

address for *IGE* is the same as the Survey's: <http://www.ucl.ac.uk/english-usage/>. Si

The Survey Parser Project (EPSRC Grant No GR/K75033)

The Survey Parser Project is for the development of an automatic syntactic parser of English and a lexical database that provides useful information for parsing. Alex Chengyu Fang was employed for the parsing component and commenced work in September 1995. Sean Wallis was employed for the lexical component and joined us in October 1995. The Survey Parser is now functioning. It is fully automatic, fast, and produces only one analysis for the input string. The current design of the parser comprises two modules: wordclass tagging and syntactic parsing. Raw texts can be fed into the parser without having to be pre-edited. The output is an indented tree structure with appropriate labels. The parser is able to process 30 words per second, a speed that is unaffected by the length of the input string. The single analysis produced can be complete or partial. A complete analysis is one where all the constituents have been successfully labelled. A partial parse has one or more missing syntactic functions. The parser has been used to parse the residue of the British component of the International Corpus of English (ICE-GB), that could not be dealt with by the TOSCA parser (Nijmegen, the Netherlands). This residue represents about 320,000 words or 28,000 sentences. The Survey Parser successfully produced either a global or a partial analysis for all the input strings. The next major development will be the interface between the parser and the lexical database. It is expected that information from the lexical database will contribute to better parsing outputs. The lexical database (LDB) currently consists of 162,989 entries. Each entry consists of a unique headword and a series of sub-entries, distinguished by the part of speech, and containing cross-references to other appropriate word forms, including inflected, derived, compounded and alternative forms. Each entry may additionally include atomic semantic information, and a list of possible collocational frames. This information has been aggregated from

a number of sources: the Collins English Dictionary, the Oxford Advanced Learner's Dictionary, and the Longman Dictionary of Contemporary English.

3. Staff

Bas Aarts was appointed Director of the Survey at the beginning of 1997.

Continuing staff from last year are Judith Broadbent, Justin Buckley, Celine Bijleveld, Marie Gibney, Isaac Hallegua, Gerry Nelson, René Quinault, Sean Wallis, and Jonathan White. Other staff who joined us this year to work on

the alignment of computerised texts are: Brian Davies, and Gunther Kaltenboeck.

We congratulate Judith Broadbent on obtaining her PhD.

We also congratulate Yibin Ni on obtaining his PhD, and on his appointment as lecturer at the University of Singapore, a post he will take up later in the year.

Alex Fang has left the Survey and now works on the EPSRC project in the Department of Phonetics and Linguistics at UCL, where he can be reached by e-mail: alex@phonetics.ucl.ac.uk.

Ken Fletcher has left the Survey for Central China, where he will be working for VSO.

Oonagh Sayce retired from the Survey in December, but still works on an occasional basis.

4. Publications based on Survey material

Aarts, B. (1996) 'The rhetorical adverb *simply* in present-day British English'. In *Synchronic Corpus Linguistics: Papers from the Sixteenth International Conference on English Language Research on Computerised Corpora (ICAME 16)*, edited by C.E Percy, C.F. Meyer and I. Lancashire, 59-68.

Aijmer, K. (1996) *Conversational routines in English: convention and creativity*. Longman: London.

Buckley, Justin (1996) 'An Outline of the Survey's ICE Parsing Scheme'. In *Comparing English Worldwide: The International Corpus of English*, edited by Sidney Greenbaum. Clarendon Press: Oxford, 125-141.

Collins, Peter C. (1996) 'Get-passives in English'. In *World Englishes*,

Fang, Alex Chengyu (1996) 'Automatically generalising a wide-coverage formal grammar'. In *Synchronic Corpus Linguistics: Papers from the Sixteenth International Conference on English Language Research on Computerised Corpora (ICAME 16)*, edited by C.E Percy, C.F. Meyer and I. Lancashire, 207-222.

Fang, Alex Chengyu (1996) 'AUTASYS: Grammatical Tagging and Cross-Tagset Mapping'. In *Comparing English Worldwide: The International Corpus of English*, edited by Sidney Greenbaum. Clarendon Press: Oxford, 110-124.

