

Sylviane Granger, Joseph Hung, and Stephanie Petch-Tyson (eds.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam and Philadelphia: John Benjamins, 2002. viii + 245 pp. ISBN 90 272 1702 5. Reviewed by **Erik Smitterberg**, Stockholm University.

In recent years, corpus linguistic methods have gained an increasingly central place in English language teaching (ELT). For instance, students use a range of materials that draw on corpus linguistics, from modern learners' dictionaries to concordances produced for the purpose of data-driven learning. The overwhelming majority of all English-language corpora used in this way consist of native-speaker English, but recent findings have shown that the compilation and analysis of learner corpora are also relevant from a pedagogical perspective (see e.g. Granger 1998), regarding fields such as materials design. There are also indications that students may benefit from analysing learner English as a complement to looking at native-speaker output. However, there is a need for studies of learner English that help to open up the field in this respect. *Computer learner corpora, second language acquisition and foreign language teaching* helps to satisfy this need. It is a collection of contributions intended both to help researchers assess the relevance of research on computer learner corpora for second language acquisition (SLA) theory, as well as ELT practice, and to "give practical insight to researchers who may be considering compiling a corpus of learner data or embarking on learner corpus research" (p. vii); this is a broad scope for a single volume, something which I shall return to towards the end of this review.

The book is divided into three sections. The first, "The role of computer learner corpora in SLA research and FLT", contains only one paper, by Sylviane Granger. She focuses on the contribution learner corpora can make to SLA and Foreign Language Teaching (FLT) research. As the title of her contribution indicates, Granger offers a brief but lucid and informative "bird's-eye view of learner corpus research". She outlines the field of corpus linguistics and the role of learner data in FLT and SLA research, and – importantly – offers a definition of learner corpora as well as commenting on aspects of that definition. Approaches such as Contrastive Interlanguage Analysis (CIA) and Error Analysis (EA) are discussed, as are the possibilities of software-aided analysis of computerized learner English. Granger is careful to point out potential pitfalls in this area, for example the fact that the accuracy rate of automatic taggers may decrease when they are applied to non-native English. Granger also discusses

practical applications of learner corpora, e.g. materials design, and future challenges, such as the need for corpus linguists and, among others, SLA specialists to co-operate (see also Hasselgård 1999: 152). Granger's contribution provides the non-specialist reader with a good deal of background information that is necessary in order to benefit fully from several of the subsequent, more specialized contributions. However, the wealth of abbreviations used is a drawback in this respect; even though they are usually explained, their frequency may discourage non-specialists.

The second part of the book, "Corpus-based approaches to interlanguage", comprises three contributions. In his analysis of Swedish students' overuse of causative *make* (e.g. *make someone happy*), Bengt Altenberg stresses that "reliable interpretations of interlanguage features require thorough knowledge of the three 'languages' involved: the learner's interlanguage, his/her mother tongue and the target language" (p. 38). Altenberg argues that the overuse is due to transfer from the students' first language (L1) rather than to overgeneralization of the main English target pattern, as French students display an underuse of causative *make*. Focusing on patterns where the complement is an adjective phrase, Altenberg uses the English-Swedish Parallel Corpus to compare the two most closely corresponding constructions in English and Swedish (i.e. English causative *make* and Swedish causative *göra*, as in *göra någon lycklig* 'make someone happy'), and to relate their use to that of other options.¹ By considering Swedish and English source texts as well as analysing translations bidirectionally, he demonstrates that causative *göra* appears to be more central in Swedish than causative *make* is in English. The results thus suggest that Swedish learners' overuse of the dominant target pattern (causative *make*) is due to transfer supported by cross-linguistic similarity where the similar pattern in the L1 (causative *göra*) is even more dominant.

Karin Aijmer also looks at advanced Swedish learners in her study of the expression of modality. Aijmer uses the Swedish component of the International Corpus of Learner English (ICLE) as her main primary material, and compares the Swedish texts with similar native English material as well as with the French and German components of the ICLE. In addition to modal auxiliaries, Aijmer examines modal adverbials (e.g. *perhaps*) and modal combinations (e.g. *would probably*). Aijmer's results reveal "a generalised overuse of all the formal categories of modality examined" (p. 72); she points out, however, that not all categories of modality were included in the study.² Aijmer sees several possible reasons for this overuse, including influence from spoken English, transfer from Swedish, and the topic of the essays. She also makes several suggestions for teaching, such as studying modal auxiliaries from a discourse perspective.

The third and last contribution to this section is Alex Housen's study of Dutch- and French-speaking learners' acquisition of parts of the English verbal system: the base form, third person singular present *-s*, the *-ing* form, and regular as well as irregular preterite/past participle forms. Housen looks at spoken language produced by learners between *c.* 9 and 17 years of age, who were divided into four proficiency groups based on lexical and grammatical criteria (a reference corpus of native English is also used).³ One of the grammatical criteria "measures a speaker's morphological accuracy against the target norm" (p. 89). The inclusion of this measure may be problematic, as the proficiency groups are later used to characterize the learners with respect to a type of morphological accuracy, *viz.* their use of English verbal morphology; to the extent that the same forms were used in the criterion and in Housen's study, there is a risk of circular reasoning here. Considerable variation is found between the forms investigated as regards overuse, underuse, etc., and Housen shows that learners frequently acquire a form without yet being able to use it correctly. Housen also analyses parts of his data in order to test the Aspect Hypothesis, which predicts that learners will initially use a verb form predominantly with the verb type with which the function of the form is chiefly associated (e.g. *-ing* with activity verbs). Again, the results point to differences between the forms investigated. Housen offers several speculations, which would clearly be worth pursuing further, on why the forms investigated do not seem to be acquired in the same way: they include differences between temporal/aspectual and grammatical markers, different learning processes for regular and irregular morphology, and L1 transfer.

The third section of the volume is devoted to "[c]orpus-based approaches to foreign language pedagogy", and contains five contributions. The first and most general of these is by Fanny Meunier; it centres on the relevance of using corpora in EFL teaching with a focus on form. While she argues that findings based on native and learner corpora have not yet brought about "major changes in EFL curriculum design" (p. 124), she shows that reference tools such as dictionaries and grammars have benefited considerably from corpus-based research. In terms of teaching, the inclusion of authentic examples in textbooks and the use of data-driven learning with concordances are important developments; in this context, the use of learner English as a complement to native-speaker data is controversial, but appears to have several advantages. Meunier also lists ongoing and possible changes in grammar teaching from a short-term (e.g. data-driven learning), medium-term (e.g. using corpus linguistic methods and tools), and long-term perspective (e.g. a discourse-based rather than sentence-based view of grammar).

Angela Hasselgren addresses the issue of assessing learners' fluency. Fluency is difficult to define and describe; on the basis of previous research as well as her own investigation, Hasselgren demonstrates that so-called smallwords⁴ may be an important indicator of learners' fluency. Like Housen, Hasselgren works with spoken material, consisting of the speech of 14- and 15-year-old Norwegian learners taking a spoken interaction test, as well as that of a British control group. Her investigation shows that the Norwegian students who were judged as relatively fluent in the test situation use fewer disruptive pauses, longer utterances, and more smallwords (in terms of both types and tokens) than the less fluent students; however, their output rarely approaches that of native speakers. Drawing on relevance theory, Hasselgren also discusses how smallwords contribute to fluency by, for instance, helping to indicate the state of success of communication; this discussion further strengthens the division between more and less fluent learners.

The contribution by Ulla Connor, Kristen Precht, and Thomas Upton illustrates a textlinguistic approach to learner English. They analyse the genre of (simulated) letters of job application written by non-native and native speakers of English: undergraduates from Belgium, Finland, and the U.S. Their analysis of the letters focuses on seven genre moves, described as "semantic/functional units of texts which can be identified first because of their communicative purposes, and second because of linguistic boundaries typical of the moves" (p. 180); for example, offering to provide more information is one such move. Overall, the results suggest "a cross-cultural consensus on the use of the majority of moves" (p. 185), but a few significant differences emerge: for instance, when arguing for the application, Belgian students tend to emphasize the benefit to the applicant more than Finnish and U.S. students, who mention benefits to the hiring company more often. The authors contend that genre-specific learner corpora will be useful for teachers, in that they make it easier to assess student needs, and that textlinguistic analyses of learner data are valuable; they also suggest that analyses such as theirs can help learners by clarifying genre characteristics, in terms of what moves are expected. Their study is interesting in that it opens up a textlinguistic perspective on learner English. However, I miss tables with raw frequencies that would enable readers to study the results in more detail, especially since the total number of letters (99) is fairly low, considering that the number of rhetorical moves, rather than the number of instances of, say, a grammatical feature, was analysed.

Quentin Grant Allan's contribution concerns the TLSC (TELEC Secondary Learner Corpus, where TELEC stands for Teachers of English Language Education Centre, Hong Kong). In 2002, the corpus, which is still under development,

contained 2.2 million words of student writing. Each text is coded for a number of parameters, which makes it possible to extract more homogeneous subcorpora (e.g. argumentative writing only). The main function of the TLSC is to form the basis for “systematic linguistic analysis of areas of English in which Hong Kong secondary students experience difficulty” (p. 200). The results are used in a hypertext database on grammar and usage aimed at teachers of English in Hong Kong: for instance, corpus extracts may illustrate incorrect or unidiomatic student output pertaining to an area of grammar, together with an explanation and correct versions. There are plans to improve and expand the corpus by, for instance, providing part-of-speech tagging, adding a spoken component, and creating a concordancer that would allow teachers to explore the corpus themselves.

Barbara Seidlhofer, finally, reports on an approach she dubs “learning-driven data”, in which advanced learners analyse a corpus which they have themselves produced collectively during a course. The learners thus work with their own output, a practice Seidlhofer links to the Pushed Output Hypothesis, which states that “pushed output, i.e. sustained output that stretches the limits of learners’ current linguistic capacity, can further their development significantly” (p. 218). The students’ short written responses to the same article are conflated into corpora controlled for topic (the individual responses now being anonymous). The students then construct questions about the corpus texts, and discuss and try to answer many of these questions with the aid of corpus data. Seidlhofer argues that the students’ motivation increased significantly as a result of their working with non-threatening texts that were already familiar to them.

Computer learner corpora, second language acquisition and foreign language teaching is a valuable and important publication. It contains several studies of great interest to corpus linguists in general, but also demonstrates the relevance of examining learner corpora both outside and in the classroom, regarding, for instance, curriculum development, materials design, and data-driven learning. Probably as a result of the broad scope of the volume, the contributions differ somewhat concerning matters such as how much detail they provide, and how much background knowledge they require, as regards, for example, terminology, corpus linguistic methods, and linguistics. However, Granger (p. 28) explicitly emphasizes the need for interdisciplinarity in research on learner corpora, and in order to bring several disciplines together some differences are probably unavoidable in this respect. The overviews that introduce the contributions help to familiarize the reader with the content of each contribution; the inclusion of a Name Index and a Subject Index is also an advantage. However, a list of abbreviations used in the volume as a whole, and perhaps a

list of explanations for specialist terms used, would have made the volume even more accessible to readers from different disciplines. The division of the contributions into sections appears logical for the most part, although the third section gives a more heterogeneous impression than the other two sections. Also, given that Aijmer's and Hasselgren's contributions both analyse learner English, compare it with native English, and suggest pedagogical implications, they might have been placed in the same section. The layout is inviting and the text usually runs smoothly, with only occasional infelicities regarding punctuation and spelling. This volume will be a definite asset to readers with an interest in learner corpus research, SLA theory and/or ELT practice.

Notes

1. The most frequent alternative is the use of a synthetic verb instead of causative *make*, e.g. *make something easier* \approx *facilitate something*.
2. Aijmer also finds occasional underuse by Swedish students: for instance, they did not use root *may* in the texts examined.
3. While most learners were only interviewed once, a few were interviewed five times, at five-month intervals (p. 83). Consequently, a small number of students contributed considerably more material than the others.
4. "Smallwords" are defined as "small words and phrases, occurring with high frequency in the spoken language, that help to keep our speech flowing, yet do not contribute essentially to the message itself" (p. 150). A total of 19 smallwords (or smallword groups) were included in the study.

References

- Granger, Sylviane (ed.). 1998. *Learner English on computer*. London and New York: Longman.
- Hasselgård, Hilde. 1999. Review of: Granger (1998). *ICAME Journal* 23, 148–152.