



Article

Humanities Research, Book Digitization, and the Problem of Linguistic Change

Karen Sobel
University of Colorado Denver
Jeffrey Beall
University of Colorado Denver

Abstract

The good news is that millions of books have been digitized and are freely available over the Internet. The bad news is also that millions of books have been digitized and are freely available over the Internet. Linguistic change presents one of the greatest hurdles to information retrieval in databases of digitized books because keyword searching of digitized materials does not guarantee discoverability. This article examines the problem of linguistic change in humanities research in full-text databases and describes the innovative solution offered by two proprietary library content providers.

Humanities research now benefits from an abundance of digitized books. Massive digitization efforts by Google, the Internet Archive, and other organizations make the record of civilization more accessible than ever before. This benefit manifests itself in ease of access to online materials, free access to materials in the public domain, and the ability to search across and within works. However, massive book digitization is also beset by the problem of linguistic change. Searches that use contemporary keywords sometimes fail to deliver information that meets a searcher's needs because of semantic and orthographic change.

This paper will show how linguistic change affects humanities research of digitized books and will describe how innovations in the world of digitization can maximize the effective use of digitized materials. In some instances, words or phrases in the digitized text have meanings different from today's usage. In other instances, words in digitized books that are now obsolete may be unknown, and therefore not discoverable to the searcher. We look at linguistic change and its two components, semantic change and orthographic change. Our work focuses on the English language, although linguistic change affects researchers in many languages.

Linguistic change has always posed a challenge in information retrieval. Controlled vocabularies, like those used in library catalogs, deal effectively with the problem. But as information retrieval has shifted from the domain of the library to the domain of the World Wide Web, the problem has re-emerged. Information retrieval and searching in Google and the Internet Archive tend to utilize full-text searching with no controlled vocabulary. In this search method, searchers enter keywords and the search engine automatically and mechanically returns documents that contain the specific words. For example, a term that has become obsolete and replaced by another term is *chrestomathy*. A chrestomathy is a reader, that is, a selection of readings, often used in learning a second language. The term rarely appears in contemporary English where the preferred usage is *reader*. However, a searcher looking for readers in Google Books will miss those that call themselves a chrestomathy; an example is the book *Assyrian grammar with chrestomathy and glossary* (Mercer, 1921). Another example of a word with a changed meaning is *demagogue*. Originally, according to the *Oxford English Dictionary*, it meant "a leader of the people; a popular leader or orator who espoused the cause of the people against any other party in the state." Today, according to the OED, it means, "a political agitator who appeals to the passions and prejudices of the mob in order to obtain power or further his own interests" (*Demagogue, Oxford English Dictionary*). Full-text searches of digitized books will provide a researcher with all books that contain the term *demagogue*; yet the searcher likely only needs those with one specific meaning. Sorting through the results and extracting only the relevant hits involves a cost, and the searcher bears that cost; ideally the search engine would bear it.

Google and other organizations are carrying out book digitization on a massive scale. By one estimate, Google alone is digitizing ten million books a year, and there are probably about 65 million unique titles of printed books in existence (*The Economist*, 2007). In several years, essentially every published work will exist in electronic form, though many proprietary works will still require either affiliation with a library or personal payment to gain access. In the United States, most books published before 1923 have reached the public domain and are therefore freely available for copying and publishing. Massive book digitization represents one of civilization's greatest achievements, for it has unearthed and made freely available many millions of heretofore difficult-to-access and inaccessible books. Nevertheless, the scholarly applications of Google's and others' accomplishments in book digitization are made somewhat less effective by the nature of full-text searching. When a needed resource fails to turn up in search results,

the efforts expended in digitizing it are wasted, and its discovery by the searcher is unfulfilled.

Previous Studies

This article expands on many of the problems in full-text searching identified by Jeffrey Garrett (2006), who pointed out that "Over the last several years, full-text searching of large text corpora has placed an extraordinarily powerful tool in the hands of humanities students and scholars" (Abstract). Garrett explained that "orthographic irregularities in early modern English, which can lead to significant omissions or skewing of search results [and] the different vocabulary registers in 17th, 18th, even 19th century English" often throw off results of full-text searches (para. 15).

The subjects of full-text searching in Google Books and the evolution of the English language have been discussed extensively in the library and linguistic literatures. An example is Bair and Carlson (2008), who stated, "Abbreviations, obsolete and regional word usage, idioms, misspellings and alternate spellings, and omissions in primary sources make keyword searching difficult" (p. 249). The popular press has also widely discussed Google Books. A few articles in scholarly library publications have discussed the benefits of being able to search for rare words using Google Books. However, few have analyzed the challenges of using full-text searching in Google Books when terminology and spellings have changed since a given book was published.

Thomas Mann, a librarian at the Library of Congress, wrote about keyword searching within Google Books, pointing out the weaknesses of full-text searching of Google books in the context of the high-quality searching that library catalogs offer (Mann 2008). He explained that Google Book's search mechanism is "expressly designed and optimized for quick information-seeking rather than scholarship" (p. 159). He continued, "It fails to retrieve literature that uses keywords other than those the researcher can specify; it misses not only synonyms and variant phrases but also all relevant works in foreign languages. Searching by keywords is not the same as searching by conceptual categories" (p. 159). Mann then focused his criticism on relevance ranking's inability to lay the topic out in an organized manner, as a list of related subject headings can. He also discussed the difficulty of certain research concepts such as finding the earliest book published in French. He demonstrated that although this sort of question is common and important, it is practically impossible to effectively research by means of full-text searching.

Semantic Change

Many of the linguistic changes negatively affecting full-text searching fall under the category of *drift* (Gorrell, 1994), a broad term for semantic change. Linguist Robert Gorrell addressed the potential confusion caused by words that have retained the same or very similar spellings across centuries, but that have undergone dramatic changes in meaning. He discussed Sapir's "three drifts 'of major importance'": reduction of syntactic cases, word order replacing inflection to clarify grammar, and "the drift toward the in-

variable word” (p. 32). While his discussions come across as highly technical, his examples ring true to those who have studied pre-20th century English literature. For example, the word *pretty* experienced a marked shift in usage in the sixteenth century, when it became a popular word for expressing moderation (*pretty far*), in addition to a degree of beauty (Gorrell, 1994).

Nineteenth Century English: Stability and Change, edited by Kytö, Rydén, and Smitterberg, provides great detail on linguistic differences between 19th century and contemporary English. A variety of linguists analyze changes in many specific parts of speech, such as relativizers (indicators of relative clauses) and the subjunctive. Some parts are discussed in context, such as "the passive in nineteenth-century scientific writing" (Gustafsson, 2006, p. 110-135). *Nineteenth Century English* also describes the etiology of spelling variation in the nineteenth century and earlier, focusing on changes caused by evolving verb tenses, applications of adjectives and adverbs, and more. Familiar examples include changing usage of the words “more” and “most.” For example, the phrase “most happiest” once emphasized the subject’s degree of happiness (Kytö & Romaine, 2006, p. 195).

Linguistic change provides a greater challenge in digital humanities research than in the science, technology, and medicine (STM) fields. STM terminology is more stable because it employs standards such as binomial nomenclature in biology (e.g., *Sciurus carolinensis* for Eastern gray squirrel), a system that guards against linguistic change and promotes stable names over time and even among languages. Indeed, the linguostylistics of the humanities often promotes language change, for it is not uncommon for humanists to introduce new words or to add new meanings to existing words.

Orthographic Change

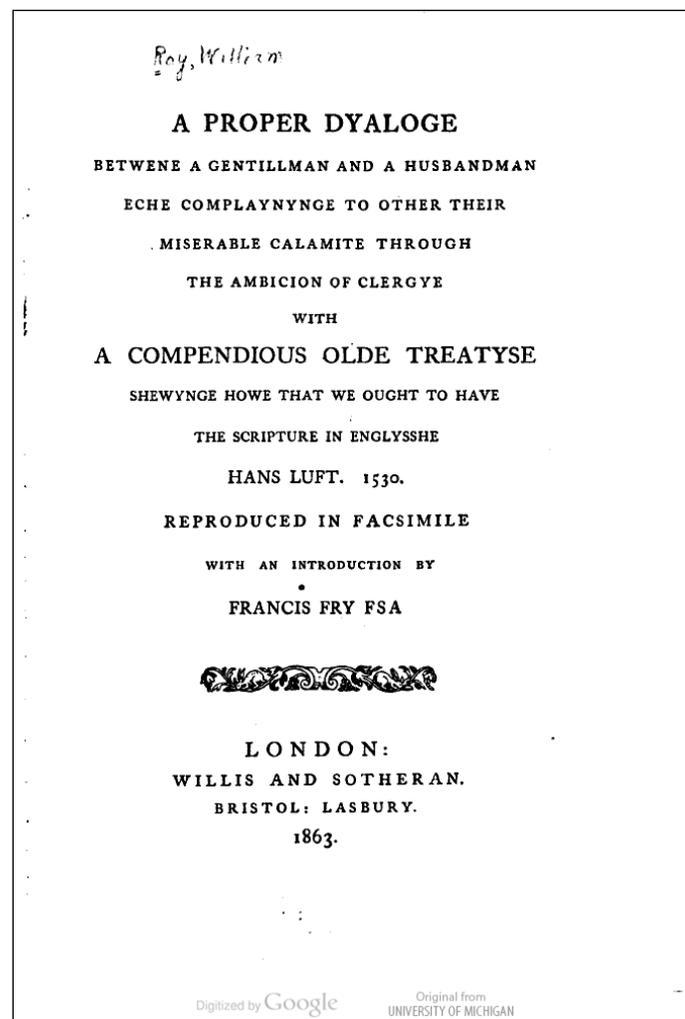
Variant spellings have always been a problem in full-text searching, and even contemporary language is beset by the problem. For example, the variant British and American spellings *labour* and *labor* can impair full-text searching when only one spelling of the word is retrieved in search results. Still, spelling changes over time and a search on a contemporary spelling of a word can result in an incomplete retrieval due to a variant spelling in digitized print documents that use an earlier spelling. An example is the word pair *tonight* / *tonite*.

Spelling change occurs purposefully, as in spelling reform, and gradually, as in the shift from *through* to *thru*. In the United States, publication of Noah Webster's *American Dictionary of the English Language* in 1828 "successfully altered the spelling of hundreds of words. For Americans, *gaol* became *jail*, *publick* became *public*, *travelled* became *traveled*, *centre* became *center*, and *valour* became *valor* -- all because of Webster's handiwork" (Wolman, 2008, p. 5). Webster's dictionary is the source of many variant American and British spellings. Prior to the dictionary's publication, British spelling predominated, but American spellings were creeping into usage. Webster's dictionary singlehandedly legitimized the new spellings, and they have been used ever since. In-

terestingly, the first edition of this dictionary, the 1828 edition, is not available in "full view" in either Google Books or the Internet Archive.

An extreme example of orthographic and semantic change we discovered recently is the book *A Proper Dyaloge Betwene a Gentillman and a Husbandman Eche Complaynyng to Other Their Miserable Calamite through the Ambicion of Clergye: With a Compendious Olde Treatyse Shewynge Howe That We Ought to Have the Scripture in Englysshe*. It is an 1863 publication of a 1530 manuscript. The spelling *Englysshe* is significant here, for it was common at the beginning of the modern period. Indeed, a search of this spelling of the word retrieves numerous hits in the Google Books database. Also significant is the word *husbandman*, an antiquated word for *farmer*. A search for *farmer* as a keyword won't retrieve this work because an antiquated synonym is used instead; researchers searching the keywords *social conditions* (miserable calamites) of *farmers* (husbandmen) of the *16th century* will miss this work. Figure 1 shows the title page of this work.

Figure 1. The title page facsimile reproduction of a 1530 manuscript. Reproduced from the Hathi Trust database (<http://www.hathitrust.org/>).



Linguistics, Searching and Digitized Books

Millie Jackson (2008) explored searching Google Books using a variety of metadata. She covered the topic from a distinctly library-focused perspective, asking such questions as "How does the metadata that Google supplies enhance research and assist researchers in locating what they need? [and] What can the library contribute from the bibliographic record and its metadata?" (p. 165). Jackson framed her discussion by comparing searches and metadata in Google Books with those in the University of Michigan Libraries' Mbooks (now Hathi Trust) project. Google Books does offer an "Advanced Book Search," that offers searching by title, author, publisher, and subject, but these searches are limited to complete and correct metadata, and this is a very small proportion of their overall collection of digitized books.

Jeffrey Beall has explored linguistic challenges and metadata in the past, with his work on the "synonym problem." He identified a number of linguistic issues that are particularly problematic for searching, including differences in British and American spellings (Beall & Kafadar, 2008), false synonyms such as "waterfall" and "cascade," and synonyms with no overlapping words, such as *dentures* and *false teeth*. He presented these challenges in the context of Internet search engines such as Yahoo. Beall (2008) also identified the homonym problem which he says "occurs in full-text searching when a single word or phrase has more than one meaning" (p. 440). Thus the homonym problem in full-text searching can be a manifestation of the linguistic shift described above. The two meanings of *demagogue* described above are an example of the problem. When a single word names two different concepts, it causes imprecise retrievals in full-text searching. Typically a searcher only wants one of the two meanings; for example, jaguar the animal, not the car. The additional, unwanted entries in the search results serve only as noise and add to the cost of the search process in time spent sorting through the results to extract only the relevant ones.

Query Expansion as a Solution

Advances in search engines ameliorate some of the problems inherent in full-text searching, especially basic ones such as spelling variation. One advance is query expansion, which occurs when a search engine automatically searches for more than what the searcher entered in the search box. Query expansion works well as a solution to orthographic variation. For example, whenever a searcher enters a search that includes the term *dialog*, a search engine may be programmed to simultaneously search *dialogue* and then combine the results in a relevancy-ranked order. Search engines can also use query expansion to simultaneously search different forms of a given term. For example, a search on *poem* would also retrieve results containing *poetry*, *poetic*, *poetics*, etc. Advanced systems employing query expansion might search synonyms of the term a searcher enters in the search box. For example, a search on *frescoes* might be augmented with results containing the term *murals* as if the searcher had included both terms in the search. The problem with query expansion is that the searcher may not

want all the extra results included with the search results. Therefore, query expansion can greatly decrease search precision.

Several proprietary databases have incorporated query expansion functionality to help resolve the problems we describe in this paper. One is the Early English Books Online (EEBO) database (<http://eebo.chadwyck.com>). Both the basic and advanced search pages of this product offer a feature that searches variant spellings of a word or words in a search box. The site provides a checkbox that says "Search using variant spellings" that is checked by default. Additionally, after entering a search term in the search box, the user is presented with a link that says "Check for variants." This opens up a dialog box that shows all variant spellings for the term and allows the user to check or un-check each one for inclusion in the search. For example, we entered the term *milk* in the search box. Upon clicking on the "check for variants" link, we were presented with the spellings *milk*, *milke*, *mylke*, *mylk*, and *milck*. In addition the database also presented forms and spellings of the word as a verb, with the option of checking each spelling for inclusion in the search. The database offers a total of 14 variant spellings of *milk*. This type of query expansion will certainly ameliorate the variant spelling problem for users.

EEBO also searches typographical variants when the "search variant spellings" box is checked. For example, in early modern texts, the word *love* sometimes appears as *loue*. The work to find and link the early spellings with the modern ones has been done as a part of the Virtual Orthographic Standardization Project at Northwestern University. Computers are unable to reliably complete the work; so much of it has been done manually, using the labor of student workers. The variant spellings functionality in EEBO is unique.

Gale Cengage Learning's *Eighteenth Century Collections Online* (ECCO), Part I and Part II, also includes a variant spellings searching functionality. However, unlike the manual solution created in the EEBO database, the ECCO solution uses algorithms to connect variant spellings in the database to the terms users enter in the search box. This feature is probabilistic, which means it lacks the precision that is achieved by human connection of variant spellings as in the EEBO database. The ECCO help page states:

A "fuzzy search" option expands the search to include words similar to search terms entered, thereby accommodating for spelling variations.

Think of a fuzzy search in terms of how similar one word is to another. To change a word into another, characters can be added, deleted or replaced. A single degree is one change to one character. Two degree is changes to two characters, and so on up to four. The higher the degree, the greater margin of error, presenting the researcher with more results (Gale Cengage Learning).

As the description indicates, this method has the drawback of decreasing search precision by generating results that do not always match the searcher's need.

The two different solutions employed by EEBO and ECCO show several things. EEBO's solution is deterministic, relying on human work to properly code the millions of variant spellings in its collection. ECCO's solution is probabilistic, an approach that avoids the costs of a manual solution but one that introduces probability into the equation: sometimes it provides incorrect results. If there were a single, effective solution, many vendors would have applied it by now. Both of these competing databases are expensive for libraries to license, and the revenue each generates has funded work and research to make their search functionality among the best available. These databases are the exceptions; most full-text databases lack the functionality needed to search variant spellings or obsolete, synonymous terms.

EEBO's decision to use a manual approach to solving variant spellings mirrors the approach that library cataloging has taken for over a century. Manual work linking variant spellings and word meanings, and obsolete terms with contemporary ones is the most reliable solution. Computers have not advanced to the stage where they have the ability to algorithmically achieve what trained catalogers achieve every day.

Strengths of Full-Text Searching

Full-text searching proves highly beneficial to scholars searching for historical information on obscure terms. They may also find that full-text searching is a relatively effective way of finding information on small phenomena that have not yet inspired an entire work, or even merited a subject heading. Jackson (2008) compared searching a book's full text versus catalog records and finds that for obsolete terms such as *pin money* (an allowance a husband gives to his wife, a term common in the nineteenth century), searching the full text is better.

Examples of How Linguistic Change can Affect Retrieval

The following sections describe examples that list the different categories of linguistic change that can affect retrieval in searching digitized books for humanities research. Some of the categories are overlapping; they are separated out to give emphasis to the different types of linguistic change that are problematic in research.

Obsolete terms.

Obsolete words and phrases can cause a problem in searching when searchers use contemporary terms rather than the terminology of the times being researched. Full-text searching groups digitized books together by the words contained in the content, not by the subject matter, thereby inhibiting discoverability for many researchers. The example already given, *chrestomathy*, is an example of an obsolete term. Another example is *photoplay*, which is an early term for motion picture, or movie. Early books on motion picture authorship included works with titles such as *The Technique of the Photoplay*

(Sargent, 1916), and *The Art of Photoplay Making* (Freeburg, 1918). Anyone researching the early history of motion picture authorship that searches using the terms *movies* or *motion pictures* might miss these works.

Homonyms with obsolete meanings.

Another challenge is the homonym problem in which a modern word had a secondary meaning in the past, which is no longer in use. Full-text searching for the term may retrieve many false hits. One example is the word *terrier*, which according to the *Oxford English Dictionary* used to mean:

A register of landed property, formerly including lists of vassals and tenants, with particulars of their holdings, services, and rents; a rent-roll; in later use, a book in which the lands of a private person, or of a corporation civil or ecclesiastical, are described by their site, boundaries, acreage, etc. Also, in extended application, an inventory of property or goods (Terrier, *Oxford English Dictionary*)

In contemporary English of course, the word *terrier* refers to a variety of dog. A searcher looking for books containing the older meaning of the term will be hounded by irrelevant results that are about small, hyperactive dogs. Another example is the word *very*, which was frequently used to indicate genuineness, such as in the phrase “the veriest knave I ever knew” (Gorrell, 1994, p. 43).

Spelling changes.

The contemporary spelling of a word may not match the spelling in many digitized texts from the early twentieth century and earlier, a situation that can lead to missed retrieval. The 1863 reproduction of the 1530 manuscript described earlier, *A proper dialoge betwene a gentillman ...* is a striking example of orthographic change, saliently contrasting the archaic spelling of the late middle ages (or perhaps, the early modern period) with current, highly standardized spellings. Spelling variation also occurs due to regional or dialectical variation (*catalog*, *catalogue*), and it occurs slowly as in the current change from *through* to *thru*.

Typographic ligatures (such as Æ and Œ), sounds symbolized by two letters linked together, are a common and problematic example. They appear frequently in Latin-based words, including some formerly used in English. They are also used in modern languages such as Norwegian. While in most contexts there is a formal or informal expectation that a ligature can be written with the two letters broken apart (for example, æ as ae), Google Books does not have a specific command for dealing with such spelling issues. This issue affects searches for some common words, such as *onomatopoea* (originally spelled *onomatopœia*), as well as the name Æsop (often spelled Aesop). Words like *encyclopedia* can be spelled three ways: *encyclopedia*, *encyclopaedia*, and *encyclopædia*. Search engines each have their own heuristic for dealing with these digraphs; some may search for words that contain only the digraph, and others may

search for words that contain the digraph and words that contain the two separate letters.

Also, typographical differences in the representation of certain letters have occurred over time. For example, the letter *w* in some early printed texts was recorded using two instances of the letter *v*. A practical example of this is the book with the title *An abstract or [sic] the lavves of Newv England, as they are novv established* (Cotton, 1641). In many contemporary search systems, a digitized version of this text would not be retrieved using contemporary spellings, for there is no match between the search and the indexed content. Similar problems occur with old typographical practices involving the letters *v*, *u*, *l*, and *j*. In some old texts, for example, the country *Ireland* is spelled *Jre-land*. Search interfaces that employ query expansion or fuzzy intelligence may in many cases be able to ameliorate the problems these typographical variations cause, but those systems are few, the programming can be imperfect, and a computer-expanded query may decrease search precision by adding irrelevant results to the search query.

Ethnic and racial groups.

Humanities research often covers the artistic and literary endeavors of various ethnic groups. Ethnic group terminology is more subject to change than any other terminology. Take the term *African Americans*, for example. In the last century in the United States, members of this ethnic group have been called *Colored*, *Colored people*, *Negroes*, *Blacks*, *Afro-Americans*, and *African-Americans*. A student searching for early examples of African-American poetry who only searches using the term *African-American* in Google books will miss important (and free) resources such as *The American Book of Negro Poetry* (Johnson, 1922). Because the term *African-American* does not occur in the text of this book, it is not retrieved in a full-text search in Google Books.

Other ethnic groups fare similarly; for example, the Iroquois Indians have also been known as the *Agoneaseah*, the *Massawomeke*, and the *Mengwe* Indians. Books tend to employ a single term, so searchers using full-text searching of textual corpora will need to enter all variants to ensure complete results.

Artistic and literary movements.

The Aesthetic Movement in literature occurred in Europe in the late 19th century. Like many movements, the name that came to describe it was coined after it was over. Contemporary literature and documents describing the movement probably didn't use the term "Aesthetic Movement" to describe it; the term was created later. So a search in a full-text database of digitized works seeking information on the Aesthetic Movement will naturally be limited to works published after the movement itself. Such a search will filter out works contemporary to the movement itself because it wasn't called that at the time. An exception might be the Decadence Movement, which got its name from early, contemporary critics.

This problem occurs for any movement or event that is named long after it occurs or starts. During the period between the time when the event starts and gets named, all the publications about it lack the defining name. Thus primary and contemporary sources are less likely to occur in search results when someone is searching for information about an event or named artistic or literary movement. For example, contemporary works on the Seven Years' War certainly didn't call it that, for no one at the time knew how long it was going to last. The name was coined after the war ended.

Conclusion

Humanities scholarship is bolstered by the sudden abundance of millions of easily-accessible digitized books and other works, but linguistic change threatens to make many of these works just as hidden and inaccessible as they were before they were digitized. A few proprietary library content providers have developed solutions to the problem of orthographic change, but these innovations are rare. Most full-text databases are unable to collocate variant spellings of a given word in search results, and semantic change in language still presents a problem in most databases. Humanists and students of the humanities still need to take semantic and orthographic change into account in their research strategies, and novice researchers need an understanding of semantic and spelling changes and how they affect searching. Linguistic change demonstrates the richness of human language yet remains one of information retrieval's greatest challenges.

References

- Bair, S., and Carlson, S. (2008). Where keywords fail: Using metadata to facilitate digital humanities scholarship. *Journal of Library Metadata* 8(3), 249-262.
- Beall, J. (2008). The weaknesses of full-text searching. *Journal of Academic Librarianship* 34(5), 438-444.
- Beall, J., and Kafadar, K. (2008). Measuring the extent of the synonym problem in full-text searching. *Evidence Based Library and Information Practice* 3(4), 18-33. Retrieved from <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/4081>
- Cotton, J. (1641). *An abstract or [sic] the lavves of Nevv England, as they are novv established*. London: F. Coules, and W. Ley at Paules Chain.
- Demagogue (1989). In *Oxford English Dictionary* (2nd edition). Retrieved June 21, 2011 from <http://www.oed.com>.
- Fairman, T. (2006). Words in English record office documents of the early 1800s.

- In M. Kytö, M. Rydén, & E. Smitterberg (Eds.) *Nineteenth century English: Stability and change* (pp. 56-88). New York: Cambridge University Press.
- Freeburg, V.O. (1918). *The art of photoplay making*. New York: Macmillan.
- Gale Cengage Learning. (2010). *Eighteenth Century Collections Online, Part I and Part II: Search Functionality*. Retrieved from <http://gdc.gale.com/products/eighteenth-century-collections-online/evaluate/search-functionality>
- Garrett, J. (2006). KWIC and dirty? Human cognition and the claims of full-text searching. *Journal of Electronic Publishing*: 9, no. 1. doi: [10.3998/3336451.0009.106](https://doi.org/10.3998/3336451.0009.106)
- Gorrell, R. (1994). Language change, usage and drift. *English Today*, 10 (3), 42-44.
- Gustafsson, L. O. (2006). The passive in nineteenth-century scientific writing. In M. Kytö, M. Rydén, & E. Smitterberg (Eds.), *Nineteenth century English: Stability and change* (pp.110-135). New York: Cambridge University Press.
- Jackson, M. (2008). Using metadata to discover the buried treasure in Google Book Search. *Journal of Library Administration* 47(1/2), 165-173.
- Johnson, J. W. (1922). *The Book of American Negro Poetry*. New York: Harcourt, Brace.
- Johansson, C. (2006). Adjective comparison in nineteenth-century English. In M. Kytö, M. Rydén, & E. Smitterberg (Eds.), *Nineteenth century English: Stability and change* (pp.194-214). New York: Cambridge University Press.
- Mann, T. (2008). Will Google's keyword searching eliminate the need for LC cataloging and classification?" *Journal of Library Metadata* 8(2), 159-168.
- Mercer, S. A. B. (1921). *Assyrian grammar with chrestomathy and glossary*. London: Luzac & Co.
- Not bound by anything. (2007, March 24). *The Economist* 382 (8521), 93-94.
- Sargent, E. W. (1916). *The technique of the photoplay*. New York: The Moving Picture World.
- Terrier (1989). In *Oxford English Dictionary* (2nd edition). Retrieved from <http://www.oed.com>.
- Wolman, D. (2008). *Righting the mother tongue: From Olde English to email, the tangled story of English spelling*. Washington, D.C.: Smithsonian Books.

Karen Sobel is a Research and Instruction Librarian at the Auraria Library, University of Colorado Denver. The majority of her research focuses on library instruction for first-year college students. For more details, please see her home page at: <http://library.auraria.edu/~ksobel>.

Jeffrey Beall is a Metadata Librarian at Auraria Library, University of Colorado Denver. His research explores the weaknesses of full-text searching and the strengths of meta-data-enabled searching. His home page is at: <http://carbon.ucdenver.edu/~jbeall/>

©2011, K. Sobel, J. Beall. *Journal of Library Innovation* is an open access journal. Authors retain the copyright to their work under the terms of the following Creative Commons license: Attribution-Noncommercial-No Derivative Works 3.0 (United States) <http://creativecommons.org/licenses/by-nc-nd/3.0/us/>