

John Benjamins Publishing Company



This is a contribution from International Journal of Learner Corpus Research 6:2
© 2020. John Benjamins Publishing Company

This electronic file may not be altered in any way. The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/content/customers/rights>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

EDITORIAL

Promoting methodological expertise, transparency, replication, and cumulative learning

Introducing new manuscript types in the
International Journal of Learner Corpus Research

Magali Paquot and Marcus Callies

As repeatedly discussed in the literature, Learner Corpus Research (LCR) is a field of enquiry whose potential will only unlock fully if the community is prepared to take further steps towards methodological reform (Paquot & Plonsky, 2017; Gries, 2018) and if more and more diverse learner corpora are compiled (e.g. Myles, 2015). Corpus compilation, however, is an extremely time-consuming task and requires a high level of expertise in many different areas which are rarely described and discussed sufficiently (e.g. protocols and ethical/legal procedures for data collection, or data transcription and annotation guidelines). Similarly, methodological transparency and documentation are key at all stages of a corpus-linguistic study, ranging from corpus compilation to the qualitative and/or quantitative analysis of corpus data, but these crucial aspects of the research process often do not receive the attention they deserve.

We believe it is essential to promote transparency and documentation of research methodology and to give researchers engaged in advancing our methodological expertise in the field the recognition they deserve. One way to achieve this is by publishing data reports that introduce open-access corpora and other types of publications on research design and methodology in academic journals. This is all the more important as the transparent documentation of methodology and data sharing are a prerequisite for reproducibility and replicability, which are considered by many to be fundamental for the progress of science (Marsden & Plonsky, 2018). Enhanced methodological discussion about the various steps involved in corpus compilation should also help answer repeated calls for more standardization and the development of best practices and shared protocols for data collection and annotation (Granger & Paquot, 2017; Tracy-Ventura, Paquot, & Myles, in press).

In order to further support and promote this growing trend towards methodological awareness and reflection and to give LCR researchers a forum to publish and gain recognition for their work in corpus development, compilation, and annotation, as well as their development of other research material and software tools, we have decided to introduce several new manuscript types in the *International Journal of Learner Corpus Research* (IJLCR) as of 2020. These are as follows:¹

- **CORPUS REPORTS** present a detailed description of new learner corpora (i.e. corpus design, collection, transcription, annotation and distribution). Such reports should focus on learner corpora that are original in their design and/or construction, widely accessible (stored online or available via a data repository) and with high potential of use by the research community.
- **MATERIALS AND METHODS REPORTS** provide details of the methods and protocols developed and materials used during a research cycle (e.g. corpus annotation schemes, learner questionnaires). Researchers can co-submit their report to IJLCR together with the original research article.
- **SOFTWARE REPORTS** present a detailed description of new software tools or sets of code and how these can be used to compile, annotate and/or analyze learner corpus data. Software reports should focus on tools that offer researchers new ways to compile, annotate or analyse a learner corpus and are available to the research community.
- **SHARED TASK REPORTS** present the cumulative results of shared tasks organized to promote research advancement by solving a problem of general interest to the community on the basis of learner corpus data (e.g. error identification and correction, native language identification).

One SHARED TASK REPORT was already published in issue 6(1): Ballier, Canu, Petitjean, Casso, Balhana, Alexopoulou, and Gaillat (2020) provide a summary of the CAP 2018 Machine Learning (ML) competition whose goal was to produce a machine learning system to predict learners' proficiency levels from written texts sampled from the L1 French component of the *EF Cambridge Open Language Database* (EFCAMDAT; Geertzen, Alexopoulou, & Korhonen, 2013). The current issue now features a CORPUS REPORT where Shatz describes the way he sampled and improved the EFCAMDAT in various ways with the help of Natural Language Processing techniques. We are also happy to include a MATERIALS AND

1. See our online guidelines (<https://benjamins.com/catalog/ijlcr>) for more information about the different manuscript types, as well as other new submission requirements related, among other things, to data citation and the reporting of corpus research findings, and Open Science initiatives encouraged by the journal.

METHODS REPORT where Larsson, Paquot, and Plonsky discuss the issue of inter-rater reliability based on insights from a collaborative study on adverb placement.

LCR is entering its fourth decade. We believe there is already scope for the field to try and reproduce findings from previous research in the form of **REPLICATION STUDIES** which have received growing importance in the field of Second Language Acquisition. The time is also ripe for assessing our methodological practices more systematically (see Paquot & Plonsky, 2017), and taking stock of learner corpus-based empirical findings, thus synthesizing the accumulated state of knowledge. As discussed by Mizumoto, Plonsky, and Egbert (in press), “massive potential exists for meta-analysis as a means to synthesize findings systematically” in corpus linguistics. More generally, there is a need for more discussion about our conceptual, theoretical, and methodological apparatus.

With these specific goals in mind, we also invite the following new types of manuscripts:

- **REPLICATION STUDIES** are empirical studies motivated by a previously published study (not necessarily published in the IJLCR). They reproduce and/or extend the methodology proposed in an initial study of significant impact and of credible methodological rigour within its domain of research. They aim to verify findings and/or examine the generalizability of the insights obtained in earlier learner corpus studies.
- **REVIEW ARTICLES** are systematic state-of-the-art analyses of research published in the field. They offer a summary of findings and make critical observations on published research to date. Manuscripts submitted in this category can be of two kinds: (1) reviews of study quality in LCR and (2) meta-analyses in a particular domain of LCR.
- **POSITION PAPERS** are shorter essays (i.e. papers not based on original empirical research) motivated by current theoretical, methodological and/or educational issues that are of general interest to the community. We also welcome responses to position papers.

We sincerely hope that these new publication formats will be welcomed by the community and that the IJLCR will contribute to advance the conceptual, theoretical, methodological and empirical state of the art in the field.

Magali Paquot & Marcus Callies
IJLCR General Editors

References

- Ballier, N., Canu, S., Petitjean, C., Casso, G., Balhana, C., Alexopoulou, T., & Gaillat, T. (2020). Machine learning for learner English: A plea for creating learner data challenges. *International Journal of Learner Corpus Research*, 6(1), 72–103. <https://doi.org/10.1075/ijlcr.18012.bal>
- Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge open language database (EFCAMDAT). In *Proceedings of the 31st Second Language Research Forum*. Somerville, MA: Cascadia Proceedings Project.
- Granger, S., & Paquot, M. (2017). *Towards standardization of metadata for L2 corpora*. Invited talk at the CLARIN workshop on Interoperability of Second Language Resources and Tools, 6–8 December 2017, University of Gothenburg, Sweden. Retrieved from https://sweclarin.se/sites/sweclarin.se/files/event_atachements/Granger_Paquot_Metadata_G%C3%B6teborg_final.pdf (accessed 1 September 2020).
- Gries, S. Th. (2018). On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies*, 1(2), 276–308. <https://doi.org/10.1075/jsls.00005.gri>
- Marsden, E., & Plonsky, L. (2018). Conclusion: Data, open science, and methodological reform in second language acquisition research. In A. Gudmestad & A. Edmonds (Eds.), *Critical reflections on data in second language acquisition* (pp. 219–228). Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.51.10mar>
- Mizumoto, A., Plonsky, L., & Egbert, J. (in press). Meta-analyzing corpus linguistic research. In M. Paquot & S. Th. Gries (Eds.), *The Practical Handbook of Corpus Linguistics*. Berlin: Springer.
- Myles, F. (2015). Second language acquisition theory and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 309–332). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.014>
- Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, 3(1), 61–94. <https://doi.org/10.1075/ijlcr.3.1.03paq>
- Tracy-Ventura, N., Paquot, M., & Myles, F. (in press). The future of corpora in SLA. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge Handbook of Second Language Acquisition and Corpora*. Abingdon: Routledge.