Sandra Götz and **Joybrato Mukherjee** (eds.). *Learner corpora and language teaching* (Studies in Corpus Linguistics 92). Amsterdam/Philadelphia: John Benjamins. 2019. iv+267 pp. ISBN 978 90 272 0236 9. Reviewed by **Sylvie De Cock**, Université catholique de Louvain.

Learner corpora and language teaching brings together a selection of thematically connected papers presented at the 12th Teaching and Language Corpora Conference (TaLC) held in Giessen in 2016. All eleven contributions centre around language-pedagogical implications of learner corpus research (LCR) involving data from L2 English learners. The papers are grouped into four different sections depending on their main focus: new learner corpora and tools (Section I), written learner corpus research (Section II), spoken learner corpus research (Section III) and language teacher education (Section IV). In their introduction, the two editors, **Sandra Götz** and **Joybrato Mukherjee**, explain that the aim of the volume is to “contribute to closing the still existing gap” (p. 1) between learner corpora and language teaching, testing and assessment.

The first part of the opening paper in Section I, written by **Dana Gablasova**, **Vaclav Brezina** and **Tony McEnery**, describes the main features of the Trinity Lancaster Corpus (TLC), a recent addition to the spoken learner corpus family. The corpus, which was transcribed and developed at the Centre for Corpus Approaches to Social Science (CASS) at Lancaster University, includes over four million words of interaction collected within the framework of the Graded Examinations in Spoken English organised by Trinity College London. Particular attention is paid to the nature of the interaction between the candidates (L2 speakers from a variety of L1 backgrounds) and the examiners (L1 users of English) with a careful description of the four speaking tasks included (presentation, interactive task, discussion and conversation) as well as of the different linguistic settings created by each task in terms of interactiveness, topic familiarity and interlocutor roles. The authors also emphasise the availability of reliable and detailed information about the participants’ proficiency (CEFR band, overall performance in the band and mark for each speaking task); this type of meta-data is regrettably often lacking in learner corpora. The major contribution of

the paper to the volume lies in its discussion of how findings from research on a spoken learner corpus like the TLC can be integrated into foreign language classroom activities and teaching materials. Three TLC-based case studies that target a different aspect of learners' pragmatic knowledge in spoken interactions (i.e. expression of disagreement, ability to adjust linguistic choices according to the linguistic setting and engaged listeningship) provide concrete examples of awareness-raising activities in which the TLC learner data is given centre stage and serves as positive evidence of language use for the learners.

In the second and final paper of Section I, the spotlight is on a new tool which automatically generates questions for English tests from the Russian Error-Annotated Learner English Corpus (REALEC, National Research University High School of Economics), a manually error-tagged corpus containing essays written by Russian first- and second-year Bachelor students. After describing and illustrating the error annotation procedure and discussing the advantages of error annotation in REALEC for both teachers and learners, **Olga Vinogradova** details the automatic test making process made possible thanks to the REALEC English Test Maker (RETM) and explains how the program handles some of the challenges that arise from special error tagging cases. Not only does RETM automatically generate questions from the annotated corpus, but it also creates tests from the generated questions in Moodle XML format, which can be made available to students after manual editing. Although fairly time-consuming especially for tests that target a wide range of testing areas, the manual editing process is essential to delete inappropriate questions (e.g. when the mistake is too difficult to detect) and to improve questions that need fine-tuning. Vinogradova offers a number of practical recommendations for the editing process and argues that the time teachers spend producing and editing tests thanks to RETM is a worthwhile investment provided insights from the error-annotated corpus are a central part of their classes.

In the first paper of Section II, **Albert Biel** looks into intermediate and advanced German EFL learners' depth of vocabulary knowledge by zooming in on the syntactic structures and collocations featuring the high-frequency verb *TAKE* in the German component of the International Corpus of Learner English (ICLE), compared with a reference corpus, namely the American English part of the Louvain Corpus of Native English Essays (LOCNESS). The results of Biel's examination of six categories of syntactic patterns with *TAKE* (monotransitive, complex-transitive, phrasal verbs, compounds and derivatives, idioms, errors) in the learner and native data contradict the hypothesis that the learners tend to overuse simple syntactic structures: there are no statistically significant differences between the two groups except for the error category, with learners often

using the wrong tense or the wrong preposition in complex phrasal constructions (*take advantage from, take care for*). Biel's findings also indicate that the German learners demonstrate collocational competence as their use of collocations with TAKE exhibit strong similarities with that of the native speaker students. To conclude, the author calls for teachers to focus more on phrasal verbs and complex phrasal constructions and stresses instructors' key role in raising learners' awareness of the whole potential of words.

The specific objective of **Meredith D'Arienzo's** quantitative and qualitative analysis of a set of linking adjuncts in the Italian component of ICLE and in LOCNESS is to identify learners' uses that can be traced to cross-linguistic influence from their L1. To this end, the study also examines a group of 'target linking adjuncts' (e.g. *consequently, eventually*) selected because they may be overused and/or misused by Italian learners as a result of the existence of cognates or false cognates in Italian. The paper gives two concrete examples of how findings from the study could be fed into teaching and reference materials designed specifically to cater for the needs of Italian learners (particularly when their EFL teachers do not speak Italian). The first is a noticing activity in which Italian learners are instructed to examine sentences containing the frequently misused *eventually* (taken from LOCNESS), to focus on the meaning of the word and compare it to that of the Italian false friend *eventualmente* and to produce three sentences that showcase the correct meaning of the word. The second resource is a 'help box' that could feature in a monolingual learners' dictionary targeted at Italian EFL learners. It is designed to address learners' overuse of *consequently* (presumably because of the Italian cognate *conseguentemente*) and to help them expand their range of resultative linking adjuncts through an examination of alternatives to *consequently* in sentences from LOCNESS.

In their exploratory paper, **Pascual Pérez-Paredes** and **María Bélen Díez-Bedmar** set out to uncover the possible contribution of two different methods of automatic analysis, namely POS keyword analysis and syntactic complexity analysis, when seeking to gain insights into different developmental stages of EFL instruction and to identify criterial features that distinguish written productions by lower- and higher-level learners. The cross-sectional data used in the study comes from the Spanish component of the International Corpus of Crosslinguistic Interlanguage (ICCI) and includes essays on the same topic written by pupils in the first two years of secondary education and in the two years of non-compulsory secondary education before university. The two automatic methods made it possible to identify the POS tags that are statistically more frequent in each corpus and to calculate fine-grained indices of phrasal complexity and of syntactic sophistication (based on the notion of Verb Argument Construc-

tions). Traditional measures of syntactic complexity (e.g. mean length of sentences, number of clauses per sentences) were also added to complete the picture. Although the study would need to be complemented with an analysis of the POS tagger and parser's performance on (low-level) learner data, the approach adopted sheds light on a series of features that can be connected with complexity development (e.g. pronouns vs. noun heads, nouns as modifiers, prepositional phrase postmodification, number of clauses per sentence). Based on their findings, the authors mention possible teaching interventions to support the development of complexity and sophistication in Spanish pupils' writing. These include explaining the substitution processes involved in the use of pronouns, providing linguistic input featuring elaborated noun phrases or encouraging the learners to try out various verb complementation patterns.

The final paper of the section on 'Written Learner Corpora and Language Teaching' is devoted to learners' academic writing. **Leonie Wiemeyer** seeks to shed light on how German BA undergraduate students of English handle intertextuality and more specifically direct quotation in academic English by scrutinising ten research papers on pragmatics written by 2nd and 3rd year students of English from the Corpus of Academic Learner English (CALE). The quantitative and qualitative analyses of the data are made possible thanks to the manual coding of intertextual links (e.g. direct quotes, summaries) and additional characteristics (e.g. use of reporting structures, evaluation) using a fine-grained annotation scheme (included on p. 136). While the learners in the corpus are clearly familiar with referencing conventions, the findings also point to their insecurity and lack of experience as novice writers (e.g. high frequency of direct quotation, overciting, restricted range of reporting structures) as well as strong individual preferences. Given the highly individual character of source-based writing, Wiemeyer recommends a local approach to the teaching of academic writing with teachers analysing corpora of their own students' productions to inform exercises that would truly meet their needs. She advocates a combination of awareness-raising and data-driven activities using both student and expert texts.

Mariko Abe's paper serves as a bridge between Sections II (focussing on written learner corpora) and III (spoken learner corpora) as it reports on a study that combines both written and spoken corpus data. Manually error-tagged spoken and written learner productions by Japanese EFL learners are compared to explore the extent to which production modes affect accuracy and error rates across three proficiency levels. The underlying aim of the analysis is to identify the errors that arguably require particular teaching time and effort as they do not disappear when more attention is paid to form during a written task. While the

spoken data comes from the National Institute of Information and Communications Technology Japanese Learner English (NICT JLE) Corpus, the written data was elicited using the same single picture description task as in the spoken corpus. It emerges that one item that would warrant particular teaching focus is third-person singular *-s* because it is one of the only items under investigation whose accuracy rate does not improve in writing. According to Abe, this finding suggests that a linguistic feature can only be affected by production modes if it is used accurately to a certain degree and stresses that teachers should provide learners with ample monitoring and automatizing practice opportunities in the classroom (e.g. using a pair-work format followed by writing activities with the guidance of a check-list) to help them reach that stage. Abe's contribution is followed by three studies that target features that are specific to spoken language.

In his paper, **Tomáš Gráf** investigates the effect of elicitation-task design on speech rate (measured in words per minute) in the Czech EFL learner subcorpus of the Louvain International Database of Spoken English Interlanguage (LINDSEI) and in the Louvain Corpus of Native English Conversation (LOCNEC). In addition, a smaller corpus of interviews in Czech by 31 of the 50 same speakers as in the LINDSEI Czech component is used to explore similarities between the learners' speech rate in English and in their L1. The study uncovers statistically significant differences between the monologic story reconstruction task based on a set of four pictures and the other two tasks (a largely monologic task on a set topic and a dialogic free discussion on everyday topics): speech rate is markedly slower in the picture-based story reconstruction task in all three corpora. An examination of individual learner variation confirms that this is true for the majority of the native speakers and the learners in their L1 and L2 (with some rare exceptions). As rightly pointed out by Gráf, these findings have a direct pedagogical implication because picture descriptions and picture-based storytelling activities are often used in the classroom and in the assessment of speech. He stresses that teachers and examiners who assess spoken skills and especially fluency need to be fully aware that picture-based story reconstructions are cognitively challenging tasks (even in a speaker's L1) which tend to elicit slower and less fluent speech.

With her paper on intonation in the speech of advanced EFL learners, which reports on some of the research conducted as part of her PhD project, **Karin Puga** ventures into the largely unexplored territory of learner prosody using learner corpus data and a Contrastive Interlanguage Analysis (CIA) approach. Her specific focus is on German and Spanish EFL learners' intonational deviances of edge tones within intonation units in the German and Spanish components of LINDSEI and in LOCNEC. Given the labour-intensive manual

prosodic annotation required for the analysis, the study is limited to ten speakers from each of the three corpora and to 500 words per speaker. The selection of the data (to ensure comparability) and the annotation procedure are carefully documented. The analysis reveals that, while the native speakers in the reference corpus essentially make use of falls within intonation phrases and intermediate phrases, the German learners tend to use rising tones and the Spanish learners frequently make use of falling tones. In view of her findings and of the fact that teachers tend to serve as models for their pupils, Puga highlights the need for prosodic training in teacher education and lists a series of awareness-raising and self-monitoring activities that foreign language teachers could use in the classroom.

Anna Rosen deals with the possible impact of acquisition context, and more specifically of instructional setting with or without additional natural exposure, on German learners' use of a set of 18 fluency-enhancing smallwords (e.g. *sort of, a bit, you know*) in the informal interviews included in LINDSEI. To this end, the German component of LINDSEI is divided into two groups of learners: those with and those without a stay abroad. Learners' use of the smallwords is compared with that of native speakers from two different reference corpora, namely LOCNEC and the Jersey Interview Corpus (which incidentally offers interesting insights into the possible effect the choice of a control corpus may have on findings). A rather complex picture emerges from Rosen's quantitative and qualitative analyses: while, on the whole, smallwords are used more frequently by the learners who went abroad than by those who did not, the two groups of learners can also be seen to overuse some smallwords or to use them to the same extent as native speakers and additional natural exposure appears to have only a limited (positive) impact on the pragmatic and syntactic use of the items under study. She concludes that her results suggest that smallwords are not all equally easy/difficult to acquire (with or without a stay abroad) and that the cases of similar use across the various groups of speakers seem to indicate that some smallwords can actually be taught and acquired in instructional settings. She recommends the development (and evaluation) of teaching material that would explicitly concentrate on the appropriate use of smallwords.

The only paper in the final section of the volume addresses the key issue of corpus literacy in language teacher education. **Marcus Callies** reports on findings from a small-scale online survey completed by 26 German secondary-school teachers of English in 2016, 15 years after a similar questionnaire-based study by Mukherjee (2004). The aim was to update Mukherjee's results following the increase in linguistic classes dealing with corpus linguistics in the curricula of modern language students in Germany since the early 2000s. The aspects

explored in the survey include teachers' familiarity and experience with corpus resources and tools, their actual use of corpora in EFL teaching, as well as the benefits of corpus use for teachers and learners. Although around a third of the participants learned about corpus linguistics at university, very few of them use corpora regularly and, when they do use them, it is mainly to check language acceptability or when marking. It also emerges that corpora are seldom used in learner-centred activities and that the teachers do not collect their own local learner corpora. Callies calls for learner and reference corpora to become part of the toolbox of language teachers and describes a two-hour module for (future) teachers (taught as part of the teacher training program at the University of Bremen) designed to make "teacher students learn about the use of corpora both from a researcher's, teacher's and learner's perspective" (p. 253). A convincing case is made for training the teachers with learner data that is fairly close to what they will experience as teachers (e.g. a corpus like ICCI could be used in a primary or secondary school teaching context).

LCR is now a well-established field of linguistic investigation (Granger et al. 2015). The papers contained in Götz and Mukherjee's volume foreground different facets of the interplay between learner corpora and language teaching. They demonstrate that the language-pedagogical implications of LCR can come in many shapes and sizes, from providing valuable insights into what learners can and cannot do at various stages of language development as well as insights into the effect of elicitation task design on fluency-related aspects, to concrete pedagogical applications like classroom activities (with learner data also serving as positive evidence of use), help boxes in reference tools, test-making tools or teacher training modules. Some of the pedagogical suggestions made by the authors underline the key role of teacher education (e.g. if teachers are to make use of local corpora to produce materials adapted to their pupils'/students' specific needs). Although there is still room for considerable development in LCR (a number of challenges are discussed), the contributions in the volume indicate that the 'still existing gap' between learner corpora and language teaching mentioned by Götz and Mukherjee in their introduction is narrowing to the benefit of learners and teachers alike. In view of the papers' wide variety of approaches and foci, *Learner corpora and language teaching* is a must-have for researchers interested in LCR and SLA, language teachers, teachers of English for Academic Purposes, teacher trainers, language examiners and materials developers.

References

- Granger, Sylviane, Gaëtanelle Gilquin and Fanny Meunier (eds.). 2015. *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press.
- Mukherjee, Joybrato. 2004. Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor and Th. A. Upton (eds.). *Applied Corpus Linguistics. A multi-dimensional perspective*, 239–250. Amsterdam: Rodopi.