March 2018

Once Upon a Genre: Distant Reading, the Newbery Medal, and the Affordances of Interdisciplinary Paradigms for Understanding Children's Literature

Melanie Griffin
University of South Florida, griffinm@usf.edu

Once Upon a Genre:

Distant Reading, the Newbery Medal, and the Affordances of

Interdisciplinary Paradigms for Understanding Children's Literature

by

Melanie Griffin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Teaching and Learning
College of Education
University of South Florida

Major Professor: Jenifer J. Schneider, Ph.D.
James King, Ed.D.
Cynthia Patterson, Ph.D.
Barbara Shircliffe, Ph.D.

Date of Approval:
March 9, 2018

Keywords: macroscopic reading, prizing, canonicity, interdisciplinarity

Dedication

In memory of Robert.

Acknowledgments

First and foremost: thank you to Dr. Jenifer Schneider, major professor extraordinaire. I would not be here, writing this dissertation, without you. Thank you for recruiting me, for giving me the space to experiment and find my voice in the field of children's literature, for always pushing me to think harder and write better, for encouragement, and for providing me with this opportunity. Thanks most especially for not letting me quit.

I would also like to gratefully acknowledge my committee members and thank them for helping me become the scholar that I am today. Thank you to Dr. James King for introducing me to the study of linguistics and encouraging me to think about why individual words matter. Thanks, too, for your incisive comments on my first full draft. Thank you to Dr. Barbara Shircliffe for the introduction to gender theory as it relates to the history of education in America. Thank you to Dr. Cynthia Patterson for signing me into your class on scholarly research and writing as an overload student and for reading and providing so much guidance on the very first draft of what turned into this dissertation project.

I took a lot (a lot) of coursework en route to the dissertation, and I thank each and every professor for their part in the education of Melanie Griffin. In particular, though, I thank Dr. Jennifer Wolgemuth for Qualitative Research Methods I and for offering me the intellectual space to explore, to experiment, to create, to fail, and to explore some more.

And last but not least, thank you to the friends and family who supported me and who cheered me on. Thanks to Megan Cross for the many mornings we spent dissertating together.

Thanks to Sarah Pennington for rescuing me in Stats II and to Ashley Reese, Jessica Cook, and Tomaro Taylor for assuring me that this, too, would pass and for enduring so many conversations about the Newbery Medal. Thanks, of course, to Mom and Dad, who have always believed I could do anything. And finally, thanks to Dori Griffin for reminding me that if I were to quit before defending, the Drs. Griffin could not go on a celebratory vacation together.

Table of Contents

List of Tables

List of Figures

Abstract

Typical critical patterns for studying children's literature, defined in this study as a written text intended for a reader up to the age of 14, make it difficult to chart generic change across a large corpus of texts. Traditionally, criticism of children's literature focuses on cherry picked archetypes, exemplars, and the standout extraordinary. This study employs interdisciplinary methods and data sources from library science, education, and literary studies to create a method for analyzing a sample corpus of children's literature more holistically vis-à-vis distant, macroscopic reading techniques.

In this dissertation, I macroscopically read the corpus of Newbery Medal-winning texts in order to identify patterns of change in the genre of prized 20th century American children's literature, seeking to animate this corpus of texts in different ways than is possible through microscopic analysis alone. The resulting analysis foregrounds the shared conventions of the text set, including descriptive elements, including bibliographic information, author information, publisher information, illustrative content, and length; structural elements, point of view, literary form, and select measures of text complexity; and thematic elements, including book summaries and subject analyses from a range of library, publisher, and social media databases. In addition, I consider various metrics for assessing popularity of the corpus as a whole and the ways in which popularity changes as time passes.

Ultimately, in this dissertation I distantly read the corpus in conversation with existing critical understandings of the Newbery Medal, which previous critics generated using microscopic, close reading techniques, in order to investigate what changes with the introduction

of distant methods. Distantly reading this corpus in conversation with existing critical understandings of the Newbery reveal that a more holistic approach to analysis paints a broader, more complete picture of the genre of prized children's literature than microscopic, close reading alone does. Further, distant reading underscores the critical importance of explicit attention to methodology. The results that distant reading uncovers are inextricably intertwined with the methodological decisions made.

Chapter 1: Introduction

Once Upon a Dissertation Study

In titling this dissertation "Once Upon a Genre," I explicitly call attention to stories, to tales told and retold, and to narrative traditions that change, particularly as the teller also changes. Rather than focusing on the "times" or "tales" that traditionally follows the introductory "once upon a…" phrase, however, I focus instead on stories about scholarly traditions, especially the stories that scholars tell about children's literature as a genre. One of the stories that inspired this dissertation study relates to perceived problems in defining children's literature as a discrete genre. Decades worth of critical debate suggests that children's literature is generically tricky. The phrase connotes literature *by* children and literature *for* children as well as the critical study of both (Nodelman, 2008). If a scholar takes "literature for children" as the denotation of the phrase, however, additional definitional problems arise. What, precisely, constitutes a child? How can a reader know if the text in hand was intended for a child? Must a text be intended for a child in order to be children's literature? Does a text remain an exemplar of children's literature when an adult reads it? Complicating these questions is the answers the stories provide: there is, of course, no one answer. Thus, some claim, a unifying definition of children's literature becomes impossible (Bator, 1983; Egoff, 1976; Rose, 1982; Townsend, 1980; Zipes, 2013). I worked against these stories and asked instead how the genre of children's literature might be understood if critics were to accept a multiplicity of genres rather than seeking one monolithic structure.

The second story that inspired this project relates to how scholars study children's literature, however "children's literature" might happen to be defined. Criticism of children's literature tells a persistent story, one in which multiple disciplinary perspectives introduce additional complexities into questions of defining children's literature. This story claims that literary studies, library science, and education all lay territorial claim to children's literature (Clark, 2003), and the scholarship of each discipline introduces unique and sometimes conflicting criteria to definitional claims. According to this axiom, library science scholarship typically informs collection development practices, education scholarship frequently focuses on informing teaching and learning practices, and literary studies scholarship more often than not focuses on the text analysis outside of the social contexts of actual readers (Nel & Paul, 2011). The three paradigms, so the story goes, compete in siloes without informing one another. This story is persistent; most recently, Bittner and Superle (2016) re-affirmed that "the often substantially different theoretical lenses used by various groups affect their beliefs about the value and purpose of children's literature" (p. 73).

I embody the same tripartite delineation to which Clark (2003) and Nel and Paul (2011) refer. I am, by training and trade, a librarian, and my professional duties include curatorial responsibility for a collection of largely historical children's literature. My undergraduate and subject-specific master's level work occurred in departments of English, with a focus on text analysis from historical perspectives. My doctoral work is situated in a college of Education, with coursework across a range of humanistic and social sciences disciplines. On a daily basis, I witness the fallacy of reducing disciplinary perspectives of children's literature to competing, rather complementary, paradigms. Therefore, through this dissertation project, I sought fruitful avenues for combining the seemingly discrete disciplinary frameworks of library science,

education, and literary studies in my analysis of children's literature in order to offer a more holistic method for analyzing these texts. In doing so, one of my goals is to disrupt the canonical story of three siloed disciplines, as cited by Clark (2003) and Nel and Paul (2011), and to use methods and data sources from library science and education to inform a macroscopic discussion of children's literature as a literary genre. My method for accomplishing this disruption relies on distant reading techniques (Moretti, 2005) in order to create a macroscopic view of a sample corpus of children's literature.

The specific sample corpus on which I have chosen to test distant reading methods is Newbery Medal winners, 1922-2017. Newbery Medal winners provide a purposive sample (Maxwell, 2009) in that I chose them because they provide the following affordances that a random sample would not. The Newbery Medal has been awarded annually since 1922, using largely unchanged criteria and resulting in a workably sized, fairly homogenous corpus. In addition, a large body of criticism exists on individual Medal-winning texts, and some criticism offers insights about the Medal in general, spanning the entire corpus. This existing criticism enabled me to interrogate existing assumptions about the corpus and test macroscopic patterns against those assumptions in a way that would be impossible in a completely random sample.

Statement of the Problem

Typical critical patterns for studying children's literature, defined in this study as a written text intended for a reader up to the age of 14, make it difficult to chart generic change across a large corpus of texts. Traditionally, criticism of children's literature focuses on cherry picked archetypes, exemplars, and the standout extraordinary. Nancy Drew, for example, frequently stands in for an entire genre of girl sleuths, while Anne Shirley represents girl orphans and, more recently, *The Fault in Our Stars* represents the quintessential modern young adult

problem novel. Genres come to be defined by the individual exemplar, and, to use an old cliché, the forest is lost with all the focus on extraordinary trees. Even within well-defined corpora of canonical children's literature, such as winners of the Newbery Award, critical studies tend to focus on single defining characteristics, such as readability (Leal and Chamberlain-Solecki, 1998) or critical race theory (Cook, 1985; Larrick, 1965; Miller, 1998). Further, more often than not, these studies focus on a small sample within the already small corpus of Newbery winners. Previous sampling strategies include Medalists during World War II (Dyson, 2007); winners and Honor books from a limited time span meeting specific content criteria (Forest, 2014; Leininger, Dyches, Prater, & Heath, 2010); or the work of a single Medal-winning author (e.g., Roggenkamp, 2008).

Franco Moretti (2000, 2005), however, challenges this notion of exclusively close, or microscopic, reading, providing instead a framework for macroscopic reading of genres that looks beyond individual exemplars and takes into account the larger contexts of generic traditions over time that become visible when corpora of texts rather than single examples are considered holistically. Inspired by Moretti in particular and digital humanities more generally, and building on Kenneth Kidd's (2007) work on the prizing of American children's literature, I exploit distant methods of reading in this dissertation to explore the descriptive, structural, and thematic characteristics of the Newbery Medal-winning titles as a sub-genre of American children's literature. I intentionally couple data from library science, education, and publishing sources with distant reading, traditionally found in the domain of literary studies, in order to consider the affordances that different disciplinary perspectives offer to the study of children's literature.

Purpose of the Study

In this dissertation, I seek to interrogate the Newbery Medal corpus as a genre of children's literature from new, frequently computational, distant perspectives in conversation with what is already known about the text set from more microscopic inquiry. I do not assert that this interrogation or the data I use is objective; rather, I seek to understand and interpret data sets in order to better understand the cultural and social work accomplished by prized American children's literature as a genre. Although most criticism of the Newbery restricts itself to the history of the award or the text of a subset of the winning books, my purpose is to examine the corpus holistically, exploring the history of the award, the descriptive and structural elements of the winning books, thematic components of the entire corpus, and popularity measures for all winners, ultimately providing insight into the defining generic characteristics of prized American children's literature over the past century. In this study, I define *generic characteristics* as the sum total of the descriptive, structural, and thematic components identified and analyzed throughout.

Further, I conducted this study in order, in part, to investigate what happens when the critic does not relegate methodology to an appendix. In addition to exploring how children's literature changes structurally over time, I seek to explore how methodological approaches might affect the conceptualization of children's literature. In order to explore these questions, I purposefully and explicitly employ different modes of inquiry, informed by Moretti's (2005) notion of distant reading, generating computational models and data visualizations of the Newbery Medal created from secondary data sets describing the corpus. As such, my resultant analyses do not offer close readings of any of the individual texts that won the Newbery or content analysis based on a cluster of texts, although they do suggest fruitful avenues for future

microscopic explorations. Instead, I foreground the shared conventions of the text set, such as descriptive elements, including bibliographic information, author information, publisher information, illustrative content, and length; structural elements, point of view, literary form, and select measures of text complexity; and thematic elements, including book summaries and subject analyses from a range of library, publisher, and social media databases. In addition, I consider various metrics for assessing popularity of the corpus as a whole and the ways in which popularity changes as time passes. Ultimately, I distantly read the corpus in conversation with existing critical understandings of the Newbery Medal, which previous critics generated using microscopic, close reading techniques (Richards, 1929), in order to investigate what, if anything, changes with the introduction of distant methods.

Scope of the Project

The American Library Association first awarded the Newbery Medal in 1922. Subsequently, they awarded a Medal every year and, as of 2017, there are 96 winners. In this project, I trace the development of the Newbery's canon of children's literature from its inception to the present day. I use this common set of texts as an instantiation of a subcategory of children's literature to perform different types of analyses in order to investigate how children's literature might change, descriptively, structurally, thematically, and in terms of popularity over time when considered holistically rather than through the lens of a single text. I examine the corpus holistically, and I use this holistic analysis to pinpoint descriptive, structural, and thematic characteristics that warrant closer, more microscopic scrutiny.

A Framework for Distant Reading

In his article "Conjectures on world literature," Franco Moretti (2000) laid the groundwork for a seminal approach to literary analysis outside of an established canon: distant

reading. Reacting against the practice of close reading, which he argued was a "theological exercise" overly reliant "on an extremely small canon" (2000, p. 57), Moretti instead advocated that the literary critic gain distance from individual exemplars by focusing on large bodies of texts in the composite through a reliance upon statistical analysis. In this model, Moretti argues:

> Distance, let me repeat, *is a condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something. (2000, p. 57, emphasis in original)

In *Graphs, Maps, Trees* (2005), Moretti explores in greater detail what, precisely, distant reading entails and how a literary critic might gain distance from individual texts using not only statistical analysis but also other tools via temporal, spatial, and morphological approaches to literature. The graph, map, and tree, or a "trio of artificial constructs," Moretti argues, allows for "the reality of the text [to undergo] a process of deliberate reduction and abstraction" (p. 1). In this abstraction, Moretti conceptualizes two centuries of European novels via quantitative graphs, reduces texts from words to symbolic maps, and charts morphological change in diagrams of the systems governing the literary survival of the British detective novel. By moving away from the individual exemplar, he shifts his perspective to larger trends observable from his distant stance. I employ a similar shift in my study by using Newbery Medal-winning texts as a sample corpus of contemporary American children's literature, asking what a different perspective affords to the existing generic model and what it misses. As such, the purpose of this study is to model a

macroscopic approach to analyzing children's literature, using Newbery Medal-winning titles as a test case.

Delimitations

Numerous definitions of what constitutes "literature" for "children" abound, and a multitude of potential corpora of "children's literature" from which I could potentially draw exists. In order to create a text set for investigation, I selected a purposive sample (Maxwell, 2009) from the larger field of contemporary American children's literature and considered one discrete corpus: Newbery Medal-winning texts. The American Library Association (ALA), the professional body that oversees the Newbery Award, specifically frames the Newbery Medal as an award for literature, noting that the Medal goes to "the author of the most distinguished contribution to American literature for children" published in the previous year (ALSC 2016). The ALA instituted the Newbery Award in 1922, resulting in nearly a century of texts, all selected according to the same criteria. Further, scholars frequently write about and critique the Medal and Medal-winning texts. Much is already known about individual exemplars of the Newbery, and there are small areas of knowledge that cover the entire corpus. My intention is not to hold the Newbery Medal-winning titles up as unquestioned exemplars of children's literature. It is also not my intention to claim that the Newbery Medal-winning titles provide a representative sample of all types of children's literature. Instead, I assert that the Newbery Medal-winning titles provide data on *one* sub-genre of children's literature, not *the* genre of children's literature. My purpose is to work with the corpus that this pre-defined sub-genre provides in order to explore ways for macroscopically understanding children's literature as genre. I use this corpus of canonical American children's literature to interrogate assumptions behind the designation of literary texts for children. Further, I explore how the Newbery Medal

as genre has both changed and remained stable over time in stylistic, structural, thematic matters. I also consider how measures of popularity have and have not changed in the corpus with the passage of time.

Significance of the Study

This study addresses two gaps in the literature: one methodological, one content-related. Distant reading, by its very nature, requires metadata – that is, data about the literary work under consideration – or databases that include encoded corpora of literary texts. Full text databases of encoded literary texts abound in some fields, particularly those that deal with literature corpora published before contemporary United States and/or European copyright law coverage, but there is a paucity of encoded corpora of modern children's literature. Due to United States copyright law, which protects texts published after 1923 (Copyright Law of the United States), an open access, encoded database of contemporary American children's literature does not currently exist. Therefore, distant reading in contemporary American children's literature requires creative approaches to locating and harvesting secondary data sets. I locate these data sets through library and publishing resources geared towards educators and readers.

Given these challenges, as well as the location of the data sources, it is unsurprising that, to date, analyses of contemporary children's literature have not made use of distant reading methods to understand children's literature as genre. In this dissertation, I model methods for finding, identifying, collecting, and analyzing data sources about one sub-genre of children's literature as a test case in order to facilitate a macroscopic understanding of that selected sub-genre in conversation with existing microscopic critical conversations. In doing so, I seek to understand the utility of applying macroscopic reading tools to the field of children's literature.

Critics and scholars of children's literature frequently analyze Newbery Medal-winning texts, but previous studies of the Newbery Medal focused on a subset of winning texts (e.g., Dyson, 2007; Forest, 2014; Leininger, Dyches, Prater, & Heath, 2010; Roggenkamp, 2008), analyzed individual thematic elements found in the corpus (e.g., Larrick, 1965; Miller, 1998), or theorized the work accomplished by prizing in the field of children's literature (Cook, 1985; Kidd, 2007). A very small number of studies investigated the entire corpus, but focused in very narrowly on a specific element, such as readability (Chatham, 1980; Clements, Gillespie, and Swearingen, 1994; Stevens, 2010), the representation of women (e.g., Houdyshell and Kirkland, 1998; Smulderes, 2015), or the exclusion of authors and characters of color (e.g., Miller, 1998). By applying distant reading methods to this corpus of texts, I explore how different perspectives might shift understandings of the Newbery Medal as a sub-genre of children's literature. Scholars currently understand the Newbery, I argue, either from an overly restrictive, small sample of texts that do not account for larger trends across the entire corpus or from a large sample that investigates one problem narrowly. In reframing how this corpus works on a holistic level, this study questions an over-reliance on individual exemplars and themes and the ways in which those exemplars and themes have come to define what is accepted as children's literature. In addition, I use the framework of distant reading to uncover the methodological ramifications that lead to definitions of genre.

Definitions

As I noted above, a persistent story about children's literature is that it is generically difficult to define (Nodelman, 2008). Similarly, the term distant reading takes on different meanings in different studies, with variations ranging from size of corpus (Bode, 2017; Booth, 2017) to tools employed (Underwood, 2017). Rather than claiming one definition is inherently

better or more correct than another, I instead rely on the operational definitions listed below throughout this study.

Children's literature. Children's literature refers to a work of literature intended for a child reader, up to the age 14. This definition encompasses works in any textual mode, including but not limited to poetry, nonfiction, drama, and fiction, originally published in traditional (i.e., print) book format. This operational definition intentionally mirrors that used by the Newbery Medal selection committee (ALSC, 2016) and, as such, relies on the publisher's designation for a given book meeting the criteria listed above. It excludes young adult literature intended for readers 14 and above.

Distant reading. Distant reading provides a method for reducing individual texts to abstraction by focusing on a corpus of texts holistically, through coding, content analysis, or statistical analysis rather than close reading of individual exemplars (Moretti, 2005).

Close reading. In this study, close reading is the opposite of distant reading and offers a focused analysis of a single text that places emphasis on the individual words, sentences, structures, and aesthetic choices found within that text (see Richards, 1929). As taught in literary analysis courses, close reading typically relies on repeated readings of a text, reader annotations, notation of patterns, and questioning why and how those patterns occurred (Kain, 1998).

Microscopic reading. Microscopic reading is a synonym for close reading that emphasizes the size of the corpus under consideration. Microscopic reading is a practice in which the unit of analysis is the individual text or a corpus of texts small enough to be examined by the human brain alone (Moretti, 2005).

Genre and generic characteristic. A genre refers to a set of literary texts sharing common conventions. These conventions may govern style, length, content, form, subject, and/or

intention. Fundamentally, genre is an organizational paradigm for literature (Frow, 2014). A **generic characteristic is** an identifying characteristic of a genre that marks a text as belonging to that genre. Examples range from the rules governing how a story is told to thematic elements, narrative strategies, the relationship between word and image, and intended audience.

Descriptive characteristic. A descriptive characteristic provides basic illustrative information about a work of literature. Descriptive characteristics help a reader identify a specific text and differentiate it from other texts. Elements include bibliographic information about the text and its creator(s).

Structure and structural characteristic. Structure refers to the set of formal features (Frow, 2014) found within a literary text. A **structural characteristic** is an individual unit of the formal features governing a literary work. Examples may be found in the text, such as point of view or perspective, or in the arrangement of the text on the page or the relationship between word and images in the creation of meaning. In this study, I limit analysis of structural characteristics to point of view, literary form, and text complexity.

Theme and thematic characteristic. Theme refers to the central idea of a literary text (Cuddon, 2012). A **thematic characteristic** is an individual unit that illustrates or describes a portion of the text's main theme. In this study, I derive thematic characteristics from controlled vocabularies and user tags providing subject access to works of literature. Therefore, in this study, theme bears a stronger relationship to subject than it traditionally does.

Measure of Popularity. A measure of popularity provides quantifiable information about the circulation density or longevity of a single title. Examples include the number of discrete editions published of a title, the number of editions that remain in library collections, as well as the number of readers who record reading a title on a social networking site.

Chapter 2: Review of the Literature

Introduction

The purpose of this study is to interrogate the Newbery Medal corpus as a genre of children's literature from new, frequently computational, distant perspectives in conversation with what is already known about the text set from more microscopic inquiry. As such, this study represents my attempt to generate distant reading tools by which critics can understand a multiplicity of children's literature genres, not a study that seeks to define children's literature as a monolithic structure. Therefore, this study both is and is not about the Newbery Medal. Focusing on the Newbery affords a test case for macroscopic analysis of contemporary children's literature and enables a consideration of how commonly employed microscopic methods may or may not result in different understandings of children's literature as a genre that previous scholars generated using primary microscopic approaches. As the Newbery Medal forms the test corpus that I use to model a macroscopic approach, I begin by tracing the history, development and purpose of the Medal before turning to previous critical approaches to the Newbery. Two facets of inquiry are important in this exploration of the Medal: existing methodological approaches to studying the Medal and Medal-winning texts, and the findings that those methodological approaches have yielded.

The macroscopic approach I employ intentionally blurs the boundaries between library science, education, and literary studies. Therefore, it is useful to consider the origins of the disciplinary silos entrenched in studies of children's literature to which scholars often allude (e.g., Clark, 2003; Nel & Smith, 2011), to trace the full range of disciplinary paradigms and their

treatment of children's literature, and to consider outliers to the siloed approach and the affordances a more interdisciplinary approach can offer to the study of children's literature.

The Newbery Medal and the Construction of Canonical American Children's Literature

Definitions of childhood and children's literature proliferate, and many scholars conclude that a unifying theory of "children's literature" is therefore impossible (Bator, 1983; Egoff, 1976; Rose, 1982; Zipes, 2013). Rather than operationally defining children's literature as a monolithic structure, I instead acknowledge the multiplicities of children's literature inspired by and reacting to the multiplicity of formats, styles, cultures, environments, and readers involved in children's literature writ large. Within this project, however, I have operationally defined children's literature as any work of literature intended for a child reader, up to the age 14. This definition encompasses works in any textual mode, including but not limited to poetry, nonfiction, drama, and fiction, originally published in traditional (i.e., print) book format. This operational definition intentionally mirrors that used by the Newbery Medal selection committee (ALSC, 2016) and, as such, relies on the publisher's designation for a given book meeting the criteria listed above. It excludes young adult literature intended for readers 14 and above.

Despite an abundance of different critical approaches, however, most critics agree that boundaries between children's and adult literature solidified through the twentieth century due to a confluence of events leading up to the development of the Newbery Medal. Rowe (1971) suggests that clearer boundaries appear to the modern eye due to expanding library and bookstore spaces devoted to the child. Similarly, Alderson (2010) argues that the rise of children's librarianship, and the corresponding shift in libraries' collections budgets, affected publishing practices, with publishers modifying their children's books "to tastes and fashions espoused by professional readers of children's books who were inclined to encourage experiment

and leave popular appeal to look after itself" (p. 39). All are important considerations, and all relate to one additional factor: the advent of prizing within the field of children's literature. Prizes for children's literature began in America with the Newbery Medal. Given the Newbery's place of prominence in the field of children's literature and the ways that the Award shapes the generic boundaries surrounding contemporary American children's literature, it is crucial to interrogate the assumptions behind prizing children's literature in general and the creation of the Newbery Medal, the Medal's definitional criteria, and the ways in which the Medal influences American literary culture in particular. In this section, I offer a brief history of the Newbery Medal and consider the roles that the prize plays in the construction of American children's literature as a genre.

The creation of the Newbery Medal. Many awards for children's literature currently exist,[1] but the Newbery was the first[2] and remains the most prestigious (Allen, 2011), with a reputation for creating the canon of children's literature (Kidd, 2007). Begun in 1921 through efforts of bookseller Frederic Melcher after his noted success at creating the first Children's Book Week celebration in 1919 with children's librarian Anne Carroll Moore (Smith, 1957, p. 16), the award is named for British bookseller John Newbery (1713-1767). Although the focus of the Newbery has always centered on American children's literature, the choice of an eighteenth century London-based publisher and bookseller as the award's namesake was intentional: Newbery has long enjoyed a reputation as the publisher who introduced pleasure and amusement

[1] The American Library Association's Association for Library Service to Children alone currently offers ten book and media awards for children's literature annually. Professional organizations in other fields also offer numerous awards, as do analogous organizations in other countries.

[2] The Carnegie Medal, awarded by England's Chartered Institute of Library and Information Professionals, was not begun until 1936.

to the previously instruction-bent genre of children's literature (Townsend, 1996) and whose business acumen made the genre economically viable (Marcus, 2008).

From its inception, the Newbery Award has represented a partnership between publishers and librarians. Melcher proved instrumental in early twentieth century American attempts to add prestige and profitability to the children's literature publishing industry (Marcus, 2008, p. 85). After noting the success of the Pulitzer Prize, which was established in 1917, Melcher proposed the idea for a children's literature award at the Children's Librarians' Section meeting at the 1921 American Library Association annual meeting (ALSC 2016a; Smith, 1957). The ALA Executive Board officially approved the proposal in 1922 (ALSC, 2009, p. 7) after Melcher contracted with the association to donate the bronze medal to the winner, and the first Medalist was named later that year. Every year since 1922, the Association for Library Service to Children[3] (ALSC), a division of the American Library Association (ALA), has presented the Newbery Award to the "author of the most distinguished contribution to American literature for children" (ALSC, 2016c).[4]

In 1922, the Newbery's inaugural year, a vote of the ALA's Children's Librarians' Section determined the Medal winner. By 1924, a committee, comprised entirely of children's librarians, oversaw the selection. In 1937, the same year that the ALA introduced the Caldecott Medal for illustration, four school librarians joined the Newbery selection committee. Beginning in 1978, membership of the committee shifted from entirely elected volunteers to mixed appointed and elected positions, with membership in the ALSC remaining a requirement for service. Although committee membership introduced some modifications, procedures and

[3] Originally the Children's Librarians' Section; the name changed to the Children's Services Division in 1958 and the current Association for Library Service to Children in 1977.
[4] This phrase is to be found in the original contract between Melcher and the ALA, and it has remained in every document outlining criteria for the Award since.

selection criteria remain largely unchanged from the Medal's inception to the present day. Minor revisions, mostly dealing with the advent of new media and non-book formats in the field of children's literature, occurred in 1978, 1987, and 2008 (ALSC, 2016c).

Procedures for selecting the Medal-winning title appear simple. A selection committee, currently consisting of 15 members from the ALSC, considers each eligible book and ultimately names the winner. Eligibility criteria include nationality of the author, with "American" interpreted quite broadly,[5] intended readership of the book, and publication date (ALSC, 2009). All committee deliberations remain confidential, and the ALSC clearly delineates definitional benchmarks for the award, noting that "'Contributions to American literature' indicates the text of the book," and the phrase "for children" denotes a book intended for a readership up to the age of fourteen (ALSC, 2016a). Other criteria for judges to consider take the form of a bulleted list, presented without commentary:

- Interpretation of the themed or concept
- Presentation of information including accuracy, clarity, and organization
- Development of a plot
- Delineation of characters
- Delineation of a setting
- Appropriateness of style. (ALSC, 2016)

The only commentary provided on these criteria points back to the idea that the text must include "distinguished qualities…[and] excellence of presentation for a child audience" (ALSC, 2016). Further, each book is to be considered as a contribution to American literature, and the

[5] American citizenship is not a requirement; rather, an author must live in America. Neil Gaiman, for example, a British citizen living in America, maintained eligibility to win the Newbery Medal in 2009.

committee is tasked with making a decision based almost exclusively on the text. The Award committee's criteria foreground the idea of literary merit and artistic achievement, noting that, "The committee should keep in mind that the award is for literary quality […] The award is not for didactic content or popularity" (ALSC 2009, p. 11). Somewhat curiously in an award for children's literature, a genre dominated by illustrated texts, the only consideration visual components receive in the criteria are negative, with the ALSC noting that "illustrations" and "overall design of the book… may be considered when they make the book less effective" (ALSC, 2015a).

The influence of the Newbery Medal. As originally conceived by Melcher, the Newbery Medal served a three-fold purpose:

> To encourage original and creative work in the field of books for children. To emphasize to the public that contributions to literature for children deserve similar recognition for poetry, plays or novels. To give those librarians, who make it their life work to serve children's reading interests, an opportunity to encourage good writing in this field. (ALSC, 2009, p. 60)

The assumptions underlying the creation of the Newbery Medal illustrate important conventions in the field of children's literature as well as the ways in which the Award continues to interact with the publishing market, schools, and libraries. First, Melcher created the Newbery with the burgeoning children's book industry in mind and with an explicit goal of promoting the creation of books for that particular market. This focus implies that the Medal, and the literature that it seeks to reward, are commercial ventures. Second, the Newbery asserts that children's literature is *literature* and worthy of being considered as such. Third, Melcher's purpose highlights the gatekeeping role long played by librarians in selecting, promoting, and preserving literature for

children. All three assumptions influence the types of literature that the Newbery prizes and therefore influence the corpus of texts that I analyze in this study. In subsequent sections, therefore, I explore the relationship between the Newbery Medal and the publishing industry as well as the role played by librarians in the construction of the Newbery canon and consider how Melcher's original assumptions are visible in the Medal-winning titles.

The Newbery as market force. It is unsurprising that the Newbery Medal, named after John Newbery, is known for its impact on the children's literature market. John Newbery's reputation, after all, rests on his status as a book publisher known for looking for books that were both instructive and enjoyable in order to boost his sales (Marcus, 2008), and Melcher himself was also involved in the publishing trade. More recently, the Newbery Medal has come to be known as "the one literary prize that can dramatically boost book sales" (Silvey, 2008, p. 39). Although this assertion arguably downplays the important role performed by other contemporary literary prizes on book sales,[6] prizing plays an additional role in the market for children's literature given the relationship between book sales and curriculum (Clark, 2003; Kidd, 2007). The Newbery Medal sticker helps guarantee sales, and it helps guarantee sales year after year (Silvey, 2008; Maughan, 2013), with the Newbery sticker almost "ensuring a permanent place on a publisher's backlist" (Clark, 2003, p. 74). This continual sales activity helps keep Newbery Medal-winning books in print longer, and ready access encourages the adoption of these texts in the classroom or as supplemental, encouraged reading in schools (Kidd, 2007).

Librarians as gatekeepers of the Newbery. In order to understand the role played by children's librarians in establishing the Newbery Medal and the criteria under which it would be

[6] The Mann Booker prize is particularly known for boosting sales; the Telegraph recently reported that Nielsen Bookscan shows sales increases of up to 1918 percent for winners (Blumsom, 2015).

awarded, it is first necessary to explore briefly the history of library services for children in America. During the 19th century, male librarians predominantly offered library services (Passett, 1993) and were focused towards adults. Libraries frequently did not even allow children inside; when libraries did grant access to children, adult accompaniment was a must (McDowell, 2014, p. 521). With the rise of public libraries aided by Carnegie grants, however, services offered in libraries began to shift, and in 1876, Minerva Sanders, librarian at the Pawtucket Public Library in Rhode Island, took the controversial step of allowing children to access library materials without adult supervision (Eaton, 2010, p. 4). At this time, however, children's materials remained inter-shelved with adult materials (Fathauer & Rogall, 2000).

Progressive Era educational reform and accompanying changes in educational policy, theory, and practice resulted in a professional shift for librarians. Librarians as well as educators began to conceptualize "childhood" as a separate status from "adulthood," with children standing in need of nurturing and protection (Tyack and Hansot, 1992). As a result, women found a niche in the field of professional librarianship: services for children (Hearne, 1996). Despite critique about the resultant feminization of the library workforce (see Matthews, 1917), children's librarianship offered women vocational opportunity that required traditional feminine values, including "piety and purity (in selecting and distributing books that would be a good influence on readers), submissiveness (in serving the public), and domesticity (in maintaining a home-like environment in the library)" (Jenkins, 1996, p. 814). As a result, libraries and the services offered in libraries began to transform. In 1887, Emily Hanaway, a school headmistress distressed by children reading questionable material in the absence of adult guidance, used donations to establish a private children's library, which allowed entrance via a ticketing system. In 1890, the Brookline Public Library in Massachusetts established the first publicly funded children's room

(Eaton, 2010), and public libraries across the country rapidly followed suit (Fathauer & Rogall, 2000). More often than not, women oversaw the management of these new children's rooms. In addition to providing dedicated physical spaces for children and adolescents, children's rooms in the libraries, and the women who staffed them, revolutionized services for the child audience, offering specialized collections, personnel, and child-centric techniques (McDowell, 2014).

A crucial component of the new children's librarian's job was selection of specialized collections for a juvenile readership. Although "librarians relied on input from their teacher colleagues, they regarded themselves as the ultimate experts in selecting materials for children's reading, and did not have any doubt of their authority to choose" (Kimball, 2012, p. 680). Collection development, including book selection, became the purview of the children's librarian.

Newly re-conceived ideas of space allowed children access to shelves of books rather than requiring them to rely on pages who would bring individual titles out from closed stacks as was the norm in adult collections at the time (Eaton, 2010). Despite this practice, the children's librarian nevertheless played an important gatekeeping role with respect to a child's access to library materials. These librarians served as "arbiters of taste for youth, who would be responsible for reforming gauche reading habits, and for shaping the minds of all children, including children of immigrants" (Martens, 2013, p. 309). Children's librarians conceptualized the selection of books for the children's collection as an important mechanism for safeguarding and nurturing the child reader's mind. As children's librarian Caroline Burnite noted in 1911, librarians "must be an active influence in the mental progress of the child" (Burnite, 1911, p. 162) through the selection of appropriate books.

Children's librarians did not comprehensively collect the burgeoning number of books published for a juvenile audience each year. Rather, librarians selectively acquired examples of what they considered the best books to nurture and guide the minds of young readers. Dime novels and series books, for example, which were thought to be corrupting influences on the young, were excluded from the children's library (West, 1985). Books that explored life in other countries, however, proved popular mainstays in children's collections (Kimball, 2012), especially as tools for demonstrating "the superiority of American democracy and the American way of life" to an increasingly diverse population (p. 681). This emphasis finds a mirror in the titles awarded the Newbery Medal, with awardees becoming "vehicles for selecting and defining that which is American – even when the books are ostensibly about other cultures" (Alberghene, 1981, p. 10).

As a corollary to their selection activities, children's librarians instituted the practice of making lists of "best books" for other libraries to purchase. A central professional function of children's librarians, as Wiegand (1986) demonstrates, consisted of creating and compiling annotated reading lists for children and their colleagues. Children's librarians saw book reviewing and critical activity as a logical extension of their selection activities and many early children's librarians, such as Anne Carroll Moore, children's librarian at the New York Public Library, established reputations in both the critical review arena and librarianship (Martens, 2013, p. 209). As a result of their dual roles as reviewers and selectors, children's librarians exerted immense influence on the children's book market, with their professional values and tastes guiding publishing decisions (Kimball, 2012). After all, if a children's librarian did not consider a book a "best book" for children to read, she simply did not purchase it for her library collection while recommending that her colleagues refrain from doing so as well.

The creation of the Newbery Award cemented the children's librarian's reputation as critic. Although the ostensible purpose of the Medal is to honor distinguished contributions to the field of children's literature, an unstated but nevertheless important purpose of the Medal as originally conceived was to reinforce the children's librarian's role in recognizing, acknowledging, and selecting the best books for children (Willett, 2001) and serving as a selection guide for teachers, children, and parents (Miller, 2014).

It is also important to consider the types of literary works prized by the first selection committees and to consider how the early Newbery Medal-winning titles reflected the professional values of the librarians who selected them. Early twentieth century children's librarians, Jenkins (1996) argues, valued "the inner workings of the child's imaginative life" in contrast with educators, who valued "the 'here and now' of children's lived experiences" (p. 819), and this professional value is reflected in the books that librarians selected for early Newbery Medals. Folklore and historical fiction set outside of the United States dominate the first two decades of Medalists, defining "what was American…through and against contact with the cultural other, usually safely removed across time and/or space" (Kidd, 2007, p. 177). At the same time, the Medal titles reinforced traditional gender values: the boys depicted in Medal-winning titles were heroes, hunters, travelers, and providers; the girls depicted were parts of families, anthropomorphized dolls, writers, and teachers (Association of Library Services for Children, 2016a).

These examples arise from the early decades of the Newbery Medal. The relationship between selection committees, librarian values, and prized books from later decades is not thoroughly documented, but the existing literature suggests a correlation continues. Willett (2001), for example, traces how librarians influenced the revision process of *Rifles for Watie*, the

1958 Newbery Medal-winning title, in order to better reflect shifting conceptions of race in response to the Civil Rights movement. Regardless of time period, Lundin (2004) argues, "librarians can be defined as canon makers who reproduce social hierarchy in a systematic act of tradition bearing" (p. 30). The books selected for the Newbery Medal, as well as the role played by librarians in establishing and codifying Newbery procedures, illustrate this process.

The effect of prizing children's literature. Despite the limitations of the Newbery, which are well established and which I consider at length below, the Medal nevertheless plays an enormously influential role in the construction of the American children's literature canon (Clark, 2003; Kidd, 2007; Marcus, 2008). Given the weight ascribed to the Medal, it is important to interrogate how the award functions and to theorize its role in the production of children's literature as a genre. English (2005) provides a framework for understanding literary and artistic prizes in general, particularly in relationship to the "cultural capital" which these prizes wield (p. 3). Prizes call out and reward artistic merit, English argues, while at the same time providing the cultural elite an establishment against which to rail. This railing, however, is an important part of the work accomplished by cultural prizes, for they "cannot fulfill their social functions unless authoritative people – people whose cultural authority is secured in part through these very prizes – are thundering against them" (p. 25). Kidd (2007) examines the Newbery Medal through the lens of English's framework, exploring the peculiar breed of cultural capital exerted by the Medal. While its selectors might serve as de facto creators of the American children's literature canon by virtue of the Medal's influence on book sales, Kidd argues that at the same time the award represents "edubrow" culture with its merging of middlebrow educational values found in public schools and libraries with the more literary aims of the Medal itself (p. 169).

This understanding of the Newbery points to an unstated double purpose of the Medal. Even though the criteria explicitly exclude didacticism from the award, part of the cultural function that the Medal has come to play is nevertheless educational. Indeed, this argument points to something critics have long known about the Newbery: there is a hidden agenda behind many, arguably most, of the texts that win, and this agenda largely promotes conservative, middle-class, white American values. Cook (1985), for example, demonstrates that the corpus betrays a "striking convergence" around the American value of individual self-reliance (p. 421), and that this convergence shifts over time as conceptualizations of individual self-reliance change. Alberghene (1981) points to the representation of American ideals in foreign settings, and Forest (2014) traces how the corpus perpetuates rags-to-riches mythologies. At the same time, librarian, educator, and publisher critics of the award rail against the most recent selections, arguing that the selection committee chose novelty and literary innovation over books that will stand the test of time (Devereaux, 2008; Silvey, 2008), helping to perpetuate the economy of the Newbery. Cummins (2016), meanwhile, considers how the Medal could use intersectionality to become more socially and culturally diverse while still privileging artistic and literary merit.

The Newbery's Definition of Children's Literature

As an abstract concept, then, the Newbery Medal exerts enormous cultural influence over the field of children's literature. What, though, of the particulars? How does the Newbery Medal conceptualize and operationalize a definition of children's literature? The committee charged with selecting the Medal-winning titles provides its own, arguably unhelpful, definition of children's literature found within the Medal's criteria. The definitional amendments appended to the Medal's criteria are brief and deserve quotation in full:

1. "Contribution to American literature" indicates the text of a book. It also implies that the committee shall consider all forms of writing—fiction, non-fiction, and poetry. Reprints, compilations and abridgements are not eligible.
2. The book displays respect for children's understandings, abilities, and appreciations. Children are defined as persons of ages up to and including fourteen, and books for this entire age range are to be considered."
3. "Distinguished" is defined as:
    - Marked by eminence and distinction; noted for significant achievement.
    - Marked by excellence in quality.
    - Marked by conspicuous excellence or eminence.
    - Individually distinct. (ALSC, 2016c)

This definition of children's literature foregrounds a simple and un-problematized understanding of the genre, focusing exclusively on format and intended audience. The first criterion delimits the purview of the award to the text of a book, despite illustrations being so commonplace in books for children of all ages that they are frequently considered a crucial component of the genre (Avery, 1994; Darton, 1932/1982). Further, it treats children's literature as a static construct and does not allow for consideration of new modalities and technologies that can – and do – radically alter the genre (Serafini, Kachorsky, and Aguilera, 2016). The second definitional criterion appears to set limits to the intended audience for the literature considered, although specific delimitations are not included. Who, for example, decides if a book "displays respect for children's understandings, abilities, and appreciations" (ALSC, 2016c)? Is it the publisher, the committee, the child reader, or the parent responsible for purchasing the child's reading material? Further, while ostensibly an award for children's literature from birth through

early adolescence, the Newbery nevertheless privileges literature for the middle grade reader (Kidd, 2007; Schafer, 1976; Schafer, 1986). The third definitional criterion is problematic in much the way that the second is: who determines what merits distinction? What, for that matter, is excellence? As Bittner and Superle (2016) note, "excellence" remains undefined, and since all committee deliberations remain confidential, the only evidence available lies in the books selected for the Medal. The Newbery Medal, ostensibly the gold standard for children's literature, largely defines what it awards through an over-reliance on vague adjectives.

Limitations of the Newbery

My selection of Newbery Medal-winning titles for analysis is not intended to suggest that the Newbery seal on a book cover provides an unquestioned stamp of literary merit. Indeed, it is crucial to acknowledge the limitations inherit within the corpus of Newbery titles. Critics previously noted many of these limitations: women authors and female protagonists have historically dominated the award (Pease, 1939; Jenkins, 1996; Clark, 2005); historical fiction is preferred above all other modes; progressive (but not too progressive) social values proliferate (Kidd, 2007), particularly in Honor rather than Medal-winning books (Cummins, 2016); and books for advanced child readers, typically grades 6 and above, predominate (Schafer, 1976; Schafer, 1986).

Importantly, the Newbery contributed to the creation and perpetuation of what Nancy Larrick (1965) termed "the all white world of children's books": white authors, white characters, and white cultures predominate in the corpus of Newbery winning titles. In fact, an African American was not a central character in a Newbery Medal-winning text until 1951, when *Amos Fortune, Free Man* won the award. Although some critics have suggested that multicultural perspectives can be found within the corpus (e.g., Gillespie et al., 1994), others have charted the

problematic approach to depictions of race embodied by the Newbery, including a tendency to treat all races other than white as the other, with individual characters typically embodying exceptionalism (e.g., Clark, 2007; Madsen and Robbins, 1981; Miller, 1998; Wilkins, 2009). Indeed, the Newbery earned such a reputation for whiteness that the ALA and other professional organizations created additional awards, including the Coretta Scott King Award, to address the limitations of the Newbery (Wilkins, 2009, p. 7). Cummins (2016) explores the limitations of identity-based awards, arguing that the Newbery's continued, persistent whiteness problematically foregrounds the idea of whiteness as a neutral identity, resulting in no noticeable change in children's literature publication practices.

It is also important to note limitations in the types of texts that the Newbery encompasses. Fiction dominates the award, despite the proliferation of high-quality, even literary, informational texts for children throughout the twentieth and twenty-first centuries (Stevenson, 2011). Although the award criteria explicitly include literature for audiences from birth to age fourteen (Association for Library Services to Children, 2016c), texts for the middle grade child reader nevertheless dominate (Kidd, 2007; Schafer, 1976; Schafer, 1986). A picture book with a publisher's designation "intended for ages 3-5" did not win the Medal until 2016, when Matt de la Peña won with *Last Stop on Market Street*. This limitation is, perhaps, unsurprising, given that the Newbery criteria explicitly excludes the consideration of visual elements and illustration from the award committee's deliberations. While there are many examples of picture books with text worthy of the award, picture books are designed to use word and image codes symbiotically (Nikolajeva and Scott, 2013; Nodelman, 1989; Schwarcz and Schwarcz, 1991). To ignore the visual elements, as the Newbery does, is to ignore at least half of the book's meaning. Finally, as Silvey (2008) and Miller (2014) note, the Newbery is known for favoring formal

experimentation and the juvenile equivalent of literary fiction rather than books children would actually choose and read independent of adult intervention. When children's reading preferences inform the selection of prize-winning texts, the resultant corpus looks very different than the Newbery's (Miller, 2014).

Methodological Approaches to the Newbery

Along with the limitations inherent within the Newbery corpus itself, there are limitations to previous methodological approaches to studying the Medal. Scholars from literary studies, education, and library science have historically approached the Newbery from different angles, sometimes using different methods, and frequently reaching different conclusions about the award or the text under consideration. The amount of attention, or lack thereof, paid to the Newbery in these fields suggests some of the differences in approach as well as value placed on the award. For example, by mid-2017, the MLA International Bibliography indexed 39 distinct works that consider the Newbery (22 journal articles, 11 dissertations, three books, and three book chapters), the ERIC database indexed 86 results (34 ERIC documents, 28 journal articles, 24 magazine articles, and 10 books), and Library Literature & Information Science Full Text indexed 374 (209 magazine articles, 125 trade publications, 72 journal articles, 26 books, 17 book reviews, 15 biographies, and one conference paper).[7] In addition to quantity, the variation in types of sources indexed by these databases speaks to disciplinary divides in approaches to studying children's literature. The MLA does not provide access to book reviews or discussion on the Newbery in trade or popular press publications, and the Library Literature & Information Science Full Text database buries academic discussions of the award under interviews and trade publications. ERIC, meanwhile, presents a balance of unpublished research on the Medal, peer

[7] Searches conducted on June 17, 2017 using the keywords "Newbery Medal."

reviewed journal articles, and books. Some overlap exists between all three databases, however, particularly in the peer reviewed journal articles indexed.

Although observable disciplinary differences exist, there are also commonalities. Beyond extensive cross-indexing of peer reviewed journal articles, the scholarship on the Medal displays some striking similarities across disciplines in methodological approaches to understanding the Newbery. Previously, most scholars approached the Newbery Medal and Newbery Medal-winning texts through small-scale studies, typically employing a range of microscopic approaches to analysis across a small sampling of texts. In the following sections, I explore the different methodological approaches employed to study the Newbery Medal and consider their affordances and limitations.

The Newbery as sampling strategy. Scholars frequently use the Newbery as a method for selecting a text to analyze or creating a sample of text sets through which they analyze elements other than the Medal-winning status of the texts. Sampling approaches include three tiers: single author or text studies, small groups of authors or texts, and larger scale studies.

Single author and single text studies. A common methodological approach considers individual Medal-winning texts or Medal-winning authors. In these studies, scholars focus on individual authors or texts as a method for exploring a specific aspect of children's literature as represented in a specific author's work (e.g., Halliday, 1999; Nodelman, 1990; Schneebaum, 1990). The award winning status of the work under consideration is typically secondary to another element: *A Wrinkle in Time*'s depiction of feminism and the construction of womanhood (Schneebaum, 1990), the depiction of gender and identity in the works of Eleanor Esetes (Smulders, 2015), the relationship between van Loon's history and fictional constructions of the past (Nodelman, 1990), the relationship between base text and film adaptation in *Sarah, Plain*

and Tall (MacLeod, 1998), the revision history of *Rifle's for Watie* (Willett, 2001), or the treatment of place in *Out of the Dust*'s narrative verse (Halliday, 1999).

In existing single author studies, the Newbery Medal-winning designation is secondary to the textual or cultural factor under consideration. The fact that the Newbery Medal committee deemed the text under analysis the most distinguished contribution to American children's literature in the year it was published is incidental. More often than not, these studies mention the Newbery Medal merely as a descriptor, with the canonical designation suggested by the award going unnoticed and unanalyzed. In fact, many examples of criticism on Newbery Medal-winning texts cannot be found with a search for the phrase "Newbery Medal." Instead, the researcher must search for the author's name or the title of the winning book.

Small group studies. Another sampling strategy relies on the Newbery to create a small cluster of authors or texts for analysis. As with single author or single text studies, these studies almost exclusively use Newbery winning texts to examine something other than the Newbery Medal. Dyson analyzes the Newbery Medalists published during World War II in order to investigate American responses to the conflict. Leininger, Dyches, Prater, and Heath (2010) selectively sample Newbery Medal and Honor books published between 1975 and 2010 for depictions of characters with disabilities. This subset of the Newbery corpus forms the basis for their analysis. Nelson (2011) reads Newbery Medal-winning texts from 1930-1950 to establish a canonical view, or "a shorthand for cultural approval" (p. 499), of children's literature for girls against which to read the non-canonical author, Sally Watson. In contrast, some studies focus on authors who produced numerous examples of children's literature, using the awards that the author won as a mechanism for narrowing down the pool of titles for consideration within the study. Roggenkamp (2008), for example, explores Cynthia Rylant's subversion of the

Appalachian hillbilly stereotype exclusively through a consideration of her Caldecott Honor and Newbery Medal-winning books, without considering Rylant's larger body of work. Jenkins (1996), analyzing early Medal-winning titles in conversation with archival research on early critiques of the Medal, provides the only small-group study to date that examines a subset of Newbery Medal-winning titles to understand the Newbery Medal.

Larger scale studies. Some studies, however, do consider a much larger sample of the Newbery Medal corpus. In an early study, Cook (1985) uses Medalists from 1941 to 1981 to chart shifting implications in social conceptualizations of self-reliance. Cook notes that his selection of the Newbery is both practical, resulting in a manageable text set, and ideologically driven as the Medalists were "highly regarded and widely distributed" (p. 425). Despain et al. (2015) examine all Newbery Medal and Honor books between 1930 and 2010 in their content analysis investigating depictions of family structures over time. Kidd (2007) considers the Medal holistically in his consideration of the cultural work accomplished by prizing children's literature. He does not, however, explicitly define which works inform his analysis or present a methodology for analysis beyond the conceptual framework offered by English's work (2005) on literary prizes for adult literature. Building on Kidd's (2007) work, additional studies consider the cultural work accomplished by the Newbery Medal. Cummins (2016) reads the Newbery against identity-based awards, while Bittner and Superle (2016) consider the role played by privileging formalism and aesthetics in creating overly restrictive canons of children's literature. Like Kidd's earlier work, these considerations do not define which works inform the analysis or present a methodology beyond the theoretical frameworks that guide the examination of the Medal.

Studies focusing on readability of Medal-winning titles frequently make use of a larger sample size than other approaches. In two separate studies, Schafer (1976, 1986) analyzed the readability of Medalists between 1940 and 1986. Clements, Gillespie, and Swearingen (1994) computed readability formulae for all Medalists up to 1991, and Stevens (2010) extended their work to consider all Medalists up to 2010.

Analytical lenses. Existing studies of the Newbery Medal and Medal-winning titles employ a range of analytical lenses. These lenses occur across studies with varying sample sizes. Very few studies, however, identify the analytical lenses adopted, requiring the reader to infer the methodological and analytical tools employed.

Textual analysis and close reading. Many of the studies mentioned above favor textual analysis enabled, at least in part, through the close reading of a small number of selected texts. Although different scholars employ close reading for different purposes, close reading in these studies typically offer insight on how individual texts, or a small group of texts, work on structural, literary, and formal levels. Some look specifically at text structure and narrative strategy (Halliday, 1999; Simon, 2008), while others investigate how the text reflects cultural assumptions and values (e.g., Nodelman, 1990; Roggenkamp, 2008; Schneebaum, 1990). Although Kidd's (2007) consideration of the Newbery Medal is largely theoretical, he nevertheless relies on close readings of individual texts, pointing to individual examples of larger thematic or generic characteristics. Given the close, microscopic focus of these studies, the extant criticism contains much information about individual Newbery Medal-winning texts and their literary qualities, beginning with Nodelman's (1990) consideration of the first Newbery Medalist, Hendrik Van Loon, and continuing through the high points of the Medal's history. Not

all Medal-winning texts are the subject of textual analysis, however, suggesting gaps in critical knowledge about some of the texts in the corpus.

Although literary studies traditionally lays claim to close reading, I note that close reading occurs across scholarship on the Newbery Medal from literary studies, library science, and education. Willet (2001), for examples, writes from a library perspective and juxtaposes close readings of revised editions of *Rifles for Watie* with archival research on the role played by librarians in the revision process. Forest (2014), meanwhile, writes from an education perspective and examines rags-to-riches stories in select Newbery Medal-winning texts, using both content analysis and close reading to inform her argument that thoughtful text selection, looking past the Newbery Medal sticker on the front of some books, in the classroom is required. The implications of library science and education studies making use of close reading may differ from literary studies, but the methods employed are the same.

Cultural studies via content analysis. Another common approach uses the Newbery Medal as a sampling tool to create a corpus for examining the relationship between children's literature and the culture in which sample texts were created. This examination occurs via content analysis, although approaches to content analysis vary greatly from study to study. Some, such as Forest (2014) and Leininger, Dyches, Prater, and Heath (2010), directly align with Krippendorff's (2004) classic conceptualization of content analysis; others, such as Dyson (2007) and Moir (1981), employ the techniques of content analysis without ever identifying them by name. As with the studies employing textual analysis and close reading discussed above, these studies privilege the relationship between culture and text over the titles' award-winning statuses.

Despite variation in methodological specificity, these studies follow a similar structure. All identify a specific cultural phenomenon under investigation and then identify a subset of the Newbery Medal-winning texts that will form the basis for exploring that phenomenon through literature. These studies then proceed to identify shared themes, patterns, and motifs across texts. Forest (2014), for example, relies on inductive coding to reveal socioeconomic statuses of characters and inform a discussion of the portrayal of upward class mobility in Newbery Medal and Honor titles from 2009-2013. Similarly, Moir (1981) analyzes the Newbery Medal and Honor winning texts from 1952-1961 to explore how the children's books of the Eisenhower era reflect society's values for children, and Dyson (2007) relies on Newbery and Caldecott Medal winners published during World War II to explore how historical fiction can be used to portray contemporary concerns. Lathey (2005) turns to Newbery Medalists to find samples of protagonists from the 1990s to compare to those from the 1950s and chart the impact of New Realism on child psychology. St. John (1981) uses Newbery winners from 1971 to 1981 to explore an increasing prevalence of social realism and the problem novel within the children's literature. Powell et al. (1998), meanwhile, rely on the Newbery to provide a text set for exploring how gender roles change in children's literature over time, and Despain et al. (2015) investigate the representation of family structures depicted in Medal-winning texts to census data on actual family compositions. Fleming and Parker (2013) perform content analysis on a random sample of Newbery Medal-winning texts to explore the representation of Biblical virtues in the corpus.

These studies cover a wide swath of the Newbery Medal, both in terms of chronological coverage and themes analyzed. These studies do not, however, offer a unified approach to the employment of content analysis, and not all of them explicate in any detail the methodological

approach adopted. Further, although content analysis of an entire corpus can provide a fruitful method for macroscopic reading (Hoyt, Long, Tran, and Hughes, 2015), all existing content analyses of Newbery Medal-winning texts rely on microscopic readings strategies. Scholars read the books under consideration; code the data, either through explicitly discussed coding strategies (e.g., Fleming and Parker, 2013; Forest, 2014) or implied coding activity (e.g., Dyson, 2007); and analyze the ways in which the texts read for the study illuminate understandings of American history and culture. None of the existing studies look at the intersection of multiple themes or content areas, none examine the relationship between formal and structural changes in the genre to thematic and content changes. Existing content analyses of the Newbery Medal exhibit the strengths of microscopic reading strategies in their reliance on human coding and its resulting accuracy. They also betray weaknesses; there are limits to the amount of information that microscopic approaches to content analyses can cover.

Content analyses of the Medal rely on a tacit understanding of the role that the Newbery serves in shaping the types of children's literature available, particularly in the classroom and the school library, but this understanding is never fully developed. With the exception of Fleming and Parker (2013), who employ a random sample of Newbery Medal-winning texts and argue that their results are therefore generalizable to the rest of the corpus, there is very limited consideration of how the findings from a limited pool of texts might inform an understanding of the rest of the corpus.

Critical considerations of race. Importantly, studies from library science and education perspectives introduced scholarship on the limitations of the Newbery Medal, particularly in terms of diversity. Nancy Larrick, an educator, was the first vocal critic of children's literature's whiteness (1965), and her observations apply to the Newbery Medal as well as to the field more

generally. Although Gillespie et al. (1994) used content analysis to argue that it is possible to locate characters of color among Newbery Medalists and that the corpus can therefore assist in multicultural education, most critics instead point to the continued whiteness of the corpus (Cummins, 2016). Using critical race theory and focusing on individual Medal-winning books, scholars have explored how Medalists tend to treat all races other than white as the other, with individual characters typically embodying exceptionalism (e.g., Clark, 2007; Madsen and Robbins, 1981; Miller, 1998; Wilkins, 2009). Others, using close reading techniques in conjunction with critical race theory, have argued that individual Medal-winning texts have whitewashed history, with narrative silences erasing the presence of non-white characters from the story (Simon, 2008). Cummins (2016) turns to identity theory to consider the social and cultural work that the Newbery Medal could, but does not, currently perform in addressing the still all too white world of children's literature in general and the Newbery in particular.

To date, most scholarship criticizing the Newbery Medal for its whiteness came from either a library science or education perspective. As such, scholars tend to suggest practical implications for their work. They encourage practicing librarians to consider the ramifications of the Medal in the provision of library services, especially collection development (Hill, 2011; Horning, 2015; Miller, 1998; Wilkins, 2009), or the crucial need for critical literacy practices in conjunction with the use of whitewashed texts in the classroom (McKoy, Lowery, and Baglier, 2016; Simon, 2008).

Readability Measures. A number of studies chart readability measures of the Newbery Medal. This work is largely quantitative, explicating methods for computing readability measures for the Newbery Medal and finding that, at least through the mid-1980s, the Newbery Medal more frequently honored more complex texts intended for older children. Readability

studies of the Newbery follow a common formula: define and identify readability measures, introduce the Newbery Medal, and compute selected measures for Medal-winning titles during certain date parameters. The readability formulae selected for computation and analysis vary across each study, with selection based in large part on the date of the study's completion, and usually based on sentence and word lengths. The earliest studies calculated Botel levels and the Fry Readability Formula (Schafer, 1976; Schafer, 1986). Later studies calculated the Fry Readability Formula, Flesch Reading Ease, Flesch-Kincaid Formula, and the Gunning Fog Index measures (Clements, Gillespie, and Swearingen, 1994), and the Flesch Reading Ease, Flesch-Kincaide Grade Level, Gunning Fox Index, and SMOG readability measures (Stevens, 2010). These studies consider nothing beyond readability of Medal-winning texts, grade levels suited for reading Newbery Medal-winning texts in general, and how text complexity in the corpus has changed over time. None of these studies connect readability to content, theme, or genre.

Leal and Chamberlain-Solecki (1998) rely on existing research on readability measures to selectively investigate readability and student interest in Medal-winning titles, arguing that text complexity alone is a poor indicator for selecting books students might be interested in reading and suggesting that a focus on content is crucial. Other research building on quantitative approaches to readability within the Newbery corpus explore the ways in which the text complexity of (Broemel, Wysmierski, and Gibson, 2014) or student interest in (Friedman and Cataldo, 2002) Medal-winning texts might affect individual students and learning types in the classroom.

Opinion pieces and reactions to new Medalists. Within library science, opinion pieces on the Newbery proliferate, particularly after an "unpopular" or controversial book wins the award, as exemplified by the reactions to *Good Masters! Sweet Ladies! Voices from a Medieval*

Village's win in 2008. These pieces point to the Newbery winners of old, suggesting that more recent committees have favored the unusual, the unconventional, and the overly quirky (Silvey, 2008) or selecting a "compromise" winner because the most distinguished work would be too controversial (Devereaux, 2008). Another common opinion-based approach includes the author profile (e.g., Bird, 2013; Horning, 2004; Hong, 2002; Imdieke, 2012; Sutton, 2009), presenting an interview with a recent Medalist and the author's reaction to his or her book's win. Finally, library science periodicals frequently examine a Medal-winning title's critical reception, especially when the reception is negative (Bosman, 2017; Schreiber, 2017). It is important to note that these are opinion pieces, not in-depth scholarly considerations. As such, they rely on cherry picked examples to make their case. There is no attempt to quantify assertions about previous Medal winners' popularity or staying power.

Bibliographies. Another important function of library science scholarship on the Newbery lies in the construction of annotated bibliographies describing individual Medal-winning titles and their authors (Peterson and Solt, 1982; Solt, 1981) as well as compendia of reference materials about the award (Association for Library Service to Children, 2001; Gillespie and Naden, 2006; Kingman, 1965; Kingman, 1975; Kingman, 1986). Although seemingly antiquated from the perspective of 2017, such works were invaluable information sources before the advent of online reference resources, particularly for collection development. They remain crucial tools for gathering large quantities of specialized information as well as materials from the mid-twentieth century that have poor coverage in online reference resources.

Critical Understandings of the Newbery and the Need for Distant Reading Methods

The extant scholarship on the Newbery Medal frequently relies on a narrow scope. Most studies of the Newbery focus on a limited number of authors, texts, themes. As such, existing

scholarship highlights critical understandings about single authors, single texts, or small clusters of texts, usually chosen for their thematic similarities or publication date. A small sliver of criticism aims to understand the theoretical and cultural work accomplished by prizing in general and the Newbery in particular, albeit frequently without defining the methodological apparati that lead to that understanding. Another commonly observed trait in existing scholarship is the use of the Newbery as a sampling strategy to investigate one element, usually unrelated to the Medal, such as the depiction of race (Clark, 2007; Madsen and Robbins, 1981; Miller, 1998; Wilkins, 2009), family structure (Despain et al. 2015), or socio-economic status (Forest, 2014). Limited larger-scale studies exist, but with the exception of Kidd's (2007) theoretical exploration of the role played by prizing in the construction of canonical American children's literature, these studies investigate a single element, usually related either to readability (Clements, Gillespie, and Swearingen, 1994; Schafer, 1976; Schafer, 1986; Stevens, 2010); a single thematic element, such as political education (Cook, 1985); or a single theoretical perspective, such as intersectionality (Cummins, 2016) or formalism (Bittner and Superle, 2016). Regardless of methodology, almost all studies ignore the implications of using the Newbery Medal as a sampling strategy.

Despite a preponderance of criticism stemming from small samples of Medal-winning texts, scholars, journalists, and practicing librarians and teachers claim to know much about the Medal in its totality: women and female protagonists dominate (Jenkins, 1996); Newbery Medal-winning titles sell better and continue to be read more than other works (Kidd, 2007; Clark, 2003); historical fiction abounds (Cook, 1985; Dyson, 2007); somewhat progressive moral and social values predominate (Alberghene, 1981; Kidd, 2007); and small town life provides a preferred setting over cities, unless the setting is exotic (Alberghene, 1981; Solt, 1981); and

recent committees favor formal experimentation at the expense of enjoyable texts that children actually read (Devereaux, 2008; Silvey, 2008).

These observations directly influence my study and the methods that I employ. As Moretti (2005) noted, canonical understandings of literary genres are overly reliant on small samples of texts. With this review, I suggest that existing understandings of the Newbery Medal are overly reliant on a small sample of texts. As a partial remediation of this trend within the criticism of children's literature, I use the entire corpus of Newbery Medal-winning texts to explore a subset of children's literature as genre, employing holistic methods. At the same time, however, it is important to stress again that although my use of the Newbery as a sampling strategy is intentional, it nevertheless irrevocably affects the implications and definitions that I posit. Using the Newbery provides data on *one* sub-genre of children's literature, not *the* genre of children's literature. This study, then, represents my attempt to generate distant reading tools by which critics can understand a multiplicity of children's literature genres, not a study that seeks to define children's literature as a monolithic structure.

Disciplinary Siloes and the Affordances of Interdisciplinary Paradigms

This study is intentionally interdisciplinary, relying on data sources, methods, and critical approaches from library science, education, and literary studies. Scholars across these three disciplines study children's literature in very different but also complementary ways. Disciplinary differences rather than confluences take prominence in existing discussions on the topic, often to the detriment of identifying interdisciplinary approaches to investigating shared critical questions. As Nel and Paul (2011) note, "children's literature" is an "umbrella term," with very little interdisciplinary crosspollination of critical vocabulary or methodology occurring (p. 1). These differences have been explored many times, but the dominant discourse reaches

conclusions following a typical pattern: studies from library science and education care more about children than literature, while literary studies focuses on text at the exclusion of the child for whom the text was originally written.

One of the earliest and still frequently cited considerations of disciplinary differences in approaches to the study of children's literature helped establish this view of siloed paradigms. As it represents the inspiration for many subsequent assertions about disciplinary approaches to the study of children's literature, this now-dated source deserves consideration. First presented at the 1980 Children's Literature Association Annual Conference as a panel session and subsequently published as three complementary articles in the conference's proceedings, this panel firmly established literary studies as the domain of text analysis in children's literature, with education and library science focusing on the children who read children's literature. Bingham (1980), representing the perspective of teaching children's literature in a college of education, noted that her course emphasized "*literature* and *children* equally" (p. 70, emphasis in original), with coursework designed to help students explore genres and gain skills needed in the classroom. These skills included reading aloud and deciding, through a visit to an elementary school and giving book talks to children, "which book might be most appropriate for a particular group of children" (p. 73). Laughlin (1980), presenting the results of a survey of library school professors who taught children's literature, described the pedagogical approaches to children's literature found in library science curricula across the United States. Once again, Laughlin described the purpose of studying children's literature as largely practical: the courses enabled students to evaluate materials critically, demonstrate familiarity with a broad range of materials, recognize characteristics and needs of children at different age levels, demonstrate storytelling techniques, demonstrate facility with selection tools, identify current trends, and understand reader's

advisory techniques (p. 76). In stark contrast, Anderson (1980) offered a perspective from an English literature department:

> I want my students, through a careful and critical scrutiny of fine children's books, to stretch for a knowledge of the quality of childhood, as well as to attempt to rediscover the child that still resides within them. In this way, I feel they may begin to recognize excellence in literature. (p. 80)

Anderson's focus was not on the child but on the text, foregrounding the finding of "questions of social value" (p. 81) and applying the techniques of "close reading and logical thinking" (p. 82) to children's literature.

Subsequent scholars, particularly those representing the literary studies side of the debate, repeat the claims that this panel made (see, for example, Clark, 2003; Kunze, 2015; Nel and Paul, 2011; Bittner and Superle, 2016). Most recently, the 2016 Children's Literature Association annual conference revisited the original panel session, with different scholars, at a session titled "Core Competencies for Students of Children's Literature: A Conversation about Disciplines, Pressures, and Priorities." The conclusions reached were strikingly similar to those from the original panel, despite the shift in terminology from pedagogical approach to core competencies.

Critics can interpret the disciplinary differences found in the scholarship about children's literature as mirroring the pedagogical differences explored above. As Clark (2003) notes, the scholarship published by librarians often looks very different from that of literary critics. From this perspective, library science typically focuses on bibliography (e.g., Kingman, 1965; Kingman, 1975; Kingman, 1986), selection criteria and the role played by librarians in selection (e.g., Eddy, 2006; Hearne, 1996; Jenkins, 1996; Kimball, 2012; Martens, 2013; Wiegand, 2005;

Wiegand, 2011) and review essays (e.g., Allen, 2011; Devereaux, 2008; Hill, 2011). Viewed superficially, library science scholarship can sometimes seem to share little common ground with the children's literature studies found in education. For educators, the focus in children's literature tends to be on matters related to pedagogy, such as readability and reading comprehension (Lysaker & Hopper, 2015; Topping, 2015), the relationship between children's literature and social justice pedagogy (Hasty, 2015; Lacy, 2015; Parsons & Castleman, 2011; Smulders, 2015), motivating reluctant readers (Fingon, 2012; Gabriel et al., 2012), and, more recently, the effect of the Common Core State Standards on the inclusion of literature in the elementary classroom (Eppley, 2015; Groth, 2015; Hiebert, 2015; Möller, 2015).

The conclusion that literary studies, library science, and education use different disciplinary paradigms when researching and teaching children's literature is, however, reductive. It reifies claims stemming largely from literary studies perspectives that English departments provide the intellectual home for text analysis (Clark, 2003; Nel and Paul, 2011) and ignores perspectives that point to the possibilities provided by interdisciplinary approaches as well as the many different ways of teaching and writing about children's literature that exist. The Children's Literature Assembly, a part of the National Council of Teachers of English, hosts a syllabus exchange website that destabilizes the canonical story of siloed pedagogical approaches. The exchange includes many examples of discipline-specific syllabi, particularly at the graduate level for courses focused on concrete professional duties in librarianship or education, but these classes also include significant emphasis on an appreciation for and understanding of the many different types of texts that constitute the field of children's literature (e.g., Vardell, 2011; Zaleski, 2011). The exchange also highlights the many different approaches to teaching children's literature in education departments. Some focus specifically on literary texts and the

children's literature marketplace (e.g., Aziz, 2011; Crisp, 2011). Other focus on undergraduate general education seminars that can simultaneously satisfy pre-service teacher training requirements (Liang, 2011) or offer special topical investigations, including but not limited to picture books (Graff, 2011), poetry (Allen, 2011), and global or diverse children's literature (Short, 2011; Wilfong, 2011). None of these examples come from departments of English. All focus on children's literature as text without foregrounding or, in many cases, even mentioning real children.

Similarly, the disciplinary differences in scholarly approaches to children's literature are more complicated than the traditional story, cited by Clark (2003), would suggest. Clark's argument relies on the observation that the professional purpose and implications of each discipline's scholarship are quite different: librarians study children's literature to understand texts and therefore make more informed decisions about the texts that they buy for their collections. Educators, meanwhile, study children's literature to understand texts and therefore help future educators make more informed decisions about the texts that they use in their classrooms, and literature scholars study texts at to understand those texts and the cultures in which they were written. Again, as with the pedagogical examples, this conclusion is reductive. As the literature I reviewed above on methodological approaches to the Newbery Medal shows, all three disciplines use methods of text analysis, content analysis, historical analysis, visual analysis, and other lenses to examine children's literature. All three disciplines also consider the function of the child reader (Nodelman, 2008), the adults who create and perpetuate the children's literature marketplace (Falconer, 2009; Griswold, 1992; MacLeod, 1994; Marcus, 2008; Murray, 1998; Nodelman, 2008; Stephens, 1992), and the relationship between children's literature and the social constructs that literary texts often reflect (Zipes, 2013).

A significant difference in scholarly approach that Clark does not note, however, relates to sample size. Traditionally, literary studies rely on a small number of texts for analysis, with the sample frequently formed from a priori categories of canonical texts. Three classic and still frequently cited examples of literary criticism on children's literature as genre provide illustrative examples regarding sampling strategy. Nodelman (2008) begins his extensive consideration of children's literature as genre with a discussion of only six canonical literary texts. The chapters that follow consider the extant scholarship on children's literature and genre in relationship to those six texts. Rose (1984) takes a single text, *Peter Pan*, and its subsequent textual incarnation as the basis for her generic explorations. Shavit (1986) offers a reading of non-canonical children's literature via a case study of only one author, Enid Blyton. The field of literary studies, however, does not hold a monopoly on small samples of texts informing larger conversations about genre. Nikolajeva (1996, 2002, 2009, 2013), covers much the same generic ground, issuing from a college of education, as Nodelman, Rose, and Shavit do from their respective departments of English. Instead of limiting her consideration of children's literature as genre to a pre-defined small number of texts, Nikolajeva creates definitional categories and provides handpicked exemplars that illustrate those categories. Nodelman, Rose, Shavit, and Nikolajeva all offer understandings of an incredibly wide range of children's literature vis-à-vis very small samplings of exemplar texts. The Newbery Medal scholarship stemming from overtly literary studies perspectives mirrors the textual balance found in these considerations of children's literature as literary genre. These studies focus on single texts, single authors, or small clusters of authors. The findings from these limited studies goes on to inspire assertions about the entire corpus.

Scholarship on children's literature from education and library science can, however, provide models for studying children's literature using larger sample sizes. Despain (2015) samples four decades worth of Medal-winning titles in an exploration of shifting depictions of American family structures. Studies charting text complexity in the Newbery Medal refer to the entire corpus at the time of study completion (Clements, Gillespie, and Swearingen, 1994; Guidry and Knight, 1976; Schafer, 1976; Schafer, 1986; Stevens, 2010). Crisp and Hiller (2011) study the entire corpus of Caldecott Medal-winning texts in their exploration of depictions of gender and sexuality in children's picture books. The conclusions reached in these studies typically point back to implications for library collection development practice or classroom pedagogy, but they also provide an important shift in perspective. Larger scale studies shift perspective from individual texts to broader patterns, with these patterns inspired by a consideration of more than a handful of texts. With these observations in mind, I turn to a consideration of distant reading methods and ask what large-scale studies can offer to an understanding of a corpus of children's literature texts.

Chapter 3: Methods for Reading the Newbery from a Distance

Literary analysis is not known for its attention to methodological detail. In fact, the opposite is true. Underwood (2016), a scholar with a reputation for applying quantitative methods to the study of literature, goes so far as to argue that "literary criticism has little reason to exist unless it's interesting; if it bogs down in a methodological preamble, it's already dead" (n.p.). A recent case study, however, suggests the potential importance of "bogging down" readers of literary criticism with more methodological detail. In a review essay, Nelson (2016) praises two literary and cultural historians who relegate their methodological explications to either an appendix or a supplementary website. He subsequently concludes, "their method takes a backseat to their argument. However innovative and technically impressive they might be, both of them treat their computer aided methods as a means to an end rather than an end in itself" (p. 135).

Although he intends this observation on a backseat approach to methodology as praise, Nelson also notes that certain questions remain unanswered, particularly in his assessment of Wilkens' (2013) reliance upon novels at the exclusion of all other types of writing in his discussion of place in American literature during the Civil War. Upon closer examination, however, Wilkens actually addresses Nelson's criticism, but he limits this consideration to the methodological appendix that Nelson praised so highly, which contains "information about the texts included and methods employed…The appendix also includes discussions of the unique cultural position occupied by novels during the period, of the quality and limits of the data involved, and of the challenges unique to corpus-based analysis" (p. 807). Moving this

potentially uninteresting information from the body of the argument, then, may result in a more readable piece of literary analysis, but it does not necessarily aid in a reader's understanding of that text or the subsequent analysis of data. This case of misunderstanding, I would like to note, occurs in the familiar domain of literary genre and selection of the type of text upon which to build an argument. How much more room for misunderstanding might there be when unfamiliar methods, such as the quantitative analyses that form the backbone of Wilkens' argument, are applied?

In this dissertation, I investigate what happens when methodology is not relegated to an appendix by intentionally foregrounding the methodological apparati employed. In addition to exploring how children's literature changes structurally, formally, and thematically over time, I explore whether shifting methodological approaches leads to different conceptualizations of children's literature. In order to explore these questions, I interrogated the Newbery Medal corpus from distant, frequently computational, perspectives, in conversation with what is already known about the text set from more microscopic inquiry. In doing so, I interpreted the data sets in order to better understand the cultural and social work accomplished by prized American children's literature as a genre.

Research Questions

The research questions guiding my study are:

1. How do descriptive, structural, and thematic variables illustrate the formal characteristics of the corpus?
2. In what ways can statistical and descriptive data be used to address common assertions about the Newbery Medal corpus?
3. How do descriptive, structural, and thematic variables vary across data sources?

4. How do different measures of popularity vary across data sources?
5. In what ways does macroscopic reading contribute to a more nuanced understanding of children's literature as genre?

Distant Methods for Analyzing Children's Literature

Conventional wisdom, which finds a voice in the decades of scholarship produced about the Newbery Medal, makes a number of assertions about the award: women authors and female protagonists dominate, leading to a lack of boys' perspectives in the corpus (Jenkins, 1996); Newbery Medal-winning titles sell better and continue to be read more than other works (Kidd, 2007; Clark, 2003); historical fiction abounds (Cook, 1985; Dyson, 2007); somewhat progressive moral and social values predominate (Alberghene, 1981; Kidd, 2007); small town life is preferred over city settings, unless the setting is exotic (Alberghene, 1981; Solt, 1981); and recent committees have favored formal experimentation at the expense of "good" children's literature that children might actually read and enjoy (Devereaux, 2008; Silvey, 2008). Scholars largely derive these observations, however, from thematic analysis of small subsets of Newbery Medal-winning texts, paying little attention to structural and formal elements of the genre as well as texts that disrupt the prevalent models. These observations also tend to lead to assertions about thematic characteristics without quantification. Are a significant number of Newbery Medal books really by women and about female characters, as early critics of the Award claimed (Jenkins, 1996), or does it just seem that way, particularly when readers consider a specific subset of Medal-winning titles? Further, what precisely constitutes a work of socially "safe" historical fiction? Have recent committees really done a poorer job than their predecessors of identifying works of children's literature with "staying power," or is hindsight selectively blind to the failures of Medal winners previous decades?

Moretti's (2005) notion of distant reading serves as a springboard for my conceptualization of distant reading, and like Moretti I seek to employ a range of techniques so as to explore the genre of Newbery Medal winners in its totality rather than relying on statistical analysis alone. In now-classic studies, Moretti employs methods that are not dissimilar from the social and even physical sciences. He relies on statistical data derived from other scholars' data sets to drive his quantitative analysis of genre (2000, p. 18), geographic models to inform his literary maps, and Darwinian theories of evolutionary biology to structure the construction of his morphological trees. It is important to note, however, that Moretti operates under a strictly postpositivist theoretical framework, and he argues that "[q]uantitative research provides a type of data which is ideally independent of interpretations…and that is of course also its limit: it provides *data*, not interpretation" (p. 9, emphasis in original). In the decade since Moretti introduced the paradigm of distant reading, this postpositivist claim to strict objectivity of data has led to sharp criticism. Prendergast (2005) charts the "positivist antecedents" (p. 45) for Moretti's theories, arguing that Moretti falsely placed "a very large bet on bringing the laws of nature and the laws of culture far closer than they are normally thought to be" (p. 56). Ascari (2014) also argues that "Moretti's tendency to regard distant reading as objective, within the framework of a purportedly scientific approach to the humanities…might be more aptly described as pseudo-scientific" as it "adopt[s] biased views of literature under the mask of objectivity" (p. 2-3). Data, these critics point out, is not neutral; it is reliant upon the human critic both for its gathering and its interpretation.

This clash between Moretti and his critics echoes a larger debate within the field of digital humanities in general: can digital tools and quantitative approaches to literature fundamentally change our understanding of literature and culture, leading to something

resembling objective Truth? Or do digital humanists simply use modern, digital, and quantitative tools to investigate the same questions that "traditional" humanists have studied for centuries, albeit from different perspectives (see, for example, Elson, Dames, and McKeown, 2010; Fitzpatrick, 2012; Michel et al., 2011; Gooding, Terras, and Warwick, 2013; Wilkens, 2015)? Reacting simultaneously to the promising potential of distant reading and accompanying anxieties surrounding the promise of objectivity, some scholars sought a middle ground between Moretti's postpositivist claims and the methodological tools he introduced. These scholars foreground the idea that distant reading methods and tools provide different avenues for exploring culturally and temporally situated artifacts, not data-driven certainties. Levy and Perry (2015), for example rely heavily on Moretti's distant reading framework in order to "reduce" Romantic-period anthologies to statistics in order to quantify the effects of feminism (p. 133). They do not, however, rely upon Moretti's postpositivist lens, arguing instead that their quantitative analysis suggests the need for "a comprehensive view of literature as a social construct" (p. 151). Other critics, meanwhile, modify Moretti's distant reading techniques and bring them to bear on a single text, demonstrating how distant reading can augment pattern recognition invisible to the human eye alone through close reading, leading to more nuanced understandings of texts and the contexts in which those texts were produced (Hayles, 2013; Held 2012).

Drouin (2014) explicitly sets out to bridge close and distant reading due to the methodological shortcomings he identifies in macroscopic and microscopic quantitative analysis; big data, he argues, fails in that it does not provide the ability to investigate the texts which it quantifies, while microscopic text mining does not acknowledge "the work's historical and discursive context" (p. 111). Further, Drouin argues that "effective digital literary study requires

the ability to process data, read well, and interpret both the numbers and the texts in light of each other" (p. 111). These scholars provide a mechanism for drawing from Moretti's methodological innovations without accepting the postpositivist paradigm within which he operates. They also suggests the limits inherit in quantitative approaches to literature alone, without resultant qualitative assessment of what those numbers mean in conversation with the texts which they describe.

Although Moretti's tools have proved fruitful across a wide range of disciplines, ranging from his own comparative literature to British literature (Held, 2012; Levy and Perry, 2015), the study of periodicals (Cordell, 2016; Drouin, 2014) and the history of the book (e.g., Gooding, Terras, and Warwick, 2013; Kirschenbaum and Warner, 2014), to cite just a few examples, they have not been applied in a systematic way to the study of children's literature. Building upon the methods first suggested by Moretti, I seek to interrogate the Newbery Medal corpus as a genre of children's literature from new, frequently computational, distant perspectives in conversation with what is already known about the text set from more microscopic inquiry. Like Moretti's critics, however, I do not assert that this data is objective; rather, I seek to understand and interpret data sets in order to better understand the cultural and social work accomplished by prized American children's literature as a genre.

Importantly, like previous scholars (e.g., Wilkens, 2015; Elson, Dames, and McKeown, 2010; Fitzpatrick, 2012; Gooding, Terras, and Warwick, 2013; and Michel et al., 2011), I do not interpret distant reading as synonymous with quantitative analysis of literature alone. Rather, I interpret distant reading as a tool set that enables the critic to answer a familiar set of questions in a different way than close reading, or, as comparative literature scholar, Wilkens, terms it, "a new set of [tools] for dealing with…abundance" in the identification of patterns and the creation

of "abstractly quantifiable model[s]" (2015, p. 11-12). The end goal of distant reading is not the creation of data or the computation of statistical models. As such, corpus size is less important than thoughtful application of distant reading tools to answer questions about the text set, since "no vantage point is sufficiently distant" to provide objective truth (Booth, 2017). Instead of seeking a perfect statistical model, distant reading enables my interpretation of the Newbery as a genre in conversation with existing models of the Newbery text set.

Utilizing a range of secondary data sources and relying upon the distant reading techniques of quantitative analysis, topic modeling, and data visualization, I generate methods for modeling the Newbery Medal as a genre. I also generate methods for exploring how different secondary data sources reflect the values and conceptualizations of the individuals and institutions responsible for creating those sources and influence the ultimate distant reading of the corpus that those data sets describe. Literary analysis's disdain for fully explicated methods presents challenges in creating a model for employing distant reading techniques on a new type of literature. As no studies in distant reading methods for contemporary corpora of children's literature currently exist, I explicate fully my methods for identifying the elements required and the tools available for distant reading before proceeding to a discussion of selecting data sources.

Drawing inferences from the studies that provide the conceptual framework for this project, I note the following elements as important to distant reading: definition and selection of a corpus (Held, 2012; Levy and Perry, 2015; Moretti, 2005); development of tools for locating, harvesting, and analyzing data (Gooding, Terras, and Warwick, 2013); location of data sources (Moretti, 2005); harvesting data; manipulating and storing data for eventual analysis (Gooding, Terras, and Warwick, 2013); and developing analytical methods that work with the data available for the project (Droun, 2014; Hayles, 2013; Held, 2012).

Tools for Reading the Newbery Medal from a Distance

In its earliest incarnation, Moretti restricted distant reading to statistical analysis of data sets, primarily comprised of secondary sources rather than full-text corpora (Moretti, 2000). Contemporary distant reading practices may still include this activity (e.g., Liddle, 2015; Michel et al. 2011), but a range of activities now exist that support reading from a distance. Other fruitful avenues for distant considerations of texts include quantitative, computer-assisted content analysis (Hoyt, Long, Tran, and Hughes, 2015); coding, including computer-assisted, manual, and hybrid manual/computer assisted approaches (Drouin, 2014; Lewis, Zamith, and Hermida, 2013); and the creation of statistically valid topic models (Buurma, 2015; Goldstone and Underwood, 2014; Long and So, 2016). Although many scholars now rely on full text corpora for distant reading, it is beneficial to remember that, in his original thought piece on distant reading, Moretti employed secondary data generated by other scholars rather than a full text corpus (Moretti, 2000; Moretti, 2005).

Data Sources: Selection, Affordances, Limitations, and Caveats

Distant reading's roots in secondary data sources rather than full text corpora are important given copyright restrictions. Regardless of the specific technique used, the distant analysis of literature requires metadata – that is, data about the literary work under consideration, such as the frequency distributions of specific words – as well as databases that include encoded corpora of literary texts or information about those texts. As Leetaru (2015) suggests, copyright law, which precludes the creation of open access databases of encoded texts, results in gaps in knowledge for contemporary data, including literature. As a result, scholars know more about pre-1923 text sets than their modern equivalents due to the advent of modern copyright laws in America. The unavailability of a full-text corpus for analysis, as is the case with contemporary

American children's literature, does not preclude the ability to employ distant reading methods. It does, however, necessitate the distant reader to find other data sources for consideration. Copyright precludes publishing and sharing full text corpora of contemporary American children's literature. Although copyright laws do not preclude the creation of full text corpora for private use by an individual scholar, such a model is unsustainable and impractical.

Despite the challenges presented by copyright law and a lack of full-text corpora, there are nevertheless other sources of metadata about contemporary children's literature available for harvest, analysis, and discussion. In this study, I create a model for finding data sources describing the Newbery Medal corpus, a sub-genre of contemporary American children's literature, from a variety of sources to enable macroscopic consideration of the corpus in conversation with existing critical conversations about the text set. I intentionally rely on data sets from library science, publishing, education, and readers sources to inform an interdisciplinary, macroscopic investigation of the generic characteristics of the Newbery Medal as a literary corpus.

Data set selection criteria. The data sources and sets that I selected for inclusion in this study provide interdisciplinary perspectives on the Medal, enabling a consideration of how different professional bodies and organizations view the text set as well as a comparison of the each data set's utility for distant reading. It is important, however, to acknowledge that these sources are not the only available data sets that describe corpora of contemporary American children's literature in general or the Newbery Medal in specific (for a full list of data sources considered, see Appendix A). Rather than attempt to analyze any and all data sources describing the Newbery Medal, I instead purposively selected the data sets used in this study for a number

of reasons, acknowledging that these decisions would lead to "conditions of knowledge" affecting the study (Moretti, 2005).

Coverage. First, I selected data sets that provide the fullest coverage possible of the Newbery corpus. This enables the most holistic interpretation of the corpus possible for each variable considered. Each data source selected covers the entire date range of the corpus, although an individual data set may exclude some titles based on the inclusion criteria used in generating the set. Non-prose works, for example, lack Lexile measures and are therefore not found in MetaMetric's Find a Book Database. Relying on chronological coverage precludes using some pre-existing data sets compiled by other scholars that cover a smaller range of Newbery medalists (e.g., Chatham, 1980; Clements, Gillespie, and Swearingen, 1994; Stevens, 2010). This decision limits the data sets available, but it also allows for analysis based on the most holistic coverage available.

Existing data sets. Second, when possible, I relied on existing data sets rather than data sets that required creation from full text. This decision reflects my desire to create a methodology for distant reading contemporary texts that could transfer to other corpora.

Structured and unstructured data formats. Third, I selected sources that represent a variety of structured and unstructured data formats in order to test which data sets prove the most efficacious for distant reading. Some rely on controlled vocabularies, some utilize folksonomies, and others present metadata in the form of natural language.

Children's literature as a genre. Finally, as I designed this study as an interdisciplinary response to discipline-bound considerations of children's literature as genre, I intentionally selected data sets from a range of perspectives, including data sets created by libraries, publishers, educational professionals, and readers.

Critical considerations. It is also crucial to acknowledge that my selection of a data set for analysis does not equate to an endorsement of products or platforms that use those data sets. Indeed, as I discuss below, the pedagogical value of some of the products and platforms that I employ in this study are debatable at best. Further, my selection of a data set also does not provide an endorsement of the controlled vocabulary or data structure included in the set. I fully acknowledge that metadata is not neutral; rather, it reflects the organizations responsible for its creation, preservation, and dissemination. A crucial component of this study, therefore, is my analysis of each data source and its efficacy for distant reading, considering the affordances offered by the data available and the limitations inherent in its creation. Again, I do not assert that distant reading is neutral act; rather, it is interpretive. The findings suggested are inherently and inextricably tied to the methodological decisions that enable distant reading as well as the data elements that enable the distant reading process.

Major data sources. Five databases provide the foundation for the majority of data harvested and analyzed in this study: WorldCat, Accelerated Reader's Bookfinder database, MetaMetric's Find a Book database, Bowker's Books in Print, and Goodreads. Information from the ALA's website devoted to the Newbery Medal completes the major data sources consulted. Implicit in each data source is the inherent purpose, and subsequent bias, of the database's creators, and this bias unavoidably colors the types of data found in each resource. In this section, I consider the affordances and limitations of each major data source as well as the types of data sets that I extract from each in order to create variables for analysis.

WorldCat. For centuries, libraries and library catalogs have provided users with metadata about books and other information objects (Pomerantz, 2015). Individual library catalogs provide metadata about that library's holdings, but WorldCat, a union catalog, provides aggregated

information about the holdings of all its participant libraries. Participant libraries enter information into a shared database, owned and managed by OCLC, and this database displays information to the public via the WorldCat interface. WorldCat is the largest union catalog in existence and includes information from its over 16,000 members in academic libraries, public libraries, research libraries, and special and corporate libraries worldwide (OCLC, 2017a). Indeed, OCLC is so large and ubiquitous that the company has been sued (unsuccessfully) for creating a monopoly on library information systems (Breeding, 2010).

Its ubiquity provides one of WorldCat's greatest strengths. This database includes information on a huge swath of libraries worldwide. Nevertheless, there are limitations to a reliance on data from WorldCat. First, membership skews heavily to the United States and Western Europe and to academic and public libraries (OCLC, 2017b). Second, with over 16,000 libraries adding records to the database, there is an unavoidable level of messiness in the data, such as duplicate records existing for the same item or less than perfect adherence to accepted cataloging practices and controlled vocabularies. Third, WorldCat presents information on a single point in time. The database does not enable a comparison of library holdings information across time. Finally, OCLC creates and markets a variety of different interfaces for accessing the database on which the WorldCat catalog is built, ranging from staff interfaces that allow for editing the database to locally customizable search interfaces that only show users records for information objects available locally. Search results vary, of course, based on which interface is used.

WorldCat relies on the Marc21 data structure to code and present information. This is an older data-encoding standard, and it is unique to libraries. Professional librarians typically enter data into WorldCat, relying on descriptive standards and rules maintained by the Library of

Congress. This element to WorldCat's data constitutes both a benefit and a limitation. It constitutes a benefit as is typically represents the work of a human, trained in analyzing and describing information objects. It constitutes a limitation due to the restrictions imposed by the descriptive rules and controlled vocabularies employed.

The application of subject headings, which exist to help users access information sources based on intellectual content of a resource not reflected in the title, illustrate this limitation. The records for Newbery Medal-winning titles in the WorldCat database most frequently employ Library of Congress Subject Headings as subject access points. In an early work describing the guiding principles behind the application of these headings, Haykin (1951) argued that the librarian should select the heading he or she thinks a user would use to find the work, requiring the librarian to postulate an "average" reader. As Marshall (1977) pointed out in an influential critique of subject access, "this guiding principle…introduces bias as the average reader was defined by catalogers an 'American/Western European, Christian, white, heterosexual, and male" (1977, p. 6).

In 1971, Sanford Berman, a cataloger in Minnetonka, Minnesota, published a tract with examples of how this guiding principle resulted in problematic and objectionable terminology found in Library of Congress Subject Headings for describing people, particularly ethnic and racial groups. Knowlton (2005) traces the suggestions Berman made and the incredibly long time it required for subsequent changes to subject access points, noting that the heading "Jewish question" existed until 1984, while "Negroes" was used until 1975. "Homesexuality" and "Lesbianism" included cross-references to the term "Sexual perversion" in the mid-1970s; the exact date of the cross-references deletion is unknown. "Idiot asylums," meanwhile, persisted as a heading until 1993 (Knowlton, 2005). Terminology describing sexual and gender identity

remain problematic in controlled vocabulary well into the 21st century as these vocabularies persist in identifying people by terms that they would not choose to use themselves (Adler, 2009).

These examples illustrate the bias inherent in many subject terms as well as the problem with considering library catalog descriptions neutral constructs. They also illustrate the idea that subject access changes over time. Before the advent of easily updatable databases, the library literature included best practices for updating subject access to works in a library's collection (Nuckolls, 1994), allowing researchers to trace changes in cataloging practices and norms. As library catalogs now rely on computer software, once updates are made, the old term effectively ceases to exist unless the change is documented elsewhere.

As I make use of data from WorldCat, I therefore make use of headings as they were applied in 2017, not as they were applied when a Newbery Medal-winning title was first cataloged. As such, I refer to the heading of "African Americans," used in 2017 to describe *M.C. Higgins, the Great* (1975) rather than the originally applied and subsequently superseded heading "Negroes." I rely on data harvested from worldcat.org, which is free for users to access and presents the same information to different users, regardless of the user's "home" library. Results in a worldcat.org search default to showing aggregated information for all editions of a book on one record. I use this aggregated record display to harvest information on bibliographic information, descriptions and summaries, subject headings, and library holdings.

Reading assessment software databases: Accelerated Reader and MetaMetrics. The reading assessment software databases produced by Accelerated Reader (AR) and MetaMetrics, the company responsible for calculating and marketing Lexile measures, provide additional, albeit controversial, data sets about the Newbery Medal text set. Both databases exist to sell a

product. The information they contain is therefore inherently biased towards describing the text in such a way as to bolster the product for sale.

The text's Lexile measure is the only descriptive information that MetaMetric's Find a Book database provides beyond basic bibliographic information, limited to title, author, publisher, and occasionally a publisher's description, is the text's Lexile measure. Accelerated Reader's Bookfinder database provides more extensive descriptive bibliographic information, including word counts and the company's own controlled vocabulary subject access points, but the main purpose of the database is the Advantage/TASA Open Standard (ATOS) level, computed in part from the Lexile measure, and the number of Accelerated Reader (AR) points possible in related quizzes.

Both databases presuppose, however, that the user wishes to match a reader to a book based on a match between the reader's skill level and the book's text complexity. This presupposition can be inferred from the often obtuse subject headings observable in the database. Accelerated Reader's subject heading of "misc./other," for example, represents a largely useless and frequently used subject access point. Both databases prominently display the book's reading level, however, although the page does not display information on what that level means.

Librarians and educators frequently debate the value of leveling books using readability formulas, as both Accelerated Reader and Lexile measures do, and using readers' functioning levels to recommend books to child readers (see, for example, American Association of School Librarians, 2011; Krashen, 2001; Krashen, 2002; Shannon, Styers, Wilkerson, & Peery, 2015; Stenner, 2001). These critics argue that focusing on reading level alone, and encouraging librarians and teachers to match children to books based on a text's complexity and the child's reading level, misses a vital point of reading: the content of the book. Reading assessment

software also treats reading as a contest, rewarding fluency and reifying notions of what constitutes a good reader (Schneider, 2016).

Further, text complexity measures are not natural, neutral, or static constructs. A third-grade reading level, for example, is not a constant but is instead an agreed upon standard definition, one subject to changing norms and educational practices (Hiebert and Mesmer, 2013). Measures of text complexity also ignore issues of text comprehension (Valencia, Wixson, and Pearson, 2014). The information in the MetaMetrics and Accelerated Reader databases bear the marks of these limitations as evidenced in the limited and problematically formatted information contained in them.

It is also important to note that many formulas exist for calculating text complexity in addition to Lexile measure and ATOS level. As Hiebert and Mesmer (2013) explain, most readability formulas rely on two measurements: vocabulary and syntax, or, more specifically, word and sentence lengths, used as proxy variables. Some formulas from previous generations of literacy research and practice, such as Spache's (1953), could provide valuable insight to parts of the Newbery text set. Spache's formula specifically addressed primary grade reading materials, and the formula is therefore validated only for texts for grades three and below. Dale-Chall's formula (Dale & Chall, 1948) frequently supplements Spache's for texts leveled at grades four and above. Therefore, neither measure would apply to the entire sample.

Further, although there are studies that consider these and other earlier readability formulas in relationship to the Newbery Medal, none are recent enough to provide measurements for the entire date range of the corpus (Clements, Gillespie, and Swearingen, 1994; Guidry and Knight, 1976; Schafer, 1976; Schafer, 1986; Stevens, 2010). No existing databases offer computations of these readability formulas for Newbery Medal-winning texts, requiring the

research to sample the text(s) and calculate the measure, either manually or using software. This is not prohibitively difficult, although copyright limitations do present an additional barrier as full text or harvestable samples of text are not readily available for analysis. As the purpose of this study is to model methods for finding and analyzing existing secondary data sources describing children's literature, not compare readability metrics, I rely on the more modern, and previously calculated, Lexile measure and ATOS level.

The Lexile Framework for Reading, which provides a measure of text complexity for prose, powers, in part, the suggested grade levels for Accelerated Reader. Two components constitute the Lexile framework: the individual student reader's Lexile range, identified through standardized testing, and a book's Lexile measure, computed from sentence length and frequency of word use (MetaMetrics, 2008b; White and Clements, 2001). ATOS levels in turn use Lexile measures to suggest grade levels for each text in the Accelerated Reader Bookfinder database, although, as Hiebert and Mesmer (2013) note, these reading levels vary over time as educational practices and conventions change. The validity of these assessments lies outside the scope of this project.

Despite the serious and significant limitations of leveling books and using book levels to suggest reading material, however, both MetaMetrics and Accelerated Reader provide extensive coverage of Newbery Medal-winning titles. Although both have a bias against older titles, with many pre-1970 titles not included, Newbery winners are more commonly assigned and found in print (Kidd, 2007). The Accelerated Reader database includes all but one Newbery Medal-winning title, excluding only Paul Fleischman's *Joyful Noise: Poems for Two Voices*. Similarly, only three early prose Newbery winners (*Waterless Mountain*, the 1932 Newbery Medal-winning

title; *Dobry*, from 1935; and *Daniel Boone*, from 1940) and non-prose Medalists lack Lexile measures from MetaMetrics.

Breadth of coverage arguably constitutes the greatest strength for these reading assessment software databases, but they also represent the pervasive marketing of publishers to both teachers and librarians. The Lexile framework is currently the most commonly used metric in the United States for identifying text complexity (MetaMetrics, 2008a), and other educational resources, including Accelerated Reader, publishers, and libraries, use Lexile measures to quantify book difficulty and make decisions on which titles are appropriate for which readers. Like the Newbery Medal itself, both Lexile measures and Accelerated Reader are ubiquitous. This ubiquity results in a more complete data set for investigation. The subject and thematic data from Accelerated Reader also provides an important point of contrast to that harvested from WorldCat as both rely on different controlled vocabularies.

For this study, I harvested the Lexile measure for each Newbery Medal-winning text from MetaMetric's Find a Book database. From Accelerated Reader's Bookfinder database, I recorded ATOS level, subject tags, word count, and description. As with data from WorldCat, in this study, I rely on information from these databases as it existed in 2017.

***Goodreads*.** Data harvested from Goodreads provides an important source of information on popularity and how contemporary readers, not scholars, publishers, or professional reviewers, interact with and describe the corpus. In this study, I employ classification schema from Goodreads as a form of altmetric data. Although altmetrics more typically provide alternative measures for assessing the impact of an academic work outside of the more conventional h-index or citation frequency count (Priem, Taraborelli, Groth, and Neylon, 2010), library science literature suggests that Goodreads can provide researchers with a source of altmetric data on

publishing trends. Zuccala, Verleysen, Cornacchia, and Engels (2015) find that, within the field of history, "reader ratings and reviews on Goodreads serve as an indicator of [readership] beyond academia" (p. 332). Similarly, librarians also explore how Goodreads can provide alternative sources of information for collection development (Hoffert, 2010), readers' advisory (Braun, 2013; Evans, 2014; Trott and Naik, 2010; Rapp, 2011) and promotional work (Davies, 2014; Ganss, 2015), with the platform offering readers a chance to crowdsource services more traditionally found in libraries.

Nakamura (2013) asserts that "[s]cholars looking to study reading culture 'in the wild' will be rewarded by a close study of *Goodreads*" (p. 241, emphasis in original). Nevertheless, there remains scant critical attention paid to the site and the folksonomies its users create. In fact, no scholarly literature to date considers how readers use Goodreads to classify children's literature or how the site's other metrics might inform an understanding of text sets.

Although many social networking platforms exist that specifically offer readers a place to interact with one another while listing, cataloging, rating, and reviewing books, including LibraryThing and Booklikes, Goodreads remains the most popular, boasting over 50 million members, 1.5 billion books added to the site, and 50 million reviews (Goodreads, 2016). Given its popularity, the Goodreads database therefore offers a large dataset for review and includes information on all Newbery Medal winners. Further, the platform encourages users to read books and reflect on what they have read, not necessarily buy books, unlike reader reviews on Amazon. Goodreads is, however, a commercial entity, and its platform offers extensive commercial integration. An Amazon company since 2013 (Kaufman, 2013), Goodreads integrates not only with their parent company but also other publishing sources. Publisher's Weekly, for example,

uses statistics from Goodreads to quantify the buzz surrounding a new book ("Bestseller Stat Shot," 2015), and WorldCat displays reader reviews from Goodreads on a book's detail page.

As in WorldCat, users of Goodreads can access information about all editions of a book in the aggregate or individual editions. Results default to showing aggregated information for all editions of a book on one record. I use this aggregated record display to harvest information on ratings and tags.

Bowker's Books in Print. Books in Print is a trade resource designed to help libraries, booksellers, and publishers locate appropriate books for their collections. Bowker is the entity responsible for registering ISBN numbers in the United States, and their Books in Print database collocates information from a variety of sources, including publisher descriptions, Lexile Measures, professional book reviews, coverage in media sources, and BISAC (from the Book Industry Study Group) and Sears controlled vocabularies. These data elements largely duplicate elements recorded in other variables for this study, but the database also includes information on publication and print status not readily found in other sources. Most significantly, Books in Print provides information on different formats and the availability of those formats for current purchase. It also indexes titles that are out of print, providing the last date that a specific text was published in the United States. As such, it provides a current snapshot of the publishing marketplace and the Newbery's place in that marketplace.

Books in Print has existed in various formats and under various titles since 1868, and Bowker, the company currently responsible for its publication, was founded in 1872 (Bloomberg, 2017). Bowker's Books in Print is currently licensed by ProQuest, and in this study I make use of data from ProQuest's platform.

Data Collection

Guided by my review of existing literature on the Newbery Medal, I identified four areas to investigate via distant reading: descriptive information about Medal-winning titles, structural information, thematic information, and information about the relative popularity of each title. I reviewed each data source selected for inclusion in this study and identified data elements present in those sources that would enable analysis of these four areas. I then harvested the identified data elements and stored them as independent variables, described below, in a tab-delimited spreadsheet to enable later analysis using SAS (a statistical analysis software package) and Voyant (a web-based tool for analyzing and creating visualizations from text-based data sets). For an example of storage techniques for raw data, see Appenix A.

Variables Analyzed

In this section, I define, summarize, and categorize the variables I created from the data sources outlined above and briefly describe how each category will inform my analysis of the corpus (see Table 3.1). These variables represent the information I considered important for analysis; they do not represent a comprehensive list of data available from each source. I also note additional limitations of particular variables, outlining the conditions of knowledge (Moretti, 2005) that affect my interpretations.

Descriptive variables. These variables provide basic information about the Newbery Medal text set and describe the Medal-winning works.

Bibliographic information. This variable identifies and describes the work that won the award. Coding levels for this variable include title, author, publisher, year of publication, and year of award. The ALSC maintains a listing of all past winners and honor books, dating from the award's inception in 1922 to the present (see Appendix B). From this listing, I harvested

bibliographic information for each Medalist, with each variable recorded in its own tab-delimited cell.

Table 3.1. Overview of variables collected and analyzed. Where multiple data sources contribute similar variables, the data source is appended at the end of the variable name.

Descriptive Variables	Structural Variables	Thematic Variables	Variables Describing Popularity
*These variables provide **basic descriptive information** about the text set.*	*These variables provide information about the **structural characteristics** of the text set.*	*These variables provide information about the **thematic elements** in the text set.*	*These variables provide **measures of popularity** for each title in the text set.*
• Bibliographic Information • Publisher: Imprint and Parent Company • Gender of Author • Race of Author • Gender of Main Character(s) • Race of Main Character(s) • Illustrative Content • Illustrator • Type of Illustration • Length (Number of Pages) • Length (Word Count)	• Point of View • Literary Form (WorldCat) • Literary Form (Accelerated Reader tags) • Literary Form (Goodreads) • Text Complexity (Lexile Measure) • Text Complexity (ATOS level)	• Description (WorldCat) • Description (Accelerated Reader) • Description (Goodreads) • Subject (WorldCat) Subject (Accelerated Reader) • Subject (Goodreads) • Setting (Geographic, WorldCat) • Setting (Geographic, Accelerated Reader) • Setting (Time Period, WorldCat) • Setting (Time Period, Accelerated Reader) • Setting (Description)	• Print Status • Editions in Circulation (Books in Print) • Library Holdings • Editions Held by Libraries (WorldCat) • Number of Goodreads Ratings • Goodreads Rating

Publisher: imprint and parent company. This two-level variable describes the publisher imprint, as found in the bibliographic information harvested from the ALSC list of Medal winners, and the publishing company that owns the imprint. The imprint describes the press as it existed at the time the text was published and won the Newbery Medal. The parent publishing company records information as listed in reference resources about the imprint and reflects the many mergers that have occurred in the twenty-first century (Fialkoff, 2013). Some publishing companies closed rather than merge, thus not all companies recorded are currently in trade.

Gender of author. This variable identifies the gender of the author for each winning text, drawn from the third-person pronoun used by the author in self-descriptions or from biographical source material. To date, all winners have identified as male or female, so the variable is recorded as having two levels.

Race of author. This variable identifies the race of the author for each winning text, as described in author biographies. Category names for authors of color reflect the descriptions found in biographical source material. White authors were almost exclusively described as "American" in biographical source material, reflecting a default assumption of "American" equaling "white" (Morrison, 1989). To confirm the race of these authors, I consulted author photographs and autobiographical writings. In this study, I describe authors who do not identify as a person of color as "white" rather than "American."

Gender of main character(s). This variable identifies the gender of the main character(s) in each winning text and is coded as female, male, group, or not applicable. The description field for each text in WorldCat (Marc field 520) provides a summary of each title, and from this summary I harvested the personal pronouns used to describe the main character. As with authors, when personal pronouns occurred they were always male or female. Some summaries did not

identify an individual main character, describing instead a group of characters. In these instances, a value of "group" was recorded. Other summaries indicated that the title was a collection of poems or short stories, and no main character was identified. In these instances, a value of "not applicable" was recorded.

Race of main character(s). This variable identifies the race of the main character(s) in each winning text. I created this variable from information found in book summaries, as recorded in Marc field 520, and subject headings as applied in the WorldCat, Accelerated Reader, and Goodreads databases.

Rather than assuming a default racial category of "white" when data sources do not specify the race of characters, I recorded a value of "not specified." Racial and ethnic identity terms are harvested directly from descriptions and headings in the data sources used for this study.

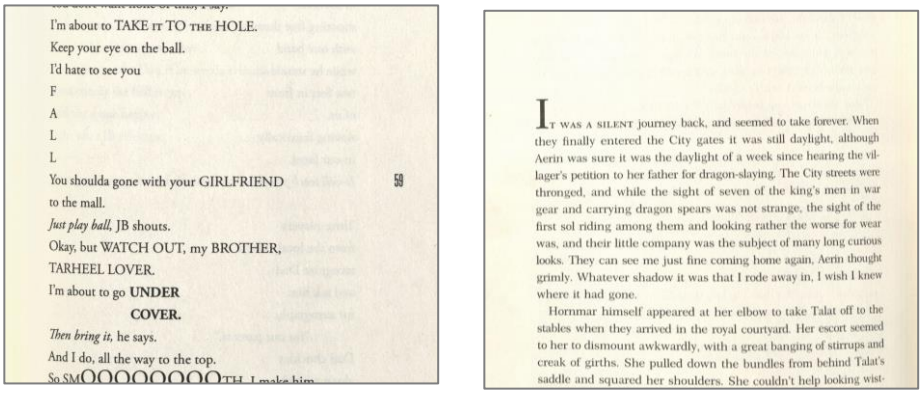
Illustrative content. This variable identifies the presence of illustrations within a Newbery Medal-winning text and is "yes" and "no." This information is drawn from the text's bibliographic description in the WorldCat database. This variable only describes the presence or absence of illustrations. It does not describe the relationship between illustration and text or qualify the role played by illustration in the text. This variable, therefore, only offers analytical information suggesting the need, or lack thereof, for future research related to illustrative content in the corpus.

Illustrator. If a text contains illustrations beyond cover art, this variable records the name of the illustrator responsible for the accompanying images. If a text does not contain illustrations beyond cover at, a value of n/a is recorded. This information comes from the text's bibliographic description in the WorldCat database.

Type of illustration. If a text contains illustrations, this variable records the type of illustration found in the text using free text. Examples include "black and white lithographic prints" and "color reproduction of water color paintings." This information comes from bibliographic information found in the WorldCat database and publisher's descriptions. If a text does not contain illustrations, a null value is recorded.

Length (number of pages). This variable records the number of pages in the first edition of each winning text. This information comes from the text's bibliographic description in the WorldCat database. As such, it represents the numbered pages within each book, including separately paginated front and end matter but not un-numbered end pages.

The number of pages in a book, however, is not always an accurate measure for determining the length of a book. Kwame Alexander's *The Crossover* (the 2015 Newbery Medalist) and Robin McKinley's *The Hero and the Crown* (the 1985 Medalist) have very similar page counts: the former is 237 pages, the latter 246. Visually comparing the pages in these books suggests that approximately 240 pages are not always created equal, and the word counts for these titles underscore this impression. *The Crossover* contains only 16,888 words, while *The Hero and the Crown* fits 87,370 words on roughly the same number of pages (see Figure 3.1).


I'm about to TAKE IT TO THE HOLE.
Keep your eye on the ball.
I'd hate to see you
F
A
L
L
You shoulda gone with your GIRLFRIEND
to the mall.
Just play ball, JB shouts.
Okay, but WATCH OUT, my BROTHER,
TARHEEL LOVER.
I'm about to go **UNDER**
COVER.
Then bring it, he says.
And I do, all the way to the top.
59

IT WAS A SILENT journey back, and seemed to take forever. When they finally entered the City gates it was still daylight, although Aerin was sure it was the daylight of a week since hearing the villager's petition to her father for dragon-slaying. The City streets were thronged, and while the sight of seven of the king's men in war gear and carrying dragon spears was not strange, the sight of the first sol riding among them and looking rather the worse for wear was, and their little company was the subject of many long curious looks. They can see me just fine coming home again, Aerin thought grimly. Whatever shadow it was that I rode away in, I wish I knew where it had gone.

Hornmar himself appeared at her elbow to take Talat off to the stables when they arrived in the royal courtyard. Her escort seemed to her to dismount awkwardly, with a great banging of stirrups and creak of girths. She pulled down the bundles from behind Talat's saddle and squared her shoulders. She couldn't help looking wist-

Figure 3.1. Page views of *The Crossover* (left) and *The Hero and the Crown* (right)

Length (word count). Given the apparent variance in text density on the sample Medal-winning texts of *The Crossover* and *The Hero and the Crown*, I also calculate length as word count. This variable records the number of words in each winning text. This information comes from the book's bibliographic description in the Accelerated Reader Bookfinder Database.

Structural variables. These variables provide information about the structural characteristics of the Newbery Medal text set and enable an analysis of the formal characteristics of the text set as a discrete genre.

Point of view. This four-level variable refers to the narrator's positionality within each text. Narrative positions coded include first person, second person, third person, and mixed. No secondary data sets exist for any portion of the Newbery Medal corpus describing point of view, so this variable constitutes the only data set that I created from primary rather than secondary sources. To generate this variable, I used the "preview" feature on Amazon and Google Books to randomly sample each text. From that sample, I recorded the narrative position(s) observed.

Literary form (WorldCat genre headings). This variable records simple generic information that suggests the narrative form for each text, such as biography, fiction, and poetry, as recorded in controlled vocabulary terms. This variable was created using Library of Congress Genre Headings (Marc field 655) and genre information from general Library of Congress subject headings (Marc field 650) as assigned to each winning text in the WorldCat database. The application of genre headings to titles in WorldCat is, at best, variable. All Newbery Medal-winning titles except for *Thimble Summer* (1939) have at least one genre heading. Many titles, however, have many genre headings, and the most recorded for any title is eight. The terms range in specificity as well as quantity. The non-descriptive heading "juvenile fiction" appears

frequently as a genre heading within the corpus, as do much more specific terms, such as “novels in verse,” “bildungsroman,” “nonsense verse,” “sea stories,” and “folklore.”

As I have combined genre information from general subject headings (field 650) with genre specific headings (field 655), this variable does not map precisely to a list of headings applied in either the 650 or 655 field. Instead, this variable represents a composite list of all genre headings observed in both fields. Duplicate terms used in both field 655 and field 650 to describe the same text were recorded only once. “Basketball stories,” for example, occurred in both field 650 and 655 as descriptors for *The Crossover*.

Literary form (AR tags). This variable records generic information as recorded in Accelerated Reader tags. Unlike WorldCat, Acclerated Reader does not isolate genre tags in a separate field from thematic headings. Therefore, this variable represents genre headings I identified and separated for analysis from thematic headings.

Literary form (Goodreads). This variable records generic information as applied by users in the Goodreads database. User tags applied in Goodreads include a wide array of genre information, although this information is not recorded in a different tag than thematic subject headings as it is in WorldCat. I isolated genre-related tags from all harvested Goodreads user tags to create a discrete set of genre tags.

Text complexity (Lexile measure). This variable provides one measure of how complex each text is, utilizing the Lexile Framework for Reading’s standardized measurements. This information is harvested from MetaMetric’s Find a Book database. The Lexile measure represents MetaMetric’s determination of how complex a text is, based on an algorithm that examines length of sentences and word frequency. Standard punctuation is a requirement of the Lexile measure, so non-prose works are excluded. Lexile measures begin at 5L and increase in

five-digit increments to 2000L. In general, a higher Lexile measure corresponds to a more complex text.

Text complexity (ATOS level). This variable provides an additional measure of how complex each text is, utilizing the ATOS level. This information is harvested from the Accelerated Reader Bookfinder database. The ATOS level represents Renaissance Learning, Inc.'s, determination of how complex a text is, based on their readability formula that considers sentence length, word length, and the difficulty of individual words (Renaissance Learning, Inc., 2017).

Thematic variables. These variables provide information about the thematic elements found in the Newbery Medal text set. These variables enable a consideration of the themes, motifs, subjects, and contents of the corpus without resorting to microscopic reading strategies. Further, they enable a consideration of these elements through the lens of various reading audiences, although not all data sources include the same types of thematic information. The creation of independent variables from multiple sources describing similar types of information allows for eventual comparison of data sources (discussed further below, in "Analysis of Data"). Comparison of various topic models derived from different data sources, in turn, enables and a discussion of how the differences and similarities in the models speak to the types and qualities of information available that enable the study of contemporary American children's literature.

Description (WorldCat). This variable records the book's summary as found in Marc field 520 in the WorldCat database.

Description (Accelerated Reader). This variable records the book's summary as found in the book's description in the Accelerated Reader Bookfinder database.

Description (Goodreads). This variable records the book's summary as found in the book's description in the Goodreads database.

Subject (WorldCat). This variable records all of the Library of Congress Subject Headings assigned to each winning text in the WorldCat database in Marc field 650 and 651. Marc field 650 records thematic subject headings (e.g., "Squirrels—Juvenile Fiction), while Marc field 651 records geographic subject headings (e.g., New Jersey – Juvenile Fiction).

All Newbery Medal winners have at least two Library of Congress subject heading (Marc field 650 or 651) applied to them in the WorldCat database. One title, *Rabbit Hill* (1945), uses an astonishing 33 subject headings to describe the contents of the novel. On average, 10.42 subject headings are applied to each title. As this study makes use of the master record for each title from WorldCat, extensive duplication in the subject headings for each title is present due to the use of genre subheadings. The following subject headings all describe characters in *The Girl Who Drank the Moon* (2017), for example: witches – juvenile fiction; witches – fiction; witches. The actual subject term is the same, but different libraries have variously described the work as piece of juvenile fiction about witches, a piece of fiction about witches, and an unspecified type of text about witches. *The Hero and the Crown* (1985), meanwhile, uses only one heading to describe the main character: tomboys – juvenile fiction.

Both approaches are technically correct, but weighting "witches" more heavily than "tomboy" due to different libraries' approaches to subject analysis biases the analysis of characterization in this microscopic example. In order to control for this variation, I normalized and simplified subject headings and consider only the root terms. As another section of this study considers genre, I stripped the generic information from subject headings that is found in subfield v of Marc fields 650 and 651. Similarly, another section of this study considers

geographic setting, so I also stripped geographic information and deleted those terms from this analysis. I deleted the exact duplicate subject terms that resulted from this cleanup. *The Girl Who Drank the Moon*, for example, has only one heading to describe the characters: "witch." Genre headings recorded in Marc field 650 were also deleted from this analysis as they were considered in the section on genre. I also deleted terms that are not authorized headings for Marc fields 650 and 651, such as "reading group guide" and "trans-world travel." In all instances where I deleted a non-authorized heading, an authorized heading also described the same work. This cleaned, de-duplicated list results in an average of 4.36 terms applied to each title, with a minimum observed value of zero and a maximum observed value of 15.

Subject (Accelerated Reader). This variable records the subject tags assigned to each text in the Accelerated Reader Bookfinder database.

Subject (Goodreads). This variable records the thematic tags from the top ten user-assigned tags for each text in the Goodreads.com database. By default, Goodreads displays up to ten top tags on a book's main page, providing users with a sense of how other readers have categorized a text. Although Goodreads labels all user tags as "genre," regardless of content, a sizable number of thematic tags also exist in the database. To create this variable, I separated thematic headings from genre tags.

Setting (geographic, WorldCat). This variable records the geographic setting for each title, as suggested by geographic information in Library of Congress Subject Headings (Marc field 651 and Marc field 650, subfield z) assigned to each winning text in the WorldCat database. This variable is extracted from the data harvested for the variable "Subject, library perspective." It includes generic places that describe communities and environments, such as "museums," "farms," "homelessness," and "islands." It excludes terms that describe people, such as "Native

Americans and "U.S. Presidents," as well as terms that describe historical events, such as "World War II," that imply place by do not provide enough specificity for analysis.

Setting (geographic, Accelerated Reader). This variable records the geographic setting for each title, as suggested by geographic information in Accelerated Reader tags. This variable is extracted from the data harvested for the variable "Subject, Accelerated Reader perspective." As with Library of Congress Subject Headings, it includes generic places that describe communities and environments, but excludes terms that describe people and historical events without specific geographic information provided.

Setting (time period, WorldCat). This variable records the time period in which each text is set, as suggested by information in Library of Congress Subject Headings (Marc field 650, subfield y) assigned to each winning text in the WorldCat database. This variable is extracted from the data harvested for the variable "Subject, library perspective."

Setting (time period, Accelerated Reader). This variable records the time period in which each text is set, as suggested by information in Accelerated Reader tags. This variable is extracted from the data harvested for the variable "Subject, Accelerated Reader perspective."

Setting (description). As Goodreads tags do not include a significant amount of information about setting, I employ instead collated geographic and setting information harvested from WorldCat, Accelerated Reader, and Goodreads' book descriptions as a method for checking for omissions in geographic and temporal coverage in WorldCat and Accelerated Reader's controlled vocabularies. This variable records information on setting extracted from book summaries found in descriptions from WorldCat, Accelerated Reader, and Goodreads. As descriptions in all three sources are variable in quality and information density, I used all three to

provide the widest range of coverage possible. The source with the most complete information was preferred in creating this variable.

Variables describing popularity. A common assertion critics make about the Newbery is that the titles that win the Medal, by and large, have staying power. They sell well, both to individuals and to libraries; they are frequently assigned in school or as supplemental reading; and they are read. As with thematic analysis, measures of a literary text's popularity prove tricky. Long-term sales data would be the ideal source for measuring popularity, but raw sales data is proprietary and infrequently released (Michel, 2016). Nielsen's BookScan, the industry standard for tracking sales data outside of publishing firms, did not appear on the market until 2001 (Magner, 2003), is not retrospective, and only reports approximately 75% of any genre's print-only sales (Michel, 2016). Thus, data from BookScan is restricted to the recent past and is incomplete. Sales data also drives best-seller lists, but these lists are generated from a single week's sales figures, skewing long-term analyses as slow but steady sellers could, over time, out sell a one-week wonder (Truitt, 1998; Miller, 2000). They are also inconsistent in their inclusion and treatment of children's literature. When the New York Times bestseller list, for example, began in 1931, and it included children's literature as individual titles earned a spot in competition with general adult fiction. It wasn't until 2000, when *Harry Potter and the Goblet of Fire*'s imminent release threatened to claim a fourth spot on the list, that the *Times* created a separate children's list (Smith, 2000).

Despite these challenges, there are nevertheless alternative sources that speak to long-term popularity of the corpus. The variables in this section provide measures of popularity for each title in the Newbery Medal text set and enable an analysis of the Newbery's enduring popularity with libraries, schools, and readers that extends beyond anecdotal evidence.

Print status. This variable records if each text is currently available in print or not. This variable records print status as listed in Bowker's Books in Print database. Books in Print does not differentiate between commercial publishers and print on demand models in their indication of print status. Therefore, this count includes both traditionally published and print on demand publications.

Editions in circulation. This variable records the number of different editions of each text currently listed with a status of "in print" in Bowker's Books in Print database as of November 2017. This count includes different physical media: print, ebook, and audio book editions, as well as omnibus editions and multipacks of books including the Newbery Medal winner. It also includes foreign language and braille editions published and available for purchase in the United States. This count excludes teacher's guides, student workbooks, curriculum guides, and vocabulary lists.

Library holdings. This variable records the number of libraries holding any edition of each text in the WorldCat database as of July 2017. This variable does not record the number of duplicative copies individual libraries might hold of the same title.

Editions held by libraries. This variable records the number of different editions of each text in the WorldCat database as of July 2017. This count includes different physical media: print book, ebook, and audio book all display as holding information on the master title record and are thus counted as an edition in circulation.

Number of Goodreads ratings. This variable records the number of Goodreads users who have rated any edition in any format, including print, audio, and ebooks, of each title on Goodreads.com as of July 2017.

Goodreads rating. This variable records the average rating assigned to each title on the Goodreads platform. Goodreads ratings take the form of stars, with possible values ranging from one to five. One-star equals "did not like it," while two stars indicate "it was ok," three "like it," four "really liked it," and five "it was amazing."

Data Analysis

The data I analyze in this study speaks to many different components of the Newbery Medal as a discrete subgenre of children's literature: its descriptive characteristics, its structural characteristics, its thematic characteristics, and its popularity. In order to explore this data, I first rely on statistical avenues for classifying the descriptive and structural characteristics of Newbery Medal corpus and then use those statistical results to inform an analysis of the corpus in conversation with existing critical conceptions of the Award.

Describing the corpus bibliographically and structurally. I compiled simple descriptive statistics in SAS on all descriptive and structural variables to illustrate the formal characteristics of the corpus itself, and I analyze these variables in two ways. In one view, I focused on the variable holistically, observing the totality of the way that, for example, gender representation in main characters occurs within the corpus. In the second view, I focused on the variable in ten-year increments, observing how the variable fluctuated or remained static over time. This shift provided balanced data sets for all but one decade (2011-2017) rather than two (1922-1929 and 2010-2017). Chi-square tests of independence test the significance of observed frequencies across descriptive and structural variables. As I observed a nonparametric distribution across all variables, I restricted analysis to descriptive statistics.

Thematic analysis: varying perspectives, varying results. Subject headings provide one mechanism for coding genre and thematic information. Publisher synopses and database

book descriptions provide a different lens through which to analyze the same questions. Regardless of source, the thematic variables collected and analyzed in this study are subjective measures that prove challenging to quantify. Moving beyond the formal classification of "children's literature – fiction" and "children's literature – non-fiction" is quite difficult due to the way that many publishers and libraries classify books for children. Review sources tend to identify works for children simply as either fiction or nonfiction, despite the rich range of genres encompassed by these two terms. Accelerated Reader, for example, classifies books by reading level, offering subject and genre access points as a secondary measure. The database does not allow users to browse or search by genre.

Similarly, library catalogs, including WorldCat and the Library of Congress, prove inconsistent in their classification of thematic genre, particularly over the range of time represented in this study. Although subject headings have been applied to works for children with varying degrees of success since the 1950s (Rue and La Plante, 1952; Vizine-Goetz, 2008), the Association for Library Collections and Technical Services, a division of the American Library Association, did not even approve the application of Library of Congress subject headings, the schema which is most commonly used in contemporary automated library catalogs (Fountain, 1996), to individual works of fiction until 1990 (ALCTS, 1990).

As a result, newer works tend to have more granular generic classification, while older works rely more heavily upon generic headings such as "juvenile fiction" or "biography." The Accelerated Reader Bookfinder database, commonly used in school reading programs, also includes controlled vocabulary describing the titles it indexes in the form topical keywords, although the topics provided are sometimes more helpful than others.

A few microscopic examples illustrate the problems inherent in both sources. The Library of Congress subject headings for *Flora & Ulysses: The Illuminated Adventures* (DiCamillo, 2013) consist of "fantasy fiction," "humorous stories," "fiction," "adventure stories," and "juvenile fiction." The Accelerated Reader tags, meanwhile, include "adventure-adventurers" and "animals-squirrels." These tags make no mention of the fact that the "animal – squirrel" in question writes poetry and possesses super strength after a life-altering trip through a vacuum cleaner. The Library of Congress subject headings and topics for *When You Reach Me* (Stead, 2009) suggest the difficulties inherent in classifying literature. Libraries classify this title as "fiction," "history," and "juvenile fiction," while Accelerated Reader provides "family life-TV viewing," "historical fiction-historical fiction (all)," and "science fiction-time travel." "Family Life-TV viewing" does not appear to be a helpful keyword for either educators or readers, and *When You Reach Me* is simultaneously an example of both historical fiction (it takes place in 1978 New York City) and science fiction (time travel plays a crucial role in the plot arc). The Accelerated Reader topics provide no guidance as to which element predominates, and library subject headings ignore the science fictional elements completely. *The Witch of Blackbird Pond* (Speare, 1959), meanwhile, has the subject headings of "fiction," history," "juvenile fiction," and "paranormal fiction" in WorldCat, while Accelerated Reader topics include "history-American," "horror/thriller–witches/warlocks." Neither catalog tags the novel as "historical fiction," and Accelerated Reader misleadingly identifies the work as horror.

These challenges are problematic, particularly on the level of the independent exemplar. Taken holistically, they undeniably introduce noise and uncertainty into the thematic data set. For this reason, I sought out and collected data from a variety of similar sources that enable cross-variable comparisons and offer a system of checks and balances to the analysis.

Comparison of topic models derived from different data sources also enables and a discussion of how the differences and similarities in the models speak to the types and qualities of information available that facilitate the study of contemporary American children's literature, and it also provides a mechanism for checking for omissions and oversights in the corpus.

Despite the challenges associated with consistently applying subject and genre classifications to children's literature, scholars frequently make assertions about the sub-genres preferred by Newbery committees over time (Kidd, 2007), noting a preponderance of historical fiction, exotic settings, and plucky children. In this portion of my study, I explicate methods that model the thematic markers of the Newbery Medal corpus holistically, including specific generic form (e.g., historical fiction, science fiction, poetry, non-fiction, etc.) and content markers, such as historical era represented, location, gender of characters, family structures, and the general activities associated with childhood in the corpus. To facilitate this analysis, I rely on Voyant Tools (https://voyant-tools.org/), an open access web platform that enables analysis and data visualizations of text corpora.

In this study, I treated each independent thematic variable as a separate corpus, uploading each to Voyant and calculating frequencies, collocations, and correlations within that individual variable's corpus. I relied on Voyant's corpus summary tool to gain an overview of each variable's corpus, investigating the number of unique words in the corpus as well as the most frequently used words. The frequencies grid allowed for further insights into most frequently used words, and I employed the collocate frequencies tool to explore terms that collocate with those most frequently used words. Word trees, meanwhile, provided visualizations of terms collocated with frequently used words in each corpus. After analyzing each variable independently, I then compared findings from each variable's corpus with those from other

variables to identify potential lacunae in the data sets. When I identified gaps in the data sets, I interrogated them against data from other sources in order to see if the gap signified an important element or change in the corpus or pointed to a limitation of the data set itself.

Coding subject headings and tags. Library of Congress Subject Headings and Accelerated Reader tags both represent examples of controlled vocabularies, and different rules govern both for application and use. Goodreads, meanwhile, relies on its users to generate their own tags to describe a work. Given this, it is unsurprising that all three platforms use different terms to describe similar concepts, generic forms, or themes. Further, many of the vocabulary terms applied to the corpus are incredibly specific. The Library of Congress Subject Heading "Arabian horse – fiction," for example, applies to only one text in the corpus, as do the headings "Elephants – fiction," "Squirrels – fiction," along with a veritable host of headings for other animals. Therefore, in addition to analyzing the actual vocabularies applied in both library catalogs and the Accelerated Reader database, I used Voyant Tools to locate infrequently used terms in the corpus and look for patterns in those terms that might benefit from the creation of broader, inductive coding categories. These codes facilitated analysis of larger trends across the corpus after reviewing the entire data set that controlled vocabularies miss.

Assessing popularity in the corpus. I used portions of the data set harvested from WorldCat, Goodreads, and Books in Print to measure the Newbery's popularity as a corpus and to investigate if the entire corpus is equally popular. Is the Newbery actually popular as a genre, or are only a few well-known individual Medal-winning titles popular? Using data from Books in Print, I compiled descriptive statistics to illustrate the number of titles still in print in 2017 as well as the number of editions still available. I contrasted these statistics with those from editions and holdings WorldCat, which provides information on editions not necessarily currently in

print. I also analyzed the number of libraries that hold a copy of any edition of each Newbery Medal-winning title. Using data from Goodreads, I compiled descriptive statistics on the number of readers on the site who have rated each Newbery Medal-winning title as well as the average rating assigned to each title. I used these measures to assess the popularity of the Newbery corpus in its entirety, rather than the individual titles that critics typically point to when discussing who the Award is commonly read or assigned.

In addition to this holistic analysis, I also considered how measures of popularity in the corpus change with time. I compiled descriptive statistics on ten-year segments of the corpus for each measure and charted variations in popularity for different points in time. Again, rather than using traditional decade markers (e.g., 1920-1930), I instead marked decades by ten-year chunks in the Medal's history. This shift provided balanced data sets for all but one decade (2011-2017) rather than two (1922-1929 and 2010-2017).

Validity and Generalizability

I studied the entire population of Newbery Medal-winning titles as a purposive sample drawn from the larger field of children's literature. The analyses conducted reveal a great deal about the characteristics of Newbery Medal text set as a corpus and as a discrete genre. My analyses do not, and are not intended to, generalize to the larger population of children's literature texts. Instead, by generating methods for understanding a subset of children's literature in the composite, I seek to explore how shifting the reader's view from the microscopic, the mode most frequently used in studying the Newbery, to the macroscopic leads to potential changes in understanding of the Newbery as genre.

The Researcher's Role in Distant Reading

Literary analysis is notorious for obfuscating research design and methodology. Unless quantitative methods provide the primary focus of the study (e.g., Goldstone & Underwood, 2014), discussions of sampling strategy, development of thematic coding vocabulary, and the like tend to be relegated to the discursive appendix (e.g., Algee-Hewitt and McGurl, 2015) or footnote (e.g., Marshall, 2012). Similarly, in the social science paradigms from which digital humanities in general and distant reading in particular largely draw inspiration, quantitative analyses are not known for researcher reflexivity; it is, instead, a hallmark of qualitative research. In this dissertation, however, I argue against the objectivist stance of distant reading, embracing instead the idea that distant reading is a tool to help answer the social, cultural, and formal questions that literary criticism has been asking, in different registers, for decades. It is a tool, moreover, wielded by a researcher, and the decisions made by the researcher affect that construction of the data set. The data set for this dissertation is not a neutral construct: I created it, I analyzed it, and I employed a range of analyses to inform my ultimate understanding of the Newbery Medal text set as subgenre of children's literature.

Chapter 4: Describing the Newbery

In this chapter, I holistically explore bibliographic and structural data sets that describe the Newbery Medal corpus. The descriptive information illustrates the formal characteristics of the basic types of books, authors, voices, and perspectives that the Medal privileges through prizing. The structural information, meanwhile, provides insights into the formal characteristics of the types of texts that the Medal privileges. The distant perspective I adopt to analyze this data allows for a consideration of the corpus holistically as well as a consideration of how these various formal elements have themselves changed with the passage of time.

Describing the Newbery: Authors and Perspectives

The bibliographic information harvested for this study indicates that a large number of different authors, represented by a wide range of publishing houses, have won the award. Only four authors have been awarded the Medal twice: Elizabeth George Speare (1959 and 1962), E. L. Konigsburg (1968 and 1997), Katherine Paterson (1978 and 1981), and Lois Lowry (1990 and 1994). The remaining 88 Medals went to different authors. Closer investigation, however, reveals that these authors look quite a lot alike. Most, particularly in more recent years, are female. Across the board, these authors are predominantly white.

Gender of authors. Critics have long argued that the Newbery Medal favors women's voices and girls' stories. The earliest such criticism comes from Howard Pease's 1939 invective against the overly feminized world of children's literature, viewed in part as an over representation of female authors and female characters in Newbery Medal-winning texts (Pease, 1939). Pease's rallying cry sparked a debate over gender representation in the Newbery Medal

that lasted through 1945 (Jenkins, 1996). Decades later, the Newbery remains known as one of the only literary prizes where women continue to dominate (Clark, 2003; Kidd, 2007).

In this section, I test these assertions against data from the entire corpus, both from the vantage point of the present day as well as the perspective of the entire corpus when Pease introduced the idea of a gender imbalance in the Medal. Using the personal pronouns found in author biographies located on Newbery winning titles where possible, and from biographies in the *Dictionary of Literary Biography* when the books themselves did not include a biographical statement, I coded the inferred gender of each Newbery Medalist, 1922-2017: 61 winners are female; only 34 male (see figure 4.1).

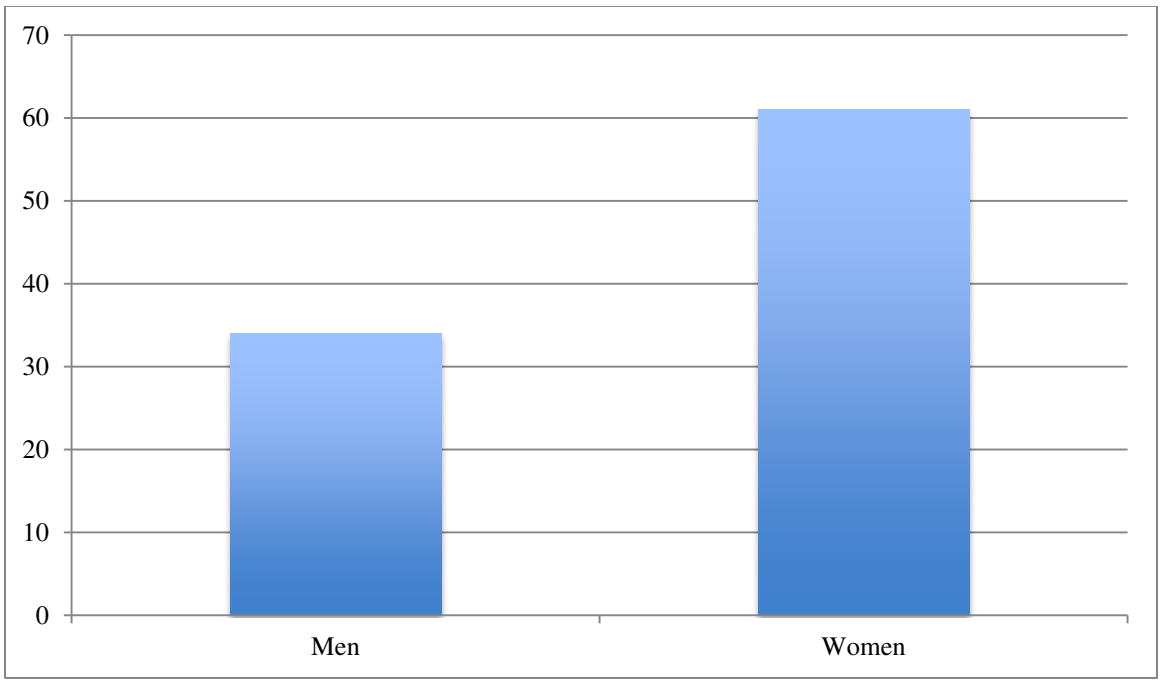


Figure 4.1: Gender of Newbery Medal-winning authors, 1922-2017.

In 2017, the of gender discrepancies of Newbery Medal-winning authors is significant (χ^2 (1, N=96) = 7.0417, p>.008). After nearly a century of Medalists, significantly more women

have won the Newbery Medal than men. What, though, of specific points in time? Pease (1939) introduced the idea of women dominating the Medal in 1939, and this conversation prevailed, particularly among librarian critics of the award, until 1945 (Jenkins, 1996). Figure 4.2 illustrates the frequency distribution of the gender of Medal-winning authors by decade, with decade defined as ten-year increments of the Medal, rather than calendar-based decades. Of particular interest are two ten-year spans: 1922-1931 and 1932-1941, the two decades of data upon which Pease based his claims. The first ten years of the Newbery Medal significantly favored male authors, at a ratio of eight male authors to 2 female (χ^2 (1, N=10) = 3.6000, p>.05). The second ten years witnessed a reverse of this trend, with 8 female authors to 2 male authors in the years 1932-1941. In the entire corpus in 1945, which marked the end of the first debate over gender representation among winners, the observed distribution matches the expected frequency precisely (see figure 4.3), with precisely the same number of male and female medalists. Thus, Pease's 1939 observation and subsequent arguments that women authors dominated the Medal came at a point in time when women's voices were entering the Medal's corpus and ending the dominance of men's voices. This argument came into existence as a reaction against the inclusion of women's voices, and it was accepted as true even when it was not.

The two ten-year spans of 1962-1971 and 1972-1981 witness the same significant inverse distributions, with 8 women authors to 2 male authors observed in 1962-1971 and 8 men to 2 female authors observed in 1972-1981 (χ^2 (1, N=10) = 3.6000, p>.05). Taken in total, these segments, comprised of 40 years, illustrate that considerable significant fluctuation in the gender composition of the corpus of Medal-winning authors is observable over time. To claim that the Newbery Medal favors women's voices over men's is to oversimplify the question, particularly in light of the critical conversations that surrounded the Medal at different points in time.

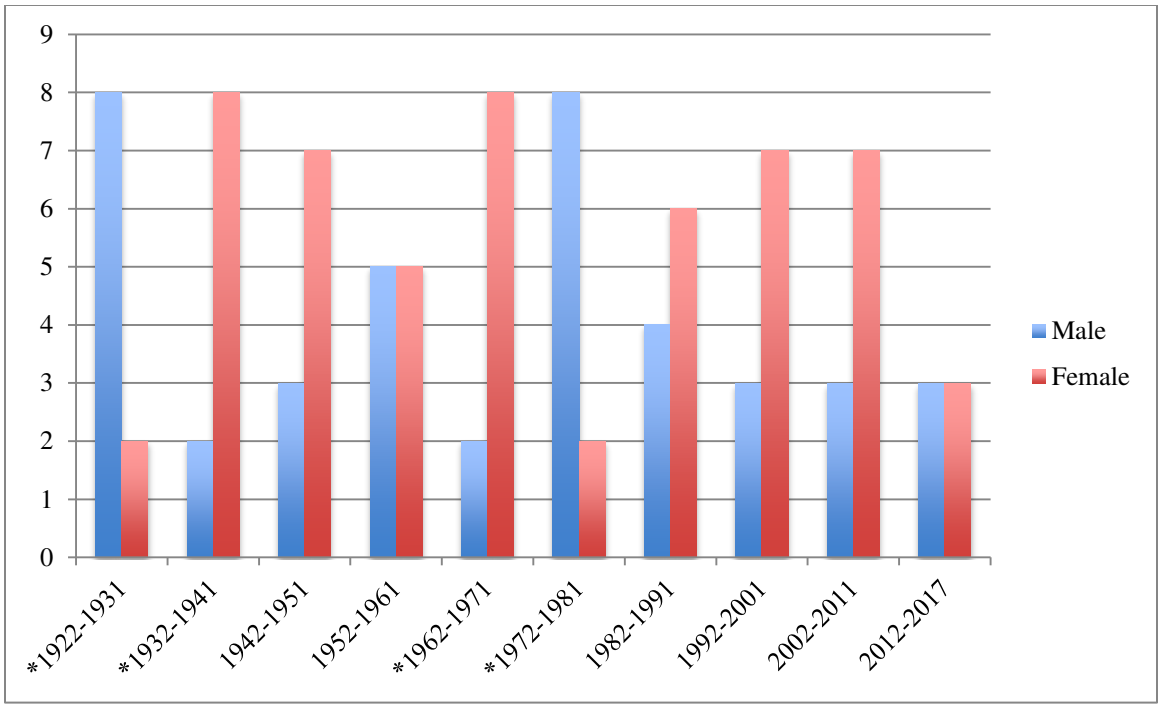


Figure 4.2: Gender of Newbery Medal Winners by decade. Distributions significant at the *p*>.05 level are designated with an asterisk.

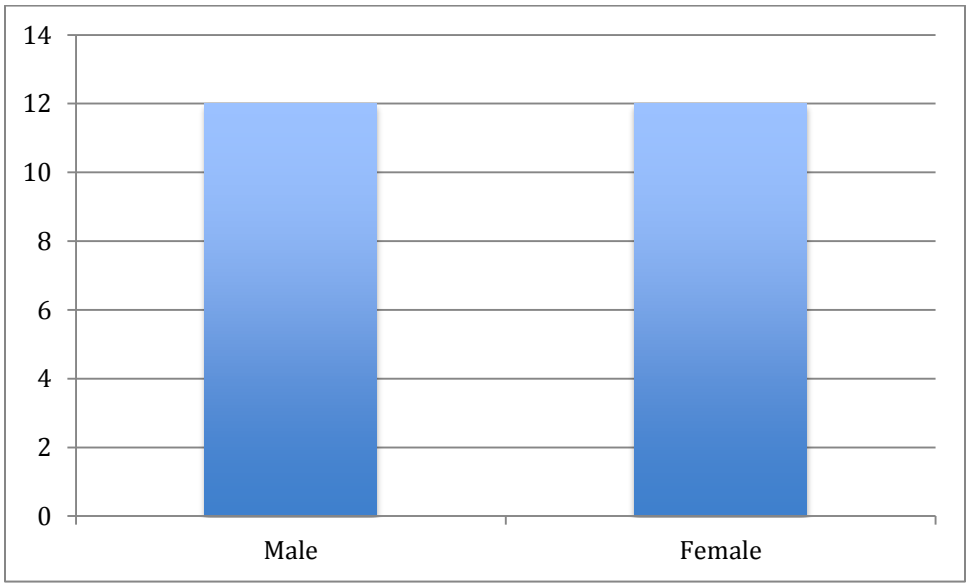


Figure 4.3: Gender of Newbery Medal winners, 1922-1945.

Depictions of gender in the corpus. Similarly, analysis of the gender of main characters in the corpus illustrates a tendency on the part of critics to overemphasize the role played by female characters, which again began with Pease (1939) and continued through subsequent generations (see, for example, Nelson, 2011; Powell, 1998). Figure 4.4 illustrates the frequency distribution of main characters' signified gender. The significance of this distribution varies based on assumptions made about the corpus. If only the 88 Newbery Medal-winning books with either a male or female character are considered, the differences in observed frequency are not significant (χ^2 (1, N=88) = 3.482, p>.06), indicating that the corpus does not favor stories about either boys or girls. When the full corpus, including winners with groups of main characters as well as no identifiable main character, is considered, the observed frequency is significant (χ^2 (3, N=96) = 57.25, p>.0001), when the expected distribution is even across all categories. If the expected distribution is adjusted to account for the probability of fewer texts featuring groups or including an unidentifiable main character, the observed frequency is not significant (χ^2 (3, N=96) = 6.734, p>.08). This series of calculations suggests that the Newbery Medal is more likely to privilege a text about a male *or* a female character (that is, an individual) than it is to prize a text about a group of characters or a text without an identifiable main character. It does not support the supposition that the Newbery privileges stories about girls over stories about boys.

Race of authors. The Newbery Medal is well known for its contributions to the all-white world (Larrick, 1965) of children's literature (see, for example, e.g., Clark, 2007; Madsen and Robbins, 1981; Miller, 1998; Wilkins, 2009). Existing scholarship focuses on representation of a single race within the corpus, such as African American authors and characters (Wilkins, 2009) or Native American imagery (Madsen and Robbins, 1981). No studies consider race holistically,

and this holistic analysis illustrates the sheer magnitude of the whitewashing that occurs in the types of texts that the Medal privileges and the lack of diversity in the corpus.

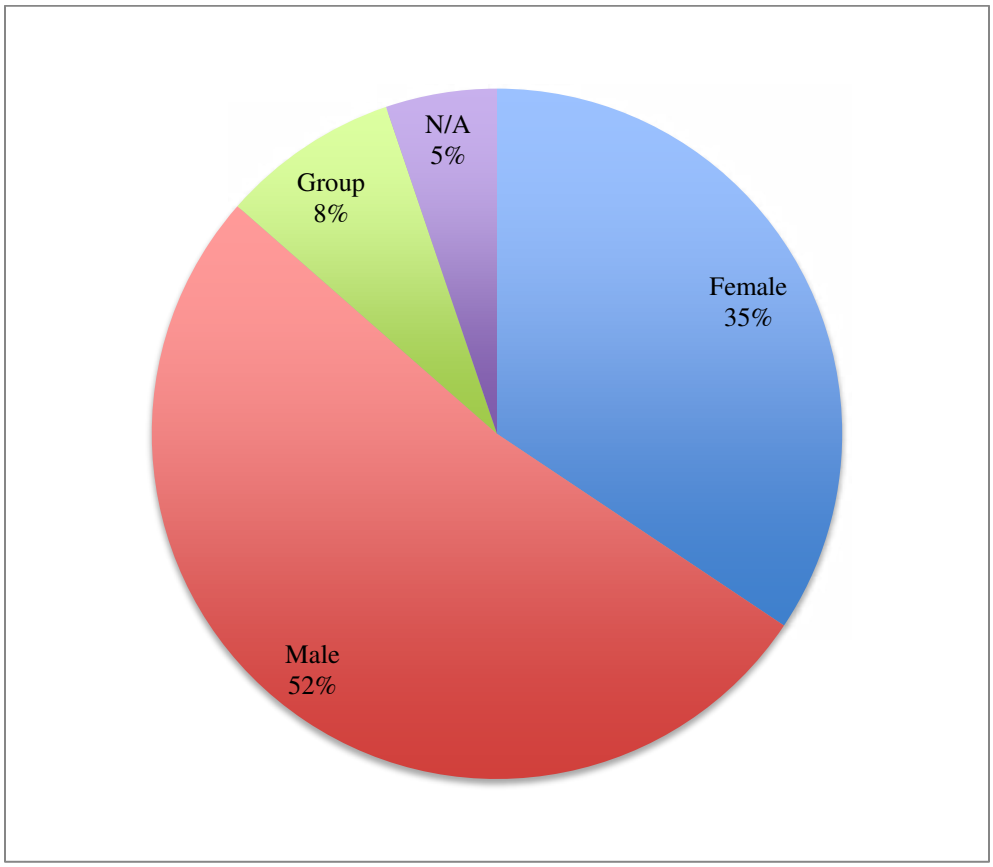


Figure 4.4: Gender of main characters in Newbery Medal corpus.

As figure 4.5 illustrates, as of 2017, the Newbery Medal has been awarded to 88 white authors (92%) and only eight authors of color (8 %), a distribution that is strongly significant (χ^2 (1, N=96) = 66.667, p>.0001). Four of the eight authors of color in the corpus identify as African American, while of the remaining four, one identifies as Hispanic, one as Indian, one as Japanese, and one as Korean (see figure 4.6). Four of the eight authors of color identify as male, and the remaining four identify as female. Five of these eight authors, however, were awarded

the Newbery in 2000 or later, meaning that only three Medal-winning authors in the first 79 years of the award identified as people of color.

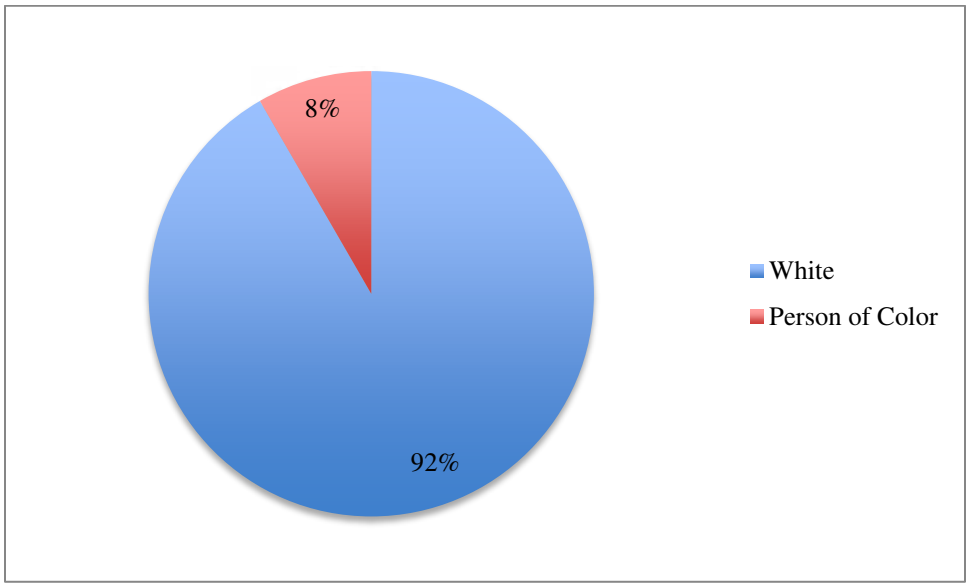


Figure 4.5: Race of Newbery Medal-winning authors.

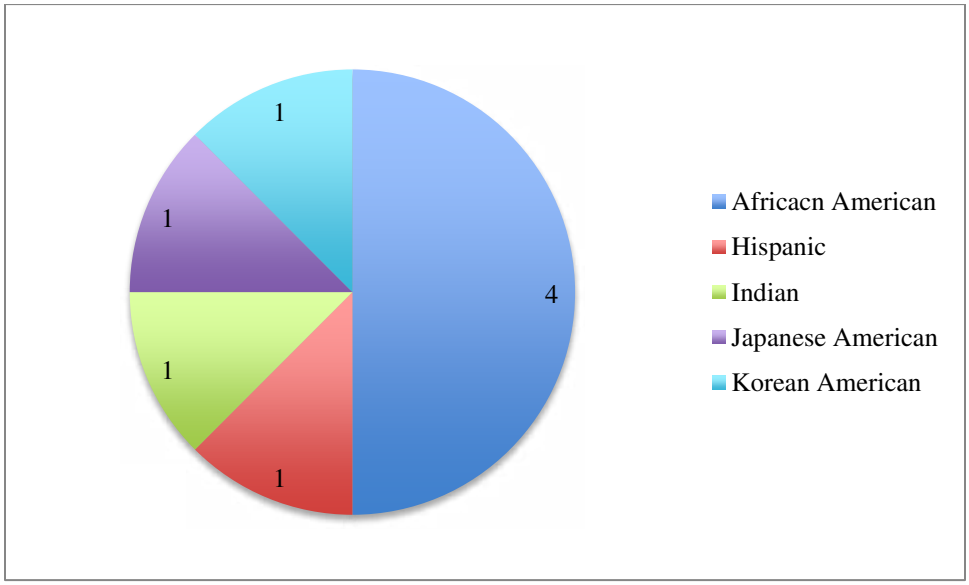


Figure 4.6: Race of writers of color in the Newbery Medal corpus, where *n*=8.

Depictions of race and ethnicity in the corpus. Unsurprisingly, given the Newbery's significant tendency to privilege texts written by white authors, the Newbery corpus also significantly features texts about white characters or texts, such as high fantasies, that do not specify the race of characters (χ^2 (1, N=96) = 18.375, p>.0001). As shown in figure 4.7, there are 27 Medal-winning texts with main characters of color or specified ethnicities beyond white American.

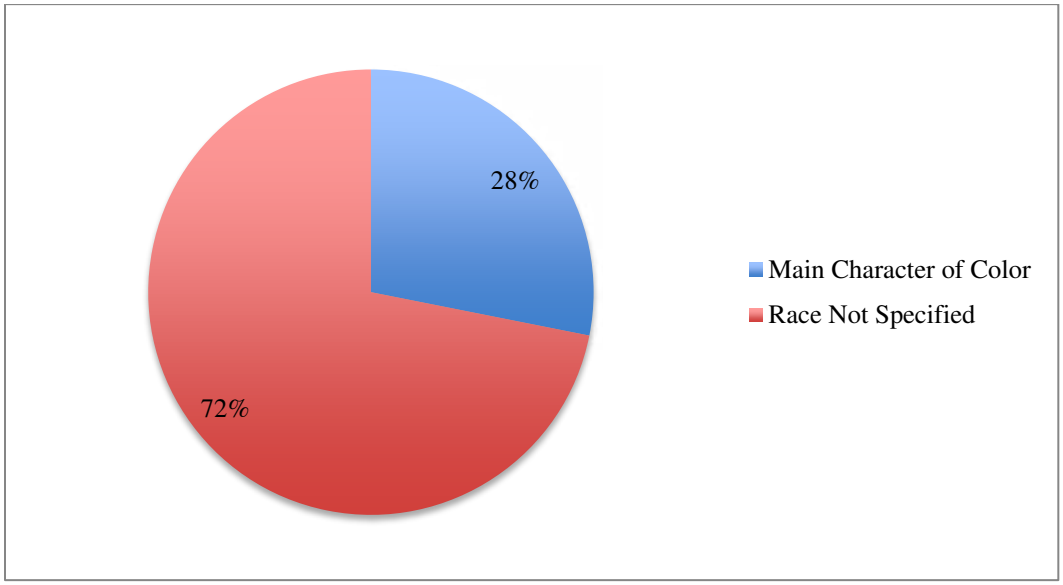


Figure 4.7: Texts featuring main character(s) of color in the Newbery Medal corpus.

Table 4.1 illustrates the race or ethnicity depicted in these 27 Newbery Medal winners in relationship to the number of authors in the corpus from those racial or ethnic groups. As this data indicates, the Newbery Medal does not privilege what Corinne Duyvis terms "#OwnVoices," or literature "about diverse characters written by authors from that same diverse group" (Duyvis, 2015). Only five titles provide fictional representations of a diverse group by authors from that same group. Recent scholarship has emphasized the importance of

#OwnVocies (e.g., Gall, 2017; Gómez, 2016; Pérez, 2017) in children's literature. In the area of #OwnVoices literature, the Newbery Medal corpus is sadly and statistically significantly lacking.

Table 4.1: Race or ethnicity of main characters in the Newbery Medal corpus.

Race / Ethnicity of Main Character	Frequency: Main Character	Frequency: Author	Frequency: #OwnVoices
African American	7	4	4
Native American	3	0	0
Chinese	2	0	0
Spanish	2	0	0
Arab	1	0	0
Bulgarian	1	0	0
Hispanic	1	1	0
Huns	1	0	0
Indian	1	1	1
Indians of South America	1	0	0
Inuit	1	0	0
Japanese	1	0	0
Japanese American	1	1	1
Korean	1	0	0
Korean American	0	1	0
Palestinian	1	0	0
Peruvian	1	0	0
Polynesian	1	0	0

Describing Newbery Medal-Winning Books

Publishers. Thirty-seven different imprints, or the trade names associated with the various arms of a publishing firm, published the works of Newbery Medal winners. These 37 imprints, however, represent 17 different publishing companies, with 69 Medal-winning titles coming from an imprint currently owned by the big five publishing companies of Hachette, Simon & Schuster, HarperCollins, Macmillan, and Penguin Random House. Houghton Mifflin, a publishing house specializing in educational as well as trade publications, published a further 14

Medal-winning titles (see figure 4.8). As of 2017, just six publishing companies account for 83 of the 96 Medal-winning titles, or 86% of the corpus. Only three companies that published a Medal-winning text are now completely defunct: Dodd, Mead, and Company; Lippincott; and Stokes.

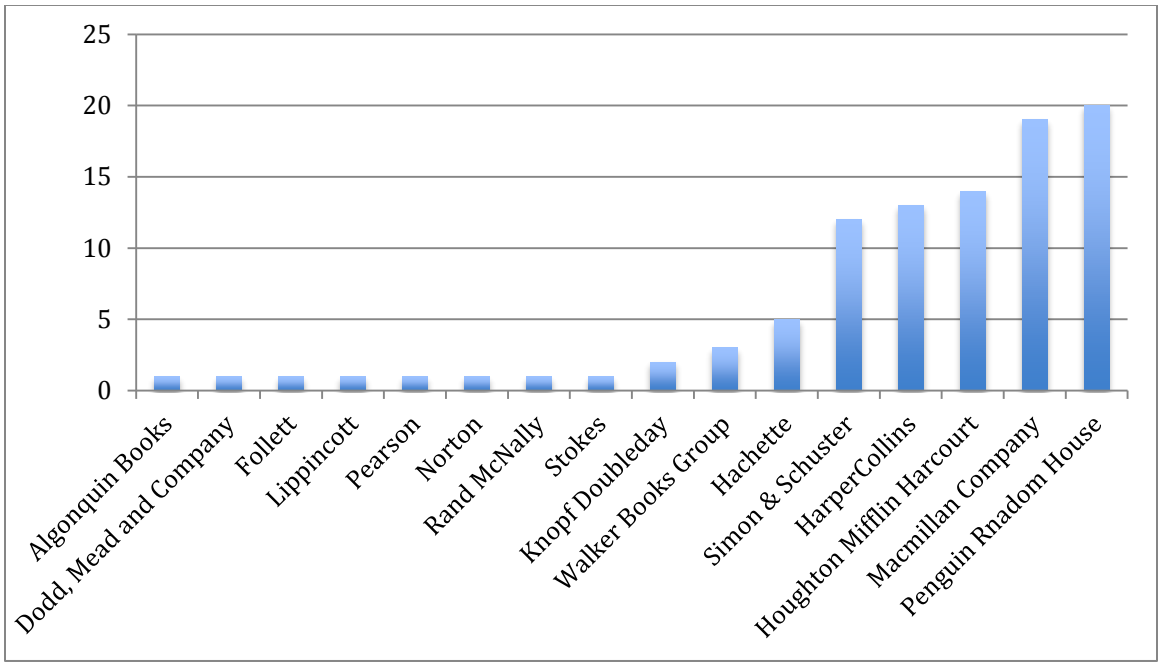


Figure 4.8: Publishing companies of Newbery Medal winners, 1922-2017.

The creation of publishing conglomerates results in works represented by a handful of corporations achieving critical and commercial success at the exclusion of works represented by smaller presses (Fialkoff, 2013; Maryles, 2015). This general trend is observable among the Newbery Medal winners as well. The works the Newbery privileges represent mainstream American publishing and publishers, with only Algonquin Books representing a small press in the corpus. Despite critics' tendency to criticize the Medal for favoring literary experimentation

and formal innovation (Devereaux, 2008; Silvey, 2008), the corpus actually favors mainstream and established venues, not experimental small presses.

Length. Descriptive statistics (see table 4.2) illustrate the relative lengths in the corpus of Newbery Medal-winners. The mean length, in terms of pages, of a Newbery Medal-winning text is 201.4 pages, while the mean length, in terms of word count, is 44,236.3. The maximum observed value for both measures occurred in 1922, with the *The Story of Mankind*, the first Medal-winning title, and the minimum observed value occurred in 2016, with the Medal's only picture book winner, *Last Stop on Market Street*. The standard deviations observed in both measures indicate a high level of variance across the corpus. In order to explore the distribution of this variance more closely, I analyzed each decade individually (see figure 4.9). Despite the ability for pages to include radically different word counts based on type size and spacing, these two measures follow very similar trajectories over time. During the Newbery Medal's first decade, works with both the largest number of pages and the largest word count occurred. After this original profusion of text, the corpus follows a similar pattern over thirty-year time spans: a sharp decline followed by resurgence in length across both measures. This pattern does not change until the most recent decade, when the mean number of pages continued to rise while the mean number of words followed the established pattern and declined. It is notable that the most recent decade's winners include a graphic novel as well as a verse novel. Both of these formats, heavily reliant of spatial relationships on the page and unusual within the corpus, could contribute to this change.

Illustrative content. The criteria used for selecting the Newbery Medal winner explicitly state that the selection committee does not consider illustrative content unless this content "make[s] the book less effective" (ALSC, 2015a). This criterion limits the Medal to

considerations of the text of a book alone, despite the crucial role illustration plays in children's literature (Avery, 1994; Darton, 1932/1982; Serafini, Kachorsky, and Aguilera, 2016). As other scholars have previously noted, illustrations in children's books work together with text symbiotically to create, extend, and complement meaning (Nikolajeva and Scott, 2013; Nodelman, 1989; Schwarcz and Schwarcz, 1991). In this section, I consider the extent to which ignoring illustration in the Newbery Medal corpus could prove problematic.

Table 4.2: Descriptive statistics illustrating relative length of Newbery Medal corpus.

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
Length (number of pages)	201.4	194.5	84.28	32	489
Length (number of words)	44236.31	40531	24210	757	170226

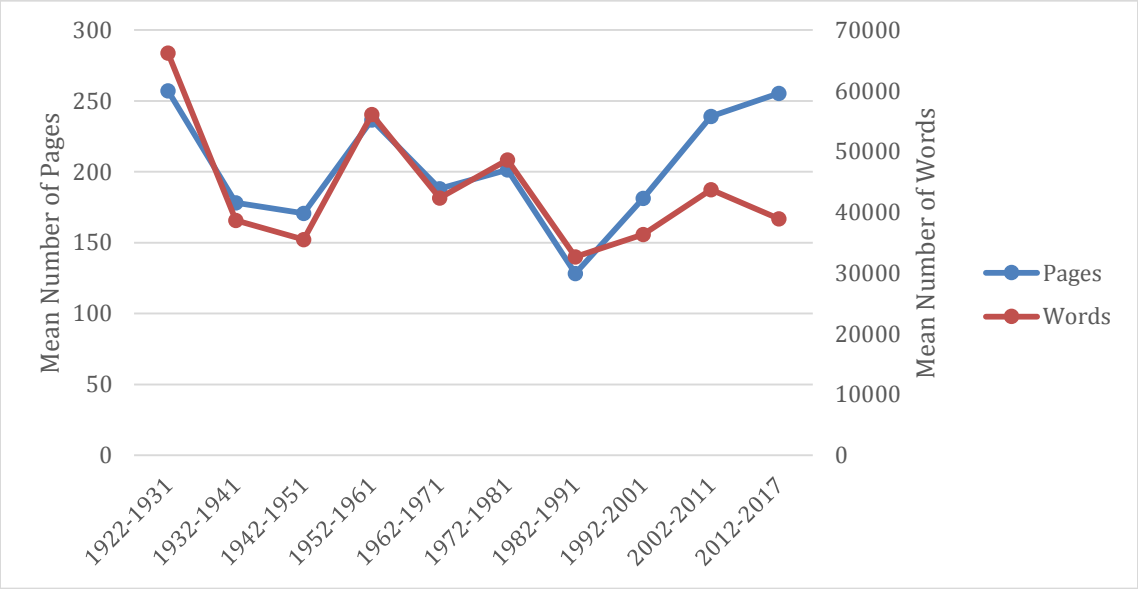


Figure 4.9: Mean length of Newbery Medal corpus by decade.

As of 2017, 65 Medal-winning titles have included illustrations and 31 have not. The observed frequency of illustration within the corpus is significant (χ^2 (1, N=96) = 12.042, p>.0005), indicating that the Newbery statistically favors illustrated texts over those without illustrations. As with the gender distribution of Medal-winning authors, illustration status has varied significantly in frequency over time (see figure 4.10). The first three decades of Medal-winning titles were all illustrated; the first non-illustrated winner did not occur until 1958, when Harold Keith's *Rifles for Watie* received the Medal. Only one decade in the Medal's history, 1992-2001, witnessed a significant number of non-illustrated titles win the award. In all other respects, as measured by descriptive and structural variables, 1992-2001 was an unremarkable decade for the Medal. There were no significant variations or even suggestive fluctuations in terms of length or text complexity.

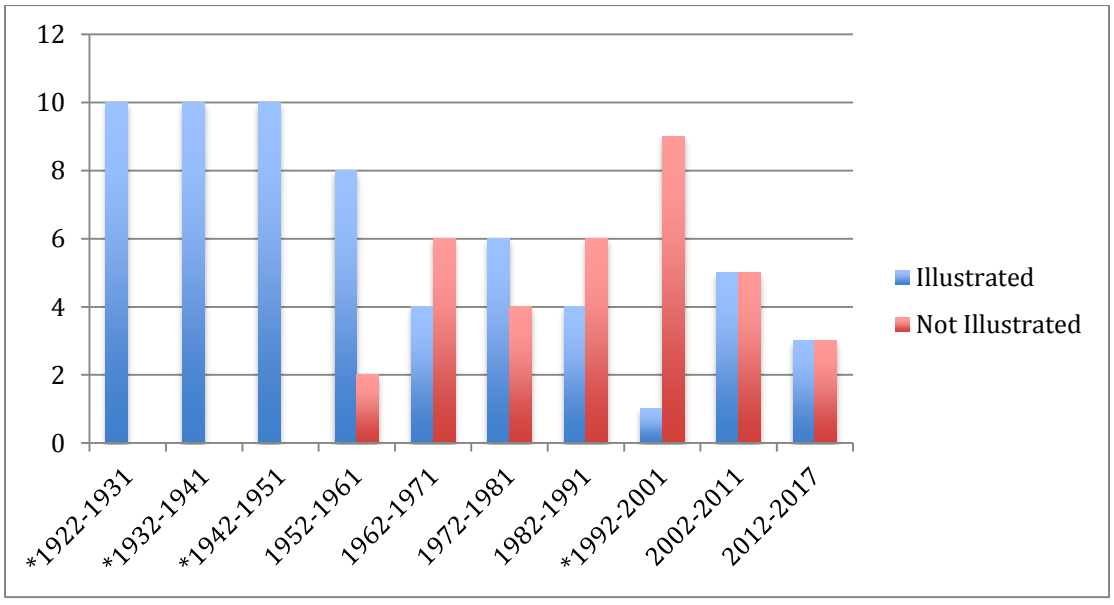


Figure 4.10: Illustrations in Newbery Medal corpus by decade. Distributions significant at the p>.05 level are designated with an asterisk.

Bibliographic information for the corpus names the illustrator for 61 of the illustrated Medal-winning texts. The illustrators for the remaining four texts are un-credited. As with authors, a large number of different illustrators are represented in the Newbery corpus. Of the 61 credited illustrators, only four illustrated two different Medal-winning titles: Lynd Ward (1931 and 1944), Kate Seredy (1936 and 1938), Robert Lawson (1943 and 1945), and Jean Charlot (1953 and 1954). 12 illustrators both wrote and illustrated the text for which they received the Newbery Medal, and the remaining 49 Medal-winning titles have different authors and illustrators. Despite claiming that illustrations are inconsequential, the Newbery nevertheless privileges illustrated texts. The presence of illustrative content in a book, however, does not necessarily indicate the importance of that illustrative content. It does not, for example, enable differentiation between full page illustrations and small illustrations at the head of chapters, nor does it provide information about the visual element's semantic meaning.

The most significant limitation to bibliographic information about illustration in the corpus relates to identifying illustration type. The physical description field in WorldCat (Marc field 300) provides the most robust information about illustration status for each title, and it has the capacity to provide large amounts of data. Sample data elements describing Newbery Medal-winning texts found in Marc field 300 include statements such as "chiefly illustrations (colour)," "frontispiece, plates, portraits," and "maps." Taken holistically, however, data from Marc field 300 in WorldCat records illustrates the dearth of relevant, descriptive bibliographic information available to describe the illustrative content of the corpus (see table 4.3). Most illustrations are described simply as "illustrations," and the most frequently used descriptor for illustrations notes the presence of color. Maps and portraits are the only type of illustration to receive a more specific description.

Table 4.3: Frequency distribution of illustration type in Newbery Medal corpus

Illustration Type	**Count**
Illustrations	43
Illustrations (some color)	7
Color frontispiece, illustrations, color plates	2
Color illustrations	2
Illustrations, map	2
Map	2
Chiefly illustrations (colour)	1
Color frontispiece, color plates	1
Frontispiece, plates, portraits	1
Illustrations (including maps; color plates	2
Illustrations, double plates	1
Illustrations, map, plates	1

The general note field in WorldCat records (Marc field 500) provides the opportunity to include additional information about illustration type. This field is used only 11 times in the entire Newbery Medal corpus to record information about illustrations, although the notes that do exist suggest the potential for the type of information not found in other records and the significance of the illustrations that remain undescribed. The most common type of note used to describe the illustrations comments on placement, including notes such as "illustrated lining-papers" (*Waterless Mountain*, 1932), "lining-papers illustrated in colors" (*Thimble Summer*, 1939), and "illustrated lining papers. Color illustrated frontispiece. Illustrated headpieces" (*Tales from Silver Lands*, 1925). Other notes draw attention to the type of illustration featured in the text, including "photographs" (*Lincoln: A Photobiography*, 1988), "map on lining-papers" (*Johnny Tremain*, 1943), and "art techniques used: whimsical gouache, pen and ink paintings" (*A Visit to William Blake's Inn: Poems for Innocent and Experienced Travelers*, 1982). This last note, in its use of the adjective "whimsical," is the only descriptive note in the corpus that suggests the emotive and thematic work accomplished by illustration.

Literary characteristics of the corpus. Two structural variables, point of view and literary form, describe some of the literary qualities that the corpus privileges. Taken together, they help assess claims that the Newbery Medal privileges formal experimentation (e.g., Devereaux, 2008; Silvey, 2008; Miller, 2014). The observed frequency distribution of point of view in the corpus suggests that the Newbery favors the traditional narrative formats of first and third person (see table 4.4). Only four titles make use of more experimental narrative strategies. *The Matchlock Gun* (1942) couples predominantly third person narration with a forward written in second person, while *The Girl Who Drank the Moon* (2017) relies predominantly on third person with scattered passages in first and second person. E. L. Konigsburg wrote two of the four Medal-winning titles that make use of a mixed narrative strategy. *From the Mixed-Up Files of Mrs. Basil E. Frankweiler* (1968) frames third person narration with a first-person letter Mrs. Frankweiler that introduces the premise of the text; Mrs. Frankweiler sprinkles first person comments throughout an otherwise straightforward third person account. *The View from Saturday* (1997), meanwhile, alternates between first and third person, with different narrators offering first person accounts in different chapters. From the perspective of narrative strategy, and in contrast to reports of narrative changes, formal experimentation is actually quite limited in the corpus.

Table 4.4: Frequency distribution of point of view in Newbery Medal corpus.

Point of View	**Frequency**
3rd person	60
1st person	31
Mixed	4

As with the other descriptive and structural variables analyzed in this chapter, the corpus displays interesting changes in point of view and narrative strategy over time (see figure 4.11). The second two decades of the Medal, 1932-1941 and 1942-1951, are the only two decades with significant (p=.05) distributions. The former is composed entirely of third person narratives, and the latter features nine third person narratives and one mixed narrative strategy. First person narration slowly builds throughout the corpus, reaching a peak between 1992 and 2011. The two most recent decades, 2002-2011 and 2012-2017, display the most balanced distribution of narrative strategies. In 2002-2011, the Newbery shifted to five first person and five third person narratives, while in 2012-2017, the winners included 3 first person, 2 third person, and 1 mixed narrative approach.

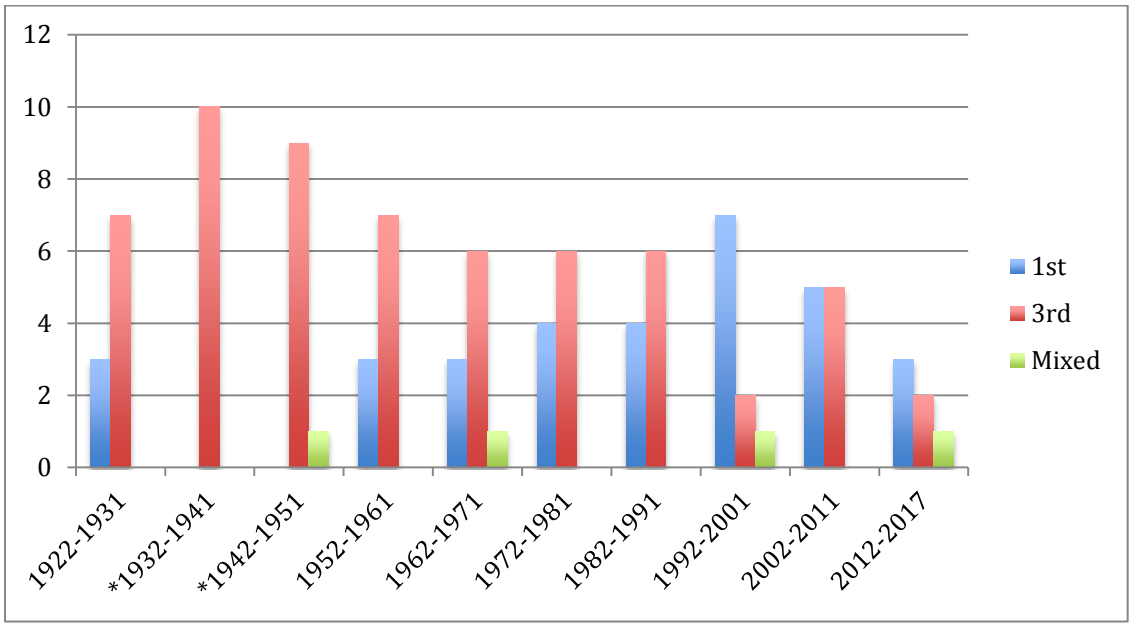


Figure 4.11: Point of view in Newbery Medal corpus by decades. Distributions significant at the p>.05 level are designated with an asterisk.

Literary form privileged by the corpus. Genres in children's literature are notoriously difficult to analyze due to the ways that libraries and publishers classify children's literature.

Frequently, publishers note that a title is either fiction or nonfiction, and scholars note in passing the wide range of genres, including but not limited to mystery, fantasy, poetry, nonfiction, drama, and science fiction, encompassed by the phrase "children's literature" without offering a mechanism for exploring the range of subgenres inherent in children's literature (Hunt, 2001; Nodelman, 2008). In this section, I test methods for moving beyond the broad classification of "fiction" and nonfiction" and understanding subgenres represented in the Newbery Medal corpus on a more granular level. This analysis also provides the ability to test which data sources provide the most productive and accurate descriptions of genre. It is important to reiterate, however, the limitations of the Newbery corpus: it includes only one picture book, a small smattering of poetry, and very limited nonfiction texts. As such, this subset of texts is not representative of literary form across a wider range of children's literature.

Genre headings in WorldCat. Library of Congress genre headings, as recorded in WorldCat, provide a mechanism for investigating previous analyses of the types and frequencies of genres represented in the corpus. All titles in the corpus except for *Thimble Summer* (1939) have been assigned at least one genre heading (Marc field 655), and most have more than one genre heading. *Carry On, Mr. Bowditch* (1956) has the most genre headings, with eight different terms describing this work (see table 4.5). Sixty-six unique genre terms have been used to describe the literary form of the Newbery Medal corpus. Only six unique terms, however, have been used five times or more as descriptors (see table 4.5), and 42 terms have been used only once. The most frequently used genre headings prove generic to the point of unhelpfulness in isolating all but the largest scale trends within the corpus. "Juvenile fiction" occurs 83 times and the even less specific "fiction" occurs 76 times, indicating a corpus largely comprised of fiction and corresponding with what is already known about the Medal (Clark, 2003; Kidd, 2007).

Table 4.5: Top six Library of Congress genre terms used to describe Newbery Medal corpus.

Library of Congress Genre Terms	Frequency
Juvenile fiction	83
Fiction	76
Juvenile works	62
Juvenile literature	17
History	15
Fantasy fiction	5

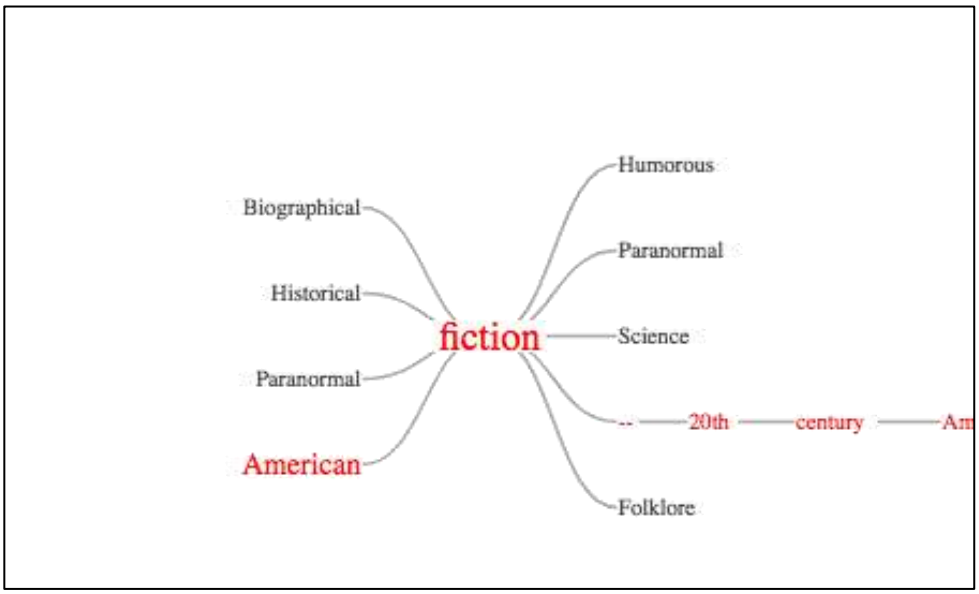


Figure 4.12: Word tree illustrating low-use genre terms and their relationship to the term "fiction." Word trees in Voyant are fixed-width, resulting in the visualization cutting off the end of the word "American." Text in red highlights related terms occurring in different contexts: "American fiction" and "fiction – 20th century – American."

"Juvenile works" is the third most frequently used term, and many other terms, such as "juvenile materials" and "children's stories," mark the text as for a juvenile audience. Collocating these terms together, a juvenile audience marker occurs 192 times. Although these terms help users isolate materials for a juvenile audience in a large bibliographic database with

many types of items represented, thereby explaining their popularity in WorldCat, they do not help analyze further generic breakdown of a corpus already known to contain only juvenile works. The terms used do suggest, however, that libraries tend to categorize Newbery Medal winners as children's literature rather than young adult as the genre term "young adult" occurs only twice. Interestingly, the term "historical fiction" occurs only twice, despite the known prevalence of historical fiction within the corpus (Kidd, 2007). WorldCat instead prefers the term "history," despite the fact that the terms are, by and large, describing works of fiction rather than nonfiction texts about historical events.

In their individuality, the remaining low-use genre terms suggests a wide range of micro fictional genres within the corpus. Frequencies within this small corpus support this supposition. The most frequently occurring word is "fiction" (10), followed by "stories" and "adventure." A word tree illustrates the relationship between other terms collocated with the term "fiction" in the corpus (see figure 4.12), showing a range from "paranormal" to "biographical." These low-use terms provide additional generic information about the type of fiction described and indicate the presence of a range of fictional stories in the corpus.

***Genre headings in Accelerated Reader*.** Accelerated Reader genre tags vary greatly in terms of their specificity and scope. Unlike Library of Congress Subject Headings in WorldCat, which uses the same five terms to describe a large percentage of the corpus, Accelerated Reader employs 39 distinct genre terms a total of 65 times, with the most frequently used term, "folklore/fables/myths-folklore/fables/myths (all)" occurring only five times. The subdivisions that Accelerated Reader database applies to its headings results in seemingly unique terms, despite terms sharing the same root. The term "adventure," for example, appears as a genre heading describing the Newbery Medal corpus 16 times. These 16 instances are further modified

by nine different subheadings to specify the type of adventure found in the text: adventures, danger, discovery/exploration, escape, abandoned, misc./other, runaway, survival, and travel. This specificity has the potential to help readers find and identify a very precise type of book, but the sub-headings employed range from the oddly worded ("abandoned") to the patently unhelpful ("misc./other"). When subheadings are deleted and headings collapsed, Accelerated Reader categorizes 14 genres within the corpus (see table 4.6). These genres largely focus on the identifying categories of books: adventure stories, fantasy, and humor, for example. No genre tags attempt to categorize books by reading level, as Library of Congress Subject Headings do with their preference for "juvenile" rather than "young adult." This is unsurprising since Accelerated Reader uses both Lexile measures and ATOS levels to suggest reading levels and, indeed, the general unhelpfulness of the subject headings in Accelerated Reader suggests that the database does not expect many users to rely on thematic headings to choose a book.

Also of note is the prevalence of tags that suggest a non-realistic genre. The Newbery is known for its preference for historical fiction, which both Library of Congress Subject Headings and Accelerated Reader tags support. Failing a historical setting, the Newbery is thought to favor realistic settings depicting white, middle class families (Kidd, 2007). Several different terms, however, suggest a fantastical genre: fantasy/imagination, folklore/fables/myth, fairy tales, and science fiction. Taken together, these terms occur 20 times across the corpus, pointing to a prominent thread of the magical, the mystical, and the decidedly not realistic in the corpus. Also of interest is the most frequently applied term: adventure. Uniquely among the three databases, Accelerated Reader uses the heading "mystery" somewhat frequently to describe subgenres in the corpus. As with "adventure," the subheadings applied to "mystery" are wide-ranging: ESP, missing persons, murder, supernatural, treasures, and who-dun-it.

The Newbery, then, is not solely a corpus of mostly historical fiction stories or stories about middle class families. It is also a corpus of adventure, the mysterious, and the fantastic. These subgenres deserve further, microscopic scrutiny in order to understand how the non-realistic functions in Newbery Medal-winning texts.

Table 4.6: Simplified genre headings applied in Accelerated Reader to describe Newbery Medal corpus.

Accelerated Reader Genre Tags (Simplified)	**Frequency**
Adventure	16
Fantasy/Imagination	13
History	10
Mysteries	6
Folklore/Fables/Myth	5
Historical Fiction	4
Poetry	3
Arts	1
Biographies	1
Classics	1
Fairy Tales	1
Horror/Thriller	1
Humor/Funny	1
Science Fiction	1

Genre tags in Goodreads. Goodreads displays user tags in its interface with the label "genre," thereby encouraging users to create tags that describe the genre observed within the work that the tag describes. As with Library of Congress and Accelerated Reader genre headings, Goodreads genre tags are numerous and wide ranging. They also strongly indicate the corpus' preference for fiction. This study considers the top 10 tags assigned to each book in the Goodreads databases, and a total of 59 different tags meeting this criterion have been applied to

the Newbery corpus 796 times. Thirteen of these terms have been used 18 times or more (see table 4.7). The remaining tags were applied to fewer than ten titles across the corpus.

Table 4.7: Most frequently applied genre tags in Goodreads describing the Newbery Medal corpus.

Goodreads Genre Tags	Frequency
Children's	94
Fiction	89
Young Adult	89
Children's -- Middle Grade	65
Classics	61
Historical Fiction	59
Children's -- Juvenile	57
Historical	38
Realistic Fiction	36
Academic -- School	31
Adventure	26
Fantasy	22
Children's -- Chapter Books	18

Unlike Library of Congress genre tags, which indicate that corpus is largely composed of works for children rather than young adults (e.g., "juvenile literature" and "juvenile works"), Goodreads genre tags provide more granularity in identifying appropriate reading levels for children's fiction. In addition to the general tag "children's," four additional terms subdivide the children's genre by appending reading level: chapter books, juvenile, middle grade, and picture books. Goodreads users also find many more young adult titles in the corpus than Library of Congress and Accelerated Reader genre tags do. "Young adult," "young adult – coming of age," and "young adult – teen" occur 102 times. Although this tag occurs less frequently than permutations on children's, which occurs 242 times, Goodreads users nevertheless suggest that

the corpus contains a significant amount of young adult fiction, pointing to a difference in the way users classify literature for older children and younger teens.

"Historical" and "historical fiction" remain the most prominent subgenre identified by Goodreads users, occurring 98 times. As seen in Accelerated Reader headings, Goodreads users also find a prominent thread of adventure and the fantastic running alongside the expected historical fiction. The fantastic, though, is here contrasted with the tag "realistic fiction." Uniquely among the three databases, Goodreads employs a tag for the genre "romance" (three occurrences). The low frequency of this term and its isolation to Goodreads warrants closer investigation to ensure that the tag is not an example of messy or inappropriately used terminology. The "romance" tag describe *Criss Cross* (2006), *Up a Road Slowly* (1967), and *The Witch of Blackbird Pond* (1959), and book summaries for all three titles indicate that romantic relationships occur in each text. The Goodreads folksonomy successfully pinpoints elements in the corpus that would remain hidden if it were not a part of the analysis.

The folksonomy found in Goodreads genre tags introduces an additional element into a discussion of genre in the corpus: assigned school reading. Two different tags, "academic – read for school" and "academic – school" were applied to 40 of the 96 Newbery Medal winners in the Goodreads database. The prevalence of this tag across the corpus quantifies Kidd (2007) and Clark's (2003) assertions that individual Newbery Medal winners are frequently assigned. At least 42% of the corpus is assigned with enough frequency to prompt Goodreads users to tag the titles as school reading. Similarly, Medal winners were tagged as "classics" 61 times, suggesting that users consider Medal winners as Literature (with an intentionally capitalized L) or, at the very least, texts that are worthy of being assigned reading.

Text complexity and implied readerships. Scholars have long debated questions about text complexity and intended age range for Newbery Medal-winning titles. As the discrepancy in the application of tags for "children's literature" and "young adult literature" in Goodreads, WorldCat, and Accelerated Reader indicates, this question extends beyond the academy and into the realm of libraries, publishers, and readers. The age range that the Medal claims to serve is incredibly large, and many critics have previously noted that the Medal does not actually represent texts for the entire age range that it purports to cover (Leal and Chamberlain-Solecki, 1998; Schafer, 1976). Early studies on readability focused on answering claims that the Newbery Medal-winning titles were frequently too difficult for children to read and finding that the winners were most often at or above a sixth-grade reading level (Schafer, 1976; Schafer, 1986). A more recent study charted change in the corpus' readability levels over time, finding a decrease in difficulty of texts awarded the Medal, on average, during each decade (Stevens, 2010). Readability and implied readerships represents the most analyzed portion of the corpus from a holistic perspective in previous studies, but only one study (Stevens, 2010) considered the entire corpus at the date of the study's completion. In this section, I consider methods for bringing existing critical conversations on text complexity in the corpus up to date as well as how methods influence understandings of complexity.

As with length, the two measures of text complexity considered in this study, Lexile measure and ATOS level, are widely variable. Descriptive statistics (see table 4.8) illustrate the relative complexity of the Newbery Medal corpus. The mean Lexile measure of a Newbery Medal-winning text is 871, and the mean ATOS level is 5.545. The maximum observed Lexile measure and ATOS level occurred in 1922 with Van Loon's *The Story of Mankind*, and the minimum observed Lexile measure value occurred in 2014 with Kate DiCamillo's *Flora &*

Ulysses: The Illuminated Adventures. The minimum observed ATOS level occurred in 2016 with Matt de la Peña's *Last Stop on Market Street*. The standard deviations observed indicate a high level of variance across the corpus. In order to explore this variance more closely, and to compare my findings with those of previous scholars, I analyzed the mean complexity as observed in Lexile measure and ATOS level during each decade individually (see figure 4.13), with decade defined as ten-year increments of the Medal. The first decade of the Newbery Medal witnessed the most complex winners in the corpus by both measures, with a mean Lexile measure of 1131 and mean ATOS level of 6.7, while the most recent decade witnessed the least complex Medalists, with a mean Lexile measure of 1131 and an ATOS level of 4.33. In the intervening decades, these two measures of text complexity followed a similar, shifting pattern of declining after the initial peak, rebounding slightly, declining, rebounding to a slightly lower level than the previous rebound, and declining again. Neither measure has come particularly close, however, to matching the first decade for difficulty.

Table 4.8: Descriptive statistics illustrating measures of text complexity in the Newbery Medal corpus

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
Lexile Measure	871	855	184.2	520	1440
ATOS Level	5.545	5.5	1.041	3.3	9.9

It is also crucial to point out the ways in which methodological decisions affect understandings of text complexity in the corpus, the area that has the most extensive previous holistic study. Schafer (1976 and 1986) relies on analysis of each title alone and the mean observed values for the entire corpus; he does not consider changes in the corpus over time.

Stevens (2010), meanwhile, analyzes change over time based on chronological decades. My own analysis relies on an analysis of change over time based on ten-year publication time spans (e.g., 1922-1931 and 1932-1941 rather than 1922-1929 and 1930-1939). The simple decision of how to count a decade leads to different results: Stephens observes no positive spikes in text complexity, noting instead a general decline. My own findings also find a general decline in text complexity but also note a decades-long waxing and waning pattern that recurs on a thirty-year cycle. Schafer (1976 and 1986) finds the Newbery Medal corpus significantly offers works for middle grade readers, grade six and above. My own findings suggest a slightly lower reading level, with an average reading level of grade 5.5 (based on ATOS level) and grade 6 (based on Lexile measure). This subtle shift reinforces previous research on leveling books. Levels are not neutral, constant measures but rather agreed upon definitions, which are subject to changing norms and educational practices (Hiebert and Mesmer, 2013). They should be interpreted as such. Despite the observed shifts, this study confirms that the Newbery privileges texts at the more complex end of the range eligible for the award rather than privileging texts for all reading levels.

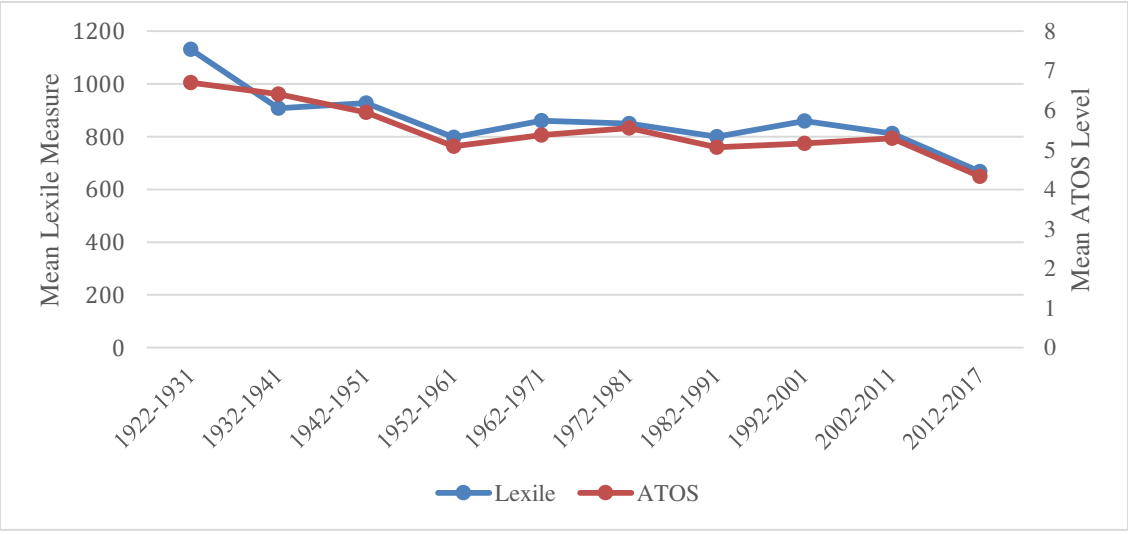


Figure 4.13: Mean Lexile measure and ATOS level by decade.

Analyzing the Holistic Description of the Newbery

The findings presented in this chapter provide insights into the descriptive and structural elements of texts that the Newbery Medal privileges through prizing. Additionally, analyzing these elements by decade intervals enables a consideration of how the privileged structures have changed over time. Taken together, the descriptive and structural variables I analyzed reveal a remarkably homogenous corpus that favors slow, gradual, and predictable change over time. The Newbery Medal prefers to honor new-to-the-corpus authors rather than rewarding the same authors over and over again. Nevertheless, the authors favored statistically tend to look alike: they are predominantly white and female. Further, a small handful of publishing companies tends to represent the homogenous author type that the Medal prefers. These companies represent standard, large presses, not small presses known for innovative practices.

Despite the Newbery Medal criteria's claim that illustration is unimportant, the Medal nevertheless privileges illustrated texts. As with authorship, the Medal prefers texts illustrated by a new illustrator rather than rewarding the same illustrator year after year. Only one decade in the Medal's history, 1992-2001, witnessed a significant number of non-illustrated titles win the award. This is an unremarkable decade in terms of length and text complexity, with neither particularly high nor particularly low values observed in any other variable. This suggests that illustration is not particularly tied to length or complexity in this corpus. At the same time, the Medal also privileges texts for the older child reader rather than picture books or early readers, which typically contain more illustrations. Frustratingly, available data sources do not provide much in the way of descriptive detail about the types or placement of illustration in the corpus despite the existence of metadata schema that would allow for this information to be recorded.

In very many ways, then, the Newbery Medal privileges a homogenous type of text.

Variation in the corpus does exist, but when it occurs, it tends to occur in predictable patterns. The peaks and valleys across length and measures of text complexity, for example, occur in familiar patterns. A slow decline across all measures of length and complexity has occurred over the past century, but routine positive spikes punctuate this overall decline.

Relying on identifiable variables allows for systematic analysis of the corpus, taking into account both the corpus in its entirety as well as smaller chunks. This systematic analysis across a range of variables brings nuance to existing critical conversations. Discussions of gender representation in the corpus, for example, have previously considered one isolated ten-year span (Jenkins, 1996), with this ten-year span coming to influence understandings of gender in the entire corpus. Similarly, attention to statistical significance introduces additional layers to critical conversations about gender, particularly relating to gender of characters. Although critics have long argued that girls' stories outweigh boys' stories in the corpus (Pease, 1939; Kidd, 2007), this is, in fact, simply perception. In terms of sheer numbers, stories about boys outweigh stories about girls in the corpus, although the observed frequency is not statistically significant. Critical complaints about female dominance in the corpus do not map to actual observed data points, illustrating the necessity of considering the entire corpus before making such assertions.

The findings also illustrate the affordances and limitations of available data sources for distantly reading the Newbery Medal corpus. The variables available for analysis undeniably introduce limitations into statistical examination of the corpus. The information describing illustrations in the corpus points to the serious limitations of existing practices in describing texts, and the genre headings applied across WorldCat, Accelerated Reader, and Goodreads also suggest the restrictions imposed on distant reading by a reliance on metadata rather than full text. The headings and tags used to describe the corpus vary so greatly that it is difficult to quantify

much beyond a preference for fiction, although tantalizing glimpses of types of texts beyond the expected historical fiction are findable. When relying on metadata, analysis can only be as specific as the description offered in that metadata. The existing frameworks in which the metadata exist also inevitably influence the resultant analysis. Library of Congress Subject Headings, for example, illustrate the ways in which WorldCat descriptions exist to help users find and locate materials for a juvenile audience within the larger database of materials for juvenile, adolescent, and adult readers. Accelerated Reader genre headings, meanwhile, betray a fundamental weakness in their formation and the vocabulary employed when analyzed for their utility as genre access points. Their odd formation, however, actually facilitates the location of patterns in distant reading, particularly in highlighting the thread of the fantastic running through the corpus. Goodreads introduces questions of audience and implied readerships into the question of genre while also illustrating the frequency with which Newbery Medal texts are assigned reading for Goodreads users. Taken together, however, the three data sources provide a much more complete picture than any single source alone. From a more expansive, methodological perspective, when metadata provides the backbone for distant reading, the use of multiple sources proves crucial.

The holistic description I have offered above of the Newbery Medal corpus underscores the importance of critical, explicit attention to questions of methodology and the need for studies to provide sufficient, and to the degree possible, explicit, detail of methodological decisions. The findings relating to text complexity provide the most salient example, contrasting as they do with existing research in this area. Previous studies have examined text complexity using a more holistic approach than any other area of the corpus, and the findings of these studies are inextricably linked to the methods used to study the question of complexity and appropriate

reading audience. Method, in short, matters, and understanding the methodological decisions underpinning a study provides crucial information for understanding the findings put forth by that study.

Chapter 5: Analyzing Theme and Content

In this chapter, I explore methods for analyzing the thematic characteristics of the Newbery Medal winners and for interrogating existing critical stories about thematic motifs against data from the entire corpus. As with the descriptive and structural elements of the corpus, I rely on variables created from a variety of data sources to foreground a holistic, macroscopic consideration of thematic elements. These variables enable analysis of the themes, motifs, subjects, and contents that the Newbery Medal privileges through its prizing. Further, they enable a consideration of these elements through the lens of various audiences.

Existing critical stories about thematic motifs in the Newbery put forth the idea of a corpus replete with socially safe historical fiction and solidly middle class values (Alberghene, 1981; Cook, 1985; Dyson, 2007; Kidd, 2007). Additionally, critics point out that the historical or exotic lens offers a mirror for contemporary American social constructs (Cook, 1985; Moir, 1981). Small towns, the country, or an exotic but real land, this story continues, provide the preferred settings for representing ideals inhabited by the quintessential middle class American child (Alberghene, 1981; Solt, 1981). The story that the descriptive and structural metadata tells about the Newbery, explored in chapter 4, suggests that these stories are true. It also suggests, however, that there are other stories, untold or glossed over, to be found in the corpus, particularly fantastical stories with no overt tie to realism.

What's This Corpus About?: Using Summaries to Analyze Theme and Content

WorldCat, Accelerated Reader, and Goodreads all provide descriptive synopses of Newbery Medal-winning titles. Each data source provides these summaries for slightly different

purposes and with different audiences in mind due to the purpose of each database, so it is to be expected that the summaries vary in length and content. None explicitly state who wrote the descriptions offered on the platform, although contextual clues provide occasional hints. Goodreads, an Amazon company, occasionally repurposes descriptions from Amazon, and WorldCat and Accelerated Reader sometimes include phrases like "publisher's description." Variation in length, style, and content across each platform suggests that each relies on variety of sources, some of which provide more useful information for distant reading than others. A few individual titles in the corpus have identical summaries across WorldCat and Accelerated Reader, and both *Ginger Pye* (1952) and *Dobry* (1935) each share the same summary across all three platforms, respectively, suggesting that some of the summaries come from the same, uncited source.

Voyant Tools enables analysis of each platform's descriptions, offering the ability to calculate word frequency and visualize how the most frequently used words occur across the corpus of summaries. Corpus size for each set of descriptions confirms that descriptions vary across platforms: Accelerated Reader descriptions contain 2,356 words, WorldCat summaries contain 3, 779 words, and Goodreads descriptions are considerably longer at 10,698 words. Each corpus contains a high percentage of unique word forms, suggesting variation in theme and content. Even though the three platforms rely on summaries from different sources, the most frequently occurring words display remarkable consistency across the three platforms (see table 5.1).

The most frequently used words in summaries provide insights into the content of the corpus, including the types of texts included, the characters featured, setting, and narrative action. The word "old" appears at or near the top of each most frequent word list. Word trees,

which display the word in context, indicate that each corpus of summaries uses this word in the same way: as an indicator of the protagonist's age (see figure 5.1). Unsurprisingly, given that this is a corpus of summaries describing children's literature, a child's age always modifies "old." With the exception of "seven-year," all of the observed collocated modifiers for "old" indicate tween or early teen ages, with age of protagonist correlating with the observed reading levels identified by the Accelerated Reader and Metametrics databases. The Newbery Medal largely consists of books for and about 10 to 14-year-olds.

Table 5.1: Most frequent words in summaries from WorldCat, Accelerated Reader, and Goodreads.

Summary: WorldCat	**Summary: Accelerated Reader**	**Summary: Goodreads**
Total words: 3,779 **Unique word forms**: 1,449	**Total words**: 2,356 **Unique word forms**: 1,081	**Total words**: 10, 698 **Unique word forms**: 3,028
Most frequent words: old (30); year (27); life (19); boy (17); new (16); girl (15); family (13); father (13); home (11); story (11); young (11); england (9); adventures (8); becomes (8); team (8); world (8); years (8); century (7); indian (7); man (7); mother (7); summer (7); become (6); children (6); death (6); finds (6); good (6); know (6); long (6); love (6); things (6); village (6); away (5); city (5); courage (5)	**Most frequent words**: old (25); year (22); life (15); book (14); boy (13); family (13); new (12); father (11); home (10); mother (10); young (10); girl (9); story (8); adventures (7); becomes (6); century (6); indian (6); man (6); city (5); courage (5); friends (5); war (5); begins (4); bring (4); collection (4); comes (4); england (4); great (4); learns (4); left (4); town (4); village (4); world (4); years (4); york (4)	**Most frequent words**: story (41); life (40); new (31); like (28); old (28); newbery (27); year (26); family (25); medal (25); father (24); boy (23); world (19); just (18); young (18); author (16); novel (16); way (16); book (15); sea (15); girl (14); mother (14); day (13); home (13); winner (13); away (12); comes (12); it's (12); man (12); tale (12); town (12); winning (12); children (11); come (11); dog (11); friends (11)

Similarly, contextual information about other frequently used words correlates to findings suggested by descriptive and structural information. The frequent usage of the words "story" and "tale," for example, triangulates with findings that the corpus favors fiction. Definite articles

most frequently precede “story,” followed by adjectives such as “powerful,” “gripping,” and “legendary.” Similarly, “adventures” finds a spot on the WorldCat and Accelerated Reader lists, as do terms suggesting the types of adventures described, such as “courage,” “war,” “sea,” and “world.” These words suggest a corpus concerned with telling exciting, adventurous stories. All three platforms list “boy” and “girl” as frequently used words, with “boy” occurring slightly more frequently than “girl” in all databases (see table 5.1). This is unsurprising given the finding that the corpus includes slightly, but not statistically significantly, more stories about boys than girls. These points of triangulation provide checks and balances for distant reading, offering corroboration of assumptions made from other data sources.

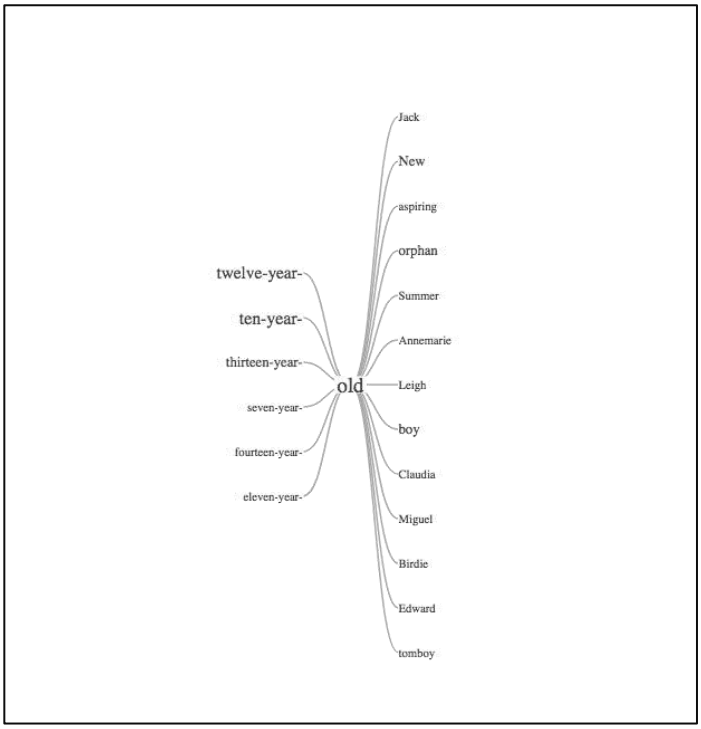


Figure 5.1: Word tree illustrating context for frequently used word “old” in WorldCat summaries. Font size in the illustration correlates to word frequency, with more frequently used terms appearing larger.

Frequently-used words in book summaries also suggest thematic elements in the corpus not implied by descriptive and structural variables. Words describing family units, including "father," "mother," "family," and "home," point to a corpus largely concerned with family. Contextual word trees indicate that happy families do not dominate the corpus, nor do families universally provide a sense of safety and stability. Modifiers such as "motherless," "struggle(s)," "away," "go," and "share" occur in tandem with "family." "Home," meanwhile, co-occurs with simple adjectival modifiers like "prairie" or "Virginia" as well as descriptors that suggest the home as locus for conflict or change, such as "foster," "permanent," "new," and "leaving" (see figure 5.2). The thematic pictured painted by modifiers describing family and home life begin to suggest a corpus with a sizable emphasis on struggle, change, and personal challenges.

Three additional terms deserve consideration in conversation with ideas raised by the use and function of family in the corpus descriptions: "death," "love," and "becomes." Death and love suggest a continued concern across the corpus with close, interpersonal relationships, particularly relationships that change or draw to a close. "Becomes," the only verb to occur as a most frequent word on all three platforms, suggests stories of catalyst and transformation, not stasis. Taken as a composite, the most frequently used words describing people and their actions in the corpus indicate a preference for privileging stories about boys and girls on the cusp of adolescence experiencing challenging changes in their family life. It is important to note, however, that these are simply the most frequently used words. The frequencies are not necessarily statistically significant, nor do these words occur in descriptions for every title represented in the corpus. Instead, they provide a general impression of adventure, change, family, and emotion.

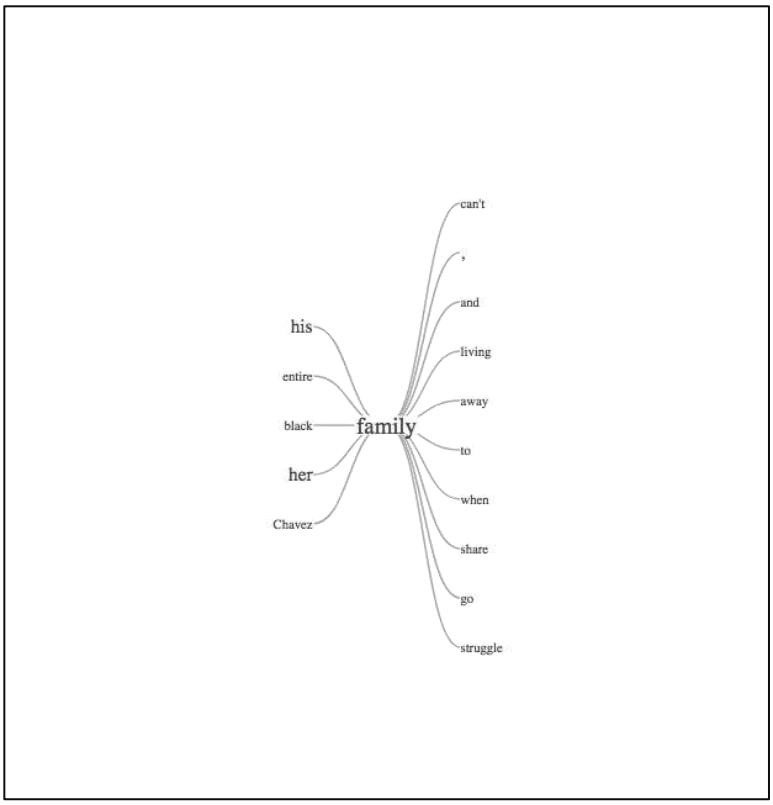


Figure 5.2: Word tree illustrating contextual modifiers for the term "family" in WorldCat. Font size in the illustration correlates to word frequency, with more frequently used terms appearing larger.

What's This Corpus About?: Using Controlled Vocabulary to Analyze Theme

Summaries on WorldCat, Accelerated Reader, and Goodreads all exist to describe the individual text in brief and help readers decide if they want to read that particular book. As such, it is unsurprising that the specific words used to describe the Newbery Medal winners contain so much variation. Subject headings and tags serve a complementary purpose to book descriptions. Headings and tags describe a specific book, but they do so using a standard set of terms. These terms, in turn, exist to help users find other books that are similar to the text in hand. In this section, I consider how the controlled vocabularies applied in WorldCat and Accelerated Reader and the folksonomies of user tags applied in Goodreads enable a distant reading analysis of theme in the Newbery Medal corpus.

Thematic controlled vocabulary in WorldCat. Despite the use of a controlled vocabulary, Library of Congress Subject Headings still utilize a wide range of terms to describe the Newbery corpus, many of which are unique or infrequently applied to more than one text. Only 20 individual words occur across all subject headings more than five times (see table 5.2). Examining collocation of terms associated with these frequently used words provides contextual information for how these headings function within the corpus, thereby providing insight into which terms appear more frequently with other terms across the entire corpus. The most frequently used word, "life" almost always appear in the context of "conduct," reflective of the heading "conduct of life." The scope note for this heading notes that it pertains to "works on standards of behavior and works containing moral guidance and advice to the individual" (Conduct of life, n.d.). Taken in tandem with the finding of a corpus favoring fiction, this heading further suggests a corpus largely concerned with stories designed in some way to reflect a "good" life. This further triangulates with previous research pointing to the moralistic, didactic tone found in many Newbery Medal winners (Cook, 1985; Kidd, 2007), despite the Medal criteria's assertion that the award is for literary merit, not didactic content.

The additional frequently used terms found in subject headings, and the terms collocated with them, provide interesting insights into the thematic elements of the corpus that are not suggested by previous research. Of particular interest given the emerging motif in descriptive summaries of struggle, strife, and potentially unhappy families is the prevalence of death and the terms collocated with death in the corpus. "Death" is most likely to appear as a heading in conjunction with the terms friendship, children, sisters, and prejudices. Death, then, appears not to be an abstract concept or something that happens to other people in Newbery Medal winners.

Instead, death is something intimate and experienced, something most likely to occur to a close loved one, a sibling, or a friend.

Table 5.2: Top 20 terms used in Library of Congress Subject Headings describing the Newbery Medal corpus.

Term	Frequency	Collocated Terms
life	33	conduct
children	15	life, friendship, death, conduct
friendship	13	death, children
sisters	10	brothers, life, families
americans	9	depressions, runaways, life, death, race, identity
brothers	9	sisters, life, families, family, african
families	9	life, brothers, sisters, friendship
family	9	life, sisters, children, brothers
african	8	americans, depressions, runaways, life, families, death, brothers
depressions	8	1929, runaway, voyages, travels,
ages	7	orphans, children
animals	7	treatment, circus, welfare
death	7	friendship, children, sisters, prejudices
middle	7	ages, orphans
conduct	6	life, children
country	6	life
identity	6	psychology, philosophical, concept
orphans	6	middle, identity, ages
runaways	6	vaughan, brian, african
survival	6	wolves, teenage, rifles, inuit, girls, courage, daniel, boone

"Death" also appears as a collocated term for "African," a term which itself is most frequently collocated with "Americans." This suggests a tendency for the Newbery to privilege stories about African Americans where death features prominently. Other significant phrases collocated with "African" include "depressions," "runaways," "life," and "families." This is not an entirely affirmative set of terms, nor do these terms point to a full or measured consideration

of race in Newbery Medal-winning titles. Instead, it suggests that the stories about African American characters prized by the Newbery are problem novels, or stories that traditionally have been read as privileging depictions of social problems over narrative complexity (Russell, 2005). Further, many of these problems novels are collocated around the term "depression," which itself collocates strongly with "1929." Coupled with knowledge that the corpus favors historical fiction, the collocation of these terms suggests that the Newbery Medal favors historical fiction about African Americans during the Great Depression. When these terms occur in the corpus is also significant: most uses of the term "African Americans" as a subject heading occur in the 1970s. From these terms, distant reading suggests that the Newbery favors historical fiction problem novels about race relations at least a generation removed from the contemporary issues associated with civil rights movements.

Terms found in subject headings also introduce new thematic motifs not found through distant readings of descriptions or descriptive and structural metadata. The most prominent of these terms is "animals," which collocates with "treatment," "circus," and "welfare." As with families, the animals found in Newbery Medal winners appear far from happy. Relying on just the terms "animal" and "animals," however, underestimates the representation of non-humans in the corpus. A large number of single use heading terms describe specific animals, including but not limited to: squirrels, rats, mice, cats, dogs, horses, cows, tiger, wolves, sheep, stork, moles, skunks, woodchuck, foxes, microtus, and pigeons. Subject headings name specific animal species 21 times in the corpus, in addition to the more general "animal" and "animals." These specific species do not collocate with "treatment" and "welfare" to the extent that "animal" does, indicating a variety of approaches to representation of animals in the corpus. Taken together, the specific and general animal headings outnumber all other terms in the corpus except for "life."

Given this prevalence, the Newbery's depiction of and preference for animals remains surprisingly un-discussed in existing examinations of Medal winners. The representation, depiction, and role played by animals in the corpus warrants further, microscopic research.

Terms found in Library of Congress subject headings provide additional, albeit limited, insights into the types of historical fiction privileged by the corpus. "Orphans" collocates with "middle" and "ages," pointing to a predominance of historical fiction about medieval era orphans. Similarly, "runaways" typically occur with "depressions" or "african," pinpointing depictions of Depression-era runaway African American children. The terms do not, however, provide much, if any, insight into the themes and motifs found in the fantastical works in the corpus. None of the most frequently used terms in the subject headings applied to Newbery Medal winners suggests the non-realistic. Less frequent terms found in the headings include "legends" (four occurrences), folklore (three occurrences), and fantasy (three occurrences). The level of specificity offered in these headings does not approach that found in headings describing the real world, suggesting either that headings do not adequately describe the fantastical or that the terms applied are so specific that they do not apply to more than one text in the corpus. Sampling individual records in the data set suggests that both problems exist. Texts described as science fiction or fantasy in the corpus use incredibly specific headings, like "extraterrestrial beings," as well as headings for individual characters, such as "Taran – fictional character," in concert with more general headings that do not particularly describe the theme of work, like "supernatural" and "folklore."

Thematic controlled vocabulary in Accelerated Reader. Accelerated Reader does not utilize quite the range of thematic headings that WorldCat does to describe the corpus, but the headings it does use are surprisingly informative for the purposes of distant reading. As with

genre headings, discussed in chapter 4, Accelerated Reader appends a wide variety of thematic subheadings to a more limited number of main headings to granularity in description. This granularity provides specificity, but it also results in 107 unique terms to describe the Newbery corpus. Of these 107 terms, only three apply to more than five works: family life – death (6 times), family life – growing up (8 times), and interpersonal relationships – friendship (9 times). With subheadings deleted and only the main heading considered, 16 general terms occur across the corpus at least twice, although only four occur more than 10 times (see table 5.3).

Table 5.3: Most frequent thematic headings describing Newbery Medal winners in Accelerated Reader.

Accelerated Reader Thematic Term (Simplified)	**Frequency**
Family Life	42
People	17
Animals	14
Interpersonal Relationships	11
Emotions	7
Community Life	5
Social Issues	5
Careers	4
Wars	4
Middle Ages / Medieval	3
Disabilities	2
Disasters	2
Magic	2
Natural Environments	2
Painting	2
Sports/Recreation	2

The simplified Accelerated Reader headings provide remarkable congruence with terms used in Library of Congress Subject Headings. Once again, these headings suggest that families and family life play a prominent role in the corpus. The collocated terms for “family life” in

Accelerated Reader speak to the range of family relationships represented in the corpus in a way that WorldCat's Library of Congress Subject Headings do not: adoption, aunts, birth, brothers, fathers, grandparents, mothers, orphans, pets, sisters, sons, and stepfamilies. This list is much broader than WorldCat's focus on sisters and brothers. In addition to describing family units, collocated terms also indicate some of the experiences that families encounter: death, coming of age, growing up, growing old, and moving to a new area. Once again, these terms do not suggest an entirely happy corpus. Instead, they speak to the challenges associated with an array of family experiences.

"Love" and "fear," meanwhile, collocate with "emotions," as do "survival," "away," "behavior-meanness," and "people-slaves" (see figure 5.3). These are powerful emotions, and again, not always positive. "Behavior-meanness" and "survival" particularly, coupled with the prevalence of fiction in the corpus, suggest the probability of problem novels featuring at least somewhat prominently. As with WorldCat's use of Library of Congress Subject Headings, Accelerated Reader headings do not provide much in the way of clarification for the thematic content found in fantastical texts. "Dragons" (one occurrence) and "witches/warlocks" (two occurrences) offer the most thematic information for fantastical elements in the corpus, although their frequencies are so low that they are relatively meaningless on a macroscopic scale.

Thematic folksonomies in Goodreads. Despite encouraging users to tag books with genre tags, users also apply limited thematic tags in the Goodreads database. Even with a generous definition of what constitutes a thematic tag, including "family," which could indicate a genre tag for a book suitable for family reading, thematic tags are not nearly as extensive in Goodreads as they are in the WorldCat or Accelerated Reader databases. The ones that do exist, however, are illuminating in what they highlight and the points of synergy they provide with

other descriptive and thematic sources. Sixty-five Newbery Medal winners have at least one thematic tag in the Goodreads database, but only six tags occur more than five times (see table 5.4). Many of these tags, particularly the most frequently used, correlate with tags also observed in WorldCat and Accelerated Reader. Once again, "families" and "animals" top the list. Unlike WorldCat and Accelerated Reader, Goodreads does not employ subheadings, and as most titles have only one thematic tag, collocation of terms is largely not possible for individual tags. Using Goodreads alone, it would be impossible to know the range of families represented in the corpus or the emotional range of experiences that these families encounter. Similarly, Goodreads tags do not provide an indication of the range of animals represented in the corpus or the types of stories associated with animals. Instead, Goodreads tags simply point to the prominence of families and animals in the corpus.

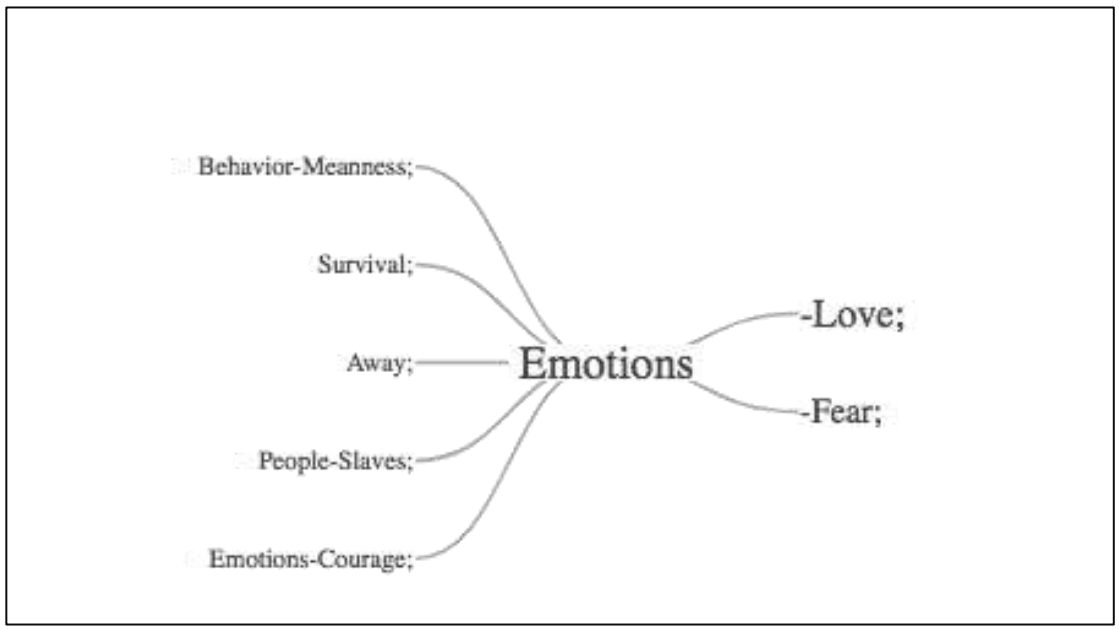


Figure 5.3: Word tree displaying terms collocated with "emotions" in Accelerated Reader thematic headings. Font size in the illustration correlates to word frequency, with more frequently used terms appearing larger.

One term among the observed Goodreads thematic tags, however, frequently does occur with subheading modifiers: "cultural." Users have identified five different subheadings to modify "cultural" in the Newbery corpus: African American, Asia, Bulgaria, China, and Spain. Although this practice of labeling books about cultures other than mainstream, middle class white America as "cultural—[specific culture modifier]" does help users find books about other cultures and races in the all too white world of children's literature, it nevertheless presupposes that books about white children and white families do not constitute a culture. Instead, Goodreads users apparently identify white culture as a default, going unremarked and untagged.

Table 5.4: Most frequent thematic headings describing Newbery Medal winners in Goodreads.

Goodreads Thematic Tag	Frequency
Animals	28
Cultural	21
Family	17
War	8
African	7
American	6

Where and When Does This Corpus Take Place? Using Controlled Vocabulary to Analyze Setting

Existing criticism on setting in the Newbery Medal focuses on the preponderance of farms, rural settings, and exotic realms of the past (Alberghene, 1981; Solt, 1981; Kidd, 2005). Word frequencies from book description summaries suggest, however, that this supposition warrants closer scrutiny. "City," "village," and "town" occur with similar frequencies across all three platforms, and contextual usage indicates that "new" frequently modifies "york city," indicating a metropolitan setting. England is the only other frequently used geographic place in the summaries. Although certainly not domestic terrain for the Newbery Medal, England is also

not the exotic setting suggested by earlier studies on place in the corpus. In this section, I consider the extent to which the controlled vocabularies employed in WorldCat and Accelerated Reader help add nuance to understandings of setting in the corpus. I also consider the ways in which information from book descriptions helps mitigate any flaws in the controlled vocabularies. As Goodreads users have not applied geographic tags to Newbery Medal-winning titles beyond those associated with the thematic tag "cultural," Goodreads data does not provide additional insights to geographic setting.

Setting-related controlled vocabulary in WorldCat. As with all other controlled headings in the corpus, repetition of precise geographic or time period terms used in Library of Congress Subject Headings to describe different books is rare. WorldCat records geographic or time period subheadings for 61 titles. Somewhat surprisingly, given existing critical understandings of the corpus as one that favors the rural or the exotic, headings for New York City and England occur the most frequently (see table 5.5), although this frequency is not significant in the statistical sense of the term. No time period subheadings exist for New York City, but those provided for England and Great Britain indicate a strong preference for the England of the Middle Ages. Also of interest are the three headings for "United States – History." Periodized subheadings indicate the range of U.S. history covered in these titles: the Revolutionary War (1775-1783), the French and Indian War (1754-1763), and the Civil War (1861-1865).

Single and two-use terms help locate 22 specific states within the United States where a text was set. This full list speaks to a range of geographic settings in the corpus: Alaska, California, Connecticut, Florida, Georgia, Idaho, Illinois, Kansas, Maryland, Massachusetts, Michigan, Mississippi, New Hampshire, New Mexico, New York, Ohio, Oklahoma,

Pennsylvania, Virginia, West Virginia, and Wisconsin. The East coast dominates this list, with a few representative samples from the Midwest peppered throughout. Outside of California and Alaska, the West coast appears to have very little representation in Newbery Medal winners. When time period information appears in conjunction with geographic headings describing a United States setting, it is almost always "History—20th century." This lack of specificity adds very little to an understanding of setting in the corpus.

Table 5.5: Most frequent geographic subheadings describing Newbery Medal winners in WorldCat.

Headings	Frequency	Collocated Time Periods
New York (N.Y.)	5	None provided
England	4	Medieval; Middle Ages
Great Britain	4	1066-1485; 1327-1377

Setting-related controlled vocabulary in Accelerated Reader. Accelerated Reader's oddly formulated headings utilize a structure that aids distant reading because the headings enable sorting geographically and locating domestic versus international settings. All titles described as set in the United States use a heading that begins with "U.S. States/Regions," while all titles described as set in a region outside of the United States use a heading that begins with "Countries/Regions." Unfortunately, headings for neither U.S. States nor other countries include additional information on time period. When it comes to texts about places outside of the United States, Accelerated Reader locates Newbery Medal winners all over the map, ranging from China to the Netherlands, Israel, and Mexico (see table 5.6). Assuming that Accelerated Reader consistently applies geographic headings to texts, the date range accompanying headings beginning "Countries-Regions" indicates a steady supply of texts in the corpus representing foreign locations until 1962, followed by a sudden discontinuation of foreign settings. The

headings describing the United States also indicate a sudden, sharp decline of domestic settings at roughly the same time, (see table 5.7), with a resurgence in the 1990s. As the genre headings analyzed in chapter 4 indicated a preponderance of real world settings in historical and realistic fiction, it is more likely that the metadata Accelerated Reader applied to titles published between 1960 and the mid-1980s simply lacks geographic headings.

Table 5.6: Accelerated Reader headings describing location of texts occurring outside of the United States.

Date	Region
1925	Countries/Regions-Central America
1926	Countries/Regions-China
1929	Countries/Regions-Poland
1931	Countries/Regions-Japan
1933	Countries/Regions-China
1943	Countries/Regions-England
1950	Countries/Regions-England
1953	Countries/Regions-Peru
1954	Countries/Regions-Mexico
1955	Countries/Regions-Netherlands
1962	Countries/Regions-Israel
1996	Countries/Regions-England
2002	Countries/Regions-Korea, North and South
2008	Countries/Regions-England

In this instance, manual coding of the descriptions proves much more effective for deriving information about setting than a reliance on controlled vocabulary alone. Coding reveals the very limited range of cities covered in the corpus. Although eight works take place in a city, four of these eight take place in New York City, with the remaining four set in Chicago, Flint, Boston, and an unnamed "bustling city." Although Flint, Michigan, does not carry the same metropolitan connotations as New York and Chicago, it is, nevertheless, the largest city in Michigan and an urban center. Summaries identify no West coast cities or cities outside of the

United States. Coding also suggests that previous understandings of the preponderance of American small towns in the corpus are slightly misleading. Small towns do certainly exist, but villages, farms, the prairie, as well as the simple descriptor "rural" outnumber towns. These terms suggest that the Newbery does not simply privilege depictions of small town life. Instead, it appears to privilege iconoclastic, archetypal stories of frontier and farm life, reminiscent of what Fellman terms a "guiding American mythology" built on veneration of idealized pioneer roots that exists to shape social and governmental policies (1996, p. 101).

Table 5.7: Accelerated Reader headings describing location of texts set in the United States.

Date	State
1932	U.S. States/Regions-Arizona
1936	U.S. States/Regions-Wisconsin
1939	U.S. States/Regions-Wisconsin
1946	U.S. States/Regions-Florida
1977	U.S. States/Regions-Mississippi
1992	U.S. States/Regions-West Virginia
1993	U.S. States/Regions-West Virginia
1998	U.S. States/Regions-Illinois
2001	U.S. States/Regions-Illinois
2005	U.S. States/Regions-Georgia
2011	U.S. States/Regions-Kansas
2012	U.S. States/Regions-Pennsylvania

Glimpses of this American mythology are also viewable in the number of descriptions suggesting rural settings and movement. Many descriptions use terms, such as "West Virginia trailer" and "prairie home," that suggest the characters live in smaller, single family dwellings. Other terms suggest that a character moves from the city to the mythologized setting of the country, the farm, or small town. Characters, for example, leave their Chicago home for rural Illinois, or move from an unnamed city to a farmhouse. Summaries do not describe movement in

the opposite direction. Instead, the corpus appears to favor texts that feature movement towards the rural as a catalyst for the text.

Table 5.8: Comparison of geographic coverage in controlled vocabularies and book summaries.

Headings for Locations in the United States			**Headings for Locations Outside of the United States**		
Library of Congress	**Accelerated-Reader**	**Book Summaries**	**Library of Congress**	**Accelerated-Reader**	**Book Summaries**
Alaska	Arizona	Alaska	China	Asia	Asia
California	Florida	California	Denmark	Central America	Bulgaria
Connecticut	Georgia	Florida	Great Britain	China	Caribbean
Florida	Illinois	Georgia	Hungary	England	Central America
Georgia	Kansas	Illinois	India	Europe	China
Idaho	Mississippi	Kansas	Japan	Israel	Denmark
Illinois	Pennsylvania	Kentucky	Korea	Japan	England
Illinois	West Virginia	Massachusetts	Krakatoa (Indonesia)	Korea, North and South	Japan
Kansas	Wisconsin	Michigan	Palestine	Mexico	Korea
Maryland		Mississippi	Peru	Netherlands	Krakow, Poland
Massachusetts		New York	Poland	Peru	Pacific Island
Michigan		Ohio	Polynesia	Poland	Peru
Mississippi		Oklahoma	South America		Polynesia
New Hampshire		Pennsylvania	Spain		South America
New Mexico		Texas	Wales		Spain
New York		West Virginia			Tropical Seas
Ohio		Wisconsin			
Oklahoma					
Pennsylvania					
Virginia					
West Virginia					
Wisconsin					

Movement does not only occur in the context of the rural. It also features prominently in descriptions of texts set outside of the United States as well as texts with an indeterminate geographic setting. Representative examples include “migration from Asia to Europe,” “leave behind the shimmering Caribbean islands,” “Africa bound ship,” “vacation in a balloon, and “voyage over tropical seas.” As with the terms suggesting movement for books set in the United States, these terms once again typically describe the catalyst of action for the book, setting the characters in motion and providing the impetus for the ensuing story. They also suggest that place and setting are somewhat fluid in the corpus, with considerable transition occurring from point of inception to conclusion.

Analyzing Thematic Elements in the Newbery from a Holistic Perspective

The findings presented in this chapter provide insights into the thematic elements present in texts that the Newbery Medal privileges. Taken holistically, these variables enable a consideration of the themes, motifs, subjects, and settings found in the corpus. The distant reading techniques employed here use broad strokes to paint the picture of thematic motifs in the corpus, suggesting areas where further, microscopic scrutiny may be of use. These findings complement and augment existing critical stories about the types of texts that the Newbery typically favors. Frequently used words help generate topic models of thematic elements the corpus. Favored topics include particularly challenging elements in family and home life, especially death and change, as experienced by children ages 10-14. Models of thematic elements also underscore long-held assertions about representations of race in the Newbery Medal. The corpus relies on particularly problematic, reductive, and restrictive representations and descriptions of race. Depictions of African Americans in the corpus are largely reduced to historical representations from the Depression Era, written at the remove of at least a generation.

Further, descriptions only mention race or ethnicity if they describe something other than white America.

Distant reading also suggests new avenues for understanding thematic elements in the corpus in addition to adding nuance to existing critical frameworks. In uncovering the proliferation of animals in the corpus, the distant reading techniques I employed reveal an area completely uncharted in previous studies of the Newbery. Given the significance of animals as characters in children's literature in general (see, for example, Nodelman, 2008; Zipes, 2013) and the current critical focus on questions of human and non-human representation in particular (Nikolajeva, 2016), this lacuna in the scholarship warrants further, microscopic consideration. Distant reading proved singularly unhelpful for analyzing the presence of the fantastic in the corpus, beyond a reminder of its existence. Once again, given the significance of the fantastic in children's literature in general (see Levy and Mendlesohn, 2017) and dominant discourses privileging realism and the historic in the Newbery, microscopic consideration of fantastical elements in the corpus and the role they play in shaping the corpus of privileged children's literature should be considered.

The findings presented here also point to the need for further consideration of setting in the corpus. Traditionally, the story told about the Newbery Medal is a tale of small towns and exotic locals. These elements certainly exist in the corpus, but the descriptors used to depict setting suggest that the conversation could benefit from more nuance as well as attention to texts that do not fit the expected mold. The privileging of texts about New York City and medieval England alone point to unexplored settings, as do potential differences in representations of towns, villages, farms, and pioneer life and implications for mythologizing the American frontier in the corpus.

The analysis of theme also underscores serious limitations inherent in the metadata used to facilitate this study. The metadata covering geographic and chronological headings in particular proved severely lacking in both specificity and consistency. Without other data points to serve as a corrective, for example, Accelerated Reader's geographic headings would suggest that geography played little role in setting the stage for Newbery Medal winners after the mid 1960s. Other sources, however, reveal that this gap points instead to the weakness of Accelerated Reader's application of controlled vocabulary. Similarly, the vocabularies describing setting, both geographic and chronological, are wildly inconsistent, especially in terms of levels of specificity. Headings range from encompassing entire continents to specific, imaginary cities. All too often, headings describing chronological setting are unhelpfully vague and do not provide enough information to enable anything resembling a considered analysis. It is critical for distant reading to rely on more than one data source.

Chapter 6: On Popularity, Sales, and Circulation

In this chapter, I consider the long-held truism that winning the Newbery Medal is a game-changer for books, essentially offering the winning text a ticket to enduring, intergenerational sales and popularity. A visit to the children's section in a bookstore or library supports this assertion. Newbery stickers positively pepper the front covers of many titles in stock, and some stores even have separate shelves to highlight award winners, where Newbery Medal-winners feature prominently. Bookseller Robert Hale describes how the sales bump provided by the Newbery sticker on the front of a book leads to increased sales, which more often than not lead to a book becoming part of a store's permanent stock (Hale, 1995, p. 364). Accelerated Reader's Bookfinder database offers the ability to browse a list of Newbery Medal-winning and honor titles as a way of selecting a book to read, and they also tag each constituent title as a Newbery Medal winner. Similarly, a number of libraries insert notes and headings into bibliographic records to mark the Medal-winning status of the title. From a purely practical standpoint, organizing databases, library catalogs, libraries, and bookstores in this way suggests that enough patrons have asked for Newbery Medal-winning titles that this arrangement makes sense. This activity presupposes users want to be able to find Newbery Medal-winning titles, indicating that they may want to identify and select a text based on its Medal-winning status rather than its content or author alone.

On the individual, microscopic level, it is easy to pinpoint the effect of the Newbery Medal on a book's reception, sales, and market penetration. Consider *Moon Over Manifest*, winner of the 2011 Newbery Medal. This historical fiction novel was largely ignored upon its

first publication. *The Horn Book Magazine*, a major review source for what the industry considers the best of children's literature, did not even review it upon its initial publication. After it won the Newbery, *Moon Over Manifest*'s modest success surged dramatically. *The Horn Book Magazine* rapidly reviewed it, and the novel even vaulted onto the New York Times children's chapter book bestseller list for January 30, 2011. In 2017, six years after winning the Medal, the novel remains comfortably popular, with 11 different editions currently in print and on the shelves of over 4,000 different library systems.

Literature about the popularity of Newbery Medalists relies upon data similar to that presented above: anecdotal reports of individual booksellers, trips to bookstores and libraries to assess the presentation of Medal-winning texts, and more granular consideration of reception history for individual books. Studies on popularity of the Newbery Medal as a whole do not consider changes in popularity over time nor the relationship between genre, theme, and popularity. Using data from Books in Print, WorldCat, and Goodreads, I considered methods for analyzing popularity across the entire corpus as well as how popularity intersects with genre. Sales data is certainly one element of popularity: books that do not sell do not stay in print. Similarly, outside of research libraries, books that do not circulate frequently do not remain in library collections. Sales data, however, do not provide a complete picture of popularity. Does the sticker on the front of the book translate to books that are actually read, or to books that people think ought to be read? Is the entire corpus popular, or are the well-known, frequently discussed winners popular, with the remaining titles boosted by the Newbery sticker on their covers, essentially riding on the coattails of the other, more popular winners?

Circulating the Newbery

As of 2017, Books in Print lists all 96 Newbery Medal-winning titles as in print. This figure is slightly misleading, however, as one title, *Daniel Boone* (1940), is available only through a homeschool curriculum company's print-on-demand service. Despite the slight padding offered by print-on-demand publishers, the figure is nevertheless remarkable. As a now-classic and still influential text on collection development notes, within 10 years, less than half of a single year's publications remain in print (Katz, 1980). Kidd (2007) notes that the typical children's book remains in circulation for a far shorter time, averaging approximately 18 months. Given these figures, the Newbery Medal significantly alters the long-term availability of its winners within the field of children's literature. In addition to remaining in print at all, Medal winners also typically have a number of editions that remain in print. *Gay Neck, the Story of a Pigeon* (1928), *Dobry* (1935), and *Daniel Boone* (1940) are the only Medal winners with only a single edition remaining in print in 2017. The first Newbery Medal winner, *The Story of Mankind* (1922), takes the distinction of having the most editions in print, with 103, followed closely by *The Voyages of Doctor Dolittle*'s (1923) 98 editions. It is worth noting, however, that these two texts represent the only titles in the corpus currently in the public domain. Therefore, they also represent the two texts that are the easiest and cheapest to republish. The mean number of editions in print for Newbery Medal winners is 22.94, with a standard deviation of 20.14. As with all other variables analyzed for this study, the standard deviation points to a high level of variance, suggesting significant difference across the corpus. The presence of two works in the public domain, *The Story of Mankind and The Voyages of Doctor Dolittle*, and the resultant proliferation of cheap editions of these texts, contributes to this variance, but it is not the only

contributing factor. The average number of editions in print for Medal winners between 1924 and 2017 is 21.29, with a standard deviation of 16.8.

Analyzing the distribution of editions remaining in print by decade indicates that although *The Story of Mankind* and *Daniel Boone* may be outliers at the opposite ends of the spectrum, Newbery Medal winners consistently remain in print in multiple editions (see figure 6.1). Unsurprisingly, given the amount of time it requires for new and different editions of the same text to proliferate, the most recent decade of Medal-winning titles averages a modest 9.5 editions. The reverse, however, is not necessarily true. Increased age does not correlate with an increase in popularity when that popularity measured by number of editions in print. A general bell curve is observable in average number of editions by decade from 1932 to the present, with the peak occurring in 1972-1981. Of the ten texts winning the Medal between 1972 and 1981, only two currently have fewer than 30 editions in print: *The Grey King* (1976, with 13 editions) and *A Gathering of Days: A New England Girl's Journal, 1830-32* (1980, with 29 editions). Further, the titles taking spots three and four on the list of texts from the entire corpus with the most editions in print are also from 1972-1981: *Bridge to Terabithia* (1978, with 70 editions) and *Roll of Thunder, Hear My Cry* (1977, 65 editions). This peak occurs after a slow growth from the Medal's least popular decade, when popularity is calculated by editions in print. The second decade of the Medal, 1932-1941, is the least popular in terms of number of editions remaining in print, with an average of 8.9.

Library holdings and editions owned by libraries, as captured in WorldCat, provide a complementary metric to print status for assessing popularity. Although library collections in the aggregate tend to skew towards newer, more recent titles, WorldCat data provides insight into editions beyond those currently available in print. Bibliographic records for editions remain in

WorldCat even after all libraries have deleted their holdings, indicating that no libraries worldwide have a copy of that particular edition in their collections but providing a record of its existence. In addition to providing a more holistic picture of editions published, holdings information affords insight into current density of representation of the corpus across multiple libraries by providing a metric to determine how many libraries currently have any edition of a specific title in their collections.

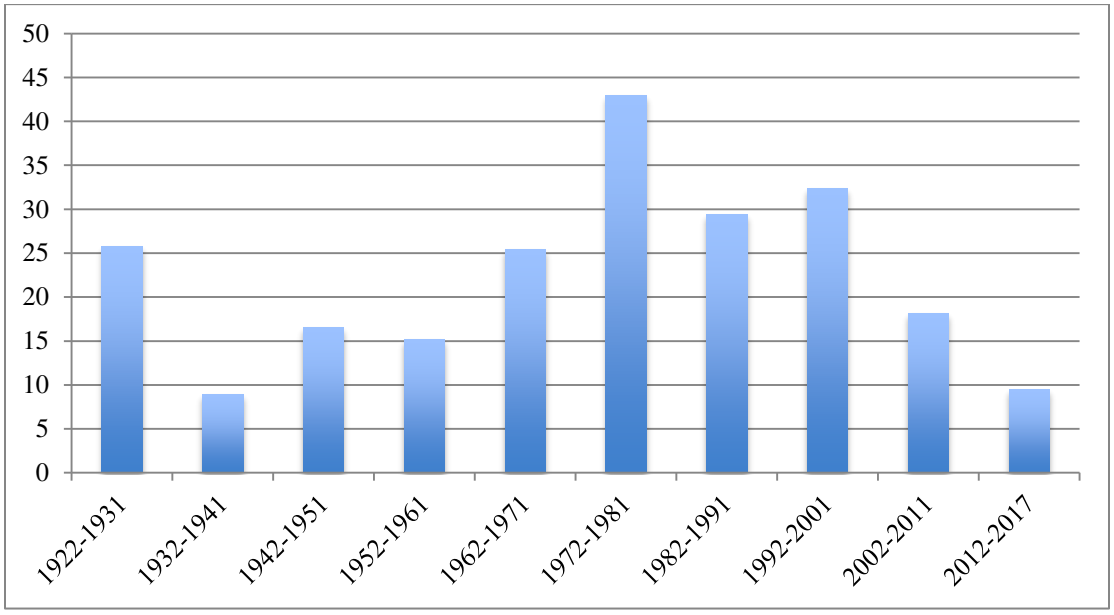


Figure 6.1: Average number of editions in print by decade.

WorldCat records a mean number of 71.98 different editions for the entire Newbery corpus, regardless of publication status, with a standard deviation of 51.4, as compared to the mean of 22.94 editions currently in print from Books in Print. The three most recently published Medal-winning titles, unsurprisingly, have the fewest number of aggregate editions regardless of print status. They simply have not had the time to accrue the publication history of other titles in the corpus. *The Story of Mankind* remains the most popular Newbery in terms of editions

published with 351 editions, followed by *A Wrinkle in Time*'s 221. The least popular titles, in terms of total number of editions published, vary from those determined by editions still in print. Aside from the 2015-2017 winners, the Newbery Medalist with the least number of editions published is *A Visit to William Blake's Inn* (1982), one of the few non-prose works to win the Medal, with 14 editions, followed by *The Witch of Blackbird Pond*'s (1959) 21 editions. *Daniel Boone* appears to have done much better in previous decades than its current status of one print-on-demand edition would suggest. Although ranking a relatively low eleventh with 25 editions published, it is nevertheless not at the bottom of the list. Indeed, WorldCat data indicates that all of the Newbery Medal-winning titles with only one or two editions currently in print performed substantially better at earlier points in time (see table 6.1). Comparing data from WorldCat and Books in Print suggests that popularity is actually a fluid metric, subject to changes over time.

Table 6.1: Comparison of number of editions in print and all editions published.

Title	Editions in Print	All Editions
Daniel Boone (1940)	1	25
Dobry (1935)	1	38
Gay Neck, the Story of a Pigeon (1928)	1	60
Invincible Louisa: The Story of the Author of Little Women	2	67
Waterless Mountain	2	62
Tales from Silver Lands	2	43

This pattern continues across the corpus. Editions in the aggregate do not follow the same general bell curve distribution as editions currently in print (see figure 6.2). Instead, the number

of editions in the aggregate follows a general negative trajectory, following a sudden spike in 1962-1971, aided in no small part by *A Wrinkle in Time*'s enormous number of republications. 1952-1961 proved an unremarkable decade for the Newbery, both in terms of total editions and editions remaining in print. Interestingly, this decade is replete with historical fiction, purportedly the genre that gives the Newbery its backbone.

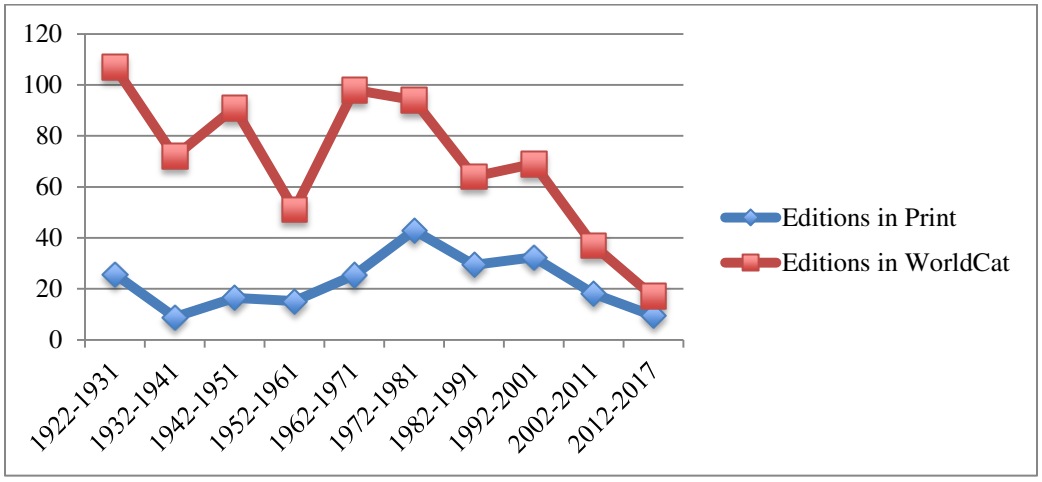


Figure 6.2: Number of editions in print compared to editions total.

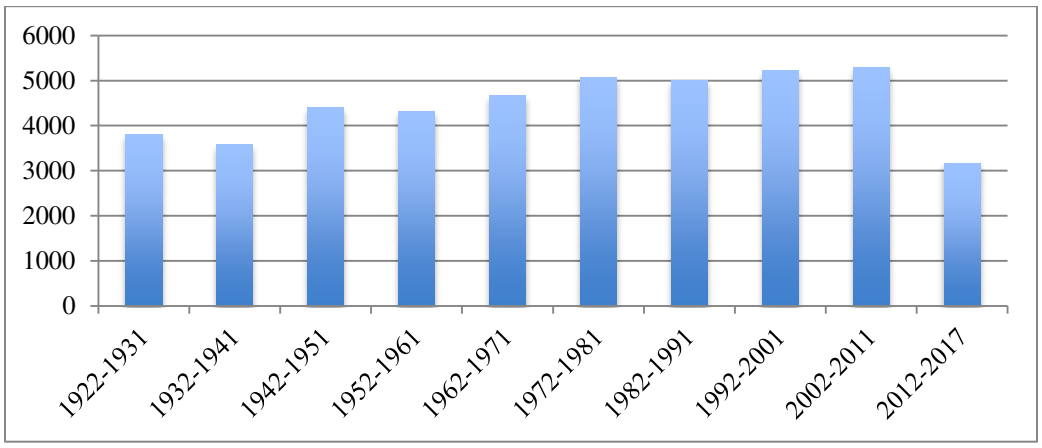


Figure 6.3: Mean number of WorldCat holdings by decade.

WorldCat holdings provide a different perspective on popularity, measuring representation of the Newbery in current library collections rather than the possibility for a library or a bookstore to purchase a book. As such, they represent books that have already been bought and offered to readers. The difference is subtle, but the observed values illustrate what this shift in perspective provides to a discussion of popularity. The Newbery corpus has a mean number of 4505.1 holdings in WorldCat, with a standard deviation 466.7. Unlike publication status or aggregate number of editions, analyzing holdings by decade indicates a slow increase, with an expected plunge in the current decade since it only includes six books rather than 10. This metric suggests that, even though older Newbery Medalists are more likely to be in print, newer winners are more likely to be in active library collections (see figure 6.3). Across the board, however, Newbery Medal-winning titles are well represented in library collections, and variation between decades is more modest than the variance observed in other measures of popularity.

The Newbery in the Wild

Publication status, print status, and inclusion in library collections provide information on popularity a step removed from readers. These metrics rely on readers, purchasers, and circulation for their existence, but they obscure reader perspectives behind sales and circulation data. Goodreads flips the equation, providing information directly from readers, although not from child readers as Goodreads terms of participation require users to be 13 years of age or older to register for an account. In addition, the metrics provided in Goodreads certainly depend on publication and print status or inclusion of titles in library collections. Readers cannot read and react to texts they cannot locate and read. Despite these limitations, Goodreads nevertheless provides information on what Nakamura calls “reading culture ‘in the wild’” (2013, p. 241).

Analyzing metrics from the platform enables a consideration of how readers, rather than librarians and publishers, measure the popularity of the Newbery Medal.

As of 2017, Goodreads users have rated and reviewed all 96 Newbery Medal-winning titles. They have done so in fairly large numbers as well. The Newbery corpus has an average of 65,523 ratings, with a standard deviation of 174,700. The average rating for the entire corpus is 3.80, with a standard deviation of 0.25.[8] *The Giver* (1994) has the most ratings of any Newbery winner with 1,333,938, while *Dobry* (1935) has the fewest with 802. *Dobry* is also the least popular title on Goodreads in terms of average rating, ranking 3.21 stars. *The High King* (1965) is the most popular, with an average rating of 4.26 stars.

The five most and least popular titles by number of ratings and average rating provide interesting points of comparison to measures of popularity found in publication history and library holdings (see Table 6.2). Less-popular titles, as calculated by both number of ratings and average rating, tend to be titles with fewer editions currently in print, with one notable exception. *The Story of Mankind* has the most editions currently in print and the most editions in WorldCat, but it is the third least popular Newbery Medalist in terms of average rating. Long-term availability does not always predict long-term popularity. The Newbery with the highest average rating, meanwhile, is *The High King*, which does not appear at the top of the list for any other popularity metric in this study. Also of note is the sheer number of ratings for *The Giver*. With over 1.3 million user ratings in Goodreads, *The Giver* has been rated more than twice as frequently as any other Newbery Medalist. As with *The High King*, *The Giver* does not appear at or near the top of any other list measuring popularity by other metrics. In addition to these

[8] As a point of comparison, consider *Harry Potter and the Sorcerer's Stone*, the most frequently rated titles on Goodreads. It has 4.9 million ratings, with an average score of 4.45. *Little House on the Prairie*, meanwhile, a text similar to Newbery winners in many ways, has 206,000 ratings, with an average score of 4.18.

standout individual titles, it is worth noting that less popular titles on Goodreads tend to be older medalists, while newer titles tend to do better, particularly in terms of average rating. Both the 2016 (*Last Stop on Market Street*) and 2017 (*The Girl Who Drank the Moon*) Medalists make the top five when the popularity is calculated by average rating.

Table 6.2: Least and most popular Newbery Medal winners on Goodreads platform.

Least Popular				**Most Popular**			
By Number of Ratings	*Number of Ratings*	*By Average Rating*	*Average Rating*	*By Number of Ratings*	*Number of Ratings*	*By Average Rating*	*Average Rating*
Dobry (1935)	802	*Dobry* (1935)	3.21	*The Giver* (1994)	1,333,938	*The High King* (1969)	4.26
Waterless Mountain (1932)	892	*Gay Neck, The Story of a Pigeon* (1928)	3.25	*Holes* (1999)	751,559	*Last Stop on Market Street* (2016)	4.24
Tales from Silver Lands (1925)	1106	*The Story of Mankind* (1922)	3.31	*A Wrinkle in Time* (1963)	634,553	*The One and Only Ivan* (2013)	4.23
Shen of the Sea (1926)	1409	*The Dark Frigate* (1924)	3.33	*Bridge to Terabithia* (1978)	348,616	*The Hero and the Crown* (1985)	4.21
Gay Neck, The Story of a Pigeon (1928)	1589	*Criss Cross* (2006)	3.34	*Number the Stars* (1990)	337,118	*The Girl Who Drank the Moon* (2017)	4.2

Popular Medal-winning titles on Goodreads are notable for the diverse range of genres represented. Two of the most popular books on Goodreads are works of science fiction, and three are fantasy. Two are general fiction, one is a picture book, and one a realistic-ish depiction of an elephant's life in captivity. Only one represents a work of historical fiction. The Newbery may privilege historical fiction over all, but currently, Goodreads users privilege other genres within the corpus. In fact, early examples of historical fiction in the corpus dominate the less popular end of the spectrum on Goodreads. In addition to being generally less available in print, these titles are also examples of historical fiction, frequently featuring representations of diverse cultures written by white men.

From a more macroscopic perspective, analyzing popularity metrics from Goodreads by decade provides additional points of contrast to metrics from WorldCat and Books in Print. The mean rating by decade (see figure 6.4) follows a generally positive trend line, indicating a preference on the part of Goodreads users for newer titles over older ones. This difference is particularly striking in the current decade as mean rating on Goodreads is the only popularity metric considered in the study that does not exhibit a sharp decline in the current decade. The average number of ratings by decade also behaves differently than other metrics describing popularity, with a small bell curve between the Medal's inception and 1982-1991, followed by an enormous spike in 1992-2001, aided in part by *The Giver*'s enormous popularity on the site. Three different measures of central tendency, mean, median, and mode, locate a significant spike at this decade, indicating a general preference among Goodreads users for Newbery Medal-winning titles published between 1992-2001. As with other popularity metrics that rely on accumulation over time, however, more recent decades witness a decline in overall number of ratings on the site.

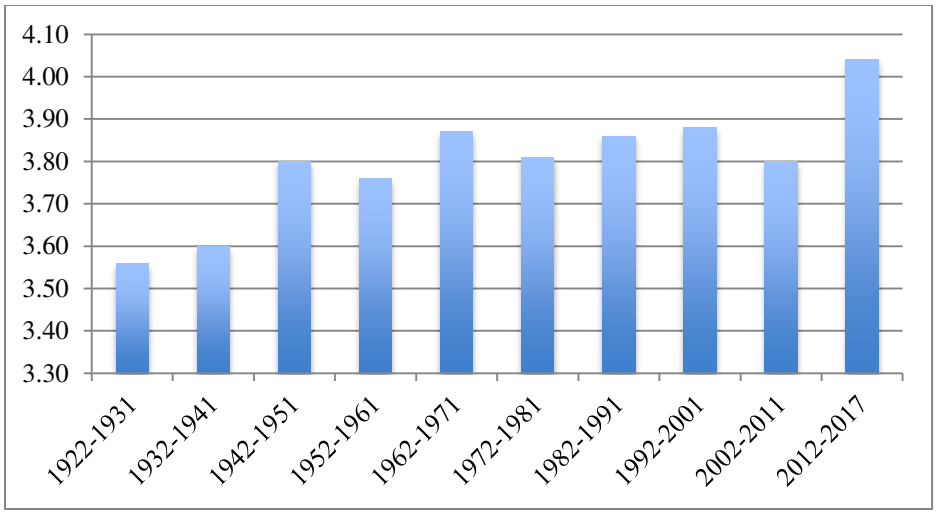


Figure 6.4: Mean ratings on Goodreads by decade.

Analyzing Popularity of the Newbery

The findings presented in this chapter provide insights into understandings of the Newbery Medal as a popular subset of children's literature. The Newbery's seal of approval undeniably has an effect on the long-term sustainability of a particular text, as evidenced by the fact that 100 percent of the corpus is currently available in print. This metric, however, relies on an overly broad stroke that does not consider fluctuations in the corpus or the effect that different metrics introduce into a conversation about popularity. All of the metrics that I consider include significant levels of variance, suggesting that not all Newbery Medal-winning titles are created equally. Further, no title performs equally well or poorly across all metrics. In order to create a more complete picture of popularity, it is necessary to consider more than one measure.

A more complete picture of the Newbery's popularity highlights changes over time and with metric. The Newbery titles most commonly found in print in 2017 are not a mirror for the entire publication history of the Medal. Instead, current print status reflects more recent reactions to the corpus, with certain texts and decades privileged more than others. Similarly, this more

complete picture of the Newbery highlights the need for skeptical reconsideration of claims that previous years' winners were better or more perennially popular than others (Devereaux, 2008; Silvey, 2008). The passage of time has cemented the place of certain winners from previous years as stand out extraordinary, but on the whole, previous decades do not always fare well with contemporary readers. On the other hand, WorldCat holdings indicate that some measures of popularity organically grow over time, offering a diminished view of recent developments. By the same token, popularity can also wane over time. Medal winners that enjoyed commercial success and many reprintings in the early decades of the Newbery now rest in near obscurity. All of the measures I considered in this chapter offer different answers to the question of what, precisely, constitutes the most popular Newbery Medal-winning title. Popularity is, in many ways, very much in the eye of the beholder.

It is also necessary to consider both the affordances and the limitations offered by each metric. Each of the metrics that I consider offers a different lens for considering popularity. Print status in 2017 suggests the long-term marketability for Newbery winners, while total number of editions published pinpoints titles that were popular closer to their original publication date but have lost marketability in the ensuing decades. Both of these metrics rely on highly abstracted readers, as print status is tied to book sales and library holdings are largely tied to circulation, but neither one truly captures reader reactions to the corpus. In contrast, data from Goodreads provides information from actual (assumed adult) readers in the aggregate, offering a snapshot of the Newbery's popularity from the vantage point of late 2017. Goodreads, in particular, provides tantalizingly different results, such as *The Giver*'s incredibly large number of ratings, without offering a mechanism for investigating the reasoning behind those differences. Goodreads provides six months' worth of change logs on a book's "stats" screen, but this limited history

does not provide the ability to analyze significant changes. Once again, the conclusions that may be reached via distant reading are only as good as the metadata that facilitates that reading.

Chapter 7: The Affordances of Interdisciplinary Paradigms

The story that I tell about the Newbery Medal is intentionally interdisciplinary, relying on data sources, methods, and critical approaches from library science, education, and literary studies. My professional and scholarly background, which includes experiences in and with all three disciplines, inspired this interdisciplinary perspective. It also offers a response to the unnecessarily stringent disciplinary boundaries that exist between the three disciplines, despite the fact that scholars across these three disciplines study children's literature in complementary ways. No one perspective provides a complete view of the Newbery Medal. Instead, each individual view complements others. Taken holistically, the composite views provide a more complete understanding of this designated text set.

In employing interdisciplinary perspectives and data sources to enable distant reading, my purpose was two-fold. First, I sought to explore the efficacy of distant reading strategies for a corpus of contemporary, copyright protected American children's literature. In order to enable this exploration, I applied distant reading strategies to the corpus of Newbery Medal-winning tests. As a result, my second purpose was to explore the Newbery Medal from a macroscopic perspective and to see what, if any, changes this distant perspective would lend to an understanding of the types of texts that the Newbery privileges. In doing so, I tested the hypothesis that scholarly understandings of the Newbery as a distinct subgenre of children's literature rely on an overly restrictive selection of texts. As such, my selection of the Newbery Medal as a test corpus was purposive (Maxwell, 2009) in that it provided me with a corpus about which much has already been written and against which I could test the findings exposed vis-à-

vis distant reading methods. In doing so, I uncovered a number of different ways in which methodological approaches to the study of children's literature affect conceptualizations of that literature. I also reaffirmed Moretti's (2005) argument that understandings of genre benefit from holistic, macroscopic approaches. I also join many of Moretti's critics, however, in reaffirming that distant reading augments microscopic, or close, reading practices. Distant reading, particularly distant reading enabled via secondary data sources, is imperfect. It does, however, highlight avenues for inquiry that would benefit from closer examination.

Distant Perspectives of the Newbery

In adopting a distant perspective to analyze the Newbery Medal, I focused on descriptive, structural, and thematic, and popularity variables to assess the formal characteristics of the corpus. I also used these variables to read the corpus in conversation with existing critical understandings of the Medal. In doing so, I found that the shape of the Newbery corpus, as seen from a distance, is familiar from existing critical understandings of the award, particularly if the existing understanding of the Medal is interdisciplinary in nature. The existing understanding of the Newbery Medal from the literary studies perspective, for example, is completely devoid of any consideration of text complexity. Distance provides a much sharper perspective in places and sheds light on blind spots in the familiar framework.

This study confirms familiar suppositions about the Newbery Medal. The Medal decidedly privileges narrative fiction for older child readers over other genres, and there is undoubtedly quite a lot of historical fiction in the mix. The Newbery exhibits a crucial lack of diverse voices and stories, and women do indeed currently outnumber men among the winners of the Medal. Further, the Newbery is a conservative corpus, witnessing stable text structures that change only slowly and in predictable ways over time. Traditional publishing houses and

imprints almost exclusively represent Newbery Medal-winning texts. Most studies (e.g., Bittner & Superle, 2016; Cummins, 2016;) do not quantify the extent of the problem when critiquing the Newbery's persistent problems in this area, relying instead on vague assertions of perceived oversights. This study provides measurable evidence for the ways in which the Newbery perpetuates a particular type of children's literature through the voices and types of texts that it privileges.

A distant perspective of the Newbery also highlights the limitations of previous stories told about the Medal and locates blind spots in existing models. In previous scholarship, there is a tendency to make sweeping assertions that do not actually reflect the entire corpus. Instead, these assertions reflect the portions of the corpus that scholars most frequently study. Gender distribution of Medal-winning authors provides an excellent case in point (Jenkins, 1996; Kidd, 2007). Women authors do currently outnumber men in the entire corpus, but they have not always done so, and they did not outnumber men when critics began decrying the prevalence of and preference for women's voices in the corpus. Telling only the story of the Newbery's preference for privileging women's voices misrepresents the entire corpus.

Despite privileging women authors at a statistically significant rate, and despite a long-held truism that the Newbery favors stories about girls (Pease, 1939), the Newbery corpus does not privilege stories about either boys or girls at a significant rate. Instead, focusing on the entire corpus reveals that the Newbery privileges stories about *individual* boys or girls over stories about groups or communal protagonists. This propensity towards privileging individualism is currently completely unstudied, as is the Medal's tendency to privilege stories about children's relationships with animals.

Distant reading techniques also shed light on the inclusion of different genres within the corpus. Traditionally, critical understandings of the Newbery have highlighted the privileging of historical fiction (Alberghene, 1989; Kidd, 2007). This study confirms that the Newbery does privilege historical fiction, as numerous scholars have pointed out previously, but it also privileges fantastical genres, including science fiction, fantasy, and horror; adventure stories; and non-prose works as well. Measures of contemporary popularity suggest that readers currently prefer Medal-winning titles from these other genres to the more studied historical fiction. To focus on historical fiction at the exclusion of the other generic forms in the corpus is to miss crucial components to the types of children's literature that the Newbery privileges and to paint an overly reductive picture of the Medal.

Existing critical attention to setting in Newbery Medal-winning texts typically focuses on two strands: one on small town America, the other exotic foreign locals. Again, a distant perspective confirms that these two strands do indeed exist within the corpus, and they exist frequently. They do not, however, represent the only two settings that frequently occur. The Newbery also privileges stories about major metropolitan areas in the United States, particularly on the East coast and in the Midwest. Further, the Newbery does not privilege stories about small towns so much as it frequently foregrounds tales of pioneer life, homesteading, and villages, or what might be collectively considered as non-urban displacement narratives. Distant reading practices are unable to locate suburbs, the West coast, and mid-size cities in the corpus at all, suggesting a lacuna in the types of stories that the Medal privileges deserving further consideration. Distant reading does, however, identify the curious prominence of medieval England in the corpus. Outside of the American frontier and New York City, no other single setting features as prominently in the corpus. As with cities, however, current models of the

Newbery Medal do not discuss why an award for American literature finds such fascination with pre-modern Britain.

Distant Perspectives for Children's Literature

I employed distant reading techniques in this dissertation as a test case, asking if distant reading can provide any nuance to existing critical models for understanding children's literature as a genre. Applying theories of distant reading to children's literature suggests that seminal studies of children's literature as genre rely on overly restrictive samples, making sweeping assertions about the entire genre from a single text (Rose, 1984; Shavit, 1986) or a small cluster of texts (Nikolajeva, 2013; Nodelman, 2008). This study highlights the ways in which an overly restrictive sample leads to overly reductive findings. Distant reading techniques did not reveal that previous understandings of the Newbery were wrong; rather, they revealed that previous understandings were incomplete because they did not consider the entire corpus.

This study employs metadata as a tool for facilitating distant reading techniques on a corpus of contemporary, and therefore copyright protected, texts. Although Franco Moretti's original thought pieces on distant reading relied on secondary data sources (2005), it is now much more common for distant reading to rely on full text corpora (Berry and Fagerjord, 2017). As a result, the literature on distant reading provides little guidance for dealing with the dearth of data for contemporary, copyright protected full text corpora or for developing strategies for employing distant reading techniques using secondary data sources as the base for analysis. This study provides a test case for using metadata to facilitate distant reading and shift the view of a corpus from individual exemplars to a more holistic understanding of the corpus as a whole. As this methodology is exploratory in nature, I selected Newbery Medal winners as a test corpus so

that I could test the findings suggested by distant reading against what is already known about a relatively well-studied corpus.

This strategy was, to a certain extent, successful. Many of the findings I presented from a distant perspective correlate with and augment existing knowledge about the corpus, and others point to lacunae in the existing framework for understanding the Newbery. Using metadata to enable distant reading in a corpus of contemporary American children's literature indeed works, although it is undeniably messy at times. The findings in this study are only as good as the metadata used to enable distant reading. Distant reading techniques reveal holes in critical understandings of the Newbery, but they also reveal holes in the very metadata used to gain those understandings. The limitations of metadata are particularly prominent when considering the role of illustration in the corpus. Metadata points to significant quantities of illustration, but it does not enable substantive analysis beyond the fact that the Newbery is an illustrated corpus. Similarly, headings for geographic setting and time period are severely lacking and inconsistently applied across metadata sources. The distant reading techniques I employed in this study were able to locate genres outside of historical fiction, but they were unable to provide much in the way of nuance in understanding those genres beyond their identification. Genre, particularly speculative genres, warrant further, microscopic consideration in the corpus.

As the limitations of each metadata standard considered in this study indicate, it is crucial to employ metadata from different sources as well as metadata created by different constituencies, as well as to read them in conversation with one another. I fully anticipated metadata from Accelerated Reader, for example, to prove the weak link in this study given its odd structure and its reliance on headings, such as "misc./other," that seem designed to obfuscate the task of locating books by content, rather than reading level. In some cases, particularly for

geographic setting, Accelerated Reader metadata did indeed prove less than helpful, especially in its lack of consistency in application over time. For providing insight into thematic content and genre, however, Accelerated Reader's headings proved surprisingly helpful due to the unique, even odd, structure of the controlled vocabulary employed in the database. Although the subheadings applied to thematic categories were unhelpful, Accelerated Reader consistently applied top-level thematic headings across the corpus, enabling a high level, consistent consideration of theme and content. Neither Library of Congress Subject Headings nor Goodreads user tags proved quite as helpful due to the wide range of different terms applied.

Similarly, asking whether the texts in a corpus are popular or not seems a simple question, but as the data from different sources indicate, it is not. Different data sources provide different perspectives in answer to this question, and triangulating results from these different data sources provides nuance to an understanding of popularity that is missed from a single data source, such as print status. This principle transfers to the other elements that I considered in this study. Seemingly simple questions and assertions are actually quite complex, and analyses that highlight only the easiest to identify elements reduce complexity in the corpus.

Methodological Matters

Distant, computational reading techniques require explicit attention to method. All of the findings that I presented in this dissertation are inextricably intertwined with the methodological decisions that I made in producing them. Method shaped the entire project, from the data sources that I chose to analyze, to the way in which I defined a decade of Medal-winning texts and the tools I employed for analysis. Different data sources provide different answers, as the discussion of popularity above indicates, and different data analysis strategies lead to different interpretations. These findings reinforce my argument that distant reading, which relies on

computational data, is interpretation, not objective truth, as Moretti (2005) would have it. They also underscore the need for explicit attention to methodology in the text. To relegate method to appendix and focus exclusively on crafting an "interesting" argument that does not "[bog] down in methodological detail" (Underwood, 2016, n.p.) is to miss the point. Methodological detail provides a crucial part of the argument.

Once Upon a Genre

I began this dissertation with a reflection on my interest in the scholarly stories surrounding children's literature, both those told about children's literature as genre as well as those told about the scholars who generate the tales. As such, I conclude with a consideration of what insights this case study on distant reading might offer to these stories. Before doing so, however, I reiterate that I did not select Newbery Medal-winning texts as a proxy for children's literature as a monolithic structure. Instead, I selected Newbery Medal-winning texts as an example of one discrete sub-genre of children's literature to serve as a test case. The findings directly related to the Newbery corpus, then, do not generalize to broader conceptualizations of children's literature.

The methods for generating these findings about the Newbery corpus, however, do generalize to the broader field. The case of the Newbery Medal illustrates how shifting the critic's perspective from the microscopic to the macroscopic leads to shifts in understanding the text set. To generate a more holistic view of even this small sub-genre, it proved necessary to examine carefully the entire corpus rather than a few examples taken as representative. Further, this holistic view benefited enormously from interdisciplinary data sources, particularly given the restrictions imposed by copyright on the availability of full text. My test case suggests that siloed data sets derived from a single discipline fatally skew the results of distant reading. The story I

tell about the Newbery Medal and children's literature, then, is that it's time for a new story. This new story is one that champions interdisciplinarity rather than hiding behind artificial barriers. It is a story that asks the teller to focus on how the telling of the story happens, and how that telling shapes the narrative. Finally, it is a story that advocates listening to as many voices as possible before deciding what the story is trying to say.

References

Adler, M. (2009). Transcending library catalogs: A comparative study of controlled terms in Library of Congress Subject Headings and user-generated tags in LibraryThing for transgender books. *Journal of Web Librarianship, 3*(4), 309-331.

Alberghene, J. M. (1981). From frontier to foreign shores: Seeing ourselves in the thirties. *Children's Literature Association Quarterly*, *6*(3), 10–12.

Ascari, M. (2014). The dangers of distant reading: Reassessing Moretti's approach to literary genres. *Genre: Forms of Discourse and Culture*, *47*(1), 1–19.

Association for Library Collections & Technical Services. (1990). *Guidelines on subject access to individual works of fiction, drama, etc*. Chicago: American Library Association.

Algee-Hewitt, M., & McGurl, M. (2015). Between canon and corpus: Six perspectives on 20th-century novels. *Literary Lab Pamphlet*, *8*. Retrieved from http://litlab.stanford.edu/LiteraryLabPamphlet8.pdf

Allen, J. (2011). *ELAN 5/7312: Teaching poetry in elementary & middle school*. Retrieved from http://www.childrensliteratureassembly.org/docs/Allen-ELAN%205-7312.pdf

Allen, K. (2011). Poetic gestures. *School Library Journal*, 61–63.

American Association of School Librarians. (2011, August 2). Position statement on labeling books with reading levels. Retrieved October 2, 2016, from http://www.ala.org/aasl/advocacy/resources/statements/labeling

Artman-Meeker, K., Grant, T. O., & Yang, X. (2016). By the book: Using literature to discuss disability with children and teens. *TEACHING Exceptional Children*, *48*(3), 151–158.

Association for Library Service to Children. (2001). *The Newbery and Caldecott Medal books, 1986-2000: A comprehensive guide to the winners*. Chicago: American Library Association.

Association for Library Service to Children. (2009). John Newbery Award committee manual. Retrieved from http://www.ala.org/alsc/sites/ala.org.alsc/files/content/newbery_manual_august2015_Current%20on%20Website.pdf

Association for Library Service to Children. (2016a). The John Newbery Medal. Retrieved April 14, 2016, from http://www.ala.org/alsc/awardsgrants/bookmedia/newberymedal/aboutnewbery/aboutnewbery

Association for Library Service to Children. (2016b). Newbery Award selection committee. Retrieved April 14, 2016, from http://www.ala.org/alsc/awardsgrants/bookmedia/newberymedal/newberycommittee/committeemembersnew

Association for Library Service to Children. (2016c). Newbery Medal terms and criteria. Retrieved April 14, 2016, from http://www.ala.org/alsc/awardsgrants/bookmedia/newberymedal/newberyterms/newberyterms

Association for Library Service to Children (ALSC). (2016d). Welcome to the Newbery Medal home page! Retrieved April 14, 2016, from http://www.ala.org/alsc/awardsgrants/bookmedia/newberymedal/newberymeda

Aziz, S. (2011). *CIED 5153: Advanced studies in children's literature*. Retrieved from http://www.childrensliteratureassembly.org/docs/Aziz-CIED-5153.pdf

Bator, R. (1983). *Signposts to criticism of children's literature*. Chicago: American Library Association.

Berry, D. M., & Fagerjord, A. (2017). *Digital humanities: knowledge and critique in a digital age*. Cambridge, England: Polity Pres.

Bingham, J. (1980). Literature for children as taught in a department of education. *Proceedings of the Children's Literature Association*, 70–75.

Bird, E. (2013). The one and only. *School Library Journal, 59*(3), 25.

Bittner, R. & Superle, M. (2016). The last bastion of aesthetics? Formalism and the rhetoric of excellence in children's literature. In K.B. Kidd & J.T. Thomas (Eds.), *Prizing children's literature: The cultural politics of children's book awards* (pp. 73-86). New York: Routledge.

Bloomberg. (2017). Company overview of R.R. Bowker LLC. Retrieved from http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=1165718

Bode, K. (2017). The equivalence of "close" and "distant" reading; Or, toward a new object for data-rich literary history. *Modern Language Quarterly, 78*(1), 77-106.

Booth, A. (2017). Mid-range reading: Not a manifesto. *PMLA, 132*(3), 620-627.

Bosman, J. (2017, February 18). With one word, children's book sets off uproar. *New York Times*. Retrieved from http://www.nytimes.com/2007/02/18/books/18newb.html

Braun, L. (2013). Can it be done? Readers' advisory in a digital world. *Voice of Youth Advocates*, *36*(4), 44–45.

Breeding, M. (2010). SkyRiver and Innovative Interfaces Sue OCLC. *Library Journal, 135*(20), 18-20.

Broemel, A., Wysmierski, J., & Gibson, I. (2014). When one size does not fit all: Making Newbery literature accessible for all students. *NERA Journal, 49*(2), 51-59.

Burnite, C. (1911). The standard selection of children's books. *Library Journal*, *36*, 161–166.

Buurma, R. S. (2015, July-December). The fictionality of topic modeling: Machine reading Anthony Trollope's Barsetshire series. *Big Data & Society*, 1-6.

Chatham, W. L. (1980). The reading level of the John Newbery books from 1945-1965. *Southern Journal of Educational Research, 14*(1), 71-78.

Clark, B. L. (1992). American children's literature: background and bibliography. *American Studies International*, *30*(1), 4–40.

Clark, B. L. (2003). *Kiddie lit: the cultural construction of children's literature in America*. Baltimore: Johns Hopkins University Press.

Clark, R. (2007). From margin to margin? Females and minorities in Newbery and Caldecott Medal-winning and Honor books for children. *International Journal of Sociology of the Family*, *33*(2), 263–283.

Clements, N., Gillespie, C., and Swearingen, R. (1994). Readability and the Newbery award winners: How do they compare? *American Reading Forum Online Yearbook, 14*, 154-163.

Conduct of life. (n.d.). In *Library of Congress subject headings*. Retrieved from http://id.loc.gov/authorities/subjects/sh85030802.html

Cook, T. E. (1985). The Newbery award as political education: Children's literature & cultural reproduction. *Polity*, *17*(3), 421–445. http://doi.org/10.2307/3234652

Copyright Law of the United States, 17 USC § 101-§1332.

Cordell, R. (2016). What has the digital meant to American periodicals scholarship? *American Periodicals: A Journal of History & Criticism, 26*(1), 2-7.

Crisp, T. (2011). *Critically reading children's and young adult literature*. Retrieved from http://www.childrensliteratureassembly.org/docs/Crisp-CRS.pdf

Crisp, T. & Hiller, B. (2011). "Is this a boy or a girl?": Rethinking sex-role representation in Caldecott Medal-winning picturebooks, 1938-2011. *Children's Literature in Education, 42*(3), 196-212.

Cuddon, J. A. (2012). *A dictionary of literary terms and literary theory* (5th ed). Oxford: Blackwell Publishing.

Cummins, J. (2016). The still almost all-white world of children's literature: Theory, practice, and identity-based children's book awards. In K. B. Kidd & J.T. Thomas (Eds.), *Prizing children's literature: The cultural politics of children's book awards* (87-103). New York: Routledge.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*, 11-20, 37054.

Darton, F. J. H. (1932/1982). *Children's books in England: Five centuries of social life*. Cambridge: Cambridge University Press.

Davies, S. (2014). Ditch the posters. *School Librarian*, *62*(3), 134–136.

de la Peña, M. (2015). *Last stop on Market Street*. New York: Putnam's.

Despain, S. M., Tunnell, M. O., Wilcox, B., and Morrison, T. G. (2015). Investigating shifts in diverse family structures in Newbery Award and Honor books utilizing U.S. census data, 1930-2010. *Literacy Research and Instruction, 54*(4), 316-340.

Devereaux, E. (2008). The same old story. *Publishers Weekly*, *255*(4), 74–74.

Drouin, J. (2014). Close- and distant-reading modernism: Network analysis, text mining, and teaching The little review. *Journal of Modern Periodical Studies*, *5*(1), 110–135.

Dunbar, R. (1997). Rarely pure and never simple: The world of Irish children's literature. *The Lion and the Unicorn, 21*(3), 309-321.

Duyvis, C. (2015). Own voices. (Blog post). Retrieved from http://www.corinneduyvis.net/ownvoices/

Dyson, J.-P. C. (2007). Children's books go to war: Newbery and Caldecott winners during World War II. In L. Ousley (Ed.), *To see the wizard: Politics and the literature of childhood* (Vols. 1–xxi, 439 , pp. 200–218). Newcastle upon Tyne, England: Cambridge Scholars.

Eaton, G. (2010). Camping out in the people's palace: Public library space for Boston children, 1895 to 1975. *Children & Libraries: The Journal of the Association for Library Service to Children*, *8*(1), 3–7.

Egoff, S. A. (1979). *One ocean touching: Papers from the first Pacific Rim conference on children's literature*. Metuchen, N.J: Scarecrow Press.

Elson, D., Dames, N., & McKeown, K. (2010). Extracting social networks from literary fiction. Proceedings from the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden.

English, J. F. (2005). *The economy of prestige: Prizes, awards, and the circulation of cultural value*. Cambridge, Mass.: Harvard University Press.

Eppley, K. (2015). Seven traps of the common core state standards. *Journal of Adolescent & Adult Literacy*, *59*(2), 207–216.

Evans, R. (2014). Beyond the e-reader: Alternative uses for the ipad in libraries. *Computers in Libraries*, *34*(5), 5–10.

Fellman, A. C. (1996). "Don't expect to depend on anybody else": The frontier as portrayed in the Little House books. *Children's Literature, 24*, 101-116.

Fialkoff, F. (2013). Penguin Random: Bigger is better – or not. *Library Journall, 138*(14), 24.

Fingon, J. C. (2012). Nontraditional texts and the struggling/reluctant reader. *Voices from the Middle*, *19*(4), 70–75.

Fleming, K., & Parker, K. (2013). Character traits depicted in Newbery Medal books from the 1920s through the 2000s. *Journal of Research on Christian Education, 22*(3), 264-282.

Forest, D. E. (2014). From rags to "rich as Rockefeller": Portrayals of class mobility in Newbery titles. *Curriculum Inquiry 44*(5), 591-619.

Fountain, J. F., & Fountain, J. F. (1996). *Subject headings for school and public libraries: an LCSH/Sears companion*. Englewood: Libraries Unlimited.

Friedman, A. A., & Cataldo, C. A. (2002). Characters at crossroads: Reflective decision makers in contemporary Newbery books. *Reading Teacher*, *56*(2), 102–112.

Frow, J. (2014). *Genre*. New York: Routledge.

Gall, E. (2017). Empower kids with #OwnVoices: My students weighed in on diverse representations in books. *School Library Journal, 63*(5), 14.

Gabriel, R., Allington, R., & Billen, M. (2012). Middle schoolers and magazines: What teachers can learn from students' leisure reading habits. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *85*(5), 186–191.

Ganss, D. (2015). Classics -- great literature and good reads. *School Librarian's Workshop*, *35*(6), 18–19.

Gillespie, C. S., Powell, J. L., Clements, N. E., & Swearingen, R. A. (1994). A look at the Newbery Medal books from a multicultural perspective. *The Reading Teacher*, (1), 40-50.

Gillespie, J. T., & Naden, C. J. (2006). *The Newbery/Printz companion: Booktalk and related materials for award winners and honor books*. Westport, Conn.: Libraries Unlimited.

Goldstone, A., & Underwood, T. (2014). The quiet transformations of literary studies: What thirteen thousand scholars could tell us. *New Literary History: A Journal of Theory and Interpretation*, *45*(3), 359–384.

Gómez, S. H. (2016). Decolonizing nostalgia: When historical fiction betrays readers of color. *Horn Book Magazine, 92*(6), 20-26.

Gooding, P., Terras, M., & Warwick, C. (2013). The myth of the new: Mass digitization, distant reading, and the future of the book. *Literary and Linguistic Computing: Journal of the Association for Literary and Linguistic Computing and The Association for Computers and the Humanities*, *28*(4), 629–639.

Goodreads Inc. (2016). About Goodreads. Retrieved October 8, 2016, from https://www.goodreads.com/about/us

Graff, J. M. (2011). *LLED 7045: Reading picturebooks*. Retrieved from http://www.childrensliteratureassembly.org/docs/Graf-LLED-7045.pdf

Groth, R. E. (2015). Royalty, racing, and rolling pigs. *Teaching Children Mathematics*, *22*(4), 218–227.

Guidry, L. J., and Knight, F. F. (1976). Comparative readability: Four formulas and Newbery books. *Journal of Reading, 19*(7), 552-556.

Hale, R. (1995). Musings. *Horn Book Magazine*, *7*(1*),* 363-364.

Hanaway, E. S. (1887). The children's library in New York. *Library Journal, 12*(5), 185-186.

Haykin, D. J. (1951). *Subject headings: A practical guide*. Washington: U.S. Government Printing Office.

Hayles, N. K. (2013). Combining close and distant reading: Jonathan Safran Foer's Tree of codes and the aesthetic of bookishness. *PMLA: Publications of the Modern Language Association of America*, *128*(1), 226–231.

Hearne, B. (1996). Margaret K. McElderry and the professional matriarchy of children's books. *Library Trends*, *44*(4), 755–75.

Held, J. R. (2012). "Distant reading" on a small scale: Lighting and ethics in Far from the madding crowd. *Hardy Society Journal*, *8*(2), 49–65.

Hiebert, E. H., & Mesmer, H. A. E. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher*, *42*(1), 44-51.

Hiebert, E. H. (2015). Changing readers, changing texts: Beginning reading texts from 1960 to 2010. *Journal of Education*, *195*(3), 1–13.

Hill, R. A. (2011). The color of authenticity in multicultural children's literature. *Voice of Youth Advocates*, *34*(5), 445–447.

Hoffert, B. (2010). Every reader a reviewer: The online book conversation. *Library Journal*, *135*(14), 23–25.

Horning, K. (2015). Milestones for diversity in children's literature and library services. *Children & Libraries, 13*(3), 7-11.

Hong, T. (2002). Repeating forbidden words: A profile of Linda Sue Park, winner of the Newbery Medal. *Bloomsbury Review*, *22*(6), 12–12.

Horn Book. (2017). *Submissions*. Retrieved July 31, 2017 from http://www.hbook.com/about-us/submissions/

Horning, K. T. (2004). The tale of DiCamillo. *School Library Journal, 50*(4), 44-47.

Houdyshell, M. L., & Kirkland, J. J. (1998). Heroines in Newbery Medal award winners: Seventy-five years of change. *Journal of Youth Services in Libraries*, *11*(3), 252–262.

Houdyshell, M. L., & Martin, C. M. (2010). Heroines in Newbery Medal award winners. *Children and Libraries, 8*(1), 25-31.

Hoyt, E., Long, D., Tran, T., & Hughes, K. (2015). Variety's transformations: Digitizing and analyzing the first thirty-five years of the canonical trade paper. *Film History: An International Journal*, *27*(4), 76–105.

Hunt, P. (2001). *Children's literature*. Oxford: Wiley Blackwell.

Imdieke, S. (2012). Newbies and Newberys. *Child & Libraries: The Journal of the Association for Library Service to Children, 10*(1), 30-36.

Jenkins, C. A. (1996). Women of ALA youth services and professional jurisdiction: Of nightingales, Newberies, realism, and the right books, 1937-1945. *Library Trends*, *44*(4), 813–39.

Kain, P. (1998). How to do a close reading. Retrieved August 3, 2017 from https://writingcenter.fas.harvard.edu/pages/how-do-close-reading

Katz, W. A. (1980). Collection development: the selection of materials for libraries. New York: Holt, Rinehart and Winston.

Kaufman, L. (2013, March 28). Amazon to buy Goodreads. *The New York Times*. Retrieved from http://www.nytimes.com/2013/03/29/business/media/amazon-to-buy-goodreads.html

Kidd, K. (2007). Prizing children's literature: The case of Newbery gold. *Children's Literature: Annual of The Modern Language Association Division on Children's Literature and The Children's Literature Association*, *35*, 166–190.

Kimball, M. A. (2012). Seeing the World from Main Street: Early Twentieth-Century Juvenile Collections about Life in Other Lands. *Library Trends*, *60*(3), 675–693.

Kingman, L. (1965). *Newbery and Caldecott Medal books: 1956-1965: With acceptance papers, biographies & related materials chiefly from the Horn book magazine.* Boston: Horn Book.

Kingman, L. (1975). *Newbery and Caldecott Medal books, 1966-1975: With acceptance papers, biographies, and related material chiefly from the Horn book magazine*. Boston: Horn Book.

Kingman, L. (1986). *Newbery and Caldecott Medal books, 1976-1985: With acceptance papers, biographies, and related material chiefly from the Horn book magazine*. Boston: Horn Book.

Kirschenbaum, M., & Werner, S. (2014). Digital scholarship and digital studies: The state of the discipline. *Book History 17*, 406-458.

Knowlton, S. A. (2005). Three decades since "prejudices and antipathies": A study of changes in the Library of Congress Subject Headings. *Cataloging & Classification Quarterly, 40*(2), 123-145.

Krashen, S. D. (2001). The Lexile framework: Unnecessary and potentially harmful. *CSLA Journal*, *24*(2), 25–26.

Krashen, S. D. (2002). The Lexile framework: The controversy continues. *CSLA Journal*, *25*(2), 29–31.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks: Sage.

Kunze, P. C. (2015). 13 ways of looking at a wild thing: Teaching children's literature in the American literature survey. *Teaching American Literature*, *7*(3/4), 23.

Lacy, M. (2015). Portraits of children of alcoholics: Stories that add hope to hope. *Children's Literature in Education*, *46*(4), 343–358.

Larrick, N. (1965). The all-white world of children's books. *Saturday Review*, *48*(11), 63–65, 84–85.

Lathey, J. (2005). Challenges then and now: A survey of protagonists in Newbery Award books, 1950s and 1990s. *Bibliographical Essay*, *3*(3), 20–24.

Leal, D. J., & Chamberlain-Solecki, J. (1998). A Newbery Medal-winning combination: High student interest plus appropriate readability levels. *Reading Teacher, 51*(8), 712-716.

Leetaru, K. H. (2015). Mining libraries: Lessons learned from 20 years of massive computing on the world's information. *Information Services & Use*, (1–2), 31-50.

Leininger, M., Dyches, T. T., Prater, M. A., & Heath, M. A. (2010). Newbery award winning books 1975–2009: How do they portray disabilities? *Education and Training in Autism and Developmental Disabilities, 45*(4), 583-596.

Levy, M., & Mendlesohn, F. (2017). *Children's fantasy literature*. Cambridge: Cambridge University Press.

Levy, M., & Perry, M. (2015). Distantly reading the Romantic canon: Quantifying gender in current anthologies. *Women's Writing*, *22*(2), 132–155.

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic media, 57*(1), 34-52.

Liddle, D. (2015). Genre: "Distant reading" and the goals of periodicals research. *Victorian Periodicals Review*, *48*(3), 383–402.

Long, H., & So, R. J. (2016). Literary pattern recognition: Modernism between close reading and machine learning. *Critical Inquiry*, (2), 235-267.

Lucas, J. (2003). *AR: Accelerated Reader and accurate reflection? A collection assessment of the Accelerated Reader collection at an elementary school in north central North Carolina* (unpublished master's thesis). University of North Carolina, Chapel Hill.

Lundin, A. H. (2004). *Constructing the canon of children's literature: beyond library walls and ivory towers*. New York: Routledge.

Lysaker, J., & Hopper, E. (2015). A kindergartner's emergent strategy use during wordless picture book reading. *Reading Teacher*, *68*(8), 649–657.

MacLeod, A. S. (1998). Writing backward: Modern models in historical fiction. *Horn Book Magazine, 74*(1), 26-33.

Madsen, J. M., & Robbins, R. (1981). Native American visual and verbal images in the Caldecott and Newbery Award books. *Minority Voices*, *5*(1–2), 17–40.

Magner, B. (2012). Behind the BookScan bestseller lists: Technology and cultural anxieties in early twenty-first-century Australia. *Script & Print: Bulletin of the Bibliographical Society of Australia and New Zealand*, *36*(4), 243–258.

Marcus, L. S. (2008). *Minders of make-believe: Idealists, entrepreneurs, and the shaping of American children's literature*. Boston: Houghton Mifflin Co.

Marshall, E. (2012). Global girls and strangers: Marketing transnational girlhood through the Nancy Drew series. *Children's Literature Association Quarterly*, *37*(2), 210–227.

Marshall, J. K. (1977). *On equal terms: A thesaurus for nonsexist indexing and cataloging*. New York: Neal-Schuman.

Martens, M. (2013). The librarian lion: Constructing children's literature through connections, capital, and criticism (1906-1941). *Journal of Education for Library and Information Science*, *54*(4), 307–319.

Maryles, D. (2015). What the numbers reveal about the 2014 bestsellers. *Publisher's Weekly, 262*(2), 21-26.

Matthews, C. (1908). The growing tendency to over-emphasize the children's side. *Library Journal, 33*, 135-138.

Maughan, S. (2013). The call that changes everything-or not. *Publishers Weekly*, *260*(2), 16–19.

Maxwell, J. (2009). Designing a qualitative study. In L. Bickman & D. J. Rog (Eds.), *The SAGE Handbook of Applied Social Research Methods* (pp. 214–252). Thousand Oaks: SAGE.

McDowell, K. (2014). Open wide the doors: The children's room as place in public libraries, 1876-1925. *Library Trends*, *62*(3), 519–529.

McKoy Lowery, R. & Baglier, T. (2016). "And the winner is…": Celebrating diverse cultural experience through award-winning children's literature. *Literacy Today, 33*(6), 10-11.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., and Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Michel, L. (2016). Everything you wanted to know about book sales (but were afraid to ask): An in-depth look at what/how/why books sell. *Electric Lit*. Retrieved from

https://electricliterature.com/everything-you-wanted-to-know-about-book-sales-but-were-afraid-to-ask-1fe6bc00aa2d

Miller, B. J. F. (1998). What color is gold?: Twenty-one years of same-race authors and protagonists in the Newbery Medal. *Journal of Youth Services in Libraries*, *12*(1), 34–39.

Miller, C. F. (2014). Precious medals: The Newbery Medal, the YRCA, and the gold standard of children's book awards. In A. E. Hubler (Ed.), *Little red readings: Historical materialist perspectives on children's literature* (pp. 57–74). Jackson, MS: UP of Mississippi.

Miller, L. J. (2000). The best-seller list as marketing tool and historical fiction. In Ezra Greenspan (Ed.), *Book History* (pp. 286-304). University Park: Penn State Press.

Moir, H. (1981). Personal courage and American values: Newbery Medal and Honor books: 1952-1961. *Children's Literature Association Quarterly*, *6*(3), 12–15.

Möller, K. J. (2015). Integrating graphic nonfiction into classroom reading and content area instruction: A critical literacy focus on selection issues. *Journal of Children's Literature*, *41*(2), 52–59.

Moretti, F. (2000). Conjectures on world literature. *New Left Review*, *1*, 54–68.

Moretti, F. (2005). *Graphs, maps, trees: Abstract models for a literary history*. New York: Verso.

Morrison, Toni. (1989). Unspeakable things unspoken: The Afro-American presence in American literature." *Michigan Quarterly Review 28*(1), 1–34.

Murray, G. S. (1998). *American children's literature and the construction of childhood*. New York : Twayne Publishers ; London : Prentice Hall International, c1998.

Nel, P., & Paul, L. (Eds.). (2011). *Keywords for children's literature*. New York: NYU Press.

Nelson, C. (2011). Jade and the tomboy tradition. In J. L. Mickenberg & L. Vallone (Eds.), *The Oxford handbook of children's literature* (p. 497-517). Oxford, England: Oxford University Press.

Nelson, R. K. (2016). Digital humanities as appendix. *American Quarterly*, *68*(1), 131–136.

Nikolajeva, M. (1996). *Children's literature comes of age: Toward a new aesthetic*. New York: Garland.

Nikolajeva, M. (2002). *The rhetoric of character in children's literature*. Lanham, MD: Scarecrow Press.

Nikolajeva, M. (2009). *Power, voice and subjectivity in literature for young readers*. London: Routledge.

Nikolajeva, M. and Scott, C. (2013). *How picturebooks work.* London: Routledge.

Nikolajeva, M. (2016). Recent trends in children's literature research: Return to the body. *International Research in Children's Literature, 9*(2), 132-145.

Nodelman, P. (1989). *Words about pictures: The narrative art of children's picture books*. Athens: University of Georgia Press.

Nodelman, P. (1990). History as fiction: The story in Hendrik Willem van Loon's Story of mankind. *The Lion and the Unicorn: A Critical Journal of Children's Literature*, *14*(1), 70–86.

Nodelman, P. (2008). *The hidden adult: Defining children's literature*. Baltimore: Johns Hopkins University Press.

Nuckolls, K. A. (1994). Subject access to diversity materials. *The Reference Librarian, 21*, 45-46, 241-251.

OCLC. (2017a). About OCLC. Retrieved July 27, 2017, from https://www.oclc.org/en/about.html

OCLC. (2017b). Directory of OCLC members. Retried July 27, 2017, from https://www.oclc.org/en/contacts/libraries.html?cmpid=md_mem_directory

Parsons, L. T., & Castleman, M. (2011). “I have a dream, too!”: The American dream in Coretta Scott King Award-winning books. *Journal of Children’s Literature*, *37*(1), 6–18.

Passet, J. E. (1993). Men in a feminized profession: The male librarian, 1887-1921. *Libraries & Culture*, *28*, 385–402.

Pease, H. (1939). Children’s books today: One man’s view. In *Proceedings of the Institute on Library Work with Children* (Jun 15-17, 1939), (pp. 5-20). Berekeley, CA: School of Librarianship, University of California.

Pérez, C. C. (2017). When Google Translate gives erroneous Español. *Horn Book Magazine, 93*(2), 44-48.

Peterson, L. K., & Solt, M. L. (1982). *Newbery and Caldecott Medal and Honor books: An annotated bibliography*. Boston, Mass: G.K. Hall.

Pomerantz, J. (2015). *Metadata*. MIT Press.

Powell, J. L., Gillespie, C. S., Swearingen, R. A., & Clements, N. L. (1998). The history of gender roles in the Newbery Medal winners. *Journal of Children’s Literature*, *24*(1), 42–57.

Prendergast, C. (2005). Evolution and literary history. *New Left Review*, (34), 40–62.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, October 26). Altmetrics: A manifesto. Retrieved August 5, 2016, from http://altmetrics.org/manifesto/

Rapp, D. (2011). Crowdsourcing RA. *Library Journal*, *136*(10), 56–57.

Renaissance Learning, Inc. (2017). Text complexity, ATOS, and Lexile measures. Retrieved from https://www.renaissance.com/products/practice/accelerated-reader-360/atos-and-text-complexity/

Richards, I. A. (1929). *Practical criticism: A study of literary judgment*. New York: Harcourt Brace.

Roggenkamp, K. (2008). Seeing inside the mountains: Cynthia Rylant's Appalachian literature and the "hillbilly" stereotype. *Lion and the Unicorn*, *32*(2), 192–215.

Rose, J. (1984). *The case of Peter Pan, or, The impossibility of children's fiction*. London: Macmillan.

Rue, E., & La Plante, E. (1952). *Subject headings for children's materials*. Chicago: American Library Association.

Russell, D. (2005). *Literature for Children*. Boston: Pearson.

Schafer, P. J. (1976). The readability of the Newbery Medal books. *Language Arts, 53*(5), 557-559.

Schafer, P. J. (1986). The readability of the Newbery Medal books, 1974-1986 (Report No. ED273929). Retrieved from ERIC database.

Schneebaum, K. (1990). Finding a happy medium: The design for womanhood in *A wrinkle in time*. *The Lion and the Unicorn, 14*(2), 30-36.

Schneider, J. J. (2016). *The inside, outside, and upside downs of children's literature: From poets and pop-ups to princesses and porridge*. Retrieved from http://scholarcommons.usf.edu/childrens_lit_textbook/

Schreiber, M. (2017). And the Newbery goes to…a picturebook? *Children & Libraries: The Journal of the Association for Library Services to Children, 15*(2), 29-31.

Schwarcz, J. H., & Schwarcz, C. (1991). *The picture book comes of age: looking at childhood through the art of illustration*. Chicago: American Library Association.

Serafini, F., Kachorsky, D., & Aguilera, E. (2016). Picture books in the digital age. *Reading Teacher*, *69*(5), 509-512.

Shannon, L. C., Styers, M. K., Wilkerson, S. B., & Peery, E. (2015). Computer-assisted learning in elementary reading: A randomized control trial. *Computers in the Schools*, *32*(1), 20–34.

Shavit, Z. (1986). *Poetics of children's literature*. Athens: University of Georgia Press.

Short, K. G. (2011). *LRC 586: Reading the world through children's and adolescent literature.* Retrieved from http://www.childrensliteratureassembly.org/docs/Short-LRC-586.pdf

Simon. L. (2008). Weaving colors into a white landscape: Unpacking the silences in Karen Hesse's children's novel *Out of the dust*. *Multicultural Education, 16*(2), 22-30.

Smith, D. (2000). The Times plans a children's best-seller list. *New York Times*. Retrieved from http://www.nytimes.com/2000/06/24/books/the-times-plans-a-children-s-best-seller-list.html

Smith, I. (1957). *A history of the Newbery and Caldecott medals*. New York: Viking Press.

Smulders, S. (2015). Dresses make the girl: Gender and identity from "The hundred dresses" to "10,000 dresses." *Children's Literature in Education*, *46*(4), 410–423.

Solt, M. L. (1981). The Newbery Medal and Honor books: 1922-1981. *Children's Literature Association Quarterly*, *6*(3), 22–28.

Spache, G. (1953). A new readability formula for primary grade reading materials. *The Elementary School Journal, 53*, 410-413.

St. John, B. A. (1981). Courage in the face of adversity: Recent Newbery winners. *Children's Literature Association Quarterly*, *6*(3), 21–22.

Stenner, A. J. (2001). The Lexile framework: a common metric for matching readers and text: a response to Stephen Krashen. *CSLA Journal*, *25*(1), 41–42.

Stephens, J. (1992). *Language and ideology in children's fiction*. Harlow, Essex, England ; New York: Longman.

Stevens, K. A. (2010). *Don't judge a book by a sticker on the cover: A readability study of the Newbery award winners, 1922-2010* (Unpublished master's thesis). University of North Carolina, Chapel Hill.

Stevenson, D. (2011). History of children's and young adult literature. In S. A. Wolf et al., *Handbook of research on children's and young adult literature* (179-194). New York: Routledge.

Sutton, R. (2009). It's good to be Gaiman. *School Library Journal, 55*(3), 30-32.

Topping, K. J. (2015). Fiction and non-fiction reading and comprehension in preferred books. *Reading Psychology*, *36*(4), 350–387.

Townsend, J. R. (1996). *Written for children: An outline of English-language children's literature*. Lanham: Scarecrow Press.

Townsend, J. R. (1980). Standards of criticism for children's literature. In Z. Sutherland, *The Arbuthnot lecture, 1970-1979* (26-33). Chicago: American Library Association.

Trott, B., & Naik, Y. (2012). Finding good reads on Goodreads. *Reference & User Services Quarterly*, *51*(4), 319–323.

Truitt, E. (1998, September 3). The Book Industry's Best-Seller Lists. *Slate*. Retrieved from http://www.slate.com/articles/news_and_politics/the_culture_culture/1998/09/the_book_industrys_bestseller_lists.html

Tyack, D. B. and Hansot, E. (1992). *Learning together: A history of coeducation in American schools*. New Haven: Yale University Press.

Underwood, T. (2016). The read problem with distant reading [Blog post]. Retrieved from https://tedunderwood.com/2016/05/29/the-real-problem-with-distant-reading/

Underwood, T. (2017). A genealogy of distant reading. *DHQ, 11*(2). Retrieved from http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html

Valencia, S. W., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *The Elementary School Journal*, *115*(2), 270-289.

Vardell, S. M. (2011). *LS 5603/20-21: Literature for children and young adults*. Retrieved from http://www.childrensliteratureassembly.org/docs/Vardell-LS-5603.pdf

Vizine-Goetz, D. (2001). Popular LCSH with Dewey numbers. *Journal of Library Administration*, *34*(3–4), 293–300.

West, M. I. (1985). Not to be circulated: The response of children's librarians to dime novels and series books. *Children's Literature Association Quarterly*, *10*(3), 137–139.

White, S., & Clements, J. (2001). Assessing the Lexile framework: Results of a panle meeting (Working Paper No. 2001-08). Retrieved August 3, 2017 from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200108

Wiegand, W. A. (1986). *The politics of an emerging profession: The American Library Association, 1876-1917*. Westport, Conn: Greenwood Press.

Wilfong, L. (2011). *C&I 67395: Special topics: Multicultural voices in children's and young adult literature*. Retrieved from http://www.childrensliteratureassembly.org/docs/Wilfong-CI-67395.pdf

Wilkens, M. (2013). The geographic Imagination of Civil War-era American fiction. *American Literary History*, (4), 803-840.

Wilkens, M. (2015). Digital humanities and its application in the study of literature and culture. *Comparative Literature*, *67*(1), 11–20.

Wilkins, B. T. (2009). *African and African American images in Newbery Award winning titles: Progress in portrayals*. Lanham: Scarecrow Press.

Willett, H. G. (2001). Rifles for Watie: Rollins, Riley, and racism. *Libraries & Culture: A Journal of Library History*, *36*(4), 487–505.

Zaleski, J. (2011). *LYST 220 Section a: Literature in the lives of young children*. Retrieved from http://www.childrensliteratureassembly.org/docs/Zaleski-LYST-220.pdf

Zipes, J. (2013). *Sticks and Stones: The troublesome success of children's literature from slovenly Peter to Harry Potter*. Hoboken: Taylor and Francis.

Zuccala, A. A., Verleysen, F. T., Cornacchia, R., & Engels, T. C. E. (2015). Altmetrics for the humanities. *Journal of Information Management*, *67*(3), 320–336.

Appendix A: Potential Variables Considered for Distant Reading

This appendix records the many data sources and variables considered for analysis in this study. An asterisk (*) denotes a variable selected for inclusion in this study.

Descriptive Variables:

*Bibliographic Information
*Publisher: Imprint
*Publisher: Parent Company
*Gender of Author
*Race of Author
*Gender of Main Character
*Race of Characters
*Illustrative Content
*Illustrator
*Type of Illustration
*Length
    *Number of Pages
    *Word Count

Structural Variables:

*Point of View
*Literary Form
    *WorldCat, Marc field 655: Genre/Form terms
    *Accelerated Reader tags
    *Goodreads User tags
*Text Complexity
    *Lexile measure
    *ATOS level
    Botel level
    Fry Readability Formula
    Flesch Reading Ease
    Flesch-Kincaid Formula
    Gunning Fog Index
    SMOG Readability measure

Thematic Variables

*Description
    *WorldCat (Marc field 520: Summary)

*Accelerated Reader (Summary)
*Goodreads (Summary)
*Subject
*WorldCat (Marc fields 650, 651, 690)
WorldCat (Children's Subject Headings)
BISAC headings
Sears headings
Books in Print (Sears Headings)
*Accelerated Reader (Subjects)
*Goodreads (User tags)
Professional Reviews Genre Categories
* Setting: Geographic
*Book summaries
*WorldCat (Marc fields 650, 651, 690, subfield z)
BISAC headings
Sears headings
*Accelerated Reader (Subjects)
*Goodreads (User tags)
Professional Reviews: free text
*Setting: Time Period
*Book summaries
*WorldCat (Marc fields 650, 651, 690, subfield y)
BISAC Headings
Sears headings
*Accelerated Reader (Subjects)
*Goodreads (User tags)
Professional Reviews: free text

Variables Describing Popularity

Presence in Libraries
*Number of editions in WorldCat
*Number of library holdings in WorldCat
*Print Status
*Number of copies in print as of 2017
*Number of copies published from original date of publication-2017
Sales Rankings
New York Times bestseller list
Amazon bestseller lists
Goodreads Metrics
*Number of ratings
*Average rating
Number of reviews
Text of reviews
Amazon metrics (sales ranking, user reviews)
Number of ratings

- Average rating
- Number of reviews
- Text of reviews

LibraryThing metrics
- Number of ratings
- Average rating
- Number of times shelved
- Number of reviews
- Text of review

Booklikes
- Number of ratings
- Average rating
- Number of times reviews
- Text of reviews

Presence on assigned reading lists

Professional Reviews
- *Horn Book Magazine*
- *School Library Journal*
- *Bulletin for the Center of Children's Literature*
- *Kirkus*
- *New York Times Book Review*

Appendix B: Example of Raw Data

This appendix presents one example of my method for storing raw data in tab-delimited spreadsheets. Here, I have presented a sample of variables harvested from WorldCat that constitute a portion of the descriptive information analyzed in this study.

Year	Author	Title	Illustrations	Illustrator	Illustration type (300 field)
2017	Barnhill, Kelly	The Girl Who Drank the Moon	no	n/a	n/a
2016	de la Pena, Matt	Last Stop on Market Street	yes	Christian Robinson	chiefly illustrations (colour)
2015	Alexander, Kwame	The Crossover	no		n/a
2014	DiCamillo, Kate	Flora & Ulysses: The Illuminated Adventures	yes	K.G. Campbell	illustrations
2013	Applegate, Katherine	The One and Only Ivan	yes	Patricia Castelao	illustrations
2012	Gantos, Jack	Dead End in Norvelt	no	n/a	n/a
2011	Vanderpool, Clare	Moon over Manifest	no	n/a	n/a
2010	Stead, Rebecca	When You Reach Me	no	n/a	n/a

2009	Gaiman, Neil	The Graveyard Book	yes	Dave McKean	illustrations
2008	Schlitz, Laura Amy	Good Masters! Sweet Ladies! Voices from a Medieval Village	yes	Robert Byrd	color illustrations
2007	Patron, Susan	The Higher Power of Lucky	yes	Matt Phelan	illustrations
2006	Perkins, Lynne Rae	Criss Cross	yes	Lynne Rae Perkins	illustrations
2005	Kadohata, Cynthia	Kira-Kira	no	n/a	n/a
2004	DiCamillo, Kate	The Tale of Despereaux: Being the Story of a Mouse, a Princess, Some Soup, and a Spool of Thread	yes	Timothy Basil Ering	illustrations
2003	Avi	Crispin: The Cross of Lead (OCLC 48559447)	no; decorative elements	n/a	n/a
2002	Park, Linda Sue	A Single Shard	no; decorative elements	n/a	n/a
2001	Peck, Richard	A Year Down Yonder	no	n/a	n/a

2000	Curtis, Christopher Paul	Bud, Not Buddy	yes	not cited	n/a
1999	Sachar, Louis	Holes	no	n/a	n/a
1998	Hesse, Karen	Out of the Dust	no; decorative elements	n/a	n/a
1997	Konigsburg, E. L.	The View from Saturday	no	n/a	n/a
1996	Cushman, Karen	The Midwife's Apprentice	no; decorative elements at head of each chapter	n/a	n/a
1995	Creech, Sharon	Walk Two Moons	no; decorative elements at head of each chapter	n/a	n/a
1994	Lowry, Lois	The Giver	no	n/a	n/a
1993	Rylant, Cynthia	Missing May	no	n/a	n/a
1992	Naylor, Phyllis Reynolds	Shiloh	no	n/a	n/a
1991	Spinelli, Jerry	Maniac Magee	no; decorative elements at head of each chapter	n/a	n/a
1990	Lowry, Lois	Number the Stars	no	n/a	n/a

1989	Fleischman, Paul	Joyful Noise: Poems for Two Voices	yes	Eric Beddows	illustrations
1988	Freedman, Russell	Lincoln: A Photobiography	yes	Alfred Whital Stern Collection of Lincolniana (Library of Congress)	illustrations
1987	Fleischman, Sid	The Whipping Boy	yes	Peter Sis	illustrations
1986	MacLachlan, Patricia	Sarah, Plain and Tall	no	n/a	na
1985	McKinley, Robin	The Hero and the Crown	no; decorative elements at head of each chapter	n/a	n/a
1984	Cleary, Beverly	Dear Mr. Henshaw	yes	Paul O. Zelinsky	illustrations
1983	Voigt, Cynthia	Dicey's Song	yes	Sarah Young	illustrations
1982	Willard, Nancy	A Visit to William Blake's Inn: Poems for Innocent and Experienced Travelers	yes	Alice and Martin Provensen	color illustrations
1981	Paterson, Katherine	Jacob Have I Loved	no	n/a	n/a

1980	Blos, Joan W.	A Gathering of Days: A New England Girl's Journal, 1830-32	no	n/a	n/a
1979	Raskin, Ellen	The Westing Game	no	n/a	n/a
1978	Paterson, Katherine	Bridge to Terabithia	yes	Donna Diamond	illustrations
1977	Taylor, Mildred D.	Roll of Thunder, Hear My Cry	yes	Jerry Pinkney	illustrations
1976	Cooper, Susan	The Grey King	yes	Michael Heslop	illustrations
1975	Hamilton, Virginia	M. C. Higgins, the Great	no	n/a	n/a
1974	Fox, Paula	The Slave Dancer	yes	Eros Keith	illustrations
1973	George, Jean Craighead	Julie of the Wolves	yes	John Schoenherr	illustrations
1972	O'Brien, Robert C.	Mrs. Frisby and the Rats of NIMH	yes	Zena Bernstein	illustrations
1971	Byars, Betsy	Summer of the Swans	yes	Ted CoConis	illustrations
1970	Armstrong, William H.	Sounder	yes	James Barkley	illustrations
1969	Alexander, Lloyd	The High King	yes	not cited	map

1968	Konigsburg, E. L.	From the Mixed-Up Files of Mrs. Basil E. Frankweiler	yes	E. L. Konigsburg	illustrations
1967	Hunt, Irene	Up a Road Slowly	no	n/a	n/a
1966	Trevino, Elizabeth Borton	I, Juan de Pareja	no	n/a	n/a
1965	Wojciechowska, Maia	Shadow of a Bull	yes	Alvin Smith	illustrations
1964	Neville, Emily	It's Like This, Cat	yes	Emil Weiss	illustrations
1963	L'Engle, Madeleine	A Wrinkle in Time	no	n/a	n/a
1962	Speare, Elizabeth George	The Bronze Bow	no	n/a	n/a
1961	O'Dell, Scott	Island of the Blue Dolphins	no	n/a	n/a
1960	Krumgold, Joseph	Onion John	yes	Symeon Shimin	illustrations
1959	Speare, Elizabeth George	The Witch of Blackbird Pond	no	n/a	n/a
1958	Keith, Harold	Rifles for Watie	yes	not cited	map
1957	Sorensen, Virginia	Miracles on Maple Hill	yes	Beth and Joe Krush	illustrations
1956	Latham, Jean Lee	Carry On, Mr. Bowditch	yes	John O'Hara Cosgrave	illustrations

1955	DeJong, Meindert	The Wheel on the School	yes	Maurice Sendak	illustrations
1954	Krumgold, Joseph	…And Now Miguel	yes	Jean Charlot	illustrations
1953	Clark, Ann Nolan	Secret of the Andes	yes	Jean Charlot	illustrations
1952	Estes, Eleanor	Ginger Pye	yes	Louis Slobodkin	illustrations
1951	Yates, Elizabeth	Amos Fortune, Free Man	yes	Nora S. Unwin	illustrations
1950	de Angeli, Marguerite	The Door in the Wall	yes	Marguerite de Angeli	illustrations (some color)
1949	Henry, Marguerite	King of the Wind	yes	Wesley Dennis	illustrations (some color)
1948	du Bois, William Pene	The Twenty-One Balloons	yes	William Pène Du Bois	illustrations
1947	Bailey, Carolyn Sherwin	Miss Hickory	yes	Ruth Gannett	illustrations
1946	Lenski, Lois	Strawberry Girl	yes	Lois Lenski	illustrations
1945	Lawson, Robert	Rabbit Hill	yes	Robert Lawson	illustrations
1944	Forbes, Esther	Johnny Tremain	yes	Lynd Ward	illustrations (some color)
1943	Gray, Elizabeth Janet	Adam of the Road	yes	Robert Lawson	illustrations, map

1942	Edmonds, Walter	The Matchlock Gun	yes	Paul Lantz	illustrations (some color)
1941	Sperry, Armstrong	Call It Courage	yes	Armstrong Sperry	illustrations
1940	Daugherty, James	Daniel Boone	yes	James Daugherty	illustrations, map
1939	Enright, Elizabeth	Thimble Summer	yes	Elizabeth Enright	illustrations (some color)
1938	Seredy, Kate	The White Stag	yes	Kate Seredy	illustrations
1937	Sawyer, Ruth	Roller Skates	yes	Valenti Angelo	illustrations
1936	Brink, Carol Ryrie	Caddie Woodlawn	yes	Kate Seredy	illustrations
1935	Shannon, Monica	Dobry	yes	Atanas Katchamakoff	illustrations, map, plates
1934	Meigs, Cornelia	Invincible Louisa: The Story of the Author of Little Women	yes	not credited	frontispiece, plates, portraits
1933	Lewis, Elizabeth	Young Fu of the Upper Yangtze	yes	Kurt Wiese	illustrations (some color)
1932	Armer, Laura Adams	Waterless Mountain	yes	Sidney Armer and Laura Adams Armer	illustrations
1931	Coatsworth, Elizabeth	The Cat Who Went to Heaven	yes	Lynd Ward	illustrations, double plates

1930	Field, Rachel	Hitty, Her First Hundred Years	yes	Dorothy L. Lathrop	illustrations (some color)
1929	Kelly, Eric P.	The Trumpeter of Krakow	yes	Janina Domanska	color frontispiece, color plates
1928	Mukerji, Dhan Gopal	Gay Neck, the Story of a Pigeon	yes	Boris Artzybasheff	illustrations
1927	James, Will	Smoky, the Cowhorse	yes	Will James	illustrations
1926	Chrisman, Arthur Bowie	Shen of the Sea	yes	Else Hasselriis	illustrations
1925	Finger, Charles	Tales from Silver Lands	yes	Paul Honoré	color frontispiece, illustrations, color plates
1924	Hawes, Charles	The Dark Frigate	yes	unknown	illustrations
1923	Lofting, Hugh	The Voyages of Doctor Dolittle	yes	Hugh Lofting	color frontispiece, illustrations, color plates
1922	van Loon, Hendrik Willem	The Story of Mankind	yes	Hendrik Willem van Loon	illustrations (including maps; color plates

Appendix C: List of Newbery Medal-Winning Titles by Year Awarded

Year	Author	Title	Publisher
2016	de la Pena, Matt	Last Stop on Market Street	Putnam's
2015	Alexander, Kwame	The Crossover	Houghton Mifflin Harcourt
2014	DiCamillo, Kate	Flora & Ulysses: The Illuminated Adventures	Candlewick Press
2013	Applegate, Katherine	The One and Only Ivan	HarperCollins Children's Books
2012	Gantos, Jack	Dead End in Norvelt	Farrar Straus Giroux
2011	Vanderpool, Clare	Moon over Manifest	Delacorte Press
2010	Stead, Rebecca	When You Reach Me	Wendy Lamb Books
2009	Gaiman, Neil	The Graveyard Book	HarperCollins
2008	Schlitz, Laura Amy	Good Masters! Sweet Ladies! Voices from a Medieval Village	Candlewick Press
2007	Patron, Susan	The Higher Power of Lucky	Simon & Schuster / Richard Jackson
2006	Perkins, Lynne Rae	Criss Cross	Greenwillow Book
2005	Kadohata, Cynthia	Kira-Kira	Atheneum Books for Young Readers
2004	DiCamillo, Kate	The Tale of Despereaux: Being the Story of a Mouse, a Princess, Some Soup, and a Spool of Thread	Candlewick Press
2003	Avi	Crispin: The Cross of Lead (OCLC 48559447)	Hyperion Books for Children
2002	Park, Linda Sue	A Single Shard	Clarion Books
2001	Peck, Richard	A Year Down Yonder	Dial

2000	Curtis, Cheristopher Paul	Bud, Not Buddy	Delacorte
1999	Sachar, Louis	Holes	Frances Foster
1998	Hesse, Karen	Out of the Dust	Scholastic
1997	Konigsburg, E. L.	The View from Saturday	Jean Karl / Atheneum
1996	Cushman, Karen	The Midwife's Apprentice	Clarion Books
1995	Creech, Sharon	Walk Two Moons	HarperCollins
1994	Lowry, Lois	The Giver	Houghton
1993	Rylant, Cynthia	Missing May	Jackson/Orchard
1992	Naylor, Phyllis Reynolds	Shiloh	Atheneum
1991	Spinelli, Jerry	Maniac Magee	Little, Brown
1990	Lowry, Lois	Number the Stars	Houghton
1989	Fleischman, Paul	Joyful Noise: Poems for Two Voices	Harper
1988	Freedman, Russell	Lincoln: A Photobiography	Clarion
1987	Fleischman, Sid	The Whipping Boy	Greenwillow
1986	MacLachlan, Patricia	Sarah, Plain and Tall	Harper
1985	McKinley, Robin	The Hero and the Crown	Greenwillow
1984	Cleary, Beverly	Dear Mr. Henshaw	William Morrow
1983	Voigt, Cynthia	Dicey's Song	Atheneum
1982	Willard, Nancy	A Visit to William Blake's Inn: Poems for Innocent and Experienced Travelers	Harcourt
1981	Paterson, Katherine	Jacob Have I Loved	Crowell
1980	Blos, Joan W.	A Gathering of Days: A New England Girl's Journal, 1830-32	Scribner
1979	Raskin, Ellen	The Westing Game	Dutton
1978	Paterson, Katherine	Bridge to Terabithia	Crowell
1977	Taylor, Mildred D.	Roll of Thunder, Hear My Cry	Dial
1976	Cooper, Susan	The Grey King	McElderry / Atheneum
1975	Hamilton, Virginia	M. C. Higgins, the Great	Macmillan
1974	Fox, Paul	The Slave Dancer	Bradbury
1973	George, Jean Craighead	Julie of the Wolves	Harper
1972	O'Brien, Robert C.	Mrs. Frisby and the Rats of NIMH	Atheneum

1971	Byars, Betsy	Summer of the Swans	Viking
1970	Armstrong, William H.	Sounder	Harper
1969	Alexander, Lloyd	The High King	Holt
1968	Konigsburg, E. L.	From the Mixed-Up Files of Mrs. Basil E. Frankweiler	Atheneum
1967	Hunt, Irene	Up a Road Slowly	Follett
1966	Trevino, Elizabeth Borton	I, Juan de Pareja	Farrar
1965	Wojciechowska, Maia	Shadow of a Bull	Atheneum
1964	Neville, Emily	It's Like This, Cat	Harper
1963	L'Engle, Madeleine	A Wrinkle in Time	Farrar
1962	Speare, Elizabeth George	The Bronze Bow	Houghton
1961	O'Dell, Scott	Island of the Blue Dolphins	Houghton
1960	Krumgold, Josept	Onion John	Crowell
1959	Speare, Elizabeth George	The Witch of Blackbird Pond	Houghton
1958	Keith, Harold	Rifles for Watie	Crowell
1957	Sorensen, Virginia	Miracles on Maple Hill	Harcourt
1956	Latham, Jean Lee	Carry On, Mr. Bowditch	Houghton
1955	DeJong, Meindert	The Wheel on the School	Harper
1954	Krumgold, Joseph	…And Now Miguel	Crowell
1953	Clark, Ann Nolan	Secret of the Andes	Viking
1952	Estes, Eleanor	Ginger Pye	Harcourt
1951	Yates, Elizabeth	Amos Fortune, Free Man	Dutton
1950	de Angeli, Marguerite	The Door in the Wall	Doubleday
1949	Henry, Marguerite	King of the Wind	Rand McNally
1948	du Bois, William Pene	The Twenty-One Balloons	Viking
1947	Bailey, Carolyn Sherwin	Miss Hickory	Viking
1946	Lenski, Lois	Strawberry Girl	Lippincott
1945	Lawson, Robert	Rabbit Hill	Viking
1944	Forbes, Esther	Johnny Tremain	Houghton
1943	Gray, Elizabeth Janet	Adam of the Road	Viking
1942	Edmonds, Walter	The Matchlock Gun	Dodd
1941	Sperry, Armstrong	Call It Courage	Macmillan
1940	Daugherty, James	Daniel Boone	Viking
1939	Enright, Elizabeth	Thimble Summer	Rinehart
1938	Seredy, Kate	The White Stag	Viking

1937	Sawyer, Ruth	Roller Skates	Viking
1936	Brink, Carol Ryrie	Caddie Woodlawn	Macmillan
1935	Shannon, Monica	Dobry	Viking
1934	Meigs, Cornelia	Invincible Louisa: The Story of the Author of Little Women	Little, Brown
1933	Lewis, Elizabeth	Young Fu of the Upper Yangtze	Winston
1932	Armer, Laura Adams	Waterless Mountain	Longmans
1931	Coatsworth, Elizabeth	The Cat Who Went to Heaven	Macmillan
1930	Field, Rachel	Hitty, Her First Hundred Years	Macmillan
1929	Kelly, Eric P.	The Trumpeter of Krakow	Macmillan
1928	Mukerji, Dhan Gopal	Gay Neck, the Story of a Pigeon	Dutton
1927	James, Will	Smoky, the Cowhorse	Scribner
1926	Chrisman, Arthur Bowie	Shen of the Sea	Dutton
1925	Finger, Charles	Tales from Silver Lands	Doubleday
1924	Hawes, Charles	The Dark Frigate	Little, Brown
1923	Lofting, Hugh	The Voyages of Doctor Dolittle	Stokes
1922	van Loon, Hendrik Willem	The Story of Mankind	Liveright