UNIVERSITY COLLEGE LONDON

Survey of English Usage

ANNUAL REPORT 1988-89

1. The 30th anniversary

This academic year marks the 30th year of the Survey's existence.
Randolph Quirk founded it at the University of Durham in 1959 and
moved it the following year to its present home at University College
London. Since 1983 Sidney Greenbaum has been Director of the Survey.

One of the principal activities of the Survey has been the
compilation and analysis of a corpus of British English used by adult
educated native speakers of the language. The corpus provides
researchers with samples of spoken and written English that represent
the range of uses to which the language is put.

The Survey corpus is now complete, having reached its original
target of a million words. Each text (sample of 5,000 words) has been
typed onto 6" x 4" slips with an overlap between each slip and
adjacent ones. For each collected feature in a text there is a marked
slip filed in a Survey filing cabinet. The Survey has collected 65
grammatical features, over 400 specified lexical items (words or
phrases), about 100 prosodic and paralinguistic features in spoken
texts, and all punctuation marks in written texts. For the most recently
collected texts, we no longer file slips for certain data, since they are
easily retrievable from our computerized corpus (see below). The data
consist of all the lexical items, some of the prosodic features, and all
the punctuation marks. Prosodic and paralinguistic features have still
to be filed for 20 of the 200 texts, and grammatical features for 15
texts.

Most of the Survey spoken texts (87) have been available since the early 1980s in machine-readable form in the London-Lund Corpus, compiled by the Survey of Spoken English at the University of Lund under the direction of Jan Svartvik. These texts are available on magnetic tape in a simplified prosodic transcription from the Norwegian Computing Centre for the Humanities, University of Bergen, P.O. Box 53, Norway. A selection of 34 of those texts appears in A Corpus of English Conversation, edited by Jan Svartvik and Randolph Quirk (Lund: Gleerup/Liber 1980). The Survey has now completed the computerization of the rest of the corpus. The computerization of 83 written texts was undertaken under the direction of Dr Geoffrey Kaye of IBM and funded by IBM. The transcription and computerization of the recorded texts that were not included in the London-Lund Corpus was funded by Longman and by an ESPRIT grant. They comprise 13 texts of spoken material and 17 texts of scripted material (i.e. read aloud from written scripts). For consistency, the recordings have been transcribed in the simplified system used in the London-Lund Corpus. The full prosodic and paralinguistic transcription is available in slips and booklets at the Survey premises.

The Survey corpus has supplied data for numerous publications. About 250 are known to us. Prominent among them is the standard reference grammar of modern English, A Comprehensive Grammar of the English Language by Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (London: Longman 1985).

The Survey annually updates its list of publications that have utilized material from the corpus. The list is available on request.

2. Progress on the computerized corpus

The corpus has been concordanced. The concordance lines may be browsed on the screen, with further context if desired, and may be printed with context. The concordance browser in present use was devised by Geoffrey Kaye, and a more advanced version has been requested and will be applied when it is available. The WordCruncher concordance browser program has been obtained and will also be used.

In addition, we are planning to have a concordance browser tailored to our specifications. For the benefit of the Survey, Professor John Campbell in the Computer Science Department at University College has kindly agreed to supervise projects that will be undertaken by his students. The first project, which will be conducted by David Elkan, is a program for a concordance browser that will allow for retrieval of lexical strings or word-class tags or a combination of the two.

By permission of Roger Garside and Geoffrey Leech of the University of Lancaster, we are planning to apply their tagging program to the Survey corpus. We are also investigating schemes for the syntactic parsing of the corpus.

We are grateful to Longman and the Sir Sigmund Sternberg Charitable Foundation for their continuing financial support.

3. International Corpus of English

The Survey is coordinating the planning of an International Corpus of English (ICE). At the time of writing, regions participating in the project are Australia, Canada, East Africa, India, New Zealand, Nigeria,

UK, and USA. In addition, there will be two nonregional supplementary corpora, compiled as analogues to the corresponding sections of the regional corpora: (a) a corpus of written translations into English from European languages, and (2) a corpus of international spoken communication. The Survey has undertaken to compile the British component of ICE.

An Advisory Board has been established consisting of scholars who have expertise in the construction of English corpora, in computational techniques for the analysis of corpora, and in English as a world language. It has been agreed that the material in the corpus should date from 1990 (though some teams may start collecting in 1991) and that we should think of the end of 1993 as the closing date for the material. Each regional corpus will be compiled and analysed along the same lines. The regional corpora will each consist of one million running words, and the texts will each contain 2,000 words.

4. **Tapes from <u>The Story of English</u>**

The Survey has acquired the complete set of the original recordings made for the BBC TV series <u>The Story of English</u>. By courtesy of the authors and the BBC, these recordings from all over the English-speaking world are available to researchers.

5. **Visitors**

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

Professor J. Algeo   University of Georgia   British & American English

Dr B. Erman	University of Sweden	discourse markers
Ms S. Grosslercher	University of Graz	conversational analysis
Professor Y. Ikegami	University of Tokyo	causative constructions
Mr Y.-C. Kim	University of Wales	passive
Professor M. Kohda	University of Osaka	syntax
Professor O. Kuj̧ore	University of Ibadan	syntax
Professor T. Mitsunaga	University of Nihon	phrasal verbs
Professor R. Pappenheim	University of Reading	discourse analysis
Dr J. Schmied	University of Bayreuth	relative constructions

Others who paid us brief visits were:

Mr. J. Amey	BBC, English by Radio and Television
Professor R. Baumgardner	The Asia Foundation, Pakistan
Professor R. Bugarski	University of Belgrade
Professor N. Enkvist	University of Helsinki
Mr N. Fukushima	University of Kokugakuin, Tokyo
Professor P. Fries	Central Michigan University
Dr R. Herwig	Schiller University, Jena
Dr R. Huddleston	University of Queensland
Mr M. Jaffe	Franklin Computers, USA
Professor S. Lee	University of Yonsei, Seoul
Mr M. Lesk	Bell Telephone Corporation
Dr C. Mair	University of Innsbruck
Mr C. McGregor	Latimer Lion Press, London
J. McCarten	Cambridge University Press
Professor S. Morag	Hebrew University, Jerusalem
Mr H. Morton	Fellow, National Endowment for the Humanities, Washington D.C.
Mr V. Starova	Institute of Foreign Languages, Moscow
Professor S. Terminasova	Moscow State University
Professor H. Chr. Wekker	University of Groningen

6. Staff

We welcomed two new people to the Survey this year: Fiona MacAlinden, who was awarded an MA in Modern English Language after taking the MA programme at University College, and Rudolph Spurling, who was awarded a BA in English from University College. Both of them were busy dealing with the backlog in the collection of prosodic and paralinguistic symbols on slips for filing. Fiona MacAlinden also typed and proofread texts in the reduced transcriptions on the computer, and was engaged in the collection of grammatical features. Rudolph Spurling is leaving the Survey to study for an MA in Old and Middle English at King's College.

Dr Jill Taylor and Rosalind Tricker completed the transcription from tape of the final texts and did some proofreading, and Rosalind typed one text on the computer in reduced transcription. We are sorry that they are both leaving the Survey, Dr Taylor to a lectureship at the University of Exeter and Rosalind Tricker to devote more of her time to a PhD in Phonetics.

Oonagh Sayce completed typing the final texts in slip form in full transcription and proofread them, and typed and proofread most of the texts in reduced transcription. She has also been working on the collection of grammatical features in the final 15 spoken texts. Andrew Rosta, who has just gained a BA in Linguistics from University College, has been working energetically on all aspects of the computational side. He is particularly active in providing liaison with the Computer Science Department on the specifications for the concordance browser program. Bas Aarts, who has been appointed as a lecturer in the English Department at University College from next session, has worked on the classification of texts and other matters. René Quinault has continued to

be responsible for keeping the records of the Survey's progress and for filing new offprints and periodicals. He has taken charge of the BBC recordings from The Story of English.

Sir Randolph Quirk lectured in Tokyo, Kyoto, Kobe, University of La Laguna (Spain), Isle of Man Postgraduate Centre, University College London, Royal Society of Medicine, and the Conference on University English (London), and gave several broadcasts by radio and television. He was elected to the Academia Europaea and received an honorary doctorate from the University of Sheffield.

Professor Sidney Greenbaum chaired a general session at the ICAME conference in Bergen as well as sessions of the ICE Advisory Board and participants. He completed his first year as Dean of Arts at University College London. He received an honorary doctorate from the University of Wisconsin-Milwaukee.

Dr Robert Ilson taught in the English Department at University College London. He continued to edit the International Journal of Lexicography, the EURALEX Bulletin, and (with Dr Joanna Channell) the Newsletter of the AILA Commission on Lexicography and Lexicology. At the Budapest Congress of EURALEX he gave a paper and was one of the four official speakers in the Presidential Debate.

Dr Bernard Lott continued to engage in research into prepositional meanings. He lectured at the annual conference of the International Association of Teachers of English and gave lectures and conducted workshops in the University of Indonesia (Djakarta) and Airlangga University (Surabaya).

Bas Aarts gave a paper at the autumn and spring meetings of the Linguistic Association of Great Britain.

7. Publications

We draw attention to a new series from Mouton de Gruyter: Topics in English Linguistics, edited by Professor Jan Svartvik (English Department, Lund University) and Professor Herman Wekker (English Department, Groningen University). The series will be open to books on language varieties.

Aijmer, K. (1988) 'Now may we have a word on this: The use of now as a discourse particle', Corpus Linguistics, Hard & Soft, eds. M. Kytö, O. Ihalainen, and M. Rissanen, 15-34. Amsterdam: Rodopi.

Akimoto, M. (1989) A Study of Verbo-Nominal Structures in English. Tokyo: Shinozaki Shorin.

Altenberg, B. (1987) 'Spoken English and the lexicon', Standardization in Computerized Lexicography: Proceedings of the IV Forum for Information Science & Practice, ed. V. Liedloff, 19-32, Saarbrücken: Institut der Gesellschaft zur Förderung der angewandten Informationsforschung e.V. an der Universität des Saarlandes.

Altenberg, B. (1987) Prosodic Patterns in Spoken English: Studies in the Correlation between Prosody and Grammar for Text-to-Speech Conversation. Lund: Lund University Press.

Altenberg, B. (1989) 'Some functions of the booster in spoken English', Proceedings from the Third Nordic Conference for English Studies, eds. I. Lindblad and M. Ljung, 5-20. Stockholm: Almqvist & Wiksell.

Bald, W.-D. (1988) 'A note on the textual distribution of one', Essays on the English Language and Applied Linguistics, eds. J. Klegraf & D.Nehls, 156-160. Heidelberg: Julius Groos.

Biber, D. (1988) Variation across Speech & Writing. Cambridge: Cambridge University Press.

Biber, D. & E. Finegan (1988) 'Drift in three English genres from the 18th to the 20th centuries: a multidimensional approach', Corpus Linguistics, Hard and Soft, eds. M. Kytö, O. Ihalainen, and M. Rissanen, 83-101. Amsterdam: Rodopi.

Biber, D. & E. Finegan (1988) 'Adverbial stance types in English', Discourse Processes 11, 1-34.

Boase, S. (1988) 'The Centipede's Progress', Degeneration on the Air?, eds. A. Juul, H.F. Nielsen, and K. Sørensen. Copenhagen: National Institute for Educational Media.

Brodda, B. (1988) 'Tracing turns in the London-Lund Corpus with BetaText', Literary and Linguistic Computing 3, 94-104.

Engels, L. (1988) 'The effect of spoken and written-to-be-spoken English on word frequency counts of written English', Essays on the English Language & Applied Linguistics, eds. J. Klegraf and D. Nehls, 407-425. Heidelberg: Julius Groos.

Geluykens, R. (1988) 'On the myth of rising intonation in polar questions', Journal of Pragmatics 12, 467-485.

Greenbaum, S. (1988) 'Syntactic devices for compression in English', Essays on the English Language and Applied Linguistics, eds. J. Klegraf and D. Nehls, 3-10. Heidelberg: Julius Groos.

Greenbaum, S. (1989) A College Grammar of English. New York: Longman.

Haase, I. (1988) 'Temporafolge und Bedingunstyp in Konditionalen Satzgefüge', Essays on the English Language and Applied Linguistics, eds. J. Klegraf and D. Nehls, 69-76. Heidelberg: Julius Groos.

Ikegami, Y. (1988) '"HAVE + object + past participle" and "GET + object + past participle" in the SEU Corpus', Meaning and Beyond: Ernst Leisi zum 70. Geburstag, eds. U. Fries and M. Heusser, 197-123. Tübingen: Gunter Narr.

Ilson, R. (1988) 'How big is your dictionary?' Verbatim 15, 9-10.

Ilson, R. (1989), consultant ed. Reverse Dictionary. London: Reader's Digest.

McCarthy, M. (1988) 'Some vocabulary patterns in conversation', Vocabulary and Language Teaching, eds. R. Carter and M. McCarthy, 181-200. London: Longman.

Mindt, D. (1988) 'Rechnergestützte Forschungen zur Angewandten Linguistik', EDV in der Angewandten Linguistik: Ziele – Methoden – Ergebnisse, ed. D. Mindt, 7-9. Frankfurt/Main; Diesterweg.

Mindt, D. and I. Haase (1988) 'Definition linguistischer Daten und Modifikation durch Variablenänderung: das englische Verbsystem', EDV in der Angewandten Linguistik: Ziele – Methoden – Ergebnisse, ed. D. Mindt, 45-53. Frankfurt/Main: Diesterweg.

Nevalainen, T. (1987) 'A quantitative survey of nuclear tone variation in English', Proceedings of the Eleventh International Congress of Phonetic Sciences, vol.3, 180-183. Tallin: Institute of Language and Literature, Academy of Sciences of the Estonian S.S.R.

Oostdijk, N. (1988) 'A corpus for studying linguistic variation', ICAME Journal 12, 3-14.

Oostdijk, N. (1988) 'A corpus linguistic approach to linguistic variation', Literary and Linguistic Computing 3, 12-25.

Pakosz, M. (1988) 'Tonic prominence and the coding of thematic-rhematic relations', <u>Papers in Contrastive Linguistics</u> 23, 119-134.

Pannek, G. (1988) 'Relativsätze im gesprochenen Englisch und in Lehrwerken für den Englischunterricht', <u>EDV in der Angewandten Linguistik: Ziele-Methoden-Ergebnisse</u>, ed. D. Mindt, 77-83. Frankfurt/Main: Diesterweg.

Quirk, R. (1988) (with Y. Ikegami and M. Toyota) <u>Words at Work</u>, 2nd ed. Tokyo: Kinokuniya (in Japanese).

Quirk, R. (1989) 'Grammar Research and English Teaching', <u>NHK Journal</u> 4, 88-105.

Quirk, R. (1989) 'Language Varieties and Standard Language', <u>JALT Journal</u> 11, 14-25.

Stenström, A.-B. (1988) 'Adverbial commas and prosodic segmentation', <u>Corpus Linguistics, Hard and Soft</u>, ed. M. Kytö, O. Ihalainen and M. Rissanen, 221-236. Amsterdam: Rodopi.

Stentröm, A.-B. (1988) 'Questioning in conversation', <u>Questions and Questioning</u>, ed. M. Meyer, 304-325. Berlin: de Gruyter.

Svartvik, J. (1988) 'On the study of spoken English', <u>Sprache und Individuum. Kongressbeiträge der 17. Jahrestagung der Gesellschaft für Angewandte Linguistik</u>, ed. W. Kuhlwein and B. Spillner, 30-37. Tübingen: Gunter Narr.

Tesch. F. (1988) '<u>Some</u> und <u>any</u> in affirmativen und negativen Kontexten', <u>EDV in der Angewandten Linguistik: Ziel-Methoden-Ergebnisse</u>, ed. D. Mindt, 59-68. Frankfurt/Main: Diesterweg.

Tottie, G. (1988) '<u>No</u>-negation and <u>not</u>-negation in spoken and written English', <u>Corpus Linguistics, Hard and Soft</u>, ed. M. Kytö, O. Ihalainen, and M. Rissanen, 245-265. Amsterdam: Rodopi.

Sidney Greenbaum

Director, Survey of English Usage

September 1989