Personal Interview:
Dr. Matthew L. Jockers
(Susan J. Rosowski Associate Professor of English, University of Nebraska, Lincoln)
Thursday, September 17, 2015
2:00 pm (MST)
via Skype™

Mitch Ingraham
ENGL 5529: "Theory and Practice of 'Doing': From Digital Humanities to Post Humanities"
Dr. Lori Emerson
University of Colorado at Boulder
Fall 2015

Mitch Ingraham
8 December 2015
ENGL 5529
Dr. Emerson

Dr. Matthew L. Jockers is the Susan J. Rosowski Associate Professor of English at the University of Nebraska, Lincoln where he currently acts as a faculty fellow in the Center for Digital Research in the Humanities and Director of the Nebraska Literary Lab. In addition to teaching courses, conducting seminars and workshops, and authoring numerous articles, his publications include: *Text Analysis With R for Students of Literature* (2014) and *Macroanalysis: Digital Methods & Literary History* (University of Illinois Press, 2013). Together with Franco Moretti, he co-founded and directed the Stanford Literary Lab, where he worked from 2010 to 2012. Dr. Jockers received his B.A. from Montana State University (1989), a M.A. from the University of Northern Colorado (1993), and his PhD. from Southern Illinois University (1997). His areas of interest/specialties include: Digital Humanities: text mining/text analysis, Irish and Irish American Literature, 20th Century British Literature, and Literature of the American West.

MI: "When did you first become involved and interested in digital humanities: specifically as related to English literature?"

Dr. J: "Well … long before DH was ever a term, that's how. I probably discovered that there was a field of people doing computational quantitative work in the humanities in around 1990. Between '90 and '93 really, just before the birth of the Internet. And, of course, there was no term digital humanities, that doesn't come along until about 2005. The people at that point called themselves computing humanists, and I certainly wasn't part of that crowd until quite awhile later. In fact, I didn't even really discover that there was such a crowd or organization at that point. I was in my MA program at that point. I was a literature grad student who was sort of fascinated by computers and had that as a side hobby. I got pretty savvy with the computer during my master's program and when I went to do my PhD, my dissertation advisor learned that I had some computer savvy and he didn't. So, he asked me to be his RA and basically bought me out of my teaching for the last two years of the four years of my PhD program. So I started working for him in 1995. Just prior to that, of course, the internet is born in about 1993. I started dabbling in HTML and those kinds of things. One of the first projects I did for him was to create a digital archive. At Southern Illinois we had one of the really good Irish Studies collections and there was interest in digitizing that material and making it available online. So I started doing that, and built the first website for our Irish Studies collection there and, I mean, it was so crude. It had all the hallmarks of the early internet, that's for sure. But, the most important part of that work was that I started doing OCR and scanning text files. Charlie Fanning was my dissertation advisor and he was doing an edited collection of stories by James T. Farrell. And so he had me […] I was literally cutting the books up because in those days the OCR was really poor quality to begin with. And so if you cut the books up you could get the pages flatter on the scanner. So, I was cutting the pages and scanning them. Then we had all this digital text, which I then handed off to him and was finally ready for the scholarly editing. But I had a small corpus of texts. However, I was yet to realize that I could connect my interest in literature and my interest in

computing. These were just two things that I was sort of doing as different tracks. I was doing a minor concentration in Anglo-Saxon, and was doing a Beowulf course where I ended up writing a paper that was analyzing patterns in the meter of Beowulf. I did it all by counting it by hand– the meter and the rhythm. And I was doing this big statistical analysis of all of this because I was trying make an argument that when the Beowulf poet makes a direct address to his audience, the meter changes. And that that symmetrical meter was somehow more powerful than the asymmetrical meter. Anyway, it was after doing that that I realized 'there's got to be a better way.' So I had all this digital text and an interest in computing and mathematics, and the two things began to come together at the end of my PhD.

My first job was a university administrative job where my title was Coordinator of Curriculum and Technology. This was at Greeley (UNC). At that job I had to learn a lot of new stuff and one of the things I learned was databases. I learned how to build relational databases. I also at that point started doing my first real programming and that was with HyperCard. From that I quickly advanced to PHP. The first text analysis programs that I wrote were in PHP. PHP would allow you to embed HTML in it. PHP stands for 'personal home page.' You could weave PHP code inside your HTML so that you could have dynamic content. And so I was teaching, but I was also an administrator and coordinator of technology. I was there for three years and then I went to Stanford for another weird, hybrid academic job as a technology specialist where my mandate was to collaborate with faculty in the English department and to help them leverage technology for their academic research. After about six months there, I was appointed as a lecturer and then I taught for the remaining ten and a half years that I was there. That was awesome because I was really forced to take on and learn a lot of new stuff."

MI: "And was most of this self-taught? I mean, DIY? Or did you have any formal training?"

Dr. J: "I took some JAVA programming courses because I wanted to learn a hard language. At one point I thought that I would start coding everything in JAVA, and then I realized what a pain in the neck that is. It's just not very agile– you can't work very quickly. So then, I got tired of PHP because it was very messy and very much geared toward web programming. And so I picked up Python at that point– Python was awesome. By this time I was already doing lots of text analysis and my own research. In either 2002 or 2003 I was working with some folks who were doing all this data analysis in R. So I would write stuff in Python, or even PHP, and save it in some object that I could then load into R. Because R was just so awesome at doing the post-processing analysis work. So, at some point, I got tired of working in two or three different languages, and I said, 'I'll bet I could just do all of this in R.' I do most everything in R now."

MI: "What advantages does R offer over other programming languages?"

Dr. J: "Well, the obvious ones to choose if you're doing text analysis […] if you really want to be hardcore, is Java. For example, the Natural Language Processing group at Stanford– they're probably, I think, the best NLP group around. Everything that they produce is done in Java. Java, though, is a pain. So then, your next choices are Python, maybe Ruby, R, PHP maybe. But, really the big contenders are Python and R. And, why did I pick one over the other . . . ? It's religion at this point. For me though, R offers better data analysis. It was designed by statisticians for doing data analysis, so that's what it's really good at."

MI: "You've collaborated closely with a number of noteworthy scholars, namely, with Franco Moretti, with whom you helped found the Stanford Literary Lab. What other sources of inspiration and influences impacted your decision to pursue data mining as a means for textual analysis?"

Dr. J: "In terms of literature, I was a Joycean first. *Ulysses*, that's my favorite book. Joyce, and I think there is a connection here: if you look at the scholarship in Joyce studies, it tends to be very meticulous, very detailed. Joyce lays out puzzles for us that have solutions and that's sort of what computing is all about: it's trying to puzzle your way through and find a solution. So I like problems and I like solving problems. There's a big difference between James Joyce and D.H. Lawrence, in terms of what they're doing. I'm attracted to Joyce. In terms of literary heroes: Joyce, and Irish literature in general. You mention Moretti, and that was an awesome partnership, those years. When I left Stanford, leaving that lab and the environment that we had fostered there, that was tough– that was the hardest thing to leave. But, before that, going back even further, there were a number of excellent scholars working in quantitative, computational text analysis who nobody had ever heard of– except for this small group that called themselves humanities computing. People like John Burroughs, who wrote the book: "Computation Into Criticism: the Study of Jane Austen," and Willard McCarty. My first entry into text analysis was authorship attribution work. I found the work that those folks were doing incredibly interesting: the idea that you could potentially extract a stylistic signal. And my first big authorship attribution paper was done in collaboration with an environmental biologist and a statistician. That was a lot of fun, and the paper was very controversial. We did a study of the Book of Mormon. That work was very rigorous, and it was quantitative which I found very attractive. The problems were problems that, I think, people are interested in and care about: not just literary people. There's considerable skin in the game, as it were."

MI: "What does digital humanities mean for you as a practice? As a concept?"

Dr. J: "I'd take you back to 2011, when Glen Worthy and I hosted the digital humanities conference at Stanford. At that time, the DH world was just sort of beginning the surge that we've seen in the past say, four years or so. There were hints of a divide in the community. And, at that time, the nature of this sort of nascent, latent divide was the difference between those who wished to study the traditional objects of the humanities using computational tools (me), and those who wished to study digital objects using the traditional methods of the humanities. Glen and I wanted very much to have a 'lovefest' where everybody would get along, so we styled the theme of the conference, "Big Tent Digital Humanities." The idea that everybody's welcome– we even branded the conference with a 'summer of love' theme. We really believed that. It's funny now, because I think that now people ridicule that whole notion of a big tent, and it's become very divisive: the different camps. So much so that . . . I'm still a believer in the 'let's all get along,' but there are a lot of folks who aren't, I think. Maybe their modus operandi is to cause trouble and stir the pot, I don't know. We have room for lots of voices, but the trouble with this is that . . . and we saw this actually at the MLA conference where there was this panel called 'the Dark Side of Digital Humanities.' This drew a massive crowd because everybody wanted to know. And then the panel got up there and basically the dark side of digital humanities was a picture of digital humanities that none of us that had been in the field for a long time thought was

digital humanities at all. It was […] there was an attack on MOOCs, well, that's not digital humanities– not the way that the field had grown up. Now there are many creation stories about the field, as well. I'm partial to mine, but others have come into the field their way too. So, there are multiple stories. The moral of the story is that we have no idea what digital humanities is, nobody knows, and frankly I think it's lost its power as a term to define anything. It's become a very useful term from a political standpoint … because it has cachet. But, personally, I don't use the term anymore."

MI: "What do you prefer instead?"

Dr. J: "Well, I think of myself as someone who does text mining. I'm a text miner with an interest in mining literary material."

MI: "Some people would consider your approach to studying literature pretty radical. Going back to what it does, how it works, what you do: you've applied it to theme, style, influence, genre, gender, what else have you considered working with?"

Dr. J: "When I wrote *Macroanalysis*, I basically focused on two of five things: style and theme. So, everything in the main part of the book is about stylistic and thematic analysis. The nationality, the gender, the time are all just metadata that we then filter. The other three things are, I mean, when we think about […] let's take ourselves out of a literature class and put ourselves in a creative writing class. What are the five things they're going to talk about? Style, theme, plot, setting, and character– and maybe emotion, or sentiment. At that point, when I was working on *Macroanalysis*, style and theme were pushing my knowledge: that's where I was and what I was competent working with at that point. But, as the book was going through editing and to the press, I was already working aggressively on setting, character, and emotion and how to mine and extract those elements from the text. So, I think the first output of that research was a twenty-minute lecture that I gave at the first meeting of the Digital Classicists Association at Buffalo. And this was a talk in which I was showing how I'd extracted the thematic stuff that I already had, and setting and emotion. I was trying to correlate those three so that when you study a corpus you can say that when writers in this corpus are writing about Ireland and the topic is tenant and landlord relationships: what is the sentiment that is associated with those moments in the text? Are they being portrayed in a positive way or a negative way?

MI: "So is that syntactic? Or … ?"

Dr. J: "You can use topic modeling to find emotions and the same approach to find settings. In *Macroanalysis* I modeled nouns to get themes. If you model adjectives you can get emotions. That's still very much what I've been working on. What happened along the way though is that I discovered that using sentiment analysis to grab the emotional language from a text, that the trajectory of sentiment in a book serves as a pretty good proxy for plot movement. And so that's what I've been working on for the past couple of years. I'm collaborating with someone else on that project now. ['Syuzhet'] was the software that I wrote for doing all this, for extracting these latent plot shapes. The so-called sequel to *Macroanalysis* is not going to happen. I had an outstanding PhD student at Stanford, and she ended up writing an awesome dissertation on bestsellers. We did a lot of computational work for that dissertation; basically trying to determine

if bestsellers had a unique signal that you could identify in a larger corpus– and, yes, we could find it. We could basically identify the bestsellers within a corpus with about 80% accuracy. So, after she finished writing her dissertation, she pulled me aside and asked if I would be interested in writing a book with her. We wrote up a proposal for a trade non-fiction book and it was bought by St. Martin's. So, we're working on the book right now. All of my work on character, plot, and setting is going to find its way into that book instead of going into another dense scholarly book."

MI: "Do you incorporate digital humanities into your pedagogical practice? And, if so, how?"

Dr. J: "Digital humanities doesn't really have a meaning for me; as a term or anything else. I mean, I get it. We have a big DH program here at Nebraska and we teach a course, which is "Readings in Digital Humanities," and we have an "Introduction to Digital Humanities" and when I teach those courses I know which books I'm going to assign. It's not like I'm rejecting the whole idea, but my research and teaching is very much focused on quantitative text analysis. And, fortunately I have four other colleagues in the English department who are also DH people and so they handle other things: Amanda Gailey works on editions and archives and Ken Price, something similar– we're very fortunate here."

MI: "What are your thoughts on the 'pushback' against the digital humanities? Those that think it's somehow inimical to humanities?"

Dr. J: "A lot of people who have the knee-jerk criticism of digital humanities don't know what it means. And, given that I don't know what digital humanities means, I can understand why they wouldn't. So, what happens is– and this is the dangerous thing– because DH is this term that has all this cachet and buzz, people attach it to all kinds of garbage. And they attach it to good things too– don't get me wrong. So, somebody who's not sympathetic sees something with the DH name attached to it and then says 'oh, this is just crap.' Honestly, I just don't want to get involved in that sort of thing. I think its folly for anyone to critique digital humanities as if it were a thing that could be critiqued. I think that the work we do, under whatever banner you want to put it under, needs to stand on its own. So, the individual paper, the individual book– that's what we should be critiquing, not some vague […] There is a huge difference between 'new criticism' and digital humanities– those are two totally different things. 'New criticism' is a label that refers to a particular set of practices. Digital humanities is nothing like that. So you could critique 'new criticism' in ways that you could never critique what's become known as, digital humanities."

MI: "If you had to speculate, how do you view the future of the digital humanities?" and/or "Are there any last take-away points that you'd like to pass along to our class?"

Dr. J: "I wouldn't want to comment on the future of the digital humanities so much as I'd rather comment on the future of text mining in the humanities […] I think that the creation of all this digital text in combination with computing power and what we know now about language and processing language […] there's no shortage of things to work on […] I'm a little nervous about what's becoming of digital humanities. Some of the divisive things that are going on– I don't

know if that's quite the right word, but there's tension. And, I have colleagues who say, 'oh, that's to be expected: anytime something gets to be a certain size, there's tension.' That tension isn't exciting anymore. I want everyone to give everyone else a fair shake: that was the old ethos in the field, anyhow. I'd like to see more of that. You mentioned something just now about those that see this kind of work as being incompatible and I think those who see it as being incompatible aren't looking at it."