

# Cambridge Elements

Elements in Digital Literary Studies

edited by

Katherine Bode

*Australian National University*

Adam Hammond

*University of Toronto*

Gabriel Hankins

*Clemson University*

## CAN WE BE WRONG? THE PROBLEM OF TEXTUAL EVIDENCE IN A TIME OF DATA

Andrew Piper  
*McGill University*



CAMBRIDGE  
UNIVERSITY PRESS

Cambridge University Press  
978-1-108-92620-1 — Can We Be Wrong? The Problem of Textual Evidence in a Time of Data  
Andrew Piper  
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/9781108926201](http://www.cambridge.org/9781108926201)  
DOI: 10.1017/9781108922036

© Andrew Piper 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 978-1-108-92620-1 Paperback  
ISSN 2633-4399 (online)  
ISSN 2633-4380 (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.

## Can We Be Wrong? The Problem of Textual Evidence in a Time of Data

Elements in Digital Literary Studies

DOI: 10.1017/9781108922036  
First published online: September 2020

Andrew Piper  
*McGill University*

Author for correspondence: [andrew.piper@mcgill.ca](mailto:andrew.piper@mcgill.ca)

**Abstract:** This Element tackles the problem of generalization with respect to text-based evidence in the field of literary studies. When working with texts, how can we move, reliably and credibly, from individual observations to more general beliefs about the world? The onset of computational methods has highlighted major shortcomings of traditional approaches to texts when it comes to working with small samples of evidence. This Element combines a machine-learning-based approach to detect the prevalence and nature of generalization across tens of thousands of sentences from different disciplines alongside a robust discussion of potential solutions to the problem of the generalizability of textual evidence. It exemplifies the way mixed methods can be used in complementary fashion to develop nuanced, evidence-based arguments about complex disciplinary issues in a data-driven research environment.

**Keywords:** digital humanities, humanities, machine learning, literary studies, natural language processing

© Andrew Piper 2020

ISBNs: 9781108926201 (PB), 9781108922036 (OC)  
ISSNs: 2633-4399 (online), 2633-4380 (print)

## Contents

Introduction, or What's Wrong with Literary Studies?	1
I Theory	8
1 Generally Speaking	8
II Evidence with Eve Kraicer, Nicholas King, Emma Ebowe, Matthew Hunter, Victoria Svaikovsky, and Sunyam Bagga	17
2 Modeling and Machine Learning	17
3 Results	36
III Discussion	51
4 Don't Generalize (from Case Studies): The Case for Open Generalization	51
5 Don't Generalize (at All): The Case for the Open Mind	60
Conclusion: On the Mutuality of Method	71
References	73