

# Cambridge Elements

Elements in Corpus Linguistics

edited by

Susan Hunston

*University of Birmingham*

## DOING LINGUISTICS WITH A CORPUS

*Methodological Considerations for  
the Everyday User*

Jesse Egbert

*Northern Arizona University*

Tove Larsson

*Uppsala University / Northern Arizona University*

Douglas Biber

*Northern Arizona University*



CAMBRIDGE  
UNIVERSITY PRESS

Cambridge University Press  
978-1-108-74485-0 — Doing Linguistics with a Corpus  
Jesse Egbert, Tove Larsson, Douglas Biber  
Frontmatter  
[More Information](#)

---

## CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,  
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of  
education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781108744850](http://www.cambridge.org/9781108744850)

DOI: 10.1017/9781108888790

©Jesse Egbert, Tove Larsson and Douglas Biber 2020

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.

First published 2020

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

ISBN 978-1-108-74485-0 Paperback

ISSN 2632-8097 (online)

ISSN 2632-8089 (print)

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

Cambridge University Press  
978-1-108-74485-0 — Doing Linguistics with a Corpus  
Jesse Egbert, Tove Larsson, Douglas Biber  
Frontmatter  
[More Information](#)

## Doing Linguistics with a Corpus

### Methodological Considerations for the Everyday User

Elements in Corpus Linguistics

DOI: 10.1017/9781108888790  
First published online: October 2020

Jesse Egbert  
*Northern Arizona University*

Tove Larsson  
*Uppsala University / Northern Arizona University*

Douglas Biber  
*Northern Arizona University*

**Author for correspondence:** Jesse Egbert, [jesse.egbert@nau.edu](mailto:jesse.egbert@nau.edu)

**Abstract:** Paradoxically, doing corpus linguistics is both easier and harder than it has ever been before. On the one hand, it is easier because we have access to more existing corpora, more corpus analysis software tools, and more statistical methods than ever before. On the other hand, reliance on these existing corpora and corpus linguistic methods can potentially create layers of distance between the researcher and the language in a corpus, making it a challenge to do linguistics with a corpus. The goal of this Element is to explore ways for us to improve how we approach linguistic research questions with quantitative corpus data. We introduce and illustrate the major steps in the research process, including how to: select and evaluate corpora; establish linguistically motivated research questions, observational units, and variables; select linguistically interpretable variables; understand and evaluate existing corpus software tools; adopt minimally sufficient statistical methods; and qualitatively interpret quantitative findings.

**Keywords:** corpus linguistics, research design, quantitative methods, qualitative methods, statistical methods

© Jesse Egbert, Tove Larsson and Douglas Biber 2020

ISBNs: 9781108744850 (PB), 9781108888790 (OC)  
ISSNs: 2632-8097 (online), 2632-8089 (print)

## Contents

1	Introduction	1
2	Getting to Know Your Corpus	4
3	Research Designs: Linguistically Meaningful Research Questions, Observational Units, Variables, and Dispersion	15
4	Linguistically Interpretable Variables	24
5	Software Tools and Linguistic Interpretability	33
6	The Role of Statistical Analysis in Linguistic Descriptions	39
7	Interpreting Quantitative Results	52
8	Wrapping Up	70
	Appendix: Why It Is Problematic to Apply NHST to Large (Corpus) Samples	74
	References	76