

Marcus Callies and **Sandra Götz** (eds.). *Learner corpora in language testing and assessment* (Studies in Corpus Linguistics 70). Amsterdam/Philadelphia: John Benjamins, 2015. 220 pp. ISBN 978 90 272 0378 6 (hardback); ISBN 978 90 272 6870 9 (E-book). Reviewed by **Signe-Anita Lindgrén**, Åbo Akademi University.

Learner corpora in language testing and assessment, coedited by Marcus Callies and Sandra Götz, consists of an introduction and a collection of eight chapters written by altogether fifteen different authors. The chapters are based on papers presented in the pre-conference workshop at the 34th ICAME conference in Santiago de Compostela, Spain, in 2013, which was entitled “(Learner) Corpora and their application in language testing and assessment”, convened by the volume editors. The volume presents a selection of original studies or pilot-studies on learners in different language contexts and of different native languages (L1s).

The volume accentuates benefits and potentials of learner corpora for the testing and assessment of L2 spoken and written proficiency and draws the attention of corpus linguists and applied linguists to contact areas between learner corpus research (LCR) and language testing and assessment (LTA). It is concerned specifically with responding to the need for a revised operational definition of L2 proficiency. The authors have been requested to discuss three contact areas between LCR and LTA, which concern, first, the quality, type, and comparability of learner corpus data, second, variation within learner corpora, and, third, the concept of proficiency level within learner corpus studies. These issues are further elaborated on in the introduction, “Learner corpora in language testing and assessment: Prospects and challenges”, co-authored by the editors. The introduction is a concise, easy read. In addition to briefly setting the scene and introducing the chapters, the authors summarize three major methodological issues, in line with the above, that challenge and intrigue researchers combining learner corpora and testing and assessment: the usefulness of corpora for testing purposes and comparisons, a more thorough recognition of variability between learners, and a description of L2 proficiency based on the corpus text.

Following the introduction, the volume is divided into two sections of four chapters each. Each chapter starts with an abstract and keywords, and ends with a reference list and acknowledgements and appendices where appropriate. The chapters themselves are unnumbered, whereas the sections within the chapters are numbered. The volume ends with contact addresses to the authors and a subject index.

Section I, entitled “New corpus resources, tools and methods”, presents tailor-made corpora for LTA, tools applicable to an automatic assessment of proficiency levels, and methods for assessment and self-assessment by way of using learner corpora.

In the first chapter, “The *Marburg Corpus of Intermediate Learner English* (MILE)” (pp. 13–34), **Rolf Kreyer** introduces a written EFL corpus (currently under compilation) of timed free text exam answers to official exams in L1 German secondary schools, grades 9 to 12, which, unlike most corpora so far, will contain *longitudinal* data from learners at an *intermediate* level of proficiency. The author explains in relative detail the two kinds of mark-up used: text-internal mark-up and multi-layer text-external mark-up, as well as the error annotation. He discusses how learner corpora could contribute to addressing linguistic descriptors in the light of the *Common European Framework of Reference for Languages* (CEFR; Council of Europe 2001), and highlights questions regarding the kinds of raw data represented in learner corpora and the types of annotations, calling for proper longitudinal data from appropriate levels of proficiency, and text types or registers/genres.

In the second chapter, “*Avalingua*: Natural language processing for automatic error detection” (pp. 35–57), **Pablo Gamallo Otero**, **Marcos Garcia**, **Iria del Río**, and **Isaac González López**, present a new automatic tool for identifying and classifying spelling, lexical, and syntactic errors, including false friends and unidiomatic word combinations in written language successfully probed on L1 and L2 learners of the Galician language. They exemplify how the tool can be used to assist writing, aid individual learning, and support teaching in written proficiency assessment and state that *Avalingua* can be implemented to other languages including English “without high computational costs” (p. 36).

In the third chapter, “Data commentary in science writing: Using a small, specialized corpus for formative self-assessment practices” (pp. 59–83), **Lene Nordrum** and **Andreas Eriksson** present a corpus of discipline specific texts (currently peer-reviewed research articles, and master’s thesis from one university in Sweden, both from the field of applied chemistry) annotated with rhetorical moves and lexico-grammatical features, thus combining top-down analyses associated with discourse analysis and bottom-up corpus analysis. They suggest

three types of corpus-informed activities to develop students' genre awareness and self-assessment abilities: "(1) teacher-designed activities on moves in data commentaries, (2) teacher-designed peer-assessment activities of master's thesis corpus data, and (3) teacher- and student-initiated activities involving students' own writing" (p. 72 ff.) The learner target group is master's students and doctoral students. An evaluation of the corpus tool, the exercises, and the learning outcome is planned (p. 78).

In the last chapter in the first section, "First steps in assigning proficiency to texts in a learner corpus of computer-mediated communication" (pp. 85–112), **Tim Marchand** and **Sumie Akutsu** propose a new method for assigning proficiency levels applied to texts in a learner corpus of computer-mediated communication consisting of comments on news articles written by non-English majors at three universities in Japan as part of a compulsory English language course. The assessment criteria are based on features of the text rather than of the learner and applied through simple binary decision trees where, in the presented study, accuracy ("Are the surface features distracting?" "Are there more than two errors with grammatical form in the first fifty words?" "Are there significant errors in lexical choice?" pp. 96–97) is more weighted than fluency ("Is the word count greater than 100?" "Is the word count greater than thirty?" "Is the longest clause more than eleven tokens long?" p. 98) than complexity ("Is personal stance evident and supported?" "Is the mean sentence length more than twelve words?" "Is sophisticated vocabulary effectively used?" pp. 99–101).

The chapters in Section II, entitled "Data-driven approaches to the assessment of proficiency", question the use of institutional status as a valid definition of proficiency levels and suggest a shift of focus to the data and key-features and patterns of the individual texts. The studies use native corpora for comparison or further corpus data to validate or complement human rating.

In "The *English Vocabulary Profile* as a benchmark for assigning levels to learner corpus data" (pp. 115–140), **Agnieszka Leńko-Szymańska** explores the success of using the English Vocabulary Profile to discover the threshold that distinguishes learners of different proficiency levels and thereby to assign learners to the relevant CEFR ranks. (The EVP links words based on individual word meanings and recurrent expressions they appear in to the six level categorization of the CEFR, A1–C2). The study uses 90 EFL essays, ten per each cohort, written under comparable conditions by Austrian students in grades 5, 8, and 11, and Polish and Spanish students in grades 6, 9, and 12. All words and phrases were manually coded with the appropriate CEFR level and, when not applicable, to the categories proper names, numerals, or code-switching. For comparison, the essays were further assigned to the CEFR levels by two raters following a holis-

tic rating scheme observing descriptors for the CEFR scales. The findings indicate that the EVP is a promising instrument for assigning proficiency levels to learner corpus texts based on the analysis of their lexical content.

In the second chapter, “A multidimensional analysis of learner language during story reconstruction in interviews” (pp. 141–162), **Pascual Pérez-Paredes** and **María Sánchez-Tornel** show that the frequency of use of several linguistic features differ in picture descriptions produced by EFL learners and native speakers. The data examined were retrieved from the Spanish subcorpus of the *Louvain International Database of Spoken English Interlanguage* (LINDSEI; Gilquin *et al.* 2010), the *Louvain Corpus of Native English Conversation* (LOC-NEC) and the British component of the *Contrastive Analysis of Orality in Spoken English* (CAOS-E) corpus. The data were POS tagged and MD-analyzed by Douglas Biber. The authors encourage the use of the research methodology they applied as a complement to parameters generally used by raters and testers in assessing the quality of L2 output.

In the third chapter in the second section, “Article use and criterial features in Spanish EFL writing: A pilot study from CEFR A2 to B2 levels” (pp. 163–190), **María Belén Díez-Bedmar** explores the order and proportion of accurate and inaccurate uses of the definite, indefinite and zero articles by 26 Spanish EFL writers by combining frequency measures and an accuracy measure. The texts analyzed were part of an English high-stakes examination and rated with the same CEFR level by two raters. The statistical analyses of this pilot study identified, among other things, three linguistic features characteristic of texts at B2 level, and the study shows how findings from corpus-driven studies can be used to inform remedial teaching to address specific types of errors.

In the fourth and final chapter, “Tense and aspect errors in spoken learner English: Implications for language testing and assessment” (pp. 191–215), **Sandra Götz** explores accuracy in the German subcorpus of the LINDSEI and establishes that there is obvious heterogeneity among the English majors (3rd and 4th year of study), but also that the top five most frequently occurring error types appear systematically across all learners and make up more than half of all errors in the data. The most proficient group of learners seems to commit mainly errors belonging to the five top categories, whereas the less proficient groups also produce errors of other types. The author further discusses the most frequent and error-prone type, i.e. misuses of either tense or aspect; interference from the L1 seems to explain some of these instances. The study shows that there is variation within the advanced proficiency level with regard to the number of errors, but also that the error types found are quite systematic. This con-

sistency in error types would allow for corpus-based descriptions of learner's error profiles at different or differentiated proficiency levels.

Learner corpora in language testing and assessment successfully represents "this new emerging research field" (p. 1) and "reflects the growing importance of learner corpora in applied linguistics and second language acquisition research (SLA)" (p. 1). It exemplifies different ways in which LCR can be applied, together with limitations and caveats that face any study. This volume is a most welcomed addition to the research community of corpus linguistics and to that of applied linguistics, and will interest readers looking for applications for data-driven corpus linguistic studies and for readers focusing on both L1 and L2 proficiency and language testing and assessment. All chapters have a successful balance between detail and depth on the one hand, and brevity and relatively short chapters on the other hand. Although some chapters are based on pilot-studies and may come across as presenting more questions and limitations than customary in a rigid study, the volume forms a complete whole and offers a commendable view into the crossing between LCR and LTA. The volume sets out to "highlight the benefits and potential of using learner corpora for the testing and assessment of L2 proficiency in both speaking and writing [...reflecting] the growing importance of learner corpora in applied linguistics and second language acquisition research" (p. 1), and so it does.

A major plus of this volume is that it presents studies of both spoken and written data, and that each study is embedded in a larger context in a way that opens up the field to readers who may not be all too familiar with the fields of (learner) corpus linguistics or language testing. All but one study investigate performances in English (as a foreign or second language, sometimes compared to native speaker performances). The reader who expects to read about studies of "a variety of different L1s" (p. 5) should, however, take into account that the volume contains only eight chapters, and that most of them include speakers/writers from only one L1 background. (The L1s represented in the written data are: English, Galician, German, Japanese, Portuguese, Swedish, and EFL pupils from Austria, Poland and Spain; regarding spoken data: English, German, and Spanish.) The volume is hoped to be of "particular interest to researchers in (applied) corpus linguistics, learner corpus research, language testing and assessment, as well as for materials developers and language teachers" (back cover). I am certain that it will be of interest to these groups, and I would also include SLA researchers and more advanced L1 and FL students interested in these fields.

References

- Council of Europe. 2001. *The common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Gilquin, Gaëtanelle, Sylvie De Cock and Sylviane Granger. 2010. *Louvain International Database of Spoken English Interlanguage*. Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain.