

# ODCZYTAĆ NIENAPISANE

**O** stole, przy którym powstają nowości, granicach literackości i polskiej humanistyce, która może podbić świat, mówi **dr Maciej Maryl**, kierownik Centrum Humanistyki Cyfrowej, z-ca dyrektora ds. ogólnych Instytutu Badań Literackich PAN.

**ACADEMIA: Co jest potrzebne do pracy humaniście cyfrowemu?**

**MACIEJ MARYL:** My zazwyczaj siedzimy przy stole.

**Tak po prostu?**

Tak. Jest taki potoczny obraz badaczy humanistów, którzy siedzą samotnie w bibliotece i wertują książki, a następnie na komputerze piszą swoje teksty. U nas najważniejszym elementem jest praca zespołowa, bo projekty cyfrowe zwykle są bardzo złożone i wymagają wielu kompetencji. Efektem rozmów przy stole jest trudny kompromis między dążeniami humanistów, informatyków i innych osób zaangażowanych w dany projekt. Franco Moretti, założyciel Stanford Literary Lab, twierdzi, że stół jest równie istotny jak drogi sprzęt.

**Jak to wygląda w praktyce?**

Po jednej stronie stołu siedzą humaniści, którzy mają jakiś pomysł badawczy, ale nie do końca wiedzą, jak go zrealizować. Po drugiej – informatycy, ludzie z ambicjami, by wykonać coś ciekawego poznawczo, i umiejętnościami programowania. A pośrodku my – osoby z zapleczem humanistycznym, w miarę zorientowane w zagadnieniach komputerowych, pełniące funkcję pośredników. Dobrym przykładem jest historia projektu dotyczącego korespondencji Skamandrytów. Wybitna edytor, prof. Beata Dorosz, zgłosiła się do nas z pomysłem edycji cyfrowej tych listów. Zespół Nowej Panoramy Literatury Polskiej pod kierunkiem dr. Bartłomieja Szleszyńskiego sformułował koncepcję technologiczną, znalazł odpowiednich informatyków i zajmuje się obecnie realizacją w ramach grantu NPRH. Docelowo listy będą nie tylko dostępne w Internecie, lecz także zostaną opracowane zgodnie z najnowszymi standardami edytorstwa cyfrowego.

Poszczególnym słowom zostaną nadane określone znaczniki zgodne z międzynarodowym standardem TEI (Text Encoding Initiative) po to, by teksty były odczytywane przez maszyny. Zakodujemy również informacje metatekstowe, np. tu było skreślenie, tu jest zastąpienie tekstu, tu była urwana kartka itd. Poza tym opracowany tekst będzie mógł być przetwarzany przy innych zastosowaniach. Można choćby w bardzo łatwy sposób wyciągnąć wszystkie nazwy miejscowe z tekstów, np. żeby naszkicować geografę danego utworu. Generalnie więc chodzi o to, żeby z jednej strony opracować i udostępnić teksty, a z drugiej, żeby – umożliwić korzystanie z nich w różnych kontekstach badawczych.

**A czy stosowane przez was metody mogą pomóc np. w identyfikacji autora tekstu?**

Tu pionierami, z osiągnięciami na skalę światową, są dwaj badacze z Krakowa, profesor Maciej Eder, dyrektor Instytutu Języka PAN, i doktor Jan Rybicki z UJ, którzy pracują nad metodą stylometryczną i rozwijają pakiet „stylo” uruchamiany w środowisku R. Jego najbardziej medialnym i ciekawym pod względem wyników zastosowaniem jest tzw. atrybucja autorska. Jej podstawą jest założenie, że autor zostawia w tekście stylometryczny odcisk palca. I nie chodzi o różne dziwne lub nietypowe określenia, a raczej o różnice frekwencyjne używania najczęstszych słów. Na tej podstawie można te teksty przyporządkować do grup, które z założenia odpowiadają poszczególnym autorom. I jeśli jakiś tekst ze zbioru sygnowanego konkretnym nazwiskiem odstaje, nie pasuje, jest to sygnał, że coś jest nie w porządku z autorstwem. Można także pracować z dziełami anonimowymi – a nuż się okaże, że ktoś ma jednak autora dającego się zidentyfikować z określonym prawdopodobieństwem.



### **Dr Maciej Maryl**

jest literaturoznawcą, socjologiem, tłumaczem i autorem słuchowisk. Absolwent Kolegium MISH UW (polonistyka i stosowane nauki społeczne), Akademii Artes Liberales oraz Szkoły Nauk Społecznych przy IFiS PAN. Stypendysta Ministerstwa Nauki i Szkolnictwa Wyższego, Fundacji na rzecz Nauki Polskiej i Komisji Fulbrighta. Członek Kolegium Redakcyjnego „Tekstów Drugich”. Interesuje się komunikacją literacką, nowymi mediami, piśmiennictwem multimedialnym, edytorstwem elektronicznym, humanistyką cyfrową oraz związkami technologii z kulturą. W CHC zajmuje się sekcją badawczą, a w szczególności wykorzystaniem narzędzi literaturoznawczych do badania nowych form piśmiennictwa elektronicznego (np. blogów).

maciej.maryl@ibl.waw.pl

## Nie tylko słowa

Badaniem ruchów oporu w krajach byłego bloku socjalistycznego zajmie się prawie 100 uczestników z 33 krajów Europy skupionych w sieci badawczej, na której utworzenie IBL PAN – jako pierwsza polska humanistyczna jednostka naukowa – zdobył unijny grant COST (European Cooperation in Science & Technology).

Projekt „New Exploratory Phase in Research on East European Cultures of Dissent” przygotowali badacze dr Maciej Maryl (kierownik projektu), dr Nina Hoffman oraz mgr Piotr Wciślik. W sieci znajdują się przedstawiciele wiodących ośrodków badawczych, instytucji dziedzictwa kulturowego, organizacji pozarządowych oraz prywatnych firm zajmujących się dostarczaniem cyfrowych usług dla nauki.

Naukowcy prowadzić będą badania nad kulturowym wymiarem dziedzictwa oporu i ruchów dysydenckich w krajach byłego bloku socjalistycznego z lat 1945–1989. Jak wskazali twórcy projektu, badania w tym zakresie utknęły w martwym punkcie z powodu stosowania rozróżnień pojęciowych zakotwiczonych jeszcze w erze zimnej wojny, ograniczania badań do danego kraju oraz zaniedbywania problemu dostępu do oryginalnych źródeł archiwalnych (np. nierównych inwestycji w ich digitalizację oraz cyfrową infrastrukturę badawczą). Dlatego w projekcie zastosowane zostaną nowe podejścia badawcze, dzięki którym projekt oceniony został jako innowacyjny.

Badacze zamierzają przyjąć perspektywę ponadnarodową i porównawczą. Chcą przy tym skupić się na badaniu zjawisk kulturowych – np. literatury lub filmu, ale też nauki i oświaty – niż zjawisk stricte politycznych. Zakres tematyczny projektu obejmie kwestie zarówno kultury niezależnej, jak i państwowego nadzoru nad kulturą, a także sfery pośredniej, czyli zjawisk „pomiędzy przystosowaniem a oporem”. Uczni podejmą się też rozwiązania wspomnianego problemu infrastruktury cyfrowej. W ramach projektu powstaną narzędzia i platformy cyfrowe przeznaczone do prowadzenia badań, upowszechniania ich wyników oraz współpracy międzynarodowej i międzyśrodowiskowej dla badaczy oraz pracowników archiwów, bibliotek, muzeów i galerii. Projekt pozwoli na przekroczenie barier w komunikacji dzielących różne kraje i różne grupy interesariuszy. Badacze zaangażowani w projekt mają nadzieję, że zapoczątkuje on nową fazę odkryć w tym obszarze europejskiego dziedzictwa.

### **Czym jest literatura z punktu widzenia humanisty cyfrowego – to tylko dzieła literackie czy właściwie wszystko, co zostało napisane?**

Granica jest umowna i w zasadzie zależy od pytania badawczego. Przywołane wcześniej badania stylometryczne ewidentnie prowadzone są na styku językoznawstwa i literaturoznawstwa.

### **Ale można dzięki nim analizować np. korespondencję osób podejrzanych o terroryzm.**

Można, tylko oczywiście wyniki są mniej pewne, bo są to przeważnie teksty krótkie, co utrudnia analizy. We

Wrocławiu był także wspólny projekt Uniwersytetu i Politechniki, w którym pracowano na smutnym materiale – listach pożegnalnych samobójców. Zadanie polegało na tym, żeby spróbować różnymi metodami językowymi odróżnić list prawdziwy od fałszywego. Tu oczywiście można sobie wyobrazić różne praktyczne zastosowania.

### **Zatem przy państwa pracy pomocna jest także psychologia?**

Tak. Prowadzone są różne badania polegające na mierzeniu frekwencji określonych grup słów, które na przykład odnoszą się do pozytywnego lub negatywnego stosunku do danych zjawisk i możemy pod tym kątem analizować na dużych próbach gorące i trudne tematy, np. stosunek do migrantów.

### **A gdyby np. wziąć na warsztat korespondencję Jarosława Iwaszkiewicza z Pawłem Hertzem. Czy można by powiedzieć, jak rozwijała się ich relacja?**

Do pewnego stopnia. Można dokonać analizy tego, jak na przestrzeni lat zmieniały się powitania obu panów czy nacechowanie emocjonalne słów, których wobec siebie używali itd. Musielibyśmy zachować pewną nieufność. Metody stylometryczne dużo lepiej się sprawdzają w skali makro – im więcej mamy danych, tym lepiej widać pewne zależności – ale z pewnością wykrylibyśmy rzeczy na pierwszy rzut oka niewidoczne.

Tego typu badania, po części pionierskie, mają docelowo służyć przetarciu pewnych szlaków, ale też opracowaniu narzędzi informatycznych, które mogą być przez badaczy wykorzystane zgodnie z ich potrzebami. Instytut Badań Literackich PAN wspólnie z Politechniką Wrocławską buduje Literacki Eksplorator Maszynowy, w skrócie LEM (<http://ws.clarin-pl.eu/lem.shtml>). Jest to platforma, która ułatwia korzystanie z różnych narzędzi do analizy tekstu, wypracowanych przez konsorcjum CLARIN-PL. W tej chwili każdy literaturoznawca może na przykład online przetworzyć interesujące go teksty w formę zlematyzowaną, czyli zamienić wszystkie słowa na formę podstawową, co ułatwia później pracę z innymi narzędziami analitycznymi, bo odmiana nie zaburza frekwencji. Docelowo będziemy zamieszczać więcej opisów i scenariuszy użycia dla badaczy literatury.

### **Żyjemy w świecie permanentnej inwigilacji, korzystając z komputerów, wystawiamy swoją prywatność na ciężką próbę. Nie boi się pan zastosowania waszych metod do takich celów?**

Oczywiście wszystko można wykorzystać przeciwko obywatelom. Ale ja się nie boję, może dlatego, że nie wyobrażam sobie do końca tego typu zastosowań. Zapewne najlepszą formą obrony przed takim scenariuszem jest jak największa wiedza o działaniu takich narzędzi i mechanizmów.

**W ramach stypendium Fulbright Senior Award będzie pan w 2018 r. realizował na Uniwersytecie Stanforda projekt „Literary transition as viewed from above. Quantitative history of Polish literature (1989–2000)”. Czy jest on rodzajem przeniesienia literaturoznawstwa w dziedzinę nauk społecznych?**

Podejście, które proponuję w tym projekcie, wyrasta z pewnych polskich tradycji badawczych, zresztą bezpośrednio rozwijanych w IBL PAN, np. szkoły socjologii literatury Stefana Żółkiewskiego. To kierunek, który polega na przyglądaniu się funkcjonowaniu literatury w kontekście instytucjonalnym, w którym stawia się pytania ze styku socjologii i literatury. Jedyłą nowością jest to, że można na większą skalę wykorzystać materiały gromadzone od lat, jak bibliografie czy bazy pełnotekstowe, bo są teraz dostępne w formie cyfrowej.

**Pan bada rynek literacki w Polsce po 1989 r. Czy może pan nie tylko uzyskać odpowiedź na pytanie, co się z nim stało, lecz także dlaczego?**

Mamy tu do czynienia ze zjawiskiem, które przez polskich krytyków literackich i badaczy literatury pierwszej dekady po przełomie jest określone jako odejście od centralizacji politycznej w kierunku rozproszenia życia literackiego i następującego po tym powrotu do centralizacji, tym razem rynkowej. W swoim projekcie chcę zobaczyć, czy na podstawie dostępnych danych cyfrowych, zwłaszcza bibliografii literackiej, można te procesy odtworzyć. Jeżeli tak, to znaczy, że w tych danych możemy szukać innych procesów dotyczących także innych okresów literackich. A czy możemy dowiedzieć się, dlaczego coś się działo? Tu wracamy do kwestii interpretacji wyników, możemy umówić się za rok i porozmawiamy, co z tego wyszło. Na razie jestem pewien, że da się zobaczyć początki tego procesu, na przykład określić grupę wydawnictw, które się stopniowo umacniały i zaczęły dyktować warunki rynku i wtedy się zastanowić, co je łączyło, dlaczego jakieś działania podjęto w tym roku, a nie w innym itd. Część z tych pytań wymaga odniesienia do rzeczywistości pozatekstowej. Stoję tu na stanowisku, że wszystkie metody ilościowe w humanistyce i tak muszą się kończyć interpretacją uzyskanych wyników lub często powrotem do samych tekstów, by lepiej zrozumieć uzyskane rezultaty.

**Lamentujemy, że humanistyka polska ma niewielkie szanse zaistnienia w świecie, że nasi badacze nie piszą po angielsku. Z drugiej strony pojawia się pytanie, kogo obchodzi polska literatura. A pan jedzie do USA na stypendium Fulbrighta zajmować się właśnie nią.**

Uważam, że polska humanistyka absolutnie ma szanse zaistnienia w świecie, a zakładanie, że to, co robimy, nikogo nie interesuje, jest błędem. Oczywiście

interpretacja nieznanego wiersza niszowego twórcy oświeceniowego może nie wzbudzić wielkiego zainteresowania badaczy zagranicznych, którzy nie znają zagadnienia. Ale już umieszczenie tych samych uwag w kontekście porównawczym czy w odniesieniu do bieżących dyskusji naukowych w dyscyplinie czyni pracę potencjalnym obiektem zainteresowania międzynarodowej publiczności. Część naszych badań polonistycznych publikujemy wyłącznie w języku polskim, bo dotyczą kwestii ważnych wyłącznie lokalnie. Jest jednak cała grupa zagadnień, w których mamy wiele do pokazania i w sprawie których powinniśmy wchodzić w dialog z badaczami zagranicznymi. Choćby badania porównawcze literatur, transferów literackich czy relacji między dawnymi literaturami na pewno mogą interesować partnerów zagranicznych. Na przykład obecnie IBL PAN jest liderem grantu COST, w ramach którego organizujemy europejską sieć naukową do badań nad kulturą niezależną za żelazną kurtyną. Trudno sobie wyobrazić

Każdy twórca zostawia w tekście stylometryczny odcisk palca. Jeśli jakiś tekst nie pasuje do innych, jest to sygnał, że z autorstwem coś jest nie w porządku.

podejmowanie takiego tematu w izolacji. Przeciwnie, nawet jeśli zajmujemy się tymi zjawiskami na gruncie polskim, to traktujemy je jako temat europejski. Jest to jednocześnie świetna okazja do pokazania szerokiej publiczności pokąźnego dorobku krajowego w tej tematyce.

Jeżeli zaś chodzi o mój projekt, to na rozmowie w Komisji Fulbrighta musiałem obronić to, że jadę do USA pracować nad polską literaturą. Wykazałem, że projekt ma charakter metodologiczny, że będę rozwijał pewne metody w bardzo zaawansowanym ośrodku. Że chcę tam pojechać także po to, by zobaczyć, jak działa współczesne laboratorium literackie – wprowadzone tam formuły chciałbym zaszczerpić tutaj.

**Zatem czeka pana dużo pracy.**

To z pewnością jest rzecz, której mi nie brakuje.

Z DR. MACIEJEM MARYLEM  
 ROZMAWIAŁY ANNA ZAWADZKA  
 I KATARZYNA CZARNECKA  
 ZDJĘCIE JAKUB OSTAŁOWSKI