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
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

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