



# Recommended Reading: Statistics for Corpus Linguistics

Priority	Text	Best for	How to use it
<b>Main recommendation</b>	<p>Brezina, V. (2018). <i>Statistics in corpus linguistics: A practical guide</i>. Cambridge University Press.</p> <p>Physical copy available at the <b>Hauptbibliothek Adenauerallee 39-41 &gt; 1. UG &gt; Freihandmagazin ; 2018/5899</b></p>	Students who want a practical corpus-linguistic introduction to statistical thinking.	Read selectively (ch. 1 plus possibly ch. 4 and 7 depending on your project). Focus on basic frequency comparison, denominators, effect sizes, and interpretation.
<b>Main accessible alternative</b>	<p>Biber, D., Conrad, S., &amp; Reppen, R. (1998). <i>Corpus linguistics: Investigating language structure and use</i>. Cambridge University Press.</p> <p>Physical copy available at the <b>Hauptbibliothek Adenauerallee 39-41 &gt; 1. UG &gt; Freihandmagazin ; 2001/7225</b> &amp; <b>Fachbibliothek Philologien Rabinstr. 8 &gt; 4. OG &gt; Freihandbereich ; ES 900 B581 C8</b></p>	Students who need the easiest replacement for Brezina for normalised frequencies and fair corpus comparison.	Read Chapter 6, "Norming frequency counts."
<b>Accessible discourse option</b>	<p>Baker, P. (2023). <i>Using corpora in discourse analysis</i> (2nd ed.). Bloomsbury Academic.</p> <p>(2006)</p> <p>Physical copy available at the <b>Hauptbibliothek Adenauerallee 39-41 &gt; 1. UG &gt; Freihandmagazin ; 2006/8797</b> &amp; <b>Fachbibliothek Philologien Rabinstr. 8 &gt; 4. OG &gt; Freihandbereich ; ES 900 B168 U8</b></p>	Students doing discourse, media, keyword, collocation, or concordance-based projects.	Use selected chapters relevant to your project. Good for moving from corpus patterns to cautious linguistic interpretation.

Priority	Text	Best for	How to use it
<b>Accessible general textbook</b>	<p>McEnery, T., &amp; Hardie, A. (2012). <i>Corpus linguistics: Method, theory and practice</i>. Cambridge University Press.</p> <p>Physical copy available at the <b>Hauptbibliothek Adenauerallee 39-41 &gt; 1. UG &gt; Lehrbuchsammlung ; Ling H 38 &amp; Fachbibliothek Philologien Rabinstr. 8 &gt; 4. OG &gt; Freihandbereich ; ES 900 M478</b></p>	Students who want a broader introduction to corpus-linguistic method, theory, and practice.	Use as a general reference for corpus design, concordancing, frequency, collocation, keyness, and interpretation.
<b>Optional: hands-on R</b>	<p>Gries, S. T. (2017). <i>Quantitative corpus linguistics with R: A practical introduction</i> (2nd ed.). Routledge.</p> <p>Available online via library access.* or on his own <a href="#">website</a>:</p>	Students who want to implement corpus-linguistic analyses in R.	Use when you need practical R code and examples. It is more technical than Brezina. If you read any of it, focus on ch. 3 and 4.
<b>Optional: advanced/self-study</b>	<p>Desagulier, G. (2017). <i>Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics</i>. Springer.</p> <p>Available <a href="#">online</a> via library access.*</p>	Students who want a fuller R-based introduction to corpus statistics.	Good for deeper self-study; not required for the current project.
<b>Optional: broader quantitative linguistics</b>	<p>Levshina, N. (2015). <i>How to do linguistics with R: Data exploration and statistical analysis</i>. John Benjamins.</p> <p>Available <a href="#">online</a> via library access.*</p> <p>Physical copy available at the <b>Fachbibliothek Philologien, Rabinstr. 8 &gt; 4. OG &gt; Freihandbereich; ES 910 L667</b></p>	Students interested in quantitative linguistics beyond corpus comparison.	Useful for longer-term development, but broader and more advanced than needed here.

\*Online manuscripts are only available when Uni Bonn VPN is activated.

## Main recommendation

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

**Why this is the best starting point:** This book is written specifically for corpus linguists who need practical statistical guidance. It is especially useful for understanding why raw counts are not enough, how to think about frequency data, why effect size matters, and how to report results cautiously.

**For this project, focus on:** the introductory chapter on statistical thinking in corpus linguistics; the chapters relevant to frequency, lexico-grammar, and change over time; and the concluding principles on interpretation and effect sizes.

**Recommended use:** Read this first. Do not worry if you do not understand every statistical detail. Concentrate on the logic of comparison: What is being counted? What is the denominator? Are the corpora comparable? What can the result actually show?

## Other accessible corpus-linguistic alternatives

Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). Bloomsbury Academic.

**Why it is useful:** This is a clear option for students whose projects are discourse-oriented. It is especially helpful for working with keywords, collocations, concordances, and the interpretation of corpus patterns in context.

**Recommended use:** Use selected chapters depending on the project. It is not primarily a statistics textbook, but it helps students connect quantitative corpus evidence with qualitative linguistic interpretation.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

**Why it is useful:** This is a broad corpus-linguistics textbook and a useful general reference for students who want more background on corpus design, corpus methods, frequency, collocation, keyness, and interpretation.

**Recommended use:** Use it as a general textbook rather than as the main reading for this particular statistics session. It is broader than Biber et al. and less narrowly focused on normalised frequency counts.

## Optional: if you want to work in R

Gries, S. T. (2017). *Quantitative corpus linguistics with R: A practical introduction* (2nd ed.). Routledge.

**Why it is useful:** This is a hands-on introduction to corpus linguistic work with R. It is particularly useful if you want to go beyond spreadsheet calculations and make your analysis more reproducible.

**Recommended use:** Use this if you want practical R guidance for basic statistics, regular expressions, data handling, and corpus-linguistic analysis. It is not required for everyone.

## Optional: advanced/self-study texts

Desagulier, G. (2017). *Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics*. Springer.

**Why it is useful:** This is a fuller R-based introduction to corpus statistics. It is helpful if you want to understand more advanced methods and see detailed corpus-linguistic case studies.

**Recommended use:** Use it for independent study or for a more ambitious project. It is probably more detailed than necessary for a small class project.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.

**Why it is useful:** This is an excellent general introduction to quantitative linguistics with R. It is broader than corpus linguistics and covers many statistical methods used in linguistic research.

**Recommended use:** Use it if you want to build broader quantitative skills in linguistics. For the current corpus project, Brezina (2018) is the more direct starting point.

## Which text should you choose?

Your need	Best choice	Reason
I need to understand normalised frequencies and denominators.	<b>Brezina (2018)</b>	Most directly relevant to responsible corpus comparison.
I want to calculate frequencies and percentages in R.	<b>Gries (2017)</b>	Good practical guide to R for corpus linguistics.

Your need	Best choice	Reason
I want to understand more advanced corpus statistics.	<b>Desagulier (2017)</b>	Detailed R-based treatment of corpus-linguistic methods.
I want broader quantitative skills for linguistics.	<b>Levshina (2015)</b>	Covers statistical analysis for many kinds of linguistic data.