Learner English on Computer

Sylviane Granger (editor)
(Université Catholique de Louvain)

London: Addison Wesley Longman
(Studies in language and linguistics,
edited by Geoffrey Leech and Jenny
Thomas), 1998, xxii+228 pp;
paperbound, ISBN 0-582-29883-0, £18.99

Reviewed by
Greg Lessard
Queen's University

This is a collection of papers edited by the founder and coordinator of the International
Corpus of Learner English (ICLE), which brings together written texts produced by
non-native speakers (NNSs) of English from a variety of European mother-tongue
backgrounds. The book is divided into three parts, each composed of several papers:
the first part is devoted to a general outline of the constitution and analysis of learner
corpora, the second to a series of case studies, most of which involve the comparison
of NNS and native speaker (NS) corpus data, and the third to the use of corpora to
enhance pedagogical tools such as dictionaries, instructional software, and teaching
manuals.

The opening paper of Part 1, by Sylviane Granger herself, is both an apology for
and a description of the use of learner corpora. With links to previous work on error
analysis, such corpora represent a sharp break with most current second-language
acquisition research, which relies primarily on elicited data or introspection, and which
often prefers oral to written data. Granger provides a good discussion of the design
criteria that underpin the ICLE corpus, as well as the methodological issues involved
in comparing NNS and NS data.

In the second paper, Fanny Meunier provides a basic overview of the range of tools
available for marking up raw corpora, from basic POS taggers to parsers to semantic,
discourse, and error taggers. Also discussed are the types of software available for
lexical frequency and concordance style analysis. The paper is complemented by a list
of software in the appendix to the volume.

The second part of the volume contains eight papers. Håkan Ringbom provides
preliminary evidence of similarities and differences in lexical frequencies of advanced
learner English in seven western-European learner corpora, as compared to a large
native speaker corpus of argumentative essays. Gunter Lorenz discusses some stylis-
tic aspects of adjective intensification in a corpus of German NNS, as compared to
English NS data, finding among other things that the former tend to use intensifi-
cation typically in thematic (clause-initial) position, whereas the latter tend to prefer
rhematic (clause-final) position. Sylvia de Cock, Sylviane Granger, Geoffrey Leech, and
Tony McEnery present software designed to capture continuous formulaic expressions
in NNS corpus data, and illustrate this using French speaker data, with special refer-
ence to vagueness tags such as *or whatever*. Bengt Altenberg and Marie Tapper discuss
the overuse and underuse of adverbial connectors such as *furthermore* (overused) and
however (underused) in a corpus of advanced Swedish learners of English, as com-
pared to a corpus of native speaker writing. Tuija Virtanen looks at the frequency of

direct questions in seven NNS corpora, as compared to a British and an American NS corpus, finding a range of variation, with most (but not all) NNSs overusing questions. Stephanie Petch-Tyson discusses markers of writer/reader visibility (including first- and second-person pronouns and referential items) in four NNS corpora (French, Dutch, Swedish, and Finnish), finding them overused in all compared to NS data. Sylviane Granger and Paul Rayson compare the relative frequencies of part of speech in French NNS and English NS texts, finding that NNSs overuse determiners, pronouns, and adverbs, and underuse conjunctions, prepositions, and nouns. Jan Aarts and Sylviane Granger compare tag sequences in POS-tagged corpora of Dutch, French, and Finnish NNSs to English NSs. Among other things, they find striking differences between learners and native speakers with respect to sentence-initial structures and sequences involving prepositions.

The third part of the volume contains five papers. Doug Biber and Randi Reppen discuss the potential contribution of corpus data to the teaching of complement clauses (based on *that, to, -ing,* and *wh-* forms) using NNS data from French, Spanish, Chinese, and Japanese learners. Patrick Gillard and Adam Gadsby from the ELT division of Longman look at the use of learner corpora in compiling English language teaching dictionaries. Przemysław Kaszubski uses a Polish perspective to discuss the usefulness of corpora to enhance writing textbooks. John Milton describes a computer-based teaching environment designed for use in the teaching of English in Hong Kong. Among other things, tagged corpora of authentic learner errors are used as doors into a tutorial system. Finally, Sylviane Granger and Chris Tribble discuss the use of learner corpus data in the foreign language classroom as the basis for an inductive approach to teaching.

As Geoffrey Leech points out in his preface, this is only the first book on the construction and use of NNS corpora. Clearly, the field is still very young, and there are many areas that will require attention. The corpora used are still relatively small, which hinders the use of statistical tests. Many of the corpora are only lightly tagged, usually only at the POS level. More syntactic, semantic, discourse, and error tagging will be required to obtain the most from software. Finally, and probably most significantly, apart from a few basic theoretical constructs such as overuse and underuse with respect to native speaker usage, there is a serious lack of contact with the notions of current linguistic theory. For example, the concept of markedness never appears in the volume.

Despite these shortcomings, normal in a new field, this volume represents a well-edited contribution to NNS corpus research and is well worth the read.

Greg Lessard is Associate Professor of French Studies and of Computing and Information Science at Queen's University. He has published in various areas of computational linguistics and humanities computing, including second-language French corpora, computer-assisted language learning, natural language generation and computer-based second-language subject testing. Lessard's address is: Department of French Studies, Queen's University, Kingston, Ontario, Canada K7L 3N6; e-mail: lessardg@qsilver.queensu.ca.