UNIVERSITY COLLEGE LONDON

Survey of English Usage

ANNUAL REPORT 1994-95

1. Research at the Survey

The main areas of research this past year have been related to parsing and to the Leverhulme Project.

1a. The Survey Parser

Alex Chengyu Fang has been developing the Survey parser, which has progressed through several versions. Progress is expected to be speeded up as a result of a substantial grant from EPSRC (the Engineering and Physical Sciences Research Council) for a two-year project that began in September 1995. ESPRC is funding two Research Fellows: Alex and Sean Wallis.

1b. ICE Tree

ICETree, devised by Nick Porter and Dennis Tech-Yong, is a Windows-based application for graphically building or editing parse trees. Delays in producing an efficient version postponed for some time the start of parsing with the Survey parser. An entirely new and much more sophisticated version of ICETree is being developed by Sean Wallis and is almost ready for use.

1c. The parsing of ICE-GB

For parsing ICE-GB (the million-word British component of the International Corpus of English) we had first used the TOSCA parser devised by the TOSCA Research Group under the direction of Professor Jan Aarts at the University of Nijmegen. The TOSCA parser required manual pre-editing and manual selection from what were generally a number of analyses. It sometimes required correction of tags to induce analyses. In all, the TOSCA parser succeeded in producing parses for 78.1 per cent of the written sentences or spoken utterances, representing 62 per cent of the total number of words in ICE-GB.

The remaining 38 per cent of ICE has been parsed by the Survey parser that is being developed by Alex Fang. The Survey parser does not require pre-editing. It offers just one parse, which is displayed on ICETree. The parses are checked manually and (if necessary) corrected. So far about 50 texts have been checked. We have recruited additional staff to speed up the checking.

We have been extending the parsing system to cater for types of constructions on which the TOSCA parser failed. Since the extensions often have consequences for earlier analyses, we intend to make some global changes in those analyses when we have finished the parsing.

1d. The Leverhulme Project

Funding for the two-year Leverhulme project finished at the end of September 1995. The project was primarily an investigation into clause relationships in spoken and written English. The results are incorporated in a number of publications by Professor Greenbaum and Dr Nelson. Two papers have been published and two others are in press. A further three papers are in progress. Results from the research also appear in a chapter in The Oxford English Grammar by Sidney Greenbaum, due to be published by Oxford University Press in February 1996.

1e. British Telecom contract

The Survey was awarded a contract by British Telecom to evaluate a speech synthesis system. The evaluation experiments were conducted by Professor Sidney Greenbaum and Alex Fang and the results were accepted by BT researchers. We are expecting to obtain further contracts.

2. ICE Documents

Three ICE Newsletters were distributed by the Survey this year.

Completely revised editions of two manuals on the ICE tagging system were produced:

Sidney Greenbaum, The ICE Tagset Manual (viii + 90 pp)

Ni Yibin, An Alphabetical Reference to the ICE Tagset (80 pp)

Documents were also produced on the new Survey ICE parsing system:

Ni Yibin, <u>The Correlation between Categories and Functions in ICE Parsing</u> (53 pp)

Ni Yibin, <u>The ICE Parsing Manual</u> (180 pp)

Ni Yibin, 'Comparison of Terminology between ICE Tagset and TOSCA for ICE' (9 pp)

The results of the British Telecom contract appear in a report:

Sidney Greenbaum and Alex Chengyu Fang, <u>Evaluating the BT Laureate Text-to-Speech System. Technical Report.</u> (121 pp)

3. Funding

The major financial contribution for 1994-5 came from the Leverhulme Trust. We are also grateful for financial support from the Michael Marks Charitable Trust, the Sir Sigmund Sternberg Foundation, and Pearson plc. (Pearson's contribution for the previous year was omitted by mistake from the last Report.) In addition, we have benefited from the British Telecom contract.

For the 1995-96 year we have so far received a grant of £176,365 from EPSRC (Engineering and Physical Sciences Research Council) for a two-year period, £10,000 from the Great Britain Sasakawa Foundation, and £10,000 from an anonymous donor. The Michael Marks Charitable Trust has loyally supported the Survey with generous donations for a number of years and is continuing to do so.

4. The Lord Quirk

Congratulations to Professor Randolph Quirk on his elevation from knighthood to peerage. The Lord Quirk was Director of the Survey of English Usage until 1983. He has not been engaged in the activities of the Survey since then, though he continues to have a benevolent interest in its progress.

5. Staff

Nick Porter left the Survey with our good wishes to join the multimedia section of Dorling Kindersley.

Continuing staff from last year are Judith Broadbent, Justin Buckley, Alex Chengyu Fang, Gerry Nelson, Oonagh Sayce, and Ni Yibin. Continuing voluntary participants are Isaac Hallegua and René Quinault. Neil Morgenstern, who was employed in the Department of Computer Science, also volunteered his services when required.

We welcome Sean Wallis, who joined Alex Fang as Research Fellow on the EPSRC project at the beginning of October 1995. Other staff joined us in October or November 1995 to work on checking the output of the parser: Celine Bijleveld, Ken Fletcher, and Andrew Neal.

Nick Porter mainly worked on ICETree and ICECUBE (the software package for preparing and installing ICECUP). Isaac Hallegua continued his activities on system and data management. In addition to his usual archive duties, René Quinault has continued working on reducing the ICE recordings to those used in the ICE corpus.

Alex Fang has devoted most of his time to developing the Survey automatic parsing program. He also spent several weeks on the BT contract. He made further improvements to AUTASYS, the automatic tagger. Jointly with Dr Mark Huckvale (Department of Phonetics and Linguistics at UCL) he is developing PROSICE, a speech database.

As during last year, Gerry Nelson has been chiefly engaged on the Leverhulme project, but has continued to provide guidance to ICE teams internationally. Judith Broadbent, Justin Buckley, and Oonagh Sayce have been contributing to the checking of the output from the Survey parser. Yibin Ni has been mainly working on recording decisions on parsing and composing detailed documents on the parsing and tagging systems.

Tony Dodd has been employed as a freelance consultant on the alignment of the parsed corpus with the tagged corpus.

Discussions on the evolving ICE parsing system are mainly the responsibility of Professor Greenbaum, Alex Fang, and Ni Yibin.

Professor Greenbaum visited the ICE project in Nigeria in January 1995 and held meetings with the leading English language scholars in the country. As a result, an ICE Advisory Board was formed and a new Director of ICE-Nigeria was appointed. During his brief visit he reviewed New English: A West African Perspective (edited by Ayo Bamgbose, Ayo Banjo, and Andrew Thomas, published 1995 by Mosuro at Ibadan) at a launch of the book at Lagos. At the ICAME conference in May 1995 in Toronto, he presented a paper jointly with Gerry Nelson, chaired an ICAME session, and chaired a meeting of ICE

4

participants. At a reception in his honour at the conference he was presented with a book dedicated to him: The Verb in Contemporary English: Theory and Description (edited by Bas Aarts and Charles F. Meyer, Cambridge University Press 1995). He gave a presentation at a postgraduate seminar in the Phonetics and Linguistics Department at UCL.

Alex Fang was appointed Deputy Director of the Survey in June 1995. He gave a paper and a software presentation of AUTASYS at the Toronto ICAME conference. He also gave a presentation at a postgraduate seminar in the Phonetics and Linguistics Department at UCL. He participated in a BBC World Service series on English internationally.

Gerry Nelson participated in a BBC World Service discussion on the future of English throughout the world.

Nick Porter demonstrated ICECUP and ICETree at the Toronto ICE conference. Jointly with Judith Broadbent and Justin Buckley he devised a poster on the Help systems for ICECUP and ICETree.

6. Visitors

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

Professor John Algeo U.S.A.	American and British Grammar
Professor Funso Akere University of Lagos	ICE systems
Anne Belan Sorbonne	nominalizations
Dr Florence Dolphyne University of Ghana	ICE systems
Francisco Gonzalvez Garcia University of Granada	modal auxiliaries
Dr Joseph Hung Chinese University of Hong Kong	ICE transcription system
Professor Takuro Ikeda Aoyama Gakuin University, Tokyo	idioms

Tom Lavelle University of Stockholm	nominalizations
Professor Magnus Ljung University of Stockholm	nonfinite adverbial clauses
David Tiomajou University of Yaounde, Cameroon.	ICE transcription system

Other visitors were:

Gordan Adgey	Mercury Telecommuncations plc
Julian Amey	Department of Trade and Industry
Janet Baker	Dragon Systems Ltd
Simon Bell	Routledge
Nigel Birch	Engineering & Physical Sciences Research Council
Andy Breen	British Telecom
John Bridle	Dragon Systems Ltd.
Kendall Clark	Oxford University Press
Dr Christopher Everett	Daiwa Foundation
Professor Jan Firbas	University of Brno, Czech Republic
Maggie Gaved	British Telecom
Martin Good	Cambridge Training and Development Ltd.
Professor Manfred Görlach	University of Cologne
Peter Hand	The Great Britain Sasakawa Foundation
John Hayward	SOROS
Susan Jellis	Cambridge Language Survey
Professor Graeme Kennedy	University of Victoria at Wellington
Dr John Kirk	Queen's University Belfast
Professor Korponay	Hungary

Peter Lawrence	Lifeline and Humanitarian Association
Sophie Lawrence	Macmillan
Mark Le Fanu	Society of Authors
Gill Macmillan	Cambridge Training and Development Ltd.
Lord and Lady Marks of Broughton	U.K.
Dr Martinez	University of Granada
Professor G. Miglior	University of Venice
Steve Minnis	British Telecom
Professor Tsukao Mitsunaga	University of Nihon, Tokyo
Dr Olga Müllerová	Czech Language Institute, Prague
Dr Paroo Nihilani	University of Singapore
Susanne Ostertag	University of Stuttgart
Nick Ostler	Linguacubun Ltd.
John Pickford	BBC World Service
Jonathan Price	Macmillan
Professor Herbert and Ute Schendel	University of Vienna
Professor Josef Schmied	Technical University of Chemnitz, Germany
Rob Scriven	Oxford University Press
Professor Jan Svartvik	University of Lund

7. Publications based on Survey material

Aarts, Flor (1994) 'Imperative sentences in A Corpus of Conversation' Leuvense Bijdragen 83, 145-155.

Aijmer, Karen (1995) 'Do women apologise more than men?', Studies in Anglistics, eds. Gunnel Melchers and Beatrice Warren, 55-69. Stockholm: Almqvist and Wiksell.

Algeo, John (1995) 'Having a look at the expanded predicate', The Verb in Contemporary English: Theory and Description, eds. Bas Aarts and Charles F. Meyer, 203-217. Cambridge: Cambridge University Press.

Cheepen, C. (1995) 'Discourse considerations in transcription and analysis', Spoken English on Computer, eds. Geoffrey Leech, Jenny Thomas, and Greg Myers, 220-223. London: Longman.

Coates, Jennifer (1995) 'The expression of root and epistemic possibility in English', The Verb in Contemporary English: Theory and Description, eds. Bas Aarts and Charles F. Meyer, 145-156. Cambridge: Cambridge University Press.

Crystal, David (1995) 'Refining stylistic discourse categories', Studies in Anglistics, eds. Gunnel Melchers and Beatrice Warren, 35-46. Stockholm: Almqvist and Wiksell.

Erman, Britt (1995) 'He is an anglist, or?', Studies in Anglistics, eds. Gunnel Melchers and Beatrice Warren, 95-107. Stockholm: Almqvist and Wiksell.

Esser, Jürgen (1994) 'Medium-transferability and corpora: Remarks from the consumer-end of corpus linguistics', Hermes 13, 45-53.

Geisler, C. (1995) Relative Infinitives in English. Stockholm: Almqvist and Wiksell.

Greenbaum, Sidney and Gerald Nelson (1995) 'Clause Relationships in Spoken and Written English', Functions of Language 2.1-21.

Greenbaum, Sidney and Gerald Nelson (1995) 'Nuclear and Peripheral Clauses in Speech and Writing', Studies in Anglistics, eds. Gunnel Melchers and Beatrice Warren,181-190. Stockholm: Almqvist and Wiksell.

Greenbaum, Sidney and Randolph Quirk (1995) A Student's Grammar of the English Language. Japanese translation. Tokyo: Kinokuniya.

Hasselgård, Hilda (1992) 'Sequences of spatial and temporal adverbials in spoken and written English', New Directions in English Language Corpora: Methodology, Results, Software Developments, ed. Gerhard Leitner, 319-328. Berlin: Mouton de Gruyter.

Hasselgård, Hilda (1993) 'Sequences of time and space adverbials in clause-initial position: A survey of ordering principles', Proceedings from the 5th Nordic Association of English Studies Conference, Reykgarvik 1992, ed. J. D'Arcy. Reykjarvik: University of Iceland Publishing Co.

Johansson, C. (1995) The relativizers 'whose' and 'of which' in present-day English. Stockholm: Almqvist and Wiksell.

Johansson, Stig (1995) '"This scheme is badly needed"': some aspects of verb-adverb combinations', The Verb in Contemporary English: Theory and Description, eds. Bas Aarts and Charles F. Meyer, 218-240. Cambridge: Cambridge University Press.

Kryk-Kastovsky, B. (1995) 'Particles revisited', Studies in Anglistics, eds. Gunnel Melchers and Beatrice Warren, 81-94. Stockholm: Almqvist and Wiksell.

Nelson, Gerald (1995) 'The International Corpus of English: Markup and Transcription', Spoken English on Computer, eds. Geoffrey Leech, Jenny Thomas, and Greg Myers, 220-223. London: Longman.

Paradis, Canta (1994) 'Compromiser - a notional paradigm', Hermes 13, 157-167.

Peppé, Sue (1995) 'The Survey of English Usage and the London-Lund Corpus: Computerizing manual prosodic transcription', Spoken English on Computer, eds. Geoffrey Leech, Jenny Thomas, Greg Myers, 187-202. London: Longman.

Slavickova, Jana (19920 'Is there a tendency in English to use abstract nouns in the singular? - A corpus-based analysis', Linguistica Pragensia 1/92, 28-35.

Stenström, Anna-Brita (1995) 'Some remarks on comment clauses', The Verb in Contemporary English: Theory and Description, eds. Bas Aarts and Charles F. Meyer, 290-301.

Svartvik, Jan and Olof Ekedahl (1995) 'Verbs in public and private speaking', The Verb in Contemporary English: Theory and Description, eds. Bas Aarts and Charles F. Meyer, 273-289. Cambridge: Cambridge University Press.

Vandepitte, Sonia (1993) A pragmatic study of the expression and the interpretation of causality: Conjuncts and conjunctions in modern spoken British English. Brussels: Paleis der Academiën.

Sidney Greenbaum
Director, Survey of English Usage
November 1995