#DLFteach • #DLFTeach Toolkit: Lesson Plans for Digital Library Instruction

Using EEBO-TCP in Research & Teaching

Heather Froehlich

Published on: Sep 19, 2019

Updated on: Aug 12, 2019

DOI: 10.21428/65a6243c.c4ffecb3

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Session Specifics

- One-shot, workshop, semester-long class, faculty collaboration, online learning, or other
- 2 hours

Instructional Partners

- Disciplinary Faculty member (English, History, both likely partners; other potential partners include Linguistics/English Language, Information Science)
- Digital Humanities Librarian (and/or liaison librarian)

Audience

Graduate students / faculty (advanced researchers); advanced undergraduates in a classroom setting

Curricular Context

Early English Books Online (EEBO) is widely taught in graduate seminars and is widely used in discussions of studying the early print record. EEBO-TCP (Early English Books Online Text Creation Partnership, full-text transcriptions of 25k records available in the public domain) is available and ripe for use by scholars, especially those not necessarily allied with digital humanities. However, knowledge about how to access, use, and search the full corpus of transcriptions is not widespread. This has the potential to really radically change how literary-historical scholarship is researched and produced, but only if scholars are taught about the access points and possibilities for the dataset.

In addition: some basic technological skills, including corpus linguistic methods (specifically, keyword in context analysis) and the use of digital interfaces to search for individual specific items will be introduced.

Each section builds on the previous, but can be presented in discrete chunks (e.g. find a text using TCP & compare to the Chadwyck Healy/JISC Historical Texts scanned image as one class lesson; searching texts as a second followup class lesson, etc).

Learning Outcomes

1. Understand the relationship between EEBO and EEBO-TCP;
2. Find and read individual texts in the TCP repository;
3. Observe frequency and context for specific words and phrases using the Corpus Query Processor (CQP) web interface to the TCP corpus;
4. Consider ways these resources can be applied towards a research question which can be answered by observing lexical items in context

Preparation

Audience

- Read Diana Kichuk, "[Metamorphosis: Remediation in Early English Books Online](#) (EEBO)," *Literary and Linguistic Computing* 22, no. 3 (2007) and Michael Gavin's "[How To Think about EEBO](#)" in *Textual Cultures* Vol 11, nos 1-2 (2019).
- Come prepared for the activity with some specific keywords which are of interest to the researcher (some potential examples include terms about religion; animals; monarchy; etc) for investigation.
- Preregister with the [Corpus Query Processor](#) using your .edu email or other institutionally-affiliated address. If interested, try out some sample searches for fun.

Instructor

- Familiarize themselves with the Corpus Query Processor (CQP) syntax (see [tutorials](#))
- Familiarize themselves with the [history of EEBO-TCP](#), Diana Kichuk, "[Metamorphosis: Remediation in Early English Books Online](#) (EEBO)," *Literary and Linguistic Computing* 22, no. 3 (2007): 295 and Michael Gavin's "[How To Think about EEBO](#)" in *Textual Cultures* Vol 11, nos 1-2 (2019).

This lesson plan pairs nicely with two essays cited in the [Folgerpedia entry](#) for Early English English Books Online:

- Ian Gadd, "The Use and Misuse of Early English Books Online," *Literature Compass* 6, no. 3 (2009): 682. DOI: [10.1111/j.1741-4113.2009.00632.x](https://doi.org/10.1111/j.1741-4113.2009.00632.x)
- Bonnie Mak, "[Archaeology of a Digitization](#)," *Journal of the Association for Information Science and Technology* 65, no. 8 (2014): 1515–1526, preprint pdf p. 22. DOI: 10.1002/asi.23061 18 (preprint).

In addition, reviewing the EEBO-TCP [working group notes](#) would be beneficial for understanding which specific decisions were made in the transformation of digitized materials into a machine-readable format.

Materials

- Computer station w/ projector (instructor)
- Laptops or access to a computer laboratory (participants)
- Optional Handout/Worksheet available under Additional Instructional Materials below, a step-by-step for the audience to move ahead if they are comfortable, while also offering an ability to have instructions in front of audience for note taking purposes.

Session Outline

Introduction

1. **EEBO and the TCP initiative** (15-20 min)
 - 125,000 mostly English works printed between 1473-1700, available as a subscription service from ProQuest/Chadwyck Healey. Potted history of EEBO for survey: http://folgerpedia.folger.edu/History_of_Early_English_Books_Online
 - A book is microfilmed microfilm image is digitized and made available online. Use Kichuk (2007) to guide this discussion
 - Hand-transcriptions by non-native speakers of the digitized images available as part of *Early English Books Online*
 - not without its flaws...but better than anything else we've got!

- Phase I: 25,368 texts entered the public domain on 1 Jan 2015 (an additional 40k forthcoming in 2020)

2. Additional Resources

- Official page <http://www.textcreationpartnership.org/tcp-eebo/>
- Transcription guidelines & other documentation
<http://www.textcreationpartnership.org/docs/>
- EEBO-TCP Tagging Cheatsheet: Alphabetical list of tags with brief descriptions
<http://www.textcreationpartnership.org/docs/dox/cheat.html>
- Text Creation Partnership Character Entity List
<http://www.textcreationpartnership.org/docs/code/charmap.htm>

Searching EEBO-TCP

1. Two access points to EEBO-TCP (10 min)

- Transcriptions vs. corpus serve different purposes: the Lancaster University Corpus Query Processer (CQPweb) corpus interface allows for more robust in-text searchability than the UMich transcriptions do (but transcriptions help you find more specific contextual information).
- Full text transcription repository apparatus available online using the [UMich EEBO-TCP](#) page.
- [CQPWeb](#) interface for EEBO-TCP (requires login, account registration is freely available).

2. Hands-on Activities: finding and using specific information in the TCP UMich repository and CQPWeb (45 min total)

- **Practice finding a specific transcription (10 min)**
- STC number vs ESTC number vs TCPID number
- STC = specific book a transcription is from; ESTC number = “English Short Title Catalogue” (see http://estc.bl.uk/F/?func=file&file_name=login-bl-estc)
TCPID = specific transcription
- Use Florio (1598) as example: STC number **11098** // TCP number A00991; “click view entire text” button; get full text.
- **Search the full-text corpus of EEBO-TCP (15 min)**

- [CPQweb EEBO search](#) (Must be logged in)
 - Example search: Friday
 - Click any of the 'filename' TCPIDs to get all available metadata
 - Go back to the UMich site, type in the TCPID and hit search to find that specific transcribed object again
 - Table of Contents > view entire text and control-F to search for your term/line to start reading
 - **In pairs/small group:** Find and record the results you get for... (15 min)
 - a specific word: king
 - A form of a word: {run}
 - Wildcard searches: s?ng, *able
 - Part of speech: *ly_RR (adjectives ending in -ly)
 - Make note of any interesting patterns and think about how these searches are *different* from each other
 - **Bonus activity: Find a specific language** (Can be done in either CQPweb or UMich, <5 min): e.g. searching Welsh: *ddg*; Dutch: *ij*
3. **On your own: try finding a specific term or concept from your research using both CQPweb and UMich repo** (20 min total)
- Choose some terms relevant to your research interests, find some results related to them – any specific patterns of use? Think about how to harness the ability to “distant” and “close” read with resources discussed so far.
 - Example research questions:
 - How does Early Modern English print talk about Catholics and Protestants?
 - What kinds of contexts do cows show up in?
 - Specific phrases, like “the X of Y”: what are most commonly x and y? Any ideas why?
4. **Bring group back together for discussion and wrap up (15 min)**– how could this be used as part of a larger research project? (discuss how to guide an analysis based on words and frequencies and time)

Assessment

Two options:

1. Feedback survey addressing questions like workshop timing, level of difficulty, participants' perceptions of the workshop's effectiveness in introducing a new tool, and participants' satisfaction with the introduction to the tool, likelihood of use outside the workshop.
2. Ask for a reflection how this could be integrated into part of a research workflow (see discussion section above: emphasize level of understanding of the strengths and weaknesses of the interface(s) as well as their perceptions of the corpus itself).

Flexible based on audience; a more research-oriented session could focus more on the reflection section whereas a more pedagogical session for those teaching EEBO-TCP may focus more on a participant's ability to find the things they need to find.

Additional Instructional Materials

Handout

[Using EEBO-TCP in Research & Teaching](#)