Chapter 8

Embracing Computer Corpora in the Language Learning Classroom and Using It in Your Classroom

Sally Durand
University of Illinois at Chicago, USA

ABSTRACT

Among the many facets of Computer Assisted Language Learning (CALL), studies using computer language corpora have risen to considerable prominence in research agendas. The author argues that corpora are useful tools for practicing teachers. However, the myth that "corpus-based research is too complicated to be useful for teachers" (Conrad, 2009) prevails in pedagogical contexts. This chapter strives to dispel that myth by synthesizing a wealth of research and their accompanying pedagogical applications. Secondly, it shares specific pedagogical activities to implement corpus data in classroom teaching. These corpus-informed classroom strategies provide concrete examples that will assist TESOL/ TEFL teachers in making their coursework authentic and therefore more meaningful to students learning language.

INTRODUCTION

What is a corpus? What is corpus-based research? When I gave my first presentation on the topic of using corpora in the classroom at Illinois' TESOL conference in early 2014, two attendees answered my opening question, *"What do corpora in the classroom mean?"* with: *"Our bodies, using the body in the classroom?"* While my presentation title and topic may have misled those two guests, the word *corpus,* in fact, does mean body in Latin. A corpus is a collection of texts and speech (written) comprised for the purpose of linguistic analysis. Corpus linguistics has roots in theology. Religious scribes scoured their bibles, coding certain words or themes, making lists of their locations together in a kind of index, in other words, collecting a corpus. Imagine how those early scholars would have benefitted from our

modern technology. In his book, *An Introduction to Corpus Linguistics*, Kennedy (1998) attributes the modern field of corpus linguistics to the capabilities offered by computerized databases. As we will see shortly, interest in corpus data has spread beyond linguists to pedagogical researchers and teachers. While linguists study language corpus data to better understand the language, corpus-informed/based language learning researchers examine corpus data to understand how language is learned and taught. This pedagogically-oriented area of inquiry, made possible by computers, is the epitome of Computer Assisted Language Learning.

The creation of computer technology has made corpus research possible, as the computer is able to store, code, categorize, and retrieve massive amounts of information. The field began with just a small group of dedicated enthusiasts, but by the end of the century, it had become a major research strand in both the fields of linguistics and language learning. Over the last two decades, educators have been finding ways to exploit the potential of corpora (the plural of corpus), channeling research data into pedagogical applications for language instruction. Research using corpus data in English language learning has become a broad field drawing from a wealth of corpora. This abundance and breadth of corpora and corpus-based research studies may contribute to the myth that it is all just too complicated for teachers (Conrad, 2009). Some teachers believe that the technology is too advanced or that the data are incomprehensible. These may be common sentiments and they are easily understandable. The sheer mass of a given database alone can be overwhelming. In my experience discussing corpora use in the classroom with peers, I have found that a few teachers are unaware of the existence of corpora. Many are vaguely to somewhat familiar with the concept of corpora and generally aware of the growing interest and research in corpus-informed language learning. Most have limited to no hands-on experience working with corpora. It is noted in the field that academic preparation programs lack any training in corpus linguistics (Granger, 2009; Hasko, 2013).

Many research studies use prominent corpora such as the British National Corpus (BNC), and the Corpus of Contemporary American English, usually referred to as COCA (Davies, 2008). These immensely huge corpora contain hundreds of millions of words from sources of news broadcasts to literary fiction. These resources allow for the analysis of language features and use such as medium and register by searches for spoken versus written language use, for example. Corpora also consist of professional and academic content, facilitating research in academic disciplines or subfields, and enabling comparisons on aspects of general English use versus English for specialized or specific purposes. Learner corpora provide insights into the language use of language learners as they collect actual learners' work. One prominent corpus of this kind is the Michigan Corpus of Upper level Student Papers (MICUSP), from the University of Michigan. Another is the International Corpus of Learner English (ICLE). Learner corpora provide insight into learners' needs and actual habits of learners' language use. The features of corpora seem limitless as does the pedagogical potential. For example, along with generating millions of words in frequency lists, certain corpora are capable of charting changes in usage patterns over time, or the geographical status of varied world Englishes.

It is not necessary for teachers to become extensively familiar with the range and navigation of extant corpora, or even one corpus, in order to benefit from this technology and area of research. Therefore, the goal of this chapter is to address the following two questions: How have corpora been used for pedagogical purposes by researchers and teachers? How can teachers use corpus data in their language learning classrooms?

16 more pages are available in the full version of this document, which may
be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/embracing-computer-corpora-in-the-language-
learning-classroom-and-using-it-in-your-classroom/188181

Related Content

Online Teaching During COVID-19: A Tale of Two Cities

Adrian Ting, Karen A. Manaigand Alberto D. Yazon (2024). *International Journal of Web-Based Learning and Teaching Technologies (pp. 1-15).*

www.irma-international.org/article/online-teaching-during-covid-19/336545

Conducting Programmatic Assessments of Online Writing Instruction: CCCC's OWI Principles in Practice

Nicki Litherland Bakerand Elisabeth H. Buck (2017). *Handbook of Research on Writing and Composing in the Age of MOOCs (pp. 385-405).*

www.irma-international.org/chapter/conducting-programmatic-assessments-of-online-writing-instruction/172598

Teaching Tradeoffs

Sorin Walter Gudea (2008). *Expectations and Demands in Online Teaching: Practical Experiences (pp. 162-182).*

www.irma-international.org/chapter/teaching-tradeoffs/18667

Research on Small Group Writing Activities Compared to Individual Flipped Model Through WhatsApp: Writing Skill

Yudhi Arifani (2022). *International Journal of Web-Based Learning and Teaching Technologies (pp. 1-10).*

www.irma-international.org/article/research-on-small-group-writing-activities-compared-to-individual-flipped-model-through-whatsapp/281724

Impact of Student Engagement in Online Learning Environments

A. V. Senthil Kumarand P. V. Praveen Sundar (2018). *Optimizing Student Engagement in Online Learning Environments (pp. 1-27).*

www.irma-international.org/chapter/impact-of-student-engagement-in-online-learning-environments/192446