

The rise of digital humanities and the visualisation of data

Simon Musgrave, Monash University

I am a linguist by training. I have been interested in applying digital technologies to linguistic research over a number of years now - I've published in the area, I've got various projects running in that area. From that, I've expanded to be more interested generally in what we're all calling digital humanities, for want of a better term.

One of the things featured on figshare is a workshop we ran where one of our PhD students in Italian Studies came along and we started talking about some of the stuff he was working on - he's working on a Sicilian writer from the turn of the 20th Century, Luigi Capuana, and he was explaining to me that one of his research resources is the catalogue we could try doing some things with that. We could start putting some geodata in there, putting stuff on maps, doing some network analyses".

"There's a copy of the actual data there, the catalog, the geographic coordinates we've added for everything."

We're starting to see some interesting things happening. There's a copy of the actual data there, the catalog, the geographic coordinates we've added for everything, and then there's a couple of the first things we've produced: a map showing where in Europe Capuana's books were published. He was Sicilian although he didn't live all his life in Sicily, but one of the things we noticed when we did that log of this guy's personal library and I thought "Oh,


was that he had quite a lot of stuff that was published in Sicily. That sort of surprised us. We didn't think of late 19th Century, early 20th Century Sicily as a hub of the publishing industry, necessarily.

So, we decided to look at that in a little bit more detail - we've got a basic network analysis which relates locations for publication in Sicily to decades of Capuana's life. It's interesting - it shows that the major centres were centres for publication throughout the whole period, but then other places popped up at different times. It seems like it was a very fluid and unstable situation where maybe the local printer decided he'd publish a few books at some point.



Cluster map of the personal library of Luigi Capuana

I hope we'll do more analysis of the catalog data. I want to try and get some nicer graphics. We're also talking about expanding to look at Capuana and his associates in terms of their correspondence, and we're also interested in expanding it beyond just the Italian context. But Capuana had a very naturalistic style and that was a movement across



Europe, so it would be interesting to look at that from a European context. I'm not a literary scholar, I'm just the person who's fiddling with the data and trying to get some nice pictures out of.

“I can easily imagine that figshare will change the way we share our research.”

I'm associated quite a lot with documentation in linguistics, field linguistics and descriptive linguistics. The kind of stuff people are doing there is now all based on collections of video, now that it's easy to acquire that stuff at good quality. There are also various text data based on that, so transcriptions, annotations of different kinds. At the moment, that data goes into recognized archives and repositories - we're just, at the moment, considering how that can be recognized as part of an academic career.

This kind of material doesn't count as conventional publication but the amount of work that's involved in getting stuff to the stage where you put it in repositories is substantial. It's really a big effort.

I can easily imagine that figshare will change the way we share our research. Almost 10 years ago now, a colleague and I did some work on responding to some scholars based in the Netherlands who used clustering algorithms on some language data to do stuff which we didn't trust very much. We wrote a response to that - we wanted to be able to expose the data matrix we used, although we were publishing in a major journal, they had no facilities to do that at that time. Obviously, nowadays, I would straightaway go to figshare for that kind of thing, and I've got a couple of projects now where I can see that I would want to do that - that I would want to make data available there in addition to through a journal.

Get in touch:

figshare.com

info@figshare.com