1. **The Survey Corpus**

The Survey Corpus is available in slip form and on computers at the Survey premises. The computerized corpus containing 200 spoken and written texts totalling one million words is supplied with two interactive concordance browsers: KAYE II and WordCruncher. All the 100 Survey spoken texts may be obtained in machine-readable form from the Norwegian Computing Centre for the Humanities, but the written texts can be accessed only at the Survey premises.

The Survey Corpus has been marked up for automatic grammatical notation, but further processing will depend on future funding.

2. **International Corpus of English**

Professor Greenbaum is coordinating the ICE project internationally, and the Survey is giving advice and technical assistance to ICE research teams in other countries. We are also devising software for a Corpus Utility that will be supplied to ICE participants. Eleven ICE Newsletters and one Supplement have so far been issued, with discussion documents and advisory material.

In October 1990 the Survey began to compile the British component of ICE. Our million-word corpus will comprise 500 texts of 2000 words each: 300 spoken texts (50 of them scripted) and 200 written texts (50 of them manuscript). At the time of writing this report, we have assembled 434 texts, for 348 of which we have obtained copyright permission for use in academic research. Most of the printed material was converted into machine-readable form by the use of an optical scanner at the University of London Computer Centre and was then proofread and corrected on the computer screens at the Survey. All our manuscript and spoken texts are keypunched at the Survey. We now have on computer 74 spoken texts and 198 written texts.

We have collaborated with the TOSCA group at the University of Nijmegen on the development of an ICE word tagset for the automatic tagging of each word in the corpus. The tagset has now been established and a manual has been written for those working on checking the
output of the tagging program and for users of the tagged corpus.

During this year the Survey has acquired a considerable amount of computer and audio equipment. We now have eleven personal computers networked to a SUN Sparc Station, two dot-matrix printers and one laser printer, four transcribing machines with headphones and foot pedals, and four cassette recorders.

We have developed a markup system for spoken texts which marks features such as self-correction, hesitation, and overlapping speech strings. We have also devised a system for computerizing the bibliographical, biographical, and sociolinguistic details of each text. This information can be retrieved together with the text, and will be used in generating sub-corpora.

Software has been written for partially automating the insertion of markup in texts as well as the selection of word-class tags from a list of alternatives. The major software development in progress is the ICE Corpus Utility Program, which will allow for sophisticated analysis of the ICE corpora. This system will provide for subcorpus selection, flexible searching, concordancing, statistical analysis, and a range of markup utilities.

3. **Visitors**

During the year we have been pleased to welcome the following scholars who have made use of our materials:

- **Prof. J. Algeo**
  University of Georgia  
  British and American grammar

- **Mr J. Ayto**
  Cambridge University  
  lexicography

- **Dr E. Antonopoulou**
  University of Athens  
  agent defocusing constructions

- **Professor L. Breivik**
  University of Bergen  
  cleft sentences

- **Mr T. Fukaya**
  Sugiyama Jogakuen University, Japan  
  prepositions

- **Ms M. Higashimori**
  Linguistics Dept., UCL  
  morphology, hyphenation
Ms M. Kolbel  
Albert-Ludwigs Universität  
Freiburg

Mr T. Ogura  
Hitotsubashi University  
Japan

Mt T. Prćić  
University of Novi Sad  
Yugoslavia

Dr M. Schulz  
Rostock University  
Germany

Ms J. Slavíčková  
Institute of Czech Language  
Prague

Mr D. Tayar  
Cambridge University Press

Ms T. Virtanen  
Åbo Akademi, Finland

several scholars have come for extended discussions on the ICE project:

Professor J. Aarts  
University of Nijmegen

Dr L. Bauer  
University of Wellington, New Zealand

Dr P. Collins  
University of New South Wales

Professor F. Davidson  
University of Illinois, Champagne-Urbana

Dr S. Granger  
University of Louvain

Professor O. Kujoře  
University of Ibadan

Professor M. Ljung  
University of Stockholm

Dr P. Mhundwa  
University of Zimbabwe
Dr P. Nihalani
Ms N. Oostdijk
Mrs M. Peters
Mr H. van Halteren

Others who paid us brief visits were:

Professor O. Alexandrova
Dr R. Baumgardner
Mr E. Chiwome
Dr B. Erman
Professor R. Filipovic
Mrs E. Gordon
Mr K. Hofland
Professor S. Johansson
Ms L. Kuipers-Alting
Ms Q. Liao
Professor M. Nahir
Mr P. Procter
Dr J. Schmied
Dr A.-B. Stenström
Ms J. Thondhlanu
Professor H. Widdowson
Professor K. Wikberg

University of Singapore
University of Nijmegen
Macquarie University, Australia
University of Nijmegen
State University of Moscow
The Asia Foundation, Islamabad
University of Zimbabwe
University of Stockholm
University of Zagreb
University of Canterbury, New Zealand
University of Bergen
University of Oslo
University of Groningen
Inst. of Linguistics, Beijing
University of Manitoba
Cambridge University Press
University of Bayreuth
University of Stockholm
University of Zimbabwe
Inst. of Education, University of London
University of Oslo
4. **Staff**

We have been joined by several new members of staff this year. Two full-time researchers arrived last October under our three-year grant from the Economic and Social Research Council: Akiva Quinn (who has a Cambridge University MPhil in Computer Speech and Language Processing) is responsible for computational developments, and Gerald Nelson (who has a PhD from University College Dublin in theories of linguistic form in the 18th century) is responsible for the compilation of the ICE corpus and is supervising the optical scanning, keypunching, and mark-up of the texts in the corpus. Yibin Ni (who received his MA in Modern English Language at University College London this past year and is studying for a PhD in the College in the same field) joined the Survey staff as a part-time member, and is working on the processing of written ICE texts. Other part-time staff who have been recruited more recently are Sue Boase (a former valued member of the Survey staff), who is recording conversations, and Yanka Gavin (an English graduate this year from the College), who is transcribing spoken material.

During the summer of 1991 David Elkan (who has a BSc in Computer Science from the College) is being employed to perfect and extend the program that he devised for his third-year project, and Laura Tollfree (who graduated this year in Phonetics and Linguistics at the College) is transcribing spoken material. Three MSc students from the Department of Computer Science - Neil McMaster, Andy Reed, and Albert Wichering - are working on Survey projects this summer under the supervision of Professor John Campbell of the Computer Science Department and Akiva Quinn.

Continuing staff from last year are René Quinault (our archivist), Oonagh Sayce, Andrew Rosta, and Vladimir Zagarac. In addition to his normal work on looking after the Survey Library and the secure storage of tapes, Rene has helped with the collection of radio recordings. Oonagh (a Survey stalwart) has worked on keypunching and proofreading written texts, while Vladimir has done similar work on the spoken texts. Andrew Rosta (who is working on a PhD in Linguistics at the College) has been actively engaged on the computational side; in particular, he has devised a system for preparing and annotating ICE texts.

Ine Mortelmans, who gave valuable service in the processing of texts for the past two years, left the Survey in July to return to Belgium. In addition to broadcasts from London and abroad, Sir Randolph Quirk gave lectures at the Complutense University, Madrid; the University of Seville; the University of Salamanca; the University of Valencia; the University of Barcelona; the University of the Balearic Islands; the Royal Society of Literature, London; the University Summer School of English, London; the University of Bologna; the University of Vienna; the National University of Mexico; the Catholic University of
Santiago; the Metropolitan University of Santiago; the Catholic University of Valparaiso; the Modern Languages Institute, Buenos Aires; the University of La Plata, Argentina; the University of Sao Paolo; the University of the Campinas, Brazil; the Nobel Symposium, Stockholm. He was awarded honorary doctorates at the University of Poznan and at the Polytechnic of Central London, and was elected to the Finnish Academy of Sciences.

Professor Sidney Greenbaum gave lectures at the University of Nottingham, the Hebrew Language Academy in Jerusalem, and the Nobel Symposium at Stockholm. He contributed to a Survey presentation at the ICAME conference in Ilkley, Yorkshire, jointly with Gerald Nelson and Akiva Quinn, and chaired there a general session and an ICE meeting. He has taken early retirement from the Quain Chair, and is now employed by the College as full-time Director of the Survey of English Usage.

Bas Aarts gave a paper at the University of Coimbra, Portugal. Rene Quinault was recently appointed a member of the Professional and Linguistic Assessment Board (PLAB) of the General Medical Council (Overseas Division), for which he has long examined. Gerald Nelson lectured at the University of Birmingham and taught at the summer school in St Patrick’s College, Maynooth, Ireland. Andrew Rosta presented a paper at the ICAME conference in Ilkley and attended a meeting in Oslo of the Text Encoding Initiative Working Group on Spoken Text.

5. Publications


Sidney Greenbaum

Director, Survey of English Usage

August 1991