

Let DH Be Sociological!

Andrew Goldstone
Rutgers University

July 11, 2014
Digital Humanities 2014, Lausanne

outline

contextualize topic modeling

outline

contextualize topic modeling

(see title)

hot topic

Robert K. Nelson, “Mining the *Dispatch*,” 2010,

<http://dsl.richmond.edu/dispatch>

Scott Weingart and Elijah Meeks, eds., “Topic Modeling,” special issue, *Journal of Digital Humanities* 2, no. 1 (2012)

Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana: University of Illinois Press, 2013)

Timothy R. Tangherlini and Peter Leonard, “Trawling in the Sea of the Great Unread: Sub-corpus Topic Modeling and Humanities Research,” *Poetics* 41, no. 6 (December 2013)

John Laudun and Jonathan Goodwin, “Computing Folklore Studies: Mapping over a Century of Scholarly Production through Topics,” *Journal of American Folklore* 126, no. 502 (Fall 2013): 455–75

motivations

A larger literary history requires other skills: sampling; statistics; work with series, titles, concordances, incipits. (Moretti, “The Slaughterhouse of Literature,” 208)

Researchers must conduct searches that not only have high precision...but that also have high recall. (Tangherlini and Leonard, “Trawling in the Sea of the Great Unread,” 726)

motivations

A larger literary history requires other skills: sampling; statistics; work with series, titles, concordances, incipits. (Moretti, “The Slaughterhouse of Literature,” 208)

Researchers must conduct searches that not only have high precision...but that also have high recall. (Tangherlini and Leonard, “Trawling in the Sea of the Great Unread,” 726)

a turn away from the singularity and richness of literary texts and a concomitant refusal of the ethical charisma of the literary translator or messenger. (Love, “Close but not Deep,” 374)

Meaning is clearly a metavalue and also metaproblem for the digital humanities. (Liu, “The Meaning of the Digital Humanities,” 411)

a longer time-scale

corpus of generalist literary studies journals

PMLA (1889–2007)

Modern Philology (1903–2013)

The Modern Language Review (1905–2013)

The Review of English Studies (1925–2012)

ELH (1934–2013)

New Literary History (1969–2012)

Critical Inquiry (1974–2013)

1-gram counts from JSTOR Data for Research <http://dfr.jstor.org>

21367 “full-length article” documents, 1000 words or longer

9045 word types retained

43,773,430 tokens

Model: MALLET LDA with 120 topics, hyperparameter optimization.

world cultural culture	world life first	speech language words	desire body form	work view nature	students language education	jonson donne sir	keats ballad song	man old house	sublime romantic rousseau	france french paris	english translation latin
form eliot new	found passage use	story narrative	jews jewish jew	name irish ireland	modern early university	coleridge wordsworth mind	two first scene	name author work	first derrida two	sense experience memory	soviet russian russia
marrjage lady wife	white black african	dr life boswell	seems way made	esthetic art work	evidence seems fact	knight romance gawain	woolf miss virginia	images image myth	line verse lines	two sir said	classical renaissance rhetoric
christian god divine	interpretation meaning theory	political war national	religious church religion	freud body medical	goethe german germany	forms form oe	century great influence	text ms manuscript	two story tale	ms fol latin	pat pe hym
pastoral death dead	theater stage audience	tale chaucer troilus	lover love passion	scientific science human	order power makes	time first written	published edition printed	first poem poems	society social public	english political state	man life world
lost paradise satan	plays play drama	first book chapter	cinema film films	hamlet shakespeare play	picture repainting art	media new time	dickens novel novels	three two number	mind nature theory	money economic labor	words word meaning
form two different	poet poetry poetic	published review article	time first two	historical history time	philosophy human philosophical	legal law justice	death man make	sonnets sonnet ode	mas spanish spain	family father mother	letters letter corresponden
dryden pope english	quant grant tant	satire swift satiric	beowulf old english	court king queen	cette plus aux	literature literary new	place city land	sun yeats earth	italy city piu	university studies academic	political new cultural
reading text writing	robinson defoe captain	man moral good	animals animal fable	faerie spenser book	biblical christ new	fiction novel novels	female women sexual	light nature beauty	new american university	nur man noch	object image space
nature philosophy ideas	human world life	musical music song	shakespeare play plays	comedy comic irony	make way just	london new sir	greek epic classical	poe whitman baudelaire	language english linguistic	browning rev new	english medieval middle

credit and blame where due

AG and Ted Underwood,
“The Quiet Transformations of Literary Studies”
forthcoming in *New Literary History*
www.rci.rutgers.edu/~ag978/quiet

credit: TU

blame: AG

a broad context: recent topics

naive sense of “recent”

1. take topic proportions by decade
2. choose cutoff date: 1980
3. “recent” if topic proportions of each decade after cutoff are all be greater than proportions of each decade before cutoff (after 1920)

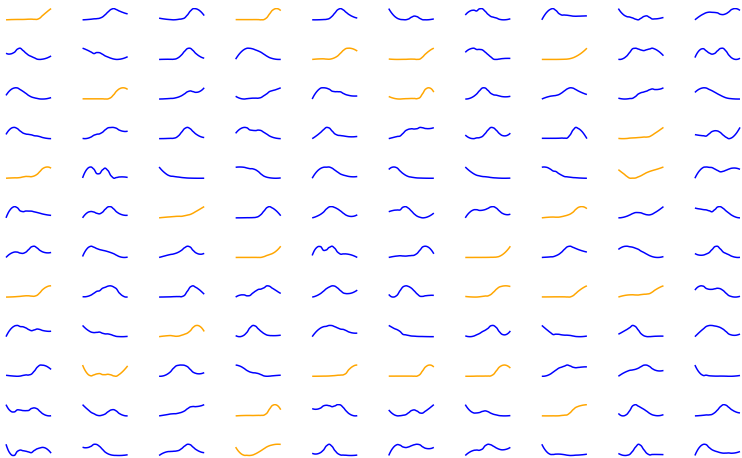


Figure: Time series of topics, orange if “recent”

A transformation of scholarship: topic most-frequent words

- cultural world culture national european
- body desire form power subject
- narrative story narrator events stories
- jewish jews jew hebrew israel
- early modern university use different
- derrida first two way question
- black white african racial slave
- war political national state military
- body freud medical psychoanalysis unconscious
- death pastoral dead life elegy
- science scientific human knowledge natural
- social society public individual class
- film cinema films camera photography
- new media time world modern
- economic money labor value economy
- history historical time past present
- human philosophy philosophical ethical life
- law legal justice rights trial
- father family mother child children
- city place land country journey
- studies university academic research scholars
- new political cultural social theory
- text reading writing reader texts
- women female sexual woman men
- image object space representation world
- way make just another point

coding scheme

S: Social or political topics, including national, ethnic, sexual, or gender identities;

T: Other thematic material, including religion, moral philosophy, love, nature, etc.;

F: Formal topics, including form, language, style, and genre;

NT: Non-thematic topics, including other languages, proper names, organizational labels, topics classifying textual studies, and clearly methodological discourses.

the shift

	not recent	recent
F	13	3
NT	56	4
S	4	14
T	21	5

the shift

	not recent	recent
F	13	3
NT	56	4
S	4	14
T	21	5

We risk adding the digital humanities to our proliferating disciplinary menus without any meaningful and substantial engagement with fields such as gender studies or critical race theory. (McPherson, “Why Are the Digital Humanities So White?”)

(aside: does the cutoff for “recent” matter?)

cutoff year	total recent	S not recent	S recent
1950	36	6	12
1960	35	3	15
1970	32	4	14
1980	26	4	14
1990	26	3	15
2000	21	10	8

polemic: no returns

polemic: no returns

dedicating the digital humanities to the soul of the humanities (Liu, “The Meaning of the Digital Humanities,” 42)

topic modeling as content analysis

Content analysis is a research technique for making replicable and valid inferences from texts...to the contexts of their use. (Krippendorff, *Content Analysis*, 24)

If the themes are sociological and the methods are sociological...