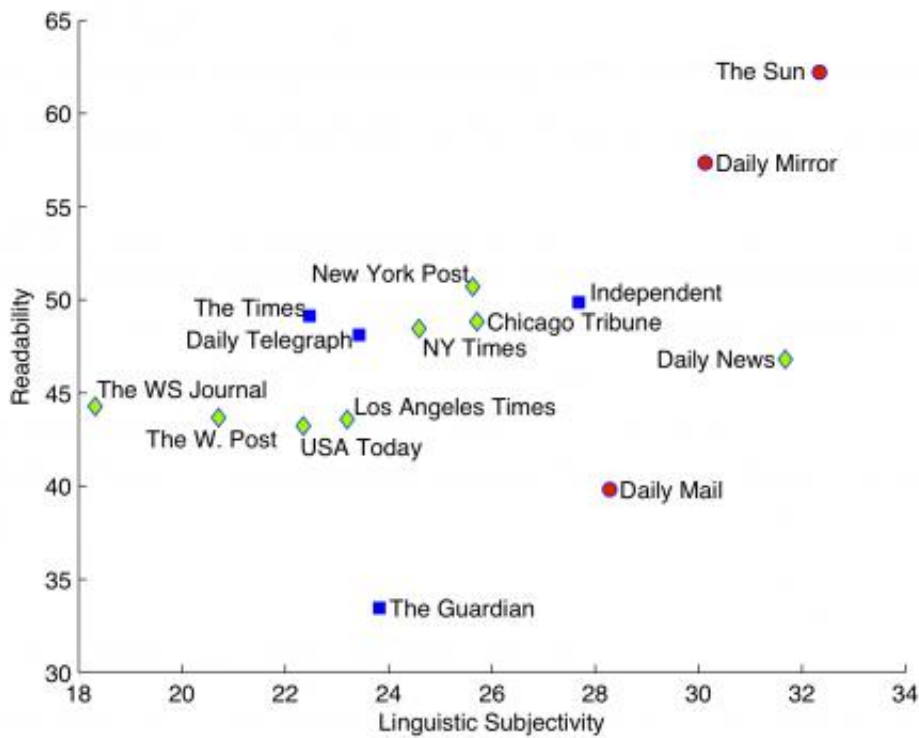# Scientists analyze millions of news articles

November 26 2012

Comparison of newspapers based on their readability and linguistic subjectivity. Credit: University of Bristol

Researchers in the UK have used artificial intelligence algorithms to analyze 2.5 million articles from 498 different English-language online news outlets over ten months.

A study led by academics at the University of Bristol's Intelligent

Systems Laboratory and the School of Journalism at Cardiff University have used Artificial Intelligence (AI) algorithms to analyse 2.5 million articles from 498 different English-language online [news](link) outlets over ten months.

The researchers found that:

• As expected, readability measures show that online tabloid newspapers are more readable than broadsheets and use more sentimental language. Among 15 US and UK newspapers, the Sun is the easiest to read, comparable to the BBC's children's news programme, Newsround, while the Guardian is the most difficult to read. 'Sport' and 'Arts' were the most readable topics while 'Politics' and 'Environment' were the least readable.

• The Sun is also the most likely to use adjectives with sentiment, while the [Wall Street Journal](link) uses the fewest emotional adjectives.

• The study found that men dominated the content of newspapers during the period analysed. The ranking of topics based on the [gender bias](link) of the articles found 'Sport' and 'Financial' articles were the most male biased, with sports news mentioning men eight times more often than women. 'Fashion' and 'Arts' were the least biased, with 'Fashion' articles being one the few topics featuring equal proportions of men and women.

• The most appealing topics to online readers were 'Disasters', 'Crime', and the 'Environment' while the least appealing topics were 'Fashion', 'Markets' and 'Prices'. The researchers also found that the popular articles tend to be more readable and more linguistically subjective.

Nello Cristianini, Professor of [Artificial Intelligence](link) at the University of Bristol, speaking about the research, said: "The automation of many tasks in [news content](link) analysis will not replace the human judgement

needed for fine-grained, qualitative forms of analysis, but it allows researchers to focus their attention on a scale far beyond the sample sizes of traditional forms of content analysis."

Professor Justin Lewis, Head of the School of Journalism, Media and Cultural Studies at Cardiff, added: "Even some of the more predictable findings give us pause for thought. The extent to which news is male dominated shows how far we are from gender equity across most areas of public life. The fact that articles about politics are the least readable might also explain widespread public disengagement."

The study is published online in *Digital Journalism*.

More information: Research methods in the age of digital journalism, Ilias Flaounas, Omar Ali, Thomas Lansdall-Welfare, Tijl De Bie, Nick Mosdell, Justin Lewis and Nello Cristianini, Digital Journalism, published online ahead of print 01 Nov 2012.
patterns.enm.bris.ac.uk/files/ … igitalJournalism.pdf

mediapatterns.enm.bris.ac.uk/A … OfMillionsOfArticles

Provided by University of Bristol