PUBLISHING THE DIGITAL HUMANITIES

Jim English, University of Pennsylvania
PRICE LAB for DIGITAL HUMANITIES

Python for Humanists

MAY 10-11

Graduate Students: Register for our Spring 2018 Python for Humanists Workshop, May 10-11

ANNOUNCEMENTS

Michael Gomer and Scott Enderle featured in Omnia
The pair talks about their English Playbills Project.

Graduate Student DH Working Group
The Price Lab’s Working Group meets biweekly to imagine generative uses of technology in student work and to sustain a community of curious young scholars. Join us!

EVENTS & SEMINARS

April 23
Whitney Trettien (University of Pennsylvania)
Mellon DH Seminar

May 10—May 11
Python for Humanists
Established Peer-Reviewed Journals in Digital Humanities

**DSH**
*Digital Scholarship in the Humanities* (print, OUP / ProQuest)

**dhq**
*Digital Humanities Quarterly* (open-access, no print)

**Frontiers in Digital Humanities** (open-access, no print)

**Journal of Digital Humanities**
*Journal of Digital Humanities* 2014-2017 (open-access, no print)
Projects at Price Lab

Currently Funded  |  Completed Funding  |  Other DH Projects at Penn

View  |  Edit

Mining Goodreads
Jim English, Lyle Ungar, Rahul H. Dhulecha, Scott Enderle
An exploration of online user reviews in the Goodreads social reading site via computational analysis.

Praxes for Race: A Catalogue
Oliver Rollins, Benjamin Wiggins
A qualitative content analysis of the seemingly mundane, normalized, and/or impartial descriptors, discourses, and practices that operate as referents for race.

The Morenziana Online Digital Edition (MODE)
Mauro Calcagno
Introduces a new model for producing and disseminating scores of Western music by integrating musical philology and digital technology.

Contemporary Fiction Database
Jim English, Scott Enderle, Daniel Sample
Uses quantitative relationships among mid-sized, hand-made datasets to map the field of Anglophone fiction from 1960 to the present.

Etymological Database for Indo-European languages
Rolf Noyer
Will produce an up-to-date cross-linguistic etymological dictionary of Indo-European languages built on a relational database format.

Machine-Aided Close Listening
Chris Mussel
A new application to empirically study the phonotextual relationship between the printed and the performed poem.
Mining Goodreads
Literary Reception Studies at Scale

Jim English
John Welsh Centennial Professor of English; Director, Wolf Humanities Center; Faculty Director, Price Lab for Digital Humanities

Lyle Ungar
Professor of Computer and Information Science

Rahul H. Dhakecha
MSc Student, Computer Science

Scott Enderle
Digital Humanities Specialist, Penn Libraries

Funding Period: June, 2016 — September, 2018

Project Manager: James Pawelski
Director of Education and Senior Scholar, Master of Applied Positive Psychology Program

Previous Research Assistants: Tianli Han (MSc SEAS), Sharvin Shah (MSc SEAS), Daniel Sample, Alex Anderson, Savannah Lambert, Amy Stidham (CAS)

Supported by the Price Lab for Digital Humanities in partnership with the Positive Psychology Center and the World Well-Being Project.
Mining Goodreads Outputs

Articles
Articles plus:
Articles *plus*:

- textual corpus (2000 contemporary novels, digitized, cleaned)
- metadata on corpus (csv file)
- database (MySQL) containing:
  - 4 million reviews scraped from Goodreads
  - metadata on reviews
  - metadata on 2000 high-volume reviewers
  - results of statistical analysis using NLP tools
- static illustrations: charts, graphs, word clouds
- interactive visualizations, linked to metadata
- code (MySQL queries; python in Jupyter notebooks)
The Transformation of Gender in English-Language Fiction

Ted Underwood, David Bamman, and Sabrina Lee / 02.13.18

Peer-Reviewed By: Heather Love
Clusters: Gender
Article DOI: 10.2294/18.009
Dataverse DOI: 10.7910/DVN/2M2DMAN
PDF: 10.7910/DVN/TEGMCI
Journal ISSN: 2371-4549


This essay explores the changing significance of gender in fiction, asking especially whether its prominence in characterization has varied from the end of the eighteenth century to the beginning of the twenty-first. We have reached two conclusions, which may seem in tension with each other. The first is that gender divisions between characters have become less sharply marked over the last 170 years. In the middle of the nineteenth century, very different language is used to describe fictional
The Transformation of Gender in English-Language Fiction (PDF)

Underwood, Ted; Bamman, David; Lee, Sabrina, 2018, "The Transformation of Gender in English-Language Fiction (PDF)". https://doi.org/10.7910/DVN/TEGMDJ, Harvard Dataverse, V1

PDF of The Transformation of Gender in English-Language Fiction (2018-02-10)

Subject
Arts and Humanities; Computer and Information Science

Keyword
characterization, fiction, literary history, gender
academic accounts of literary history, we might predict that the prominence of female characters in this tradition would have increased in the nineteenth and twentieth centuries—with, to be sure, some well-known interruptions. There is a notorious backlash against first-wave feminism in the middle of the twentieth century, for instance—dated by Gayle Greer to the 1940s and 50s. But we might hope at least to see an overall story of progress across two centuries.

If we start with that broadly positive expectation, we will be disappointed. In fact, from the nineteenth century through the early 1960s we see a story of steady decline. In Figure 1, we have plotted the proportion of words used in characterization that describe women. “Words used in characterization” here includes verbs a character governed, and nouns they possessed (like “spirits”), as well as adjectives that modify a character. Dialogue spoken by characters is not counted, but including it would not materially change the pattern. Characters of unknown gender have been excluded from the total, so the proportion being plotted (here, and elsewhere in the article) is simply the ratio of words that describe women to words that describe either men or women. This proportion actually declines from the middle of the nineteenth century to the middle of the twentieth—in the very period when we might expect to see the effects of first-wave feminism.

The trend reverses around 1970, for reasons we will investigate later.

![Figure 1. The percentage of words used in characterization that describe women.](image)

This picture is counter-intuitive enough that our first instinct was to ask whether...

Replication Data for: The Transformation of Gender in English-Language Fiction

Underwood, Ted; Bamman, David; Lee, Sabrina, 2018, "Replication Data for: The Transformation of Gender in English-Language Fiction", https://doi.org/10.7910/DVN/ZM2MAN, Harvard Dataverse, V1

Description
Code and data for research on characterization published as "The Transformation of Gender in English-Language Fiction," Ted Underwood, David Bamman, and Sabrina Lee, Cultural Analytics 2018. The zipped character_repositories is a frozen version of the code also available at https://github.com/tedunderwood/character. The character_data files contain words associated with characters in novels. For metadata that will link characters to novels, see the readme file in the repository. (2018-01-03)

Subject
Arts and Humanities; Social Sciences

Keyword
characterization, fiction, literary history, gender
Cultural Analytics

Dataverse Repository: Replication Data

Files

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>character-repository.tar.gz</td>
<td>A frozen version of <a href="https://github.com/redunrewood/character">https://github.com/redunrewood/character</a>, as it existed on Jan 6, 2018. Contains Python code used for analysis, R code used to produce figures, and metadata, along with a sample of data. If the github repo is still up, you might prefer to consult that; this is archival. Will unzrp into a folder.</td>
</tr>
<tr>
<td>character_table_18c19c.tsv</td>
<td>Raw tabular data, one character per line. Keyed to filtered_fiction_metadata in the metadata folder of code repository.</td>
</tr>
<tr>
<td>character_table_post1900.tsv.zip</td>
<td>Raw tabular data, zipped to save space; unzips to 3GB. Keyed to filtered_fiction_metadata in the metadata folder of code repository.</td>
</tr>
<tr>
<td>data_dictionaries.md</td>
<td>Data dictionaries that explain columns in the character_tables and filtered_fiction_metadata.</td>
</tr>
</tbody>
</table>

Terms

Search this dataset... Find

Versions
**Topic Modeling the Hàn diǎn Ancient Classics**

Colin Allen, Hongliang Luo, Jaimie Murdock, Jianghuai Pu, Xiaohong Wang, Yanjie Zhai, and Kun Zhao / 10.12.17

Peer-Reviewed By: Paul Viertbuler

**Clusters:** Data

Article DOI: 10.2218/18.018

Dataset DOI: 10.7938/DV7/2CGXX9

There is a small but growing literature on large-scale statistical modeling of Chinese language texts. Ouyang analyzed a corpus of over 40,000 ancient documents downloaded from multiple sources. This was used to plot the temporal distributions of word frequencies and geographic distributions of authors. ¹ Huang and Yu modeled the SongCi poetry corpus, first converting it to tonally marked pinyin to conserve poetically important pronunciation information. ² Nichols and colleagues reported initial modeling of the Chinese Text Project corpus ³ in a conference paper. (Further below, we describe differences between this corpus and the Hanidian.) With additional collaborators, this group has now conducted two studies that are currently unpublished but under review. In the first, they...
Supplemental Materials for "Topic Modeling the Hàn dân Ancient Classics (漢典古籍)"

Description
The "Hanlian" corpus (汉典 or Han dân, i.e., the "Han canon" or "Han classics") contains over 18,000 classics of ancient Chinese philosophy, as well as documents of historical and biographical significance, and literary works. The versions of the documents presented here are derived from www.zdic.net under their permissive Creative Commons 1.0 Public Domain Dedication. These significant cultural texts are modeled here and published in the Journal of Cultural Analytics. The Dataverse repository contains the models for the InPhO Topic Explorer (handlian-ca.tar), installation instructions (README.md), and supplemental materials (handlian-ca-supplemental.pdf). These works are released under the Creative Commons Attribution-Share Alike 4.0 International License (CC BY-SA 4.0).

Subject
Arts and Humanities

Keyword
Ancient Chinese, topic models, open source software, visualization, dataset, digital humanities

Related Publication
Beyond Poet Voice: Sampling the (Non-)Performance Styles of 100 American Poets
Marit J. MacArthur, Georgia Zellou, And Lee M. Miller

Crossing Over: Gendered Reading Formations at the Muncie Public Library, 1891-1902
Lynne Tatlock, Matt Erlin, Douglas Knox, And Stephen Pentecost

The Identity Issue
Introduction By Susan Brown
And Laura Mandell, Special Issue By The NovelTM Research Group

Send us your null results
Andrew Piper

The Transformation of Gender in English-Language Fiction
Ted Underwood, Devid Bamman, And Sabrina Lee
Null results section
Send us your null results
Andrew Piper / 03.03.18

Journal ISSN: 2371-4549
Cite: Andrew Piper, "Send us your null results," Cultural Analytics March 5, 2018.

A considerable amount of work has been produced in quantitative fields addressing what has colloquially been called the "replication crisis."[1] By this is meant three related phenomena: 1) the low statistical power of many studies resulting in an inability to reproduce a similar effect size; 2) a bias towards selecting statistically "significant" results for publication; and 3) a tendency to not make data and code available for others to use.

What this means in more straightforward language is that researchers (and the public) overwhelmingly focus on "positive" results; they tend to over-estimate how strong their results are (how large a difference some variable or combination of variables makes); and they bury a considerable amount of decisions/judgments in their research process that have an impact on the outcomes. The graph in Figure 1 down below represents the first two dimensions of this problem in very succinct form (see Simmons et al for a discussion of the third).[2]
Collections as Data: Laurie Allen, Thomas Padilla, Stewart Varner

Always Already Computational

Collections as Data is an Institute of Museum and Library Services supported effort that aims to foster a strategic approach to developing, describing, providing access to, and encouraging reuse of collections that support computationally-driven research and teaching in areas including but not limited to Digital Humanities, Public History, Digital History, data driven Journalism, Digital Social Science, and Digital Art History.
Always Already Computational

Collections as Data is an Institute of Museum and Library Services supported effort that aims to foster a strategic approach to developing, describing, providing access to, and encouraging reuse of collections that support computationally-driven research and teaching in areas including but not limited to Digital Humanities, Public History, Digital History, data driven Journalism, Digital Social Science, and Digital Art History.
THRESHOLDS is a digital zine. We custom-design each issue, generating provocative collisions of art and ideas. The result is creative/critical digital handiwork, a hand-stitched sampler of texts and ideas.

Issue #2, Traces, is a collective writing experiment loosely based on the work of media theorist Siegfried Zielinski.

CONTRIBUTORS: Zoe Beloff, Emily Bloom, Devon Dill, Lara Farina, Daniel Irgang, Aaron Jaffe, Heather Love, Dathalim O’Dea, Jussi Parikka, Melissa Rogers, Gay Stevenson, TC Tolbert, Darren Wershler, Siegfried Zielinski

COVER ART: Detail from Sedimental No. 9, Jody Alexander, 2011.
The eye is inhabited.

We tend to think the eye resides in a universe, but there is also a universe living in the eye, a universe that has its own laws, its own gravitational pull. Our lines of sight are directed.

The kingdom in the cornea is monotheistic. A single reality lends itself to simplicity, to safety, to salvation: preconfigured, predictable tomorrows. Loyalty, we are told, guarantees life. "Do not acknowledge other gods" = do not look past or beyond. Do not, in fact, think.

Audience to a scene, we interpret its significations in 1/20th of a second.

The rest is time-travel, chiseling a past to concoct a future, to concoct resemblance, a semblance, again.

Somebody once said we see what we want to see. But also, we see what the political present wants us to see. We are a magic. It distracts us to the spectacular, so we do not watch the flick of the
Provoking more noisy t(h)inkering in sound studies, digital humanities, and the audio arts.
Cut/Copy/Paste
Whitney Trettien (Minnesota/Manifold 2018)

- Cut/Copy/Paste
  Whitney Trettien
  Updated April, 2018

- Dark Deleuze
  Andrew Culp
  Updated April, 2018

- Deep Mapping the Media City
  Shannon Mattern
  Updated April, 2018

- The End of Man
  Joanna Zylinska
  Published March, 2018

- The Geek's Chihuahua
  Ian Bogost
  Updated April, 2018

- Going the Rounds
  Ryan Cordell, David Smith, Abby Mullen, Jonathan D. Fitzgerald
  Updated April, 2018

- How Noise Matters to Finance
  Nicholas A. Knouf
  Updated April, 2018

- In/visible Archives of the 1980s
  Margaret Galvan
  Updated April, 2018
Welcome to Manifold Scholarship.
The intuitive, collaborative, open-source platform for scholarly publishing you’ve been waiting for.

Why Manifold?

Manifold receives $789,000 phase two grant from The Andrew W. Mellon Foundation!