Literature in the Digital Age: From Close Reading to Digital Reading

Video transcript

What does distant reading look like?

PHILIPP SCHWEIGHAUSER:

In 2010, Franco Moretti founded the Stanford Literary Lab. The name of the institution, Stanford Literary Lab, already indicates that it engages with scientific practices that are alien to many a literary scholar: experiments. Because that’s what you do in labs: scientific experiments. And in their experiments Moretti and his colleagues aim for, what Moretti calls for, namely the analysis of thousands of literary texts to identify recurring patterns and large-scale historical developments. The findings of the Stanford Literary Lab are published on its website in its pamphlet series. So let’s have a look at the first pamphlet published in this pamphlet series.

The first pamphlet of the Stanford Literary Lab was published in 2011. Its title is ‘Quantitative Formalism: An Experiment’. It’s co-authored by five scholars – or should I say scientists? The question the research group sought to answer was this: Could computers recognise literary genres? For instance, if you let a computer programme run through Nathaniel Hawthorne’s The Scarlet Letter, will it be able to identify it as what it is, namely, a Gothic novel?

The researchers used a programme called Docuscope that consists of 200 million words and phrases, each of which is assigned to one of 101 linguistic categories. Thus the words ‘he’ and ‘she’ are assigned to third person, and the words ‘sad’ and ‘happy’ are assigned to emotion. Now, if you let a computer programme run through your literary text, it will be able to tell you which of these linguistic categories occur with what frequency.

And then you can do the same thing with another literary text, and yet another literary text, and so on. And chances are that those literary texts that contain the same linguistic categories with a similar frequency belong to the same literary genre. Docuscope’s task was to match Gothic novels to Gothic novels, historical novels to historical novels, industrial novels to industrial novels, and so on, from within a corpus of 36 novels. And Docuscope, the computer programme, did its job very well. It could identify, it could recognise literary genres. Consider the one passage that Docuscope, the computer programme, identified as the most Gothic of all passages in the corpus.

This passage is not from Hawthorne’s The Scarlet Letter, but from Ann Radcliffe’s Gothic novel A Sicilian Romance. Let me read it to you.

‘A moment deserted him. An invincible curiosity, however, subdued his terror, and he determined to pursue if possible the way the figure had taken. He passed over loose stones through a sort of court till he came to the archway. Here, he stopped, for fear returned upon him. Resuming his courage, however, he went on, still endeavouring to follow the way the figure had passed, and suddenly found himself in an enclosed part of the ruin, whose appearance was more wild and desolate than any he had yet seen. Seized with unconquerable apprehension, he was retiring, when the low voice of a distressed person struck his ear. His heart sunk at the sound. His limbs trembled, and he was utterly unable to move.

The sound, which appeared to be the last groan of a dying person, was repeated. Hippolytus made a strong effort and sprang forward, when a light burst upon him from a shattered casement of the building, and at the same time, he heard the voices of men. He advanced softly to the window and beheld in a small room, which was less decayed than the rest of the edifice, a group of men who, from the savageness of their looks and from their dress, appeared to be banditti. They surrounded a man who lay on the ground wounded and bathed in blood, and who, it was very evident, had uttered the groans heard by the count.’

How would I identify this passage as Gothic? I would say that it’s the atmosphere of fear, terror, and distress. It’s the wild, desolate, and decaying setting. It’s the presence of archways and ruins, the protagonist’s trembling, the groans of the dying man, the dying man’s blood, and so on. All of this identifies this passage as Gothic.

Now, Docuscope found different things in the very same passage. The one level where Docuscope’s and my judgement agreed was that Docuscope also found that this passage contains many expressions connoting fear and sadness. But for all the rest, Docuscope found very different things. Docuscope identified this passage as Gothic because it reports many events. And it identified it as Gothic because it contains many personal pronouns such as ‘he‘ and ‘him‘. And it identified it as Gothic because it indicates many expressions that denote shifts in time. And Docuscope also considered it Gothic because it has many shifts back in time, flashbacks.

So what did the researchers find? They found not only that humans and computers can assign the same texts to the same genre, and they not only found that computers do this very, very differently from human beings. What they also found – and I think that’s the most significant finding – what they also found is that literary genres are characterised by shared patterns at the deepest linguistic level, that human eyes cannot see.