UNIVERSITY COLLEGE LONDON

Survey of English Usage

ANNUAL REPORT 1990-91

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,

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

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

Ms M. Kolbel
Albert-Ludwigs Universität
Freiburg   similes

Mr T. Ogura
Hitotsubashi University

Japan   gerund, collocations

Mr T. Dráč

Dr M. Schulz
Rostock University   quantifying nouns and nominal
Germany   expressions

Ms J. Slavíčková
Institute of Czeck Language
Prague   scholarly discocourse

Mr D. Tayar
Cambridge University Press   lexicography

Ms T. Virtanen
Åbo Akademi—Finland   adverbial placement

Dr P. Nihalani	University of Singapore
Ms N. Oostdijk	University of Nijmegen
Mrs M. Peters	Macquarie University, Australia
Mr H. van Halteren	University of Nijmegen

Others who paid us brief visits were:

| Professor O. Alexandrova | State University of Moscow |

| Dr R. Baumgardner | The Asia Foundation, Islamabad |
| Mr E. Chiwome | University of Zimbabwe |
| Dr R. Erman | University of Stockholm |

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
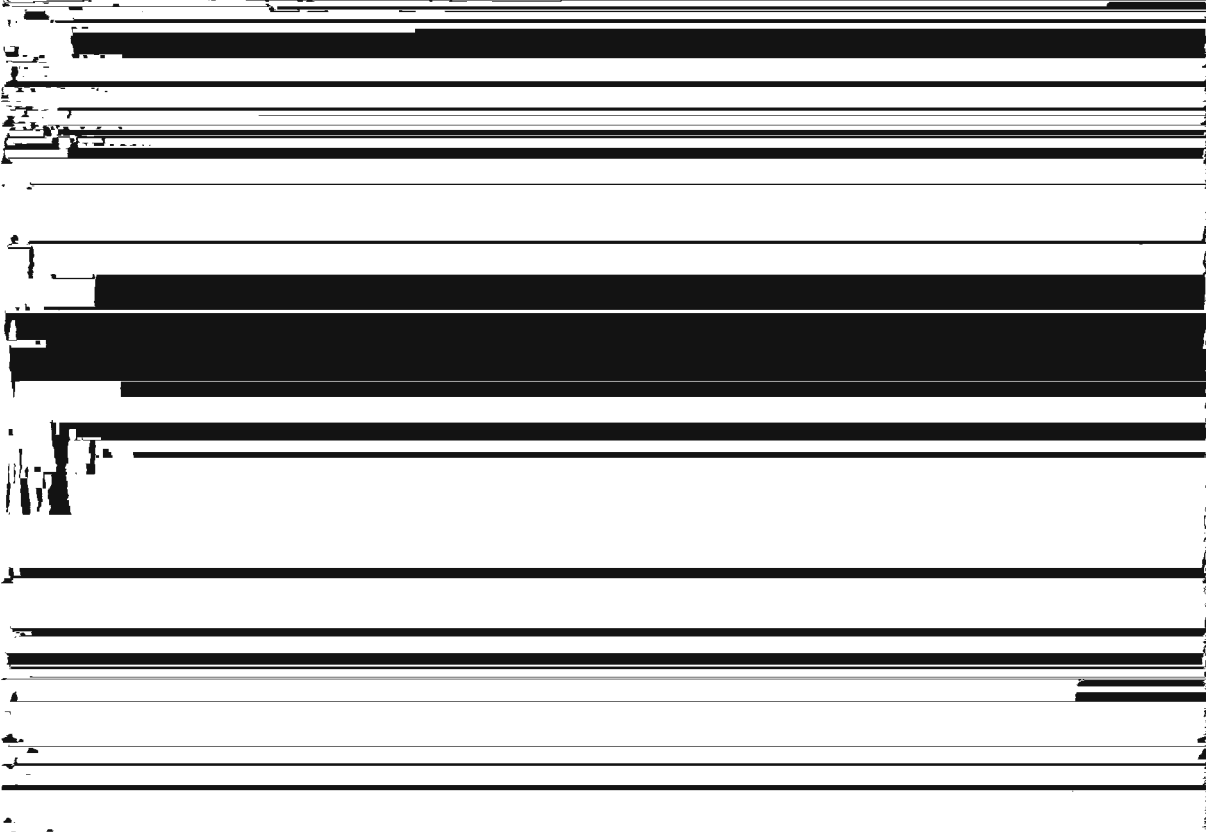
linguistic form in the 18th century) is responsible for the compilation of the ICE corpus and is supervising the optical scanning, keypunching, and mark-up of the texts in the corpus. Yibin Ni (who received his MA in Modern English Language at University College London this
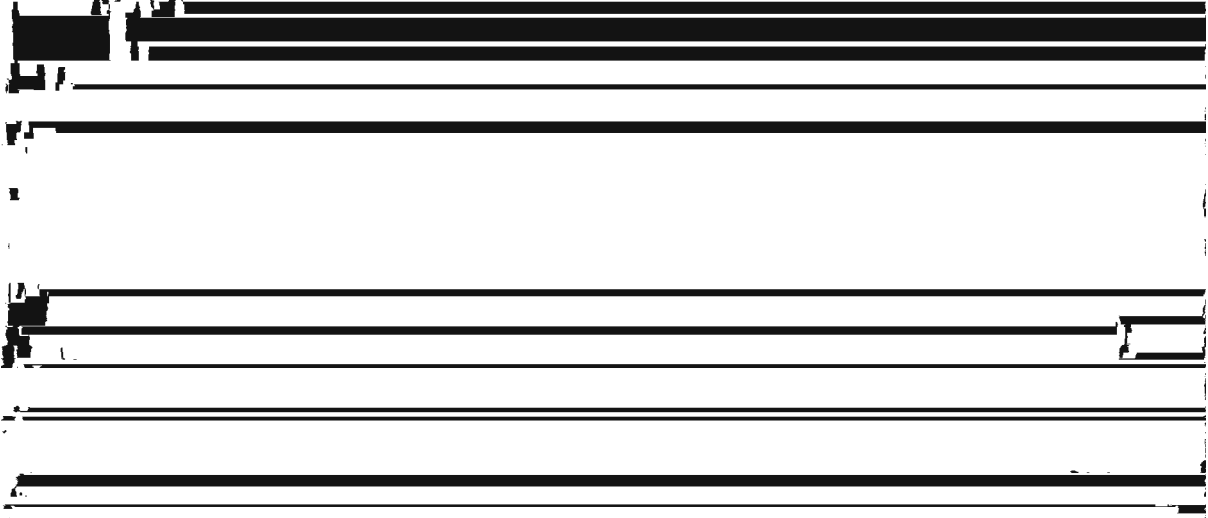
Santiago; the Metropolitan University of Santiago; the Catholic University of Valparaiso; the Modern Languages Institute, Buenos Aires; the University of La Plata, Argentina; the University of Sao Paolo; the University of the Campinas, Brazil; the Nobel Symposium, Stockholm. He was awarded honorary doctorates at the University of Poznan and at the Polytechnic of Central London, and was elected to the Finnish Academy of Sciences.
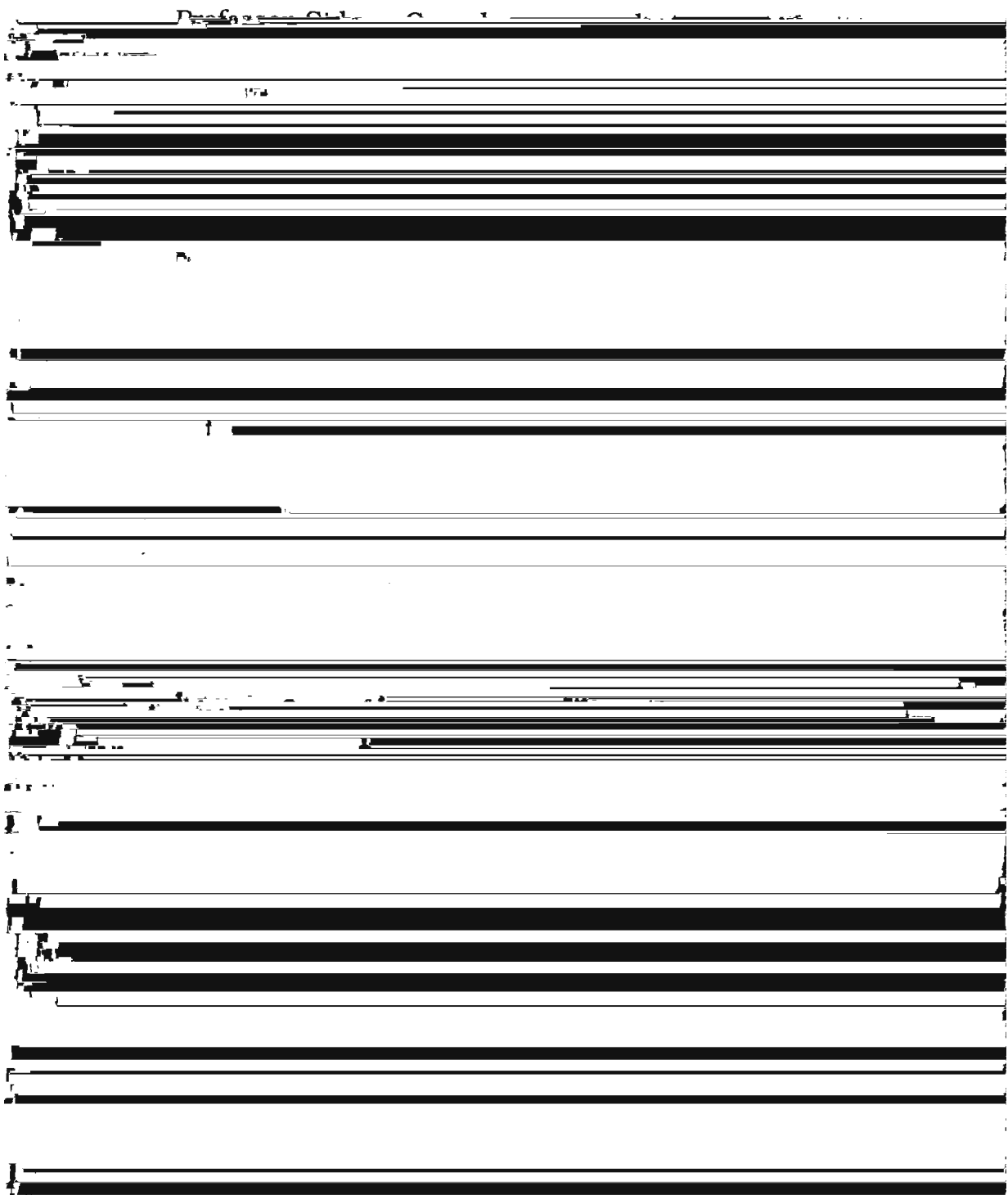
Mair, C. (1991) 'Quantitative or qualitative corpus analysis? Infinitival complement clause in the Survey of English Usage corpus', Johansson and Stenstrom, 69-80.

Transformation, ed. S.Marks, 1-6. London: Institute of Commonwealth Studies.

Quirk, R. (1991) 'Language Needs in the Single Market' Shoreline 6, 6-7

Quirk, R. (1991) 'The Issue of Standard English', Languages and