

Cambridge University Press

978-0-521-76987-7 - Using Corpora in the Language Classroom

Randi Reppen

Frontmatter

[More information](#)

---

## Using Corpora in the Language Classroom

Cambridge University Press  
978-0-521-76987-7 - Using Corpora in the Language Classroom  
Randi Reppen  
Frontmatter  
[More information](#)

---

CAMBRIDGE LANGUAGE EDUCATION

Series Editor: Jack C. Richards

*In this series:*

**Agendas for Second Language Literacy** *by Sandra Lee McKay*

**Reflective Teaching in Second Language Classrooms** *by Jack C. Richards and Charles Lockhart*

**Educating Second Language Children: The Whole Child, the Whole Curriculum, the Whole Community** *edited by Fred Genesee*

**Understanding Communication in Second Language Classrooms** *by Karen E. Johnson*

**The Self-Directed Teacher: Managing the Learning Process** *by David Nunan and Clarice Lamb*

**Functional English Grammar: An Introduction for Second Language Teachers** *by Graham Lock*

**Teachers as Course Developers** *edited by Kathleen Graves*

**Classroom-Based Evaluation in Second Language Education** *by Fred Genesee and John A. Upshur*

**From Reader to Reading Teacher: Issues and Strategies for Second Language Classrooms** *by Jo Ann Aebersold and Mary Lee Field*

**Extensive Reading in the Second Language Classroom** *by Richard R. Day and Julian Bamford*

**Language Teaching Awareness: A Guide to Exploring Beliefs and Practices** *by Jerry G. Gebhard and Robert Oprandy*

**Vocabulary in Language Teaching** *by Norbert Schmitt*

**Curriculum Development in Language Teaching** *by Jack C. Richards*

**Teachers' Narrative Inquiry as Professional Development** *by Karen E. Johnson and Paula R. Golombek*

**A Practicum in TESOL** *by Graham Crookes*

**Second Language Listening: Theory and Practice** *by John Flowerdew and Lindsay Miller*

**Professional Development for Language Teachers: Strategies for Teacher Learning** *by Jack C. Richards and Thomas S. C. Farrell*

**Second Language Writing** *by Ken Hyland*

**Cooperative Learning and Second Language Teaching** *edited by Steven G. McCafferty, George M. Jacobs, and Ana Christina DaSilva Iddings*

**Using Corpora in the Language Classroom** *by Randi Reppen*

**English Language Teaching Materials: Theory and Practice** *edited by Nigel Harwood*

Cambridge University Press

978-0-521-76987-7 - Using Corpora in the Language Classroom

Randi Reppen

Frontmatter

[More information](#)

---

# Using Corpora in the Language Classroom

*Randi Reppen*

Northern Arizona University



CAMBRIDGE  
UNIVERSITY PRESS

Cambridge University Press  
 978-0-521-76987-7 - Using Corpora in the Language Classroom  
 Randi Reppen  
 Frontmatter  
[More information](#)

CAMBRIDGE UNIVERSITY PRESS  
 Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,  
 São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press  
 32 Avenue of the Americas, New York, NY 10013-2473, USA  
[www.cambridge.org](http://www.cambridge.org)  
 Information on this title: [www.cambridge.org/9780521146081](http://www.cambridge.org/9780521146081)

© Cambridge University Press 2010

This publication is in copyright. Subject to statutory exception  
 and to the provisions of relevant collective licensing agreements,  
 no reproduction of any part may take place without the written  
 permission of Cambridge University Press.

It is normally necessary for written permission for copying to be obtained  
*in advance* from a publisher. The lists in Appendix A on pages 74, 75, and  
 76 of this book are designed to be copied and distributed in class. The normal  
 requirements are waived here, and it is not necessary to write to Cambridge  
 University Press for permission for an individual teacher to make copies for  
 use within his or her own classroom. Only those pages which carry the wording  
 '© Cambridge University Press' may be copied.

First published 2010

Printed in the United States of America

*A catalog record for this publication is available from the British Library.*

*Library of Congress Cataloging-in-Publication data*

Reppen, Randi.  
 Using corpora in the language classroom / Randi Reppen.  
 p. cm. – (Cambridge language education)  
 Includes bibliographical references and index.  
 ISBN 978-0-521-76987-7 – ISBN 978-0-521-14608-1 (pbk.)  
 I. Corpora (Linguistics) 2. Computational linguistics. I. Title.  
 II. Series.  
 P128.C68R47 2010  
 407.1–dc22 2009052595

ISBN 978-0-521-76987-7 Hardback  
 ISBN 978-0-521-14608-1 Paperback

Cambridge University Press has no responsibility for the persistence or  
 accuracy of URLs for external or third-party Internet Web sites referred to in  
 this publication and does not guarantee that any content on such Web sites is,  
 or will remain, accurate or appropriate.

Cambridge University Press

978-0-521-76987-7 - Using Corpora in the Language Classroom

Randi Reppen

Frontmatter

[More information](#)

---

*To my parents, Frank and Doris Reppen, who taught me to love  
languages and encouraged me to play with language.*

Jeg elsker deg.

# *Contents*

|   |  |
|---|--|
| Series editor's preface   | ix   |
| Preface   | xi   |
| Acknowledgments   | xiii   |
| <br>  |  |
| 1   | Corpora and language learning 1  |
| What is a corpus?   | 2  |
| Why use a corpus with language learners?                          | 4  |
| What are some ways to use a corpus with language learners?        | 5  |
| What will corpus-based materials look like?                       | 13   |
| Putting it all together   | 17   |
| Additional reading  | 17   |
| <br>  |  |
| 2   | Using corpus studies to inform language teaching 19                          |
| What corpus-based materials are available?                        | 19   |
| How can I develop materials and activities using corpus research? | 21   |
| Putting it all together   | 29   |
| Additional reading  | 29   |
| <br>  |  |
| 3   | Using corpus Internet resources in the classroom 31                          |
| Evaluating online resources                                       | 31   |
| Using available resources   | 33   |
| Using online corpora  | 37   |
| Using online corpora for classroom activities and materials       | 42   |
| Putting it all together   | 51   |
| <br>  |  |
| 4   | Using corpus material in the classroom and creating corpora for class use 52 |
| Creating corpus materials for classroom use                       | 52   |
| Developing hands-on activities                                    | 53   |
| Creating corpora for classroom use                                | 54   |
| Putting it all together   | 59   |

Cambridge University Press

978-0-521-76987-7 - Using Corpora in the Language Classroom

Randi Reppen

Frontmatter

[More information](#)

---

viii *Contents*

5 Bringing it into the classroom: Example activities and descriptions of corpora in the classroom 61

Example activities 61

Examples from the classroom 67

Putting it all together 71

Appendix A Resources for activities and instructions for using COCA and MonoConc Pro 73

Appendix B Corpus resources and tools 83

Selected bibliography of corpus linguistics and language teaching 89

References 99

Index 101

## *Series editor's preface*

A major challenge in developing language teaching materials and resources has always been to provide learners with language input that accurately reflects the way language is used in the real world. Critics of traditional language teaching materials have rightly pointed out that the information they contain about the use of English – whether it be information about grammatical usage, vocabulary, or conversational discourse – has often been based on conventional wisdom or on the intuitions of materials' developers, information that was often inaccurate or misleading. The “authentic materials” movement in language teaching that emerged in the 1980s attempted to address this problem by advocating a greater use of real-world or “authentic” materials – materials not specially designed for classroom use – since it was argued that such materials would expose learners to examples of natural language use taken from real-world contexts. More recently the emergence of corpus linguistics and the establishment of large-scale databases or corpora of different genres of authentic language have offered a further approach to providing learners with teaching materials that reflect authentic language use. The present book provides a comprehensive yet very accessible introduction to the use of corpora in language teaching and will be welcomed by teachers and materials developers wanting to know how they can make use of corpora in their language classes.

Drawing on her extensive experience in developing and using corpora in language teaching, Dr. Randi Reppen offers a masterly survey of the nature of language corpora and their practical uses in language teaching. She provides numerous examples of how corpus-informed teaching materials can be developed and used in teaching at many different levels and with students in many different contexts. *Using Corpora in the Language Classroom* together with its companion Web site will enable teachers new to corpus-informed teaching to overcome possible inhibitions about the use of language corpora, and provides them with the essential knowledge, tools, and skills needed to make use of the rich resources made possible in language teaching through the use of language corpora.

Jack C. Richards



## Preface

For many years I have been fascinated by corpus linguistics and how it can help me understand language better. My interest in corpus linguistics as a vehicle to better understand language has blossomed, and with that, a keen interest in how to use corpus linguistics to make me a more effective language teacher and teacher trainer. Corpus linguistics allows teachers and learners to be confident that they are learning the language they will encounter when they step outside the language classroom and into the real world of language use. Corpus linguistics provides a vehicle for bringing natural language into the classroom in a way that involves learners through hands-on activities interacting with “real” language. This book, *Using Corpora in the Language Classroom*, is designed to help teachers and teacher trainers better understand corpus linguistics and to help them bring the resources of corpora and hands-on learning with authentic materials into the language classroom.

Many teachers are eager to use corpora in their classrooms but lack the training and resources to accomplish this task. Teachers who would like to include corpus activities in the classroom are often overwhelmed by the task of locating corpora that are appropriate for their students, and by the task of creating activities for their students. This book addresses both of these challenges in four ways:

1. By providing an overview of corpus linguistics and detailed examples of corpus-based activities and materials, with case studies of class use that include hands-on activities.
2. By providing background information and principled instructions for creating a range of materials and activities that can be brought into the classroom, including how to create corpora to address specific class needs.
3. By providing lists of available corpora and Web sites that have corpus-based activities relevant to different teaching contexts and specific instructions on the use of existing Internet corpus resources.
4. By providing a companion Web site that includes links to online resources, frequency lists, and concordance lines (read Chapter 1 to find out what concordance lines are) as a springboard for activities; detailed corpus-based lessons and activities; and last but not

Cambridge University Press

978-0-521-76987-7 - Using Corpora in the Language Classroom

Randi Reppen

Frontmatter

[More information](#)

---

xii *Preface*

least, detailed instructions for using some of the more popular online corpora.

My goal in writing this book is to provide the step-by-step information needed for teachers to be able to successfully bring corpora and corpus resources into their language classrooms. All of the activities and examples in this book have been used in language classes. Every chapter has a strong hands-on component. Each chapter includes Your Turn boxes where you, the reader, are asked to interact with the material being presented or to do an activity.

Although only my name appears on the cover, this book certainly would not exist without the support and efforts of many people. A big thank-you to Don Miller for reading and commenting on portions of the book; your comments were very helpful. I owe a special debt of gratitude to Stacey Wizner who undertook the tedious task of combining my reference lists and standardizing the format of the Bibliography. I am especially grateful to Kathleen Corley who continued to believe in me and this project, regardless of the time zone that I was using for deadlines. A huge thanks to Carol-June Cassidy for her eagle-eye editing, helpful suggestions, and her positive energy. Doug Biber also provided tremendous support through engaging conversations on corpus linguistics. His confidence in me carried me through the ebb and flow of writing. Finally, thank you to the students in my MA, PhD, and ESL classes who have provided valuable insights that helped shape this book.

With such a multifaceted topic as corpus linguistics and using corpora to teach, there will always be aspects that are left unaddressed, but hopefully this book will serve to whet your appetite and curiosity for using corpora in your language classroom and provide you with some ways to accomplish this goal.

Randi Reppen

## *Acknowledgments*

MonoConc Pro screen shots for Figure 1.1 on page 6 and Figure 1.2 on page 9 are used by permission. Thank you to Michael Barlow for allowing this use.

Figure 1.4 on page 15 from M. McCarthy and O'Dell, F., *Basic Vocabulary in Use* (Cambridge: Cambridge University Press, 2010, 7). Used by permission.

Figure 1.5 on page 16 from M. McCarthy, McCarten, J., and Sandiford, H., *Touchstone* Level 2, (New York: Cambridge University Press, 2004, 15). Used by permission.

Figure 3.1 on page 39, screen shot from R. C. Simpson et al., *The Michigan Corpus of Academic Spoken English* (Ann Arbor, MI: The Regents of the University of Michigan, 2002). Used by permission.

Figure 3.2 on page 40 and Figure 3.4 on page 46, screen shots from Mark Davis, 2008–, COCA: Corpus of Contemporary American English (400 million words, 1990–2009). Available online at <http://americancorpus.org>. Used by permission. Thanks to Mark Davies for allowing this use and for the great corpus resources that he makes available online.

Figure 3.3 on page 41 and Figure 3.5 on page 50, screen shots from Mark Davis, 2007–, *Time Magazine Corpus* (100 million words, 1920s–2000s). Available online at <http://corpus.byu.edu/time>. Used by permission.

Microsoft product screen shots reprinted with permission from Microsoft Corporation.

Table 5.1 on page 65 is excerpted from D. Biber et al., *Longman Grammar of Spoken and Written English* (Harlow, Essex: Pearson Education, 1999), Chap. 13, by permission of the author.