Department of French and Italian Studies
University of Washington

FRENCH 379/499
Eighteenth-Century France Through Digital Archives and Tools
Winter 2018; MW 3:30-5:20
Denny 111

Geoffrey Turnovsky (gt2@uw.edu)
Padelford C-255; 685-1618
Office Hours: M 10:30-12:30pm and by appointment

Description. The last decade or two has witnessed an extraordinary migration of texts and data onto digital platforms, where they can be accessed, in many cases, by anyone anywhere. This is a terrific benefit to students and teachers alike, who otherwise wouldn't be able to consult these materials, and it has transformed the kind of work and research we can do in the French program and in the Humanities. We can now discover obscure, archival documents that we would never have been able to find or access in the past. And we can look at classic works in their original forms, rather than in contemporary re-editions that often substantially change and modernize the works.

Yet this ease of access brings significant challenges: to locate these resources on the web, to assess their quality and reliability, and to understand how to use them, as primary sources, as "data", and as new research technologies. The PDF of a first edition downloaded through Google Books might *look* a lot like the historical printed book it reproduces; but it is *not* that printed book. It is a particular image of one copy of it, created under certain conditions, and it can be a grave mistake to forget or ignore the difference.

In this course, we'll study a variety of digital archives, databases and tools that are useful for studying French cultural history. We'll use these tools and resources to explore the eighteenth century and the intellectual culture of the Enlightenment in the decades leading up to the French Revolution. In particular, we'll explore:

- a new database of Revolutionary pamphlets recently created by the Newberry Library
- two competing databases of book orders received in the 1770s and 1780s by a publishing house in Neuchâtel, Switzerland, which shipped clandestine texts into France
- text analysis tools that leverage the power of computer processing to help us see patterns – relations between words or between elements in a literary text – which we might, as "human readers," miss or be unable to perceive at scale

We'll discuss current debates on digital scholarship and the digital humanities. These often address the underlying question of what it means to read and to interpret a text: computers can be programmed to search through massive corpuses of texts in seconds to find patterns

or locate elements. How might this new technological possibility change or challenge our notions of what reading is? We'll discuss this in parallel with discussions about how reading practices in 18th-century France were evolving as they, too, were driven by analogous technological and sociological changes in the world of the printed text.

We'll consider others aspects of the digitization of texts that we might not normally think about as we scour the internet for materials: how are the texts we find chosen and edited? In what ways do copyright and intellectual property law determine what we can and can't access and how we can use the materials? Finally, we will explore, hands-on, the rudiments of digital publishing. We will prepare a short digital reader of Revolutionary pamphlets using materials in the Newberry collection. We will experiment with scanning, OCR (Optical Character Recognition: a tool that automates the conversion of print into editable digital text), and different approaches to transcription (do we respect historical spelling or modernize the text?). We will then encode our texts using the widely used guidelines of the Text Encoding Initiative, and begin to style and transform our encoded text into HTML for publication on the web.

Learning Goals.

* to discover and explore in detail some of the growing number of archives and tools becoming available for studying many aspects of French cultural history and literature.

* to better understand the information these archives present; how the archives are constructed and what kinds of research they lend themselves to. What do they effectively help us know better, but what might they also obscure? In this case, through these tools, you will develop a sense of the reading culture of the late 18th century: what and how French people were reading, as the Revolution and the end of the Old Regime loomed.

* to better understand digital texts; as PDFs and as searchable text; to learn basic principles of editing, and specifically digital editing using the TEI protocol.

* to develop a sense of the burgeoning field of the digital humanities, including the debates it has generated. This new field has transformed graduate education, but its effects are only slowly seeping into the undergraduate curriculum. This will change, however.

* finally, to reflect on new relationships between technology, computation, and humanities research. How has digitization transformed the way we read? What new insights do new computational tools enabling us to analyze massive corpuses of text in seconds (which would take an individual years to read) provide us?

Texts and materials. Readings will be delivered via the Canvas website as PDFs or will be accessible online. **You will need to bring a laptop to each class**.

Work. This course will have many workshop and hands-on aspects, when we are exploring a tool and, especially, developing our final projects in the 2nd half of the quarter. Thus

preparation for and participation in classroom sessions will be an inegral part of your final grade. **Bring your laptops to each class**.

The written work consists of
* **six** 1/2- to 1-page exercises to be uploaded to Canvas
* **2 individual exercises**, in preparation for the group TEI project
* **group TEI project**: this will entail a collection of documents: above all, the XML mark-up of your Revolutionary pamphlet
* **a short final reflection paper** (2-3 pages)

Breakdown.		
	Participation	20%
	6 class exercises	25%
	2 individual TEI Project exs	20%
	TEI Project (in groups of 2-3)	20%
	Final reflection paper	15%

Syllabus

Wed, 1/3   **The Digital Revolution in Reading and Research**
Katherine Hayles, "Close, Hyper, Machine," in *ADE Bulletin* 150 (2010) (https://ade.mla.org/content/download/7915/225678/ade.150.62.pdf)
Tim Carmody, "10 Reading Revolutions Before E-Books," in *The Atlantic*, Aug 25, 2010 (http://www.theatlantic.com/technology/archive/2010/08/10-reading-revolutions- before-e-books/62004/)

Think of 2-3 digital archives, databases, repositories or collections you've used in your work for your French courses. What did they cover? What lessons were you able to draw from them? What did you find useful about them?

Mon 1/8   **A Quick Introduction to the Eighteenth-Century Francophone World and to Some Useful Digital Resources**
Colin Jones, chaps 6 and 7 (up to p.193) from *Cambridge Illustrated History of France*

Skim through a few of the following digital projects, which focus (mostly) on the 18th century.
http://republicofletters.stanford.edu (Mapping the Republic of Letters)
http://blogs.memphis.edu/salonsproject/ (The Salons Project)
http://www.e-enlightenment.com (The Electronic Enlightenment database of 18th-century correspondence. If you're off campus, try this: http://www.e-enlightenment.com.offcampus.lib.washington.edu)

https://www.youtube.com/watch?v=YP_1eHeyo4 (Visite de Paris au XVIIIe Siècle)

https://archive.org/details/newberryfrenchpamphlets (Newberry Collection of French Revolution Pamphlets)

http://publications.newberry.org/digitalexhibitions/exhibits/show/marie/intro (Many Faces of Marie-Antoinette)

http://publications.newberry.org/smith/exhibits/fe/fe.html (Mapping the French Empire in North America)

http://chnm.gmu.edu/revolution/imaging/home.html (Imaging the French Revolution)

https://colonyincrisis.lib.umd.edu (A Colony in Crisis: The Saint-Domingue Grain Shortage of 1789)

http://www.marronnage.info/en/index.html (Marronage in Saint-Domingue [Haïti]. History, Memory, Technology)

https://www.rousseauonline.ch/tdm.php (Rousseau Online)

https://artfl-project.uchicago.edu (ARTFL Project – see especially the ARTFL *Encyclopédie*)

http://cfregisters.org (Comédie-Française Registers Project)

http://obvil.paris-sorbonne.fr/projets/mercure-galant (*Mercure Galant* at Obvil)

Exercise 1: pick one of the digital resources listed above that seems interesting to you; delve more deeply into the site and write up a short report on it (up to 1 page) following these guidelines:

First, describe the site in general terms: what does it address? what can one learn from it? How does it help us understand French/Francophone history better?

Second, discuss in a bit more detail one idea/fact/concept you gleaned from the site, which you found especially interesting?

Third, address what information the site *does not* provide.

And **finally**, in a concluding paragraph/sentence, say what kind of research you would expect to do with this site: what kinds of projects would it be especially useful for. **Be ready to present this in class**.

Wed 1/10   **Case Study: The Enlightenment and the Origins of the French Revolution. News, Politics, and the Enlightenment Reading Public.**
Voltaire, chapter one, *Treatise on Tolerance*
Alexis de Toqueville, from *The Old Regime and the French Revolution* (see Canvas for versions of Voltaire and Tocqueville in French)
Robert Darnton, "The Forbidden Bestsellers of Prerevolutionary France," *Bulletin of the American Academy of Arts and Sciences* 43, 1 (Oct 1989): 17-45.

Mon 1/15   Martin Luther King, Jr Day

Wed 1/17   **Case-Study, cont'd: An 18th-Century Publisher's Registers and Two Databases**
Chartier, "Do Books Make Revolutions?" in *The Cultural Origins of the French Revolution* (1991), 66-91.
Mark Curran, "Beyond the Forbidden Best-Sellers of Pre-Revolutionary France," *The Historical Journal* 56, 1 (March 2013): 89-112.

Explore the following two websites, based on the sales and correspondence of a publisher in Neuchâtel, Switzerland (*Société typographique de Neuchâtel* – STN)
The French Book Trade in the European Enlightenment (FBTEE): http://fbtee.uws.edu.au/main/. The database is here: http://fbtee.uws.edu.au/stn/interface/.
A Literary Tour of France: http://www.robertdarnton.org

Exercise 2: **First**, skim through list of books in which the STN did business, as they appear in the FTBEE database (browse books by title and simply look at the titles on the drop-down menu): what are you impressions of this list? How much do you recognize?

Second, use the *FBTEE* database to look at the business the STN did with Rousseau's influential novel *La Nouvelle Héloïse.* Look at the editions it bought or sold; how many copies did it send out? where to? Repeat the exercise for *Candide*.

Finally, look at the records for *a same bookseller* in the *FBTEE* and *A Literary Tour of France* databases. For the latter (which is a much smaller list based on the travels of one of the STN's travelling salespeople, named Favarger), click "Tour de France" and scroll down. You'll see a table with towns, booksellers, and numbers of letters. Click on the name of a bookseller and you'll see a short bio with links to PDFs, including one for the "Full Order List." Compare this full order list with the results you find when you "Browse by Client" in the *FTBEE*: find the name of the same bookseller in the drop-down menu, then click "books bought by this client." You might try La Tourmy in Orléans; Buchet in Nîmes; or Pavie in La Rochelle... But any bookseller will work. How would you characterize this booksellers business generally? Do you see differences between results form the two databases?

Mon 1/22   **Visit of Christophe Schuwey, Professor of French at the University of Fribourg to discuss *Digitized Dissent: Three Anti-Louis XIV Pamphlets* and other digital projects**
Deegan and Sutherland, from *Transferred Illusions. Digital Tehnology and the Forms of Print* (2009), 110-118

"How Long Do Users Stay on Web Pages?"
https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/

Take a look at the following sites:
www.nouvellesnouvelles.fr
www.dissidence17.fr/alcoran.html
www.idt.paris-sorbonne.fr/corpus/finalengine.php?search=true (old
    version)
www.idt.paris-sorbonne.fr/ (new version)
webpages.ursinus.edu/jlionarons/wulfstan/wulfstan.html

Wed 1/24 **Benefits, Pitfalls, and Trade-Offs of Digitization; and introduction to our corpus of Revolutionary Pamphlets. In Suzzallo Instructional Lab.**
Visit of Verletta Kern (Digital Scholarship Librarian) and Beth Lytle (Learning Technologies)
Stephen Nichols, "Materiality and Mimesis: Anatomy of an Illusion," from *From Parchment to Cyberspace. Medieval Literature in the Digital Age* (2016), 43-54.
Alexis Madrigal, "What is a Book," in *The Atlantic*, May 7, 2014: https://www.theatlantic.com/technology/archive/2014/05/what-is-a-book/361876/

Look at the Newberry Collection of French Revolution Pamphlets: https://archive.org/details/newberryfrenchpamphlets. We'll be developing a short "reader" of a small selection of excerpts from these documents.

Mon 1/29 **Editing, Digital Editing, and Document Analysis: some general principles and preliminary considerations**
Leah Marcus, "Textual Scholarship" from *Introduction to Scholarship in Modern Languages*, 143-159.
"What is XML and Why Should Humanists Care: An Even Gentler Introduction to XML": http://dh.obdurodon.org/what-is-xml.xhtml. Read through the section "Is every document really a hierarchy?"

TEI PROJECT ex 1: Pick one of the pamphlets from the Newberry collection (you can pick from my list of suggestions or find one yourself; keep in mind that you'll be developing an "edition" of this document in groups of 2 or 3). For class, read the pamphlet. Try to understand as much about it as you can, as a historical text and a physical document. Write down some notes in answer to some of the questions which follow, loosely categorized (it is *very likely* that you won't be able to answer all, or even most of these questions):

-- who wrote this document? when? who published it?
-- what is the text trying to say? how does it say it? what kind of form does it take as a text?

-- what kind of form does it take as a physical document (inasmuch as you can determined this from the PDF): why does the typography look like?
-- who do you think its readership was? Is there any explicit address to the reader? Any implicit indications about who this text was intended for?

Mark-up parts of the text that seem obscure to you – especially references to things that the text doesn't explain. What annotations do you think might be needed for a modern reader?

Submit prelimary answers to these questions, **but keep working on this for Feb. 7** when you'll start working on with 1-2 partners on an edition of the first part of the pamphlet.

Wed 1/31   **The Legal Framework – Copyright, Fair Use: Transforming a Text**
               Visit of Maryam Fakouri, Scholarly Publishing Outreach Librarian
US Copyright Office Circular: https://www.copyright.gov/circs/circ01.pdf
Maryam's Copyright Guide:
     http://guides.lib.uw.edu/research/copyright/introduction
Author's Guild vs. Google (edited version of decision):
EFF discussion of Google Books Opinion:
     https://www.eff.org/deeplinks/2015/10/big-win-fair-use-google-books-lawsuit
Carla Hesse, "The Rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance," *Dardalus* 131, 2 (spring 2002): 26-45 (read up to p.33 for some historical backdrop, which is relevant to the period we're studying).

Mon 2/5   **Copying Text: Adventures in Transcription and Optical Character Recognition (OCR). Plus preliminary Intro to Mark-Up, XML, and TEI**
Daniel Pollack-Pelzner, "Why We (Mostly) Stopped Messing With Shakespeare's Language," *New Yorker*, October 6, 2015
Irvine, Marcellesi and Zomorodian, "Digitizing 18th-Century French Literature: Comparing Transcription Methods for a Critical Edition Text," *Workshop on Computational Linguistics for Literature* (2012): 64-68.
"What is XML and Why Should Humanists Care: An Even Gentler Introduction to XML": http://dh.obdurodon.org/what-is-xml.xhtml. Read from "Could I please see some XML already" through "Other Web Standards"

Early Modern OCR Project (http://emop.tamu.edu)

Exercise 3: Experiment with OCR. There are many OCR tools out there. A very accessible one is through Google Drive. To use this, upload a PDF to the Drive; then open that PDF in Google Docs. The PDF will be converted to editable text. You can also use the OCR function in Acrobat Pro, which is on the computers in Odegaard. First, OCR a text from a modern edition. Start from scratch: scan a page from a book; then run it through an OCR process

(Google; Acrobat Pro or another). Second, do your pamphlet: save it as a text or Word file (we'll come back to it). What kind of results do you see?

Upload two files (text or Word) to the Ex 3 assignment:
    1. the output of the page you scanned. Add the reference to the source at the top of the page (what book is this from?). And after the processed text, write a paragraph highlighting issues you noticed.
    2. the output of the OCR'd pamphlet, with a short reflection (brief paragraph) on the issues you see.

Wed 2/7   **Visit to Special Collections** to see our collection of Revolutionary Pamphlets.

    TEI exercise 2. Fun with Transcription. Transcribe the title page and first 2-3 paragraphs of your pamphlet; type the text into Word or Google Docs. You'll be making numerous formatting and linguistic decisions (relating to page layout, spelling, punctuation, etc). Make them and think about why you did. Describe a few examples: explain what choices you felt you had to make and why you made them.

Mon 2/12   **Digital Editing, cont'd: Text Encoding. A Very Basic Introduction to the Text Encoding Initiative (TEI) and XML**
    "What is the TEI?"
    "Getting Started with TEI" and "Technical Background"

    Some examples of TEI-encoded corpuses:
    "The TEI by Example Project": http://teibyexample.org
    The *Mercure Galant* Project by Obvil: http://obvil.paris-sorbonne.fr/corpus/mercure-galant/.
    Molière at Obvil: http://obvil.paris-sorbonne.fr/corpus/moliere/critique/

    For class, download the oXygen XML editor (https://www.oxygenxml.com (Links to an external site.)Links to an external site.). You have a 30-day free trial. If you download the morning of class, this will take us through the date for submission of materials (Feb 12 to March 12 is 28 days).

    If you want to start experimenting, you can look at the directions here: http://tei.oucs.ox.ac.uk/Talks/2010-07-oxford/exercise-01-basic.xml (Links to an external site.)Links to an external site. (but we'll cover this in class).

    We'll talk about other text editors as well. There are many options. oXygen's key advatange is that it is made for XML, and includes pre-set templates (validation schema) for TEI. But it is a bit cluttered and complex. Some possibilities for Mac and Windows:

TextWrangler: https://itunes.apple.com/cv/app/textwrangler/id404010395?mt=12 (Links to an external site.)Links to an external site. For Mac. (Note that this isn't supported by Bare Bones Software anymore, and hasn't been updated to work with High Sierra, if you have that installed. They've replaced TextWrangler with BBEdit, which works well, too: https://www.barebones.com/products/textwrangler/ (Links to an external site.)Links to an external site.. If you try BBEdit, you'll get a full set of features for 30 days, then will still be able to use its basic functions indefinitely). Notepad ++: https://notepad-plus-plus.org (Links to an external site.)Links to an external site.. For Windows. It's always at the top of "best free editors" lists.
Editey: through Google Drive

Bring your updated editorial notes and transcriptions (the OCR version and your hand transcription) to class: we'll start working in groups.

Exercise 4. At the Molière at Obvil site, you'll notice when you go to a document that there's an option to download the TEI document at the top left. Go to Grimarest's Vie de Molière (http://obvil.paris-sorbonne.fr/corpus/moliere/critique/grimarest_vie-moliere (Links to an external site.)Links to an external site.); right-click the TEI option and save to your hard drive. Open up the downloaded file (which will have an .xml extension) in oXygen (if you've downloaded this) or in another text editor (TextWrangler, Notepad ++, Editey... you can also use the default text editors that come with Macs [TextEdit] and Windows [Notepad].

Another option: open up the XML file in a browser: Firefox or Chrome (Safari is trickier). If you don't get the TEI (the page is blank or just a stream of text), then right click and choose "View Page Source." This should give you the XML document that you've asked the browser to open.

Try to identify the structure of the document and the various tags and elements that you see. A TEI document is divided into two elements: the <teiHeader>, which contains all the metadata about the text; and the <text>, which includes as content (as you might imagine) the actual text. Find these two elements (together they comprise the entire TEI document, under the root element <TEI>).

Compare the <text> content in TEI with the actual document that's being encoded: https://books.google.com/books?id=lDo0Ix2OLOYC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false (Links to an external site.)Links to an external site..

Focusing especially on the first pages/paragraphs of the text, what do you see? What decisions have been made about what the "text" is and how it should be marked up in the XML? What's left out? (Ignore the website for

now. Just look at the XML document). Upload a short write-up of your initial, quick impressions about what you were able to decipher and how this corresponded to the original text.

Wed 2/14   **Workshop: Creating a TEI XML File**

In class, working in your groups, we'll begin to create a TEI-XML document of your pamphlet. Bring your updated editorial notes and transcription.

We'll also talk about automated processes for converted text to XML/TEI. See TEI's tool, OxGarage: http://www.tei-c.org/oxgarage/
And Odette: http://obvil.lip6.fr/Odette/

Mon 2/19   PRESIDENTS DAY

Wed 2/21   **What to Do with Digital Text? Quantitative Approaches to the Literary Archive. Some General Principles and Applications to 18th-Century France.**
Franco Moretti, "Conjectures on World Literature," *New Left Review* (Jan/Feb 200): 54-68 (https://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature)
Jean-Baptiste Michel, et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science*. 16 Dec. 2010
Keith Michael Baker, "Revolution 1.0," in *Journal of Modern European History* 11, 2 (2013): 187-219

Exercise 4: Experiment with Google Books Ngram viewer (https://books.google.com/ngrams) and with Philologic 4 on ARTFL.
First, using the Google Ngram viewer and selecting the French corpus between 1500 and today, check for Revolutionary keywords "liberté," "égalité," "droits," and "citoyen," as well as one or two other terms you can think of. Write up your results and impressions.

Second, see if you can run similar tests using the Philologic 4 tool on ARTFL: http://artfl-project.uchicago.edu.offcampus.lib.washington.edu/ARTFL.html. This is a little complex; but see if you can find some interesting results. Note them down and we'll discuss and do more in class (the ARTFL-FRANTEXT is the big database of French-language texts).

Mon 2/26   **More Text Analysis: Tools and Goals.**
Ted Underwood, "Seven Ways Humanists are Using Computers to Understand Text" (https://tedunderwood.com/2015/06/04/seven-ways-humanists-are-using-computers-to-understand-text/)
"Getting Started," with Voyant: http://docs.voyant-tools.org/start/.
Voyant: http://voyant-tools.org

Exercise 6. Here's a plain text version of *Candide* 🔍. Try to upload in Voyant and you'll see five panels with five tools: a word-cloud called Cirrus, a Reader (which is just the text itself), a tool called Trends, with frequencies of terms graphed across the work, a Summary, and a panel called Contexts, which is a concordance tool (it's a KWIC tool showing the KeyWord In Context [i.e. KWIC]). Contexts also has an option called Correlations, which charts pairs of words from highest to lowest correlation (often appearing together vs. never appearing together). You can pick terms in the boxes underneath each panel. You can also use other tools: hover at the top right of each tool and you'll see what look like Microsoft Windows logos (four panels at an angle): click and you'll see a bunch of other options. Collocates is similar to what we saw in ARTFL, looking at how often terms are associated with each other.
No need to come up with anything in particular here. Read the Underwood and we'll talk more about quantitative approaches and do some experiments with Voyant and some other tools.
We'll do some exercises in class.

Wed 2/28 **Workshop and presentations: continue collaborative work on your TEI projects.**

Mon 3/5 **Web Publishing: HTML, CSS, TEI Boilerplate, and XSLT**. **Styling and transforming our TEI documents.**
Do Miriam Posner's short tutorial in HTML and CSS: http://miriamposner.com/classes/dh101f16/tutorials-guides/web-publishing/html-css/. You can do this in oXygen. Open up a new document and in the folder marked "New Document," you'll see options for both a new HTML and a new CSS document. Or in any other text editor.
"What is XML and Why Should Humanists Care: An Even Gentler Introduction to XML": http://dh.obdurodon.org/what-is-xml.xhtml. Read the section "Practicing what we preach."

Wed 3/7 **Putting It All Together.**
Conclusions.

Mon 3/12 **FINAL REFLECTION PAPER DUE as well as all materials relating to your TEI project, by 5pm.**