

# Understanding Digital Humanities:

Week 2

**Close reading** versus **Distant Reading**

- Google's goal: "to organize the world's information and make it universally accessible and useful." (1999)
- In 2004, Google launched the Google Print Project, renamed the **Google Book Search Project** in November 2005
- Google's intention was to digitize the world's collection of books
- Google planned to scan as many books as possible and make these scans available to users online
- Potentially revolutionary for research/for society?

Page 360

She coloured as she spoke; but neither that, nor any thing else, awakened a suspicion of the **truth**. The evening passed quietly, unmarked by any thing ...

No preview available for this page. [Buy this book.](#)

[Page 5 »](#)

**I**t is a **truth** universally acknowledged, that a single man in possession of a good fortune must be in want of a wife. However little known the feelings or views of such a

[Page 197 »](#)

cusation, of having injured Mr. Wickham, I can only refute it by laying before you the whole of his connection with my family. Of what he has *particularly* accused me I am ignorant; but of the **truth** of what I shall relate I can summon more than one witness of undoubted veracity. Mr. Wickham is the son of a very respectable man, who had for many years the

[Page 200 »](#)

could not be in your power, and suspicion certainly not in your inclination. You may possibly wonder why all this was not told you last night. But I was not then master enough of myself to know what could or ought to be revealed. For the **truth** of every thing here related, I can appeal more particularly to the testimony of Colonel Fitzwilliam, who, from our near relation-

# Utopian dreams

- “Scan This Book!, Kevin Kelly, *The NYT Magazine*, 2006
- “What happens when all the books in the world become a single liquid fabric of interconnected words”
- No book will be an island

# Ngram/ <https://books.google.com/ngrams>

- **Google Books Ngram Viewer** is an online search engine that charts frequencies of any set of search strings using a yearly count of [n-grams](#) found in sources printed between 1500 and 2019.
- **Culturomics**: a form of computational lexicology that studies human behavior and cultural trends through the quantitative analysis of digitized texts. **Measure culture?**
- <http://lab.kbresearch.nl/find/Ngrams>
- <https://bookworm.htrc.illinois.edu/develop/>

# Google Books Ngram Viewer

Q Bovary,Sherlock Holmes,Frankenstein

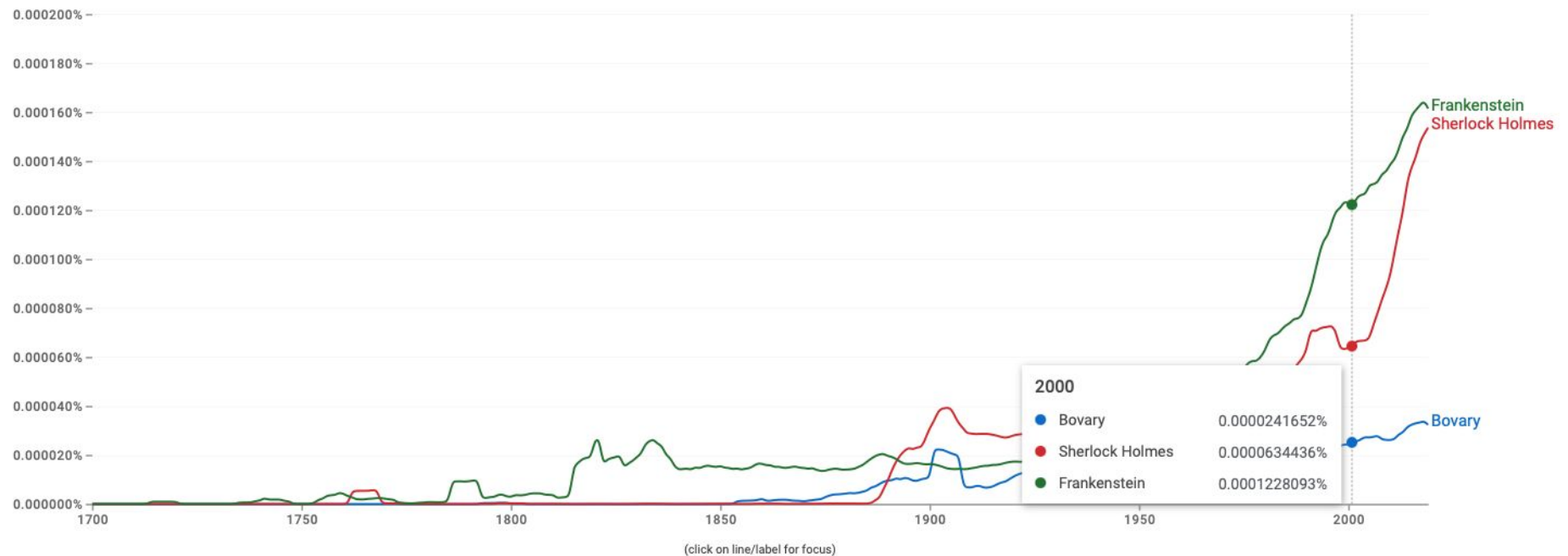


1700 - 2019

English (2019)

Case-Insensitive

Smoothing



Moretti & *distant reading*!  
(2000)

Moretti: *Stop reading books!*

# Franco Moretti & his Distant Readings

- What does Moretti mean by that term?
- Does it generate...
  - new science?
  - new research questions?
  - new methods? (can we read a text by not reading texts?)
  - different answers/new knowledge?



# “Observed from a distance”

Primo Levi's short story (1968-70)

“3. NON-ASTRONOMICAL PERIODICITY. A certain number of phenomena observed on Earth follow a seven-day rhythm. [...] An astronomical rhythm is certainly not involved: as is well-known, neither the Earth's month (synodic or sidereal) nor its year (solar or sidereal) contains a number of days in multiples of seven.”

# “Observed from a distance”

Primo Levi’s short story (1968-70)

“The phenomena that we call SDE (Seventh Day Events), which primarily concern the Cities and their immediate environs, take place simultaneously across the entire surface of the Earth—all, of course, subject to their local times.

...activity and rest (or vice versa) follow each other in segments of six days and one day.”

# “Observed from a distance”

Primo Levi's short story (1968-70)

“SDE ANOMALIES. During the past few months, in some zones of North Africa, Southern Asia, and the Malay Archipelago, the SDE phenomena appear two days earlier with respect to the rest of Earth, and one day earlier on the narrow isthmus that links Asia with Africa. In the British Isles they appear instead over the sixth and seventh days”

By the way....

- What is close reading?

*Close reading is a method for:*

- analyzing (themes, style, tone, argument...)
- Uncover layers of meaning that lead to deep comprehension
- appreciating the context of a given piece of writing

# Moretti on *close reading*

- “At bottom, it’s a **theological** exercise - very solemn treatment of very few texts”.
- Distant reading: distant is a condition of knowledge

# Moretti: *distant reading*

Understanding literature not by studying particular texts, but by aggregating and analyzing massive amounts of data.

Since no feasible amount of reading can fix that, what's called for is a change not in scale but in strategy.

Change of **how**, not of **what** we read

- “Focus on units that are much smaller or much larger than the text: devices, themes, tropes-or genre and systems”
- And if, between the very small and the very large, the text itself disappears...
- **Less is more/accept loss** (reality is rich, while concepts are abstract and poor)





# “formalism without close reading”

- “identifying a discrete formal trait and then following its metamorphoses through a whole series of texts.”
- E.g: **Devices** and **genres**, two formal units
- Focus not on texts: texts are real objects - but not objects of knowledge.

(From: Moretti 2000, “The slaughterhouse of literature”)

# “The slaughterhouse of literature”

- “Reading” the 99,5%
- the “great unread”

# Go against the grain

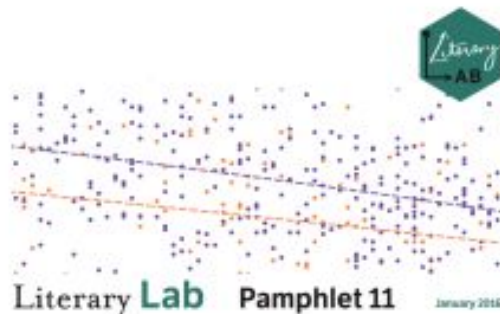
- “literature was indeed a system—but a system of *variations*. The system was one, not uniform.”
- “What quantitative methods have to offer the historians of literature [is] a *reversal of the hierarchy between the exception and the series*, where the latter becomes—as it is—the true protagonist of cultural life. A history of literature as a history *of norms*.”

# Lit Lab

- The Stanford Literary Lab is a research collective that applies computational criticism, in all its forms, to the study of literature.

<https://litlab.stanford.edu/>

## 11. Canon/Archive. Large-scale Dynamics in the Literary Field



Of the novelties introduced by digitization in the study of literature, the size of the archive is probably the most dramatic: we used to work on a couple of hundred nineteenth-century novels, and now we can analyze thousands of them, tens of thousands, tomorrow hundreds of thousands. It's a moment of euphoria, for quantitative literary history: like having a telescope that makes you see entirely new galaxies. And it's a moment of truth: so, have the digital skies revealed anything that

changes our knowledge of literature?

<https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf>

# New Galaxies

- Of the novelties introduced by digitization in the study of literature, the size of the archive is probably the most dramatic: we used to work on a couple of hundred nineteenth-century novels, and now we can analyze thousands of them, tens of thousands, tomorrow hundreds of thousands. It's a moment of **euphoria**, for quantitative literary history: **like having a telescope that makes you see entirely new galaxies.** And it's a moment of truth: so,
- **have the digital skies revealed anything that changes our knowledge of literature?**

- We may end up studying 200,000 novels instead of 200; but, they're all still novels. Where exactly is the novelty? 199,000 books that no one has ever studied – runs the typical answer – **how could there not be novelties?**  
**It's a whole new dimension of literary history**



- We must still **learn to ask the right type of questions**. But before doing so, something needs to be clarified. *Canon and archive*: what do we mean, by these two words?

- The published, the archive, and the corpus.
- The **published** is simple: it's the totality of the books that have been published (the plays that have been acted, the poems that have been recited, and so on). This literature that has become "public" is the fundamental horizon of all quantitative work (though of course its borders are fuzzy, and may be expanded to include books written but kept in a drawer, or rejected by publishers, etc.)
- The **archive** is that portion of published literature that has been preserved – in libraries and elsewhere – and that is now being increasingly digitized.
- The **corpus**, finally, is that portion of the archive that is selected, for one reason or another, in order to pursue a specific research project. **The corpus is thus smaller than the archive, which is smaller than the published.**

But with digital technology, the relationship between the three layers has changed: the corpus of a project can now easily be (almost) as large as the archive, while the archive is itself becoming – at least for modern times – (almost) as large as all of published literature.

When we use the term “archive”, what we have in mind is precisely this potential convergence of the three layers into one. (In theory/not in practice)

**Table 2. Most frequent words in titles for English, Spanish, Portuguese, Turkish, and French with English translations.**

	English		Spanish			Portuguese			Turkish		
	Words	Rel. freq.	Words	Translation	Rel. freq.	Words	Translation	Rel. freq.	Words	Translation	Rel. freq.
1	love	171.42	amor	love	125.69	amor	love	159.45	aşk	love	242.04
2	one	76.54	vida	life	70.62	Amigo	friend	143.09	hayat	life	70.02
3	life	57.10	amigos	friends	60.26	vida	life	96.71	kız	girl	66.10
4	girl	43.27	sueño	dream	46.79	diário	diary	86.30	sözler	words	46.66
5	shot	33.51	siempre	forever/ always	42.03	garota/ menina	girl	73.84	küçük	small	40.07
6	reader	31.27	nueva	new	41.55	querido	dear/ beloved	47.46	güzel	beautiful	39.54
7	boy	27.97	chica	girl	32.56	melhor	best	45.05	günlüğü	diary	38.36
8	friendship	25.73	corazón	heart	32.20	dias	days	35.35	hayal	dream	38.20
9	bad	23.42	enamorada	in love	29.54	sonhos	dreams	34.23	yeni	new	37.04
10	heart	21.27	día	day	28.92	estrellas	stars	32.64	karanlık	dark	30.28
11	new	20.84	chico	boy	28.84	novo	new	31.09	yaz	summer	29.86
12	little	18.70	sin	without	24.18	mundo	world	27.34	okul	school	28.48
13	lost	17.26	nunca	never	22.72	nada	nothing	25.70	çocuk	child	26.83
14	harry	17.17	quierer	to want/ love	21.72	nunca	never	24.84	son	last	26.28
15	BTS	16.62	mundo	world	20.42	irmão	brother	24.28	siyah	black	25.88
16	time	16.13	pensamientos	thoughts	19.36	sin	without	23.21	gerçek	real	25.56
17	best	14.99	dos	two	18.75	frases	sentences	22.68	ilk	first	24.66
18	never	14.59	ojos	eyes	18.42	coração	heart	22.41	sözleri	remarks	23.83
19	world	14.38	realidad	reality	16.71	vez	time	22.14	korku	fear/horror	21.46
20	secret	14.30	yaoi	Yaoi	15.54	filha	daughter	20.36	bts	BTS	21.25

### # 1 Pride and Prejudice (1813)



Fig 15. Graph with emotional arcs for *Pride and Prejudice* text and comments.

# Lev Manovich

- Database as a Symbolic Form (1998)
- The Language of New Media (2001)
- Soft Cinema (2005)
- Software Takes Command (2013)
- **Cultural Analytics/Software studies**



Theory:

- How can work with "big cultural data" help us to question standard assumptions and concepts in humanities and social sciences?
- How can we combine computational techniques and analysis of massive cultural data with more traditional humanities methodologies?
- What new theoretical concepts and models do we need to deal with the mega-scale of born-digital culture?
- What would "science of culture" driven by massive data look like, and what will be its limitations?
- How we can study big cultural data without aggregating and summarizing data (the standard approach of both classical statistics and contemporary data science)?
- How can analysis of human cultural artifacts and processes focus on variability and diversity, rather than only "typical" and "most popular"?

## Cultural analytics

Cultural analytics refers to the use of computational methods for the analysis of massive cultural data sets and flows

Software Studies Initiative/  
Software Studies lab

Breakout session



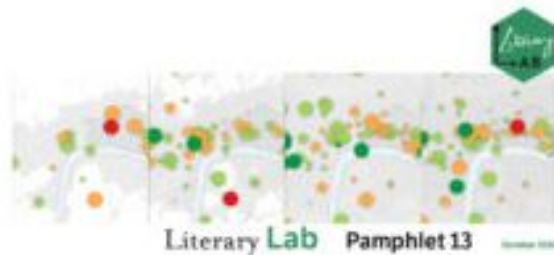
# 6 case studies

- What is the central research question?
- How large is the corpus?
- What method is used
- Do the author claim new findings?
- How do you evaluate this project?

# Output: 5 minutes oral report

1. brief presentation of the aim of the project and tools/methods used
2. explanation of why the project can be considered an example of distant reading
3. usefulness and limitation of using the concept of distant reading for the project: E.g. can this concept describe all the activities/parts of the project? Are other reading modalities used? How are they used together with distant reading?

## 13. The Emotions of London



“The Emotions of London”, written by Ryan Heuser, Franco Moretti, and Erik Steiner, inaugurates a new field of work for the Literary Lab — that of literary and cultural geography. Working on a corpus of 5,000 novels, and covering the two centuries from 1700 to 1900, this pamphlet charts the uneven development of social spaces and fictional structures, bringing to light the long-term connection between emotion and class in narrative representations of London.



Pamphlet 9  
→ March 2015

Literary Lab

Bankspeak:  
The Language  
of World Bank Reports,  
1946-2012

Franco Moretti  
Dominique Pestre

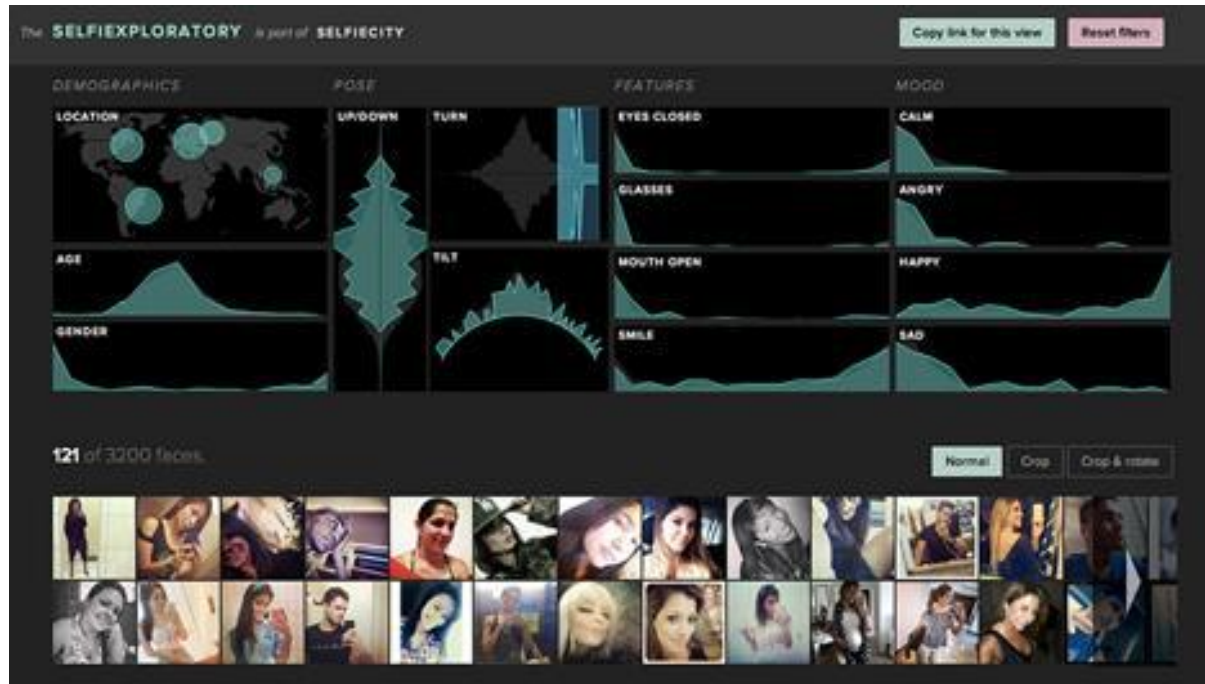
## 9. Bankspeak: The Language of World Bank Reports, 1946-2012

A literary historian and a sociologist of science analyze how the World Bank presents – and justifies – its role within the global economy. Focusing on the Bank’s semantic and grammatical patterns, Moretti and Pestre provide a path-breaking study of an “institutional” language, and of the neo-liberal rhetoric of recent decades.

<https://litlab.stanford.edu/LiteraryLabPamphlet9.pdf>

3

# selfiecity



<http://selfiecity.net>



## China Biographical Database Project (CBDB)


[HOME](#)
[About Us ▾](#)
[Methodology ▾](#)
[Sources ▾](#)
[Events ▾](#)
[Download ▾](#)
[Blog ▾](#)
[中文](#)

### METHODOLOGY

[Prosopography](#)
[GIS: Mapping and Analyzing](#)
[Social Networks](#)
[HOME](#) / [METHODOLOGY](#) /

## GIS: Mapping and Analyzing

One challenge in using CBDB data in prosopographical research is seeing the patterns in large amounts of data. To do this we can use statistical software and we can use geographic information system (GIS) software. For China's history the China Historical GIS (CHGIS) project makes datasets of the administrative system between 221 BC and 1911 AD and major non-administrative towns for 1820 and 1911 freely available. Using GIS software such as ArcGIS or MapInfo (or even GoogleEarth) we can combine CBDB output with CHGIS datasets. This paper illustrates some ways in which CBDB data can be linked and mapped in a GIS.

<https://projects.iq.harvard.edu/cbdb/gis-mapping-and-analyzing>

<https://fairbank.fas.harvard.edu/research/blog/introducing-the-china-biographical-database-cbdb/>




 Search for names, offences, places, Record ID or Life Archive ID

## Tracing London Convicts in Britain & Australia, 1780-1925

This website allows you to search millions of records from around fifty datasets, relating to the lives of 90,000 convicts from the Old Bailey. Use our site to search individual convict life archives, explore and visualise data, and learn more about crime and criminal justice in the past.

In **December 2019** this website was updated with the addition of [two datasets on convict tattoos](#) and a new search feature allowing you to **search by convict occupations**. A substantial number of additional **punishment outcomes** have been linked to trials. We have made a number of corrections to convict **genders and ages**, and new visualisations have been added to the [visualisation gallery](#).

**58,002 convicts in this database had tattoos.** You can now [search for specific tattoos](#) and for [specific types of tattoos on specific parts](#)

### Life of the week

#### Mina Jury b. c.1828-1890

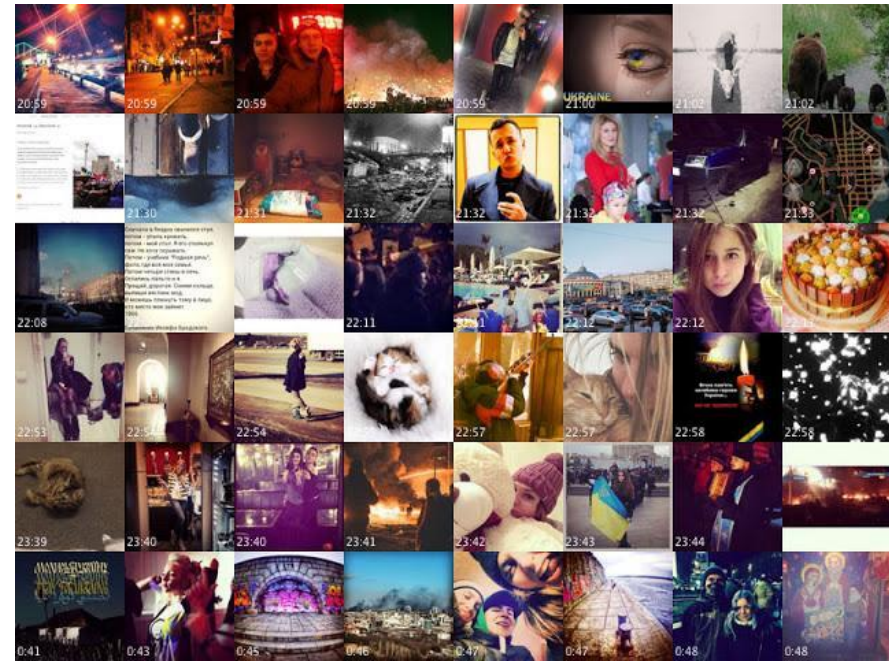
Mina jury was a habitual thief whose criminal career spanned more than four decades. Mina experienced both transportation to Australia and



<https://www.digitalpanopticon.org>

## "The Exceptional and the Everyday: 144 Hours in Kyiv"

Our project also **explores how during the exceptional events, the exceptional co-exists with the everyday**. The images of Maidan clashes, political slogans, and burned cars and buildings appear right next to everything else: selfies, parties, cultural events, etc. Through a number of different visualizations, we explore how the extra-ordinary and ordinary are connected visually (the use of images) and textually (tags).



<http://www.the-everyday.net/p/iconography-of-revolution.html>



Some reflections

- Tim Hitchcock, “Historyonics:  
Big Data for Dead people”

We are in danger of giving ourselves over to what sociologists refer to as 'problem closure' - the tendency to reinvent the problem to pose questions that available tools and data allow us to answer - or in Lewis Mumfords words, ask questions we know that computers can answer.

<http://historyonics.blogspot.nl/2013/12/big-data-for-dead-people-digital.html>

Tim  
Hitchcock,  
“Historyonics:  
Big Data for  
Dead people”

It feels to me as if our practise as humanists and historians is **being driven by the technology, rather than being served by it.**

And really, the issue is that while we have a strong theoretical base from which to critique the close reading of text - we know how complex text is - we do not have the same theoretical framework within which to understand how to read a space, a place, an object, (...) - **all suddenly mediated and brought together by code** - or to critique the reading of text at a distance.

Tim  
Hitchcock,  
“Historyonics:  
Big Data for  
Dead people”

For myself, I am wondering how we might **relate non-text to text more effectively**; and how we might combine - for historical purposes - **close and distant reading** into a single intellectual practise; how we might **identify new objects of study**, rather than applying new methodologies to the same old bunch of stuff.

.....how to ask questions of computers, without simply asking questions we know computers can answer.



## **The Proceedings of the Old Bailey, 1674-1913**

A fully searchable edition of the largest body of texts detailing the lives of non-elite people ever published, containing 197,745 criminal trials held at London's central criminal court. If you are new to this site, you may find the [Getting Started](#) and [Guide to Searching](#) videos and tutorials helpful.

<https://www.oldbaileyonline.org/>

A portrait of a woman with dark hair, wearing a dark shawl. The shawl has the text "1808 5" printed on it. The background is a mottled, aged grey.

Sarah Durrant, Convicted to two years imprisonment at Wandsworth Gaol, for the theft of £2000 worth of banknotes on 9 January 1871.

## The Proceedings of the OLD BAILEY

London's Central Criminal Court, 1674 to 1913

---


[Home](#)
[Search](#)
[About The Proceedings](#)
[Historical Background](#)
[The Project](#)
[Our Writ](#)
[Contact](#)

---

[In this Section...](#)

**SARAH DURRANT, Theft > pocketpicking, 9th January 1871.**

**Reference Numbers:** M271049-120  
**Offence:** Theft > pocketpicking  
**Hearings:** [Hears](#) > [in substance](#)  
**Periodicals:** [Contemporary](#) > [in substance](#)  
**Date NCR:** [Contemporary](#) [NCR Information](#)



127. **SARAH DURRANT (AKA)**, Stealing two Bank of England notes, for the payment of 1000£ each, the property of John Quincy House, and others: *David Cook—Pleading*.


M271049, PLEADING and/or testimony available for the Proceedings online. M271049 WILLIAM Mc Duffee

(JERRY FOLGER : I was a collecting clerk, in the service of Messrs. Barclay, Henry & Co., bankers, of Lombard Street—on the morning of Saturday, 9th February last, I went out on my collecting rounds—I received, at Messrs. Coutts' bank notice to bid—that one party paid by four notes, of 1000£ each, two of those notes were numbered: B2302 and B2303—when I received notes I write on them the name of the bank where I receive them, I did so in this case—where (gradually) are the two notes, and I wrote on them "Coutts, 9th February"—they were perfectly clean and quite new when I received them—besides the 1000£ note which I had received from Coutts' I had as my collect some more notes, amounting to 1000£ odd—I went from Coutts' to the British Bank, Chancery Lane—I put my cheque on the counter, and called out: "That's a charge!"—I took out my notes and ran quickly through to see that they were right, and then put them in my own coat—I found them right, and put the amount on a slip of paper—the note was open on the counter at that time, and my hand was on it—some person touched me and said "You have dropped something!"—I looked down and saw a piece of paper, like a cheque, all my feet—I stooped down and picked it up—there was nothing on it, and I put it down again—it was blank paper—I got up directly, and I noticed I saw a hand leaving the seat—I looked in the door and saw the notes were gone—I called out, and then ran to the door, and looked up and down the street, but I could see no one—the bank is close to some steps, where you go down into St Giles's Inn—the number of the notes was obtained from the different banks, and advertisements were published and bills issued, offering rewards to any person who should discover the thief.

[In this Section...](#)

**SARAH DURRANT, Theft > pocketpicking, 9th January 1871.**

**Reference Numbers:** M271049-120  
**Offence:** Theft > pocketpicking  
**Hearings:** [Hears](#) > [in substance](#)  
**Periodicals:** [Contemporary](#) > [in substance](#)  
**Date NCR:** [Contemporary](#) [NCR Information](#)



127. **SARAH DURRANT (AKA)**, Stealing two Bank of England notes, for the payment of 1000£ each, the property of John Quincy House, and others: *David Cook—Pleading*.


M271049, PLEADING and/or testimony available for the Proceedings online. M271049 WILLIAM Mc Duffee

(JERRY FOLGER : I was a collecting clerk, in the service of Messrs. Barclay, Henry & Co., bankers, of Lombard Street—on the morning of Saturday, 9th February last, I went out on my collecting rounds—I received, at Messrs. Coutts' bank notice to bid—that one party paid by four notes, of 1000£ each, two of those notes were numbered: B2302 and B2303—when I received notes I write on them the name of the bank where I receive them, I did so in this case—where (gradually) are the two notes, and I wrote on them "Coutts, 9th February"—they were perfectly clean and quite new when I received them—besides the 1000£ note which I had received from Coutts' I had as my collect some more notes, amounting to 1000£ odd—I went from Coutts' to the British Bank, Chancery Lane—I put my cheque on the counter, and called out: "That's a charge!"—I took out my notes and ran quickly through to see that they were right, and then put them in my own coat—I found them right, and put the amount on a slip of paper—the note was open on the counter at that time, and my hand was on it—some person touched me and said "You have dropped something!"—I looked down and saw a piece of paper, like a cheque, all my feet—I stooped down and picked it up—there was nothing on it, and I put it down again—it was blank paper—I got up directly, and I noticed I saw a hand leaving the seat—I looked in the door and saw the notes were gone—I called out, and then ran to the door, and looked up and down the street, but I could see no one—the bank is close to some steps, where you go down into St Giles's Inn—the number of the notes was obtained from the different banks, and advertisements were published and bills issued, offering rewards to any person who should discover the thief.

[In this Section...](#)

**SARAH DURRANT, Theft > pocketpicking, 9th January 1871.**

**Reference Numbers:** M271049-120  
**Offence:** Theft > pocketpicking  
**Hearings:** [Hears](#) > [in substance](#)  
**Periodicals:** [Contemporary](#) > [in substance](#)  
**Date NCR:** [Contemporary](#) [NCR Information](#)



127. **SARAH DURRANT (AKA)**, Stealing two Bank of England notes, for the payment of 1000£ each, the property of John Quincy House, and others: *David Cook—Pleading*.


M271049, PLEADING and/or testimony available for the Proceedings online. M271049 WILLIAM Mc Duffee

(JERRY FOLGER : I was a collecting clerk, in the service of Messrs. Barclay, Henry & Co., bankers, of Lombard Street—on the morning of Saturday, 9th February last, I went out on my collecting rounds—I received, at Messrs. Coutts' bank notice to bid—that one party paid by four notes, of 1000£ each, two of those notes were numbered: B2302 and B2303—when I received notes I write on them the name of the bank where I receive them, I did so in this case—where (gradually) are the two notes, and I wrote on them "Coutts, 9th February"—they were perfectly clean and quite new when I received them—besides the 1000£ note which I had received from Coutts' I had as my collect some more notes, amounting to 1000£ odd—I went from Coutts' to the British Bank, Chancery Lane—I put my cheque on the counter, and called out: "That's a charge!"—I took out my notes and ran quickly through to see that they were right, and then put them in my own coat—I found them right, and put the amount on a slip of paper—the note was open on the counter at that time, and my hand was on it—some person touched me and said "You have dropped something!"—I looked down and saw a piece of paper, like a cheque, all my feet—I stooped down and picked it up—there was nothing on it, and I put it down again—it was blank paper—I got up directly, and I noticed I saw a hand leaving the seat—I looked in the door and saw the notes were gone—I called out, and then ran to the door, and looked up and down the street, but I could see no one—the bank is close to some steps, where you go down into St Giles's Inn—the number of the notes was obtained from the different banks, and advertisements were published and bills issued, offering rewards to any person who should discover the thief.

[In this Section...](#)

**SARAH DURRANT, Theft > pocketpicking, 9th January 1871.**

**Reference Numbers:** M271049-120  
**Offence:** Theft > pocketpicking  
**Hearings:** [Hears](#) > [in substance](#)  
**Periodicals:** [Contemporary](#) > [in substance](#)  
**Date NCR:** [Contemporary](#) [NCR Information](#)



127. **SARAH DURRANT (AKA)**, Stealing two Bank of England notes, for the payment of 1000£ each, the property of John Quincy House, and others: *David Cook—Pleading*.

M271049, PLEADING and/or testimony available for the Proceedings online. M271049 WILLIAM Mc Duffee

(JERRY FOLGER : I was a collecting clerk, in the service of Messrs. Barclay, Henry & Co., bankers, of Lombard Street—on the morning of Saturday, 9th February last, I went out on my collecting rounds—I received, at Messrs. Coutts' bank notice to bid—that one party paid by four notes, of 1000£ each, two of those notes were numbered: B2302 and B2303—when I received notes I write on them the name of the bank where I receive them, I did so in this case—where (gradually) are the two notes, and I wrote on them "Coutts, 9th February"—they were perfectly clean and quite new when I received them—besides the 1000£ note which I had received from Coutts' I had as my collect some more notes, amounting to 1000£ odd—I went from Coutts' to the British Bank, Chancery Lane—I put my cheque on the counter, and called out: "That's a charge!"—I took out my notes and ran quickly through to see that they were right, and then put them in my own coat—I found them right, and put the amount on a slip of paper—the note was open on the counter at that time, and my hand was on it—some person touched me and said "You have dropped something!"—I looked down and saw a piece of paper, like a cheque, all my feet—I stooped down and picked it up—there was nothing on it, and I put it down again—it was blank paper—I got up directly, and I noticed I saw a hand leaving the seat—I looked in the door and saw the notes were gone—I called out, and then ran to the door, and looked up and down the street, but I could see no one—the bank is close to some steps, where you go down into St Giles's Inn—the number of the notes was obtained from the different banks, and advertisements were published and bills issued, offering rewards to any person who should discover the thief.

[In this Section...](#)

**SARAH DURRANT, Theft > pocketpicking, 9th January 1871.**

**Reference Numbers:** M271049-120  
**Offence:** Theft > pocketpicking  
**Hearings:** [Hears](#) > [in substance](#)  
**Periodicals:** [Contemporary](#) > [in substance](#)  
**Date NCR:** [Contemporary](#) [NCR Information](#)

[illegible]

<https://www.oldbaileyonline.org/browse.jsp?foo=bar&path=sessionsPapers/18710109.xml&iv=t18710109-137&xml=yes>

locate her words, and her image, her imprisonment and experience, both in 'text' and in the notes of the administration of a trial, as marked up in the XML

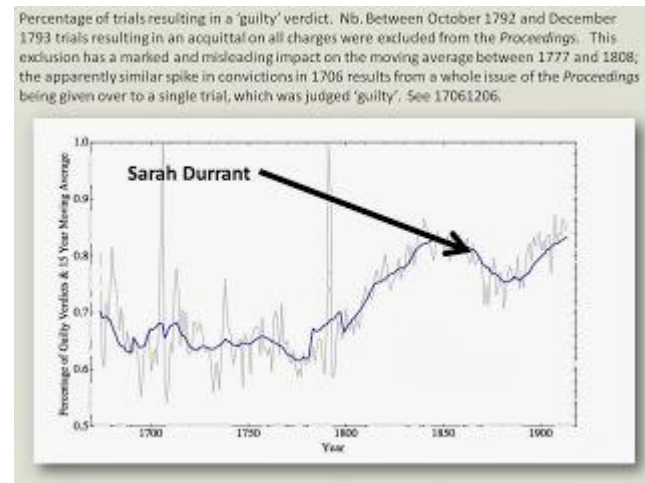
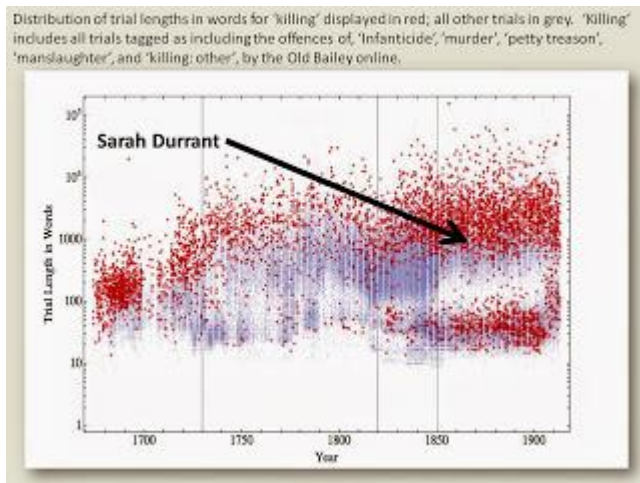








But we can go further still. We could **contextualise Sarah's experience** among that of some **240,000 defendants** like her, brought to trial over 239 years at the Old Bailey, and reported in 197,475 different accounts. We can visualize these trials by length, and code them for murder and manslaughter, or we could just as easily do it by verdict, or gender, punishment, or crime location. The following material is the outcome of a joint research project with William Turkel at the University of Western Ontario.



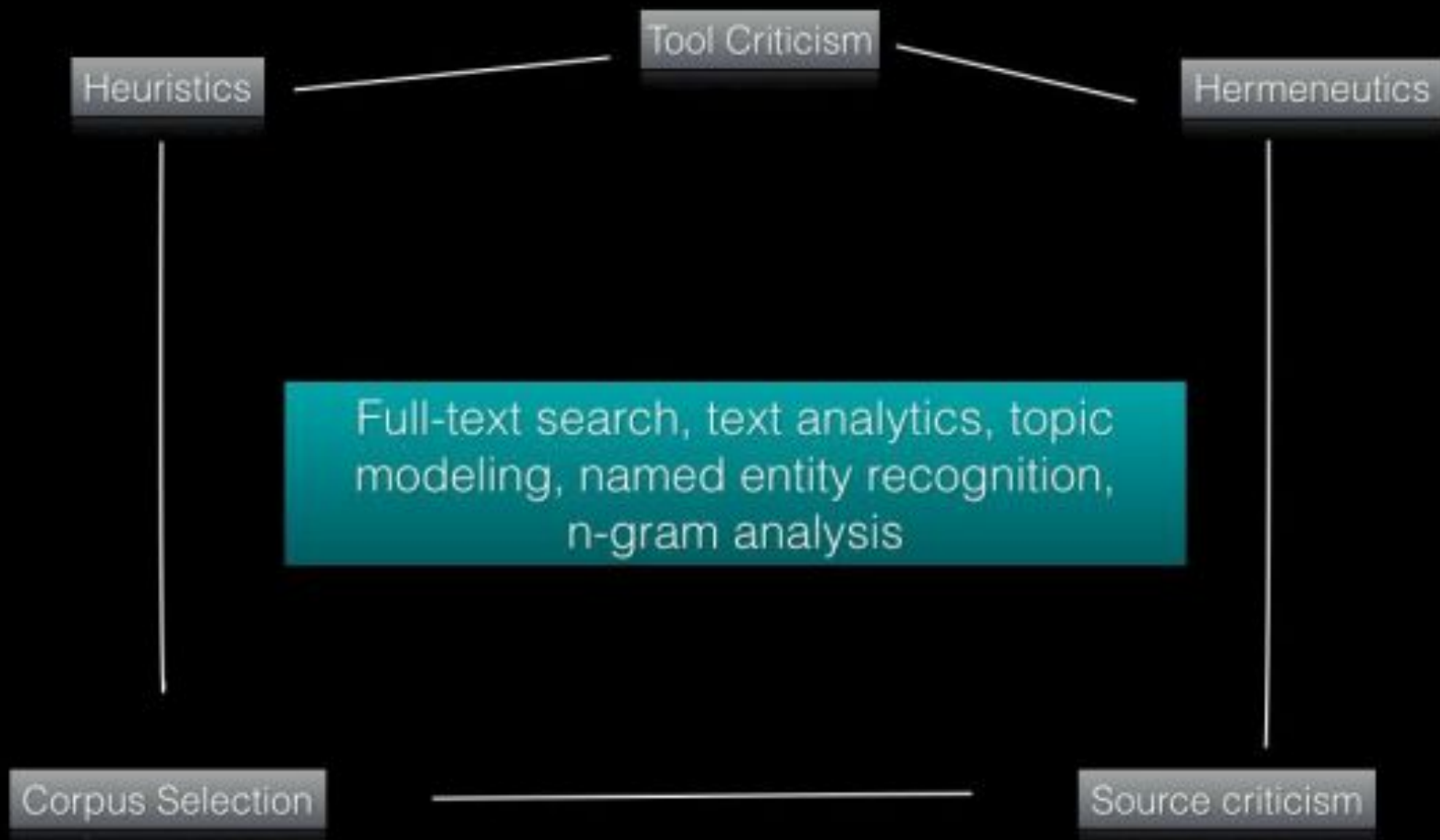
<http://criminalintent.org/2011/08/white-paper-on-criminal-intent-project/>

In other words, the creation of new tools and bodies of data, have allowed us to 'read' this simple text and the underlying bureaucratic event that brought it into existence, and arguably some of the social experience of a single individual, in a series of new ways. We can do 'distant reading', and see this trial account in the context of 127 million words - or indeed the billions of words in Google Books; and we can do a close reading, seeing Sarah herself in her geographical and social context.

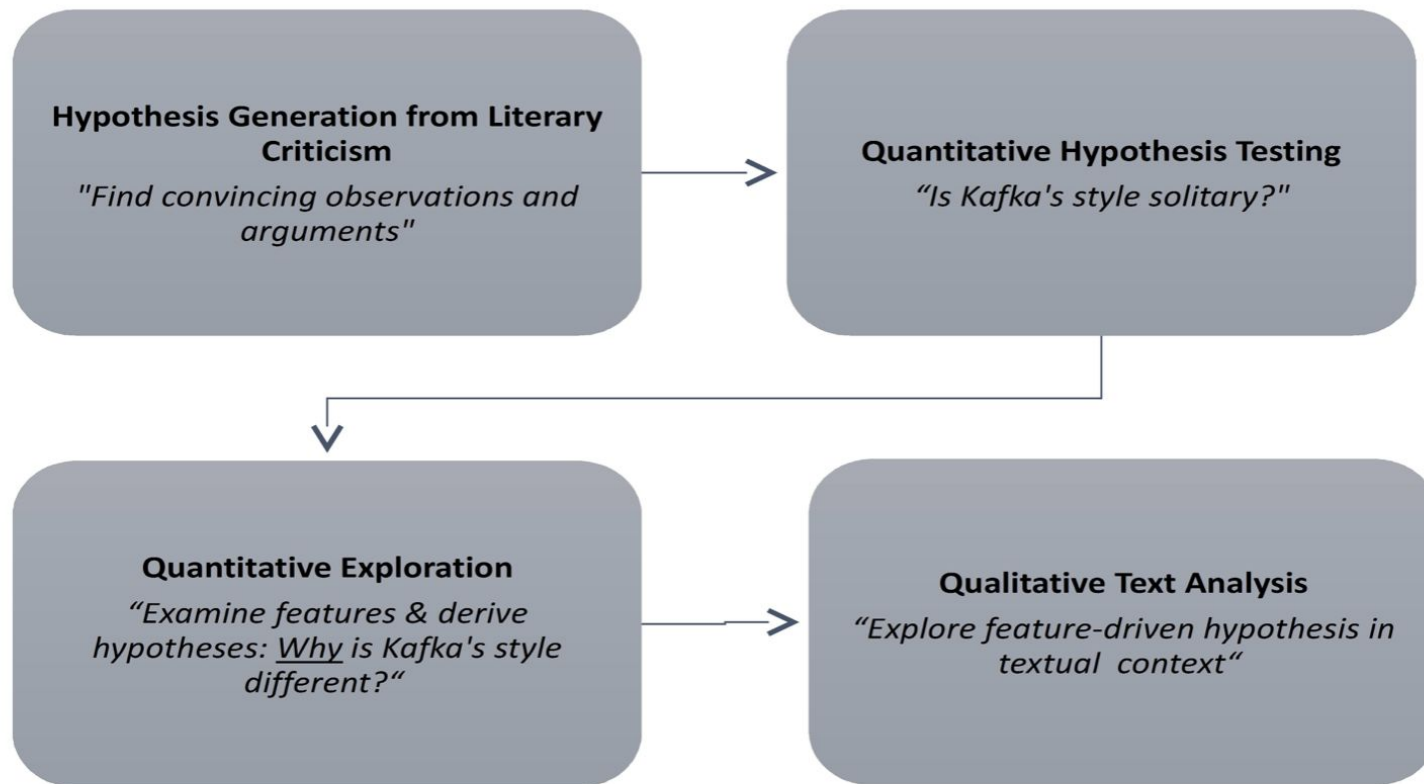
The **Digital Humanities Cycle** is a response to the plea made by numerous scholars that, after the Digital or Computation Turn (Nicholson 2013; Berry 2011), one should obtain ‘hybridity’, that is combine classic and digital historical research (Zaagsma 2013). When applied in combination, these techniques might yield patterns in the dataset, without missing “the power of the particular” (**Hitchcock** 2014). The Digital Humanities Cycle leads to a fruitful cooperation between “counting and understanding” (Rieder & Röhle 2012) – one of the prominent challenges when using digital tools.

(Jesper Verhoef and Melvin Wevers, University of Utrecht:  
<http://www.dhbenelux.org/wp-content/uploads/2015/04/20.pdf>

Figure 1: The Digital Humanities Cycle



# Mixed-methods in DH



**Figure 1.** Mixed-Methods Style Analysis of Kafka

J. B. Hermann, 2017. "In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics"  
<http://www.digitalhumanities.org/dhq/vol/11/4/000341/000341.html>

# Report

In the final report the main question should be:

- How do you evaluate the concept of distant reading as introduced by Moretti, Manovich, and Hitchcock and what are the main promises or obstacles you envision?

You should discuss this issue in relation to a specific DH project of your choice (you cannot choose one of the 6 projects presented in class). The report should include:

- brief presentation of the aim of the project and tools/methods used
- explanation of why the project can be considered an example of distant reading
- usefulness and limitation of using the concept of distant reading for the project: E.g. can this concept describe all the activities/parts of the project? Are other reading modalities used? How are they used together with distant reading?

You should make use of the required readings. Please use APA as your reference system. Upload this assignment A2 on Brightspace before Wednesday, September 20, 5PM. Max words: 600. You are encouraged to use figures and tables to present the chosen projects.