report No. 7 (prepared under Contract N62558-2214 for the Information Systems Branch, U.S. Office of Naval Research), Jerusalem, November 1961.

(3) Bar-Hillel, Y. "Some Recent Results in Theoretical Linguistics," in *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science,* pp. 379-385. Stanford, Calif.: Stanford University Press, 1962.

(4) יהושע בר-הלל [Yehoshua Bar-Hillel]. סיבוכו של פסוק מהו? סוגיה ["What is a Complex Sentence]" בבלשנות הקומבי נאטורית [*Study in Combinatorial Linguistics*], לשוננו [*Our Language*], vol. 25, no. 3-4, 1961, pp. 150-164.

† *See* Glossary.

HUMBOLDT-UNIVERSITÄT ZU BERLIN 5.3.10

Institut für Phonetik und Kommunikationswissenschaft,
Universitätstrasse 7, Berlin W. 8 (Eastern Sector), Germany
GEORG F. MEIER, *Director*

Beginning in the autumn of 1961, various work in applied linguistics was accelerated. Among other special investigations, the following research has been started.

Studies are being made of statistical research on syntax and lexical items of the German language for the purpose of preparing Markov-chains for mechanical synthesis.

An optimal intermediate language for multilingual MT is being developed by comparing the grammatical categories and position types of 47 languages. The studies are directed primarily toward algorithms of the grammatical information for synthesis of any target language without direct information from the source language.

Investigations for a polyglottic vocabulary on the basis of exact definitions for nonscientific terms are being advanced. A special system of ciphered hierarchy (with 8 classes, 8 subclasses, etc.) is being prepared and will be preliminarily programmed for computer in the beginning of 1963. Investigations are being made for theoretical purposes (especially to find out minimum semantic units) and for practical purposes (in order to gather materials for an intermediate Language Lexicon for MT and for logical processes on computers).

In communication research, human decoding processes are being investigated. These investigations are conducted on the feedback mechanisms of the brain during decoding on the different speech levels. Especially, additional redundancy by memory feedback is being studied.

Using a collection of linguistic bibliography (at present about 300,000 titles), the possibility of mechanical extracts is being investigated.

For speech automation and acoustical MT input, German syllable structure is being studied by different methods of experimental phonetics.

INDIANA UNIVERSITY 5.3.11

Box 70, Bloomington, Ind.
F. W. HOUSEHOLDER, *Chairman of the Committee on Linguistics, and* J. P. THORNE, *Research Associate*

Research is being carried out on the mechanical analysis of natural language (scientific English) with a view to determining the most efficient means of storing a large corpus of natural-language text so that information relevant to particular questions can be retrieved from it.

Work is underway on construction of FLEX, an artificial language of minimum syntactic organization and maximum semantic organization.

Essentially this is a device to facilitate the weighting of partial semantic matches.

Programs previously constructed for use on the IBM 650 computer are currently being changed for use on the IBM 709 computer.

A reverse-order English dictionary containing some 80,000 entries has also been produced.

Clause routines and ambiguity routines are now in semifinal form. Work is beginning on procedures for semantic analysis, on the improvement of FLEX, and on easily calculable measures of semantic agreement between sentences.

The research is sponsored by the U. S. Air Force (Rome Air Development Center).

References:

(1) Householder, F. W., Jr., and J. P. Thorne. *Seventh Quarterly Report on Automatic Language Analysis,* Indiana University, March-May, 1962. Contract AF 30 (602)-2185.

(2) Michie, Donald. *U. S. Air Force Project for the Mechanical Analysis of Language: Report on Consultations with Mr. J. P. Thorne,* Indiana University, July 1962. Contract AF30 (602)-2185.

5.3.12 INDIANA UNIVERSITY

Research Center in Anthropology, Folklore, and Linguistics, Bloomington, Ind.

THOMAS A. SEBEOK, *Chairman*

The project involving exploration of a number of basic psycholinguistic hypotheses has been completed. The project was sponsored by the National Science Foundation.

References:

(For additional earlier publications in this field, see *CRDSD,* No. 9, Statement No. 5.3.9.)

(1) Saporta, S., ed. *Psycholinguistics: A Book of Readings.* New York: Holt, Rinehart and Winston, 1961.

(2) Sebeok, T. A. "The Informational Model of Language: Analog and Digital Coding in Animal and Human Communication," in *Natural Language and the Computer,* ed. by Paul L. Garvin. New York: McGraw-Hill Book Co., Inc. (In press)

(3) Sebeok, T. A., and V. J. Zeps. "Computer Research in Psycholinguistics: Towards an Analysis of Poetic Language," *Behavioral Science,* vol. 6, 1961, pp. 365-369.

(4) Sebeok, T. A. "Notes on the Digital Calculator as a Tool for Analyzing Literary Information," in *Poetics,* pp. 571-592. Warsaw: Literary Research Institute, Polish Academy of Sciences, 1961.

(5) Sebeok, T. A., and V. J. Zeps. *Janua Linguarum, Series Maior, No. 8: Concordance and Thesaurus of Cheremis Poetic Language.* The Hague: Mouton & Co., 1961.

(6) Sebeok, T. A. "Materials for a Typology of Dictionaries," to be published in *Lingua,* vol. 11, 1962.

(7) Sebeok, T. A., ed. *Style in Language.* New York and London: The Technology Press of Massachusetts Institute of Technology and John Wiley & Sons, Inc., 1960.

(8) Sebeok, T. A. "The Texture of a Cheremis Incantation," *Mémoires de la Société Finno-Ougrienne.* (In press)

(9) Sebeok, Thomas A. "Coding in the Evolution of Signalling Behavior," *Behavioral Science,* vol. 7, no. 4, October 1962, pp. 430-442.

INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION 5.3.13

Conservatoire National des Arts et Métiers, Paris, France

ERIC DE GROLIER[1]

The study of present trends in modern linguistics, particularly structural linguistics and semantics, from the point of view of their use in the better design of information retrieval systems continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the U. S. Air Force Office of Scientific Research.

Reference:

(1) de Grolier, Eric. *Etude sur les Categories Generales Applicables aux Classifications et Codifications Documentaires.* UNESCO, Comité Consultatif international pour la documentation et la terminology dans les sciences pures et appliquées, March 1960.

[1] Present address: Conseil International des Sciences Sociales, 6 rue Franklin, Paris 16e, France.

5.3.14

ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ, ИНСТИТУТ ЯЗЫКОЗНАНИЯ АН СССР
[INSTITUTE OF LINGUISTICS, LENINGRAD DEPARTMENT, ACADEMY OF SCIENCES U.S.S.R.]

Group on Mathematical Linguistics, Leningrad, U.S.S.R
Н. Д. Андреев [N. D. ANDREEV], *Leader*

The independent Group on Mathematical Linguistics is working on the testing of statistico-combinatorial models with a dozen languages: Russian, English, French, German, Bulgarian, Ukrainian, Italian, Hausa, Swahili, Indonesian, Vietnamese, Bengali, Hindi, Czech, Burmese, and Korean. The algorithm for statistico-combinatorial modeling works without any use of word and phrase meanings and without any kind of preliminary grammatical information. Only texts (divided into phrases and words) are given and investigated within a frame of strictly determined programs which can be fully executed through the use of computers. The first results have already been obtained: the algorithm for modeling has produced four of the Russian, two of the English, two of the French, two of the German, two of the Hausa, two of the Bulgarian, and one of the Ukrainian morphological types.

Work has been recently begun on a statistico-combinatorial investigation of a small semantic field; only Russian texts are considered. Statistico-combinatorial methods also are applied to obtaining a more precise picture of syntax: the functional classes of words are characterized by their probabilistic and combinatorial syntactic properties, which permit one to pass from a less exact system of classes to a more exact one. Such approximative analysis is made on materials of many languages: those named above (with the exception of Swahili, Czech, and Hausa), Spanish, Polish, Lettish, Chinese, Burmese, Estonian, Turkish, and Kirghiz.

References:

(1) Андреев, Н. Д. [Andreev, N. D.]. "Алгоритмы статистико-комбинаторного моделирования морфологии, синтаксиса, словообразования и семантики," Материалы по математической лингвистике и машинному переводу, т. 2, Ленинград, 1962.

["Algorithms of Statistical-combinatorial Modeling of Morphology, Syntax, Word Formation, and Semantics," *Materials on Mathematical Linguistics and Machine Translation,* vol. 2, Leningrad, 1962.]

(2) Andreev, N. D. "Model Construction as a Tool for Development of Linguistic Theory," *Word,* vol. 18, no. 1-2, June 1962,

(special issue on the occasion of the Ninth International Congress of Linguists).

5.3.15
ИНСТИТУТ МАТЕМАТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ АКАДЕМИИ НАУК СССР
[INSTITUTE OF MATHEMATICS, SIBERIAN BRANCH, ACADEMY OF SCIENCES U.S.S.R.]

Novosibirsk 72, U.S.S.R.

Э. В. Евреинов, Ю. Г. Косарев, В. А. Устинов
[E. V. EVREINOV, YU. G. KOSAREV, *and* V. A. USTINOV]

The application of electronic computers to investigations in the humanities continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Евреинов, Э. В., Ю. Г. Косарев, В. А. Устинов [Evreinov, E. V., Yu. G. Kosarev, and V. A. Ustinov]. "Вычислительная техника в историкофилологических исследованиях" ["Computing Technique in Historical-Philological Investigations"], Вестник Академии Наук СССР [*Herald of the Academy of Sciences U.S.S.R.*], no. 1, 1962, pp. 80-83.

(2) Евреинов, Э. В., Ю. Г. Косарев, В. А. Устинов [Evreinov, E. V., Yu. G. Kosarev, and V. A. Ustinov]. Исследование рукописей древних майя с помощью электронной математической машины: Том 1. Мадридская рукопись, 397 стр.; Том 2. Дрезденская рукопись, 370 стр.; Том 3. Систематизированный каталог иероглифических знаков, 400 стр. [*Investigation of the Ancient Maya Manuscripts With the Aid of an Electronic Computer*: Vol. 1. *Madrid Manuscript,* 397 p.; Vol. 2. *Dresden Manuscript,* 370 p.; Vol. 3. *Classified Catalog of Hieroglyphs,* 400 p.], Publishing House of the Siberian Branch of the Academy of Sciences USSR, Novosibirsk, 1961.

(3) Устинов, В. А. [Ustinov, V. A.]. "О применении электронных математических машин в исторической науке ["Application of Electronic Computers in the Historical Sciences"], Вопросы истории [*Problems of History*], no. 8, 1961. Translation in JPRS 11066, 13 November 1961, available from Office of Technical Services, Department of Commerce, Washington 25, D. C.

The following papers, presented at the Conference on Information Processing, Machine Translation and Automatic Text Reading, which

was sponsored by the Institute of Scientific Information of the Academy of Sciences U.S.S.R. in Moscow, January 24-31, 1961, have appeared as separates under the general cover title Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста [*Reports at the Conference on Information Processing, Machine Translation and Automatic Text Reading*]:

(4) Устинов, В. А. [Ustinov, V. A.]. Исследование рукописей древних майя с помощью электронной вычислительной машины (Анализ письменности) [Investigation of Ancient Maya Manuscripts with the Aid of an Electronic Computer (Analysis of the Written Language)]. 27 p.

(5) Евреинов, Э. В., Ю. Г. Косарев, В. А. Устинов [Evreinov, E. V., Yu. G. Kosarev, and V. A. Ustinov]. Исследование рукописей древних майя с помощью электронной вычислительной машины (Методы исследования) [Investigation of Ancient Maya Manuscripts with the Aid of an Electronic Computer (Methods of Investigation)]. 11 p. Translation in JPRS 10508, 16 October 1961, available from Office of Technical Services, Department of Commerce, Washington 25, D. C.

(6) Евреинов, Э. В., Ю. Г. Косарев, В. А. Устинов [Evreinov, E. V., Yu. G. Kosarev, and V. A. Ustinov]. Исследование рукописей древних майя с помощью электронной вычислительной машины (алгоритмы и программы) [Investigation of Ancient Maya Manuscripts with the Aid of an Electronic Computer (Algorithms and Programs)]. 9 p. Translation in JPRS 10508, 16 October 1961, available from Office of Technical Services, Department of Commerce, Washington 25, D. C.

(7) Евреинов, Э. В., Ю. Г. Косарев, В. А. Устинов [Evreinov, E. V., Yu. G. Kosarev, and V. A. Ustinov]. Исследование рукописей древних майя с помощью электронной вычислительной машины (предварительные результаты) [Investigation of Ancient Maya Manuscripts with the Aid of an Electronic Computer (Preliminary Results)]. 65 p. Translation in JPRS 10508, 16 October 1961, available from Office of Technical Services, Department of Commerce, Washington 25, D. C.

5.3.16 IBM-FRANCE

5, place Vendôme, Paris 1, France
R. MOREAU

A statistical analysis of written French has been started in order to determine an automatic coding scheme which will reduce overloading of the channels.

The results obtained thus far are related to the frequency of letters and groups of letters, the distribution of word length, and the redundancy of the language at these main levels. The work was done on the IBM 650 computer.

The Laboratory of Lexicological Analysis of Besançon (see 5.3.18) will cooperate in the continuation of the study. The phonetical structure of the language will be studied.

References:

(1) Moreau, R. "Sur la distribution des r-grammes en français," in *Comptes-rendus de l'Académie des Sciences,* vol. 252, pp. 3384-3385, meeting of May 29, 1961.

(2) Moreau, R. "Sur la distribution des unités lexicales dans le français écrit," in *Comptes-rendus de l'Académie des Sciences,* vol. 253, pp. 2626-2628, meeting of December 4, 1961.

(3) Moreau, R. "Quelques remarques en vue d'un codage automatique dans les télécommunications," *Automatisme,* November 1962.

KARLOVA UNIVERSITA [CHARLES UNIVERSITY] 5.3.17

Institute of Numerical Mathematics, Faculty of Mathematics and Physics, Linguistic Group, Krasnoarmějcu 2, Prague 1, Czechoslovakia

PETR SGALL, *Leader*

The problems of the relationship between form and function in natural languages are being studied by means of the axiomatic method. The principal question concerns the hierarchy that exists in the structures of form and function in connection with the levels of natural languages (1).

Reference:

(1) Nebeský, Ladislav, and Petr Sgall. "Vztah formy a funkce v jazyce" ["The Relationship Between Form and Function in the Natural Language"], *Slovo a slovesnost,* vol. 23, 1962, pp. 174 ff.

LABORATOIRE D'ANALYSE LEXICOLOGIQUE 5.3.18

Faculté des Lettres et Sciences Humaines, Université de Besançon, France

BERNARD QUEMADA, *Director*

Analyses of linguistic references concerning the national vocabulary continue, with no change reported since the previous statement [*Ed.*].

References:

(1) Quemada, B. "L'Inventaire mécanique des dictionnaires bilingues," *Bulletin d'Information,* fasc. 4, 1961, pp. 13-50. (ronéotypé)

(2) "Actes du Colloque sur le Mécanisation de Recherches Lexicologiques (Besançon, 6-10 Juin 1961)," *Cahiers de Lexicologie,* no. 3, 1961, pp. 1-220.

5.3.19

LIBRASCOPE DIVISION OF GENERAL PRECISION, INC.

808 Western Avenue, Glendale 1, Calif.

R. M. WORTHY

Research ultimately aimed at man-machine communication in natural, idiomatic English continues, with no change reported since the previous statement [*Ed.*].

5.3.20 MANCHESTER UNIVERSITY

FAMULIS (Faculty of Arts Manchester University Lexicographical and Indexing Service), Manchester 13, England

P. J. WEXLER

Work on the applications of computers to linguistics in subject, name, and exhaustive word indexes continues, with no change reported since the previous statement [*Ed.*]. A library of English and French texts on Telex, Linotype, and computer tape from other sources is being built up.

5.3.21

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cooperative Computing Laboratory, Cambridge 39, Mass.

MICHAEL P. BARNETT

Work on the theory of string transformations and on non-numeric applications of computers continues. A formal language for the description of relationships between strings has been defined (1). This formal scheme A language is a development of earlier work on mechanized syntactic analysis (2). Programs to effect string transformations of certain types, that are expressed in a mnemonic form of the scheme A language, are being coded for the IBM 709 computer (3). A hypothetical machine that could perform syntactic analyses and string transformations of certain types, and which consists of associative memory units organized in regular two-dimensional arrays, has been described (4). An IBM 709 computer has been used to perform

tedious algebraic reductions in some problems of celestial mechanics and molecular quantum chemistry. Work continues on the mechanized scanning of chemical formulas, both by use of the Shadow program (2) for "linearized" formulas (5) and by other methods for formulas that are treated as two-dimensional structures.

This research program is supported by a grant from the National Institutes of Health.

References:

(1) Barnett, M. P. *A Formal Language for the Description of String Relationships,* C.C.L. Technical Report No. 15, M.I.T., August 1962.

(2) Barnett, M. P., and R. P. Futrelle. "Syntactic Analysis by Digital Computer," to be published in *Communications of the ACM.*

(3) Barnett, M. P., and J. J. Levine. *Scheme A Programs for String Transformations,* C.C.L. Technical Report No. 9, M. I. T., June 1962.

(4) Barnett, M. P. "Some Comments Suggested by a Consideration of Computers," in *Macromolecular Specificity and Biological Memory,* ed. by F. O. Schmitt. Cambridge, Mass.: M.I.T. Press, 1962.

(5) Fehder, P. *Scanning Chemical Formulae by Use of the Shadow Subroutine,* C.C.L. Technical Report No. 8, M.I.T., May, 1962.

5.3.22

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Research Laboratory of Electronics, Cambridge 39, Mass.
NOAM CHOMSKY, MORRIS HALLE, *and* ROGER BROWN,
Principal Investigators

Work continues on the general theory of linguistic structure, with no change reported since the previous statement [*Ed.*]. The project is supported by the National Science Foundation.

References:

(1) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 60,* January 15, 1961, pp. 199-212.

(2) Halle, M. "On the Role of Simplicity in Linguistic Description," in *Proceedings of Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects,* ed. by Roman Jakobson, pp. 89-94. Providence, R.I.: American Mathematical Society, 1961.

(3) Chomsky, N. "On the Notion 'Rule of Grammar,'" in *Proceedings of Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects,* ed. by

Roman Jakobson, pp. 6-24. Providence, R.I.: American Mathematical Society, 1961.

(4) Chomsky, N. "Some Methodological Remarks on Generative Grammar," *Word*, vol. 17, no. 2, August 1961.

(5) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 63*, October 15, 1961, pp. 139-169.

(6) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 64*, January 15, 1962, pp. 231-238.

(7) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 65*, April 15, 1962, pp. 187-197.

(8) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 66*, July 15, 1962, pp. 295-297.

(9) M.I.T., Research Laboratory of Electronics. *Quarterly Progress Report No. 67*, October 15, 1962, pp. 177-180.

5.3.23 MATEMATIKMASKINNÄMNDEN

Box 6131, Stockholm 6, Sweden

OLLE DOPPING *and* HANS KARLGREN

The purpose of the project is to program computers to insert corrections, addenda, deletions, etc., in a punched text, yielding a tape with updated text that can be fed into a teletypesetter reading unit without further manual processing. The result should relieve the editor, the proofreader, and the printing office, and should effect more rapid delivery.

The project requires some linguistic analysis since most changes of the original text entail new division of the text into printed lines; thus the computer must be programmed so that it can divide words appropriately. The Swedish language is being used.

A successful preliminary computer program has been made, and work is now underway to improve and enlarge it.

One phase of the project, that of automatic hyphenation of words at the end of a line, has been discussed in a recent publication (1).

Reference:

(1) Thorelli, Lars Erik. "Automatic Correction of Errors in Text," *BIT (Nordisk Tidskrift for Informations-Behandling)*, vol. 2, no. 1, 1962, pp. 45-52.

5.3.24

NATIONAL LANGUAGE RESEARCH INSTITUTE

Inatuke-Nisiyamatyō, Kita-ku, Tokyo, Japan
ETUTARŌ IWABUTI, *Director*

The project to explore the basic vocabulary of modern Japanese continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) *Gendai Zassi 90 Syu no Yōgo, Yōzi* (Vocabulary and Chinese Characters in Ninety Magazines of Today), vol. I, March 1962. 321 p. (Reports of the National Language Research Institute, 21)

5.3.25

RADIO CORPORATION OF AMERICA

Data Systems Center, 4922 Fairmont Avenue, Bethesda 14, Md.
W. D. CLIMENSON, *Project Leader*

Work is continuing on the analysis and processing of natural-language text for activities within information storage and retrieval systems. The basic approach taken involves the use of those structural properties of natural language which enable a sentence or text to establish its processing logic for the computer with minimum reference to outside tables of constructions or vocabulary. The first two programs developed under this program of research determine and display various syntactic characteristics of sentences and their parts (based on the initial work of Z. Harris and his associates). The first is a sentence parsing program with particular emphasis on the preparation of sentences for further processing such as simple transformations. The second is a text condensation program which systematically reduces the size of individual sentences of a document by dropping the (grammatically) peripherally integrated structures of the sentence. Both of these programs operate on English and were developed for the RCA 501 computer.

Linguistic research is also being performed on text units larger than the sentence. Rules are being developed for identifying units of discourse comparable to the paragraph as marked in English prose. Like the paragraph, these discourse units have a major topic to which other topics are subordinated. Because of their unity of topic, these units can be considered as the basic retrievable elements to be indexed. This research, called "paragraph analysis," makes use of the sentence descriptions produced by the syntax analysis program. In addition, work has been initiated in "scope analysis," which identifies the context

necessary to relate the occurrences of word spellings within an individual document and organize them into connected lists which can become semantically interesting and useful in information retrieval. Both of these techniques impose restrictions on the acceptable outputs of the syntax analysis program. Plans for future work include the incorporation of these restrictions in the program. It is believed that a more adequate notion of sentence well-formedness will result which will greatly reduce the permissiveness of the syntax analysis program.

References:

(1) *Automatic Syntax Analysis,* Report Number ACSI-MATIC SR-60-2 for OACSI, Department of the Army, February 1961.

(2) Climenson, W. D., N. H. Hardwick, and S. N. Jacobson. "Automatic Syntax Analysis in Machine Indexing and Abstracting," in *Machine Indexing: Progress and Problems.* Washington, D. C.: American University, 1962. Also published in *American Documentation,* vol. 12, no. 3, July 1961, pp. 178-183.

(3) Jacobson, S. N. "Lexical Coding and Grammatical Analysis," presented at the Symposium on Written English for Information Processing Systems, American Psychological Association, New York, September 4, 1961.

5.3.26 THE RAND CORPORATION

1700 Main Street, Santa Monica, Calif.

DAVID G. HAYS *and* THEODORE W. ZIEHE, *Project Leaders*

The objectives of the project are to develop a general-purpose program package for the IBM 7090 computer as a tool for linguistic research to be conducted both at RAND and elsewhere, and also for use in the construction of various automatic language-data processing (ALDP) systems; to invent, refine, and test methods of linguistic research; to contribute to theoretical and empirical linguistic knowledge; to stimulate the development of practical ALDP systems; and to encourage the dissemination of computational methods in linguistic research.

The file of 750,000 running words of Russian scientific text on magnetic tape has been completely checked for accuracy. The glossary, based on ⅓ of this text, is being corrected and updated by glossary corrector routines for the IBM 7090 computer. Work continues on the design and experimental use of a data-handling system for ALDP. The translated text file (containing complete sentence structure description for ¼ of the total text) is now available for retrieval of linguistic information. A ranking of all nouns used in this text with respect to their proneness to enter into syntactic combination with

certain modifiers (adjectives and genitive nouns) has been obtained. Work is continuing on morphemic analysis of word forms in the RAND glossary and in a 100,000-word dictionary. Of the latter wordlist, 18,000 entries have been segmented by machine; rules for isolation of both derivational and flexional morphemes for the whole wordlist are being developed on the basis of the above.

Substantial additions have been made to the file of English text, without keypunching, from two sources: 7,500,000 running words of text from a national news service and 2,250,000 words from a weekly news magazine. Other sources are being sought.

A new system for automatic parsing, based on dependency grammar, has been initiated. The grammar code will be in the form of a bit pattern, describing each glossary form in terms of its possible functions as both governor and dependent. The parsing routine will proceed by matching the bit patterns of pairs of occurrences in a sentence. Experimentation with the application of this system to English is underway. This program receives continuing support from U.S. Air Force Project RAND.

References:

(RAND publications can be obtained from the authors; all of RAND's research projects, including text and glossary on magnetic tape, can be obtained by other research workers under appropriate circumstances. For additional listings of RAND publications in this field, see *CRDSD*, No. 9, Statement Nos. 3.22 and 5.3.22.)

(1) Worth, D. S. *Suprasyntactics,* RAND Memorandum RM-3161-PR, June 1962.

(2) Worth, D. S. *Studies in Russian Morphology. I. The Suffix—AGA,* RAND Memorandum RM-3235-PR, August 1962.

SUMMER INSTITUTE OF LINGUISTICS 5.3.27

Box 1960, Santa Ana, Calif.
JOSEPH E. GRIMES, *Principal Investigator*

The concordance-making program continues.

The SIL Concordance Program has been programmed in two versions for the IBM 1410 computer. One version gives card output similar to that of the original IBM 650 version but at around seven times the speed. In addition, it gives two kinds of concordance lines: one based on blank or hyphen as delimiter (thus permitting simultaneous construction of a word concordance and a morpheme concordance), with the delimiter symbol at the center of the line; the other based on period or comma as delimiter for use in syntactic analysis, with the delimiter symbol at the extreme left end of the line. Both types of

cards are developed in a single pass, but they are stacked separately. A more advanced version of the program develops longer context lines (120 characters) and more reference information which it stores on magnetic tape for transfer to a tape merge sort routine. This means that all sorting is handled internally by the machine and the final output is in alphabetized form. The IBM 1410 system, because of its variable word-length format, renders programming much simpler than the original programming for the fixed-length IBM 650 computer. Both 1410 programs are well along in the debugging stage.

To date, concordances have been prepared for linguistic analysis projects in Chinanteco and Huichol, languages of Mexico. Input data have been prepared in Mixteco and Sierra Nahuat, also of Mexico, for use with the 1410 version of the concordance as soon as it is ready. The Summer Institute of Linguistics hopes eventually to make this kind of processing of linguistic data standard for all its field projects, which at present include 266 languages and dialects in 14 countries.

5.3.28 TUSKEGEE INSTITUTE

Computing Center, School of Engineering, Tuskegee Institute, Ala.
E. ROBERT ASHWORTH, *Project Leader*

A study is being conducted to determine the statistical moments of words and their correlations between a source and target language.

Subject matter will be selected from Russian, German, French, Spanish, and N. T. Greek. Programming for an IBM 1620 computer, 20K card system will be employed for this investigation. It is postulated that there exists a unique statistic which, upon a suitable transformation, may yield either a source or target language. The letters are transformed into a binary code. The first 10 statistical moments are computed for each word. The same process is repeated for the target language. Comparisons are then made to measure the degree of relationship between the languages.

5.3.29 UNIVERSITÄT BONN

Institut für Phonetik und Kommunikationsforschung,
Koblenzer Strasse 98a, Bonn, Federal Republic of Germany
H. SCHNELLE, *Coordinator*

I. Automatic Linguistic Analysis (Dr. K. Peters-Holger, Dr. E. Gerresheim, and D. Krallmann, *Principal Investigators*)

In cooperation with the Philosophisches Seminar A at the University of Bonn, research is underway on automatic linguistic analysis, using as a text the works of the philosopher I. Kant. The complete works of Kant are being keypunched. Frequency-lists of words (glossaries) and

of concordances will be automatically constructed for retrieving proposed words and complex word combinations throughout the whole text. A formal and statistical characterization of the text will be made. In cooperation with Prof. Dr. W. Fucks (see 5.4.18), studies of style will be undertaken for the purpose of dating passages of the text whose time composition is unknown. The project will also include experiments in automatic abstracting.

II. Semantic Analysis (Dr. G. Kandler and Dr. H. Schnelle, *Principal Investigators*)

Research on semantic analysis is underway in cooperation with the Sprachwissenschaftliches (Linguistic) Institut at the University of Bonn. The possibility of using ordinary dictionaries in automatic semantic analysis and transformation of texts is being studied. Different words of known semantic relationship in a given context are being substituted. Other phases of the study include automatic generation of different contexts for the same word and automatic derivation of a class of kernel sentences from classes of sentences with similar semantic context.

III. Word Segmentation (G. Schweisthal and A. Kranzhoff, *Investigators*)

A study is being made of segmentation of words and word compounds in cooperation with the Zentralstelle für Terminologieforschung (Center for Research on Terminology) of Dr. G. Kandler. Currently the center is preparing a dictionary of German words. Among other information, this dictionary should indicate the morphemic constituents of words. In cooperation with the Center, automatic segmentation will be corrected manually and recorded.

IV. Mathematical Linguistics and Theory of Programming Language (Dr. H. Schnelle, *Principal Investigator*)

A project of mathematical linguistics and theory of programming languages, undertaken in cooperation with the Institut für angewandte Mathematik (Institute for Applied Mathematics), includes (a) descriptions of linguistic structures (generalization of descriptions for syntactic structures according to Chomsky, Bar-Hillel, Hays, Gaifman, and Lecerf for the following purposes: application to semantic structures, dynamic interpretation of the static descriptions, and morpho-phonemic studies as preliminary research on automatic character and speech recognition); (b) application of (a) to the theory of programming languages; (c) machine translation of programs written in ordinary language into programs in machine languages (or programming languages); and (d) machine learning of structural descriptions of texts by the methods of problem solving.

5.3.30 UNIVERSITY OF CALIFORNIA

Linguistics Department and Computer Center, Berkeley, Calif.
SYDNEY M. LAMB, *Associate Professor of Linguistics*

Research is being conducted on the use of the computer as an aid in linguistic analysis, and linguistic programs and subroutines intended to facilitate linguistic analysis are being written for the IBM 7090 computer. Consultation is also being provided to various linguists in order to assist them in taking advantage of the possibilities afforded by the computer.

The first concrete job of this project has been the construction of a general concordance-making program. It is planned that some of the routines previously prepared by the Machine Translation Project (see 3.38) and also those worked on in connection with the Seminar in †Mechanolinguistics offered by the Linguistics Department will be generalized for widespread application.

The tools and services of this project are currently being used in studies of (a) the development of grammatical patterns in child language, (b) differences between spoken and written Hindi, and (c) French literary style.

Reference:

(1) Lamb, Sydney M. "The Digital Computer as an Aid in Linguistics," *Language,* vol. 37, 1961, pp. 382-412.

5.3.31 UNIVERSITY OF CAMBRIDGE

Cambridge, England
R. WISBEY, *Downing College,* and P. SWINNERTON DYER,
Mathematical Laboratory, Corn Exchange Street

Work on word indexes, concordances, and lexicographical material for medieval German literary texts of the 11th and 12th centuries continues, thus creating a growing reservoir of coded material for use in linguistic research by computer.

All texts are transferred first to seven-hole punched paper tape and then to magnetic tape for sorting by the EDSAC 2 computer. The limitations of the Flexowriter keyboard make necessary some small degree of coding where unusual characters and signs are concerned. However, this difficulty is likely to be only temporary.

All indexes and concordances will be of the classified type; i.e., the intermediate output, which is a strictly alphabetical index of word forms, is rearranged according to nonalphabetical criteria with homonyms being separated. Important variant readings are included;

† *See* Glossary.

the degree of completeness depends on the manuscript situation with regard to a particular text. The sorting programs have been designed to provide statistical information on word and letter frequencies, word lengths, etc. The first index in this series is of the *Wiener Genesis,* an Early Middle High German text of over 28,000 words. An account of this work has been published (1). 150,000 words of early German sermons, including the complete *Speculum Ecclesiae,* have been punched. Progress is also being made, under the auspices of the Jakob Boehme Society Inc. and in cooperation with Dr. Hans Popper, University College of Swansea, England, with concordances and indexes to the works of Jakob Boehme. Well over 100,000 words have been punched to date. Good progress has been made at the Mathematical Laboratory with the development of a Monotype reader.

The described work has aroused considerable interest within the University, and at least one similar project for the Italian language has been undertaken by Dr. C. P. Brand. A research center is being set up to coordinate these activities.

References:

(1) Wisbey, R. A. "Concordance Making by Electronic Computer: Some Experiences with the *Wiener Genesis,*" *Modern Language Review,* vol. 57, April 1962, pp. 161-172.

(2) Wisbey, R. A. "Mechanization in Lexicography," *Times Literary Supplement,* March 30, 1962. Reprinted in *Freeing the Mind.* London: Times, 1962.

UNIVERSITY COLLEGE LONDON **5.3.32**
Gower Street W.C.1, London, England
RANDOLPH QUIRK, *Director*

The examination of the grammar of spoken and written English continues, with no change reported since the previous statement [*Ed.*]. The survey is being financed by Naturmetodens Sproginstitut (Denmark), the Clarendon Press, and Longmans, Green and Co., Ltd.

References:

(1) Quirk, R. "Towards a Description of English Usage," *Transactions of the Philological Society,* 1960, pp. 40-61.

(2) Quirk, R., and A. P. Duckworth. "Co-existing Negative Preterite Forms of *dare,*" *Language and Society,* Copenhagen, 1961, pp. 135-140.

(3) Quirk, R. *The Use of English.* London: Longmans, Green and Co., Ltd., 1962.

5.3.33 UNIVERSITY COLLEGE LONDON

Communication Research Centre, Gower Street, London, W.C.1, England

A. H. SMITH, *Honorary Secretary*

The Centre's present program of research is concerned with describing and determining the effectiveness of trade, professional, and similar "dialects" of English. Detailed studies of advertising and Civil Service English are now in progress, and it is hoped to initiate research on other "dialects" (such as the English of Law and of the Sciences) soon. The nature and aims of the Centre's program are discussed at length in (1).

References:

(1) Warburg, Jeremy. *The Best-Chosen English: Introduction to a Programme of Research,* Communication Research Centre, 1961.

(2) Wason, P. C. *Psychological Aspects of Negation: An Experimental Enquiry and some Practical Applications,* Communication Research Centre, 1962.

5.3.34 UNIVERSITY OF EDINBURGH

Department of English Language and General Linguistics, 40b George Square, Edinburgh 8, Scotland

ANGUS McINTOSH *and* M. A. K. HALLIDAY

The foundational interrelations between various types of logics and linguistics are being examined. Specific projects involve the application of symbolic logical notations to statements of linguistic theories, and a description of the functional logic which operates in everyday spoken English. Emphasis is being placed on the adequate descriptive nature of statements, and on the setting up of distributional, rather than conceptual, classifications.

Work is proceeding on a unified general theory of comparative linguistics. The general statement of the categories of Modern English grammar is nearing completion and is already being applied in detailed descriptions and comparisons of particular texts. Preliminary plans for using large-scale computer techniques for grammatical, and particularly lexical, classificatory and analytic studies are under discussion. An IBM 7090 computer program for an equi-domain criteria selection procedure is being written; this will have very wide linguistic applicability.

The work is supported in part by IBM-United Kingdom Ltd.

References:

(1) Ellis, J. O. "Some Problems in Comparative Linguistics," in

Proceedings of the University of Durham Philosophical Society, 1961, pp. 54-62.
(2) Halliday, M.A.K. "Linguistics and Machine Translation," *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft,* 1962.
(3) Halliday, M.A.K. "Linguistique générale et linguistique appliquée dans l'enseignement des langues vivantes," *Études de Linguistique Appliquée,* vol. 1, 1962, pp. 5-42.

UNIVERSITY OF GOTHENBURG 5.3.35

Department of English Language and Literature, Gothenburg, Sweden
ALVAR ELLEGÅRD

Work is in progress on mechanical linguistic analysis of English texts, on the basis of teletypesetter tapes used by Hunt Barnard & Co., Aylesbury, Bucks., England, for printing paperbacks. A start has been made with Daniel Defoe's *Moll Flanders* (Corgi Books edition). The tape is at present being processed on the Alwac unit at the ADB Institute, Chalmers Tekniska Högskola, Gothenburg.

The program is still in its initial stages, but the following steps are envisaged: A full alphabetical concordance will be compiled, covering every word without exception. Statistics will be drawn up on the vocabulary distribution in successive 1,000-word portions of the text to yield answers to the following questions: How many different words are there in each portion of text? How many new words are introduced in each successive portion? Which words are most (least) evenly distributed throughout the text? An attempt will be made to establish word classes purely on the basis of frequency and distribution criteria; a pilot project on this has been started. The sentences in the text, classified according to word-class structure, will be listed. Word-class structure and immediate constituent structure will be compared.

References:

(1) Ellegård, Alvar. "Estimating Vocabulary Size," *Word,* vol. 16, no. 2., 1960, pp. 219-244.
(2) Ellegård, Alvar. *A Statistical Method for Determining Authorship,* Gothenburg Studies in English, 13, Gothenburg, 1962.

UNIVERSITY OF PENNSYLVANIA 5.3.36

Department of Linguistics, Philadelphia 4, Pa.
A. F. BROWN

In order to facilitate research on English, extensive English word-lists are being prepared and made available at cost to others. So far, the available lists, alphabetized normally and in reverse order, com-

† *See* Reverse-Alphabetized List in Glossary.

prise about 355,000 entries selected from an unabridged dictionary and four technical dictionaries. The lists have been prepared by alphabetizing- and merging-routines on a UNIVAC I computer and reproduced on paper in 8-volume sets of limited supply. Tape version duplications are being provided on UNIVAC I and IBM tapes. More extensive listings are being prepared by the inclusion of entries from several more specialized dictionaries and glossaries.

Phonemicized versions of syllabic subsets are being developed on the basis of the present orthographic lists, and will be used in studies of the voids and predictabilities within the phonological rules for syllabic subsets, and in studies of mechanical recognition of morpheme and word boundaries. Research is also being carried out on word-derivational processes and the transformational equivalents of compound structures.

Support is provided by the U. S. Air Force Office of Scientific Research.

5.3.37 UNIVERSITY OF PENNSYLVANIA

Department of Linguistics, Philadelphia, Pa.
ZELLIG S. HARRIS, *Professor of Linguistics*

The computable analysis of language structure and the detailed investigation of linguistic transformations continue. It is expected that, in addition to its theoretical interest, this work may have a new kind of application to information retrieval. The project has resulted so far in a large body of syntactic studies on English, and in a working UNIVAC program which analyzes a syntactic (constituent) structure of English sentences (without the assistance of any human editing). At present, a transformation program capable of reducing any English sentence to component sentences is being put on a computer. In addition, work is being done on the detailed English transformations, on a general theory of transformations, and on the transformations of other languages.

The project is supported by the National Science Foundation.

References:

(1) The main results of the project are given in Papers on Formal Linguistics, a series of monographs published by Mouton and Co., The Hague, Netherlands. Those published or in press:

PoFL 1. *String Analysis of Sentence Structure,* by Zellig S. Harris, 1962. (A revised version of TDAP 15, below)

PoFL 2. *Discourse Analysis Reprints,* by Zellig S. Harris. (In press)

(2) More detailed or tentative presentations of results are given in the series, Transformations and Discourse Analysis Papers (TDAP). The major numbered reports to the National Science Foundation to date are:

2. *Introduction to Transformations* (Reprint, 1956), Zellig S. Harris.
3, 4. *Discourse Analysis Reprints,* Zellig S. Harris.
5. *Generation of the Nominal Compound in English,* by R. B. Lees. This paper was later expanded into a thesis which was published as "The Grammar of English Nominalizations," *International Journal of American Linguistics,* Part II, vol. 26, no. 3, July 1960. Available from Director of Publications of the Research Center in Anthropology, Folklore, and Linguistics, Indiana University, Bloomington, Ind. Price: $4.00.
11. *Preparing a Text for Retrieval* (Tentative), Zellig S. Harris.
12. *An Effective Procedure for Mechanical Preparation of a Scientific Text for Information Retrieval,* Ruth Zeitlin Koenigsberg.
14. *A Computer Program for Syntactic Analysis,* Zellig S. Harris.
15. *Computable Syntactic Analysis,* Zellig S. Harris. (Revised version published as PoFL 1, above)
16. *Word and Word-Complex Dictionaries,* Lila Gleitman.
17. *Elimination of Alternative Classifications,* Naomi Sager.
18. *Recognition of Local Substrings,* Aravind K. Joshi.
19. *Higher-order Substrings and Well-formedness,* Bruria Kaufman.
20. *Iterative Computation of String Nesting (Fortran Code),* Bruria Kaufman.
21. *Syntactic Completion Analysis and Theories of Grammatical Categories,* Henry Hiż.
22. *Pre-adjectivals in the English Nominal Phrase,* Seymour Chatman.
27. *Procedure for Left-to-Right Recognition of Sentence Structure,* Naomi Sager.
28. *Problems in the Computation of English Sentence Structure (Substring analysis),* Naomi Sager.
29. *The Intuitions of Grammatical Categories,* Henry Hiż.
30. *English Transformation List,* Zellig S. Harris.
31. *Order of Adjectives,* Zeno Vendler.

32. *A Procedure for Decomposing a Complex Sentence Into Its Kernels,* Aravind K. Joshi.
33. *A Decision Method for a Class of Formal Systems, and its Application to Transformational Analysis,* Hans Herzberger.
34. *Questions and Answers,* H. Hiż.
35. *The Classification of English Verbs by (Non-Sentential) Object Types,* Seymour Chatman.
38. *The Transformational Status of the Definite Articles in English,* Beverly Robbins.
39. *An Analysis of Container Sentences,* Ruth Z. Koenigsberg.
40. *Conjunction with* and, Lila R. Gleitman.
41. *Transformational Decomposition Systems,* Hans Herzberger.
42. *A Procedure for a Transformational Decomposition of a Complex Sentence,* Aravind K. Joshi.
44. *On Some Connections Between Transformations,* Zbigniew Lis.

Paper 15 gives an information presentation of a general theory and method for syntactic recognition. Papers 16-19 give the actual flow charts of each section of the syntactic analysis program.

(3) Harris, Zellig S. "Linguistic Transformations for Information Retrieval," in *Proceedings of the International Conference on Scientific Information,* pp. 937-950. Washington, D. C.: National Academy of Sciences-National Research Council, 1959.

(4) Hiż, Henry. "Steps Toward Grammatical Recognition," in *Advances in Documentation and Library Science,* vol. 3, part 2, pp. 811-822. New York-London: Interscience Publishers, Inc., 1961.

(5) Joshi, Aravind K. "Computation of Syntactic Structure" (TDAP No. 25), in *Advances in Documentation and Library Science,* vol. 3, part 2, pp. 831-840. New York-London: Interscience Publishers, Inc., 1961.

(6) Gleitman, Lila. "The Isolation of Elements for Grammatical Analysis," in *Advances in Documentation and Library Science,* vol. 3, part 2, pp. 823-830. New York-London: Interscience Publishers, Inc., 1961.

(7) Chomsky, Noam. "On Certain Formal Properties of Grammars," *Information and Control,* vol. 2, 1959, pp. 137-167.

(8) Hiż, Henry. "Congrammaticality, Batteries of Transformations and Grammatical Categories," in *Proceedings of the Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects,* ed. by Roman Jakobson, pp.

43-50. Providence, R. I.: American Mathematical Society, 1961.

(9) Herzberger, Hans G. "The Joints of English," in *Proceedings of the Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects,* ed. by Roman Jakobson, pp. 99-103. Providence, R. I.: American Mathematical Society, 1961.

(10) Goodman, Nelson. "Graphs for Linguistics," in *Proceedings of the Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects,* ed. by Roman Jakobson, pp. 51-55. Providence, R. I.: American Mathematical Society, 1961.

(11) Hiż, Henry. "Grammatical Category as a Relation," in *Abstracts of Contributed Papers at the International Congress for Logic, Methodology and Philosophy of Science* (Stanford University, Stanford, Calif.), 1960, pp. 147-148.

(12) Hiż, Henry. "The Intuitions of Grammatical Categories," *Methodos,* vol. 12, no. 48, 1960, pp. 1-9.

(13) Hiż, Henry. "Modalities and Extended Systems," *The Journal of Philosophy,* vol. 58 (1961), pp. 723-731.

(14) Hiż, Henry. "Questions and Answers," *The Journal of Philosophy,* vol. 58 (1962), pp. 253-265.

UNIVERSITY OF PITTSBURGH 5.3.38

Computation and Data Processing Center,
817 Cathedral of Learning, Pittsburgh 13, Pa.
DALE W. ISNER, *Principal Investigator*

Investigation underway concerns the feasibility of developing a programming language compatible with the needs of psychologists, educators, etc., in retrieval and analysis of natural text material. Existing languages in this field are being compared, and the requirements which such a language must have to be beneficial and also usable are being determined. The languages worked with are COMIT and Wegstein's String Manipulations in ALGOL. The IBM 7070/1401 system is used. Possible implementation on several computers is planned.

Thus far, samples of existing languages have been studied to determine their deficiencies, and various people using textual material as input to computers have been interviewed.

It is planned to evaluate existing languages and to innovate new ideas to determine specifications.

Partial support is provided by the International Business Machines Corporation.

Reference:

(1) Kehl, William B., John F. Horty, Charles R. T. Bacon, and David S. Mitchell. "An Information Retrieval Language for Legal Studies," *Communications of the ACM*, vol. 4, no. 9, September 1961, pp. 380-389.

5.3.39 UNIVERSITY OF WISCONSIN

Department of Linguistics, Madison, Wisc.
MURRAY FOWLER

The purpose of a project underway is a complete statistical analysis of the Etruscan language. Some 9,000 Etruscan inscriptions are being entered into punchcards. Programming of the corpus for the CDC 1604 computer will be completed by March 1, 1963. The corpus will be run through the computer in from 45 to 55 operations.

The various analyses to be made include studies of (a) the frequency and transitional probability of letters in the alphabet; (b) word lengths; (c) marks of punctuation; and (d) the rank frequency, linguistic environment, and recurrent patterns of words. The assumption is made that Etruscan is an inflected language. Word endings will be cut off and studied.

5.3.40 UNIVERSITY OF WISCONSIN

Department of English, Madison, Wisc.
A. HOOD ROBERTS[1]

An investigation was made into the etymological composition of English and the frequency of occurrence of phonemes and phoneme sequences in American English according to the principles of quantitative statistical analysis.

The 10,000 words in Ernest Horn's *A Basic Writing Vocabulary* were put into sentence frames and recorded on tape by a speaker of Midwestern idiolect. From the tapes the words were transcribed phonemically according to the Trager-Smith Notation. The proximate etymological sources of the words were determined. The phonemic transcriptions of the words, their etymological sources, and their frequencies of occurrence were punched on IBM cards. These data were processed through the CDC 1604 digital computer, and the following were obtained: (a) the total number of frequencies of occurrence of the words and phonemes in the corpus; (b) the average word length in phonemes and in syllables by frequency and by number; (c) the etymological composition of English; (d) the relative frequencies of the segmental phonemes for the entire corpus, the number and frequency of phonemes and their standard error, the standard error of the dif-

[1] Present address: University of South Florida, Tampa, Fla.

ference between successive phonemes, and the standard error deviation for each decile and the entire corpus. These computations were also made separately for vowels and for consonants and semivowels; (e) transitional probabilities for all sequences of two phonemes; (f) transitional probabilities for word-initial sequences of three phonemes, and estimated probabilities for sequences of four phonemes; (g) the entropy and redundancy for each decile and for the entire corpus (these computations were based on phoneme frequencies, word length in phonemes, and word length in syllables); (h) the relative frequencies of occurrence for arrangements according to the classification of phonemes as consonant, vowel, or semivowel, and as to the manner of production and the points of production of the phonemes, and (i) the frequencies for the 60 initial, 408 intervocalic, and 152 final consonants and consonant clusters.

This investigation was supported by the Research Committee of the Graduate School of the University of Wisconsin.

Reference:

(1) Roberts, A. Hood. *Frequencies of Occurrence of Segmental Phonemes in American English.* Doctoral Thesis. Madison, Wisc.: University of Wisconsin, 1961. 683 p. (Available from University Microfilms, Ann Arbor, Mich.)

(INDEPENDENT WORK) **5.3.41**

str. 13 Decembrie 7, Timişoara IV, Rumania
S. ABRAHAM

Research is being conducted on the possibility of processing a scientific text with the aid of a digital computer for (a) revealing the synonymy of expressions, and (b) recognizing the logical deducibility of certain expressions from others (within the limits of the same text).

For realizing the aim stated in (a), a convenient formal theory of synonymy of expressions has been elaborated. By expression is meant a set of dependent (on different levels) words, in connection with which a so-called "grammar of dependence" is being elaborated for the English language.

Synonymy is distinguished on lexical, morphological, and syntactical levels.

An algorithm has been worked out for machine recognition of synonymy on the basis of so-called "transformation rules" of the English language, i.e., rules whose application to a given expression gives another synonymous expression. It is planned to program and machine-test this algorithm before the end of 1962 on either a MECIPT-1 or a CIFA computer.

After completing the first stage of the work, principal attention will be paid to imparting greater flexibility to the bases of the algorithm (on some topological considerations) and to removing the limitations to the nature of the text.

As far as (b) is concerned, the main attention is paid to the problems of registering a given text in a form suitable for the computer, which, it is hoped, can be realized by a corresponding completion of the logical language.

5.3.42 (INDEPENDENT WORK)

132-35 Sanford Avenue, Flushing 55, N. Y.
EDWIN S. SPIEGELTHAL

As of July 1962, the program to extract the salient semantic information contained in English sentences is virtually complete. This program analyzes sentences in "reduced" English (specifically, a 32-word English lexicon) and derives therefrom such semantic data as: Who or what does the sentence explicitly mention or implicitly allude to, and what interrelationships does the sentence state, or imply, as subsisting among the things referred to?

The underlying methodology is apparently adequate to the task of eliciting all the possible semantic linkages, using the data in a novel type of dictionary, and has proven reasonably effective, using algorithmic techniques exclusively, in deleting those "possibilities" which were, in fact, incorrect or redundant. It was hoped that algorithmic techniques alone would suffice for this task, but it has become apparent that recourse will have to be had to those heuristic techniques first developed for solving Double-Crostic puzzles by machine.

To be suitable as a "production program" for such tasks as automatic indexing and document categorization, the present research program clearly requires work in four major directions: expanding the dictionary to useful proportions, strengthening the syntactic sophistication of the program, adding the requisite heuristic techniques to take up from where the algorithmic techniques leave off, and recasting the program so as to maximize operating efficiency rather than research-oriented completeness of post-mortem data.

It is not yet clear whether further work in this direction will take the form of another research project or of a production program.

Reference:

(1) Spiegelthal, E. S. "Redundancy Exploitation in the Computer Solution of Double-Crostics," in *Proceedings of the Eastern Joint Computer Conference,* December 1960.

(INDEPENDENT WORK) 5.3.43

Seminar für Sprache und Kultur Japans, Universität Hamburg, Hamburg 13, Federal Republic of Germany
GUNTHER WENCK

Japanese chemical texts (1,000 sentences as primary, i.e., fully analyzed, material) have been submitted to syntax analysis, with the goal of finding out the fundamental methods and of locating the principal problems for a fully automatic (i.e., without preediting and with a view to the use of a reading machine) Japanese-English machine translation. As to the syntax within a simple sentence or clause, the salient point is that it cannot be fully described in terms of syntactical morphemes only, but that the lexical meaning of nonsyntactical morphemes is relevant for structure determination, too (e.g., class meaning in the case of nouns, immanent sentence patterns in the case of verbs). As to the syntax of compound sentences (clause syntax), studies have been concentrated on the problems of determining the boundaries of inserted clauses and of defining rules for nonlinear clause arrangement (e.g., clause grouping, leap-over relation).

The project, which is supported by Euratom, was begun in the spring of 1961.

A program has been written, tested and rewritten several times since 1959 for a punchcard analysis of Japanese literary texts throughout the ages, with the goal of permitting exact language description of individual texts as well as functional comparison between the languages of different texts and different times. The card used is an ALLFORM card with a total of 408 holes. A preliminary report is in preparation.

The project is supported by the Deutsche Forschungsgemeinschaft.

5.4 *ARTIFICIAL INTELLIGENCE*

Artificial intelligence is concerned with the problem of making machines behave intelligently; that is, of exhibiting thought processes and the mechanisms that underlie them. Much of the work in the artificial intelligence field is concentrated on simulating the complex information processing activities involved in human cognition and the neurophysiological mechanisms that underlie them. Mathematical, physical, neurophysiological and psychological models of these processes and mechanisms are being constructed, and computers and computer simulation techniques are being used as the primary research tools for testing these models and gaining new insight into the nature of these processes and mechanisms. The potential impact of advances in artificial intelligence research on the information retrieval field becomes obvious when information retrieval is regarded as a highly complex intellectual information processing task involving such processes as experience, learning, classification, searching, and recognition. This research is also closely related to work on pattern recognition described in subsections 5.1 and 5.2.

With the exception of three projects, the work described in this subsection is a continuation of previously reported work and represents a wide variety of research activity aimed at design, construction, and use of problem-solving machinery. Some work in relevant mathematical, psychological, and physiological domains is included.

A new project at **Princeton University** (5.4.19) is investigating computer applications to music and musicology involving the development of a data processing "language" suitable for the treatment of written musical information.

Research underway at **Rand Corporation** (5.4.20) is attempting to clarify and gain insights into the question of organizational principles for intelligent systems (both real and artificial).

National Bureau of Standards (5.4.14) is making a comparative study of four list processing languages (COMIT, FLPL, IPL-V, and LISP).

5.4.1 AERONUTRONIC

Division of Ford Motor Co., Ford Road, Newport Beach, Calif.
J. K. HAWKINS, *Self Organizing Systems Department*

Research is being conducted on the problem of learning in cascade adaptive network structures. The goal is to develop learning rules suitable for simple adaptive circuit elements, when such elements must be interconnected in cascade, or multi-level networks, in order to realize complex digital functions which are beyond the capabilities of single-level networks.

A special-purpose electronic system has been constructed for performing a variety of learning experiments for the purpose of this investigation. Checkout of the system has been completed, and preliminary experiments conducted.

Two types of two-level network learning structures have been investigated. Both are characterized by the fact that first-layer elements can only progress in the direction of the desired overall network function. The results were largely negative, and indicate a need for learning rules which cause divergence of behavior among first-level elements. Techniques possessing this property are being investigated, and various versions of them will be prepared for experimental trial.

The research is supported by the U.S. Office of Naval Research and the U.S. Air Force (Rome Air Development Center).

5.4.2 AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

CRRBA, L. G. Hanscom Field, Bedford, Mass.
STANLEY R. PETRICK *and* A. LEWIS BASTIAN, Jr.

Work on the chess program of Mr. A. L. Bastian, Jr., and the contract bridge-bidding project continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Irons, E. T. "A Syntax Directed Compiler for Algol 60," *Communications of the ACM*, vol. 4, no. 1, January 1961, pp. 51-55.

(2) Brooker, R. A., and D. Morris. "A General Translation Program for Phrase Structure Languages," *Journal of the Association for Computing Machinery*, vol. 9, no. 1, January 1962, pp. 1-10.

(3) Warshall, S. "A Syntax Directed Generator," in *Computers: Key to Total Systems Control.* New York: The Macmillan Co., 1961.

5.4.3 ARMOUR RESEARCH FOUNDATION

10 West 35th Street, Chicago, Ill.
SCOTT H. CAMERON

Research on both the theoretical and hardware aspects of adaptive or self-organizing networks of threshold operators continues, with no change reported since the previous statement [*Ed.*]. The project is supported by the U. S. Office of Naval Research.

Reference:

(1) Cameron, Scott H. *Self-Organizing Networks,* Annual Report to Office of Naval Research for the period February 15, 1961-February 14, 1962.

ASTROPOWER, INCORPORATED 5.4.4

A Subsidiary of Douglas Aircraft Company, Electronic Systems Division, P. O. Box 500, Costa Mesa, Calif.
R. D. JOSEPH *and* S. S. VIGLIONE

The general goal of the activities in progress is the development of electronic data processing subsystems having capabilities comparable to the sensor-and-brain combination possessed by the higher order animals.

Portions of the program in the Analysis Department, under the direction of R. D. Joseph, are oriented toward unanswered questions in neurophysiology with little emphasis on immediate applications. In addition, efforts are being made to apply these perceptron principles to the analysis of problems in photographic interpretation and signature analysis.

Other portions of the program in the Electro-Optics Department, under the direction of S. S. Viglione, are heavily oriented toward the solution of present military problems, mainly in the surveillance area. This work is oriented toward the development of hardware to perform the photo interpretation and signature analysis tasks using self-organizing techniques. Adaptive elements are also being investigated. A recognition device, the result of a self-organizing procedure which was simulated on a digital computer, has been constructed and demonstrated to illustrate the feasibility of self-organizing techniques. An advanced guidance for deep space probes is in the preliminary design stages.

The long-range research indicates the practicality of major extensions of the perceptron concept. These extensions have previously been suggested but not carried through in a rigorous fashion.

References:

(1) Joseph, R. D. *Contributions to Perceptron Theory.* Doctoral Thesis. Ithaca, N. Y.: Cornell University, 1961. (Extended version to be published)
(2) Daly, J., R. D. Joseph, and P. M. Kelly. "Self-Organizing Logic Systems," to be published in *Astronautica Acta.*
(3) Joseph, R. D., P. M. Kelly, and S. S. Viglione. "An Optical Decision Filter," presented at WESCON, August 1962.
(4) Joseph, R. D. "On Predicting Perceptron Performance," in *Record of IRE National Convention* Part 2, New York, 1960.
(5) Viglione, S. S., and H. F. Wolf. "Star Field Recognition for Space Vehicle Orientation," presented at the East Coast Conference on Aerospace and Navigational Electronics, October 1962.

(6) Joseph, R. D., S. S. Viglione, and H. F. Wolf. "Application of Distributed Memory Systems to Space Navigation," presented at the Spaceborne Computer Engineering Conference, October 1962.

AUTONETICS 5.4.5

A Division of North American Aviation, Inc.,
P. O. Box R-3, Anaheim, Calif.
RICHARD K. OVERTON

Basic research in learning, and in problems of mechanizing learning, is underway. This research has led to the development of a prototype learning machine: a device capable of solving, as a result of experience, problems for which it has not been programmed. The device is called the Autonetics General Information Learning Equipment (AGILE).

AGILE has learned several skills, including the playing of simple games such as tic-tac-toe and the writing of short programs for computers. It has also been taught, using severely restricted vocabularies, to translate Russian and English, and English and Spanish. Translation is taught by giving AGILE a sample of material in one language and a translation in another. On the basis of this material, AGILE makes up rules by which it translates new material.

The computer used in the AGILE system, the Recomp II, has a memory which is too small for practical translation of natural languages. Current study concerns the possible use of large auxiliary memories. Also under study is the possibility of using AGILE as an adaptive, English-language information retrieval system in special situations.

References:

(1) "Learning Machine Works on Information 'Globs,'" *Control Engineering,* May 1962, pp. 24-25.
(2) Overton, Richard K. "Some Data and Comments on Brain and Computer Memory Capacities," in *Transactions,* San Diego Conference on Medical Electronics and Bionics, 1961.

CAMBRIDGE LANGUAGE RESEARCH UNIT 5.4.6

20 Millington Road, Cambridge, England
E. W. BASTIN, *Information Structures Unit*

Research is underway on a deductive approach to structured arrays of information, particularly in relation to the use of self-organizing systems in solving problems in information handling and information retrieval.

Research has been continued on the use of the CASPAR control system as a model for a physical theory of elementary particle systems.

An algebra of matrices over the binary field J_2 has been used to establish an algebraic hierarchy of levels called the Descriptive Hierarchy. This algebra has been used to make some calculations (interpreted in terms of the control system model) of basic dimensionless ratios relating to the stable elementary particle system (the proton-electron system).

The usefulness of a control system as a model is due to the fact assumed in conventional quantum theory: In making measurements, the observer is forced to interact with the physical world (this interaction being the origin of the Heisenberg uncertainty). The theory being described is an analysis of the conditions under which such an interaction can be represented within the theory of automata, and the deductions mentioned above are derived in the course of this analysis.

The algebra of binary matrices as applied to the stable particles is a degenerate case of the control model, and future work will concern the nondegenerate cases which describe instability in particle systems.

The research is supported by the U. S. Air Force Office of Scientific Research.

Reference:

(1) Bastin, E. W. *A Theory of the Origin of Mass within a Control Model of the Elementary Particles,* Information Structures Unit Technical Note III, 1962.

5.4.7 CORNELL AERONAUTICAL LABORATORY, INC.

Computer Research Department,
P. O. Box 235, Buffalo, N. Y.
H. R. LELAND, *Head, Cognitive Systems Section*

Work on the study to determine the feasibility of automating parts of the photo interpretation task has continued. Using the photographic input device and the IBM 704 computer, work during the past few months has been concentrated on the detection of objects of interest in aerial photography by means of *outlining.* This detection procedure, which is carried out in the separate steps of gradient determination and gap filling, has been partially successful. Other effort in this area is that of defining, with relatively high statistical confidence, the capabilities of the recognition portion of the entire automatic photo interpretation system. The study is sponsored by the Geography Branch of the U. S. Office of Naval Research.

The development of perceptron applications has continued. Recent work includes the development of a theory for selection of initial weights in a simple perceptron which takes into use the presence of noise and distortion in the input patterns. The project is sponsored by the U. S. Office of Naval Research.

Automatic target recognition work has been continued with emphasis on studies of the detection and identification of ribbons in aerial photography. This work, which has been carried out by implementing special-purpose devices within the IBM 704 general-purpose digital computer, is sponsored by the U.S. Air Force (Microwave Technology Section, Aeronautical Systems Division).

A new program has been initiated which has the objective of automatically recognizing audio signals derived from doppler radar systems and providing a classification of these signals into three basic target groups (vehicles, personnel, and clutter). The current phase of this work is concerned with collection of a well-documented set of input patterns. The Research and Development Laboratory of the U. S. Army Electronics Supply Agency sponsors the program.

Another new program, sponsored by the U. S. Air Force, has as its objective the suggestion of areas in which information processing techniques can be applied to the threat evaluation and action selection problem. Among the techniques to be studied are adaptive networks, perceptrons, heuristic programming, and information storage and retrieval techniques.

The program in the mathematical analysis of perceptrons has continued. A report, to be issued soon, describes a theory for the learning process in perceptrons which has as its setting the theory of measure and function space.

The project to apply pattern recognition techniques to the submarine detection problem has continued. A report describing preliminary results has been issued. This work is conducted under a no-cost, policy agreement with the Bureau of Ships.

References:

(1) Holmes, W. S., H. R. Leland, G. E. Richmond, and J. L. Muerle. "Recognition of Mixed-Font Imperfect Characters," in *Optical Character Recognition* (proceedings of the ONR-NBS Symposium on Optical Character Recognition, Washington, D. C., January 1962), ed. by George L. Fischer, Jr., et al., pp. 213-225. Washington, D. C.: Spartan Books, 1962.

(2) Holmes, W. S., H. R. Leland, G. E. Richmond, and M. G. Spooner. *Phase II Interim Report—Perceptron Applicability to Photo Interpretation.* Contract No. Nonr-3161 (00).

(3) Holmes, W. S., H. R. Leland, and G. E. Richmond. "Design of a Photo Interpretation Automaton," presented at fall 1962 meeting of the American Federation of Information Processing Societies.

(4) Rosenblatt, Frank. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Washington, D. C.: Spartan Books, 1962. 636 p. Price: $6.50. Also issued as C.A.L.

Report No. VG-1196-G-8, March 15, 1961. Contract No. Nonr-2381 (Project PARA) at C.A.L., and Contract No. Nonr-401(40) at Cornell University.

5.4.8 CORNELL UNIVERSITY

Cognitive Systems Research Program, Hollister Hall, Ithaca, N. Y.
FRANK ROSENBLATT, *Director*

Work continues on theoretical analysis and digital simulation of perceptrons and related brain models, with no change reported since the previous statement [*Ed.*]. The work is supported by the U. S. Office of Naval Research.

References:

(1) Rosenblatt, Frank. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Washington, D. C.: Spartan Books, 1962. 636 p. Price: $6.50. Also issued as C.A.L. Report No. VG-1196-G-8, March 15, 1961. Contract No. Nonr-2381 (Project PARA) at C.A.L., and Contract No. Nonr-401(40) at Cornell University.

(2) Block, H. D. "The Perceptron: A Model for Brain Function," *Review of Modern Physics,* January 1962, pp. 123-135.

(3) Block, H. D., B. W. Knight, and F. Rosenblatt. "Analysis of a Four Layer Series-Coupled Perceptron," *Review of Modern Physics,* January 1962, pp. 135-142.

5.4.9 ИНСТИТУТ КИБЕРНЕТИКИ АКАДЕМИИ НАУК УССР [INSTITUTE OF CYBERNETICS, ACADEMY OF SCIENCES UKRAINIAN S.S.R.][1]

4 Lysogorskaya, Kiev 28, U.S.S.R.

В. М. Глушков [V. M. GLUSHKOV], *Academician of the Academy of Sciences Ukrainian S.S.R.*

A computer program simulating the process of learning to distinguish intelligent and meaningless sentences in Russian was developed in collaboration with A. A. Stognii in 1961 (1). The initial vocabulary and the complexity of the grammatical constructions were limited by the memory capacity of the "Kiev" computer used for simulation. During the learning process, the computer generated new concepts and extrapolated the properties of one concept to another. Therefore, the complete learning (i.e., the correct discrimination of meaningful and meaningless sentences) was achieved long before the computer received all intelligent sentences of the given grammatical construction composed of the present vocabulary. In other words, the computer was able to

[1] Formerly the Computing Center, Academy of Sciences Ukrainian S.S.R.

classify correctly not only the sentences, the intelligence of which was indicated, but also the ones not appearing in the previous experience. By changing some numerical parameters in the program, different types of learning may be simulated: from simple cramming to the most exaggerated tendency to generalizations and analogies.

Another program elaborated by A. A. Letichevskiĭ simulates an association of living beings which are developing according to the laws of biological evolution (2). This program represents the process of cognition and adaptation to a law of nature which controls the displacement of food in space. Heredity, mutation, death due to starvation and age, and the reproduction of living beings by fission were simulated. The initial rules for reacting to environmental changes were given for each being at random. Subsequently they changed in conformity with the laws of natural selection. After several tens of thousands of generations, a species of beings having sufficiently expedient behavior was formed. This species proved to be stable and did not change until a new alteration of the food displacement law occurred.

References:

(1) Глушков, В. М., Н. М. Грищенко, А. А. Стогний [Glushkov, V. M., N. M. Grishchenko, and A. A. Stognii]. "Алгоритм распознавания осмысленных предложений" ["Algorithm for the Recognition of Intelligible Sentences"], Принципы построения самообучающихся систем [*Design Principles for Self-Learning Systems*], pp. 19-26. Kiev: State Publishing House for Technical Literature of the Ukrainian S.S.R., 1962.

(2) Летичевский, А. А., А. А. Дородницын [Letichevskiĭ, A. A., and A. A. Dorodnitsyn]. "Моделирование естественного отбора на вычислительной машине" ["Computer Simulation of the Process of Natural Selection"], Принципы построения самообучающихся систем [*Design Principles for Self-Learning Systems*], pp. 39-44. Kiev: State Publishing House for Technical Literature of the Ukrainian S. S. R., 1962.

5.4.10

LIBRASCOPE DIVISION OF GENERAL PRECISION, INC.
808 Western Avenue, Glendale 1, Calif.
HAROLD J. HAMILTON, *Director, and* RICHARD F. REISS, *Principal Investigator*

I. Nerve-Net Theory Research

Research in the field of nerve-net theory (1-3) is concentrated on two problem areas: the behavior of small groups of neurons having very specific connections and functions and the development of appropriate simulation techniques.

Over the past year, research has been concentrated on the development of flexible neural analog models using both conventional components and gas ion breakdown mechanisms. The dynamics of sodium-potassium fluxes in neural membrane have guided the development of the model. The sub-threshold behavior of the synaptic membrane is of primary interest in these models. These models are also being used in the study of small networks of neurons, particularly certain rhythmic and resonant phenomena in collections of 2 or 3 neurons.

The work is being materially aided by advice from Prof. T. H. Bullock (UCLA) and support from the U. S. Air Force Office of Scientific Research.

II. Research on Hypothetical Mechanisms Based on Classical Association Psychology Concepts

Research continues in the area of hypothetical mechanisms based originally on the concepts of classical association psychology (circa 1750-1900). A series of "abstract machines" which constitute formal, quantitative interpretations of associationist postulates are being studied in an attempt to discover the logical consequences of those postulates. Results of the first phase of this study are reported in (4).

Current work is concentrated on two areas: (a) further elaboration of the postulation system, taking into account modern psychological evidence to some extent, and analysis of the resulting "machine" behavior in relatively simple cases; and (b) development of statistical tools for dealing with large directed and undirected graphs of the types required to represent elaborate association networks. Both efforts are aimed at acquiring knowledge and analytical techniques which will make practical a digital simulation study of large association systems in the near future. It is believed that the ultimate capabilities of association mechanisms cannot be accurately estimated until systems involving 10,000 to 1 million associated objects have been investigated, and that digital simulation is the only tool adequate for such an investigation.

This research is supported by the U. S. Air Force Office of Scientific Research.

References:

(1) Josephson, R., R. Reiss, and R. Worthy. "A Simulation Study of a Diffuse Conducting System based on Coelenterate Nerve Nets," *Journal of Theoretical Biology,* vol. 1, 1961, pp. 460-487.

(2) Reiss, R. F. "A Theory and Simulation of Rhythmic Behavior Due to Reciprocal Inhibition in Small Nerve Nets," in *AFIPS Proceedings 1962 Spring Joint Computer Conference,* pp. 171-194. Palo Alto, Calif.: National Press, 1962.

(3) Reiss, R. F. "The Digital Simulation of Neuro-Muscular Organisms," *Behavioral Science,* vol. 5, 1960, pp. 343-358.

(4) Reiss, R. F. "An Abstract Machine Based on Classical Associa-

tion Psychology," in *AFIPS Proceedings 1962 Spring Joint Computer Conference*, pp. 53-70. Palo Alto, Calif.: National Press, 1962.

5.4.11

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Research Laboratory of Electronics, Cambridge 39, Mass.
M. L. MINSKY

Research leading toward ways of making machines behave intelligently continues, with no change reported since the previous statement [*Ed.*].

References:

(1) Minsky, Marvin. "A Selected Descriptor-Indexed Bibliography to the Literature on Artificial Intelligence," *IRE Transactions on Human Factors in Electronics,* vol. HFE-2, no. 1, 1961, pp. 39-55.

(2) Minsky, Marvin. "Steps toward Artificial Intelligence" (Special Computer Issue), *Proceedings of the IRE,* vol. 49, no. 1, January 1961, pp. 8-40.

(3) Minsky, Marvin, and O. G. Selfridge. "Learning in Random Nets," in *Symposium on Information Theory. Fourth, London, 1960,* ed. by C. Cherry. London: Butterworths Scientific Publications, and New York: Academic Press Inc. Publishers, 1961.

(4) Minsky, Marvin. "Recursive Unsolvability of Post's Problem of 'Tag' and other Topics in Theory of Turing Machines," *Annals of Mathematics,* vol. 74, no. 3, November 1961, pp. 437-455.

(5) McCarthy, John. "A Basis for a Mathematical Theory of Computation," in *Proceedings of the Western Joint Computer Conference,* May 1961. (Available from ACM)

(6) McCarthy, John. "Proof-checking by Computer," in *Proceedings of Symposium on Recursive Function Theory,* April 1961, American Mathematical Society.

(7) Slagle, James. *A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus.* Doctoral Thesis. Cambridge, Mass.: Massachusetts Institute of Technology, 1961.

5.4.12

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Electronic Systems Laboratory, Department of Electrical Engineering, Cambridge 39, Mass.
DOUGLAS T. ROSS, *Head, Computer Applications Group*

A program of research is underway on the application of the concepts and techniques of modern data processing to the design of mechanical

parts, and the further development of automatic programming systems for numerically controlled machine tools.

Current basic research is leading toward a man-machine system in which the human and computer can work together on creative design problems. Ultimately it is hoped that the output from the system can feed directly into advanced APT Systems to achieve a direct and automatic means for going from the conception of a part to the finished product.

The project presently is attacking the areas of structuring abstract problem elements, translation and compiling techniques, language design, automatic drafting and graphical techniques, and stress analysis.

In order to realize the very general verbal- and picture-language system for Computer-Aided Design, a penetrating analysis and understanding of language is required. During the past 6 months, an Algorithmic Theory of Language has been begun (an outgrowth of the First-Pass Algorithm studies), which starts from a number of very simple basic assumptions or axioms about the "most natural" way for linear languages to operate. On the basis of these assumptions a series of algorithms has been written for parsing and inserting a "precedence string" which shows how meaning is to be obtained from a parsed statement.

Whereas the First-Pass Algorithm was based upon a hierarchy of binding strengths of operators, the algorithmic theory depends only upon *datatypes,* i.e., what kind of thing does a word represent, and *word likes.* A binary matrix shows whether an individual word likes to be attached to another word, and the combination of checking the matching of datatypes, followed when necessary by checking a one-bit entry in the Like Matrix, provides a mechanism for parsing a very broad class of languages including ALGOL, probably COBOL, and a large class of English sentences. Semantics in the language influences the laying down of the precedence string, and this information in turn is based upon *precedence likes* which say whether the meaning of one word likes to be preceded by the meaning of another word. This information is contained in a binary Plike Matrix.

A series of extremely compact algorithms for parsing, precedence, and treatment of modifiers have been written and merged to make one algorithm which not only serves as the definition of the class of languages which can be handled at the present time but also provides an extremely efficient computer program. Additions to this algorithm for resolving ambiguities, etc., will be made in order to expand the realm of applicability of the algorithmic theory. At the same time the algorithms of the theory are being coded in the Bootstrap Plateau Language for constructing an ALGOL Compiler. Subsequent growth of that system will trend toward the overall Computer-Aided Design System which is the objective of the project.

In addition to the Algorithmic Theory of Language, additional related studies may be termed the beginning of the Theory of Meaning. It has been found that, with a properly parsed statement with precedence string inserted, if "rules" are defined to operate on atomic configurations and transform them into other atomic configurations, then these rules may be applied to a "precedence string follower" to transform the original language statement into another language statement which also is properly parsed. This procedure has been applied manually to many problems of symbol manipulation, including calculus derivatives, tree structure manipulations, factoring of algebraic expressions, factoring of Boolean expressions, basic natural-language translation, and various problems associated with efficient compilation of computer programs. The principle involved appears to be of broad generality and will be incorporated as an integral part of the Computer-Aided Design System to permit elaborate problem manipulation.

Several graphical language systems are being developed independently, and later results will be culled and incorporated in the final Computer-Aided Design System. All of the activities make use of "light pens" and computer output oscilloscopes so that displayed elements may be manipulated manually. One study is concerned with techniques for building up a general system for constructing pictures for using "light buttons" or control spots to indicate words in the language. Another is concerned with constructing three-dimensional pictures by manipulations of orthographic projections. The third and most elaborate system permits construction of very complex pictures by means of the "light pen" and a brief vocabulary of pushbuttons. Lines and circles can be combined in many ways with automatic satisfaction of geometric constraints; pictures may be used as subpictures for building up additional pictures; and rotation, translation, and change of scale and many other features are also incorporated. These graphical data processing studies will form an integral part of the verbal- and picture-language system for Computer-Aided Design.

The project is a cooperative venture between the Computer Applications Group of the Electronic Systems Laboratory and the Design and Graphics Division of the Mechanical Engineering Department, and is sponsored by the Division of Advanced Systems Technology of the U. S. Air Force Aeronautical Systems Division.

References:

For a complete list of publications, see *CRDSD*, No. 9, Statement No. 5.4.9. Electronic Systems Laboratory reports and technical memoranda are available through ASTIA and the Office of Technical Services, U. S. Dept. of Commerce. Bibliography on request to Publications Department, Electronic Systems Laboratory, M.I.T., Cambridge 39, Mass.

5.4.13 THE MITRE CORPORATION

Bedford, Mass.
EDWARD M. BENNETT, ROLLIN P. MAYER, JAMES BARTLETT, *and* DONALD E. WALKER

Work on a number of command language projects continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Bennett, Edward, and Joseph Spiegel. "Document and Message Routing Through Communication Content Analysis," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.

5.4.14 NATIONAL BUREAU OF STANDARDS

Data Processing Systems Division, Washington 25, D.C.
R. A. KIRSCH *and* B. K. RANKIN, III

The overall purpose of a program of research in English-language data processing and pictorial data processing is to combine the results of automatic text and diagrammatic picture processing and to provide, within the framework of a conceptual system called a Picture Language Machine, the capability of retrieving facts from documents that contain both linguistic and pictorial information.

Several tasks more or less directly connected with the Picture Language Machine are underway. A grammar is being written for a language called "Boole" which consists of the geometric arrays of symbols used in various textbooks for denoting Boolean algebraic expressions. Although this language is nominally artificial, it has some of the properties of informal, natural languages. A grammar of this language will be useful in processing the geometric arrays of "Boole" as input to a machine.

A comparative study of four list processing languages (COMIT, FLPL, IPL-V, and LISP) is being made. This study will compare the practical features of these languages that affect ease of programming or object language efficiency.

A survey is being made of certain parts of the automata theory literature having potential connection with information retrieval. An analysis of these potential connections along with an annotated bibliography of the literature is being prepared. A novel feature of the annotated bibliography will be a grammar for that part of English included in the annotations. Syntactical analysis of the annotations with respect to the grammar will be included (see 2.84).

A study is continuing of the formal properties of electronic circuit diagrams viewed as syntactical structures. It is intended that this study will eventually enable circuit diagrams to be analyzed at a more

macroscopic level than the single component level which represents the state of the character recognition art (see 2.114).

Work is continuing on the use of COMIT to provide a general-purpose syntactical analyzer for English sentences in a discontinuous constituent structure grammar.

The research is supported by the National Science Foundation and the U. S. Patent Office.

NORTHWESTERN UNIVERSITY 5.4.15

Technological Institute, Evanston, Ill.
GILBERT K. KRULEE, *Associate Professor of Industrial Engineering*

A study is being made of problem solving and learning behavior by means of computer simulation techniques, using an IBM 709 computer system. Emphasis is on areas of problem solving for which algorithms do not exist, and on the possibility of describing processes common to more than one area of problem solving. In addition, attention will be given to the study of programs that in turn can modify performance programs and therefore classify as adaptive learning processes. Interest in computer programming is secondary; primarily, an attempt is being made to develop better models of human behavior and to test the models against appropriate human performances.

One current effort is on a system to accept problem statements in a natural language (English) and to extract the relevant meaning from this problem statement. Following the phrase structure analysis of Chomsky (see 5.3.22), considerable progress has been made in the development of such a system. The writing of a computer program to simulate its behavior is currently underway. In addition, some experiments with human subjects are in progress in order to explore the relevance of Chomsky's theory to pertinent areas of human behavior.

The research is supported by the National Science Foundation.

PHILCO CORPORATION 5.4.16

Scientific Laboratory, Blue Bell, Pa.
J. S. BRYAN *and* L. N. KANAL

The application of a class of multidimensional, stochastic approximation methods to the adjustment of parameters of pattern recognition networks implementing linear and higher order classification functions is being investigated. The use of cascaded, linear decision functions (majority logic) in the recognition of two-dimensional visual patterns is being studied by means of simulation on the Philco 2000 computer using special-purpose input equipment (1). Emphasis is being placed on the use of cascaded decisions to achieve useful intermediate abstractions, the choice of optimum decision criteria, and techniques

which exploit statistical dependence among recognition criteria with a minimum amount of computation and hardware.

Reference:

(1) Hoffman, Allen. "The Whirling Dervish; A Simulation Study in Learning and Recognition Systems," in *1962 IRE Convention Record.*

5.4.17 PHILCO CORPORATION

Scientific Laboratory, Blue Bell, Pa.
J. B. CHATTEN *and* T. J. HARLEY, JR.

A research program is underway to determine techniques and criteria for subsequent development of a semiautomatic device for screening tactical surveillance imagery. The program includes experimental investigation of optical and electronic techniques for image-data processing and the application of various decision logics. This program is sponsored by the U. S. Army Electronics Research and Development Laboratory and is expected to be completed early in 1964.

A family of electronic correlation techniques for analyzing and interpreting military surveillance photography is also being developed. Systems based on the cross-correlation of transformed, abstracted, and spatially quantized representations (called feature maps) of the image data have been simulated on the Philco 2000 computer. The approach to target identification by means of processing the image data through several layers of abstracting logic has been simulated through three layers of feature maps. Changes of information in repeat cover aerial photography are detected by cross-correlation of the abstracted image data. The ability of this technique to detect significant change has been demonstrated with computer simulations under U. S. Air Force (Rome Air Development Center) sponsorship. Studies of various image abstraction and correlation techniques, employing both statistical and deterministic logics, are being carried out. The application of these techniques to other object recognition problems, such as personnel identification, is also being investigated.

5.4.18 PHYSIKALISCHES INSTITUT DER TECHNISCHEN HOCHSCHULE AACHEN

Templergraben 55, Aachen, Federal Republic of Germany
WILHELM FUCKS, *Director*

Work on (a) mathematical analysis of music, and (b) problems concerning the interpretation of the writings of Kant, using statistical linguistics, continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) "Mathematical Analysis of Music and Random-Sound-Sequences," translation from *Gravesaner Blätter.* Gravesano/Tessin, Switzerland: Hermann Scherchen. (To be published)

PRINCETON UNIVERSITY 5.4.19

Department of Music, Princeton, N. J.
MILTON BABBITT, *Adviser, and* MICHAEL KASSLER, *Principal Investigator*

A project concerned with the applications of digital computers to music and musicology was initiated in the summer of 1962. Preliminary work has been concentrated on compiling a bibliography of the pertinent literature and on programming explanations of music theory. Computer implementation of this latter work involves development of a data-processing "language" suitable for the treatment of written musical information; this in part has already been done (1).

Three general areas of research have been outlined for immediate study when the preliminary stage is completed: (a) the specification of an automatic "equivalence-preserving" conversion from the acoustic domain to the notational domain of conventional (Western) music, (b) the mechanization of numerous "secretarial" tasks now performed manually by musicologists, and (c) the programming of explications of particular theories of music so that they may be tested against a significantly large corpus of compositions for "goodness of fit."

References:

(1) Kassler, Michael. "A System for the Automatic Reduction of Musical Scores," in *Papers Presented at the Seminar in Mathematical Linguistics,* vol. 6 (1960), on deposit at Widener Library, Harvard University, Cambridge, Mass.

(2) Kassler, Michael. *The Decision of Arnold Schoenberg's Twelve-note-class System and Related Systems,* Princeton University, 1961.

THE RAND CORPORATION 5.4.20

1700 Main Street, Santa Monica, Calif.
M. E. MARON

The purpose of research underway is to clarify and gain insights into the question of organizational principles for intelligent systems (both real and artificial).

The work is supported by the U. S. Air Force, Project RAND.

Reference:

(1) Maron, M. E. *Mechanisms Underlying Predictive Behavior for an Intelligent Machine,* RM-3011 PR, February 1962.

5.4.21 THE RAND CORPORATION

Santa Monica, Calif.

and

CARNEGIE INSTITUTE OF TECHNOLOGY

Pittsburgh 13, Pa.

ALLEN NEWELL,[1] JOHN C. SHAW,[2] *and* HERBERT A. SIMON[1]

Research into information processes underlying human intellectual, adaptive, and creative abilities—i.e., the domain of complex information processes—is being conducted. The fundamental technique of the project is the synthesis of intelligent systems on digital computers; the technique also is being developed by creation of new programming languages. The effort of the project is divided into three parts: (a) psychology of mental processes (1) (4-6); (b) heuristic programs, working primarily with formal tasks like chess and proving theorems in elementary symbolic logic (3) (7) but also some management science tasks (8); and (c) information processing languages, focusing on the "list" as the unit of information, with such processes as insertion, deletion, and copying a list as the basic processes (2).

The work arrangement between RAND and the Carnegie Institute of Technology is informal. Within the Carnegie Institute of Technology the work is supported in part by a grant from the Carnegie Corporation.

References:

(1) Feigenbaum, E., and H. A. Simon. "Forgetting in an Association Memory," in *Proceedings of the 1961 National Conference of the ACM,* August 1961, pp. 202-205. (RAND P-2311)

(2) Newell, A. (ed.). *Information Processing Language V Manual.* Englewood Cliffs, N. J.: Prentice-Hall Book Co., Inc., 1961. 244 p. (RAND P-1897 and P-1918)

(3) Newell, A., J. C. Shaw, and H. A. Simon. "A Variety of Intelligent Learning in a General Problem Solver," in *Self-Organizing Systems,* ed. by C. Yovits and S. Cameron, pp. 153-189. New York: Pergamon Press, Inc., 1960. (RAND P-1742)

(4) Newell, A., and H. A. Simon. "Computer Simulation of Human Thinking and Problem Solving," in *Management and the Computer of the Future,* ed. by M. Greenberger, pp. 95-131. New York: John Wiley & Sons, Inc., 1962. (RAND P-2312. Also reprinted in *DATAMATION,* June-July 1961.)

(5) Newell, A., and H. A. Simon. "Computer Simulation of Human Thought," *Science,* vol. 134, no. 3495, December 1961, pp. 2011-2017. (RAND P-2276)

[1] Carnegie Institute of Technology
[2] The RAND Corporation

(6) Newell, A., and H. A. Simon. "GPS, A Program that Simulates Human Thought," in *Lernende Automaten* (proceedings of a conference at Karlsruhe, Germany, April 1961), ed. by H. Billing, pp. 109-124. Munich: R. Oldenbourg, 1961. (RAND P-2257)
(7) Simon, H. A. *Experiments with the Heuristic Compiler,* RAND P-2349, June 1961. 85 p.
(8) Tonge, F. M. *A Heuristic Program for Assembly Line Balancing.* Englewood Cliffs, N. J.: Prentice-Hall Book Co., Inc., 1961. 114 p. (Carnegie Institute of Technology Ph.D. dissertation. Also RAND P-1993.)

ROCKFORD RESEARCH INSTITUTE INC. 5.4.22

140½ Mount Auburn Street, Cambridge 38, Mass.
RAY J. SOLOMONOFF

Study of the theory of inductive inference continues.

A paper summarizing the work on inductive inference during the past several years has been prepared and submitted for publication.

Another paper (1) presents a description of a new method of inductive inference in arithmetic learning. This topic was the subject of much investigation in an earlier paper (2). Some quantitative results have been obtained in applying the present theory of inductive inference to continuous data. In particular, the theory applied to the fitting of curves to empirical data seems to give intuitively reasonable results which are compatible with results obtained by other means.

This research is supported by the U. S. Air Force Office of Scientific Research and the National Institutes of Health.

References:

(1) Solomonoff, R. J. "Training Sequences for Mechanized Induction," presented at the Conference on Self Organizing Systems, Chicago, May 22-24, 1962.
(2) Solomonoff, R. J. "An Inductive Inference Machine," *I.R.E. Convention Record,* 1957 (section on Information Theory).

STICHTING STUDIECENTRUM VOOR ADMINISTRATIEVE AUTOMATISERING 5.4.23

[Netherlands Automatic Data Processing Research Center]

Stadhouderskade 6, Amsterdam, The Netherlands
MAX EUWE, *Project Director*

Work on the project to analyze the thinking process of a chess player and to simulate the process by means of a computer continues, with no change reported since the previous statement [*Ed.*]. The project is supported by contract with Euratom.

Reference:

(1) Toumanov, V. "The Best Move," *Chess Bulletin Moscou,* vol. 8, April 1961.

5.4.24 SYSTEM DEVELOPMENT CORPORATION

2500 Colorado Avenue, Santa Monica, Calif.

FRANK N. MARZOCCO, *Head, Artificial Intelligence Research Staff*

I. Complex Search Processes In Intelligent Automata and Man-automata Combinations (George P. DeFlorio)

Complex search processes in intelligent automata and man-automata combinations are being investigated (1)(2). Two immediate objectives are (a) program models of automata that efficiently perform search and selection operations within complex environments whose structures are initially unknown, and (b) laboratory realization of man-automata combinations, cooperatively solving difficult search and decision problems in real time.

Progress has been made toward both these goals. Several programs have been written in JOVIAL for both pure-machine and real-time operations on a Philco 2000 computer. A real-time display and command capability has been implemented in SDC's Systems Simulation Research Laboratory. A program that permits real-time human intervention during the search process has been developed; it has the same characteristics as the programs used for pure machine investigations. Real-time behavior of the search automaton is displayed on a 14-inch CRT. Weighted graph structures are used in displaying both problem space and decision sequences. In man-aided search, the problem environment, decision sequence, and strategy state of the automaton are displayed via the CRT. The human operator can aid or direct the simulated automaton via a command keyboard on a strategy, single-decision, or subgoal level. Current effort is directed toward the extension and exploration of the capabilities of these present programs within well-defined problem environments.

II. Complex Learning and Problem Solving By Machine (Aiko Hormann, L. E. Travis, and S. S. Shaffer)

Investigation of complex learning by machine continues. Emphasis is placed on problems of utilizing and exploiting the self-modifying capabilities of stored-program computers. These capabilities appear to provide the basis for interesting and powerful learning by machine.

Work thus far on the task of constructing a system capable of synthesizing simple puzzle-solving programs and of improving them on the basis of experience has been reported (3). The system is structured as a hierarchy of program-generating units, each of which manifests a feed-

back structure and thus is capable of testing tentative output before reporting that output to controlling, higher level units.

A program system has been constructed which synthesizes programs for a very simple computer best described as a memoryless vector adder. Though this can hardly be called self-programming because the program-synthesizing (IPL-V) computer is a far more complex and powerful machine than the object machine (the vector adder), constructing the system and experimenting with it have helped to illuminate several aspects of the general problem of mechanical sequence synthesis. A report is being prepared.

Because of the far-reaching analogy between theorem proving and program writing, theorem proving is being studied and a computer-based theorem proving system is being constructed. Part of this work has been concerned with painstaking observation of how relatively skilled human beings go about the task of synthesizing a derivation in the propositional calculus. Some of the more important methodological and substantive conclusions to be drawn from this observation have been reported (10).

Another part of the research on self-programming is aimed at making clear the potential methodological function of learning machines as a vehicle for clarifying and amplifying learning theories. Based on this, an attempt is being made to clarify and amplify the particular learning theory of E. C. Tolman in terms of information processing, programming, and self-programming.

The programming involved in the research is being done in IPL-V. To make this possible, an IPL-V system for the Philco 2000 computer has been constructed (5) (6).

III. Contiguity Models (Frank N. Marzocco)

The stimulus sampling computer program has been written so that additional code may be added easily for carrying out particular investigations. One class of investigations is aimed toward improvement of techniques for predicting the behavior of organisms. For this purpose, the Philco 2000 computer is being used to produce data simulating bar pressing behavior of rats in a Skinner box, and the data are being used in turn to determine optimal settings of the program parameters. Another class of investigations is concerned with ways in which intelligent behavior may be built up from simple components. The computer has been programmed to simulate both players in relatively simple board games. The aim is to determine whether the computer can discover appropriate strategies if given only the information as to legal moves on any play and knowledge of win, lose, or draw for each player at the end of the game.

IV. Computer Synthesis of Human Language Behavior (Robert F. Simmons and Sheldon Klein)

The Synthex research project is devoted to the development of general-purpose computer systems for the synthesis of human-type cognitive functions (4)(7-9).

As an initial research vehicle a prototype language processing system, Protosynthex I, has been programmed for the IBM 7090 computer and translated to the Philco 2000 computer and to a large military computer. Protosynthex I reads and indexes English text and selects answers in response to ordinary English questions about the text. In order to accomplish this task it uses four basic subsystems: (a) the indexer, (b) the question analyzer and information retrieval unit, (c) the grammatical analysis system, and (d) the answer evaluation system. In general, the system works well with simple questions of fact. As keypunched data base, 16 volumes of the *Golden Book Encyclopedia,* an extract of *Scientific American* text, and two articles from the *Encyclopedia Americana* have been used. The system is general enough to handle hundreds of thousands of words of ordinary English text although in a simple-minded fashion.

Most recent work on the Synthex project includes the early-stage development of a system that translates English commands into computer actions—namely, the movement of blips on a display console in response to such English statements as "Fly to Boston." Following Yngve's recent work, a system to generate grammatically correct but nonsensical statements has been developed. Programs have also been written for the generation of coherent discourse which is a paraphase of an input text. At this writing the results are promising although not completely satisfactory.

References:

(1) Bartram, P. R. *A Computer Program that Operates with Networks.* SDC document SP-798, May 1962. 17 p.

(2) DeFlorio, G. P. *Complex Search Processes in Automata and Man-Automata Combinations.* SDC document SP-842, June 1962. 8 p.

(3) Hormann, A. M. *Programs for Machine Learning,* SDC document TM-669/000/01, May 29, 1962. 76 p.

(4) Klein, S., and R. F. Simmons. *A Computational Approach to Grammatical Coding of English Words,* SDC document SP-701, February 22, 1962. 25 p. (To be published in *Journal of the Association for Computing Machinery,* January 1963)

(5) Shaffer, S. "Current Status of IPL-V for the Philco 2000 Computer," *Communications of the ACM,* vol. 5, no. 9, September 1962, p. 479.

(6) Shaffer, S. *Program Description of Philco Transac Implementation of IPL-V,* SDC document FN-6627, June 11, 1962. 9 p.
(7) Simmons, R. F., S. Klein, and Keren McConlogue. "Toward Computer Synthesis of Human Language Behavior" (SP-466), *Behavioral Science,* July 1962, pp. 402-407.
(8) Simmons, R. F. "Synthex: Computer Synthesis of Human Language Behavior," in *Computer Applications in the Behavioral Sciences,* ed. by H. Borko, pp. 360-393. Englewood Cliffs, N. J.: Prentice-Hall Book Co., Inc., 1962.
(9) Simmons, R F., and Keren McConlogue. *Maximum-Depth Indexing For Computer Retrieval of English Language Data,* SDC document SP-775, April 10, 1962. 22 p. (To be published in *American Documentation,* January 1963)
(10) Travis, L. *Observing How Humans Make Mistakes to Discover How to Get Computers to do Likewise,* SDC document SP-776, June 25, 1962. 21 p.

TECHNISCHE HOCHSCHULE KARLSRUHE 5.4.25

Institut für Nachrichtenverarbeitung und Nachrichtenübertragung, Karlsruhe, Federal Republic of Germany
K. STEINBUCH, *Director*; U. PISKE *and* H.-J. HÖNERLOH, *Project Leaders*

A research program on learning machines is primarily concerned with the realization and applications of the learning matrix (LM). Applications of the LM are seen in character recognition, automatic speech recognition and translation, information retrieval, medical diagnosis, meteorology (weather forecasting), error-correcting devices, self-correcting circuits, self-programming, and other problems of artificial intelligence. A special kind of LM is able to process continuous information and shows properties similar to those of human perception such as recognition and classification of patterns like chords, shapes, or graphs, the pattern being invariant to affine transformation, parallel shifts, and, to some extent, to distortions.

Simple demonstration models have been constructed. At present, a set of standardized LM units, using ferromagnetic cores with nondestructive readout, is under construction. Interactions between coupled LMs will be investigated. One main goal is the realization of inexpensive LMs with large numbers of intersection points.

The project, supported by the Deutsche Forschungsgemeinschaft, was started in 1960.

References:

(1) Steinbuch, K. "Die Lernmatrix" ["The Learning Matrix"], *Kybernetik,* vol. 1, no. 1, January 1961, pp. 36-45.

(2) Hönerloh, H.-J., and H. Kraft. "Technische Verwirklichung der Lernmatrix" ["Technical Realization of the Learning Matrix"], in *Lernende Automaten.* Munich: R. Oldenbourg, 1961.

(3) Piske, U. "Demonstrationsversuche mit der Lernmatrix" ["Demonstrations on the Learning Matrix"], in *Lernende Automaten.* Munich: R. Oldenbourg, 1961.

(4) Goerke, W., H. Kazmierczak, and S. W. Wagner. "Anwendungen der Lernmatrix" ["Applications of the Learning Matrix"], *Lernende Automaten.* Munich: R. Oldenbourg, 1961.

(5) Steinbuch, K., and H. Frank. "Nichtdigitale Lernmatrizen als Perzeptoren" ["Non-digital Learning Matrices Acting as Perceptors"], *Kybernetik,* vol. 1, no. 3, December 1961, pp. 117-124.

(6) Steinbuch, K. *Automat und Mensch. Über menschliche und maschinelle Intelligenz* [*Automata and Men. On Human and Artificial Intelligence*]. Berlin: Springer-Verlag, 1961. 253 p.

5.4.26 UNIVERSITÀ DEGLI STUDI DI MILANO

Centro di Cibernetica e di Attività Linguistiche, Milan, Italy
SILVIO CECCATO, *Principal Investigator*

A project was undertaken in July 1960 for the construction of a machine capable of performing the human operations of perception, representation, †mental categorization, †thought, and speech. This "Talking Automaton" will constitute a kind of miniature reporter, able to observe and to describe in words the events that take place on a stage situated in front of it. This stage will measure 3 meters by 2 meters, and on it there will be plates and glasses, fruits, and vegetables that can be moved on the ends of strings like puppets; there will also be a remote-controlled toy tortoise. The machine will have to describe the events that take place on this stage, and to do this it will have to adopt one of six possible *attitudes* to the situation that confronts it: the purely descriptive, the scientific (or explanatory-predictive), the ethical, aesthetic, sympathetic, and antagonistic.

The project is based on the results of a long critical inquiry into the traditional notions (the philosophical approach and the psychological one derived from it) concerning the human mind and its working. For project purposes, not only how man thinks but also what he thinks is considered operationally and analyzed in terms of operations; the results of these studies are never indicated in terms that are either

† *See* Glossary.

irreducibly metaphorical or contradictory. For if the relevant indications were such (as in fact, philosophical descriptions have hitherto been), the things indicated by them would have to be put into the machine ready made—which, in any case, would be quite impossible with the "abstract entities" that are characteristic of the traditional notions concerning the human mind.

The operational approach has made possible the reduction of all contents of thought to three kinds of operation; i.e., any content can always be obtained by means of one of these three operations or by a combination of them. The three kinds of operation are †differentiation, †figuration, and †mental categorization. Thinking, then, constitutes a fourth kind of operation which consists of giving the results of the other three kinds a temporal order, and this is called correlating.

This project is being sponsored by Euratom through its Commission de Recherche et Enseignement (Brussels); it is expected that the first model will be constructed within the next 2 years.

As a corollary to this project, the Italian Government, through its National Research Council, has sponsored a team of neurologists and anatomists to work under the direction of Professors Ceccato, Bairati, and Visintini on the twofold problem of coordinating neuroanatomical research and the analytical studies of thought and language carried out by the Centro di Cibernetica.

References:

(1) Ceccato, Silvio. *"Il linguaggio con la tábella di Ceccatieff"* ("Actualitiés Scientifiques et Industrielles"). Paris: Hermann & Cie Editeurs, 1951. (Italian and English text)

(2) Ceccato, Silvio. "La machine qui pense et qui parle," in *I Congrès International de Cybernétique (Namur, 1956)*, pp. 288-299. Paris: Gauthier Villars, 1959.

(3) Ceccato, Silvio. "Tappe nello studio dell'uomo," *Primo quaderno di 'Methodos,'* pp. 1-84. Milan: Feltrinelli Editors, 1959.

(4) Marietti, Enrico. "The Mechanisation of Perception," presented at the Nonnumerical Data Processing Symposium, Blaricum (Holland), 1961.

(5) Incarbone, Salvatori. *Programma di Inseguimento di Contorni,* Euratom, Cetis. (In press)

(6) Ceccato, Silvio. "La Meccanizzazione delle Attività Umane Superiori," *Civiltà delle Macchine,* vol. 9, no. 4, Rome, 1961, pp. 22-29.

(7) Ceccato, Silvio. "L' Osservazione nell' Uomo e nella Macchina," *ibid,* vol. 10, no. 2.

† *See* Glossary.

(8) Ceccato, Silvio. "Il capitolo più difficile della cibernetico," *Rivista Pirelli,* vol. 15, no. 3, Milano, 1962, pp. 82-84.
(9) Ceccato, Silvio. "Cibernetica e Linguistica," in *Cultura e Scuola.* (In press)
(10) Ceccato, Silvio. "Suggestions for Anthropology: The Machine Which Observes and Describes," presented at the symposium: The Use of Computers in Anthropology, Wenner Gren Foundation for Anthropological Research, Burg Wartenstein, summer 1962. (Proceedings in press)
(11) Ceccato, Silvio. "La macchina che ossezro e descrive," *La Ricerco Scientifico,* vol. 2, no. 3-4 (March-April 1962), CNR, Rome.

5.4.27 UNIVERSITY OF CALIFORNIA

Center for Research in Management Science, Berkeley 4, Calif.
EDWARD A. FEIGENBAUM *and* JULIAN FELDMAN

Research is being conducted into (a) human learning, decision making, and problem solving, and (b) problems of artificial intelligence.

The research strategy for the work on simulation of cognitive processes is to specify information-processing models of behavior in the form of computer programs. The behavior of the models is studied by using them as subjects in psychological experiments simulated on the computer. To evaluate and improve the models, this behavior is then compared with human behavior in corresponding real experiments (4).

In the current phase of the research, models of human verbal learning and of two-choice behavior are being extended to include behavior that earlier versions of these models could not predict. For the verbal learning model (called EPAM, for Elementary Perceiving and Memorizing machine), these extensions are in the following directions: processes for the learning of meaningful verbal material, more elaborate discrimination learning processes, and a general process for associating earlier learned stimulus objects into later learned, more complex objects (1-3). For the model of two-choice behavior, these extensions will provide ability to predict various types of patterned sequences. Work is also being done on the problems involved in testing and evaluating computer models of thinking processes. Current research on artificial intelligence models includes the implications of the EPAM model for certain problems in the formulation of machine pattern recognition schemes and the use of the basic scheme of the Newell-Shaw-Simon General Problem Solver for pattern recognition and prediction in the two-choice model. Research also continues on more effective information-processing computer languages for research on simulation of cognitive processes and artificial intelligence.

Most of the programs are written in IPL-V (7). The experiments have been conducted on the IBM 704, IBM 7090, and Philco S-2000 computers.

The major source of support for this research is a grant from the Carnegie Corporation of New York. Substantial support is also provided by the RAND Corporation and the System Development Corporation. Supplementary support has been given by the Western Management Science Institute and the Computation Center of the University of California, Berkeley. The verbal learning research is a joint effort with H. A. Simon of Carnegie Institute of Technology (see 5.4.21).

References:

(1) Feigenbaum, E. A. "Simulation of Verbal Learning Behavior," in *Proceedings of the Western Joint Computer Conference,* vol. 19, 1961, pp. 121-132.

(2) Feigenbaum, E. A., and H. A. Simon. *Performance of a Reading Task by an Elementary Perceiving and Memorizing Machine,* RAND Corporation Paper P-2358, July 1961.

(3) Feigenbaum, E. A., and H. A. Simon. "Generalization of an Elementary Perceiving and Memorizing Machine," presented at the International Federation of Information Processing Societies conference, Munich, August 1962. (Also RAND Corporation Paper P-2555, March 1962)

(4) Feldman, Julian. "Computer Simulation of Cognitive Processes," in *Computer Applications in the Behavioral Sciences,* ed. by H. Borko. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962.

(5) Feldman, Julian. "Simulation of Behavior in the Binary Choice Experiment," in *Proceedings of the Western Joint Computer Conference,* vol. 19, 1961, pp. 133-144.

(6) Feldman, Julian, Fred Tonge, and Herschel Kanter. "Empirical Explorations of a Hypothesis-Testing Model of Binary Choice Behavior," *Simulation of Cognitive Processes Project Working Paper #2,* September 1961 (revised December 1961, April 1962). Also appeared as SP-546, Systems Development Corporation, December 19, 1961.

(7) Newell, Allen, ed. *Information Processing Language V Manual.* Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1961.

5.4.28 UNIVERSITY OF CALIFORNIA

School of Business Administration, Berkeley 4, Calif.
and
SYSTEM DEVELOPMENT CORPORATION
2500 Colorado Avenue,
Santa Monica, Calif.
JOHN T. GULLAHORN *and* JEANNE E. GULLAHORN

Research is continuing[1] on refining a model for human behavior in simple social situations. The model derives from the work of George C. Homans (3). Computer programs have been written in IPL-V to enable computer-simulated group participants to receive and evaluate stimuli, to predict the reinforcement value of alternative responses and select socially profitable activities, and to modify their perceptions of group members during the course of interaction so that their responses will exhibit increasing adaptation to the goal structure of the situation.

Simulation of actual small group experiments using the model is in the planning stages and will be programmed for the computer and carried out during the coming year. Programs of two persons interacting in an office situation are now being run on the Philco 2000 computer.

References:

(1) Gullahorn, John T., and Jeanne E. Gullahorn. *A Computer Model of Elementary Social Behavior,* SDC document SP-741, March 1962. 15 p.
(2) Gullahorn, John T., and Jeanne E. Gullahorn. *The Computer as a Tool for Theory Development,* SDC document SP-817, June 1962. 22 p.
(3) Homans, George C. *Social Behavior.* New York: Harcourt, Brace and World, Inc., 1961.

5.4.29 UNIVERSITY OF CHICAGO

Committee on Mathematical Biology, Chicago 37, Ill.
PETER H. GREENE, *Principal Investigator*

Investigations are continuing with the aim of understanding some of the features of human perception, by learning how they might arise in rudimentary form in mathematically described models.

Attempts are being made to model on a digital computer some of the processes whereby an infant develops skilled actions involving objects and spatial relations.

The next phase of the project to be undertaken will attempt to develop a theory of what must be represented in the memory of a

[1] Project reported in *CRDSD,* No. 10, under Michigan State University.

skilled device. For example, if it has to manipulate an object in a fixed position, the memory might have to represent a numerical function giving the relation between a machine variable and the resulting effect on the object. But to be able to handle objects wherever they may be, the machine must represent some sort of bundle of functions, each corresponding to a particular potential position of the object. There are structural relations among the parts of this bundle, and the task is to find a representation of the whole relational structure. At present, no one really even knows the significant issues to talk about with regard to these structures. Some recent developments in mathematics offer promise of giving some insight into what some of these issues may be.

This work is supported by the U. S. Office of Naval Research.

References:

(1) Greene, P. H. "Some Biological Ideas," in *Bionics Symposium. Living Prototypes—The Key to New Technology,* WADD Technical Report 60-600. Wright-Patterson Air Force Base, Ohio: Directorate of Advanced Systems Technology, Wright Air Development Division, A.R.D.C., U. S. Air Force Base, December 1960.

(2) Greene, P. H. "An Approach to Computers that Perceive, Learn, and Reason," in *Proceedings of the Western Joint Computer Conference,* San Francisco, 1959, pp. 181-186. (ASTIA No. AD-213 855)

(3) Greene, P. H. "On the Representation of Information in Neural Net Models," in *Self-Organizing Systems* (proceedings of the Conference on Self-Organizing Systems, Chicago, May 1962). Washington, D. C.: Spartan Books, 1962.

UNIVERSITY OF ILLINOIS **5.4.30**

Department of Electrical Engineering
Urbana, Ill.

HEINZ VON FOERSTER, *Director, Biological Computer Laboratory*

Computational principles of living organisms are studied as a point of departure for analysis and construction of cognitive systems.

Theoretical work is directed toward the study of properties of systems with large numbers of local equilibria to identify their general statistical features and to relate these to such behavioral phenomena as multistability and the conditioned reflex.

A logical study of the reliability of automata and automatic error correction has been undertaken. Special emphasis is placed on ideas concerning stochastic switching circuits, metric reliability, multivalued logics, and ordinal logics.

Experimental work is concerned with the design and construction of property detector nets which extract certain relevant characteristics from an n-dimensional pattern and encode these properties in a manner which can be utilized for perception and recognition purposes. A dynamic signal analyzer has been constructed for the purpose of studying auditory signals, which may be considered as a special class of patterns. The dynamic signal analyzer consists of three stages of 96 special-purpose, continuous-variable computers operating in parallel, which extract certain properties from the auditory signals. These properties are the dynamic frequency spectrum distribution (3-cycle response time) and the second and fourth derivatives (with respect to log frequency) of that distribution.

The work is sponsored in part by the National Science Foundation, the U. S. Air Force (Wright Air Development Division), and the U. S. Office of Naval Research.

References:

(1) Ashby, W. R. "Cybernetics Today and Its Future Contribution to the Engineering Sciences," presented before the annual meeting of the Foundation for Instrumentation Education and Research, Inc., Washington, D. C., October 1961. New York: Foundation for Instrumentation Education and Research. 16 p.

(2) Ashby, W. R. "Homeostasis," in *Encyclopedia of Biological Sciences,* ed. by P. Gray. New York: Reinhold Publishing Corp., 1961.

(3) Ashby, W. R. "Principles of the Self-Organizing System," in *Principles of Self-Organization,* ed. by H. Von Foerster and G. W. Zopf, Jr., pp. 255-278. New York: Pergamon Press, Inc., 1962.

(4) Babcock, M. L., et al. *A Dynamic Signal Analyzer,* Electrical Engineering Research Laboratory, University of Illinois, Urbana, 1962. 77 p.

(5) Gunther, G. "The Tradition of Logic and the Concept of a Trans-classical Rationality," *Alg. Nederlands Tijds. Wijsbegeerte en Psychol.,* vol. 54, 1962, pp. 194-200.

(6) Gunther, G. *Cybernetic Ontology and Transjunctional Operations,* Electrical Engineering Research Laboratory, University of Illinois, Urbana, 1962. 113 p.

(7) Hsieh, H. S., et al. *Determination of Equivalence Classes by Orthogonal Properties,* Electrical Engineering Research Laboratory, University of Illinois, Urbana, 1962. 103 p.

(8) Inselberg, A., et al. *Property Extraction in Linear Networks,* Electrical Engineering Research Laboratory, University of Illinois, Urbana, 1962. 65 p.

(9) Lofgren, L. "Limits for Automatic Error Correction," in *Principles of Self-Organization,* ed. by H. Von Foerster and G. W. Zopf, Jr., pp. 181-228. New York: Pergamon Press, Inc., 1962.

(10) Lofgren, L. *A Theory of Uniform Switching Nets,* Electrical Engineering Research Laboratory, University of Illinois, Urbana, 1962. 53 p.

(11) Mullin, A. A. "On the Nature of the Reliability of Automata," in *Redundancy Techniques for Computing Systems.* Washington, D. C.: Spartan Books, 1962.

(12) Von Foerster, H. "Bio-Logic," in *Biological Prototypes and Synthetics Systems,* ed. by E. E. Bernard and M. R. Kare, pp. 1-12. New York: Plenum Press, 1962.

(13) Von Foerster, H. "Communication Amongst Automata," *American Journal of Psychiatry,* vol. 118, 1962, pp. 865-871.

(14) Von Foerster, H., and G. W. Zopf, Jr., eds. *Principles of Self-Organization.* New York: Pergamon Press, Inc., 1962. 541 p.

(15) Weston, P. "Photocell Field Counts Random Objects," *Electronics,* September 22, 1961, pp. 46-47.

(16) Zopf, G. W., Jr. "Attitude and Context," in *Principles of Self-Organization,* ed. by H. Von Foerster and G. W. Zopf, Jr., pp. 325-346. New York: Pergamon Press, Inc., 1962.

THE UNIVERSITY OF MICHIGAN 5.4.31

Mental Health Research Institute, Ann Arbor, Mich.
JAMES G. MILLER, *Director*

I. HEURISTIC PROGRAMMING (Merrill M. Flood)

Work on the development of mathematical models that can be used to explore the decision-making process in humans, and of techniques for solving problems that have not previously been susceptible to numerical analysis, continues, with no change reported since the previous statement [*Ed.*] (1) (2).

II. MODELS OF THOUGHT PROCESSES (John W. Gyr and Albert C. Cafagna)

The purpose of continuing research is to test various models of thought processes by their simulation on a high-speed digital computer and then by comparing the computer output with results from problem-solving experiments with humans (3-5). A new extension of this work involves concept-forming processes which do not necessarily generate highly abstract logical behavior. Thus there has been an interest in the thought processes of children, in the area of thought usually labeled as "common sense," as well as in highly abstract logical

thought, and in the interrelations between these processes. This extension has been made in order to broaden and generalize the theories of intellectual processes.

The continuing part of the study involves computer programming to generate behaviors which may be compared with the problem-solving data obtained from human subjects. The new series of experiments concerns the creation of learning situations in which, by careful manipulation of the physical environment, the subject (i.e., either the child or the computer) is taught to form concepts which are less and less dependent on specific environmental stimuli for their evocation, but rather are evolved contextually and through purely symbolic means. Ultimately, this will lead to the generation of abstract concepts which refer to objects or events never as yet observed by the subject.

In its ongoing phase, the research has consisted of perfecting the computer program which was written to test the second model of intelligent processes: the model in which it is assumed that the problem-solver is a "pragmatist" who will concentrate on those hypotheses which are most likely to be true, given certain evidence, and will discard other possibilities until forced to consider them by new evidence. One of the problems with the program, namely, its showing a much higher repeat-of-old-trials rate than do human subjects, is being resolved by fairly fundamental changes in the underlying process. The new phase of the research is still in the planning stage.

This project has been partially supported by the Social Science Research Council.

III. Information Processing in the Nervous System (W. J. Horvath and B. Peretz)

Continuing research is directed at understanding the operation of the nervous system as a processor of information.

In work on the visual system of the frog, recordings are being made from single fibers in the optic nerve, and the measurements will be analyzed and related to the visual stimuli presented to the animal.

Electronic models are being used to simulate neurons. Six models, patterned after the design developed by Leon Harmon of the Bell Telephone Laboratories, have been constructed and tested. It is planned to use these units to simulate the function of cells in the somatic sensory system.

References:

(1) Flood, Merrill M. "Adaptive System Models," in *Proceedings of the Second International Conference on Operational Research,* pp. 11-20. London: English University Press, Ltd., 1961. (MHRI Preprint 37)

(2) Flood, Merrill M. "Stochastic Learning in Rats with Hypothalamic Implants," *Annals of the New York Academy of Sciences,* 1961, vol. 89, Art. 5, pp. 795-822.

(3) Gyr, J. "An Investigation Into, and Speculations About, the Formal Nature of a Problem Solving Process," *Behavioral Science,* vol. 5, 1960, pp. 39-60.

(4) Gyr, J., and A. Cafagna. *Simulation of Three Models of Cognitive Behavior,* Mental Health Research Institute Preprint 72.

(5) Gyr, J., J. Thatcher, and G. Allen. "Computer Simulation of a Model of Cognitive Organization," *Behavioral Science,* vol. 7, 1962, pp. 111-117.

UNIVERSITY OF PENNSYLVANIA **5.4.32**

The Moore School of Electrical Engineering, Philadelphia, Pa.

GEORGE W. PATTERSON *and* HARRY J. GRAY

A study has been completed of the feasibility and utility of "building into" the equipment design of a computer the capability of interpreting and executing order structures expressed in mechanical languages (4). The study was also directed to determine the system organization of such a computer and to continue the development of the theory of mechanical languages and computing systems with reference to Air Force problems.

At first, research was concerned with the design of a processor to handle string-processing languages which appear as a special instance of linear sequential mechanical languages. The processor is assumed to use string languages at various levels in its internal structure. ALGOL, COBOL, and FORTRAN are examples of linear sequential mechanical languages and are also referred to as problem-oriented computer languages. Although it is not strictly a string-processing language, ALGOL has been modified by Wegstein (1) and others to make it such a language. Since the linear sequential mechanical language appears to be of greater scope than the string languages, it was necessary to alter the scope of the research to include the processor which could handle the more general linear sequential mechanical languages.

At present, "mechanical-language processor" means a processor the input language of which is a linear sequential mechanical language and the internal structure of which is such that it can learn to recognize and execute programs written in any such language. It is considered possible to construct such a machine from the known fact that a human being can learn and accept instructions in mechanical language as well as in natural languages.

To begin with, mechanical-language processing requires that data and instructions be called for by name in order to free the programmer from the housekeeping task of location assignment. The processing must also allow for an expandable language structure so that previously undefined processes may be added to the language and then may be called for by name during future processing. The expandable language structure will also enable the processor to handle the housekeeping tasks needed for retrieval by name. Thus, there must be some means for building the expandable language structure.

This type of processing can be done by a general-purpose computer operating with a suitable interpreter; however, a more specialized machine design may be a way of increasing the efficiency of operation of the processor. The general-purpose computer operating in an interpretive mode may operate so slowly that the advantages of mechanical-language processing may be lost. A special mechanical-language processor can be designed which will be more compatible with these mechanical-language processes and, thus, make full use of their capabilities.

The project was sponsored by the U. S. Air Force (Rome Air Development Center).

References:

(1) Wegstein, J. H., and W. W. Youden. "A String Language for Symbol Manipulation Based on Algol 60," *Communications of the ACM,* vol. 5, no. 1, January 1962, pp. 54-61.

(2) Landauer, W. I., and N. S. Prywes. "A Growing Tree for Descriptor Language Translation," presented at the Symposium on Symbolic Languages in Data Processing, March 1962.

(3) Lowe, Thomas C. *State Reduction in Partially Specified Sequential Machines,* Technical Note prepared for Rome Air Development Center under contract AF 30 (602)-2382, Interactions of Computer Languages and Machine Design, May 18, 1962.

(4) University of Pennsylvania, Moore School of Electrical Engineering. *Study of Interactions of Computer Languages and Machine Design,* Final Report on Rome Air Development Center contract AF 30 (602)-2382, September 30, 1962.

UNIVERSITY OF SOUTHAMPTON 5.4.33

Southampton, England
K. R. McLACHLAN *and* J. P. CLEAVE

Work on the development of an analog-digital machine capable of probabilistic, iterative adaption in control problems continues, with no change reported since the previous statement [*Ed.*]. Some financial support has been received from the Department of Scientific and Industrial Research.

UNIVERSITY OF TEXAS 5.4.34

Department of Psychology, Austin 12, Tex.
ROBERT K. LINDSAY

Research into the simulation of human language behavior continues, with no change reported since the previous statement [*Ed.*].

(INDEPENDENT WORK) 5.4.35

Department of Astronomy, University of Glasgow, Scotland
ARCHIE E. ROY

Work on a method of storing information which has properties resembling many of the basic phenomena associated with learning by living organisms continues, with no change reported since the previous statement [*Ed.*].

Reference:

(1) Roy, Archie E. "On a Method of Storing Information. II. A Further Study of Model Properties," *Bulletin of Mathematical Biophysics,* vol. 24, March 1962, pp. 39-69.

5.5 *PSYCHOLOGICAL STUDIES*

The psychological studies reported here deal mainly with the nature and development of information processing in human perception, memory, learning, and problem solving, and with certain aspects of human management of information. New knowledge gained in this area will have an influence on the future design and use of information systems and equipment.

One new project is reported in this subsection. In a psychological research project whose central focus is the analysis of verbal behavior in interactive situations, methods have been designed for the automatic extraction and computer processing of a variety of vocal parameters of spontaneous speech at the William Alanson White Institute (5.5.10).

In communications research at Humboldt-Universität zu Berlin (5.3.10) human decoding processes are being studied.

The project of the Medical Research Council (5.5.4) on psychological factors relevant to the training of electronic trouble-shooters has been discontinued.

The work at Harvard University (5.5.3) on computer applications to content analysis, which was previously reported in Section 2, is now included in this subsection.

5.5.1 ADVANCED INFORMATION SYSTEMS, INC.

3002 Midvale Avenue, Los Angeles 34, Calif.
JOHN A. POSTLEY *and* GARY CARLSON

The impact of behavioral factors on the relative effectiveness of an information system is being studied (see 1.1).

Investigation of the development and use of heuristic routines to predict human behavior is continuing. This approach is based on computer programs which have been developed for predicting errors made in listing bank checks. These routines successfully predict errors four times better than chance. The existing programs are being used as a base to analyze and predict errors in filing cards and documents. The investigation will proceed by clearly defining the particular behavior to be predicted. An iterative procedure will then be followed, analyzing the behavior in fine detail to adapt and improve the heuristic predictive routines.

The study of the problem of predicting which customers will next interrogate a large file is continuing. Although prediction in this case is much more difficult than the other two problems, there are indications that high predictive accuracy can be obtained. The source data are the customer files of a large utility company. Accurate prediction of where inquiries will come from should allow greater flexibility in

maintaining the main file, with a small auxiliary file being able to handle most requests for information. It is planned to extend this work later to the general file problem.

Reference:

(1) Carlson, G. *Analysis of Listing Errors Made in a Bank Central Office.* Doctor's Thesis. Los Angeles: University of California, June 1962. (Unpublished)

HARVARD UNIVERSITY 5.5.2

Center for Cognitive Studies, 61 Kirkland Street, Cambridge 38, Mass.
JEROME S. BRUNER *and* GEORGE MILLER

Research on facilitation and interference in human information processing continues, with no change reported since the previous statement [*Ed.*]. The work is supported by the U. S. Public Health Service and the National Science Foundation.

HARVARD UNIVERSITY 5.5.3

Laboratory of Social Relations, Cambridge 38, Mass.
PHILIP J. STONE, R. FREED BALES, J. ZVI NAMENWIRTH, DANIEL M. OGILVIE, DEXTER C. DUNPHY, *and* WILLIAM R. McPHERSON

In the spring of 1961, an IBM 7090 computer system known as the General Inquirer was developed for the content analysis of written notes and transcripts of taped discussions; data were collected in the process of studying small group interaction. Since then, the system has proved to have useful applications in a wide variety of research problems involving content analysis procedures.

A manual describing the system has been published (1). Written text enters the machine and is separated by the computer into words and sentences. The computer then removes regular word endings, looks each word up in a dictionary, and tags words with labels, indicating each word's membership in one or more categories specified by the investigator. The sentences, together with their tag labels, are then stored on binary tape for repeated use in inquiry procedures. If the keypunched text is also marked with a simplified form of syntactic preediting, the user not only can make inquiries about the co-occurrence of certain word categories but also can specify the syntactic relationship between them.

In addition to being a mechanical aid, the General Inquirer helps make content analysis operations highly explicit and repeatable. The dictionary represents an explication of the investigator's theory or frame of reference. The questions represent the "rules" he has de-

veloped with that theory for discriminating one kind of text from another. Both "rules" and "frame of reference" can be revised repeatedly without necessitating any recoding of the original data.

During the past year, considerable effort has gone into improving the original 3,500-word psycho-sociological dictionary and trying it out on a variety of research problems in the social sciences. The first dictionary (1) was a loose, intuitively written compendium, drawing freely on many theories. The new dictionary (2), a much more conservative, tightly argued document, represents the work of many more people and incorporates much of the experience gained from the first dictionary retrievals. While the old dictionary was quite successful in locating discriminators that were both statistically significant and theoretically meaningful, the new dictionary should represent a further improvement. The new dictionary has already been applied to two projects (3) (4) with considerable success. Several other projects, including a rerun of a project (5) done on the first dictionary, are also about to be processed.

In addition to the development of the psycho-sociological dictionary, several other dictionaries have been or are being developed. A 2,800-word dictionary has been developed by Dr. B. N. Colby (6) for the analysis of anthropological folk materials. A second is being planned at Stanford University for the analysis of tension levels in political documents (7) (8). Both of these projects are concerned with vocabularies and theoretical variables that are quite different from those handled by the psycho-sociological dictionary. The General Inquirer has no limits to either dictionary size or number of permitted variables.

While the General Inquirer is being applied to a wide variety of problems, its original use for studying small group material is still being continued. Over 20,000 cards of data have been coded and punched and are now being subjected to an extensive between-individual and across-time analysis. In addition, General Inquirer tag count printouts are being used this year as feedback to continuing discussion groups, thus supplying them with a continuing analysis of their written comments about the group.

Another new development has been the marriage of the General Inquirer with the Hunt-Hovland concept formation programs (9) (10) in order to create a computer procedure for automatic theme analysis (11). Each text sentence is described by the General Inquirer in terms of its syntactic specific tag labels. The Hunt-Hovland "Concept Learner" then builds a tree for finding a near minimum set of questions that will differentiate the sentences in document A from the sentences in document B.

While the automatic theme analysis system now operates and is being used, further improvements are planned for the coming year.

One problem is caused by the fact that the General Inquirer is programmed in COMIT (12) while the Concept Learner is programmed in IPL-V (13). This creates "interlingual" problems that tend to consume a large proportion of the total computer time. An investigation has revealed that the problem is not insurmountable inasmuch as the Concept Learner theoretically can be reprogrammed rather easily in COMIT.

The development and testing of the General Inquirer were done under a grant from the National Institutes of Health. The analysis of small group interaction is being continued under this grant. Other applications are under the direction of the projects using the results. Sources of financial support for these projects include the Social Science Research Council, the Ford Foundation, the Peace Corps, and the National Institutes of Health. The merger with the Concept Learner was made under the support of the U. S. Office of Naval Research. The U. S. Navy (China Lake Division) also sponsored a conference in the summer of 1962 at the Center for Advanced Behavioral Studies, where the possible application of the General Inquirer to research problems in the political sciences was considered in detail. IBM 7090 computer debugging and testing have been done with the cooperation of the computation centers at Massachusetts Institute of Technology, Harvard University, University of California at Los Angeles, and Stanford University.

References:

(1) Stone, P. J., R. F. Bales, J. Zvi Namenwirth, and D. M. Ogilvie. "The General Inquirer: A Computer System for Content Analysis and Retrieval Based on the Sentence as a Unit of Information," *Behavioral Science,* vol. 7, no. 4, 1962.

(2) Dunphy, D. C., W. R. McPherson, and J. Zvi Namenwirth. *Working Paper on a Revised Psycho-Sociological Dictionary for the General Inquirer,* Laboratory of Social Relations, Harvard University, Cambridge, Mass., October 1962. (dittoed paper)

(3) Ramallo, L. I. *Working Paper on the Application of the General Inquirer (a Computer System) to the Analysis and Evaluation of Reports from Field Service Volunteers,* Center for Research in Personality, Harvard University, October 1962. (dittoed paper)

(4) Ogilvie, D. M., D. C. Dunphy, C. Smith, P. J. Stone, with E. Shneidman and N. Farberow. *Some Characteristics of Genuine vs. Simulated Suicide Notes as Analyzed by a Computer System Called the General Inquirer,* Laboratory of Social Relations, Harvard University, August 1962. (dittoed paper)

(5) Psathas, G. *An Application of the General Inquirer, a Computer Method for Content Analysis, to Two Psychotherapeutic Interviews,* Department of Sociology, Indiana University, May 1962. (dittoed paper)
(6) Colby, B. N., and S. K. Postal. *Comparison of Themes in Folktales,* Laboratory of Social Relations, Harvard University, October 1962. (dittoed paper)
(7) Quant, W. B. *The Use of the General Inquirer System for the Retrieval and Scaling of Hostile Statements in Diplomatic Documents,* Stanford Studies on International Conflict and Integration, Stanford University, June 1962. (mimeographed paper)
(8) Douglass-Jensen, B. *A System of Content Analysis by Computer for Selected Categories of Action,* Stanford Studies on International Conflict and Integration, Stanford University, July 1962. (dittoed paper)
(9) Hunt, E. B. *The Development of Decision Trees in Concept Learning: Model and Basic Results,* working paper no. 6, Western Management Science Institute, U.C.L.A., April 1962.
(10) Hunt, E. B. *Concept Formation.* New York: John Wiley and Sons, Inc., 1962.
(11) Stone, P. J., and E. B. Hunt. "The General Inquirer Extended: Automatic Theme Analysis Using Tree Building Procedures," presented at the International Federation of Information Processing Societies conference, Munich, August 1962.
(12) Yngve, V. H. *COMIT Programmers' Reference Manual.* Cambridge, Mass.: MIT Press, 1962.
(13) Newell, A., ed. *Information Processing Language-V Manual.* Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1961.

5.5.4 MEDICAL RESEARCH COUNCIL

Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England
H. C. A. DALE

The project on psychological factors relevant to the training of electronics troubleshooters has been discontinued.

5.5.5 MONTANA STATE UNIVERSITY

Department of Speech, Missoula, Mont.
FORREST L. BRISSEY, *Principal Investigator*

A series of investigations concerned with the human management of information is in process. Each of the studies in the series is designed fundamentally to assess the usefulness of a particular set of experi-

mental procedures in controlled investigation of communication. Visual displays are presented to subjects about which they are then required to engage in communication employing ordinary English. Apparatus is so designed that the displays may be reconstructed by the subjects as a series of binary decisions. Communicative effectiveness is assessed in terms of the decision performances of the transmitting and receiving subjects.

The first formal investigation has now been completed. The purpose of this study was to obtain preliminary data with respect to the influence of "situational feedback" on the effectiveness of communication, and to examine differences in decision adequacy under conditions in which the subject was required to complete the display, as opposed to conditions in which the subject was permitted to stop at his own discretion.

Preliminary findings support the following tentative conclusions: (a) With high initial message adequacy the presence of situational feedback contributes relatively little to the decision effectiveness of the receiver, whether he is permitted to stop at his own discretion or required to complete the display; (b) with low-to-moderate initial message adequacy the absence of situational feedback is associated with a marked decline in the receiver's decision effectiveness. This effect is particularly marked under conditions of required display completion; and (c) the apparatus and procedures employed appear to lend themselves to an empirically meaningful assessment of certain factors associated with the human management of information.

The next study in the series will be concerned with assessment of the effects of systematic deletion of information from a message of high initial adequacy under conditions of feedback present and absent.

These investigations are sponsored by the U. S. Air Force Office of Scientific Research.

Reference:

(1) Brissey, F. "The Factor of Relevance in the Serial Reproduction of Information," *The Journal of Communication*, vol. 11, no. 4, December 1961, pp. 211-219.

UNIVERSITY OF CALIFORNIA **5.5.6**

Department of Psychology, Los Angeles 24, Calif.

IRVING MALTZMAN, *Principal Investigator*

An investigation of the conditions and principles governing original problem-solving behavior is underway.

A variety of different experiments have been conducted to investigate influencing performance on problems having a unique solution, where only one uncommon response is correct. The problem materials em-

ployed were taken from Mednick and King's Remote Associates Test (RAT).

An important feature of the test employed in these experiments is that the past associative history of the relevant responses which constitute a solution to the problems is known. Various methods employed to increase originality were effective to the extent that they evoked responses with known associative connections with the solution. Additional experiments with these materials showed that prompting was as effective as the active solution of problems in facilitating subsequent solution of similar problems.

The next phase of the experimental problem will be concerned with the effects of different kinds of instructions on original problem-solving behavior.

Financial support for this research is provided by the U.S. Office of Naval Research. The duration of the project is February 1, 1958, to January 31, 1963.

Reference:

(1) Maltzman, I., S. Simon, and L. Licht. Verbal Conditioning of Common and Uncommon Word Associations, *Psychological Reports,* 1962, no. 10, pp. 363-369.

5.5.7 UNIVERSITY OF CONNECTICUT

Department of Psychology, Storrs, Conn.
W. A. BOUSFIELD, *Principal Investigator*

The project dealing with studies of associative processes involved in the mediation of verbal behavior with special reference to the nature of organizational functions is nearing completion. The research has centered on studies of verbal generalization and associative clustering. Several experimental methodologies have been developed for these studies, and the findings have contributed to an extension of S-R (Stimulus-Response) theory which attributes mediation to partial response identities.

A final report will be submitted to the project sponsor, the U. S. Office of Naval Research, in early 1963.

5.5.8 UNIVERSITY OF MARYLAND

School of Medicine, Baltimore 1, Md.
MURRAY GLANZER, *Principal Investigator*

The aim of the project is to analyze the information-processing aspects of problem solving and to determine the mechanisms used for such information processing. To meet this aim, the solving of concept problems has been subjected to detailed logical and experimental

analysis. Ancillary work has also been carried forward on the storage of information, with experiments on the learning and retention of verbal material.

A series of experiments was carried out that isolated two functions in the solution of concept problems: dimension selection and storage of dimension values. Recent experiments have demonstrated that the load on these two functions could be varied independently.

In the series of experiments designed to isolate these functions, the following variables have also been explored: the number of dimensions required to define the concept, use of negative or positive instances, order of the instances, addition of redundant information, and rate at which irrelevant dimensions or hypotheses are eliminated by the series of instances. Recent experiments have also demonstrated that manipulation of storage and selection load induces systematic, predictable changes in EEG responses.

The project is sponsored by the U. S. Army, Office of the Surgeon General.

References:

(1) Glanzer, M. "Grammatical Category: A Rote Learning and Word Association Analysis," *Journal of Verbal Learning and Verbal Behavior,* vol. 1, no. 1, July 1962, pp. 31-41.

(2) Glanzer, M., and Stanley C. Peters. "Re-Examination of the Serial Position Effect," to be published in *Journal of Experimental Psychology.*

(3) Glanzer, M. "Toward a Psychology of Language Structure," to be published in *Journal of Speech and Hearing Research.*

(4) Glanzer, M., J. Huttenlocher, and W. H. Clark. *Systematic Operations in Solving Concept Problems: A Parametric Study of a Class of Problems,* Office of The Surgeon General Technical Report, August 1962.

THE UNIVERSITY OF MICHIGAN 5.5.9

Mental Health Research Institute, Ann Arbor, Mich.
JAMES G. MILLER, *Director, and* ANATOL RAPOPORT, *Principal Investigator*

Studies of learning continue in experiments with model languages, with no change reported since the previous statement [*Ed.*]. The work is supported by the National Science Foundation.

References:

(1) Rapoport, Anatol, and A. Horowitz. "The Sapir-Whorf-Korzybski Hypothesis," *ETC.,* vol. 17, 1960, pp. 346-363.

(2) Livant, William P. "Productive Grammatical Operations. I: The Noun Compounding of Five-Year-Olds," to be published in *Language Learning.*

(3) Livant, William P. "On the Distribution of Adverbial Multipliers," in *Perception and Motor Skills.* (In press)

5.5.10 WILLIAM ALANSON WHITE INSTITUTE

Research Department, 12 East 86th Street, New York 28, N. Y.
JOSEPH JAFFE *and* STANLEY FELDSTEIN, *Principal Investigators*

The central focus of a psychological research project is the analysis of verbal behavior in interactive situations, both experimental and naturalistic. At present, the research is concerned with delineating lexical and nonlexical correlates of psychological states. Such investigations pose major problems of data acquisition and reduction. Methods have been designed for the automatic extraction and computer processing of a variety of vocal parameters of spontaneous speech. In addition, automated content analysis of transcribed lexical data is being pursued.

AVTA (Automatic Vocal Transaction Analyzer), an instrument which automatically monitors the continuous speech of two interacting participants, has been developed. The device extracts and records complex temporal speech patterns directly onto punchcards without human intervention. An IBM 7090 computer program then extracts certain nonlexical dimensions from the AVTA data in terms of cumulative curves and histograms of their occurrence per specified time units. Further development is directed toward extending the use of AVTA to small groups of up to eight participants. In addition, the feasibility of higher speed sampling by substitution of paper tape for the automatic keypunch is being explored.

Possible implications of the research include improvement of lexical (text) retrieval operations by addition of concomitant vocal codings which are present in speech although only implied in written documents.

Instrumentation and programming phases are essentially concluded, and application to the psycholinguistic study of verbal interaction is in progress.

The project is supported by the National Institute of Mental Health.

References:

(1) Cassotta, L., S. Feldstein, and J. Jaffe. "A Device for Automatic Extraction and Quantification of Vocal Behavior in Interviews," presented at Annual Meeting, Eastern Psychological Association, Atlantic City, April 1962.

(2) Feldstein, S., and J. Jaffe. "An IBM 650 Program for the Computation of Speech Disturbances Per Time, Speaker and Group," to be published in *Behavioral Science.*

(3) Feldstein, S., J. Jaffe, and L. Cassotta. "Psychodiagnostic Implications of Automated Speech Analysis," in *Data Acquisition and Processing in Biology and Medicine* (proceedings of Rochester Conference on Data Acquisition and Processing in Biology and Medicine, 1961), ed. by Kurt Enslein. New York: The Macmillan Co., 1962.

(4) Jaffe, J. "Electronic Computers in Psychoanalytic Research," in *Science and Psychoanalysis,* vol. VI, ed. by J. H. Masserman. New York: Grune and Stratton, Inc. (In press)

(5) Jaffe, J., S. Feldstein, and L. Cassotta. *An IBM 7090 Program for Analyzing Vocal Parameters of Dyadic Interaction,* William Alanson White Institute, 1962. (Unpublished manuscript)

Glossary

This glossary is not intended to be inclusive for the field of scientific documentation but rather is intended to serve as an aid to understanding some of the specialized terminology used in project descriptions in this issue of the report series. In all instances the definitions are presented essentially as provided by the project investigators. Each definition is followed by section-article numbers which denote the descriptive statement(s) in which the term is used.

CATEGORICAL RULES—Rules of absolute validity which can be used to reject an unfeasible solution. (3.37)

COMBINATION CODES—Codes in which the code symbols are some combination of bits, in contrast to direct codes where only one dedicated bit is used. Generic term for exclusive, superimposable, and free codes. Combination codes usually make more efficient use of storage media than direct codes. (2.108)

CORRELATION—A correlation is constituted by three elements following in succession according to a certain rhythm, as in a counterpoint of two notes against one. The *correlator* lasts twice as long as each of the two *correlata* which succeed one another during the presence of the correlator. An example of correlation is "John and James," where "John" and "James" are the correlata and "and" is the correlator. (3.36)

CORRELATIONAL NET—The name given to a correlation in which at least one of the components is itself a correlation. (3.36)

DATA PROCESSING CHAIN—A significant order of work in data processing, which makes possible a frictionless transfer between manual, semimechanized, mechanized, and electronic methods. The prerequisite for this is an analysis of the total process including the individual parts of the work process. The purpose of the development of data processing chains is to find similar part processes and combine these with bigger connected processes in an effort to unify methods and as a result make possible a greater degree of mechanization. (2.22)

DENSITY OF 1-BITS—The fraction: $\dfrac{\text{1-bits}}{\text{1 bits} + \text{0-bits}}$ (2.108)

DIFFERENTIATION—A change of state whose results are the *differentiata,* which are taken as the indivisible elements in the analysis of mental activity. They can be described in terms of the change in the organ which produces them, or by indicating the opposite differentiatum, or by indicating the conditions under which we carry out this particular change of state. Examples of differentiata are "light," "dark," "hot," and "cold." (5.4.26)

EXCLUSIVE CODES—Codes in which groups of code symbols mutually exclude each other and in which a field is reserved for one code symbol out of every group. (Cherenin uses the term "local codes"; others use "fixed-field codes." However, since the fields in direct and in superimposable codes are also fixed in their location or position, the term "exclusive codes" was introduced by Körner to avoid misunderstandings.) *See also* Position Codes. (2.108)

FEATURE WORD—The binary word which results from the processing of a pattern. Each position of the feature word designates a test on the pattern, and a "1" is used if the test is satisfied, a "0" otherwise. (5.1.20)

FIGURATION—A change of place; its results are *figures* or *forms.* The dimensional aspect of these figures, however, is obtained from mental categorization. (5.4.26)

FREE CODES—Codes in which the code symbols are not restricted as to the position of a coding field they have to occupy (in contrast to position codes). Also called free-field codes (Cherenin: "non-local codes"). When free codes are used, it is not known in which positions of the code medium to search. (2.108)

HEURISTIC RULES—Hints used to scan a set of hypotheses in as efficient a sequence as possible. (3.37)

INFORMATION INTERVIEW—Discussion between a representative of an Information or a Documentation Service and a subscriber or inquirer. The purpose of the discussion is to obtain as expeditiously as possible information on a particular subject in a manner that best meets the needs of the inquirer, at the same time keeping the cost of the search to the minimum. The information covers not only the prime object of the query but also pertinent details or variations of the subject. The information extends equally the degree of pertinence to the inquirer's viewpoint. The discussion can be carried on through personal contact, correspondence, or telephonic communication. It is one of the principal means to acquire, for the record, reference data on the inquirer in relation to his particular fields of interest, certain facts concerning him (considered

pertinent by the Service) as well as the various questions answered for him. [Translated from the French—*Ed.*] (2.118)

INTERMEDIARY LANGUAGE (IL)—An artificial language whose words are semoglyphs, whose morphology is expressed by 22 standard units (formoglyphs), and whose syntax is reduced to pair configurations with a governing node and a governed one (the tree number of the governing node being a tectoglyph of the governed). (2.72, 3.24)

KOINOGLYPH—A configuration of semoglyphs, corresponding to a term in the given special branch of terminology. (2.72)

LEX—A representation of a lexeme. (3.38)

LEXEME—Within the structure of a single language: a morphemic representation of a sememe. In a machine translation system: a minimal unit of the morphemic stratum which is capable of occurring in new combinations. (3.38)

MECHANOLINGUISTICS—The interdisciplinary field which links computer science with linguistics and whose chief concern is the mechanization of linguistic processes. (3.38, 5.3.30)

MENTAL CATEGORIZATION—The temporal combination of differentiata of presence (or attention or consciousness). These differentiata are those produced by a person when someone says to him "look!" or "listen!" or "hey!" The simplest mental category is that which is obtained by maintaining one of these differentiata while another one succeeds it. This gives the category indicated by the word "etwas" in German, or "qualcosa" in Italian, or (less exactly) by "something" in English. Examples of other mental categories are "and," "or," "cause," "effect," "but," "not," "singular," "plural," "space," and "time." (5.4.26)

MICRO-ABSTRACTS—Concise abstract of article, giving important information regarding content of article which is not provided in the title. (2.116)

PHOTOCHROMICS—Chemical compounds which exhibit reversible spectral absorption effects (i.e., color changes) resulting from exposure to radiant energy in the visible, or near visible, portions of the electromagnetic spectrum. One class of photochromic compounds consists of light-sensitive organic dyes. These photochromic dyes are used by the National Cash Register Company in their PCMI process to produce very high-density microimages. (4.20)

POSITIONS ABSTRACT—An abstract which elucidates and reproduces the contents of a document in a fixed, organized logical sequence of aspects and makes it possible to determine and prepare by machine

the individual positions by a suitable signaling of the individual positions. (2.22)

POSITION CODES—Codes in which the code symbols have to occupy certain positions in a coding field (in contrast to free codes). Generic term for direct, exclusive, and superimposable codes. When position codes are used, it is always known in which positions of the storage medium to search. (2.108)

PROFILING—Finding the limits of clauses and phrases. (3.28)

QUERY TRANSLATOR—Part of the system which transforms a user's description of data desired into a characterization suitable for driving an automatic search procedure. (2.24)

RETRIEVAL ABSTRACT—A short abstract which characterizes the contents of a document only with such information as is necessary to be able to locate it again in the storage. (2.22)

RETRIEVAL LANGUAGE (RL)—An artificial language, whose words are numbers of classified notions, and whose syntax reflects semantic relations between the notions. (2.72)

REVERSE-ALPHABETIZED LIST—A vertical sequence of words printed with the letters in normal order flush to the right, but alphabetized from the right rather than the left; e.g., *fuzz* would be among the last entries, *zebra* among the first. (5.3.36)

SEMEME—The fundamental structural element of meaning, arrived at by applying structural linguistic analysis in the area of semantics; sememes exist on the sememic stratum, and are represented by morphemes, combinations of morphemes, and features of arrangement of morphemes and combination of morphemes. *See* Semology.

SEMEMICS—*See* Semology.

SEMOGLYPH—A five-digit number of the ring of translational substitutions (00001=u=and=et=und=y=e/ed=). (2.72, 3.24)

SEMOLOGY—Study of semantic structure; includes the two branches: sememics, which deals with the morphemic representation of sememes, and semotactics, which is concerned with the arrangement of sememes. (3.38)

SEQUENTIAL LANGUAGES—A subclass of simple (context-free) phrase structure languages of particularly transparent construction. (5.3.9)

STRATIFICATIONAL THEORY OF LANGUAGE—A linguistic theory which views language as made of a series of representational levels or strata, each of which is related to its neighboring strata by a complex code consisting of a set of representational rules; any natural language has three primary strata: the sememic, or the stratum of

content; the morphemic, in the middle; and the phonemic (for spoken languages) or graphemic (for written languages). (3.38)

STRUCTURAL ABSTRACT—An abstract which represents predominantly the results of experimental work in sciences with quantitative data in an optically clear manner; it cannot, however, be used by machine. (2.22)

TACTIC ANALYSIS—Analysis of arrangements of linguistic items. (3.38)

THOUGHT—The process of opening and closing correlations. (5.4.26)

TOPOLOGICAL MEMORY UNITS—A group of numbers or symbols for representing a structure made up of a number of building blocks joined together in a specific way (for example, the atoms of a chemical compound), without regard for their (incidental) topographic position, which is, moreover, to a great extent selected arbitrarily in the drawing of the structural configuration. The building blocks are numbered in arbitrary sequence and for each an indication is given of what others it is directly connected with. Each structure can be represented by a large number of topological memory units, which are externally different but equivalent in content; each topological memory unit, however, corresponds to only one definite structure. [Translated from the German—*Ed.*] (2.10)

TRANSFORM—Transform of an image is the set of binary numbers generated in the machine as a result of the input image. (5.1.7)

UNKNOWN WORD—An "unknown" word is an as-yet-unclassified member of a finite vocabulary. (5.2.17)

List of Acronyms and Abbreviations

ACSI-MATIC—A large-scale intelligence information processing system for the Assistant Chief of Staff, Intelligence (U.S. Army)

AEC—Atomic Energy Commission

AFOSR—Air Force Office of Scientific Research

AGILE—Autonetics General Information Learning Equipment

ALDP—Automatic Language-Data Processing

ALGOL—Algorithmic Language

AMNIP—Adaptive Man-machine Non-arithmetic Information Processing

APT—Automatically Programmed Tool

ARAL—Automatic Record Analysis Language

ASLIB—Association of Special Libraries and Information Bureaux

ASTIA—Armed Services Technical Information Agency

AUTOTRAN—Automatic Translation System

AVTA—Automatic Vocal Transaction Analyzer

B.A.S.I.C.—Biological Abstracts' Subjects In Context

BSCP—Biological Sciences Communication Project

CADA—Centre d'Analyse Documentaire pour l'Archéologie [Center for Document Analysis in Archeology]

CASPAR—Cambridge Analog Simulator for Predicting Atomic Reactions

CBAC—*Chemical-Biological Activities*

CEA—Commissariat à l'Energie Atomique [Atomic Energy Commission], France

CETAG—Centre d'Études pour la Traduction Automatique, Section de Grenoble [Center for the Study of Automatic Translation, Grenoble Section]

CETAP—Centre d'Études pour la Traduction Automatique, Section de Paris [Center for the Study of Automatic Translation, Paris Section]

CETIS–Centre de Traitement de l'Information Scientifique [Center for Scientific Information Processing]

CHEMTRAN–designation for a complete grammar of organic chemical nomenclature

CID–Centre d'Information et de Documentation [Center for Information and Documentation], Belgium

COBOL–Common Business Oriented Language

COMIT–A user-oriented general-purpose symbol-manipulation programming language

CRIS–Command Retrieval Information System

DICADAP–Dictionary Adapter

DOCA–Section de Documentation Automatique [Automatic Documentation Section]

EPAM–Elementary Perceiving and Memorizing Machine

EURATOM–Communauté Européenne de l'Energie Atomique [European Community for Atomic Energy]

FAP–Fortran Assembly Program

FDA–Food and Drug Administration

FLEX–An artificial language of minimum syntactic organization and maximum semantic organization

FORTRAN–Formula Translator

FROLIC–Formal, Retrieval-Oriented Language for Indexing Content

GAT–Georgetown Automatic Translation

HAYSTAQ–Have You Stored Answers to Questions? (a computer search program)

ICIREPAT–International Cooperation in Information Retrieval among Examining Patent Offices

IDEP–Interservice Data Exchange Program

IFG–Information File Generator

ILAS–Interrelated Logic Accumulating Scanner

IPL–Information Processing Language

JEIPAC–JICST Electronic Information Processing Automatic Computer

JICST–Japan Information Center for Science and Technology

JOVIAL–Jules Own Version of International Algebraic Translator

KVAL—Kvantitativ Lingvistik
[Quantitative Linguistics], actually
Forskningsgruppen för Kvantitativ Lingvistik
[Research Group for Quantitative Linguistics], Sweden

KWIC—Keyword-in-Context

LISP—List Processor (programming language)

LM—Learning matrix

LOGOS—Formal, English-like language

MEDLARS—Medical Literature Analysis and Retrieval System

META—Methods of Extracting Text Automatically

MIRA—*Monthly Index of Russian Accessions*

MNA—Matematikmaskinnämnden—The Swedish Board for Computing Machinery

MULTITRAN—Multiple Translation System (translating one language into several target languages)

NASA—National Aeronautics and Space Administration

NIH—National Institutes of Health

NLM—National Library of Medicine

NSA—*Nuclear Science Abstracts*

OATS—Original Article Tear Sheet

OCR—Optical Character Recognition

ONR—Office of Naval Research

OTS—Office of Technical Services

PACIR—*See* SECIR

PCMI—Photochromic Microimage

PERIOR—Pour l'Enregistrement et la Recherche de l'Information Organisee
[For the recording and search of organized information]

PERT—Program Evaluation Review Technique

RADC—Rome Air Development Center

RADIR—Random Access Document Indexing and Retrieval

RAT—Remote Associates Test

RICASIP—Research Information Center and Advisory Service on Information Processing

RUSDIC—Russian Dictionary in MT

RUSTAN—Russian Text Analyzer

SABIRS—Semi-Automatic Bibliographic Information Retrieval System

SACE—Semi-Automatic Chain Encoding Device

SAD—Section d'Automatique Documentaire
[Section for Automatic Documentation]

SDI—Selective Dissemination of Information

SECIR—Semi-Automatic Encoding of Chemistry for Information Retrieval (formerly identified as PACIR)

SEMA—Société d'Économie et de Mathématique Appliquées

SEMANTRAN—Semantic Transfer

SESD—Service d'Études Semiologiques et Documentaires
[Center for Semiological and Documentation Studies], France

SHARP—Ships Analysis and Retrieval Project

SIL—Summer Institute of Linguistics

SITE—Search Information Tape Equipment

SYNTOL—Syntagmatic Organization Language

TAB—*Technical Abstract Bulletin*

TABLEDEX—A coordinate index displayed in tabular form

TRAC—"text reckoning and compiling" system

UDC—Universal Decimal Classification

UNIDEK—A mechanized application of the UDC

Index of Individuals and Organizations

The statements have been assigned consecutive numbers within the various sections. The "Section-Article" column below denotes the numbers of the section and article from which the indexed entry was taken. The letter *r* indicates an author in the list of references following statements.

Research projects being conducted outside the United States are indexed in two places: by title of organization and by country.

	Section-Article
Abdian, Gregory	1.23
Abe, Koji	2.67*r*
Abercombie, David	5.2.26
Abraham, C. T.	2.62*r*
Abraham, S.	5.3.41
Académie de la République Populaire Roumaine	5.1
Advanced Information Systems, Inc. (Hayes)	2.1
Advanced Information Systems, Inc. (Postley)	1.1, 5.5.1
Aeronutronic	5.4.1
Air Force Cambridge Research Laboratories (Petrick)	5.2.1, 5.4.2
Air Force Cambridge Research Laboratories (Wathen-Dunn)	5.2.2
Alexander, S. N.	2.82
Allard, M., *S. J.*	4.2*r*
Allen, G.	5.4.31*r*
Altman, James W.	2.4
Alt, Franz L.	3.28*r*
Alvarez, Manuel	3.33*r*
American Bar Foundation	2.2
American Bible Society	5.3.1
American Brake Shoe Co.	4.1
American Chemical Society	1.2
American Diabetes Association, Inc.	2.3, 2.128
American Institute of Biological Sciences	1.3, 1.57
American Institute of Chemical Engineers	1.4, 1.23, 2.93
American Institute of Physics	1.5, 2.131
American Institute for Research	2.4
American Legal Data Processing Association	2.5
American Mathematical Society	1.6, 3.46
American Medical Association	2.6
American Meteorological Society	1.7
American Museum of Natural History	1.8
American Psychological Association	1.9, 1.53
American Society for Metals Documentation Service	2.128
American Society of Tool and Manufacturing Engineers	1.52
American University	1.3

Section-Article

Anderson, Sydney 1.8
Andreev, N. D. 2.72, 3.24, 5.3.14
Angel, N. L. 1.41r
Applications Research Corp. 2.33
Armed Services Technical Information Agency 2.7, 2.8, 2.27, 2.99
Armour Research Foundation (Cameron) 5.4.3
Armour Research Foundation (McCully) 2.8
Arndt, Joseph H., Jr. 4.14r
Arnold, Jane B. 5.2.15r
Arsenault, William R. 4.3
Artandi, Susan 2.99
Arthur D. Little, Inc. 1.10, 5.1.1, 5.3.2, 5.3.6
Ashby, W. R. 5.4.30r
Ashby-Pickford, A. G. 1.48r
Ashworth, E. Robert 5.3.28
Aslib. *See* Association of Special Libraries and Information Bureaux.
Association of Special Libraries and Information Bureaux 2.81
Association of Special Libraries and Information Bureaux (Cleverdon) 2.9, 2.46, 2.83, 2.128
Association of Special Libraries and Information Bureaux (Phillips) 1.11
Association of Special Libraries and Information Bureaux (Slater) 1.12
Association of Special Libraries and Information Bureaux (Whyte) 1.13
ASTIA. *See* Armed Services Technical Information Agency
Astropower, Inc. 5.4.4
Atherton, Pauline 1.5
Auroux, A. 3.5r
Autonetics (Bledsoe) 5.2.3
Autonetics (Overton) 5.4.5
Babbitt, Milton 5.4.19

Babcock, M. L. 5.4.30r
Backus, Oswald P., III 1.54r
Bacon, C. R. T. 2.126r, 5.3.38r
Badische Anilin- & Soda-Fabrik AG. 2.10
Bagg, T. C. 2.84r, 4.18
Bailey, M. J. 1.32r
Bairati, Prof. 5.4.26
Bales, R. Freed 5.5.3
Balestic, Ch.-J. 2.17r
Banerji, R. B. 5.1.6
Bar-Hillel, Yehoshua 5.3.9, 5.3.29
Barker, Dale L. 1.21
Barnes, Norman F. 2.96
Barnes, R. F. 2.65, 2.133r
Barnett, Michael P. 1.32, 5.3.21
Bartels, Wolfgang 2.38
Bartlett, James 5.4.13
Bartlett, Lynn J. 2.53
Bartlett, Wayne D. 2.39
Bartram, P. R. 5.4.24r
Barus, Carl 5.1.40

Section-Article

Bastian, A. Lewis, Jr. 5.4.2
Bastin, E. W. 5.4.6
Batchelor, Harold W. 2.111
Baumann, D. M. 4.14, 5.1.21
Baxendale, Phyllis 2.59
Becher, R. 2.37
Beck, Jacob 5.1.3
Behringer, F. A. 2.22r
Belgium
Université Libre de Bruxelles 3.37
Université Catholique de Louvain 2.117
Belknap, Richard H. 1.2
Bell, G. A. 1.48
Bell Telephone Laboratories, Inc. 5.1.2, 5.2.4, 5.4.31
Beltran, Sergio F. 3.35
Belzer, Jack 2.128r
Benbow, J. M. 2.39
Benkard, J. Philip 5.1.45
Bennett, Edward M. 5.4.13
Bennett, Frank G., IV 2.50r
Bennett, J. 2.59r
Bensing, Robert C. 2.128r
Benz, William 2.5r
Bernick, M. D. 2.106r
Bernstein, H. H. 2.22r
Bertram, S. 5.2.23
Berul, L. H. 1.24r
Bibbero, Robert J. 5.1.4
Biet, F. 2.23r
Biological Abstracts 1.14
Birch, Robert L. 2.100
Birkbeck College (University of London) 3.2
Black, Henry 1.18, 2.26
Blanyer, Carl G. 4.1
Bledsoe, J. D. 5.2.3
Bledsoe, W. W. 5.1.27
Blick, A. R. 1.48r
Bliss, Henry E. 2.122
Block, H. D. 5.4.8r
Blois, J. 3.37r
Blumberg, Donald F. 2.30
Bodin, L. 5.1.26r
Boeing Co. 2.11
Boek, Walter E. 1.25
Bogusz, J. F. 5.1.28
Bohnert, H. G. 2.62r
Bolt Beranek and Newman, Inc. 2.12
Boni, Charles 5.1.25
Bonniwell, Marion E. 2.14
Booth, Andrew D. 3.2
Booth, D. C. 3.2r

Section-Article

Borko, Harold 2.106
Bouillut, Jean 2.118r
Bousfield, W. A. 5.5.7
Bowker, J. Kent 4.11
Braffort, P. 2.21, 2.22, 3.6, 3.37r
Brain, Alfred E. 5.1.38
Brand, C. P. 5.3.31
Brandenberg, W. 2.32r, 2.63r
Braverman, E. M. 5.1.12
Brevatome 2.22
Brick, Donald B. 5.1.41
Brissey, Forrest L. 5.5.5
Brooker, R. A. 5.4.2r
Brooks, Glidden L. 2.13
Brooks, K. 5.1.4r
Bross, John S. 3.27r
Brown, A. F. 5.3.36
Brown, Roger 5.3.22
Brown University (Kingsland) 2.13
Brown University (Kucera) 5.3.3
Brunenkant, Edward J. 1.50
Bruner, Jerome S. 5.5.2
Bryan, J. S. 5.4.16
Buck, C. P. 4.17r
Buckland, L. F. 2.131r
Budd Electronics 5.1.3
Buell, D. N. 5.1.7
Bullock, T. H. 5.4.10
Bulova Research & Development Laboratories 5.2.5, 5.2.17
Bunnow, L. R. 2.32
Buntrock, H. 2.22r
Bureau of the Census 2.43
Bureau of Ships 2.14, 2.28, 5.4.7
Bureau of Social Science Research, Inc. 1.15
Burroughs Control Corp. 5.1.4
Burroughs Corp. 5.1.4, 5.1.5
Busa, Rev. Roberto, *S. J.* 5.3.4
Buscher, William C. 2.128r
Bushor, William E. 4.10r
Buydens, J. 3.37r

Cafagna, Albert G. 5.4.31
Cahoon, Lynda L. 1.35
Cambridge Language Research Unit (Bastin) 5.4.6
Cambridge Language Research Unit (Masterman) 3.3
Cambridge Language Research Unit (Needham) 2.15
Cameron, Scott H. 5.4.3
Campbell, D. T. H. 1.42
Canada
  Pulp and Paper Research Institute of Canada 2.93
Carlson, Gary 5.5.1

Section-Article

Carnegie Institute of Technology 5.4.21, 5.4.27
Case Institute of Technology 5.1.6
Cassotta, L. 5.5.10*r*
Ceccato, Silvio 3.36, 5.4.26
C-E-I-R, Inc. 2.16
Centre d'Analyse Documentaire pour l'Afrique Noire 2.76
Centre d'Etudes Nucléaires de Saclay 2.17
Centre National de la Recherche Scientifique (CADA) 4.2
Centre National de la Recherche Scientifique (SAD) 2.18
Centre National de la Recherche Scientifique (Sestier) 3.4
Centre National de la Recherche Scientifique (Vauquois) 3.5
Centro per l'Automazione dell'Analisi Letteraria 5.3.4
Ceskoslovenská Akademie Ved 5.3.5
Chalmers Tekniska Hogskola 5.3.35
Chang, Hsien W. 3.18
Chao, Y. R. 5.41
Charles University. *See* Karlova Universita.
Charles W. Adams Associates, Inc. 5.3.6
Charney, Elinor K. 3.27*r*
Chatman, Seymour 2.106*r*, 5.3.37*r*
Chatten, J. B. 5.1.29, 5.4.17
Cheatham, T. E., Jr. 2.24
Chemical Abstracts Service 1.16, 2.19
Cheney, P. W. 5.1.35
Ching, Eugene 3.30*r*
Chlouba, V. 3.44*r*
Chomsky, Noam 3.30, 5.3.22, 5.3.29, 5.3.37*r*, 5.4.15
Chonez, A. 2.17*r*
Chow, C. K. 5.1.5
Christophe, J. 4.2.*r*
Chrysler Corp. 5.1.7
Claridge, P. R. P. 2.46*r*
Clark, W. H. 5.5.8*r*
Clarke, Robert F. 1.39
Clayden, D. O. 5.1.24
Cleave, J. P. 5.4.33
Cleverdon, Cyril 2.9, 2.128*r*
Climenson, W. D. 2.94*r*, 5.3.25
Clowes, M. B. 5.1.24
Cognitronics Corp. 5.2.6
Colby, B. N. 5.5.3
Coleman, R. P. 5.1.5
Colilla, R. A. 2.94*r*
College of Aeronautics 2.9
Collins, G. O., Jr. 2.24*r*
Colombia
  Universidad Industrial de Santander 2.116
Columbia University (Harris) 5.2.7
Columbia University (Lazarsfeld) 1.17
Commission on Professional and Hospital Activities 2.20

Section-Article

Communauté Européenne de l'Energie Atomique (Euratom) (Braffort) 2.21, 3.6, 3.11, 3.13, 3.37
Communauté Européenne de l'Energie Atomique (Euratom) (Meyer-Uhlenried) 2.22
Communicable Disease Center 2.128
Compagnie de Saint-Gobain 2.23
Computer Associates, Inc. 2.24
Computer Concepts, Inc. 3.7
Computer Usage Co. 3.8
Computing Center, Academy of Sciences Ukrainian S.S.R.
See Institute of Cybernetics, Academy of Sciences Ukrainian S.S.R.
Connecticut Agricultural Experimental Station 1.36
Conrad, G. Miles 1.14
Conseil International des Sciences Sociales 5.3.13
Conze, K. 2.22r
Cooper, Paul W. 5.1.22
Corbe, Michael 3.48
Cornell Aeronautical Laboratory, Inc. (Leland) 5.2.8, 5.4.7
Cornell Aeronautical Laboratory, Inc. (Roetling) 2.25
Cornell University 5.4.8
Courtney, Robert H., Jr. 2.60
Coyaud, M. 2.18r
Craft, J. L. 3.18r
Creager, William A. 1.23
Creedmoor Institute for Psychobiologic Studies (Black) 2.26
Creedmoor Institute for Psychobiologic Studies (Whittier) 1.18
Cremer, M. 2.22
Crockett, Robert G. 4.21
Cros, R.-C. 2.76r
Crosland, Dorothy M. 1.22
Crouse, Eleanor 1.25r
Culbertson, Don S. 2.121
Čulik, Karel 2.70r
Czechoslovak Academy of Sciences. *See* Československá Akademie Věd.
Czechoslovakia
  Československá Akademie Věd 5.3.5
  Karlova Universita 2.70, 3.20, 3.21, 5.3.17
  Výzkumný Ústav Matematických Strojů 3.44

Dale, H. C. A. 5.5.4
Daly, J. 5.4.4r
Dameniya, M. F. 3.15r
Darlington, Jared 3.27r
Dasher, B. J. 3.12
Datatrol Corp. 2.27
Davenport, John Scott 1.38, 3.32
David Taylor Model Basin 2.14, 2.28
Davis, David F. 1.57
Davison, P. S. 2.101
Day, A. M. 3.29
DeBacker, M. 4.23r

Section-Article

DeFlorio, George P. 5.4.24
deGrolier, Eric 5.3.13
Dekker, J. 2.88
Denes, P. 5.2.25r
Dennis, J. B. 5.2.15r
Dennis, Sally F. 2.2
Department of Defense 2.7, 2.29
Deribere-Desgardes, M. L. 2.23
Dersch, W. C. 5.2.11r
de Silva, Clare 2.116
Detant, M. 2.21r
Dickinson, W. E. 5.2.11
Diebold Group, Inc. 1.31, 2.30
Documentation Inc. 2.31
Documentation Research and Training Centre 2.75
Domonkos, Erika 3.1
Donaghue, Hugh P. 2.27
Dopkowski, Philip L. 1.3r
Dopping, Olle 5.3.23
Dorodnitsyn, A. A. 5.4.9r
Doshita, Shuji 5.2.12r
Dostert, Léon 3.11
Dougherty, Ching-Yi 3.38r
Dougherty, Richard M. 2.99
Douglas Aircraft Co., Inc. (*See also* Astropower, Inc.) 2.32
Douglass-Jensen, B. 5.5.3r
Doyle, L. B. 2.106r
Doyle, Patrick E. 2.82
Driese, E. C. 5.1.37
Dubester, Henry J. 1.31, 2.74
Dubinsky, E. 3.39r
Duckworth, A. P. 5.3.32r
Duflos, Marie-Thérèse 2.118r
Dunphy, Dexter C. 5.5.3
Dupuis, L. 3.4r
Dyer, P. Swinnerton 5.3.31
Dyson, G. Malcolm 1.16, 2.19

Easter, Robert R. 1.46
Eastern Illinois University 2.122
Ecole Pratique des Hautes Etudes, VIe Section 2.76
Edgerton, Germeshausen & Grier, Inc. 4.15
Edmundson, H. P. 2.109, 3.31r, 3.34
Ehrman, Libert 1.45
Eldridge, William B. 2.2
Electro-Optical Systems, Inc. 2.33
Electrotechnical Laboratory (Iijima) 5.1.8
Electrotechnical Laboratory (Tadenuma) 3.9
Eliashvili, A. I. 3.15
Eliseyev, V. K. 5.1.13r
Ellegard, Alvar 5.3.35

Section-Article

Ellis, J. O. 5.3.34r
Elziere, M. 4.2r
Engelbart, Douglas C. 1.43
Engineers Joint Council 2.34
England
Association of Special Libraries and Information Bureaux 1.11, 1.12, 1.13, 2.9
Birkbeck College (University of London) 3.2
Cambridge Language Research Unit 2.15, 3.3, 5.4.6
Hatfield College of Technology 2.46
Library Association 2.73
Manchester University 5.3.20
Medical Research Council 5.5.4
National Book League 2.81
National Physical Laboratory 2.87, 3.29, 5.1.24
"Shell" Research Ltd. 1.41
United Kingdom Atomic Energy Authority 1.48, 1.49
University of Cambridge 4.22, 5.3.31
University College London 5.2.25, 5.3.32, 5.3.33
University of Southampton 5.4.33
Independent Work (Farradane) 2.129
Esso Research & Engineering Co. (Jahoda) 2.35
Esso Research & Engineering Co. (Rupp) 2.36
Euratom. *See* Communauté Européenne de l'Energie Atomique
Euwe, Max 5.4.23
Evreinov, E. V. 5.3.15

F. Hoffman-La Roche & Co. Ltd. 2.37
Fairthorne, Robert A. 2.128r
Fallon, H. C. 2.32r
Fant, Gunnar 5.2.21
Farbenfabriken Bayer AG 2.38
Farberow, N. 5.5.3r
Farley, Earl 1.54, 2.123
Farradane, J. 2.129
Farrington Electronics, Inc. 5.1.9
Fasana, P. J. 2.131r
Faverge, Jm. 3.37
Fazar, Willard 1.23
Federal Republic of Germany
Badische Anilin- & Soda-Fabrik Co. 2.10
Farbenfabriken Bayer AG 2.38
Gmelin Institute Documentation Center 2.44
Physikalisches Institut der Technischen Hochschule Aachen 5.4.18
Technische Hochschule Karlsruhe 5.1.42, 5.4.25
Technische Hochschule München 2.108
Universität Bonn 5.2.24, 5.3.29
Independent Work (Wenck) 5.3.43
See also Germany.
Federal Scientific Corp. 5.2.9
Fehder, P. 5.3.21r

Section-Article

Feigenbaum, Edward A. 5.4.21r, 5.4.27
Feldman, Alfred 4.23
Feldman, Julian 5.4.27
Feldstein, Stanley 5.5.10
Ferguson, Malcolm 2.128r
Fergusson, E. S. 2.97
FID. *See* International Federation for Documentation.
Fillmore, Charles J. 3.30r, 5.2.27r
Finland
  Kasvatusopillinen Korkeakoulu 3.22
Firschein, O. 5.1.20
First Research Center, Defense Agency 3.10
Fischler, M. 5.1.20
Flood, Merrill M. 5.4.31
Floyd, William 5.2.13
FMA, Inc. 4.3
Focht, L. R. 5.2.18
Food and Drug Administration 2.82
Forbes & Waite, Inc. *See* Information Dynamics Corp.
Ford, J. D., Jr. 2.106r
Ford Motor Co. *See* Aeronutronic.
Forgie, Carma D. 5.2.14
Forgie, James W. 5.2.14
Forsen, G. E. 5.1.38r
Forskningsgruppen för Kvantitativ Lingvistik (KVAL) 5.3.7
Foskett, D. J. 2.73
Fossum, Earl G. 2.115r
Fournier, M. 2.23
Fowler, Murray 5.3.39
Fox, A. 2.42r
France
  Centre d'Etudes Nucléaires de Saclay 2.17
  Centre National de la Recherche Scientifique 2.18, 3.4, 3.5, 4.2
  Compagnie de Saint-Gobain 2.23
  Ecole Pratique des Hautes Etudes, VIe Section 2.76
  Institut National des Techniques de la Documentation 5.3.13
  IBM-France 2.23, 2.64, 3.19, 5.3.16
  Laboratoire d'Analyse Lexicologique 5.3.18
  Maison des Sciences de l'Homme 2.76
  Société d'Économie et de Mathématique Appliquées (SEMA) 2.102
  Université de Paris 2.118
  Independent Work (Corbe) 3.48
Frank, A. I. 5.1.30
Frank, H. 5.4.25r
Freeman, Herbert 5.1.26
French Commissariat à l'Energie Atomique (C.E.A.) 2.17
Fry, D. B. 5.2.25
Fucks, Wilhelm 5.3.29, 5.4.18
Fujimura, O. 5.2.15r
Futrelle, R. P. 5.3.21r

Section-Article

Gaifman, C. 5.3.29
Galli, E. J. 3.18r
Galliher, Herbert P. 1.33
Gardelle, Marie-Claude 2.118r
Gardin, J.-C. 2.18, 2.76, 4.2
Gardin, Natacha 2.76r
Gardiner, Russell K. 3.38r
Gardner, L. 5.1.26r
Garfield, Eugene 1.26, 2.58
Garland, J. L. 2.60r
Garvey, William D. 1.9
Garvin, Paul L. 2.109, 2.110, 3.34, 5.2.23
Gaucher, T. A. 2.127r
Geinitz, D. 2.22r
General Dynamics/Electronics (Kanal) 5.1.10
General Dynamics/Electronics (Slaymaker) 5.2.10
General Electric Co. (Bartlett) 2.39
General Electric Co. (Goostree) 4.4, 5.3.8
General Electric Co. (Hubbell) 2.40
General Electric Co. (Spangler) 1.19
General Electric Co. (Thompson) 2.41
General Electric Co. (Wang) 2.42
General Electric Co., Defense Systems Operation 2.86, 4.21
General Electric Research Laboratory (Glenn) 4.5
General Electric Research Laboratory (Morrison) 1.20
General Precision, Inc. *See* Librascope and Link Divisions
George Washington University 2.27
Georgetown University 3.11, 3.37
Georgetown University School of Medicine 2.43
Georgia Institute of Technology (Barker) 1.21
Georgia Institute of Technology (Crosland) 1.22
Georgia Institute of Technology (Dasher) 3.12
German Patent Office 2.114
Germany
  Humboldt-Universität zu Berlin 5.3.10
  See also Federal Republic of Germany.
Gerresheim, E. 5.3.29
Gerstman, Louis J. 5.2.17
Giallanza, Frank 1.51
Gillis, P. P. 3.37
Gillum, T. L. 2.7r
Giuliano, Vincent E. 1.10, 5.1.1, 5.3.2
Gladkiĭ, A. V. 3.17
Glantz, R. S. 2.65
Glanzer, Murray 5.58
Glasersfeld, Ernst v. 3.36r
Glass, J. 5.1.26r
Gleitman, Lila 5.3.37r
Glendinning, J. I. 1.29r
Glenn, W. E. 4.5

Section-Article

Glushkov, V. M. 5.4.9
Gmelin Institute Documentation Center 2.44
Goerke, W. 5.4.25r
Goffman, William 2.128r
Goldberg, J. 2.105r
Goldman, E. H. 3.18r
Goldwyn, Alvin J. 2.128r
Gomery, Thomas D. 2.80
Goodman, Nelson 5.3.37r
Goostree, L. W. Jr. 4.4, 5.3.8
Gottsch, Lida G. 2.56
Gould, Sydney W. 1.36
Gray, Harry J. 2.125, 5.4.32
Gray, Stephen B. 5.1.41
Greanias, Evon C. 5.1.14
Green, L. S. 5.2.19r
Greene, Peter H. 5.4.29
Greenly, Robert B. 5.1.18
Grems, M. 2.62r
Griffin, Marjorie 2.61
Griffith, Belver C. 1.9
Griffith, John H. 2.20
Grimes, Joseph E. 3.33, 5.3.27
Grishchenko, N. M. 5.4.9r
Gross, M. 3.4r
Guilloux, R. 2.17
Gullahorn, Jeanne E. 5.4.28
Gullahorn, John T. 5.4.28
Gunther, G. 5.4.30r
Gurk, H. M. 2.94r
Gyr, John W. 5.4.31

Halle, Morris 5.2.15r, 5.3.22
Hallicrafters Co. 4.6
Halliday, M. A. K. 5.3.34
Halligan, Robert F. 4.6
Hamilton, Harold J. 5.4.10
Hammond, William 2.27
Hannan, W. J. 5.1.33
Hanson, C. W. 1.26r
Harary, F. 5.2.27r
Hardwick, N. H. 2.94r, 5.3.25r
Harley, T. J., Jr. 5.4.17
Harlow, Neal 2.99
Harmon, Leon D. 5.1.2, 5.2.4, 5.4.31
Harper, K. E. 3.31r
Harrar, Helen Joanne 1.40
Harrington, R. W. 2.107
Harris, Cyril M. 5.2.7
Harris, Zellig S. 5.3.25, 5.3.37
Hartley, Alice K. 5.1.19

Section-Article

Harvard University 5.1.3
Harvard University (Bruner) 5.5.2
Harvard University (Oettinger) 3.13, 3.29
Harvard University (Salton) 2.45
Harvard University (Stone) 5.5.3
Hatfield College of Technology 2.46
Hattery, Lowell H. 1.3r
Hawkins, J. K. 5.4.1
Hayes, R. M. 2.1
Hays, David G. 3.31, 5.3.26, 5.3.29
Hayward, H. Winston 2.114r
Healy, L. D. 5.1.20r
Heard, Leroy 2.5r
Heart, F. E. 2.77r
Heatwole, Henry T. 4.7
Hebrew University 5.3.9
Hecker, M. H. L. 5.2.15r
Heike, G. 5.2.24
Heiliger, Edward M. 2.121
Heinritz, Fred J. 2.99
Heinz, J. M. 5.2.15r
Hensley, C. B. 2.32r, 2.63r
Herman, Carlton M. 1.57
Herner and Co. 1.23, 1.25, 2.47, 4.7
Herner, Mary 1.23, 2.47
Herner, Saul 1.23, 4.7
Herzberger, Hans G. 5.3.37r
Hessinger, Richard W. 5.1.31
Higley, Henry G. 2.85
Hillman, Donald J. 2.71
Hirschberg, L. S. 3.37
Hiż, Henry 5.3.37r
Hoff, M. E. 4.16r
Hoffman, Allen 5.4.16r
Hoffmann, Bernhard A. 2.37r
Holland, D. B. 4.23r
Holm, B. E. 1.4r
Holmes, W. S. 5.4.7r
Homans, George C. 5.4.28
Hönerloh, H.-J. 5.4.25
Hormann, Aiko 5.4.24
Horn, Ernest 5.3.40
Horowitz, A. 5.5.9r
Horty, John F. 2.126, 5.3.38r
Horvath, W. J. 5.4.31
Houde, R. A. 5.2.10
Hours, F., *S. J.* 4.2r
House, A. S. 5.2.15r
Householder, F. W. 5.3.11
Houston Fearless Corp. 4.8
Hovland, Carl I. 5.5.3

Section-Article

Hsieh, H. S. 5.4.30r
Hubbell, J. R. 2.40
Hughes Aircraft Co. 2.48
Human Relations Area Files, Inc. 2.49
Humboldt-Universität zu Berlin 5.3.10
Hungary
  Institute of Library Science 2.57
  Országos Mezögazdasági Könyvtár És Dokumentációs Központ 1.37
Hunsberger, I. Moyer 2.79
Hunt Barnard & Co. 5.3.35
Hunt, E. B. 5.5.3
Huttenlocher, J. 5.5.8r

Iijima, Taizo 5.1.8
Inadomi, K. 3.23r
Incarbone, Salvatori 5.4.26r
Index and Retrieval Systems Inc. 2.50, 4.9
India
  Documentation Research and Training Centre 2.75
  Library Research Circle 2.75
Indian Statistical Institute 2.75
Indiana University 3.39
Indiana University (Householder) 5.3.11
Indiana University (Sebeok) 2.51, 5.3.12
Information Dynamics Corp. 1.24, 2.52
Information for Industry, Inc. 2.53
Information Retrieval Corp. 4.10
Ing. C. Olivetti & C., S.P.A. 5.1.11
Inomata, S. 5.2.15r
Inselberg, A. 5.4.30r
Institut für Dokumentationswesen 2.22
Institut Za Eksperimentalnu Fonetiku 3.14
Institut International des Brevets 2.114
Institut National des Techniques de la Documentation 5.3.13
Institute for Advancement of Medical Communication (Orr) 1.25
Institute for Advancement of Medical Communication (Schultz) 2.54
Institute for Advancement of Medical Communication (Welt) 2.55
Institute of Automatics and Telemechanics, Academy of Sciences U.S.S.R. 5.1.12
Institute for Bio-Medical Computer Research 2.56
Institute of Cybernetics, Academy of Sciences Ukrainian S.S.R. 5.1.13, 5.4.9
Institute of Electronics, Automatics and Telemechanics, Academy of Sciences Georgian S.S.R. 3.15
Institute for Experimental Phonetics. *See* Institut Za Eksperimentalnu Fonetiku
Institute of Library Science 2.57
Institute of Linguistics, Academy of Sciences, U.S.S.R. 3.17
Institute of Linguistics, Leningrad Department, Academy of Sciences U.S.S.R. 5.3.14
Institute of Mathematics, Siberian Branch, Academy of Sciences U.S.S.R. (Evreinov) 5.3.15
Institute of Mathematics, Siberian Branch, Academy of Sciences U.S.S.R. (Gladkiĭ) 3.16

Section-Article

Institute of Paper Chemistry 2.93
Institute of Physics and Mathematics, Academy of Sciences, Lithuania 3.43
Institute for Scientific Information 1.26, 2.58
Intectron, Inc. 4.11
International Business Machines Corp. 2.2, 2.120
International Business Machines Corp. (Baxendale) 2.59
International Business Machines Corp. (Courtney) 2.60
International Business Machines Corp. (Dickinson) 5.2.11
International Business Machines Corp. (Greanias) 5.1.14
International Business Machines Corp. (Griffin) 2.61
International Business Machines Corp. (King) 3.18
International Business Machines Corp. (Kochen) 2.62
International Business Machines Corp. (Nienburg) 2.63
International Business Machines Corp. (Rohland) 5.1.15
International Business Machines Corp. (Shelton) 5.1.16
International Business Machines Corp. (Stevens) 1.27
IBM-France (Levéry) 2.23, 2.64
IBM-France (Moreau) 5.3.16
IBM-France (Tabory) 3.19
International Council of Scientific Unions 1.28
International Federation for Documentation (FID) 1.28
Ireland, Luis G. 3.38r
Irish, Robert E. 2.98
Irons, E. T. 5.4.2r
Isner, Dale W. 5.3.38
Israel
  Hebrew University 5.3.9
Italy
  Centro per l'Automazione dell'Analisi Letteraria 5.3.4
  Communauté Européenne de l'Energie Atomique (Euratom) 2.21, 2.22, 3.6
  Ing. C. Olivetti & C., S.P.A. 5.1.11
  Università degli Studi di Milano 3.36, 5.4.26
Itek Corp. (Kuipers) 2.65, 4.12
Itek Corp. (O'Brien) 2.66
Iwabuti, Etutaro 5.3.24
Izard, Françoise 2.76r
Izzo, Joseph 2.3

Jacobsen, William H., Jr. 3.38r
Jacobson, Arvid W. 3.46r
Jacobson, S. N. 2.94r, 5.3.25r
Jacobus, David P. 4.23
Jaffe, Joseph 5.5.10
Jahoda, G. 2.35, 2.36
Janaske, Paul C. 1.3r
Janes, Marion 1.26r
Janiotis, Amelia 3.46r
Japan
  Electrotechnical Laboratory 3.9, 5.1.8
  First Research Center, Defense Agency 3.10

Japan (continued) Section Article

Japan Information Center of Science and Technology 2.67
Kyoto University 5.1.17, 5.2.12
Kyushu University 5.25
National Language Research Institute 5.3.24
Radio Research Laboratories 5.1.34, 5.2.20
Japan Information Center of Science and Technology 2.67
Jensen, Raymond A. 1.34
Johanningsmeier, Walter 1.23, 2.47
John Crerar Library 1.52
John Diebold and Associates, Inc. *See* Diebold Group, Inc.
John I. Thompson and Co., Inc. 1.29
Johns Hopkins University 2.68
Johnson, C. Douglas 5.38*r*
Johnson, David L. 1.55
Johnson, Eugene P., Jr. 4.1*r*
Johnson, H. Thayne 2.48
Johnson, Loren F. 2.11*r*
Jonas, Ronald W. 5.40*r*
Jones, Edgar A. 2.130
Jones, Paul E. 1.10*r*, 5.1.1, 5.3.2
Jonker Business Machines, Inc. 2.69
Jonker, Frederick 2.69
Joseph, R. D. 5.4.4
Josephson, R. 5.4.10*r*
Joshi, Aravind K. 5.3.37*r*
Josselson, Harry H. 3.46
Julesz, B. 5.2.4*r*

Kailath, Thomas 2.97
Kanal, L. N. 5.4.16
Kanal, Laveen 5.1.10
Kandler, G. 5.3.29
Kanter, Herschel 5.4.27*r*
Karlgren, Hans 5.3.7, 5.3.23
Karlova Universita (Novák) 3.20, 3.21, 3.44
Karlova Universita (Sgall) 2.70, 3.20, 3.21, 3.44, 5.3.17
Karush, W. 2.110*r*
Kaskey, Gilbert 2.115
Kassler, Michael 5.4.19
Kasvatusopillinen Korkeakoulu 3.22
Katter, Robert V. 2.106*r*
Kaufman, Bruria 5.3.37*r*
Kazmierczak, H. 5.1.42, 5.4.25*r*
Kehl, William B. 2.126, 5.3.38*r*
Kelley, K. L. 132*r*
Kelly, P. M. 5.4.4*r*
Kennedy, James 1.51
Kent, Allen 2.128
Kerimov, D. A. 2.72*r*
Kessler, M. M. 2.77
Kilgour, Frederick G. 1.58

Section-Article

Kincaid, Harry V. 1.44
Kincaid, William H. 2.20
King 5.5.6
King, Gilbert W. 2.74, 3.18
Kingsland, Lawrence C., Jr. 2.13
Kirsch, R. A. 5.4.14
Kita, J. 3.9r
Klein, D. B. 5.2.22r
Klein, Sheldon 5.4.24
Klima, E. S. 3.27r
Klingbiel, Paul H. 2.7r
Knapp, T. B. 5.1.35
Knight, B. W. 5.4.8r
Knowlton, K. C. 2.59r
Kochen, Manfred 2.62
Koelewijn, G. J. 2.88
Koenigsberg, Ruth Zeitlin 5.3.37r
Koizumi, Shinkichi 5.1.34r
Koller, Herbert R. 2.82r, 2.84r
Konečná, Dana 3.20r
Konigsberg, J. 4.17
Koriagin, Gretchen W. 2.32
Körner, Horst G. 2.108
Korvas, Z. 3.44r
Korvasová, Květa 3.44
Kosarev, Yu. G. 5.3.15
Kostíc, Djordje 3.14
Koutsoudas, Andreas 3.39
Kovalevskiĭ, V. A. 5.1.13
Kozak, A. S. 3.31r
Kraft, Donald 2.2r
Kraft, H. 5.4.25r
Kraft, Robert 3.7
Krallman, D. 5.3.29
Kranzhoff, A. 5.3.29
Kremneva, N. D. 3.24r
Krulee, Gilbert K. 5.4.15
Kubilius, J. 3.43
Kucera, Henry 5.3.3
Kuhns, J. L. 2.109
Kuikka, Teuvo 3.22
Kuipers, J. W. 2.65, 2.133r, 4.12
Kuno, S. 3.13r
Kurihara, T. 3.23
Kurmey, William J. 2.120
KVAL. *See* Forskningsgruppen för Kvantitativ Lingvistik.
Kyle, A. B. 2.50, 4.9
Kyle, Barbara 2.81
Kyoto University 5.1.17, 5.2.12
Kyushu University 3.23

Section-Article

Laboratoire d'Analyse Lexicologique 5.3.16, 5.3.18
Lamb, Sydney M. 3.38, 5.3.30
Landauer, W. I. 2.125r, 5.4.32r
Larkin, W. D. 5.2.10
Larsen, P. W. 4.10
Lattes, Robert 2.102
Lawlor, Reed C. 2.130
Lazarow, Arnold 2.3, 2.128r
Lazarsfeld, Paul F. 1.17
LeBrun, T. Q. 5.1.36
Lecerf, Y. 3.6, 5.3.29
Ledley, R. S. 2.80
Lees, R. B. 5.3.37r
Lehigh University (Hillman) 2.71
Lehigh University (Wuest) 1.30
Lehiste, I. 5.2.27r
Lehmann, W. P. 3.40
Leland, H. R. 5.2.8, 5.4.7
Lemaine, Gérard 2.118r
Lemke, G. A. 5.1.21r
Lemmon, A. 2.45r
Leningrad State University imeni A. A. Zhdanov 2.72, 3.24
Lenoir, M. 2.22r
Leonard, Gene F. 2.24r
Leroy, A. 2.21, 3.6
Letichevskiĭ, A. A. 5.4.9
Levéry, F. 2.23, 2.64
Levine, J. J. 5.3.21r
Lévy, F. 2.18r, 2.76r
Lewandowski, Frank P. 5.1.18
Lewis, J. L. 1.29r
Lewis, P. 2.59r
Library Association 2.73
Library of Congress (Dubester) 1.31, 2.30
Library of Congress (King) 2.74
Library Research Circle 2.75
Librascope Division of General Precision, Inc. (Hamilton) 5.4.10
Librascope Division of General Precision, Inc. (Worthy) 5.3.19
Licht, L. 5.5.6r
Licklider, J. C. R. 2.12
Lieberman, Philip 5.2.2.r
Lindsay, Robert K. 5.4.34
Link Division of General Precision, Inc. 5.1.18
Linney, D. S. 3.3
Lintell, R. J. 4.8
Lipetz, Ben-Ami 2.65r, 2.131
Lis, Zbigniew 5.3.37r
Lithuania
  Vilniaus Valstybinis V. Kapsuko Vardo Universitetas 3.43
Litton Systems, Inc. 5.1.19, 5.2.13
Livant, William P. 5.5.9r

Section-Article

Locke, W. N. 2.77
Lockheed Missiles & Space Co. (Firschein) 5.1.20
Lockheed Missiles & Space Co. (Rudin) 3.25
Loewenthal, Ariadne Lukjanow 3.26
Lofgren, L. 5.4.30r
Lowe, Thomas C. 5.4.32r
Lowenschuss, O. 5.1.35
Lu, C. W. 3.41r
Luckett, George R. 2.113
Luhn, H. P. 1.54, 3.41r
Lustig, G. 2.21r, 2.22r
Luxenberg, H. R. 4.8
Lynch, Michael F. 1.16r

McCarthy, John 5.4.11r
McConkle, C. C. 5.2.3r
McConlogue, Keren 5.4.24r
McCully, C. R. 2.8
McDaniel, J. 3.29
McEuen, Ronald 2.5r
McIntosh, Angus 5.3.34
McKinney, Norris P. 5.2.27r
McLachlan, K. R. 5.4.33
McPherson, William R. 5.5.3
Machine Translation, Inc. 3.26
Maesono, Ken-ichi-i 5.2.20r
Magnavox Research Laboratories 4.13
Mahany, Paul J. 1.23
Maison des Sciences de l'Homme 2.18, 2.76
Maloney, Clifford J. 2.111r
Maltzman, Irving 5.5.6
Manceaux, Harold L. 4.23r
Manchester University 5.3.20
Manley, Harold J. 5.2.22
Marckworth, Lois 2.61r
Marcus, S. 3.1r
Marden, Ethel 2.82, 2.84r
Maretti, Enrico 5.4.26r
Margvelani, L. 3.15r
Maron, M. E. 2.106, 5.4.20
Marquardt, Prof. Dr. 2.22
Martin, Samuel E. 3.38r
Marzocco, Frank N. 5.4.24
Massachusetts Institute of Technology (Barnett) 1.32, 5.3.21
Massachusetts Institute of Technology (Baumann) 4.14, 5.1.21
Massachusetts Institute of Technology (Chomsky) 5.3.22
Massachusetts Institute of Technology (Forgie) 5.2.14
Massachusetts Institute of Technology (Galliher) 1.33
Massachusetts Institute of Technology (Kessler) 2.77
Massachusetts Institute of Technology (Minsky) 5.4.11
Massachusetts Institute of Technology (Ross) 5.4.12

Section-Article

Massachusetts Institute of Technology (Scott) 4.15
Massachusetts Institute of Technology (Stevens) 5.2.15
Massachusetts Institute of Technology (Teager) 2.78
Massachusetts Institute of Technology (Yngve) 3.27
Masterman, Margaret 3.3
Matematikmaskinnämnden 5.3.23
Mathews, M. V. 5.2.25r
Mattson, R. L. 5.1.20
Mayer, Rollin P. 5.4.13
Meadow, H. R. 2.60r
Medical Research Council 5.5.4
Mednick, Sarnoff 5.5.6
Meeker, Willard F. 5.2.19
Meetham, A. R. 2.87
Meier, Georg F. 5.3.10
Mel'chuk, I. A. 3.16
Melnick, Donald A. 2.16
Melpar, Inc. 5.1.35
Melpar, Inc. (Cooper) 5.1.22
Melton, Jessica S. 2.128r
Memistor Corp. 4.16
Merck Sharp & Dohme Research Laboratories 2.124
Mergenthaler Linotype Co. 4.17
Merry, D. W. 1.56
Mersel, Jules 3.34
Mexico
Universidad Nacional Autonoma de Mexico 3.35
Meyer, Ernst 2.10, 4.23r
Meyer, Richard F. 5.1.1
Meyer-Uhlenried, K. H. 2.21r, 2.22
Micarelli, M. 2.22r
Michaels, Sheldon B. 5.2.2r
Michie, Donald 5.3.11r
Mikó, A. 1.37
Miller, Eugene 2.31
Miller, George 5.5.2
Miller, James G. 5.4.31, 5.5.9
Milton-Williams, J. 5.2.25r
Minker, Jack 2.94
Minsky, M. L. 2.97, 5.4.11
Mintz, L. 5.1.4r
Mitchell, David S. 2.126r, 5.3.38r
Mitchell, W. B. 2.90
Mitre Corp. 5.4.13
Mitsuhashi, Hiroshi 2.67r
Miwa, Daisaku 2.67
Mohr, D. V. 3.31r
Moisil, Gr. 3.1
Mommens, J. 3.37r
Montana State University 5.5.5
Montgomery, C. A. 2.109, 3.34

Section-Article

Mooers, Calvin N. 2.7r, 2.10r, 2.97
Moore, Frank W. 2.49
Moore, R. T. 2.82r
Moreau, R. 5.3.16
Morlet, E. 3.37r
Morris, D. 5.4.2r
Morris, Thomas G. 2.128r
Morrison, W. Adair 1.20
Morse, Rollin D. 1.4
Moss, J. H. 1.49r
Mote, L. J. B. 1.41
Motorola, Inc. 5.2.16
Mt. Alto VA Hospital 2.43
Muerle, J. L. 5.4.7r
Mühlen, W. 2.22r
Mullin, A. A. 5.4.30r
Mumbower, L. E. 2.60r
Munger, Sara J. 2.4r
Murdock, George P. 2.49r
Muth, E. P. 2.60r
Myers, Wilbur C. 4.20

Nakata, Kazuo 5.2.20
Namenwirth, J. Zvi 5.5.3
National Academy of Sciences-National Research Council 2.79, 2.99
National Aeronautics and Space Administration 1.24, 2.99
National Agricultural Library and Centre for Documentation. *See* Országos Mezögazdasági Könyvtár És Dokumentációs Központ
National Biomedical Research Foundation 2.80
National Book League 2.81
National Bureau of Standards (Alexander) 2.82, 2.99, 2.114
National Bureau of Standards (Bagg) 4.18
National Bureau of Standards (Kirsch) 5.4.14
National Bureau of Standards (Rhodes) 3.28
National Bureau of Standards (Stern) 2.83, 4.19
National Bureau of Standards (Stevens) 2.84, 5.1.23
National Cash Register Co. 4.20
National Chiropractic Association 2.85
National Federation of Science Abstracting and Indexing Services 1.34
National Institutes of Health 1.35, 2.47
National Language Research Institute 5.3.24
National Library of Medicine 2.3, 2.86, 4.21
National Physical Laboratory (Clayden) 5.1.24
National Physical Laboratory (Day) 3.13, 3.29
National Physical Laboratory (Meetham) 2.87
National Science Foundation 2.7
Nebeský, Ladislav 5.3.17r
Needham, R. M. 2.15
Neelameghan, A. 2.75

Section-Article

Nelson, A. L. 5.2.19r
Netherlands
International Federation for Documentation (FID) 1.28
Octrooiraad (Netherlands Patent Office) 2.88
Stichting Studiecentrum Voor Administratieve Automatisering 5.4.23
Netherlands Automatic Data Processing Research Center. *See* Stichting Studiecentrum Voor Administratieve Automatisering.
Neuts, Marcel F. 2.117
Newell, Allen 5.4.21, 5.4.27
Newill, Vaun 2.3, 2.128r
Newman, S. M. 2.59r
New York Botanical Garden 1.36
New York University (Boni) 5.1.25
New York University (Freeman) 5.1.26
New York University (Gerstman) 5.2.5, 5.2.17
Nida, Eugene A. 3.3.1
Nielsen, Carl E., Jr. 4.1r
Nienburg, R. E. 2.63
Nilsson, N. J. 5.1.38r
Nippon Electric Co. Ltd. 5.2.12, 5.2.20
Niwa, Yasujiro 2.67
Nobbs, P. M. 2.93
Nordiska Sommaruniversitetet 3.22
North American Aviation, Inc. *See* Autonetics.
Northwestern University 5.4.15
Novák, Pavel 3.20
Novikoff, Albert B. 3.20r
Novozamská, Helena 3.20r

Oblonský, J. 3.44r
Obrecht, Dean H. 5.2.28
O'Brien, B. 2.66
O'Connor, J. S. 2.127r
O'Connor, John 2.124
Octrooiraad (Netherlands Patent Office) 2.88, 2.114
O'Donnell, Theodore R. 1.29
Oettinger, Anthony G. 3.13
Office of Technical Services 2.7, 2.82, 2.99
Ogata, Yoshiharu 5.1.34, 5.2.20
Ogilvie, Daniel M. 5.5.3
Ohio State University 3.30
Ohlenbusch, Cord W. 5.1.21r
Öhman, Sven 5.2.21r
Olney, John 2.106r
Oono, Y. 3.23
Opler, Ascher 3.8
Orosz, G. 2.57
Orr, Richard H. 1.25
Országos Mezögazdasági Könyvtár És Dokumentációs Központ 1.37
Osgood, Charles E. 1.53
Oswald, V. A., Jr. 2.90

Section-Article

Overmyer, LaVahn 2.128r
Overton, Richard K. 5.4.5

Pacific Southwest Forest and Range Experiment Station 2.89
Pagès, Robert 2.118
Painter, Ann F. 2.99
Pak, G. A. 3.24r
Palek, Bohumil 2.70r, 3.44r
Panevová, Jarmila 3.21r
Panoramic Research, Inc. 5.1.27
Parker-Rhodes, A. F. 3.3r
Parks, J. R. 5.1.24
Pataky, E. 2.57r
Patterson, George W. 5.4.32
Paul, A. P. 5.2.15r
Paul, R. N. 2.122
Payne, Dan 2.4
Pazukhin, R. V. 3.24r
Pendergraft, E. D. 3.40
Perdue, Robert E., Jr., ... 2.112
Peretz, B. 5.4.31
Perles, M. 5.3.9r
Perotto, Giorgio 5.1.11
Perriault, J. 3.4r
Perry, E. G. 5.1.36
Perry, James W. 2.119
Peterhansl, E. 2.22r
Peters, Stanley C. 5.5.8r
Peters-Holger, K. 5.3.29
Peterson, Gordon E. 5.2.27
Petrick, Stanley R. 5.2.1, 5.4.2
Petrovici, Em. 3.1
Pfeffer, H. 2.82r
Philco Corp. (Bogusz) 5.1.28
Philco Corp. (Bryan) 5.4.16
Philco Corp. (Chatten) 5.1.29, 5.4.17
Philco Corp. (Frank) 5.1.30
Philco Corp. (Teacher) 5.2.18
Phillips, Moira 1.11
Photon, Inc. 4.21
Physikalisches Institut der Technischen Hochschule Aachen 5.4.18
Piault, Colette 2.18r
Pick, George G. 5.1.41
Picot, G. 2.23
Pietsch, E. H. E. 2.22, 2.44
Pike, J. L. 4.18
Pine, Miles C. 2.48r
Piovia, Frank 2.16
Pipberger, Hubert V. 2.43
Pirtakhalava, M. D. 3.15r
Piske, U. 5.4.25

Section-Article

Planning Research Corp. 2.90
Poindron, Paul 1.28
Poland
  Uniwersytet Warszawski 3.42
Popper, Hans 5.3.31
Post Office Department 5.1.9, 5.1.31
Postal, S. K. 5.5.3r
Postley, John A. 1.1, 5.5.1
Powers, E. N. 5.1.28
Prather, Rebecca 5.1.45
Pray, R. F. 4.17r
Price, W. L. 3.29
Princeton University (Babbitt) 5.4.19
Princeton University (Tukey) 2.91
Project Lawsearch 2.92
Prywes, Noah S. 2.125, 5.4.32r
Psathas, G. 5.5.3r
Pulp and Paper Research Institute of Canada 2.93
Pustula, J. H. 3.31r

Quant, W. B. 5.5.3r
Quemada, Bernard 5.3.18
Quillian, Ross 3.27r
Quirk, Randolph 5.3.32

Rabinow Engineering Co., Inc. 5.1.32
Rabinow, Jacob 5.1.32
Radio Corp. of America (Climenson) 5.3.25
Radio Corp. of America (Hannan) 5.1.33
Radio Corp. of America (Meeker) 5.2.19
Radio Corp. of America (Minker) 2.94
Radio Research Laboratories 5.1.34, 5.2.20
Ragland, Evan L. 5.216
Rahman, Abdul 2.75
Raisig, L. Miles 1.58
Ramallo, L. I. 5.5.3r
RAND Corp. (Hays) 3.31, 5.3.26
RAND Corp. (Maron) 5.4.20
RAND Corp. (Shaw) 5.4.21
Ranganathan, S. R. 2.75, 2.122
Ranganathan, T. 2.75
Rankin, B. K., III 5.1.23, 5.4.14
Rapoport, Anatol 5.5.9
Rath, G. J. 2.63r
Ray, L. C. 2.109r
Raytheon Co. 5.1.35
Reading Chemists' Club 2.95
Recognition Equipment Inc. 5.1.36
Reed, Virginia R. 1.52
Reed, W. G. 2.125r
Rees, Alan 2.128r
Reeves, Pamela W. 2.128r

Section-Article

Reifler, Erwin 3.41
Reisner, P. 2.62r
Reiss, Richard F. 5.4.10
Reitz, Gerhard 3.34
Renaut, Mme. 2.18
Rensselaer Polytechnic Institute 2.96
Research, Inc. 1.38, 3.32
Research Information Center and Advisory Service on Information Processing 2.84, 4.18
Research Institute for Mathematical Machines. *See* Výzkumný Ústav Matematických Strojů.
Resnick, A. 2.63r
Reuschlen, P. 5.1.42r
Rhodes, Ida 3.28
Richards, T. W. 2.60r
Richmond, G. E. 5.4.7r
Rigby, Malcolm 1.7
Ristinen, Elaine K. 2.51
Robbins, Beverly 5.3.37r
Robert Heller & Associates 1.34
Roberts, A. Hood 5.3.40
Robinson, Herbert 2.16
Rockford Research Institute Inc. (Mooers) 2.97
Rockford Research Institute Inc. (Solomonoff) 5.4.22
Roetling, Paul G. 2.25
Rohland, William S. 5.1.15
Rome Air Development Center 2.98
Ronco, Paul 1.47
Rosen, Charles 5.1.39
Rosenblatt, Frank 5.4.7r, 5.4.8
Rosenborg, Staffan 2.27
Rosenfeld, Azriel 5.1.3
Rosetti, Al. 3.1
Ross, Douglas T. 5.4.12
Rothkirch-Trach, Graf. 2.22
Rotolo, L. S. 2.80
Roy, Archie E. 5.4.35
Roy, Bernard 2.102
Roy, Robert H. 2.68
Royal Institute of Technology 5.2.21
Royse, D. R. 1.27
Rudin, B. D. 3.25
Ruechardt, Edward 3.40
Rumania
  Académie de la République Populaire Roumaine 3.1
  Independent Work (Abraham) 5.3.41
Rupp, W. H. 2.36
Rutgers, The State University (Clarke) 1.39
Rutgers, The State University (Harlow) 2.99
Rutgers, The State University (Harrar) 1.40
Rybakova, M. V. 3.17

Section-Article

Sabel, C. S. 1.49
Sager, Naomi 5.3.37r
Saila, Saul B. 2.127
St. George, Emery, Jr. 4.1r
Sakai, Itiroo 3.8, 3.10
Sakai, Toshiyuki 5.1.17, 5.2.12
Salton, Gerard 2.45
Sams, Burnett H. 2.94r
Sandel, T. T. 5.2.15r
Sandoz Ltd. 2.37
Santucci, Daniel 2.5
Sato, Nobukatsu 5.1.34r
Savage, T. R. 2.32r, 2.63
Schafer, Emil 2.48r
Schäring, Lennart 3.22
Scheffer, Emilie 3.6r
Schenk, H. R. 2.37
Schick, Frank L. 2.128r
Schneider, John H. 2.132
Schnelle, H. 5.2.24, 5.3.29
Schultheiss, Louis A. 2.121
Schultz, Claire K. 2.54
Schwartz, Mortimer D. 2.123
Schweisthal, G. 5.3.29
Science Index Group 2.100
Scientific Documentation Centre Ltd. 2.101
SCOPE Inc. 5.1.37
Scotland
  Scientific Documentation Centre Ltd. 2.101
  University of Edinburgh 5.2.26, 5.3.34
  Independent Work (Roy) 5.4.35
Scott, B. J. 3.31r
Scott, P. B. 5.2.19r
Scott, Peter 4.15
Sebeok, Thomas A. 2.51, 5.3.12
Sebestyen, George 5.1.19, 5.2.13
Selfridge, O. G. 5.4.11r
SEMA. *See* Société d'Économie et de Mathématique Appliquées.
Semenovskiĭ, A. G. 5.1.13r
Sendhoff, Hw. 5.2.24
Service d'Etudes Sémiologiques et Documentaires 2.18
Sestier, A. 3.4
Sgall, Petr 2.70, 3.21, 5.3.17
Shaffer, S. S. 5.4.24
Shamir, E. 5.3.9r
Shaw, John C. 5.4.21, 5.4.27
Shed'ko, T. I. 3.17
Sheinberg, I. 5.1.36
"Shell" Research Ltd. 1.41
Shelton, G. L., Jr. 5.1.16
Shepard, David H. 5.2.6

Section-Article

Shepherd, Clayton A. 2.54r
Sher, Irving H. 1.26, 2.58
Shera, J. H. 2.128
Sherry, Murray E. 5.3.6
Shields, Joanna E. 1.3r
Shilling, Charles W. 1.3
Shneidman, E. 5.5.3r
Shoup, J. E. 5.2.27r
Siegel, Milton 2.28
Simmons, Robert F. 5.4.24
Simon, Herbert A. 5.4.21, 5.4.27
Simon, S. 5.5.6r
Singer, Jerome R. 5.1.43
Sivertsen, E. 5.2.27r
Slagle, James 5.4.11r
Slater, M. 1.12
Slaymaker, F. H. 5.1.10r, 5.2.10
Slee, Vergil N. 2.20
Smith, A. H. 5.3.33
Smith, Allan H. 3.45
Smith, C. 5.5.3r
Smith, C. H. 3.31r
Smith, D. 5.1.10r
Smith, Howard W. 2.11
Smith, Jean 2.128r
Smith, Steven B. 3.34
Smoke, W. 3.39r
Société d'Économie et de Mathématique Appliquées (SEMA) 2.102
Société Phototechnique 4.2
Solomonoff, Ray J. 5.4.22
Southwestern Legal Foundation 2.103
Sowarby, A. J. 2.32r, 2.63r
Spangler, Marshall 1.19
Sparck Jones, K. 2.15r
Sparks, D. E. 2.131r
Spencer Chemical Co. 2.104
Spencer, R. A. 2.114
Sperry Rand Corp. *See* UNIVAC Division.
Spiegel, Joseph 5.4.13r
Spiegelthal, Edwin S. 5.3.42
Spinks, Paul 2.113
Spooner, M. G. 5.4.7r
Standard Oil Co. 1.42
Stanford Research Institute (Brain) 5.1.38
Stanford Research Institute (Engelbart) 1.43
Stanford Research Institute (Kincaid) 1.44
Stanford Research Institute (Rosen) 5.1.39
Stanford Research Institute (Younker) 2.105
Stanford University 5.5.3
Statulevičius, V. 3.43
Steinbuch, K. 5.1.42, 5.4.25

Section-Article

Stern, Joshua 2.88, 4.19
Stevens, Kenneth N. 5.2.15
Stevens, M. E. 2.84, 5.1.23
Stevens, L. D. 1.27
Stiassny, S. 2.63r
Stichting Studiecentrum Voor Administratieve Automatisering 5.4.23
Stiles, H. Edmund 2.29
Štindlová, Jitka 5.3.5
Stognií, A. A. 5.4.9
Stone, Eileen 2.106r
Stone, Philip J. 5.5.3
Strasel, H. C. 2.4r
Strohm, W. B. 3.18r
Sugiura, S. 3.9r
Summer Institute of Linguistics 3.33, 5.3.27
Surveys and Research Corp. 1.45
Suszko, Roman 3.42
Suzuki, Jouji 5.2.20r
Svoboda, A. 3.44r
Swanson, Don R. 2.109, 3.34
Swanson, R. W. 2.59r
Swarthmore College 5.1.40
Sweden
  Royal Institute of Technology 5.2.21
  Forskningsgruppen för Kvantitativ Lingvistik (KVAL) 5.3.7
  Matematikmaskinnämnden 5.3.23
  University of Gothenburg 5.3.35
Swift, Wayne B. 2.16
Switzerland
  F. Hoffmann-La Roche & Co. Ltd. 2.37
  Sandoz Ltd. 2.37
Sylvania Electronic Systems 5.1.22
Sylvania Electronic Systems (Gray) 5.1.41
Sylvania Electronic Systems (Manley) 5.2.22
Syracuse University 1.46, 2.98
System Development Corporation (Borko) 2.106
System Development Corp. (Gullahorn) 5.4.28
System Development Corp. (Harrington) 2.107
System Development Corp. (Marzocco) 5.4.24
System Development Corp. (Prather) 5.1.45
Szanser, A. J. 3.29

Tabory, R 3.19, 3.48r
Tadenuma, R. 3.9
Tailleur, O. 3.4r
Taine, Seymour I. 2.86
Tamati, T. 3.23
Tamarashvili, L. 3.15r
Tanimoto, T. 5.1.35
Tanner, Wilson P., Jr. 5.1.44
Tasman, J. E. 2.95

Section-Article

Tauber, A. S. 4.20r
Teacher, C. F. 5.1.29, 5.2.18
Teager, Herbert M. 2.78
Technische Hochschule Karlsruhe 5.1.42, 5.4.25
Technische Hochschule München 2.106
Te Nuyl, Th. W. 2.51r
Terry, J. E. 1.49
Tersoff, Abraham I. 5.1.9
Thatcher, J. 5.4.31r
Thode, Edward F. 2.93
Thomas, William H. B. 2.92
Thompson, Frederick B. 2.41
Thompson, Mendall S. 2.80
Thompson Ramo Wooldridge Inc. (Bertram) 5.2.23
Thompson Ramo Wooldridge Inc. (Edmundson) 2.109, 3.34
Thompson Ramo Wooldridge Inc. (Garvin) 2.110
Thorelli, Lars Erick 5.3.23r
Thorne, J. P. 5.3.11
Thornton Research Centre 1.41
Timisoara Polytechnical Institute 3.1
Tokyo Shibaura Electric Co. 2.67
Tolman, E. C. 5.4.24
Toma, Peter 3.7
Tonge, Fred M. 5.4.21r, 5.4.27r
Toumanov, V. 5.4.23r
Trantanella, F. M. 5.1.21r
Travis, L. E. 5.4.24
Tufts University 5.1.14
Tufts University (Ronco) 1.47
Tukey, John W. 2.91
Tuskegee Institute 5.3.28

Uffelman, M. R. 5.1.37r
Uhr, Leonard 5.1.45
Uldall, E. T. 5.2.26r
Unesco 2.81
Union of International Associations 1.28
U.S.S.R.
  Institute of Automatics and Telemechanics, Academy of Sciences U.S.S.R. 5.1.12
  Institute of Cybernetics, Academy of Sciences Ukrainian S.S.R. 5.1.13, 5.4.9
  Institute of Electronics, Automatics and Telemechanics, Academy of Sciences Georgian S.S.R. 3.15
  Institute of Linguistics, Academy of Sciences U.S.S.R. 3.16
  Institute of Linguistics, Leningrad Department, Academy of Sciences U.S.S.R. 5.3.14
  Institute of Mathematics, Siberian Branch, Academy of Sciences U.S.S.R. 3.17, 5.3.15
  Leningrad State University imeni A. A. Zhdanov 2.72, 3.24
United Kingdom Atomic Energy Authority (Bell) 1.48, 1.49
United Kingdom Atomic Energy Authority (Sabel) 1.48, 1.49
U. S. Army Biological Laboratories 2.111

Section-Article

U. S. Atomic Energy Commission 1.50, 2.27, 2.99
U. S. Department of Agriculture 2.112
U. S. Naval Postgraduate School 2.113
U. S. Patent Office 2.82, 2.88, 2.114
U. S. Post Office 5.1.9, 5.1.31
U. S. Public Health Service 1.25
UNIVAC Division of Sperry Rand Corp. 2.115
Universidad Industrial de Santander 2.116
Universidad Nacional Autonoma de Mexico 3.35, 3.38
Università degli Studi di Milano 3.36, 5.4.26
Universität Bonn (Heike) 5.2.24
Universität Bonn (Schnelle) 5.3.29
Université Catholique de Louvain 2.117
Université de Paris 2.118
Université Libre de Bruxelles 3.37
University of Arizona 2.119
University of California, Berkeley (Feigenbaum) 5.4.27
University of California, Berkeley (Gullahorn) 5.4.28
University of California, Berkeley (Lamb) 3.35, 3.38, 5.3.30
University of California, Berkeley (Singer) 5.1.43
University of California, Livermore (Giallanza) 1.51
University of California, Los Angeles 2.119, 2.130, 5.4.10, 5.5.3
University of California, Los Angeles (Maltzman) 5.5.6
University of Cambridge (Wilkes) 4.22
University of Cambridge (Wisbey) 5.3.31
University of Chicago (Greene) 5.4.29
University of Chicago (Kurmey) 2.120
University of Chicago (Reed) 1.52
University College London (Fry) 5.2.25
University College London (Quirk) 5.3.32
University College London (Smith) 5.3.33
University College of Swansea 5.3.31
University of Connecticut 5.5.7
University of Edinburgh (Abercrombie) 5.2.26
University of Edinburgh (McIntosh) 5.3.34
University of Gothenburg 5.3.35
University of Illinois (Osgood) 1.53
University of Illinois (Von Foerster) 5.4.30
University of Illinois Library 2.121
University of Kansas 1.54, 2.123
University of London (Birkbeck College) 3.2
University of Maryland 5.5.8
University of Michigan (Koutsoudas) 3.39
University of Michigan (Miller) 5.4.31, 5.5.9
University of Michigan (Paul) 2.122
University of Michigan (Peterson) 5.2.27
University of Michigan (Tanner) 5.1.44
University of Michigan (Uhr) 5.1.45
University of Minnesota 2.3, 2.128
University of Oklahoma 1.54, 2.123
University of Pennsylvania (Brown) 5.3.36

University of Pennsylvania (Harris) 5.3.37
University of Pennsylvania (O'Connor) 2.124
University of Pennsylvania (Patterson) 5.4.32
University of Pennsylvania (Prywes) 2.125
University of Pittsburgh (Horty) 2.126
University of Pittsburgh (Isner) 5.3.38
University of Rochester 2.3, 2.128
University of Rochester (Obrecht) 5.2.28
University of Rhode Island 2.127
University of South Florida 5.3.40
University of Southampton 5.4.33
University of Texas (Lehmann) 3.40
University of Texas (Lindsay) 5.4.34
University of Washington (Johnson) 1.55
University of Washington (Reifler) 3.41
University of Wisconsin (Fowler) 5.3.39
University of Wisconsin (Roberts) 5.3.40
Uniwersytet Warszawski 3.42
Ustinov, V. A. 5.3.15
Utica College, Syracuse University 2.98

Vallance, John A. 2.107r
Valvoda, Mary Alice 2.128r
Vandeputte, N. 2.17r
Vanderbilt University 2.132
Vann, Col. James O. 2.7
Vardy, P. 2.22r
Vataire, Danièle 2.118r
Vaughn, Lou E. 2.48r
Vauquois, Bernard 3.5
Veillon, G. 3.5r
Vendler, Zeno 5.3.37r
Verheyden, Jean 3.6r
Veterans' Administration 2.107
Veyrunes, J. 3.5r
Vickers, T. 2.42r
Vickery, B. C. 2.28r
Viglione, S. S. 5.4.4
Vilniaus Valstybinis V. Kapsuko Vardo Universitetas 3.43
Visintini, Prof. 5.4.26
Volkov, N. N. 3.24r
Von Foerster, Heinz 5.4.30
Vossler, C 5.1.45r
Výzkumný Ústav Matematických Strojů 3.44

Wächter, G. 2.22r
Wada, H. 5.1.8r
Wagner, S. W. 5.4.25r
Wahlgren, John H. 3.38r
Waite, David P. 1.24, 2.52
Waldo, W. H. 4.23r
Walker, Donald E. 5.4.13

Section-Article

Walker, Gordon L. 1.6
Walker, W. 5.1.10*r*
Wall, Eugene 1.4*r*, 2.7*r*, 2.34
Wall, Robert E., Jr. 2.33
Wallace, Everett M. 2.48*r*
Walsh, G. W. 4.17*r*
Walsh, Warren B. 1.46
Walter Reed Army Medical Center 4.23
Walton, Thomas S. 2.28
Wang, Fang Yu 3.18, 3.47
Wang, T. L. 2.42
Wang, William S.-Y. 3.30, 5.2.27*r*
Warburg, Jeremy 5.3.33*r*
Warren Spring Laboratory 2.46
Warsaw University. *See* Uniwersytet Warszawski.
Warshall, S. 2.24, 5.4.2*r*
Washington State University 3.45
Wason, P. C. 5.3.33*r*
Wathen-Dunn, Weiant 5.2.2
Wayne, Ivor 1.15
Wayne State University 3.46
Webb, K. W. 2.60*r*
Wegmüller, F. 2.37
Wegstein, J. H. 5.3.38, 5.4.32
Weiss, Mark 5.2.9
Weissblum, Walter 5.1.41
Welt, Isaac D. 2.55
Wenck, Gunther 5.3.43
Wenke, Klaus 2.10*r*, 4.23*r*
Wennerberg, A. P. 2.8
Western Electric Co., Inc. 1.56
Western Reserve University 1.15, 2.3, 2.128
Westgard, Rolf E. 4.13
Weston, P. 5.4.30*r*
Wexler, P. J. 5.3.20
Whaley, Fred R. 2.6
Wheater, R. H. 2.104
Whelan, S. 3.29
Whittier, John R. 1.18
Whyte, I. D. 1.13
Widrow, Bernard 4.16
Wildberger, August Martin 2.113*r*
Wilde, D. 2.78*r*
Wildlife Disease Association 1.57
Wilkes, M. V. 4.22
William Alanson White Institute 5.5.10
Williams, T. M. 2.65*r*, 2.133
Willis, M. P. 2.4*r*
Wilson, Harold H. 2.107*r*
Wilson, J. B. 2.80
Wilson, Robert A. 2.103

Section-Article

Winter, R. E. 2.60r
Wisbey, R. 5.3.31
Wiswesser, William J. 2.95
Wojtasiewicz, Olgierd 3.42
Wolf, H. F. 5.4.4r
Wolfe, Hugh C. 1.5
Wolfe, J. Edmond 4.5r
Wong, E. 2.62r
Worth, D. S. 5.3.26r
Worthy, R. M. 5.3.19, 5.4.10r
Wright, G. H. 2.46
Wright, Mary T. 1.20
Wright-Patterson Air Force Base 2.16
Wuest, Francis J. 1.30
Wyllys, R. E. 2.106r

Yahn, George 2.5r
Yale Medical Library 1.58
Yale University 3.47
Yates, D. M. 3.29
Yerke, Theodor B. 2.89
Yngve, Victor H. 3.27, 5.4.24, 5.5.3r
Yoneyama, Kazuhiko 5.1.34r
Yosimura, A. 3.23r
Youden, W. W. 5.4.32r
Younker, E. LeRoy 2.105
Yovich, J. C. 1.5r
Yugoslavia
Institut za Eksperimentalnu Fonetiku 3.14

Zacharjan, M. 3.43
Zambrzhitskii, V. L. 3.24r
Zator Co. 2.97
Zeps, V. J. 5.3.12r
Ziehe, Theodore W. 3.31, 5.3.26
Zopf, G. W., Jr. 5.4.30r
Zusman, F. S. 2.80
Zygouris, Radmila 2.76r

Index of Sponsors

The statements have been assigned consecutive numbers within the various sections. The "Section-Article" column below denotes the numbers of the section and article describing the research sponsored by the organization listed.

Section-Article

Autonetics, Minuteman High Reliability Components Program 2.42

Beckman Instruments, Ltd. 2.101
Bellingham and Stanley, Ltd. 2.101
Beratungsstelle für Dokumentation und Information 2.22
British Drug Houses, Ltd. 2.101

Carnegie Corp. 5.4.21, 5.4.27
Clarendon Press 5.3.32
Communauté Européenne de L'Energie Atomique. *See* Euratom
Council on Library Resources, Inc. 1.57, 2.12, 2.74, 2.92, 2.109, 2.121, 2.126, 4.11, 4.15

Department of Scientific and Industrial Research 5.4.33
Deutsche Forschungsgemeinschaft 2.108, 5.1.42, 5.3.43, 5.4.25

Euratom 2.102, 2.117, 3.4, 3.36, 3.37, 5.3.43, 5.4.23, 5.4.26

Fife County Council Trust Fund 2.101
Ford Foundation 2.76, 2.126, 5.5.3

Gmelin Institute 2.97

Information for Industry, Inc. 2.16
International Business Machines Corp. 2.90, 3.8, 3.47, 5.3.38
IBM-United Kingdom Ltd. 5.3.34

Jacok Boehme Society, Inc. 5.3.31
Japanese Ministry of Education 3.23, 5.2.12

Longmans, Green and Co., Ltd. 5.3.32

Ministry of Agriculture, Hungary 1.37
Ministry of Aviation, England 5.2.25
Modern Medicine 1.44

National Aeronautics and Space Administration 1.7, 1.24, 2.31
National Research Council, Italy 5.4.26
National Science Foundation 1.1, 1.3, 1.5, 1.6, 1.7, 1.9, 1.10 1.15, 1.17, 1.22, 1.23, 1.26, 1.29, 1.30, 1.31, 1.33, 1.34, 1.36, 1.45, 1.46, 1.47, 1.53, 1.57, 2.1, 2.9, 2.15, 2.16, 2.19, 2.27, 2.30, 2.34, 2.52, 2.68, 2.71, 2.77, 2.78, 2.79, 2.80, 2.81, 2.82, 2.94, 2.91, 2.95, 2.114, 2.121, 2.128, 2.131, 3.13, 3.27, 3.30, 3.34, 3.38, 3.40, 3.41, 5.1.40, 5.2.15, 5.2.28, 5.3.12, 5.3.22, 5.3.37, 5.4.14, 5.4.15, 5.4.30, 5.5.2, 5.5.9

Section-Article

Naturmetodens Sproginstitut 5.5.32
North Atlantic Treaty Organization 2.73

RAND Corp. 5.4.27

Smith Kline and French Research Institute 2.101
Social Science Research Council 5.4.31, 5.5.3
Sport Fishing Institute 2.127
System Development Corp. 5.4.27

Unesco 1.28

U. S. Department of Commerce
- National Bureau of Standards 1.24
- Office of Technical Services 2.99
- Patent Office 5.4.14
- Weather Bureau 1.7

U. S. Department of Defense
- Office of Director of Defense Research and Engineering
  - Advanced Research Projects Agency 5.1.35
- Department of the Air Force 1.7, 4.12, 5.2.15, 5.2.25, 5.4.7
  - Aeronautical Systems Division 2.16, 2.48, 5.2.19, 5.4.7, 5.4.12, 5.4.30
  - Air Force Systems Command 2.62, 5.3.2, 5.3.6
    - Electronic Systems Division
      - Rome Air Development Center 2.1, 2.4, 2.25, 2.31, 2.33, 2.98, 2.105, 2.109, 3.18, 3.34, 3.36, 4.8, 4.17, 5.1.25, 5.1.28, 5.1.33, 5.1.37, 5.2.9, 5.2.13, 5.2.22, 5.3.11, 5.4.1, 5.4.17, 5.4.32
  - Office of Aerospace Research
    - Air Force Cambridge Research Laboratories 2.45, 2.62, 2.65, 5.1.19
    - Air Force Office of Scientific Research 1.23, 1.39, 1.40, 1.43, 1.55, 2.31, 2.47, 2.51, 2.62, 2.69, 2.97, 2.99, 2.110, 2.115, 2.119, 2.124, 2.128, 3.3, 5.1.3, 5.1.26, 5.1.44, 5.2.7, 5.2.27, 5.3.13, 5.3.36, 5.4.6, 5.4.10, 5.4.22, 5.5.5
  - Project RAND 5.3.26, 5.4.20
- Department of the Army 1.7
  - Office of the Assistant Chief of Staff for Intelligence 2.94
  - Office of the Surgeon General 5.5.8
  - U. S. Army Materiel Command
    - U. S. Army Electronics Research and Development Laboratory (formerly U. S. Army Signal Research and Development Laboratory) 3.40, 5.1.38, 5.2.10, 5.4.17
    - U. S. Army Electronics Supply Agency (formerly U. S. Army Signal Supply Agency) 5.4.7
  - U. S. Army Research Office 3.28
  - U. S. Army Signal Corps 3.28, 5.1.4, 5.1.23
- Department of the Navy
  - Bureau of Naval Weapons 2.8
  - Bureau of Ships 2.28, 4.18
  - Bureau of Supplies and Accounts 2.125
  - Office of Naval Research 2.78, 2.124, 2.125, 3.3, 3.30, 3.46, 4.1, 4.14, 5.1.21, 5.2.27, 5.3.9, 5.4.1, 5.4.3, 5.4.7, 5.4.8, 5.4.29, 5.4.30, 5.5.3, 5.5.6, 5.5.7

Section-Article

U. S. Department of Health Education and Welfare
Food and Drug Administration 2.82
Office of Education 2.128
Public Health Service 5.1.45, 5.5.2
National Institutes of Health 1.3, 1.16, 1.18, 1.25, 1.26, 1.32, 1.58, 2.3, 2.13, 2.20, 2.31, 2.47, 2.54, 2.56, 2.97, 2.126, 2.128, 5.2.27, 5.3.21, 5.4.22, 5.5.3
National Heart Institute 2.43, 2.55
National Institute of Mental Health 1.23, 2.49, 2.55, 5.5.10
National Library of Medicine 1.25, 4.21
U. S. Department of the Interior
Fish and Wildlife Service 2.127
U. S. Department of State
Peace Corps 5.5.3
U. S. Post Office Department 5.1.9
University of Freiburg 2.22

Western Management Science Institute 5.4.27

Equipment Index

Part I consists of an alphabetical index to all equipment, devices, and storage media mentioned in the statements. Part II classifies equipment according to some of the general categories of use. The numbers following the index entries denote the section and article in which reference to the specific equipment is made.

PART I

ACE computer, 3.29
Addressing machine equipment, 4.9
ALWAC computer, 5.3.35
AN/GSQ-16, 3.8
AN/GSQ-16 (XW-1), 3.18
Analog-digital machine, 5.4.33
Automatic Vocal Transaction Analyzer, 5.5.10
Automaton for pattern recognition, 5.1.42

Besk computer, 5.3.7
BIMA "P" tape typewriter, 2.22
Burroughs 220 computer, 5.1.5, 5.1.6

CDC 1604 computer, 2.113, 5.3.39, 5.3.40
CIFA computer, 5.3.41
Command Retrieval Information System (CRIS), 4.10
Computer printer, 4.21
Computers (Specific computers are indexed by designation, e.g., IBM 650 computer)
  (not otherwise specified), 2.55
  use of. *See* Part II of this index
"Crossfiler," special-purpose logic control device, 4.12

Dynamic signal analyzer, 5.4.30

Edge-punched cards, 2.57
EDSAC 2 computer, 2.15, 3.3
Electrochemical storage device, 4.16
EPOS 1 computer, 3.44
Experimental machine for pattern recognition, 5.1.14

Facit computer, 5.3.7
FileSearch system, 4.3
Film storage device. *See* Photographic storage device
Flexowriter keyboard, limitations of, 5.3.31
Flexowriter tape typewriter, 1.32, 2.17, 2.22, 2.32, 2.37, 2.59, 2.104, 2.128
Flying spot scanner, 5.1.5, 5.1.10, 5.1.11, 5.1.13, 5.1.14, 5.1.31
Formant Vocoder, 5.2.10
FOSDIC System, 2.43

Gamma 60 computer, 2.18
GE 225 computer, 1.19, 2.42
Graphic Arts Composer, 2.86

High-density storage device, 4.1, 4.5, 4.14
High-speed scanning storage device, 4.1, 4.13, 4.14

IBM 047 tape-to-card converter, 2.22
IBM 082 sorter, 1.7
IBM 108 card proving machine, 2.36
IBM 305 RAMAC computer, 2.114
IBM 407 tabulator, 1.7
IBM 514 Reproducing Punch, 5.1.14
IBM 650 computer, 2.63, 2.126, 3.36, 3.46, 5.3.11, 5.3.16, 5.3.27
IBM 704 computer, 1.7, 3.36, 3.38, 5.4.7, 5.4.27
IBM 705 computer, 2.23
IBM 709 computer, 1.32, 1.55, 2.51, 2.60, 3.40, 5.3.11, 5.3.21, 5.4.15
IBM Cardatype 858 Accounting Machine, 2.67
IBM 870 Document Writer System, 1.7, 2.22

IBM 1401 computer, 1.7, 1.16, 1.51, 1.54, 2.6, 2.10, 2.16, 2.17, 2.19, 2.21, 2.22, 2.27, 2.28, 2.32, 2.36, 2.37, 2.58, 2.60, 2.61, 2.63, 2.65, 2.68, 2.80, 2.103, 2.113, 2.114, 2.119, 2.120, 3.8, 3.46, 5.3.38
IBM 1403 printer, 1.16, 2.19
IBM 1405 Disc Storage unit of 1401 Data Processing System, 2.125
IBM 1410 computer, 2.120, 2.123, 5.3.27
IBM 1620 computer, 2.68, 2.104, 2.125, 4.1, 4.14, 5.1.26, 5.3.28
IBM 7070 computer, 2.18, 5.3.3, 5.3.38
IBM 7090 computer, 1.51, 2.1, 2.2, 2.11, 2.16, 2.18, 2.22, 2.28, 2.32, 2.36, 2.41, 2.43, 2.58, 2.59, 2.61, 2.62, 2.63, 2.68, 2.76, 2.80, 2.106, 2.119, 2.120, 2.125, 2.130, 3.7, 3.11, 3.13, 3.25, 3.28, 3.36, 3.37, 3.38, 3.40, 3.46, 4.14, 5.1.7, 5.1.20, 5.1.21, 5.2.3, 5.2.22, 5.3.26, 5.3.30, 5.3.34, 5.4.27, 5.5.3, 5.5.10
IBM photostore-7090-console complex, 2.62
IBM 7090-1401 computer complex, 2.58
ICT-Samas punched-card equipment, 2.17
ILAS (Interrelated Logic Accumulating Scanner), 2.88, 2.114
JEIPAC computing system, 2.67
KDC-1 computer, 5.2.12
Kiev computer, 5.4.9
KT-1 computer, 3.23
KT-2 computer, 3.23
Lexical-Graphical Composer-Printer System, design of, 4.17
Linotype tape, 5.3.20
Lodestar, 2.84
Magnacard system, 4.13
Magnascriber for input to Magnacard system, 4.13
Magnavue, combined photographic and magnetic card, 4.13
Magnetic cards containing photographic information, 4.13
Magnetic storage device, 2.94, 4.7, 4.13
Magnetic tape searcher and correlator, 4.7
Marginal punched cards, 2.57
MECIPT-1 computer, 3.1, 5.3.41
Microcite machine, 2.84, 4.19
Microfilm as storage device. *See* Photographic storage device
Microfilm reader, 4.15
Microfilm selector, 4.18
Microimage viewer, 4.20
Minneapolis-Honeywell 800 computer, 1.35
Monotype reader, 4.22, 5.3.31
NEAC 1103 computer, 3.10, 5.2.20
NEAC 2203 computer, 5.1.34
Optical character recognition equipment. *See* Equipment for Pattern Recognition in Part II of this index
PB 250 computer, 5.1.26
PCMI Camera Recorder, prototype model, 4.20
PDP-1 computer, 2.65
Peek-a-boo card system, 2.38, 2.92, 2.112, 2.114, 4.2, 4.19
Philco 2000 computer, 2.106, 5.4.16, 5.4.17, 5.4.24, 5.4.27, 5.4.28; Model 210, 2.125
Phonetic typewriter, 5.2.12
Photocell scanning device for pattern recognition, 2.10
Photographic input device, 4.1, 5.4.7
Photographic storage device, 2.49, 2.66, 3.18, 4.1, 4.3, 4.6, 4.8, 4.10, 4.11, 4.14, 4.15, 4.18, 4.20
Photomemory. *See* Photographic storage device
Photon machine, 1.6, 1.32, 2.54, 4.21
Photoscopic disc memory device, 3.18
Portable speech recognition unit "Shoebox," 5.2.11
Print reader for typed numerals, 5.1.13
Punch device for Microcite peek-a-boo cards, 4.19
Punched-card equipment (not otherwise specified), 1.19, 1.36, 2.8, 2.19, 2.22, 2.34, 2.48, 2.57, 2.59, 2.61, 2.114, 5.5.3, 5.5.10
Punched cards, 1.32
Punched paper tape, 1.35
Rabinow Reader, 5.1.32
RADIR (Random Access Document and Index Retrieval) system, 4.6
Rapid access storage devices. *See* High-speed scanning storage device
Rapid Selector, 2.84, 4.18
RCA 501 computer, 5.2.19, 5.3.25
Reading machine, 5.1.33; experimental, 5.1.8

RECOMP II computer, 5.2.3, 5.4.5
Remote facsimile devices, 2.5
Remington UCT, 2.37
Replica copies, systems for retrieving, state-of-the-art review, 2.84
RPC 4000 computer, 4.23
Scanners and scanning devices for pattern recognition, 5.1.14, 5.1.31
Search Information Tape Equipment (SITE), 4.4
Selecto cards. *See* Peek-a-boo card system
Sequential card-camera devices, 1.35, 2.55
Siemens teleprinters, 2.97
Sonograph, 5.2.8
Special-purpose electronic computer, 5.4.1
Special-purpose scanner, 5.1.5
Spectrograph, 5.2.21
Spectrograph, sound, 5.2.27
Spectrum analyzer, 5.2.10
Speech synthesizer, electric analog, 5.2.15
Stenotype machine, 3.18
Storage devices. *See* High-density storage device, Magnetic storage device, Photographic storage device, *and* Electrochemical storage device
SUPERTYPER 8500 tape typewriter, 2.22
Sylvania 9400 computer, 2.94
Systematics C 750 converter, 2.22
Systematics' Card-to-Tape Converter, 2.82
Systematics' Universal Code Tape-to-Card Converter, 2.82

Table lookup machine, 3.18
Tape, monotype, 2.128
Teletypesetter reader, 5.3.23
Teletypesetter tapes, 5.3.35
Telex Mass Memory Module, 2.94
Telex tape, 5.3.20
Termatrex, 2.8
Thermoplastic recording, 4.5
TOSBAC 4134 computer, 2.67
TX-0 computer, 5.2.15
TX-2 computer, 5.2.14
Typed Page Reader, 5.1.4
Typewriter
  phonetic, 5.2.12
  tape, 2.22, 2.46, 2.55, 2.97, 4.12, 4.23. *See also* Flexowriter

UNIVAC I computer, 2.54, 2.124, 5.3.36
UNIVAC Larc, 2.125
URAL computer, 3.24

Voice coder, 5.2.19
Voice frequency analyzer, 5.2.27

PART II

Equipment for artificial intelligence
  analog-digital machine, 5.4.33
  IBM 704 computer, 5.4.27
  IBM 709 computer, 1.55, 5.4.15
  IBM 7090 computer, 5.4.27
  Kiev computer, 5.4.9
  Philco 2000 computer, 5.4.24, 5.4.27, 5.4.28
  RECOMP II computer, 5.4.5
  special-purpose electronic computer, 5.4.1

Equipment for automatic abstracting and indexing
  Flexowriter tape typewriter, 2.59
  IBM 1401 computer, 2.120
  IBM 1410 computer, 2.120, 2.123
  IBM 7090 computer, 2.59, 2.120
  punched-card equipment (not otherwise specified), 2.59
  UNIVAC I computer, 2.124

Equipment for bibliographic compilation
  GE 225 computer, 1.19
  IBM 1401 computer, 2.80
  IBM 7090 computer, 2.61, 2.80
  punched-card equipment (not otherwise specified), 1.19

Equipment for composition and printing
  comparison study of, 1.24
  Flexowriter tape typewriter, 1.32
  Graphic Arts Composer, 2.86
  IBM 709 computer, 1.32
  Lexical-Graphical Composer-Printer System, design of, 4.17
  Photon machine, 1.6, 1.32, 2.54, 4.21
  teletypesetter reader, 5.3.23

Equipment for index production
  addressing machine equipment, 4.9
  computers (not otherwise specified), 2.55
  Flexowriter tape typewriter, 2.17, 2.104
  GE 225 computer, 1.19

Equipment for index production—Contd.
IBM 082 sorter, 1.7
IBM 407 tabulator, 1.7
IBM 704 computer, 1.7
IBM Cardatype 858 Accounting Machine, 2.67
IBM 870 Document Writer System, 1.7
IBM 1401 computer, 1.7, 1.16, 1.51, 1.54, 2.16, 2.36
IBM 1403 printer, 1.16
IBM 1620 computer, 2.104
IBM 7090 computer, 2.16, 2.36, 2.61
IBM 7090-1401 computer complex, 2.58
ICT-Samas punched-card equipment, 2.17
Photon machine, 1.6, 2.54
punched-card equipment (not otherwise specified) 1.9, 1.36
sequential card-camera devices, 1.35, 2.55
typewriter, tape, 2.55
UNIVAC I computer, 2.54

Equipment for information transmission
remote facsimile devices, 2.5
Siemens teleprinters, 2.97

Equipment for input preparation
Flexowriter tape typewriter, 2.32, 2.37, 2.104
Systematics' Universal Code Tape-to-Card Converter, 2.82
typewriter, tape, 2.46, 2.55, 4.12

Equipment for library operations
IBM 1401 computer, 2.61

Equipment for linguistic research
ALWAC computer, 5.3.35
Beak computer, 5.3.7
CDC 1604 computer, 5.3.39, 5.3.40
CIFA computer, 5.3.41
computers (not otherwise specified), 5.3.14
EDSAC 2 computer, 5.3.31
IBM 650 computer, 5.3.11, 5.3.16, 5.3.27
IBM 709 computer, 5.3.11
IBM 1401 computer, 5.3.38
IBM 1410 computer, 5.3.27
IBM 1620 computer, 5.3.28
IBM 7070 computer, 5.3.3., 5.3.38
IBM 7090 computer, 2.106, 3.40, 5.3.26, 5.3.30, 5.3.34
Linotype tape, 5.3.20

Equipment for linguistic research—Contd.
MECIPT-1 computer, 5.3.41
Monotype reader, 5.3.31
Philco 2000 computer, 2.106
RCA 501 computer, 5.3.25
teletypesetter tapes, 5.3.35
Telex tape, 5.3.20
UNIVAC I computer, 5.3.36

Equipment for mechanical translation
ACE computer, 3.29
AN/GSQ-16, 3.8
AN/GSQ-16 (XW-1), 3.18
EDSAC 2 computer, 3.3
EPOS 1 computer, 3.44
IBM 650 computer, 3.36, 3.46
IBM 704 computer, 3.36, 3.38
IBM 709 computer, 3.40
IBM 1401 computer, 3.8, 3.46
IBM 7090 computer, 3.7, 3.11, 3.13, 3.25, 3.28, 3.36, 3.37, 3.38, 3.46
KT-1 computer, 3.23
KT-2 computer, 3.23
MECIPT-1 computer, 3.1
NEAC 1103 computer, 3.10
photoscopic disc memory device, 3.18
stenotype machine, 3.18
table lookup machine, 3.18
URAL computer, 3.24

Equipment for pattern recognition
automaton, 5.1.42
Burroughs 220 computer, 5.1.5, 5.1.6
Dynamic signal analyzer, 5.4.30
experimental machine, 5.1.14
flying spot scanner, 5.1.5, 5.1.13, 5.1.21
FOSDIC system, 2.43
IBM 514 Reproducing Punch, 5.1.14
IBM 704 computer, 5.4.7
IBM 1620 computer, 5.1.26
IBM 7090 computer, 5.1.7, 5.1.20, 5.1.21, 5.2.3
NEAC 2203 computer, 5.1.34
optical character recognition equipment, 5.1.11, 5.1.42
PB 250 computer, 5.1.26
photocell scanning device, 2.10
Philco 2000 computer, 5.4.16, 5.4.17
Rabinow Reader, 5.1.32
reading machine, 5.1.33; experimental, 5.1.8
RECOMP II computer, 5.2.3
scanners and scanning devices, 5.1.14

Equipment for pattern recognition—Continued
special-purpose scanner, 5.1.5
survey of development of, 5.1.23
Typed Page Reader, 5.1.4
Equipment for searching and selecting
BIMA "P" tape typewriter, 2.21
CDC 1604 computer, 2.113
Command Retrieval Information System (CRIS), 4.10
EDSAC 2 computer, 2.15
FileSearch system, 4.3
Flexowriter tape typewriter, 2.17, 2.22
Gamma 60 computer, 2.18
GE 225 computer, 2.42
IBM 047 tape-to-card converter, 2.22
IBM 108 card proving machine, 2.36
IBM 305 RAMAC computer, 2.114
IBM 650 computer, 2.63, 2.126
IBM 705 computer, 2.23
IBM 709 computer, 2.51, 2.60
IBM 1401 computer, 1.51, 2.6, 2.10, 2.17, 2.19, 2.21, 2.22, 2.27, 2.28, 2.32, 2.37, 2.60, 2.63, 2.65, 2.103, 2.113, 2.114
IBM 1403 printer, 2.19
IBM 1405 Disc Storage unit, 2.125
IBM 1620 computer, 2.104
IBM 7070 computer, 2.18
IBM 7090 computer, 1.51, 2.1, 2.2, 2.11, 2.22, 2.28, 2.32, 2.41, 2.61, 2.62, 2.63, 2.76, 2.125, 2.130, 5.5.3
IBM photostore-7090-console complex 2.62
ILAS (Interrelated Logic Accumulating Scanner), 2.88, 2.114
JEIPAC computing system, 2.67
Magnacard system, 4.13
Magnetic tape searcher and correlator, 4.7
Microcite machine, 4.19
Microfilm selector, 4.18
Minneapolis-Honeywell 800 computer, 1.35
PDP-1 computer, 2.65
Equipment for searching and selecting—Continued
Peek-a-boo card system, 2.92, 2.112, 2.114, 4.2
Photographic storage device, 4.3
punched-card equipment (not otherwise specified), 2.57, 2.114, 5.5.3
RADIR (Random Access Document and Index Retrieval) system, 4.6
Rapid Selector, 4.18
replica copies, systems for retrieving, state-of-the-art review, 2.84
RPC 4000 computer, 4.23
Search Information Tape Equipment (SITE), 4.4
Selecto cards, 4.2
SUPERTYPER 8500 tape typewriter, 2.22
Sylvania 9400 computer, 2.94
Termatrex, 2.8
TOSBAC 4134 computer, 2.67
typewriter, tape, 4.23
Equipment for speech analysis
Automatic Vocal Transaction Analyzer, 5.5.10
Dynamic signal analyzer, 5.4.30
Formant Vocoder, 5.2.10
IBM 7090 computer, 5.2.22, 5.5.10
KDC-1 computer, 5.2.12
NEAC 1103 computer, 5.2.20
portable speech recognition unit "Shoebox," 5.2.11
punched-card equipment (not otherwise specified), 5.5.10
RCA 501 computer, 5.2.19
Sonograph, 5.2.8
spectrograph, 5.2.21
spectrograph, sound, 5.2.27
spectrum analyzer, 5.2.10
speech synthesizer, electric analog, 5.2.15
TX-0 computer, 5.2.15
TX-2 computer, 5.2.14
voice frequency analyzer, 5.2.27

Subject Index

The numbers following the index entries denote the section and article to which the entry refers. The letter *r* refers to the list of references following statements.

Abstract bulletin, mechanized preparation of, 2.104

Abstract Bulletin of the Institute of Paper Chemistry as a test corpus for retrieval, 2.93

Abstract journal, study of, 1.5

Abstracting
- consistency of, 2.4, 2.128
- guidelines, formulation of, 2.4
- principles, development of, 2.4

Abstracting, automatic, 2.4, 2.39, 2.45, 2.62, 2.63, 2.70, 2.94, 2.106, 2.109, 2.120, 3.9, 3.40, 5.3.10, 5.3.29
- compared with conventional abstracting, 2.120
- evaluation, 2.98, 2.120

Abstracting and indexing services
- for mental health literature, study of, 1.23
- review of present system, 1.34

Abstracting service(s)
- in chemical-biological activities, 1.16
- current index to, 2.50

Abstracts
- analysis of preparation, 2.4
- of chemical engineering literature, 1.4
- encoding of. *See* Index entries, selection of
- indexing time for, 2.83
- machine-usable, 2.22, 2.128
- manual, evaluation of, 2.4
- retrieval of, 2.18, 2.23, 2.93, 2.104
- standardization, 2.7
- test for consistency, 2.4
- test of user value, 2.4

Acoustic analysis, 5.2.27

ACSI-MATIC system, 2.94

Adaptive networks. *See* Machine learning

AEC subject headings, compared with ASTIA descriptors, 2.27

Aeronautics as test corpus in study of information systems, 2.9

AIBS Bulletin, permuted title index to, 1.3

Algebra as subject in mechanical translation, 3.43

ALGOL programming system, 5.3.38

Alloys, patent information system for, 2.114

Ambiguity in mechanical translation, 3.1, 3.13, 3.47

Announcement media, 1.24, 2.82, 2.104
- comparative effectiveness of, 1.45

Anthropology, classification scheme for, 2.15

Aquatic sciences, information system for, 2.127

Article, insertion of (MT), 3.36

ASM-SLA classification codes, 2.67

Association between pairs of index terms, 2.29

Association psychology. *See* Psychology, association

Associative clustering, 5.5.7

Associative memory, 2.125, 2.128, 5.3.21. *See also* Memory

ASTIA descriptors
- compared with AEC subject headings, 2.27
- use frequency, 2.115

ASTIA Thesaurus of Descriptors, Mooers revision of, 2.97

Atomic energy
- information system for, 2.44
- as subject in mechanical translation, 3.7

Autoabstracts. *See* Abstracting, automatic

Automata, 2.97, 5.4.33
- reliability of, 5.4.30
- theory, 5.4.10, 5.4.11; survey of, 2.84, 5.4.14

Automatic abstracting. *See* Abstracting, automatic
Automatic documentation. *See* Information retrieval processes, mechanization of, *and* Information systems, mechanized
Automatic indexing. *See* Indexing, automatic
Automatic parsing, 3.3, 3.4, 3.8, 3.13, 3.18, 3.23, 3.25, 3.28, 3.29, 3.46, 3.48*r*, 5.3.25, 5.3.26, 5.3.37, 5.3.42, 5.4.24.
See also Predictive analysis
Automatic programming, 2.80, 3.27, 3.40, 5.4.12
Automatic text analysis, 2.106, 3.13, 3.38, 3.40, 5.3.12, 5.3.14, 5.3.43, 5.4.24, 5.5.3
Automatic translation. *See* Mechanical translation
Automaton, talking, 5.4.26
AUTOTRAN system, 3.7
Axiomatic method, 5.3.17
B.A.S.I.C. (*Biological Abstracts' Subjects in Context*), 1.14
Basic Research Resumés, 1.23
Behavioral factors in information systems, development of tests to measure, 1.1
Behavioral sciences, information system for, 2.49, 2.76
Bibliographic coupling, 2.77; experiments with, 1.5
Bibliographic services, mechanized, outline of problems in, 2.26
Bibliographic sources in physics, analysis of, 2.77
Bibliographie Scientifique Hebdomadaire, 2.17
Bibliographies, mechanized production of, 1.51, 2.61, 2.80, 2.86
Bibliography format standardization, 2.118
Biochemistry
  classification scheme for, 2.132
  information system for, 2.132
Biological evolution, principles of, in optimizing complex systems, 5.1.27
Biological information, dissemination of, 1.3
Biological processes, simulation, 5.4.9
Biology
  information exchange in, 1.3
  information system for, 2.82
  as subject in mechanical translation, 3.38
Bliss classification, study of, 2.122
Boehme, Jakob. keypunched text of works of, 5.3.31
Bracketing (syntactic analysis). *See* Automatic parsing
Cancerology as subject in information retrieval experiment, 3.37
Carburetors, information system for, 2.88
Cardiovascular agents, information system for, 2.55
Cardiovascular diseases, machine code for data on, 2.43
Case history code, 2.56
Catalog card production
  mechanization of, 4.12
  from punched tape, 4.12
Catalogs
  book, machine-preparation of, 2.61
  book and card, cost studies of, 2.99
Cerebrovascular Bibliography, keyword index to, 2.47
Cerebrovascular literature, experimental publication for, 1.25
Character recognition. *See* Pattern recognition *and* Fonts and forms in pattern recognition
Character recognition, state-of-the-art review, 2.84*r*
Character sets, 2.19, 2.86
Chemical Abstracts, 2.8
Chemical-Biological Activities (CBAC), 1.16
Chemical compound structure codes, 2.10, 2.19, 2.53, 2.58, 2.79, 2.101, 2.114, 2.132, 4.23
  decoding to conventional line formula, 2.95
Chemical compounds
  catalog of, 2.114
  indexing of new, 2.58
Chemical engineering
  information system for, 2.116
  thesaurus, 1.4, 1.23, 1.93
Chemical Engineering Progress, 1.23

Chemical Engineering Thesaurus, 1.4, 1.23, 2.93
Chemical formulas, mechanized scanning of, 5.3.21
Chemical information searching, computer programs for, state-of-the-art review, 2.84*r*
Chemical nomenclature, computer translation of, 2.58
Chemical patents, index to, 2.16
Chemical structures, device for encoding, 4.23
Chemical Titles, 1.16
Chemistry
  coding system for, 2.67
  information system for, 1.26, 2.19, 2.37, 2.82, 2.101, 2.114
  publications counting in, 1.20
  as subject in mechanical translation, 3.38
Child Language, 5.3.30, 5.5.9
Chinese characters, 3.38, 3.41. *See also* Fonts and forms in pattern recognition
Citation count
  as index of use, 1.121, 1.58
  as index of value of information, 1.21
Citation index, 1.5, 1.26, 2.17, 2.91
  evaluation experiment, 2.131
  experimental, 2.65
  formats, 2.91, 2.131
  to medical genetics, 1.26
  to molecular chemistry, 1.26
  to physics, 2.131
  to psychopharmacology, 1.26
Citation patterns, 2.62
Citations as aid to determining content, 2.45
Classification
  of abstract objects, 5.1.6
  of documents, automatic, 2.60
  experiments, 2.15
  hierarchical, for mechanized retrieval systems, 2.3
  of languages, 5.3.3
  mathematics of, 2.15, 2.62, 2.106
  of patterns in groups, 5.1.10
  relations in, 2.129
  semantic, 3.37*r*
  studies, 2.100
  techniques, 2.62
Classification—Continued
  of terms, 2.15
  theory of, 2.75
  thesaurus-type, 2.15
Classification system
  construction of, 2.73, 2.75, 2.81
  design of, 2.85
  faceted, 1.41, 2.81
  mathematical derivation of, 2.106
  study or comparison of, 2.9, 2.40, 2.45, 2.81, 2.89, 2.122
Clumps, Theory of. *See* Classification, mathematics of
Clustering of terms, 2.27. *See also* Classification, mathematics of
COBOL programming system, 5.4.32
Code conversion
  Flexowriter tape to Photon film, 1.6
  Justowriter tape to Photon film, 1.6
  of punched tape from different language keyboards, 2.22
Codes
  case history, 2.56
  mnemonic, 2.93
  prime-number, 2.108
  redundancy of, 2.108
Coding systems
  design of forms for, 2.56
  development of, 2.56, 2.57, 2.63, 2.67, 2.85, 2.88, 2.91, 2.114. *See also* Chemical compound structure codes
  study of, 2.119
Coding techniques, 2.117
Cognitive processes
  human. *See* Human learning
  mechanization of. *See* Machine learning
Colon classification system, 2.75, 2.122
COMIT programming system, 2.59, 3.27, 5.4.14, 5.5.3
Command and control systems, 2.24, 5.3.2, 5.3.6, 5.4.13
Command Retrieval Information System (CRIS), 4.10
Communication
  nonwritten scientific, study of, 1.9
  among scientists, 1.3
  technical writing as form of, 1.47
Communication processes
  in humans, 5.5.5
  mathematical model of, 1.38
Comparison of. *See* subject of study, e.g., Indexing systems, comparison of

Compatibility
of indexing vocabularies, 2.99
of materials information retrieval systems, 2.16
of thesauri, 2.93
Compiler, 2.24, 5.4.12
Computational linguistics. *See* Linguistics, computational
Computer abstracts as test corpus in information retrieval experiment, 2.106
Computer music, 5.4.18
Computer program (s)
for abstracting experiments, IBM 1401, 2.120; IBM 1410, 2.120; IBM 7090, 2.120
for indexing experiments, IBM 1401, 2.123; IBM 1410, 2.123
for Keyword in Context, 2.120
for Subject Word Out of Context, 2.120
Computer simulation
in artificial intelligence, 5.4.6, 5.4.8, 5.4.10, 5.4.15, 5.4.17, 5.4.21, 5.4.24, 5.4.27, 5.4.28, 5.4.29, 5.4.31
of behavior in group situations, 5.4.28
of human perception processes, 5.4.29
in information system design, 2.107, 2.119
in music analysis and synthesis, 5.4.18
of organism behavior, 5.4.24
in pattern recognition, 5.1.5, 5.1.13, 5.1.15, 5.1.21, 5.4.17
of psychological association systems, 5.4.10
in speech analysis, 5.2.22
Computer use, human attitudes toward, survey of, 1.55
Computers (Specific computers are indexed by designation, e.g., IBM 650 computer, in Equipment Index)
use of. *See under* specific headings in Equipment Index
Concepts and their relationships, 2.13, 2.27, 2.40, 2.62, 2.64, 2.88, 2.106, 2.114. 2.129, 5.5.3
Concordance-making by computer, 5.3.31
Concordances, 3.33, 3.37, 5.3.12, 5.3.18, 5.3.27, 5.3.29, 5.3.30, 5.3.31, 5.3.35
Conditional reflexes in machines. *See* Machine learning
Conference papers and proceedings, study of, 1.28
Content analysis, 2.18, 2.106, 5.5.10
classification scheme in, 5.5.3
computer method for, 5.5.3
information system for, 5.5.3
Conversion
of data, 2.65
of Flexowriter tape into Photon machine code, 1.32
of punched cards, 2.22
of punched tape, 2.22
Cooperative research in information retrieval, 2.114
Cooperative storage libraries, economic study of, 1.40
Coordinate indexing, 2.8, 2.11, 2.22, 2.36, 2.48, 2.80, 2.92, 2.93, 2.112, 2.114
Correlation, automatic fact, 2.33
Cost studies. *See* subject of study, e.g., Indexing, cost of
Critical incident technique, 1.30
Cultural anthropology, information system for, 2.76
Current awareness media, 2.63
Current Bibliography on Science and Technology (Japanese), 2.67
Current Contents, 1.26
CWIK notation; test of, 2.95
Data exchange, interservice program for, 2.7
Dependency grammar, 5.3.26
Dependency theory (of syntax), 3.3, 3.4, 3.37*r*
Descriptors, 2.51, 2.113
grouping of, 2.129
in subject indexes, 2.7, 2.44
thesaurus of, 2.7
Descriptors-in-context, 1.35
Diabetes literature, information system for, 2.3, 2.128
Diary-work-sample method, 1.30
Dictionary
in content analysis, psycho-sociological, 5.5.3
in information retrieval systems, 2.23, 2.32, 2.44, 2.61, 2.76, 2.94, 2.105, 2.113, 2.114, 2.116, 2.128
in linguistic research, 3.35, 5.3.26, 5.3.29, 5.3.30, 5.3.36; reverse-order, 5.3.11
in mechanical translation, 3.1, 3.4, 3.7, 3.10, 3.11, 3.13, 3.18, 3.21, 3.23, 3.27, 3.28, 3.29, 3.35, 3.37, 3.38, 3.41, 3.43,

Dictionary—Continued
in mechanical translation—Cont'd.
3.44, 3.46; automatic compilation of, 3.13
typology of, 5.3.12*r*
Dictionary entries, segmentation of, in mechanical translation, 3.38
Dictionary items, indexing of, 5.3.5
Dictionary search
in information retrieval, 2.109
in mechanical translation, 3.38
Disease vector control, information system for, 2.128
Display techniques. *See* Output
Dissemination of information, 2.52
automatic, 2.108
automatic selective, 2.32, 2.63
in biology, 1.3
degree of centralization, study, 1.10
in government research reports, 1.29, 2.7, 2.82
selective, by citation index, 1.26
Document classification, automatic, 2.60
Document descriptions, sets of, mathematical properties of, 2.65
Document storage and retrieval system, 2.29, 2.115, 4.6, 4.8
Document storage techniques, 4.20
Documentation, automatic. *See* Information retrieval processes, mechanization of, *and* Information systems, mechanized
Documentation glossary, multilingual, machine preparation of, 2.22
Documentation journals, cumulative index to, 2.22
Documentation language. *See* Language, documentation
Documentation methods and systems. *See* Information systems
Documentation, training in, 1.22, 2.75, 2.76
Documents, automatic analysis of, 2.18

Economic studies (cost studies). *See* subject of study, e.g., Indexing, cost of
Economics, classification scheme for, 2.81
Edge-punched card system, effectiveness compared with peek-a-boo card system, 2.38
Education research literature, information system for, 2.128
Electrical engineering information
coding system for, 2.67
Electrical engineering information—Continued
information system for, 2.116
Electronic circuit diagrams, study of formal properties, 5.4.14
Electronic circuits, information system for, 2.114
Electronic computers, indexing of, 2.88
Electronics as subject in mechanical translation, 3.13, 3.24, 3.44
Electrotechnical literature as subject in mechanical translation, 3.20, 3.21
English usage survey, 5.3.32
Evaluation of. *See* subject of study, e.g., Retrieval effectiveness, evaluation of

Faceted classification systems, 2.75, 2.81, 2.122
Fact correlation. *See* Correlation
File
design, formatted, 2.60
generation, 2.115
organization, 2.115; principles of, 2.1
preparation, system for semiautomatic, 2.114
use, prediction, 5.5.1
FileSearch System, 4.3
Filing errors, prediction of, 5.5.1
Film
document storage and retrieval system, 4.3, 4.6, 4.10, 4.18
output from information system. *See* Output, visual
Filmorex coding, 2.46
FLPL programming system, 5.4.14
Fonts and forms in pattern recognition
alphanumeric symbols, 5.1.15, 5.1.36
characters on paper tape, 5.1.17
Chinese characters, 3.38, 5.1.23
cursive (handwritten) script, 2.78, 5.1.2, 5.1.14, 5.1.34
Cyrillic alphabet, 5.1.33
general study of, 5.1.11
geometric configurations, 5.1.23, 5.1.26, 5.1.35
graphical data, 5.1.7, 5.1.45
hand-drawn alphabet, 5.1.40
handprinted alphanumeric symbols, 5.1.4, 5.1.5, 5.1.13, 5.1.29
handprinted characters, 5.1.26*r*
hatched characters, 5.1.42

Fonts and forms in pattern recognition—Continued
intermixed fonts, 5.1.9, 5.1.32, 5.1.36
machine-printed addresses, 5.1.30
machine-printed alphanumeric symbols, 5.1.5
mixed fonts, 5.1.40
multifont alphanumeric characters, 5.1.15
multifont print, 5.1.16, 5.1.18, 5.1.24, 5.1.36
numerals, 5.1.9, 5.1.11, 5.1.34, 5.1.42
printed characters, 5.1.1, 5.1.34
printed letters, 5.1.7, 5.1.31
printing fonts, conventional, 5.1.9
Russian printing styles, 5.1.25
typescript, 5.1.4, 5.1.8, 5.1.9, 5.1.30, 5.1.31, 5.1.42; Cyrillic, 5.1.13
FORTRAN programming system, 2.1, 2.22, 2.62, 2.130
Frequency, word, 2.44, 2.106, 2.120, 2.124, 5.3.29, 5.3.31
Fused salts, properties of, information system for, 2.8

Games, simulation of, 5.4.2, 5.4.5, 5.4.11, 5.4.23, 5.4.24
GAT method, 3.11, 3.37
Generalized mechanical translation procedure. *See* Mechanical translation, generalized procedures for
Genetics, information system for, 1.26
Glossary. *See* Dictionary
Government research report series
standardization of names of, 1.23
standardization of symbols of, 1.23
Grammar
dependency, 5.3.41
formalized, 3.4, 3.19
Modern English, 5.3.34
Grammatical structure, models of, 5.3.22
Grammatical transformation. *See* Transformation, grammatical
Graph theory, 2.62, 2.65, 2.133
Graphic composition, 4.17, 4.21
Graphic format data, information system for, 2.25, 2.60

Hard-copy output. *See* Output, hard-copy
Harvard Automatic Dictionary, 3.13, 3.29
HAYSTAQ. *See* Patent searching, mechanized system for
Heuristic programming, 5.3.42, 5.4.21
Heuristics, 3.19
Homograph resolution, 3.34
Human behavior
machine behavior compared to, 5.4.27
simulation of, 5.4.15, 5.4.26, 5.4.27, 5.4.34
Human decoding processes, 5.3.10
Human indexers, study of techniques used by, 2.109
Human learning, 5.3.15, 5.4.15, 5.4.29, 5.4.34, 5.4.35, 5.5.2, 5.5.7, 5.5.8, 5.5.9
simulation of, 1.55, 5.4.31
theory of, 5.4.24
Human spine and related structures, information system for, 2.85
Humanities, computer applications in, 5.3.15
Hyphenation, automatic, 5.3.23

Image analysis and correlation techniques, 2.60
Index
bound-volume form of, different formats, 2.80
to chemical engineering literature, 1.4
to chemical patents, 2.16
cumulative, machine preparation of, 1.7, 1.16
to drugs and drug therapy literature, 2.6
to engineering information, 2.35
evaluation methods, study of, 2.50
interdisciplinary science, 2.50
on magnetic tape, 2.23, 2.53
to medical literature, 2.86
to metal cutting literature, 1.52
to molecular formulas, 2.58
multilingual, 2.51
to operations research literature, 1.33
permuted subject, 2.47
to physics literature, 1.5
to plant names, 1.36
requirements, techniques for determining, 2.35
to research projects, 1.35, 1.50; DOD-sponsored, 2.7
separate from document storage, 4.10
storage and searching device, 4.7
subject heading, 1.35
subject-word-out-of-context, 2.120
Index-abstracts, computer retrieval of, 2.55

Index Chemicus, 1.26, 2.58
Index entries
author review of, 1.35
selection of, 2.17, 2.59, 2.63, 2.65, 2.67, 2.109, 2.114, 2.116, 2.118
Index-Handbook of Cardiovascular Agents, study of, 2.99
Index Medicus, 2.85, 2.86
Index production
comparison of methods, 2.47
mechanized, 1.6, 1.7, 1.19, 1.26, 1.35, 1.36, 1.50, 1.51, 2.16, 2.17, 2.54, 2.55, 2.58, 2.67, 2.86, 2.104. *See also* Permuted title index
Indexer qualifications, 2.63
Indexing
based on word-frequency distribution, 2.120
consistency of, 2.54, 2.99; between indexers, 2.60; test of, 2.81
cost of, at different levels, 2.3
depth of, 2.6, 2.23, 2.32, 2.44, 2.59, 2.83, 2.88, 2.92, 2.105, 2.111, 2.112, 2.114
duplication of effort among Government agencies, 2.99
duplication in metal cutting literature, 1.52
efficiency, test of, 2.81
by experienced indexers, 2.54
human factors in, 2.54, 2.81, 2.83
by keywords, 1.19, 1.35, 2.67, 2.104, 2.109
methods, 2.69; study of, 2.9
rules, possibilities of, 2.124
semimechanized, 2.49
services, duplication in, 1.52
standards for chemical engineering, 1.4
statistical approach to, 2.59
by subject specialists, 2.54
syntactic approach to, 2.59
techniques, 2.28
test procedures, 2.9
time, Aslib findings, experimental test on, 2.83
based on word-frequency distribution, 2.120
Indexing, automatic, 1.19, 2.17, 2.22, 2.39, 2.45, 2.59, 2.62, 2.63, 2.67, 2.94, 2.104, 2.106, 2.109, 2.114, 2.124, 5.3.42. *See also* Permuted title index
based on thesaurus techniques, 2.109
of books, 2.99
Indexing, automatic—Continued
compared with conventional indexing, 2.120
cost of, 2.109
evaluation, 2.120
by statistical method, 2.64
of technical texts, 2.64
of teletype news dispatches, 2.109
Indexing language, 2.14, 2.69, 2.115
control of, 2.9
formulation of, 2.28
machine-generated, 2.87
standardized, 2.36
study of, 2.27, 2.47, 2.69
Indexing systems
comparison of, 2.9, 2.35
compatilibity of, 2.99
efficiency of, 1.56, 2.9, 2.89
evaluation of, 2.92
for linguistic literature, development of, 2.51
study of, 1.5
Indexing vocabularies, compatibility of, 2.27, 2.99
Inductive inference
mechanization of, 5.4.22
theory, 5.4.22
Industrial engineering, information system for, 2.116
Information
human management of, 5.5.5
rate of growth, 2.62
Information centers, evaluation methods, 1.10
Information exchange in psychology, pattern of, 1.53
Information flow
quantitative aspects, 2.30
in various organizations, 2.30
Information needs
of agricultural specialists, 1.37
of chemical engineers, 1.4
of chemists, 1.2, 1.30; for Soviet translations, 1.23
of engineers in an electrical firm, 1.56
of Food and Drug Administration, 2.82
of industrial and academic organizations, 1.12
market analysis of, 1.27
of metallurgists, 1.30
of neuropsychiatric research workers, 1.18

Information needs—Continued
of petroleum research scientists, 1.41
of physicians, 1.25, 1.44
of psychologists, 1.9
of scientists, 1.3, 1.11
in a technical organization, 1.42
Information needs and uses, study of, development of methodology for, 1.30
Information network models, 1.10
Information processes, complex, 5.4.21
Information processing
activities of government agency, study of, 2.82
bibliography, preparation of, 2.98
human. *See* Human learning
in the nervous system, 5.4.31
state-of-the-art review, 1.10
Information retrieval
application of linguistic research to, 5.37, 5.40, 5.5.11, 5.5.13
application of list processing techniques to, 2.125
associative, 5.5.2*r*
cooperative research in, 2.125
in law research, 2.130
principles of, 2.111*r*
processes, mechanization of, 2.21
system, trial operation, 2.111
Information services
for agricultural specialists, development of, 1.37
in atomic energy research establishment, 1.48, 1.49
centralized, 1.10, 2.67
to civilian organizations, 2.7
cooperative use of, 2.7
cost factor, 2.128
for educational research materials, 2.128
on information processing, 2.84
for lawyers, 2.5
network, 2.7
for physicians, 2.5
relationship to needs, 1.42
in social sciences, survey of, 1.17
specialized, for biology, 1.14
study of, 1.9, 2.7
subject specialized and geographically regionalized network of, comparison of, 2.52
use patterns, study of, 1.12, 1.48, 1.49
Information sources and services, mechanized register, development of indexing system for, 2.16
Information storage
in humans, 5.5.8
in living organisms, 5.4.35
Information storage and retrieval
bibliography, indexed, 1.19
manual versus automatic methods, comparison of, 2.98
permuted title index to literature of, 1.19
photoelectric technique in, 2.101
theories and models, study of, 2.71
Information systems
adaptive, 2.24
analysis, 1.10
behavioral factors in, study of, 1.1, 5.5.1
comparison of, 2.9
cost factors, 1.10
costs, comparison of, 2.38
criteria for design and operation of, 2.119
criteria for evaluation, 2.96
design, 2.28
efficiency of, 2.9, 2.128
evaluation of, 1.56, 2.32, 2.38, 2.87, 2.114, 2.128
manual versus mechanized, comparison of, 1.56, 2.26
mathematical models of, 1.10
measure of effectiveness, 2.96, 5.5.1
mechanized, 2.11, 2.14, 2.28
models, 2.71, 2.77, 2.119
for specific subject fields. *See* subject in question, e.g., Chemistry, etc.
theory, 2.41
theories of, study of, 2.71
Information technology, research and development, 2.31
Input
decentralized preparation of scientific information, 2.22
graphical, 2.78
keyboard, 4.10
plotter, 2.78
typewritten, 2.78
Integral geometry methods in pattern recognition, 5.1.39
Intelligence
application of information technology to, 2.60, 2.94
extracted from data, 2.42

Intelligence analysis, fact correlation for, 2.33
"Intelligent" behavior in machines. *See* Machine learning
Intelligent systems, organization principles for, 5.4.20
Interest profiles, 2.32, 2.63
Interfix in coding system, use of, 2.88
Interlinguality. *See* Mechanical translation, generalized procedures for
Intermediate language. *See* Language, intermediate
International Abstracts in Operations Research, experimental index to, 1.33
International Political Science Bibliography, 2.81
IPL-IV programming system, 2.1
IPL-V programming system, 2.62, 5.1.6 5.4.14, 5.4.24, 5.4.27, 5.4.28, 5.5.3
Isotopy index, 5.3.3

Journal articles, indexing time for, 2.83
JOVIAL programming system, 5.4.24
Juridical literature. *See* Legal literature

Kant, I., keypunched text of works of, 5.3.29
"Keyword-in-Context" index, 1.16, 1.51, 2.120, 2.123
  compared with subject heading classification system, 2.2
  See also Permuted title index
Keyword index, 2.13, 2.47, 2.64
  permuted, 2.104
  truncated, 2.104
Keyword indexing. *See* Indexing by keywords
Keywords
  in abstract, 2.3
  in author summary, 2.3
  in telegraphic abstract, 2.3
  in text, 2.3
  in title, 2.3
Keywords Index to U. S. Government Technical Reports, 1.45
Koinoglyph, 2.72
Kyle classification, 2.81

Language(s)
  artificial, 2.67, 3.1, 3.40, 5.3.11, 5.4.12, 5.4.13, 5.4.19, 5.4.21, 5.4.32
  automatic coding scheme for, 5.3.16
  in command and control systems, 5.3.2, 5.3.6, 5.4.13
Language(s) —Continued
  documentation, 2.9, 2.18, 2.24, 2.46, 2.59, 2.62, 2.65, 2.67, 2.72, 2.75, 2.76, 2.82, 2.94, 2.97, 2.102, 2.114, 2.133, 5.3.41
  formalized, 3.40
  graphical, 2.78, 5.4.12
  in indexing. *See* Indexing language
  information processing, 2.1, 2.62, 5.1.6, 5.4.14, 5.4.21, 5.4.24, 5.4.27, 5.4.28, 5.5.3
  intermediate, 2.70, 2.72, 2.128, 3.3, 3.17, 3.20, 3.21, 3.24, 5.3.10
  invariant, mechanical translation. *See* Mechanical translation, generalized procedures for
  list processing, comparative study of, 5.4.14
  model(s), 2.59, 2.62, 2.70, 2.102, 3.42, 5.3.12, 5.3.14, 5.3.22; statistico-combinatorial, 5.3.14
  natural, as input to information system, 1.51, 2.24, 2.37, 2.94, 2.103, 2.109, 5.4.15, 5.4.24, 5.5.3; form and function relationship in, 5.3.17
  phrase structure, 5.4.2, 5.4.12, 5.4.15
  poetic, analysis of, 5.3.12*r*
  programming, 2.1, 2.59, 2.62, 2.63, 5.1.6, 5.3.21, 5.3.29, 5.3.38, 5.4.24, 5.4.27, 5.4.32, 5.5.3; theory of, 5.3.29. *See also* Programming systems *and* Language information processing
  semantic structure of, 2.110
  standardized, for information systems. *See* Language, documentation
  symbolic, 3.24
Language-data processing, automatic, 5.3.26
Language statistics. *See* Statistical data on language
Languages in linguistic research
  Cheremis, 5.3.12*r*
  Czech, 5.3.3
  English, 3.20, 5.3.20, 5.3.25, 5.3.26, 5.3.35, 5.3.37, 5.3.40, 5.3.41
  Etruscan, 5.3.39
  French, 5.3.16, 5.3.20, 5.3.28
  German, 5.3.3, 5.3.10, 5.3.28
  Greek, 5.3.28
  Mayan, 5.3.15
  Russian, 5.3.3, 5.3.26, 5.3.28
  Spanish, 5.3.28
  Swedish, 5.3.7

Languages in mechanical translation
Arabic, 3.24
Arabic-English, 3.27
Bengali, 3.24
Bulgarian, 3.24
Burmese, 3.24
Chinese, 3.24, 3.30
Chinese-English, 3.18, 3.34, 3.38, 3.41, 3.47
Chinese-Spanish, 3.38
Czech, 3.24
Dutch, 3.6
English, 3.3, 3.15, 3.19, 3.23, 3.24, 3.30, 3.37
English-Czech, 3.20, 3.21, 3.44
English-Danish, 3.7
English-Dutch, 3.7
English-French, 3.7, 3.48
English-German, 3.7
English-Italian, 3.7
English-Japanese, 3.9
English-Latin, 3.36
English-Norwegian, 3.7
English-Portuguese, 3.7
English-Rumanian, 3.1
English-Russian, 3.25
English-Spanish, 3.7, 3.35, 3.38, 5.4.5
English-Swedish, 3.7
English-Turkish, 3.11
Estonian, 3.24
Finnish, 3.24
Finnish-Swedish, 3.22
French, 3.4, 3.6, 3.19, 3.24
French-English, 3.8
Georgian, 3.15
German, 3.4, 3.6, 3.12, 3.23, 3.24, 3.36
German-English, 3.27, 3.40
German-French, 3.5
German-Russian, 3.17
Hausa, 3.24
Hindi, 3.24
Indonesian, 3.24
Italian, 3.6, 3.24, 3.36
Japanese, 3.23, 3.24
Japanese-English, 3.45, 5.3.43
Japanese-French, 3.5
Kirghiz, 3.24
Korean, 3.24
Lettish, 3.24
Mongol, 3.24
Polish, 3.24
Rumanian, 3.24
Languages in mechanical translation—Continued
Russian, 3.4, 3.15, 3.31
Russian-English, 3.7, 3.11, 3.15, 3.18, 3.24, 3.25, 3.28, 3.29, 3.34, 3.36, 3.38, 3.46, 5.4.5
Russian-French, 3.5, 3.37
Russian-Georgian, 3.15
Russian-Lithuanian, 3.45
Russian-Spanish, 3.38
Spanish, 3.12, 3.24
Spanish-English, 3.35
Swahili, 3.24
Swedish, 3.24
Tamil, 3.24
Turkish, 3.24
Vietnamese, 3.24
Languages in speech synthesis, Arabic, 5.2.28
Lattice structure of terms. *See* Classification of terms
Law literature. *See* Legal literature
Law of precedent, mathematical theory of, 2.130
Learning machine, simulation, 5.4.5
Learning-matrix technique
for machine learning, 5.4.25
in pattern recognition, 5.1.42
Legal decision-making, computer aid to, 2.130
Legal literature, information system for, 2.27, 2.72, 2.92, 2.103, 2.123, 2.126, 2.128; mechanization of, 2.2, 2.5
Lexicography, 3.11, 3.41, 5.3.18, 5.3.29, 5.3.31
computers in, 5.3.10
mechanization of, 5.3.5
Libraries, evaluation of cooperative storage, 1.40
Library
automatic stratification of, 2.125
automation survey, 2.74
biomedical, statistical data on activity of, 2.1
concepts and problems, 2.12
operations, cost study of, 2.68; mechanization of, 2.14, 2.22, 2.32, 2.61, 2.65, 2.68, 2.74, 2.121, 2.131; mechanized, test and evaluation of, 2.121
operations and problems, 1.40, 2.14, 2.68
use study, 2.68

Life sciences, serial publications in, survey of, 1.3
Linguistic analysis of
Arabic, 3.27
Bengali, 5.3.14
Bulgarian, 5.3.14
Burmese, 5.3.14
Cheremis, 5.3.12*r*
Chinanteco, 5.3.27
Chinese, 3.18, 3.30, 3.38, 3.47
Czech, 5.3.14
Dutch, 3.6
English, 3.12, 3.13, 3.18, 3.20, 3.27, 3.30, 3.38, 3.40, 3.42, 3.48, 5.2.10, 5.3.2, 5.3.6, 5.3.11, 5.3.14, 5.3.19, 5.3.26, 5.3.32, 5.3.33, 5.3.34, 5.3.35, 5.3.36, 5.3.37
Finnish, 3.22
French, 3.4, 3.5, 3.6, 3.18, 3.27, 3.42, 5.3.14, 5.3.16, 5.3.18
Georgian, 3.15
German, 3.4, 3.5, 3.6, 3.12, 3.27, 3.40, 5.3.10, 5.3.14, 5.3.29
Hausa, 5.3.14
Hindi, 5.3.14, 5.3.30
Huichol, 5.3.27
Indonesian, 5.3.14
Italian, 3.6, 5.3.14, 5.3.31
Japanese, 3.5, 5.3.24, 5.3.43
Korean, 5.3.14
Medieval German, 5.3.31
Mixteco, 5.3.27
Polish, 3.42
Russian, 3.4, 3.5, 3.15, 3.16, 3.18, 3.27, 3.38, 3.42, 5.3.14, 5.3.26
Serbo-Croatian, 3.14
Sierra Nahuat, 5.3.27
Swahili, 5.3.14
Swedish, 3.22, 5.3.7, 5.3.23
Ukrainian, 5.3.14
Vietnamese, 5.3.14
Linguistic analysis, computers in, 3.2, 3.27, 3.34, 3.38, 3.40, 5.3.3, 5.3.4, 5.3.7, 5.3.11, 5.3.12*r*, 5.3.14, 5.3.15, 5.3.16, 5.3.20, 5.3.29, 5.3.30, 5.3.31, 5.3.35, 5.3.37, 5.3.39, 5.3.40, 5.3.41, 5.3.43
Linguistic methodology, development of, 2.110
Linguistic research, 2.17, 2.62, 2.65, 2.94, 2.106, 2.110, 2.114, 2.133, 3.4, 3.5, 3.12, 3.21, 3.25, 3.31, 3.35, 3.37, 3.40, 3.42,
Linguistic research—Continued
3.46, 5.3.2, 5.3.4, 5.3.5, 5.3.6, 5.3.7, 5.3.8, 5.3.9, 5.3.10, 5.3.11, 5.3.12, 5.3.14, 5.3.15, 5.3.19, 5.3.20, 5.3.24, 5.3.25, 5.3.26, 5.3.28, 5.3.31, 5.3.33, 5.3.34, 5.3.35, 5.3.36, 5.3.37, 5.3.39, 5.3.40, 5.3.41, 5.3.43, 5.4.13, 5.4.14, 5.5.3, 5.5.9
axiomatic method in, 5.3.17
study of trends in, 5.3.13
See also Semantics
Linguistic theory, models in, 5.3.14
Linguistic transformation. *See* Transformation, grammatical
Linguistics
classification scheme for, 2.51
computational, 3.38, 5.3.26, 5.3.30, 5.3.37, 5.3.41
information system for, 2.51
mathematical, 3.1, 3.13, 3.27, 3.42, 5.3.3, 5.3.14, 5.3.29
structural, 5.3.13
theoretical, 5.3.9, 5.3.17, 5.3.22, 5.3.34
LISP programming system, 5.4.14
List processing, 2.125, 5.4.21
Literature searching
adequacy of, 1.11
mechanized, 2.32

Machine abstracting. *See* Abstracting, automatic
Machine code system, development of, 2.43
Machine indexing. *See* Indexing, automatic
Machine language. *See* Language, artificial
Machine learning, 1.55, 5.1.12, 5.1.19, 5.1.37, 5.1.40, 5.1.45, 5.3.29, 5.3.42, 5.4.1, 5.4.2, 5.4.3, 5.4.4, 5.4.5, 5.4.7, 5.4.9, 5.4.10, 5.4.11, 5.4.12, 5.4.22, 5.4.23, 5.4.24, 5.4.25, 5.4.26, 5.4.27, 5.4.29, 5.4.30
Machine translation. *See* Mechanical translation
Magnetic cards. *See under* Equipment Index
Magnetic tape storage. *See under* Equipment Index
Magnetic storage device. *See under* Equipment Index

Man-machine communication, 1.1, 1.43, 1.55, 2.12, 2.62, 2.77, 2.78, 2.94, 2.106, 5.2.14, 5.2.20, 5.3.19, 5.4.12, 5.4.13, 5.4.24
Materials, electrical and electronic properties of, information system for, 2.48
Materials data, information system for, 2.11, 2.16
Mathematical linguistics. *See* Linguistics, mathematical
Mathematical model. *See* Models
Mathematical Reviews, 1.6
Mathematics as subject in mechanical translation, 3.13, 3.17, 3.46
Maya manuscripts, 5.3.15
Mechanical engineering, information system for, 2.116
Mechanical translation, 3.2, 5.1.45, 5.3.5, 5.3.7, 5.3.10, 5.3.42, 5.3.43, 5.4.5
  evaluation of output quality, 3.32
  generalized procedures for, 3.3, 3.7, 3.10, 3.37, 3.38
  output, automatic detection of errors in, 3.34
  proposed system, 3.18, 3.19
  simulated, 3.22
Mechanization of thought. *See* Machine learning
Mechanized indexing. *See* Indexing, automatic
Medical data
  coding system for, 2.56
  information system for, 2.43, 2.107; mechanization of, 2.20
  standardization of, 2.107
Medical doctoral theses, citation use of, 1.58
Medical literature, index to, 2.86
Medical periodicals, citation use of, 1.58
Medicine
  information system for, 2.5, 2.13, 2.20, 2.37, 2.86
  news media, 1.23
  as subject in mechanical translation, 3.7
Memory
  association in, 1.55
  associative, 2.128
  human. *See* Human learning
  organization, 1.55
Metal cutting literature, indexing in, 1.52
Metalanguage. *See* Language, symbolic, *and* Language, intermediate
Metallurgical engineering, information system for, 2.116
Metallurgy, information system for evaluation of, 1.15, 2.11
  user satisfaction, 1.15
Metallurgy and ceramics, publications counting in, 1.20
Metals literature
  ASM-SLA classification scheme, 2.67
  information system for, 2.67
Meteorological and Geoastrophysical Abstracts, 1.7
Meteorological and Geoastrophysical Titles, 1.7
Microcards, publication on, 1.57
Microfilm, indexed, 4.13
Microfilm
  design, 1.24
  publication, 1.57
  unitized film records, automatic techniques for handling, 2.66
Microimage
  matrix for microcite, 4.19
  photochromic, 4.20
  reduction ratio, 4.19, 4.20
  techniques, 4.20
Microphotographic processes in storage and retrieval. *See* Photographic storage devices *in* Equipment Index
Microphotography, high-reduction, study of, 4.11
Mnemonics in classification, 2.75
Models
  of behavior, information processing, 5.4.27
  of cognitive processes, 5.4.31
  of human behavior, 5.4.27
  of information retrieval processes, 2.117
  of information systems, 2.1, 2.71, 2.102, 2.119
  of learning, 5.4.31
  linguistic, 3.40
  of pattern recognition, 5.1.44, 5.1.45
  of retrieval systems, 2.45
  two-choice behavior, 5.4.27
  verbal learning, 5.4.27
Monthly Index of Russian Accessions, 1.54
  user reactions to, 1.31
  user study of, 2.30

Multiple Instantaneous Response File (MIRF), experimental model, 2.105
Multiple-list system. *See* List processing
Multiple-path analysis. *See* Automatic parsing
Music, computer analysis and synthesis, 5.4.18, 5.4.19
Nachrichten für Dokumentation, cumulative index to, machine preparation, 2.22
Natural history collection, information system for, 1.8
Near-synonyms, semantic grouping of sets of, 2.15
Nerve-net theory, 5.4.10
Networks, information retrieval, requirements of types of, 2.69
Neural network, simulation of, 5.4.31
Neurons, network of artificial. *See* Perceptron devices
Northeastern Reporter as material for indexing experiment, 2.2
Notation. *See* Coding systems
Notation systems for chemical compounds. *See* Chemical compound structure codes
Nuclear energy literature as test corpus in automatic indexing experiment, [redacted]
Nuclear [redacted] *Abstracts*, 1.50, 2.83
as test corpus in automatic indexing, 2.22
Organic ch[redacted] 2.58
information system for, 2.19
See also Chemistry
Ortho-language. *See* Language, symbolic
Orthonormal function analysis in speech recognition, 5.2.22
Output
abstracts, 2.116
film aperture card, 4.10
hard-copy, 2.36, 4.3, 4.6, 4.10, 4.19
titles, 2.116
visual, 2.36, 2.65, 2.66, 4.3, 4.6, 4.10, 4.13, 4.19
visual display of graphs, 5.1.13
Oxford Decimal Classification for Forestry, 2.89
Parallel text processing, 3.33
Parsing. *See* Automatic parsing.
Patent claims, mathematical theory of, 2.130
Patent Gazette, 2.16
Patent information, use pattern in chemical industry, 2.53
Patent searching, mechanized system for, 2.82, 2.88, 2.114
Patents
information system for, 2.22, 2.53, 2.104, 2.114
as test corpus for retrieval experiment, 2.23
Pattern recognition, 5.1.12, 5.4.9
adaptive, 5.1.37
of aerial photographs, 5.1.37
of audio signals, 5.4.7
classes of objects, 5.1.6
classification in, 5.1.10, 5.1.26; statistical, 5.1.22
cascaded decision functions, 5.4.16
conditioned-reflex theory, 5.1.37
correlation techniques, 5.1.1, 5.1.34
decision-making portion of, 5.1.20
of geometrical solids, 5.1.35
Gestalt method, 5.1.1
of graphical data, 5.1.7
high-speed scanning technique, 5.1.9
learning process in, 5.1.40, 5.1.42, 5.1.45
mathematical theory of, 5.1.22
normalization problem in, 5.1.8
n-tuple method, 5.1.27
optical methods, 5.1.7, 5.1.9, 5.1.11, 5.1.18, 5.1.33
photographs, 5.1.28, 5.1.45, 5.4.4
in photo interpretation, 5.4.17
of real objects, 5.1.35
system model, 5.1.43
techniques, 5.1.10, 5.1.20, 5.1.24, 5.1.26, 5.1.29, 5.1.35, 5.1.39, 5.1.40, 5.1.41, 5.4.7, 5.4.17, 5.4.25, 5.4.30; applied to language translation, 5.1.45
of textual material, 5.1.7
theory of, 5.1.8, 5.1.19, 5.1.44, 5.1.45
of three-dimensional forms, 5.1.35
of visual forms, 5.1.37, 5.4.16
weighted-area scanning technique, 5.1.21
Peek-a-boo card system, effectiveness compared with edge-punched card system, 2.38

Perception
form, 5.1.45
human, 5.4.29, 5.5.2
of speech. *See* Speech recognition
visual form, 5.1.3
Perceptron devices, 5.4.4, 5.4.7, 5.4.8
Perceptron theory, 5.4.4, 5.4.7, 5.4.8
Permuted subject index. *See* Index, permuted subject
Permuted title index, 1.3, 1.7, 1.14, 1.16, 1.19, 1.45, 1.51, 2.7, 2.17, 2.82
AIBS Bulletin, 1.3
B.A.S.I.C., 1.14
Chemical Titles, 1.16
Meteorological and Geoastrophysical Titles, 1.7
to Slavic journals, 1.54
of transliterated Slavic titles, 1.54
Pesticides, patent information system for, 2.114
Petroleum technology, faceted classification scheme for, 1.41
Pharmacology
classification scheme for, 2.75
information system for, 2.37
Phonemic transcription, 5.3.3
Phonetic characterization of Mandarin Chinese, 3.30
Phonetics, experimental, 5.3.10
Phonological component, 5.3.3
Phonology, 5.3.3; Mandarin Chinese, 3.30
Photocomposition, 2.54, 4.21
automatic justification in, 4.17
automatic hyphenation in, 4.17
experimental, 1.6, 1.32
Photographic copying
economics of, 1.39
effects on publication sales, 1.39
Photographic interpretation, 5.1.3
with perceptron device, 5.4.7
Photographs, aerial, pattern recognition of, 5.1.28
Phrase structure grammars, 5.3.9
Physics, publications counting in, 1.20
Physics Abstracts, 1.5
Physics articles, English-language, citation in foreign-language physics journals, 2.131
Physics journals, foreign-language, 2.131
Physics literature
classification scheme for, 1.5
Physics literature—Continued
guides to, 1.5
index to, 1.5
Physiology, information system for, 2.18, 2.76
Pictures and diagrams as input to information system, 5.4.14
Plant names, information system for, 1.56
Plant pathology, information system for, 2.38
Plant protection, information system for, 2.38
Plants, utilization, information system for, 2.112
Political science, classification scheme for, 2.81
Potential field method in pattern recognition, 5.1.42
Predictive analysis, 3.13, 3.28. *See also* Automatic parsing
Pre-reports information, 2.7
Printing, computer-controlled, 1.32, 2.65, 5.3.23. *See also* Photocomposition
Problem solving
by humans, 1.43, 5.4.15, 5.4.24, 5.4.31, 5.5.2, 5.5.4, 5.5.6, 5.5.7
by machines, 1.55, 5.4.12, 5.4.21, 5.4.24, 5.4.27
original behavior in, 5.5.6
Program Evaluation and Review Technique (PERT), 2.7
Programming language. *See* Language, programming
Programming systems
ALGOL, 5.4.32, 5.3.38
COBOL, 5.4.32
COMIT, 2.59, 3.27, 5.3.38, 5.4.14, 5.5.3
FLPL, 5.4.14
FORTRAN, 2.1, 2.62, 2.190, 5.4.32
IPL-IV, 2.1
IPL-V, 2.62, 5.1.6, 5.4.14, 5.4.24, 5.4.27, 5.4.28, 5.5.3
JOVIAL, 5.4.24
study of, 5.3.28
Projectivity hypothesis, 3.37r
Proto-synthex (language processing system), 5.4.24
Psycholinguistics, 5.3.12, 5.5.9. *See also* Linguistic research *and* Semantics
Psychological abstracts as test corpus in information retrieval experiment, 2.108

Psychology
association, applied to machine learning, 5.4.10
information exchange in, 1.9; pattern of, 1.53
information system for, 2.76
of mental processes, 5.4.21, 5.5.7
Psychopharmacology, information system for, 2.55
Psychopharmacology literature, characteristics of, 2.55
Publication
of conference papers, 1.2
experiments, 1.2, 1.4, 1.25, 1.35, 1.50, 1.57
practices of chemists, 1.2
punchcards, 4.2
studies, 1.5, 1.26
study on life of, 2.38
Publications counting, 1.2, 1.20, 1.58
Punched-card equipment, indexing of, 2.88
Punched-card system, cost study of, 1.8
Punched cards, conversion of, 2.22
Punched tape, exchange of, 2.22
See also Text, machine-usable
Punctuation (MT), 3.36, 3.37*r*

Radio-electronics, information system for, 2.72
Reader reactions. *See* User reactions
Real-time computation, 2.78
Redundancy in communication, 2.118
Reference questions, analysis of, 2.118
Relevance of a document, probabilistic determination of, 2.109
Relevancy
of article titles. *See* Titles, relevancy of
of document to request, 2.9
Reliability data, information system for, 2.42
Report literature, information system for, 2.82
Report summaries, effect of format restriction on, 2.106
Reports collection, automated information system for, 2.14
Reprints, use of, 1.18
Requests, verbalization, 2.118
Research Grants Index, 1.35
Research projects
index, 1.35, 2.7
information system for, 1.35, 1.50
Research reports, government
announcement of, 1.29
availability of, 1.29, 2.7
dissemination, 2.7
information system for, 2.82, 2.99
regional repositories, 2.7
Research support, identification of sources of, 1.20
Resuscitators, patent information system for, 2.114
Retrieval
of abstracts, 2.18, 2.23, 2.93, 2.104
of documents, 2.22
Retrieval effectiveness
evaluation of, 2.23, 2.28, 2.109, 2.128
at various indexing levels, 2.3
statistical analysis of, 2.93
Retrieval modes, comparative efficiency, study of, 2.99
Retrieval patterns, 2.60
Retrieval system (s)
classification scheme for, 2.108
evaluation of. *See* Information systems, evaluation of
punched-card, 2.124
reference, 1.5, 2.59
state of the art, 1.56
theory, 2.69, 2.71, 2.108
See also Information storage and retrieval *and* Information systems
Retrieval time, study of, 2.38
Reverse-sort wordlist for linguistic research, 5.3.36
Role indicators, 2.14
Roles and links, evaluation of, 1.23
Rotaform Index ®, 2.58
Rotation criteria in pattern recognition, 5.1.43

Scanning techniques for pattern recognition, 5.1.9
Science news media, survey of, 1.27
Screen, viewing. *See* Output, visual
Search (ing)
full-text, by computer, 2.39
processes, man-automata combinations, 5.4.24
questions, analysis of, 2.35
request, translation into machine language, 2.24
strategies, 2.59, 2.62, 2.69, 2.93, 2.128, 5.4.24, 5.5.4; study of, 2.1

Search (ing) —Continued
techniques, 2.26, 2.27, 2.28, 2.58, 2.62, 2.82, 2.92, 2.114, 2.115, 4.3, 5.4.31; manual versus machine, 2.114; study of, 2.9, 2.128
SECIR Project (information system), 2.114
Self-organizing systems, 2.24, 2.62, 2.114, 3.40, 5.1.43r, 5.4.3, 5.4.4r 5.4.5, 5.4.6, 5.4.15, 5.4.21, 5.4.24, 5.4.25, 5.4.30
models, testing and analysis of, 2.1
Semantic grouping of sets of near-synonyms, 2.15
Semantics, 2.72, 2.98, 2.106, 2.110, 2.133, 3.3, 3.10, 3.12, 3.14, 3.17, 3.23, 3.27, 3.34, 3.36, 3.37, 3.38, 5.3.10, 5.3.11, 5.3.13, 5.3.14, 5.3.26, 5.3.29, 5.3.42, 5.5.9. *See also* Linguistic research *and* Semology
Seminar in Mechanolinguistics, 5.3.30
Semoglyph, 2.72
Semology, 3.38
Sign systems, theory of, 3.42r
Significant word index in context. *See* Permuted title index
Simulation
of file activity, 2.1
of human cognitive processes. *See* Human learning *and* Machine learning
of work environment, 1.30
Small groups, methods of studying, 5.5.3
Social anthropology, classification scheme for, 2.81
Social sciences
classification scheme for, 2.81
information system for, 2.49, 2.76, 2.118
Sociological data, computer processing of, 2.18
Sociology,
classification scheme for, 2.81
information system for, 2.46, 2.76
Soviet literature
use of translated, 1.25
use of translations of, 1.46
Space law, subject classification analysis, 2.123
Speaker identification, 5.2.10
Spectra, 2.101
availability, 2.101
information system for, 2.101
Spectrographic analysis in speech recognition, 5.2.10
Spectroscopic data. *See* Spectra
Speech
acoustic data, 5.2.3
acoustical signals, 5.2.2
compression, 5.2.10, 5.2.21
data translator, 5.2.20
invariant features of, 5.2.18
phonetic transcription of, 5.2.13
phonological component, 3.30
pitch perception, 5.2.2
Speech analysis, 5.2.2, 5.2.4, 5.2.6, 5.2.7, 5.2.9, 5.2.10, 5.2.13, 5.2.14, 5.2.15, 5.2.17, 5.2.18, 5.2.19, 5.2.20, 5.2.21, 5.2.22, 5.2.23, 5.2.25 5.2.26, 5.2.27, 5.3.10, 5.4.30, 5.5.10
automatic segmentation programs, 5.2.19
of German, 5.2.24
identification of English consonant phonemes, 5.2.8
of Japanese, 5.2.12
of Swedish, 5.2.21
Speech production
acoustics of, 5.2.15
human, physiology of, 5.2.27
Speech recognition, 5.1.10, 5.1.45, 5.2.3, 5.2.4, 5.2.6, 5.2.9, 5.2.10, 5.2.11, 5.2.12, 5.2.14, 5.2.15, 5.2.17, 5.2.18, 5.2.21, 5.2.22, 5.2.23, 5.2.24, 5.2.25, 5.2.26 5.2.27, 5.2.28, 5.4.30
computer programs for, 5.2.1
criteria, study of, 5.2.19
of Japanese, 5.2.20
Speech synthesis, 5.2.10, 5.2.13, 5.2.15, 5.2.20, 5.2.21, 5.2.27
digital-to-voice conversion, 5.2.16
State-of-the-art reviews, preparation of, 2.84
Statistical data on language, 2.44, 2.62, 3.13, 3.21, 3.41, 3.43, 3.44, 5.3.3, 5.3.7, 5.3.10, 5.3.16, 5.3.18, 5.3.24, 5.3.29, 5.3.31, 5.3.35, 5.3.39, 5.3.40, 5.4.18, 5.5.3
Statistical properties of text, 2.106, 5.3.16
Statistics, citation index to, 2.91
Statutes, health. *See* Legal literature
Storage, image versus digital, analysis of, 2.30
Storage media, thermoplastic recording technique, 4.5
String transformations, theory of, 5.3.21
Structural formulas, code for. *See* Chemical compound structure codes
Stylistics in French, 5.3.30

Subject analysis, 1.82; depth of, 2.3
Subject headings, assignment, consistency of, 1.7, 2.3, 2.27, 2.99; comparison with descriptors, 2.27
Subjects in mechanical translation
algebra, 3.43
atomic energy, 3.7
biological sciences, 3.11
biology, 3.38
chemistry, 3.11, 3.38
cybernetics, 3.11
electronics, 3.11, 3.24, 3.44
electrotechnical literature, 3.21
mathematics, 3.13, 3.17, 3.46
medicine, 3.7
physics, 3.11
physics, high energy, 3.37
social sciences, 3.11
Survey research, 1.12, 1.56
on abstracts, preparation of, 2.4
by diary collection, 1.11, 1.25
on human attitudes toward computer use, 1.55
by interview, 1.3, 1.25, 1.31, 1.55
by questionnaire, 1.3, 1.11, 1.12, 1.17, 1.25, 1.31, 1.46, 1.55, 2.68
by telephone, 1.31
Symbol manipulation, 5.4.12
Synonymy, problems of, 2.118
Syntactic analysis. *See* Automatic text analysis *and* Automatic parsing
Syntactic complexity, formal notion of, 5.3.9
Syntax Analysis Program, 5.3.25
Synthesis, automatic
of English, 3.7, 3.18
of French, 3.3
of Lithuanian, 3.43
SYNTOL (Syntagmatic Organization Language), 2.18, 2.76
Systems design, behavioral factors influencing, 1.1

Talking automaton. *See* Automaton, talking
Tape
magnetic, interchange of, 2.7
punched, conversion of, 2.22
Tape Typewriter, plan, 2.22, 2.97
Tape typewriter, use
in documentation, study of, 2.22
evaluation of, 1.23
Technical Abstract Bulletin, 2.8
in-house printing of, 2.7
permuted title index to, 2.7
Technical articles, in machine-usable form, 2.39
Technical report writing
effectiveness of, 1.47
review of research on, 1.47
Test facility for information storage and retrieval systems, 2.98
Test methods in study of information systems, evaluation of, 2.9
Text
analysis, 2.133
computer processing of, 2.39, 2.124
machine-usable, 2.13, 2.22, 2.39, 2.76 2.103, 2.104, 2.106, 2.120, 2.124, 2.128, 3.11, 3.27, 3.40, 4.22, 5.3.20, 5.3.29, 5.3.31, 5.4.24; conversion to, 2.55; teletypesetter tape, 5.3.23, 5.3.26
natural-language, 2.45, 2.70; as input to information system, 2.39, 2.109
statistical properties of, 2.62, 2.106
Textual data, unformatted, techniques for exploiting, 2.90
Theorem proving, 5.4.24
Theory of Clumps. *See* Classification, mathematics of
Thermonuclear reactions, information system for, 2.17. *See also* Atomic energy
Thesaurus, 1.4, 1.16, 2.8, 2.14, 2.28, 2.44, 2.54, 2.85, 2.94, 2.97, 2.103 2.109, 2.110, 3.3, 5.3.12
compatibility, 2.93
for engineering, compilation of, 2.34
evaluation of, 1.23, 2.7
of forestry terminology, 2.93
as indexing guide, 1.35
for legal literature, 2.2
model, 3.3
for psychopharmacology, test of, 2.55
of pulp and paper terminology, 2.93
Thesaurus of ASTIA Descriptors, 2.7
Thought, analysis of, 5.4.26
Titles
of publications, retrieval usefulness, 2.100
relevancy of, 2.9, 2.63, 2.109
Title-wording studies, 2.100
Training of science information specialists, 1.3, 1.22

Transformation, grammatical, 3.4, 5.3.37, 5.3.41
Transistor circuits, information system for, 2.114
Translation
automatic. *See* Mechanical translation
of periodicals, cover-to-cover, 1.46
problems, analysis of, 5.3.1
theory of, 3.27, 3.39, 3.40
user evaluation of, 1.25
Transmission of information, 2.5
Transmission networks in an information service, 2.7
Turing machine, 2.97
Type-composition, photographic. *See* Photocomposition
Typesetting, computer-controlled. *See* Printing, computer-controlled
Typographic composition, mechanized, user need for, 1.24

Uniterm Index to U.S. Chemical Patents on magnetic tape, 2.53
Uniterms, 2.115
Universal Decimal Classification (UDC), 2.89
study of, 1.7, 2.81
U.S. Government Research Reports, 1.45, 2.8
permuted title index to, 2.82
Use
of information, by physicians, 1.25; by psychologists, 1.9
of information services in industry, 1.41
of literature by scientists, 1.3
of medical library books and journals, 1.58
of *Monthly Index of Russian Accessions*, 2.30
of Soviet biological translations, 1.3
of Soviet literature by American scientists, 1.3
Use—Continued
of Soviet translations by American chemists, 1.23
study, library and information services, 1.12
of technical information, 2.30
of translated Soviet literature, 1.25, 1.46
of world chemical literature by Russians, 1.21
Use patterns, 2.118
study of, 1.30
Usefulness of microfilmed museum catalogs, 1.8
User behavior as experimental criterion, 1.25
User needs, interview research, unstructured, 1.18
User reactions
to *Monthly Index of Russian Accessions*, 1.31
to permuted title index, 2.17
study of, 1.2, 1.15, 1.45
User study methodology, development of, 1.25

Vocabulary studies, 1.14, 2.13, 2.27, 2.28, 2.34, 2.64
Voice communications, reliability, 5.2.2
Voice input to computers, 5.2.14, 5.3.10
Wildlife Disease, microform publication, 1.57
Word
associations, 2.27, 2.106, 2.120, 2.124, 2.128; analysis of, 5.5.8
clustering, 2.64
frequencies. *See* Frequency, word
indexes, 5.3.20, 5.3.31
length, 3.1
order, rearrangement of, 3.1, 3.10, 3.28, 3.29
segmentation, 3.47
usage, 5.3.24

Xerography, 2.37