



# ***The IALLT Journal***

*A publication of the International Association for Language Learning Technology*

## **LEGAL ISSUES & LLT**



Judy Shoaf  
*University of Florida*

### **GOOGLIZING COPYRIGHT**

In November 2013, Judge Denny Chin of the Southern District of New York filed a summary judgment against Authors Guild, which had sued Google for its massive text-scanning program.<sup>1</sup> He followed Judge Harold Baer's similar judgment, in the same court in October 2012, favoring HathiTrust, a nonprofit educational group which offers access to an alternative version of Google Books'<sup>2</sup> database of scanned texts.<sup>3</sup> Both judges relied on the argument that the

---

<sup>1</sup> Judge Chin's ruling of 11/14/13: <http://www.documentcloud.org/documents/834877-google-books-ruling-on-fair-use.html>.

<sup>2</sup> Google's database of digitized books has gone under a number of names: Google Print, Google Library Partners, Google Book Search, Google Books. The more commercial entity has been Google Editions or Google Ebooks, now part of Google Play: <https://play.google.com/store/books>. Book search and book sales functions are closely linked now.

<sup>3</sup> Judge Baer's ruling of 10/10/12: [http://www.tc.umn.edu/~nasims/HathivAG10\\_10\\_12.pdf](http://www.tc.umn.edu/~nasims/HathivAG10_10_12.pdf). HathiTrust is a partnership of academic institutions which curates a growing searchable digital

gigantic, searchable database of texts was *transformative* in nature, and therefore fair use of the originals.<sup>4</sup> This is a complex case with complex issues, going right to the core of the relationship between physical media—libraries, in this case—and “the cloud.” It has been heavily discussed since 2004 and will continue to be of the greatest interest as Google exploits the court decision to keep scanning, developing ever larger databases of good sentences in every language.

### *Spilt Milk*

Google Books arose from the decision—not only of Google but of the research university libraries which loaned it the books to scan—to scan first and ask permission later. (Authors who discovered that their works had been digitized could “opt out” by contacting Google.<sup>5</sup>) The right to make and keep a digital copy belongs to the copyright holder, not to the owner (or borrower) of a particular copy. Digitizing or otherwise copying a movie, song, or book is not made legal by this decision. *But size matters*: a huge database of searchable materials is more than the sum of its parts.

Google began in the early 2000s by working with publishers, who provided or authorized scans so that Google could offer full-text searches and “previews” of selected pages, linked to sites where the book could be purchased. This project evolved into “let’s scan all the books in the world,”<sup>6</sup> beginning with books loaned by university libraries, which Google compensated by giving the libraries digital copies of their own books. While some libraries loaned only public-domain books, others offered all their books, without regard to copyright status. This is what, in 2005, triggered the lawsuits from Authors Guild (and also the American Association of Publishers).

There are things you can do with a physical book that you can’t do with Google Books—and its search index does not replace a librarian’s analysis of holdings. But Google Books offers unique possibilities for research on specific

---

database which includes books originally loaned to Google and returned in this form. Its practices with respect to copyrighted items are similar to those of Google Books at the time of Judge Chin’s decision regarding Fair Use.

<sup>4</sup> Authors Guild is appealing both cases.

<sup>5</sup> This is comparable to the legal mechanism for removing copyrighted material from YouTube. However, in the case of Google Books each book has gone through the hands of a single entity (Google), whereas YouTube videos come from many sources. The role of librarians in selecting books to be scanned may also be questioned.

<sup>6</sup> After about 9 years, Google has “over 20,000,000” individual titles in its database, or possibly over 30,000,000. In 2010, Google estimated that there are about 130 million books in the world; <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html>.

topics, and particularly for a statistical understanding of the evolution of terms and usage. Consider the Ngram,<sup>7</sup> which allows quantitative tracking over time of distinctive names or phrases. Judge Chin commented, “Google Books has become such an important tool for researchers and librarians that it has been integrated into the educational system—it is taught as part of the information literacy curriculum to students at all levels” (11/14/13, p. 10).

The long lawsuit has allowed Google Books to reach a critical size, and allowed both the Google mathematicians and casual users to explore its value, which depends on that size. Likewise, HathiTrust is so strong because it combines the scanned works from many libraries. Both judges emphasized the transformative nature of the databases. The genie is not going back into the bottle.

### ***Who Benefits? And Who Profits?***

In these cases, one can identify three parties whose needs the judges had to consider:

1. The public. Besides the research possibilities I have mentioned, scanned copies of texts can be converted into tactile or auditory books for those who can’t use the original books, a notable advantage.

2. Authors and publishers, libraries and booksellers. Google Books develops a market for books which readers identify in searching; the interface includes links to sales websites. Authors Guild can, though, complain that Google and HathiTrust have prevented them from collecting fees paid for the act of making and keeping a copy. The biggest problem is that the huge scope of Google Books means that this party includes *all copyright holders in the world*.

3. Google. Google has several plans for monetizing its digital library by selling tailored advertising, subscriptions to the database itself, and copies of out-of-print books. These plans have had to change in the course of the lawsuit, as Google backed away from profit to emphasize Fair Use. Other benefits to Google may include: tracking users to develop profiles for targeted advertising; using the huge database of well-written sentences to develop its statistics-based Google Translate program; aspects of Artificial Intelligence training.<sup>8</sup>

---

<sup>7</sup>See <https://books.google.com/ngrams>.

<sup>8</sup> See Ken Hillis, Michael Petit, and Kylie Jarrett, *Google and the Culture of Search* (Routledge, 2012), 151; and <http://research.google.com/pubs/ArtificialIntelligenceandMachineLearning.html>.

### ***Settlement vs. Fair Use***

Ironically, the judge's decision in Google's favor, while it leaves a lovely glow on the company's brand and plans, raises questions about how it can use its database.

Google and Authors Guild in fact reached a Settlement Agreement in 2008, but it was rejected (even in a 2009 revised version) by Judge Chin. Google would pay damages and costs (an acknowledgement of sorts that the copying was not Fair Use) and set up a system whereby Google would sell both subscriptions to Google Books and copies of individual works, paying royalties to authors; a fund would be set up for royalties pertaining to *orphan works* (see below). Authors Guild, as one party to the Settlement, represents all the copyright holders (including, for example, the heirs of the author of some obscure out-of-print book), but it can't in fact speak on behalf of foreign authors. (Google had been challenged by the China Written Works Copyright Society (CWWCS) and blocked from scanning French books by French courts).<sup>9</sup> Moreover, the Settlement would "make Google the sole vendor of most of the books published in the twentieth century."<sup>10</sup> Although other sellers would not be excluded from arranging to publish out-of-print books, Google would be in the best position to do so, as a party to the Amended Settlement.<sup>11</sup>

Judge Chin, having rejected the Settlement and the Amended Settlement, considered the four Fair Use factors carefully in his judgment on the case. Factor 1 has to do with the nature of the use, and the key terms here are educational (fair), transformative (fair), and commercial (against fairness). Judge Chin emphasized the transformative nature of Google Books, which in principle makes of the database a new work; he also commented on its educational importance. With respect to Google's for-profit status, the judge noted that Google "does not engage in the direct commercialization of copyrighted works" by selling scans or using the interface for advertising revenue—though in fact Google had planned to do precisely that. Google also points out that a high proportion of the works it has scanned—and especially of those being searched—is non-fiction, which is in

---

<sup>9</sup> See <http://www.forbes.com/2009/10/21/chinese-writer-google-markets-copyrights.html>; and <http://articles.latimes.com/2009/dec/19/world/la-fg-france-google19-2009dec19>.

<sup>10</sup> Siva Vaidhyanathan, *The Googlization of Everything (And Why We Should Worry)* (University of California Press, 2012), 154. Vaidhyanathan's argument is that a database of this kind should be a public trust, not a private venture.

<sup>11</sup> A handy overview of the issue from the Google founders' point of view is Sergey Brin, "A Library to Last Forever," *New York Times*, Opinion, 11/8/2009, <http://www.nytimes.com/2009/10/09/opinion/09brin.html>.

its favor also (factor 2). Although 100% of any book can be searched, works for which Google has no copyright permission cannot be read online—all one sees is a line or two of text at most, and the page number for reference (factor 3); the next step is to locate a hard copy at a library or bookseller. In fact, from the very beginning the interface has directed users who want to purchase a book to an appropriate site, thus enhancing rather than detracting from the market value of the book (factor 4). Judge Chin could conclude that Google “advances the progress of the arts and sciences, while maintaining respectful consideration for the rights of authors and other creative individuals, and without adversely impacting the rights of copyright holders” (p. 26).

Judge Chin thus establishes that Google did not infringe copyright by scanning, indexing, and maintaining the database of books copyrighted in the U.S. without asking the permission of the authors.<sup>12</sup> Moreover, because it is a “transformative” use, Google acquires rights in the ever-growing database itself (just as 2 Live Crew acquired rights to its parody version of Roy Orbison’s “Pretty Woman”<sup>13</sup>). Although he expressed concern in his opinion on the Amended Settlement that Google would have a monopoly in selling subscriptions to such a database, Judge Chin does not consider that problem in the Fair Use decision.

### ***Orphan Works***

I want to say a few words about the concept of the orphan work, because it’s relevant to LLTI. Orphans are works under copyright whose rights holder cannot be contacted for permission to copy or use them. The out-of-print or hard-to-find book, the VHS tape of a movie for which no DVD has been released (or none subtitled in English), the ancient copy of a still-useful audio recording—these are examples of orphans I at least have run into.

Media fall into three big categories:

1. Works in the public domain, which can be freely copied; works published in the US before 1923 qualify, as well as some published 1923-1963.

---

<sup>12</sup> Compare the NSA and its databases of information, as revealed by Edward Snowden in 2013; collecting, indexing, storing, and searching the recordings of private conversations by American citizens is not presumed to require a court order; accessing the conversations themselves requires one. See *New York Times* 6/13/13, “NSA Chief Says Phone Record Logs Halted Terror Threats,” <http://www.nytimes.com/2013/06/13/us/nsa-chief-says-phone-record-logs-halted-terror-threats.html>.

<sup>13</sup> *Campbell v. Acuff-Rose Music, Inc.*

2. Copyrighted works whose rights owners watch like a hawk to make sure they control any use of the work.
3. Works published after 1923 which may or may not still be under copyright, and whose rights owners are hard to identify.

Group 3 is the “orphans.” Of course, the very existence of Google searches and social media has made it easier to find authors, publishers, and so on, in order to ask permission to copy a work. But the original author may have died, or the work may have been done for hire, or some other event may have obscured the question of whom to contact about an out-of-print work. But isn’t it in the public interest for educators and others to be able to use these works?

Librarians, editors, and authors who wish to prepare derivative works are bound to “due diligence,” a genuine effort to contact the copyright holders of works they want to use. Judge Chin’s rejection of the Settlement with Authors Guild emphasized that the Authors Guild could not represent authors with whom it had no contact at all. It was not enough for Google, with or without Authors Guild support, to say that a rights holder could “opt out” of Google’s planned books-on-demand business; rather, the rights holder had to be asked for permission.

Judge Chin’s November 2013 Fair Use decision does not authorize Google to sell copies of individual works without permission, or even to show more than the “snippet view” without permission. Only by locating the book’s rights holder and reaching an agreement with him or her can Google (or another group, e.g. HathiTrust) offer more than a snippet view. As the judge said in rejecting the Settlement in 2011, “A copyright owner’s right to exclude others from using his property is fundamental and beyond dispute.”<sup>14</sup>

Orphan works continue to pose a problem for librarians, educators, and other users. Judge Chin wrestled with the issue in his discussion of the Amended Settlement Agreement, but finally referred the problem to Congress, saying that the courts can’t make a useful decision based on current law (p.23).

## ABOUT THE AUTHOR

Judy Shoaf has a Ph.D. in French and Medieval Literature from Cornell. She has directed the Language Learning Center at the University of Florida since 1993.

---

<sup>14</sup> Judge Chin’s rejection of the Amended Settlement Agreement, 3/22/2011, [http://thepublicindex.org/docs/amended\\_settlement/opinion.pdf](http://thepublicindex.org/docs/amended_settlement/opinion.pdf), p. 32.

*Legal Issues & LLT*

She maintains a website on copyright law & educational media that can be accessed at <http://www.clas.ufl.edu/lc/Copyright>.

### **ABOUT THE COLUMN**

Legal Issues & LLT is a column dedicated to examination of the legal considerations of copyright, fair use and ownership within the context of language teaching and l