

Webinar Title:

Creating Data-Driven Pedagogical Materials Using Learner Corpora: A Guide for
Language Teachers

Learner Corpora

- The Multilingual Corpus of Second Language Speech (**MuSSeL**; release date: May 31st, 2021) https://l2trec.utah.edu/multi-Lingual_Speech_Corpus.php
- The Written Corpus of Utah Dual Language Immersion (**CUDLI**; release date: Aug 30th, 2021) <https://l2trec.utah.edu/corpus-of-utah-dual-language-immersion/>
- The Multilingual Academic Corpus of Assignments – Writing & Speech (**MACAWS**) <https://sites.google.com/email.arizona.edu/macawswebinar/home>
- Corpus & Repository of Writing (**CROW**) <http://crow.corporaproject.org>
- Michigan Corpus of Upper-Level Student Papers (**MICUSP**) <https://elicorpora.info/>
The International Corpus Network of Asian Learners of English (**ICNALE**)
http://language.sakura.ne.jp/icnale/icnale_online.html
- English as a Lingua Franca Corpus (**ELFA**)
<https://www2.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/elfa-corpus>
- Learner Corpora around the World <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>
- Spanish Learner Language Oral Corpora (**SPLLOC**) <http://www.splloc.soton.ac.uk/>
- Corpus del Español (BYU) <http://www.corpusdelespanol.org/>
- French Learner Language Oral Corpora (**FLLOC**) <http://www.flloc.soton.ac.uk/>
- KanDeL (The Kansas Developmental Learner corpus - German) <https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/research/kandel>

Other Corpora

- The English Corpora <https://www.english-corpora.org/>
- Corpus of Contemporary American English (**COCA**) <https://www.english-corpora.org/coca/>
- Michigan Corpus of Academic Spoken English (**MICASE**)
<https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>

- Corpus of Journal Articles (**CJA**) <http://rcpce.engl.polyu.edu.hk/CJA2014/default.htm>
- Corpus of Historical American English (**COHA**) <https://www.english-corpora.org/coha/>
- British Academic Written English Corpus (**BAWE**)
<https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/>
- British National Corpus (**BNC**)-Spoken and Written <https://www.english-corpora.org/bnc/>
- Business Letter Corpus (**BLC**) <http://www.someya-net.com/concordancer/>
- State of the Union Corpus: <http://stateoftheunion.onetwothree.net/>
- Time Magazine Corpus: <https://www.english-corpora.org/time/>
- Wikipedia Corpus: <https://www.english-corpora.org/wiki/>
- Lists of Corpora by Language <https://www.sketchengine.eu/corpora-and-languages/>
- Lists of Specialized Corpora in Non-English Languages <https://www ldc.upenn.edu/new-corpora>
- Persian Speech Corpus <http://fa.persianspeechcorpus.com/>

Concordancing Program

- AntConc Download: <https://www.laurenceanthony.net/software/antconc/>
- AntConc Tutorial Videos by Laurence Anthony
https://www.youtube.com/playlist?list=PLiRIDpYmiC0Ta0-Hdvc1D7hG6dmiS_TZj

Corpus-Informed Sources for Language Teachers

- Crow Resources for English Teachers <https://writecrow.org/crow-for-teachers/>
- MICASE Kibbitzers
<https://web.archive.org/web/20120328132411/http://micase.elicorpora.info/micase-kibbitzers>
- Using Corpora to Teach German to English-Speaking Learners <https://corpora.ku.edu/>

References

- Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and Writing in the University: A Multidimensional Comparison. *TESOL Quarterly*, 36(1), 9–48. <https://doi.org/10.2307/3588359>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman Grammar of Spoken and Written English. In *Pearson Education Limited*.
- Bogorevich, V., & Kia, E. (In press). Corpus-Informed materials for teaching academic group discussions. In V. Viana (Ed.), *New ways in teaching with corpora*. Alexandria, VA: TESOL International Association.
- Bunting, J., Diniz, L., & Reppen, R. (2019). *Grammar and Beyond Essentials 4*. Cambridge University Press.
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 359-370). Routledge. <https://doi.org/10.4324/9780203856949>
- Götz, S. (2020). Analyzing a learner corpus with a concordancer. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (1st ed., pp. 68-89). Routledge. <https://doi.org/10.4324/9781351137904>
- Ishikawa, S. (2014). Design of the ICNALE Spoken: A new database for multi-modal contrastive interlanguage analysis. *Learner corpus studies in Asia and the world*, 2, 63-76.
- McCarthy, M., & O'Keeffe, A. (Eds.). (2010). *The Routledge Handbook of Corpus Linguistics* (1st ed.). Routledge. <https://doi.org/10.4324/9780203856949>
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Shin, J., Velázquez, A. J., Swatek, A., Staples, S., & Partridge, R. S. (2018). Examining the Effectiveness of Corpus-Informed Instruction of Reporting Verbs in L2 First-Year College Writing. *L2 Journal*, 10(3). <https://doi.org/10.5070/1210337022>

Simpson, R. C., Lee, D. Y., Leicher, S., & Ädel, A. (2002). MICASE manual. *Ann Arbor, MI: English Language Institute, The University of Michigan.*

Staples, S., Novikov, A., Picoral, A., Sommer-Farias, B. (2019-). Multilingual Academic Corpus of Assignments - Writing and Speech.

The following sources aim to address a few of the questions asked in the Q&A section:

Automatic Error Tagging

Díaz-Negrillo, A., & García-Cumbreras, M. A. (2007). A tagging tool for error analysis on learner corpora. *ICAME Journal*, 31(1), 197–203.

Campillos Llanos, L. (2014). A Spanish learner oral corpus for computer-aided error analysis. *Corpora*, 9(2), 207–238. <https://doi.org/10.3366/cor.2014.0058>

Crosthwaite, p. (2013). An error analysis of L2 English discourse reference through learner corpora analysis. *Linguistic Research*, 30(2), 163–193.
<https://doi.org/10.17250/khisli.30.2.201308.002>

Rayson, P., & Baron, A. (2011). Automatic error tagging of spelling mistakes in learner corpora. *A Taste for Corpora: In Honour of Sylviane Granger*, 109–126.
<https://doi.org/10.1075/scl.45.09ray>

POS Taggers for African Languages

Adedjouma, S., Aoga, J., & Igue, M. A. (2013) Part-of-speech tagging of yoruba standard language of niger-congo family. *Research Journal of Computer and Information Technology Sciences*, 1(1), 2–5 <http://hdl.handle.net/2078/211882>

POS Tagger for Urdu

https://www.cle.org.pk/software/langproc/POS_tagger.htm