Cracow, 21 January 2016

The world's greatest literature reveals multifractals and cascades of consciousness

James Joyce, Julio Cortazar, Marcel Proust, Henryk Sienkiewicz and Umberto Eco. Regardless of the language they were working in, some of the world's greatest writers appear to be, in some respects, constructing fractals. Statistical analysis carried out at the Institute of Nuclear Physics of the Polish Academy of Sciences, however, revealed something even more intriguing. The composition of works from within a particular genre was characterized by the exceptional dynamics of a cascading (avalanche) narrative structure. This type of narrative turns out to be multifractal. That is, fractals of fractals are created.

As far as many bookworms are concerned, advanced equations and graphs are the last things which would hold their interest, but there's no escape from the math. Physicists from the Institute of Nuclear Physics of the Polish Academy of Sciences (IFJ) in Cracow, Poland, performed a detailed statistical analysis of more than one hundred famous works of world literature, written in several languages and representing various literary genres. The books, tested for revealing correlations in variations of sentence length, proved to be governed by the dynamics of a cascade. This means that the construction of these books is in fact a fractal. In the case of several works their mathematical complexity proved to be exceptional, comparable to the structure of complex mathematical objects considered to be multifractal. Interestingly, in the analyzed pool of all the works, one genre turned out to be exceptionally multifractal in nature.

Fractals are self-similar mathematical objects: when we begin to expand one fragment or another, what eventually emerges is a structure that resembles the original object. Typical fractals, especially those widely known as the Sierpinski triangle and the Mandelbrot set, are monofractals, meaning that the pace of enlargement in any place of a fractal is the same, linear: if they at some point were rescaled x number of times to reveal a structure similar to the original, the same increase in another place would also reveal a similar structure.

Multifractals are more highly advanced mathematical structures: fractals of fractals. They arise from fractals 'interwoven' with each other in an appropriate manner and in appropriate proportions. Multifractals are not simply the sum of fractals and cannot be divided to return back to their original components, because the way they weave is fractal in nature. The result is that in order to see a structure similar to the original, different portions of a multifractal need to expand at different rates. A multifractal is therefore non-linear in nature.

"Analyses on multiple scales, carried out using fractals, allow us to neatly grasp information on correlations among data at various levels of complexity of tested systems. As a result, they point to the hierarchical organization of phenomena and structures found in nature. So we can expect natural language, which represents a major evolutionary leap of the natural world, to show such correlations as well. Their existence in literary works, however, had not yet been convincingly documented. Meanwhile, it turned out that when you look at these works from the proper perspective, these correlations appear to be not only common, but in some works they take on a particularly sophisticated mathematical complexity," says Prof. Stanislaw Drozdz (IFJ PAN, Cracow University of Technology).

The study involved 113 literary works written in English, French, German, Italian, Polish, Russian and Spanish by such famous figures as Honore de Balzac, Arthur Conan Doyle, Julio Cortazar, Charles Dickens, Fyodor Dostoevsky, Alexandre Dumas, Umberto Eco, George Elliot, Victor Hugo, James Joyce, Thomas Mann, Marcel Proust, Wladyslaw Reymont, William Shakespeare, Henryk Sienkiewicz, JRR Tolkien, Leo Tolstoy and Virginia Woolf, among others. The selected works were no less than 5,000 sentences long, in order to ensure statistical reliability.

To convert the texts to numerical sequences, sentence length was measured by the number of words (an alternative method of counting characters in the sentence turned out to have no major impact on the conclusions). The dependences were then searched for in the data – beginning with the simplest, i.e. linear. This is the posited question: if a sentence of a given length is x times longer than the sentences of different lengths, is the same aspect ratio preserved when looking at sentences respectively longer or shorter?

"All of the examined works showed self-similarity in terms of organization of the lengths of sentences. Some were more expressive – here The Ambassadors by Henry James stood out – while others to far less of an extreme, as in the case of the French seventeenth-century romance Artamene ou le Grand Cyrus. However, correlations were evident, and therefore these texts were the construction of a fractal," comments Dr. Pawel Oswiecimka (IFJ PAN), who also noted that fractality of a literary text will in practice never be as perfect as in the world of mathematics. It is possible to magnify mathematical fractals up to infinity, while the number of sentences in each book is finite, and at a certain stage of scaling there will always be a cut-off in the form of the end of the dataset.

Things took a particularly interesting turn when physicists from the IFJ PAN began tracking non-linear dependence, which in most of the studied works was present to a slight or moderate degree. However, more than a dozen works revealed a very clear multifractal structure, and almost all of these proved to be representative of one genre, that of stream of consciousness. The only exception was the Bible, specifically the Old Testament, which has so far never been associated with this literary genre.

"The absolute record in terms of multifractality turned out to be Finnegan's Wake by James Joyce. The results of our analysis of this text are virtually indistinguishable from ideal, purely mathematical multifractals," says Prof. Drozdz.

The most multifractal works also included A Heartbreaking Work of Staggering Genius by Dave Eggers, Rayuela by Julio Cortazar, The US Trilogy by John Dos Passos, The Waves by Virginia Woolf, 2666 by Roberto Bolano, and Joyce's Ulysses. At the same time a lot of works usually regarded as stream of consciousness turned out to show little correlation to multifractality, as it was hardly noticeable in books such as Atlas Shrugged by Ayn Rand and A la recherche du temps perdu by Marcel Proust.

"It is not entirely clear whether stream of consciousness writing actually reveals the deeper qualities of our consciousness, or rather the imagination of the writers. It is hardly surprising that ascribing a work to a particular genre is, for whatever reason, sometimes subjective. We see, moreover, the possibility of an interesting application of our methodology: it may someday help in a more objective assignment of books to one genre or another," notes Prof. Drozdz.

Multifractal analyses of literary texts carried out by the IFJ PAN have been published in Information Sciences, the prestigious journal of computer science. The publication has undergone rigorous verification: given the interdisciplinary nature of the subject, editors immediately appointed up to six reviewers.

The Henryk Niewodniczański Institute of Nuclear Physics (IFJ PAN) is currently the largest research institute of the Polish Academy of Sciences. The broad range of studies and activities of IFJ PAN includes basic and applied research, ranging from particle physics and astrophysics, through hadron physics, high-, medium-, and low-energy nuclear physics, condensed matter physics (including materials engineering), to various applications of methods of nuclear physics in interdisciplinary research, covering medical physics, dosimetry, radiation and environmental biology, environmental protection, and other related disciplines. The average yearly yield of the IFJ PAN encompasses more than 450 scientific papers in the Journal Citation Reports published by the Thomson Reuters. The part of the Institute is the Cyclotron Centre Bronowice (CCB) which is an infrastructure, unique in Central Europe, to serve as a clinical and research centre in the area of medical and nuclear physics. IFJ PAN is a member of the Marian Smoluchowski Kraków Research Consortium: "Matter-Energy-Future" which possesses the status of a Leading National Research Centre (KNOW) in physics for the years 2012-2017. The Institute is of A+ Category (leading level in Poland) in the field of sciences and engineering.

CONTACTS:

Prof. **Stanisław Drożdż**
The Institute of Nuclear Physics of the Polish Academy of Sciences
tel. +48 12 6628220
email: stanislaw.drozdz@ifj.edu.pl

SCIENTIFIC PAPERS:

"Quantifying origin and character of long-range correlations in narrative texts"; S. Drożdż, P. Oświęcimka, A. Kulig, J. Kwapień, K. Bazarnik, I. Grabska-Gradzińska, J. Rybicki, M. Stanuszek; Information Sciences, vol. 331, 32–44, 20 February 2016; DOI: 10.1016/j.ins.2015.10.023

LINKS:

http://www.ifj.edu.pl/
The website of the Institute of Nuclear Physics of the Polish Academy of Sciences.

http://press.ifj.edu.pl/en/
Press releases of the Institute of Nuclear Physics of the Polish Academy of Sciences.

IMAGES:

IFJ160121b_fot01s.jpg   **HR:** http://press.ifj.edu.pl/news/2016/01/IFJ160121b_fot01.jpg
Multifractality in literature. The numbers correspond to the position of a given literary work in the pool (as in the detailed list available in the scientific publication). The color red indicates works traditionally belonging to the genre of stream of consciousness. On the vertical axis is the width of the multifractal spectrum – in those closer to unity, the text has a more multifractal structure. The horizontal axis is the degree of persistence H, in this case in the length of the sentences. A higher value of H means that there is a greater probability of a long sentence immediately following a long one, and a short one following another short. H = 0.5 means the lack of such a tendency. (Source: IFJ PAN)

IFJ160121b_fot02s.jpg   **HR:** http://press.ifj.edu.pl/news/2016/01/IFJ160121b_fot02e.jpg
Multifractal analysis of Finnegan's Wake by James Joyce. The ideal shape of the graph is virtually indistinguishable from the results for purely mathematical multifractals. The horizontal axis represents the degree of singularity, and the vertical axis shows the spectrum of singularity. (Source: IFJ PAN)

IFJ160121b_fot02s.jpg   **HR:** http://press.ifj.edu.pl/news/2016/01/IFJ160121b_fot03e.jpg
Sequences of sentence lengths (as measured by number of words) in four literary works representative of various degrees of cascading character. (Source: IFJ PAN)