

## Formula to detect an author's literary 'fingerprint'

December 10 2009



Using literature written by Thomas Hardy, DH Lawrence and Herman Melville, physicists in Sweden have developed a formula to detect different authors' literary 'fingerprints'.

New research published today, Thursday 10 December, in <u>New Journal</u> of <u>Physics</u> (co-owned by the Institute of Physics and German Physical Society), describes a new concept from a group of Swedish physicists from the Department of Physics at Umel' University called the meta book which uses the frequency with which authors use new words in their <u>literature</u> to find distinct patterns in authors' written styles.

For more than 75 years George Kingsley Zipf's maxim, based on a carefully selected compilation of American English called Brown Corpus, suggested a universal pattern for the frequency of new words



used by authors. Zipf's law suggests that the frequency ranking of a word is inversely proportional to its occurrence.

New research suggests however that the truth behind word frequency is less universal than Zipf asserted and has more to do with the author's linguistic ability than any over-arching linguistic rule.

The researchers first found that the occurrence of new words in the texts by Hardy, Lawrence and Melville did begin to drop off in their texts as their book gets longer, despite new settings and plot-twists.

Their evidence also shows however that the rate of unique word drop-off varies for different authors and, most significantly, is consistent across the entire works of any one of the three authors they analysed.

The statistical analysis was applied to entire novels, sections from novels, complete works and amalgamations from different works by the same authors - they all had a unique word-frequency 'fingerprint'.

By using the statistical patterns evident from their study, the researchers have pondered the idea of a meta-book - a code for each author which could represent their entire work, completed or in the mental pipeline.

As the researchers write, "These findings lead us towards the meta book concept - the writing of a text can be described by a process where the author pulls a piece of text out of a large mother book (the meta book) and puts it down on paper. This meta book is an imaginary infinite book which gives a representation of the word frequency characteristics of everything that a certain author could ever think of writing."

More information: The meta book and size-dependent properties of written language, Sebastian Bernhardsson et al. 2009 *New J. Phys.* 11 123015 (15pp); doi:10.1088/1367-2630/11/12/123015



## Source: Institute of Physics (<u>news</u> : <u>web</u>)

Citation: Formula to detect an author's literary 'fingerprint' (2009, December 10) retrieved 28 April 2024 from <u>https://phys.org/news/2009-12-formula-author-literary-fingerprint.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.