

Robot readers: what's the point?

'How would a robot read a novel?' was an event at this year's LSE Literary Festival that came about because I was interested in the possibility of using the methods of the social sciences to study fiction. Over at LSE's Methodology Institute (MI), researchers were using text analysis software to summarise the content of large bodies of text – government reports that were hundreds of pages long, transcripts of political speeches. The software, called Alceste, uses a proprietary algorithm to break down whatever text you feed it into between three and eight clusters of linked words. These clusters usually correspond to what we would call the main themes, or primary subject matter of the text.

The advantages of using Alceste are that it can digest a huge amount of text very quickly, and supply an accurate picture of what proportion of the text is devoted to what issues. Importantly, Alceste does this dispassionately: it won't be swayed by subject matter or be biased towards one group or another. It is a wholly impartial reader. We know that the program does a good job with non-fiction, I wondered: what would it make of a novel? The MI kindly agreed to feed its machine works of fiction. We tried lots of books including *Oliver Twist*, *Frankenstein*, *Pride and Prejudice* and even *The Da Vinci Code* – books which took days to read, Alceste read in minutes.

So how does a robot read a novel? In one sense, surprisingly well. The reading of *Moby Dick*, for example, separated out episodes in ways that tally quite well with readers' experiences of the novel. The period in Nantucket before Ishmael goes to sea is parcelled off into its own rather small cluster (occupying only 8 per cent of the text), whereas fully 42 per cent of the book is about whales. No surprises there. But the program is subtle enough to parse the history of whaling as a human institution (22 per cent) from the cultural and natural history of whales as an organism (20 per cent). So despite being sufficiently brusque to lump all of the reported speech in the novel into a single cluster, there are ways in which the reading can appear to be sensitive and astute.

But are these results really 'readings'? If we take that to mean, 'has the robot in any sense understood what it has read?' – then no, not at all. Alceste is meaning insensitive, it doesn't know what the words mean, just that they occur together in particular patterns. This makes its successes all the more remarkable. Of course,

the clusters still require interpretation, but you now have very good grounds on which to claim that *Moby Dick* is 40 per cent whales and whaling, and 18 per cent speech and dialogue.

The program allows you to have a synoptic, birds-eye view of the text. You can get a good sense of what the book is 'about,' but you won't necessarily be able to reproduce the plot from this: the robot reader has no interest in narrative order. This is really a major weakness, as the sequence in which information is disclosed to a reader is crucial to the experience of reading fiction. A murder mystery, for example, may hinge on the revelation of a single name. The robot would never pick up on this.

So what's the point? The point is you can use the methods of the social sciences to examine fiction, but the types of answers it offers are unlikely to answer the types of questions fiction raises for us. Equally, the summaries that Alceste usefully produces of non-fiction texts entirely subvert the aims of fiction. Why, after all, would you want an instant summary of a novel? Fiction isn't something we read in order to get at the message it contains. In fiction, the process of reading is the experience we seek. There are lots of tasks we might want to have automated because they are obstacles preventing us from pursuing our real interests. Reading fiction is one of those interests, not one of the obstacles. ■



Jon Adams

is a former research officer on the 'How Well Do "Facts" Travel?' project at LSE and is currently producing online videos for Communications.



SM/AIUEO/GETTY

LSE's second Literary Festival was held from 11 to 13 February 2010. Podcasts are available for many of the events, see lse.ac.uk/events