

ENGL84031/HISTGU4031/COMS4495 | Spring 2019

**Transforming Texts: Textual Analysis, Literary Modeling, and Visualization**

Wed 2:10–4pm

Fri Lab 1:10–3pm

**Instructors**

Professor Dennis Tenen

Professor Steven Feiner

Dr. Tianna Uchacz

Professor Pamela Smith

Terry Catapano

Dr. Carmine Elvezio

**M&K Project Team**

Naomi Rosenkranz

Caroline Surman

**Course Assistants**

Samantha Siu

Raj Biswas

**COURSE RESOURCES:**

- [Piazza](#)
- Courseworks
- M&K Github Repo
- Digital Edition of BnF Ms. Fr. 640

**COURSE DESCRIPTION:**

This hybrid course, designed for graduate and advanced undergraduate students in the social sciences and humanities, and computer science students interested in VR and AR visualization, is situated at the crossroads of historical exploration and computer sciences. Students will be exposed to digital literacy tools and computational skills through the lens of the Making and Knowing Project.

The edition will draw on collaboration with and research done by the Making and Knowing Project <http://www.makingandknowing.org/> on an anonymous sixteenth-century French compilation of artistic and technical recipes (BnF Ms. Fr. 640). You will be working from the encoded English translation of the manuscript, prepared by the Spring 2017 course “*HIST GR8975 What is a Book in the 21st Century? Working with Historical Texts in a Digital Environment.*” This course will also utilize the concepts and prototypes developed by the Computer Graphics and User Interfaces Lab (CGUI) that were partly inspired by the work of computer science students in the Spring 2018 “*COMS W4172: 3D User Interfaces and Augmented Reality (AR).*”

This rare French manuscript resulted from the compilation of craft knowledge over time, followed by its subsequent “disassembly” in a late sixteenth-century workshop by an

author-compiler-practitioner who experiments on techniques contained in the manuscript's "recipes." While this course will focus on this intriguing manuscript and the research that has been carried out on it, the skills you will learn over the course of the semester are widely applicable to other types of Digital Humanities projects, and indeed, in many fields outside of traditional academic study.

For the final project, students will collaborate to investigate linguistic features of Ms. Fr. 640 using natural language processing and text mining techniques. These projects will shed light on topics of interest within the manuscript and uncover connections within the textual data. By using the tools prototyped by CGUI and working alongside COMS students, groups will learn to adapt and recode data sets, as well as view them in a variety of visualization formats.

### **ASSESSMENT:**

- 25 % Class Participation
- 25 % Online Participation
- 25 % Assignments
- 25 % Final project

### **SCHEDULE:**

#### **Week 1, January 23**

Introduction to BnF Ms Fr 640, the M&K Project, the CGUI lab, xpMethods group.  
M&K Folder structure.

Lab: Setting up the environment. Find the command line. Intro to files and paths.

#### **Week 2 , January 30**

Character, code, letter.

- Sixth Book, Chapter I, pages 247-259 in *The Advancement of Learning* by Francis Bacon.
- "Codes in Combinations" and "Braille and Binary Codes" in *Code* by Charles Petzold.
- "Think Tape" and "Reveal Codes" in *Track Changes* by Matthew Kirschenbaum.
  - *FUN OPTIONAL VIEWING:* "[The Paperwork Explosion](#)" (1967) 5-min promo video for IBM directed by Jim Henson (!) described in "Think Tape" (p. 173-74)
- [The Making and Knowing Project Lab Videos](#) (Watch all 11 videos in the playlist (00-10) - c. 40 minutes)
- Donna Bilak, Jenny Boulboulé, Joel Klein, and Pamela H. Smith, "The Making and Knowing Project - Reflections, Methods, and New Directions," in *New Directions in*

*Making and Knowing*, a special issue guest edited by Smith of *West 86th: A Journal of Decorative Arts, Design History, and Material Culture*, 23.1 (2016): 35-55.

Lab: Encoding workshop.

### Week 3, February 6

Format, metadata.

- Tanselle, G. Thomas. "The Concept of Format." *Studies in Bibliography* 53 (2000): 67–115.
- Sterne, Jonathan. "Format Theory." In *MP3: The Meaning of a Format*, 1–31. Durham: Duke University Press, 2012.
- Ch. 7, "Historical Interlude" in *The Elements of Typographic Style* by Robert Point Roberts Bringham.
- "From A to Screen," by Johanna Drucker in *Comparative Textual Media*.
- *OPTIONAL READING*
  - Recent Twitter thread: Is paper better, or merely different?

**Lab:** Intro to Markdown, Latex, Pandoc.

Before lab, watch: [Chancery Papermaking](#) - University of Iowa Library;

Optional: quill-cutting video <https://www.youtube.com/watch?v=36h1vt-9sss>

### Week 4, February 13

Words, types, tokens I

- Christopher D. Manning, selections from the "Introduction," in *Foundations of Statistical Natural Language Processing*, 19–34
- Stanford Literary Lab, "[Quantitative Formalism](#)"
- Anne Carson, ed., *If not, winter. Fragments of Sappho*, 159–225.
- Pamela H. Smith and The Making and Knowing Project, "Historians in the Laboratory: Reconstruction of Renaissance Art and Technology in the Making and Knowing Project," *Art History*, special issue on Art and Technology (2016)

**Lab:** Word clustering

### Week 5, February 20

Words, types, tokens II

- Christopher D. Manning, selections from "[Corpus-Based Work](#)," in *Foundations of Statistical Natural Language Processing*, 117–47

- John F. Burrows, "[Introduction](#)," in *Computation into Criticism*, 1–12
- Bettina Fischer-Starcke, "[Keywords and concordance lines](#)," in *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*, 65–107
- Dennis Tenen, "Toward a Computational Archaeology of Fictional Space," <https://doi.org/10.7916/D8QC1M5D>

**Lab:** Type-to-token ratios and vocabulary richness

## **Week 6, February 27**

Sentence, grammar, punctuation.

- Andrew Piper, "Punctuation (Opposition)," In *Enumerations: Data and Literary Study* (Chicago: University of Chicago Press, 2018),
- Noam Chomsky, *Syntactic Structures* (The Hague: Mouton Publishers, 1957), 5–33.
- Stanley Fish, *How to Write a Sentence: And How to Read One* (New York: Harper Collins, 2011), 1–44.
- Familiarize yourself with the chief lexical categories of the English language and the process of parsing simple sentences into syntax trees, as laid out in this Introduction to syntax

**Lab:** Dynamics of Agency in Texts

Please skim Dennis's article, "Distributed Agency in the Novel," before Friday's lab

## **Week 7, March 6**

Paragraph.

- The History of the English Paragraph by Edwin Herbert Lewis, pp 9-65.
- Lovegrove, W., and D. Elliman. "Text Block Recognition from Tiff Images." In *IEEE Colloquium on Document Image Processing and Multimedia Environments*, 4/1-4/6, 1995.
- Tebeaux, Elizabeth. "Technical Writing and the Development of the English Paragraph 1473–1700." *Journal of Technical Writing and Communication* 41, no. 3 (July 1, 2011): 219–53.
- Richard Southall, "Visual Structure and the Transmission of Meaning," in *Document Manipulation and Typography*, edited by J.C. van Vliet (Cambridge University Press, 1988), 35–45.

**Lab:** Making & Knowing Lab visit

## Week 8, March 13, 2019

Page.

- The Chapter in Western Literature by Nicholas Dames.
- Meehan, James Richard. "The Metanovel: Writing Stories by Computer."
- Rolf Ingold, Rene-Pierre Bonvin, and Giovanni Coray, "Structure Recognition of Printed Documents," in *Document Manipulation and Typography*, edited by J.C. van Vliet (Cambridge University Press, 1988), 59–70.

NO LAB THIS WEEK, but Spring Break is a chance for you to make up for any worksheets you did not complete. **Any outstanding work is due by Sunday night, March 24.**

**Spring Break: March 18-22**

## Week 9, March 27, 2019

Document, book, edition.  
& Project design sprint and sign-up

- "Near Print and Beyond Paper: Knowing by \*.pdf" in *Paper Knowledge* by Lisa Gitelman.
- King, James C. "A Format Design Case Study: PDF." In *Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia*, 95–97. HYPERTEXT '04. New York, NY, USA: ACM, 2004
- Kimura, Gary D., and Alan C. Shaw. "The Structure of Abstract Document Objects." In *Proceedings of the Second ACM-SIGOA Conference on Office Information Systems*, 161–69. COCS '84. New York, NY: ACM, 1984.

Lab: Intro to version control with GitHub.

## Week 10, April 3

Library, archive.

- Mbembe, Achille. 2012. "The Power of the Archive and Its Limits." In *Refiguring the Archive*, 19–27. Springer Science & Business Media.
- Selections from *Dust: The Archive and Cultural History* by Carolyn Kay Steedman.
- Schechter, S. 1998. "Genizah Specimens." *The Jewish Quarterly Review* 10 (2): 197–206.

Lab: RBML visit & collection policies.

In lieu of Worksheet for Week 10, 3 project proposals due Sunday 11:59 pm

### **Week 11, April 10**

Interface.

- "Recondite Surfaces" in *Plain Text*, by Tenen.
- Selections from Laurel, Brenda. "Interface as Mimesis." In *User Centered System Design: New Perspectives on Human-Computer Interaction*, edited by Donald A Norman and Stephen W Draper, 68–85. Hillsdale, N.J.: L. Erlbaum Associates, 1986.
- Selections from *The Interface Effect* by Alexander Galloway.

Lab: Topic to be TBD

After lab, 15-min meetings with teams and instructors to refine a final project plan

### **Week 12, April 17**

Platform.

Lab: Topic TBD

### **Week 13, April 24**

Project workathons. Labs on Wednesday and Friday.

### **Week 14, May 1**

Final Project launch.

---

### **Statement on Academic Integrity**

The intellectual venture in which we are all engaged requires of faculty and students alike the highest level of personal and academic integrity. As members of an academic community, each one of us bears the responsibility to participate in scholarly discourse and research in a manner characterized by intellectual honesty and scholarly integrity.

Scholarship, by its very nature, is an iterative process, with ideas and insights building one upon the other. Collaborative scholarship requires the study of other scholars' work, the free discussion of such work, and the explicit acknowledgement of those ideas in any work that inform our own. This exchange of ideas relies upon a mutual trust that sources, opinions, facts, and insights will be properly noted and carefully credited.

In practical terms, this means that, as students, you must be responsible for the full

citations of others' ideas in all of your research papers and projects; you must be scrupulously honest when taking your examinations; you must always submit your own work and not that of another student, scholar, or internet agent.

Any breach of this intellectual responsibility is a breach of faith with the rest of our academic community. It undermines our shared intellectual culture, and it cannot be tolerated. Students failing to meet these responsibilities should anticipate being asked to leave Columbia.

### **Disability Related Accommodations**

In order to receive disability-related academic accommodations, students must first be registered with Disability Services (DS). More information on the DS registration process is available online at [www.health.columbia.edu/ods](http://www.health.columbia.edu/ods). Faculty must be notified of registered students' accommodations before exam or other accommodations will be provided. Students who have (or think they may have) a disability are invited to contact Disability Services for a confidential discussion at (212) 854-2388 (Voice/TTY) or by email at [disability@columbia.edu](mailto:disability@columbia.edu).