ACADEMIA | Focus on Literary Studies

Reading the Unwritten

Dr. Maciej Maryl, Head of the Digital Humanities Center and Deputy Director at the Institute of Literary Research, Polish Academy of Sciences, talks about the table where innovations are born, the limits of literariness, and Polish humanities projects that can conquer the world.

ACADEMIA: **What does a digital humanist need to do his work?**
MACIEJ MARYL: We usually work at a table.

Just like that?
Yes. Most people imagine humanist researchers sitting alone in a library, poring through books, and then typing their papers on a computer. Here we focus on teamwork because digital projects are usually very complex and require all sorts of expertise. Conversations at the table facilitate compromise between humanists, IT specialists and other people involved in a given project. Franco Moretti, founder of the Stanford Literary Lab, claims that the table is just as important as expensive equipment.

How does this look in practice?
On one side of the table sit the humanists who have a research idea, but they do not quite know how to implement it. On the other side sit IT specialists, people with programming skills and ambitions to do something interesting. And in the middle there is us, people with a humanities background and some computer skills, acting as intermediaries. A good example is the Scamander correspondence project. A leading editor, Prof. Beata Dorosz approached us with the idea of digitizing these letters. Experts from the New Panorama of Polish Literature under the guidance of Dr. Bartłomiej Szleszyński developed the technological concept, found appropriate IT specialists and are currently working on the project under a grant from the National Programme for the Development of the Humanities. Ultimately, the letters will not only be available on the Internet, but will also be digitized in accordance with the latest digital editing standards. Individual words will be encoded according to the international TEI (Text Encoding Initiative) standard, so that the texts can be read by machines. We will also encode metatextual information, such as deletions or replacements in the text, missing pages, etc. Texts prepared according to those standards could be further processed for research purposes. For example, one can easily extract all the place names from a text in order to map the geography of a given work. Generally, the idea is to on one hand digitize, annotate and publish the texts online, and on the other to enable their reuse in various research contexts.

Can your methods help identify the author of the text?
Here I should mention the internationally recognized and groundbreaking work of two researchers from Kraków, Professor Maciej Eder, Director of the PAS Institute of the Polish Language, and Dr. Jan Rybicki from the Jagiellonian University, who are working on developing the stylometric method and the "stylo" package used in the R environment. By far its most popular application is authorship attribution, which is based on the assumption that the author leaves a stylometric fingerprint in the text. And we are not talking here about unique or unusual terms, but rather about the simple frequencies of most common words used in the text. For a corpus of texts a matrix of those frequencies is created, which allows particular works to be attributed to individual authors. And if a certain text from a group of works considered to be by a specific author stands out and does not seem fit, this could be a sign that its authorship was wrongly attributed. You can also work with anonymous texts and sometimes it is possible to identify their author with a certain degree of probability.

Maciej Maryl, PhD
is a literary scholar, sociologist, translator and author of radio programs. A graduate of an individual track of study in the humanities at the University of Warsaw (combining Polish studies and applied social sciences), the Artes Liberales Academy, and the School of Social Sciences at the PAS Institute of Philosophy and Sociology. Recipient of scholarships from the Polish Ministry of Science and Higher Education, Foundation for Polish Science, and the Fulbright Commission. Member of the Editorial Board of the literary journal Teksty Drugie. His areas of interest include literary communication, new media, multimedia literature, electronic editing, digital humanities and the connections between technology and culture. At the DHC, he works in the research department with a focus on using literary tools to study new forms of electronic literature (such as blogs).

maciej.maryl@ibl.waw.pl

Not just words

Dissident movements in former communist bloc countries are being studied by nearly 100 researchers in 33 European countries, members of a research network established by the PAS Institute of Literary Research, the first Polish humanities research institution which received a grant from the COST Association (European Cooperation in Science & Technology).

The project "New Exploratory Phase in Research on the East European Cultures of Dissent" was submitted by the research team: Dr. Maciej Maryl (project chair), Dr. Nina Hoffman and Piotr Wciślik. The network includes representatives of leading research centers, cultural heritage institutions, non-governmental organizations, and private companies providing digital services for science. Researchers will study the cultural heritage of resistance and dissident movements in former communist bloc countries between 1945 and 1989. The proposers of this Action believe that, after a period of growth and consolidation, this field of study and the respective domain of cultural heritage have stalled and fallen short of its true significance. They believe that this state of affairs results from the inheritance of Cold War-era conceptual distinctions, the confinement of research within national silos and neglecting the problem of access to original archival sources for digitally enabled research. Thus the Action aims to overcome those shortcomings and to trigger the next discovery phase of this legacy.

Researchers intend to adopt a transnational and comparative perspective. At the same time, they want to focus on studying cultural phenomena, such as literature or film, but also science and education, rather than concentrating on political events. The scope of the project will cover the issues of both independent culture and state-controlled culture, as well as the middle ground between adaptation and resistance. Researchers will also take on the aforementioned problem of digital infrastructure. The project will employ digital tools and platforms for conducting research and disseminating the results. The Action will spark international and interdisciplinary collaboration between researchers and professionals working in archives, libraries, museums and galleries. The project will make it possible to cross the communication barriers between national research groups and various stakeholders.

What is literature from the point of view of a digital humanist? Does it include only literary works or everything that has ever been written?

The border between the literary and non-literary is conventional and basically depends on the research question one wants to address. The stylometric studies referred to earlier teeter on the border of linguistics and literary studies.

But they can be used analyze the correspondence between terrorism suspects, for example.

They can, but the results will be less certain, because these are mostly short texts, which makes analysis difficult. In Wrocław there was a joint project between the University and the Technical University, which focused on the somber task of analyzing suicide notes. The aim was to try differentiating between real and fake letters by using various language methods. Here we can imagine various practical applications.

So, psychology is also helpful in your work?

Yes. There are various studies in sentiment analysis, which measure the frequency of specific groups of words, which, for example, relate to positive or negative attitudes to certain phenomena. We can apply those approaches to analyze popular and difficult topics, such as attitudes towards immigrants.

What if we studied the correspondence between Polish poets Jarosław Iwaszkiewicz and Paweł Hertz, for example. Would this help us determine how their relationship developed?

To a certain degree. It is possible to analyze the change over the years in how the two men greeted each other and the types of words they used when referring to one another. We would have to maintain some level of doubt, however.

Stylometric methods fare much better where we have more information to work with. The more data we have, that is the larger the textual corpus, the better we can observe certain relationships, but we would certainly detect things that are invisible at first glance.

This type of research, which is still in its early stages, is intended to open new doors, but also to develop IT tools that can be used by researchers based on their needs. The PAS Institute of Literary Research, in collaboration with the Wrocław University of Technology, is working on creating a Literary Exploration Machine (http://lem.clarin-pl.eu/). It is a platform that facilitates the use of various text analysis tools developed by the CLARIN-PL consortium. Currently literary scholars can, for example, lemmatize texts online, that is, convert all words into a basic form, which facilitates their further processing with other analytical tools.

Ultimately, we will be publishing more descriptions and usage scenarios for literature researchers.

We live in a world of permanent surveillance and computers are threatening our privacy. Do you not worry that your methods might end up being used for such purposes?

Of course, everything could be used against citizens. But I'm not worried, maybe because I really cannot think of these methods being used in this way. Probably the best form of defense against such a scenario is having as much knowledge as possible about how such tools and mechanisms work.

As part of the Fulbright Senior Award scholarship, in 2018 you will be working at Stanford University on the project entitled "Literary transition as viewed from above: A quantitative history of Polish literature (1989–2000)". Is this a way of merging literary studies and social sciences?

The approach that I propose in this project arises from certain Polish research traditions, that had been developed at the PAS Institute of Literary Research, such as the Stefan Żółkiewski school of literary sociology. It involves looking at literature in an institutional context, in which questions are posed from the junction between sociology and literature. The main novelty is that today one can use materials collected over many years on a larger scale, such as bibliographies or full-text databases, which are now available in digital form.

You are currently researching the literary life in Poland after 1989. Will you be able to find out not only what happened to it, but also why?
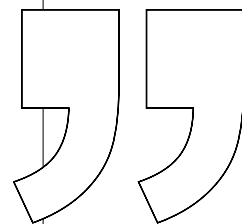
In this study I am dealing with what Polish literary critics and literary researchers of the first decade after the fall of communism define as a departure from political centralization towards the fragmentation of literary life and the subsequent return to centralization, dictated by the market. In my project I want to see whether on the basis of available digital data, especially literary bibliography, these processes can be recreated. If so, it means that we can use this data to search for other processes concerning different literary periods. As for finding out why something happened, we have to go back to the issue of results interpretation. We should meet in a year and discuss how it went. For now, I am sure that we will be able to see the beginnings of this process, such as define a group of publications that gradually gained strength and began to dictate market conditions, and then we can try to figure out what they had in common, why some actions were taken that year and not another, etc. The answers to some of these questions will have to be found outside of the textual environment. I believe that results obtained by any quantitative method in humanities must in the end be subject to interpretation, which often requires returning to the texts themselves in order to better understand the results.

We complain that Polish humanities have little chance of being recognized on a global scale and that our researchers do not write in English. On the other hand, we might wonder who cares about Polish literature. And here you are traveling to the United States on the Fulbright scholarship to work on this very topic.

I believe that Polish humanities have a very good chance to be recognized by the rest of the world, and

assuming that our work is of no interest to anyone is a mistake. Of course the interpretation of an unknown poem by a niche author of the Enlightenment era may not arouse great interest in foreign researchers who do not know anything about the topic. But placing this interpretation in the comparative context, or in relation to current academic discussions in the field, makes the work a potential object of international interest. We will obviously continue to publish in Polish too, because some studies may be only of local interest. However, there are many issues on which we have much to say, and which could serve as a valuable input to international scholarly debates. Comparative studies of literatures, literary transfers, or relations between old literatures can certainly be of interest to our foreign colleagues. For example, the PAS Institute of Literary Studies is currently the leader of a COST Action which has established a European research network for studies on East European cul-

> Every author leaves a stylometric fingerprint in the text. If a certain text does not match the others, that could be a sign that someone else wrote it.

tures of dissent. It is hard to imagine taking on such a topic in isolation. On the contrary, even if we are dealing with these phenomena in the Polish context, we are approaching them as a European question. At the same time, it serves as a great opportunity to present our substantial national achievements in this area to a wider audience.

As for my project, during my interview with the Fulbright Commission I had to defend my reasons for going to the United States to work on Polish literature. I explained that it is a methodological project, and that I will be developing various methods in a very advanced institution. I also want to see how a modern literary laboratory works in order to put this experience to use back here at home.

There is much work to be done then.

I certainly cannot complain about any lack of work.

<div align="right">
INTERVIEW BY ANNA ZAWADZKA
AND KATARZYNA CZARNECKA
PHOTOGRAPHY BY JAKUB OSTAŁOWSKI
</div>