

Załącznik 2

Autoreferat

1. Jan Rybicki

2. Dyplom

Stopień doktora nauk humanistycznych uzyskany na podstawie pracy doktorskiej, pisany pod kierunkiem prof. dr hab. Jana Prokopa pt. *A Computer-Assisted Comparison of Two English Translations of Henryk Sienkiewicz's Trilogy*, Akademia Pedagogiczna w Krakowie, 2000.

3. Informacje o dotychczasowej pracy zawodowej.

- Tłumacz literatury anglojęzycznej, od 1987
- Asystent, Instytut Neofilologii, Wyższa Szkoła Pedagogiczna, Kraków, 1991-2000
- *Assistant Professor* (Stypendium Fundacji Kościuszkowskiej), Dept. of German and Slavic Studies, Rice University, Houston, USA, 1996-1997 i 2000-2001.
- Adiunkt, Instytut Neofilologii, Akademia Pedagogiczna/Uniwersytet Pedagogiczny, Kraków, 2001-2011.
- Adiunkt, Instytut Filologii Angielskiej, Uniwersytet Jagielloński, od 2011.

4. Osiągnięcie naukowe w rozumieniu art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm. w Dz. U. z 2016 r. poz. 1311):

a) Cykl powiązanych publikacji na temat: **Sygnaly stylometryczne w języku literackim w oryginale i przekładzie.**

b) Lista publikacji wchodzących w skład cyklu (w porządku chronologicznym):

- Prace metodologiczne:
 1. J. Rybicki, M. Eder, *Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?*, "Literary and Linguistic Computing" 2011, nr 26 (3), s. 315-332.

Mój udział w publikacji oceniam na 50% (przygotowanie tekstów elektronicznych, udział w analizie komputerowej, analiza wyników, redakcja tekstu).

2. M. Eder, J. Rybicki, *Do birds of a feather really flock together, or how to choose training samples for authorship attribution*, "Literary and Linguistic Computing" 2013, nr 28 (2), s. 229-236.

Mój udział w publikacji oceniam na 50% (przygotowanie tekstów elektronicznych, udział w analizie komputerowej, analiza wyników, redakcja tekstu).

3. J. Rybicki, *Success Rates in Most-frequent-word-based Authorship Attribution. A Case Study of 1000 Polish Novels from Ignacy Krasicki to Jerzy Pilch*, "Studies in Polish Linguistics," 2015, nr 10 (2), s. 87–104.
4. M. Eder, J. Rybicki, M. Kestemont, *Stylometry with R: A Package for Computational Text Analysis*, "R Journal" 2016 nr 8 (1), s. 107-121.

Mój udział w publikacji oceniam na 20% (przygotowanie tekstów elektronicznych, wykonanie analizy komputerowej, analiza wyników).

- Prace interpretacyjno-klasyfikujące:

5. J. Rybicki, *Ślady żony tłumacza. Alma Cardell Curtin i Jeremiah Curtin*, „Przekładaniec” 2011, nr 24, s. 90-110.
6. J. Rybicki, *The Great Mystery of the (Almost) Invisible Translator: Stylometry in Translation*. W: *Quantitative Methods in Corpus-Based Translation Studies*. Red. M. Oakley, M. Ji. John Benjamins, Amsterdam, 2012, s. 231-248.
7. J. Rybicki, *Tabakowska kontra Pawelec? Stylometria polskich przekładów twórczości N. Daviesa*. W: *Kompetencje tłumacza. Tom dedykowany prof. dr hab. Elżbiecie Tabakowskiej*. Red. M. Piotrowska, A. Czesak, A. Gomola, S. Tyupa. Tertium, Kraków, 2012, s. 315-336.
8. J. Rybicki, *Stylometryczna niewidzialność tłumacza*, "Przekładaniec" 2013, nr. 27, s. 61-87.
9. J. Rybicki, M. Heydel, *The Stylistics and Stylometry of Collaborative Translation: Woolf's 'Night and Day' in Polish*, "Literary and Linguistic Computing" 2013, nr 28 (4), s. 708-717.

Mój udział w publikacji oceniam na 60% (przygotowanie tekstów elektronicznych, wykonanie analizy komputerowej, analiza wyników, redakcja tekstu).

10. J. Rybicki, *Stylometric translator attribution: Do translators leave lexical traces?* W: *The Translator and the Computer*. Red. T. Piotrowski, Ł. Grabowski. Wydawnictwo Wyższej Szkoły Filologicznej, Wrocław, 2013, s. 193-204.
11. J. Rybicki, D. Hoover, M. Kestemont. *Collaborative authorship: Conrad, Ford and rolling delta*, "Literary and Linguistic Computing" 2014, nr 29 (3), s. 422-431.
Mój udział w publikacji oceniam na 45% (przygotowanie tekstów elektronicznych, wykonanie analizy komputerowej, analiza wyników, redakcja tekstu).
12. J. Rybicki, *Pierwszy rzut oka na stylometryczną mapę literatury polskiej*. „Teksty drugie” 2014, nr 2, s. 106-128.
13. J. Rybicki, *Vive la différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies*, "Digital Scholarship in the Humanities" 2016, nr 31 (4), s. 746-761.
14. J. Rybicki, *Drugi rzut oka na stylometryczną mapę literatury polskiej*. „Forum poetyki” 2017 nr 8, s. 6-21.
15. M. Choiński, J. Rybicki, *Is God Really Angry at Sinners? A Stylometric Study of Jonathan Edwards's Representations of God*. W: *The Global Edwards*. Red. R. S. Bezzant. Wipf and Stock, Eugene, 2017, s. 349-359.
Mój udział w publikacji oceniam na 50% (przygotowanie tekstów elektronicznych, wykonanie analizy komputerowej, analiza wyników, redakcja tekstu).
16. J. Rybicki, *Partners in Life, Partners in Crime? W: Drawing Elena Ferrante's Profile*. Red. A. Tuzzi, M. Cortelazzo, Padova University Press, Padova, 2018, s. 109-119.

c) omówienie celu naukowego ww. prac i osiągniętych wyników.

Badania sygnałów stylometrycznych w języku literackim w oryginale i przekładzie, które stanowią główny temat wymienionych powyżej publikacji, opierają się na kilku podstawowych założeniach. Po pierwsze – za pionierem tej dyscypliny i twórcą pierwszego

„prawa stylometrycznego”,¹ Wincentym Lutosławskim² – zakładam, że ilościowa analiza porównawcza cech stylu tekstów (w tym przypadku literackich) może nie tylko potwierdzać to, co wiemy o tych tekstach z tradycyjnej lektury i analizy jakościowej, ale i wskazywać nowe, dotąd niebadane aspekty, momenty i relacje. Po drugie, że badanie bardzo wielu i bardzo obszernych tekstów przez analizę statystyczną takich elementów językowych jak frekwencje słów czy n-gramów słownych uzasadnia stosowanie wobec języka modelu „*bag of words*”. Po trzecie – tym razem za Johnem Burrowsem – że szczególnie przydatna w takich badaniach jest analiza nie tylko słów „znaczących”, ale może przede wszystkim słów „gramatycznych”, „nieznaczących”, „synsemantycznych” – a to z uwagi na to, że „jedna trzecia, dwie piąte czy nawet połowa materiału” to „odpowiednio dwadzieścia, trzydzieści, pięćdziesiąt najczęstszych słów ze słownika autora”.³ Słowa te „realizują swe pełne znaczenie w związkach, które tworzą z mniej częstymi wyrazami”, więc „można uznać je za wystarczający opis tych związków i wszystkiego, co z tego wynika” dla formy i treści utworu.⁴ Często stosowana jest metafora „stylometrycznego odcisku palca”, którą po raz pierwszy posłużył się prawdopodobnie Anthony Kelly: linie papilarne nie mówią zbyt wiele o osobowości człowieka, a jednak skutecznie go identyfikują; idiosynkratyczne użycie najczęstszych słów może wcale nie przesądza o wszystkich cechach tekstu, ale też wystarczy, by rozpoznać np. dwa różne teksty tego samego autora.⁵

Słowa funkcyjne podobnie poważnie traktuje językoznawstwo kognitywne, więc nic dziwnego, że pojawiają się próby kognitywistycznego wytłumaczenia zjawisk obserwowanych doświadczalnie przez stylometrów.⁶ Dowodów na przydatność analizy ilościowej w badaniach nad tekstem dostarcza też psycholingwistyka, choćby za pośrednictwem Jamesa Pennebaker, według którego „zaimki, przedimki, przysłowki i różne inne, nie rzucające się w oczy słowa mogą powiedzieć wiele o naszej osobowości, sposobie

¹ W. Lutosławski, *Principes de Stylométrie appliqués a la Chronologie des Œuvres de Platon*, „Revue des Études Grecques”, 1898, t. 11, nr 41, s. 61-81.

² W. Lutosławski, *The origin and growth of Plato's logic: with an account of Plato's style and of the chronology of his writings*, Longmans & Green, London 1897.

³ J. Burrows, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*, Clarendon Press, Oxford 1987.

⁴ W. McKenna, J.F. Burrows, A. Antonia. *Beckett's Trilogy: Computational Stylistics and the Nature of Translation*. „Revue Informatique et Statistique dans les Sciences Humaines” 1999, nr 35(1-4), s. 171.

⁵ A. Kenny, *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*, Pergamon Press, Oxford 1982, 12.

⁶ L. Connors, *Computational stylistics, Cognitive Grammar, and the Tragedy of Mariam: combining formal and contextual approaches in a computational study of early modern tragedy*, rozprawa doktorska, University of Newcastle, Newcastle (NSW), 2013.

myślenia, stanach emocjonalnych i stosunku do innych ludzi”.⁷ W stylometrycznych badaniach literackich ujawniają się nie tylko sygnały autora, ale również tłumacza, redaktora, rodzaju i gatunku literackiego, płci (autora, narratora czy bohatera), ewolucji pisarskiej, tak na poziomie pojedynczego twórcy jak i w obrębie całych grup i tradycji literackich.

Model „*bag of words*” i analiza statystyczna sprawdzają się najlepiej w badaniach nad wielkimi zbiorami tekstów, bo wtedy w tej powodzi wyrwanych z kontekstu słów ich częstości stają się pseudo-losowe. Tak potężnego materiału nie potrafi już objąć tradycyjna, czyli jakościowa analiza literacka. Wbrew obiegowym opiniom czytanie na dystans Franco Morettiego⁸ radzi sobie z tym trochę inaczej, ale humanistyczne *big data* analizuje stylometria już od pewnego czasu, stosując metody dopiero niedawne nazwane przez Matthew Jockersa „makroanalizą”. Zasadniczym czynnikiem jest tutaj właśnie rozmiar tekstowego dziedzictwa ludzkości. Jak prowokacyjnie stwierdza autor *The Bestseller Code*:

W XXI wieku badacz literatury nie może już ograniczać się do anegdotycznych przesłanek – do tych wszystkich dość przypadkowo wybranych motywów nielicznych, choć może i "reprezentatywnych" tekstów. Nowych i ciekawych odczytań musimy szukać w znacznie szerszym kontekście - choćby w olbrzymiej masie znacznie "gorszych" tekstów literackich.⁹

Dlaczego? Dlatego, że o ile kanon literatury np. wiktoriańskiej to, powiedzmy, sto powieści kilkorga „reprezentacyjnych” autorów pokroju George Eliot czy Charlesa Dickensa – i z takim materiałem poradzi sobie tradycyjny historyk literatury – to epoka wiktoriańska pozostawiła po sobie ponad dwanaście tysięcy tytułów powieściowych. Lektura tyłu książek zabrałaby jednemu człowiekowi pięćdziesiąt lat (zakładając, że ten ktoś czytałby w tempie jednej powieści na dwa dni); jeżeli mamy zadawać jakiegokolwiek pytania takiej masie danych kulturowych, musimy je zadawać w sposób nietradycyjny, czyli właśnie ilościowy. Może będą to inne odpowiedzi i inne pytania, niż te, które daje i zadaje analiza jakościowa, ale przynajmniej – będą.

Zakres moich badań zawarty w prezentowanym dorobku rozpięty jest więc między mikroskopem językoznawczego badania leksykalnych elementów tekstów a teleskopem analizy sieciowej powiązań między tekstami literackimi; między zapewne pozornym

⁷ J. Pennebaker, *The Secret Life of Pronouns: What Our Words Say about Us*, Bloomsbury Press, New York, 2011, ix.

⁸ F. Moretti, *Wykresy, mapy, drzewa. Abstrakcyjne modele na potrzeby literatury*, tłum. T. Bilczewski i Anna Kowalczak-Pawlik, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, 2016.

⁹ M. Jockers, *Macroanalysis. Digital Methods & Literary History*, University of Illinois Press, Champaign 2013, 9.

obiektywizmem statystyki słów (jak pisze mój wielokrotny współautor, Maciej Eder)¹⁰ a całkowicie humanistycznymi interpretacjami tej statystyki; i wreszcie między poszukiwaniami optymalnych metod i narzędzi analizy tekstu a opartych na nich studiami nad wybranymi zestawami tekstów literackich.

Właśnie ten ostatni podział – na prace metodologiczne i prace analityczno-klasyfikacyjne – wydaje się najdogodniejszy do prezentacji mojego dorobku naukowego.

Do pierwszej, mniejszej grupy należą prace, których współautorami byli najczęściej Maciej Eder (obecny w trzech z nich) i Mike Kestemont (w jednej), oraz jedna praca mojego własnego autorstwa, choć oparta w znacznej mierze na metodzie wypracowanej przez Edera. Warto wspomnieć przy tej okazji, że jest to bardzo charakterystyczne zjawisko tak w stylometrii jak i w całej humanistyce cyfrowej, w których przewaga prac zbiorowych nad indywidualnymi rośnie z każdym rokiem czy to na konferencjach, czy w czasopismach naukowych tej transdyscypliny. Druga, większa grupa prac w prezentowanym cyklu to w przewadze moje indywidualne dokonania, choć i tu zdarzają się studia pisane wspólnie z partnerami polskimi (Magda Heydel, Michał Choiński – po jednym artykule) i zagranicznymi (David Hoover i Mike Kestemont – jedna wspólna praca).

Prace metodologiczne miały na celu poszerzenie wiedzy o metodzie wielowymiarowej analizy leksykalnej tekstu, wprowadzonej przez wspomnianego już Johna Burrowsa. Tzw. Delta Burrowsa opiera się na porównywaniu ciągów częstości najczęstszych słów występujących w danym zestawie tekstów. Porównanie to przeprowadza się we wszystkich parach tekstów, dla których wylicza się miarę odległości (czyli stopień różnicy) według klasycznego wzoru:

$$\Delta(T, T_1) = \frac{1}{n} \sum_{i=1}^n |z(f_i(T)) - z(f_i(T_1))|,$$

gdzie

$$z(f_x(T)) = \frac{f_x(T) - \mu_x}{\sigma_x},$$

oraz

¹⁰ M. Eder, *Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii*, „Teksty drugie” 2014, nr 2, s. 90-105.

$f_x(T)$ = częstość słowa x w danym tekście T ;

μ_x = średnia częstości słowa x w całym zestawie tekstów;

σ_x = odchylenie standardowe częstości słowa x w danym tekście T .¹¹

Procedura Delty i wzór zaproponowanej przez jej twórcę miary odległości zostały stworzone głównie z myślą o atrybucji autorskiej – i rzeczywiście okazuje się, że najmniejsze odległości występują zazwyczaj między tekstami tych samych autorów. Podobnemu zagadnieniu poświęcone były więc pierwsze trzy prace z prezentowanego cyklu: chodziło o sprawdzenie, czy powodzenie atrybucji autorskiej można zwiększyć – lub zmniejszyć – stosując różne parametry, różne zestawy tekstów lub badając zbiory tekstów w różnych językach.

I tak pierwsza z tych prac¹² sprawdzała wpływ na powodzenie atrybucji autorskiej 1) rozmiaru listy słów, na podstawie której wyliczano miarę odległości, 2) miejsca na liście frekwencyjnej pierwszego i następnych słów używanych w analizie, 3) rodzaju i gatunku literackiego oraz 4) języka tekstów. Obliczenia zostały dokonane za pomocą zestawu skryptów, pisanych dla środowiska oprogramowania statystycznego R (patrz poniżej), które wykonywały wszystkie etapy procedury obliczeniowej, systematycznie różnicując pierwsze dwa parametry dla zestawów tekstów trzech głównych rodzajów literackich (proza, poezja, dramat) w siedmiu językach: angielskim, polskim, francuskim, łacińskim, niemieckim, węgierskim i włoskim. I choć zestawy tekstów nie były zbyt obszerne – szczególnie według obecnych standardów „czytania na dystans”, była to prawdopodobnie pierwsza zakrojona na tak szeroką skalę seria doświadczeń z komputerową atrybucją autorską i pozostaje do dziś najczęściej cytowaną publikacją spółki Eder/Rybicki (67 cytowań do 1.08.2018). Wykazała, po pierwsze, że powodzenie atrybucji autorskiej może zależeć od języka badanych tekstów – przy czym jest ono zazwyczaj wyższe w językach mało- lub niefleksyjnych, takich jak angielski czy niemiecki; stosunkowo gorzej „zgadywały się” teksty polskie i łacińskie. Niefleksyjne języki wykazywały też znacznie wyższą tolerancję dla zmian długości list słów używanych w analizie i dla pobierania tych słów z dalszych miejsc rangowej listy frekwencyjnej. Warto wspomnieć, że w testowaniu oprogramowania znaczny udział mieli po

¹¹ J. Burrows, 'Delta' – A Measure of Stylistic Difference and a Guide to Likely Authorship, „Literary and Linguistic Computing” 2002, nr 17 (3), s. 267–287.

¹² J. Rybicki, M. Eder, *Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?*, „Literary and Linguistic Computing” 2011, nr 26 (3), s. 315-32.

raz pierwszy studenci prowadzonego przeze mnie seminarium magisterskiego, co potem stało się regułą w mojej działalności naukowej.

Wyniki powyższej pracy nie wyjaśniły oczywiście całości zagadnienia. Choć użyte w niej zestawy były porównywalne pod względem liczby utworów, trudno było utrzymać dokładnie te same proporcje tekstów i autorów. Kolejne parametry, który należało przebadac doświadczalnie, to wpływ na powodzenie atrybucji autorskiej liczby tekstów w korpusie, liczby tekstów tego samego autora i liczby samych autorów. Dlatego też kolejna praca wykonana przez ten sam zespół¹³ i ogłoszona w tym samym co poprzednia czasopiśmie z tzw. „listy A” polegała na wielokrotnym wykonywaniu analiz na tym samym korpusie przy zmiennej liczbie tekstów i autorów. Badania zostały wykonane na zestawach powieści w pięciu językach po 60-70 utworów. Okazało się, że najlepsze i najbardziej stabilne wyniki uzyskano dla korpusu angielskiego; z kolei korpus polski cechował się najgorszą zgadywalnością i w dodatku największym rozrzutem wyników powodzenia atrybucyjnego. Ponieważ doświadczenia przeprowadzono za pomocą uczenia nadzorowanego (oprogramowanie losowo wybierało „wzorcowy” tekst danego autora), rozrzut wyników skuteczności atrybucji wskazywał na konieczność właściwego doboru owych najbardziej reprezentatywnych utworów każdego z badanych pisarzy. Bardzo często zdarzało się, że nie były to wcale teksty powszechnie za takie uznane w tradycyjnej historii literatury. Nie udało się też wykluczyć, że odmienna skuteczność atrybucji w poszczególnych korpusach językowych jest efektem nie tylko językowym, ale historycznoliterackim: słaba skuteczność w języku polskim mogła być spowodowana przez fakt, że na korpus polski złożyły się głównie powieści polskiego pozytywizmu i współczesnych mu autorów, wychowanych na podobnej literaturze i w podobnej formacji społeczno-kulturowej. Praca była dotąd cytowana 32 razy.

Obie wyżej wymienione prace wskazywały na konieczność znacznego rozszerzenia zestawów badanych tekstów, by zminimalizować wpływ ewentualnego wyboru mało reprezentatywnych tekstów i ogólnie małej reprezentatywności badanego korpusu dla literatury w danym języku. Dlatego właśnie moje kolejne, tym razem jednoautorskie studium¹⁴ dotyczyło jednej co prawda literatury – polskiej – ale za to przebadanej na

¹³ M. Eder, J. Rybicki, *Do birds of a feather really flock together, or how to choose training samples for authorship attribution*, „Literary and Linguistic Computing” 2013, nr 28 (2), s. 229-236.

¹⁴ J. Rybicki, *Success Rates in Most-frequent-word-based Authorship Attribution. A Case Study of 1000 Polish Novels from Ignacy Krasicki to Jerzy Pilch*, „Studies in Polish Linguistics,” 2015, nr 10 (2), s. 87–104.

stosunkowo obszernym korpusie napisanych przez 250 autorów 1000 tekstów od *Mikołaja Doświadczyńskiego przypadków* po *Biegunów*, czyli na ponad dwustu latach twórczości literackiej po polsku. Dla zachowania mniej więcej równego podziału tekstów między autorów liczba utworów pisanych jednym piórem nie mogła przekroczyć 12; również epoki literackie były reprezentowane przez możliwie zbliżone liczby tekstów. Oprócz porównywania skuteczności atrybucji autorskiej uzyskanej dla różnej liczby autorów, różnej liczby tekstów i różnej liczby tekstów na jednego autora sprawdziłem, czy atrybucja autorska dokonana za pomocą najczęstszych słów jest rzeczywiście skuteczniejsza od tej dokonanej na najczęstszych n-gramach słownych.

Już na początku doświadczeń okazało się, że n-gramy są znacznie gorszym papierkiem lakmusowym sygnału autorskiego niż pojedyncze słowa; inaczej mówiąc, skuteczność atrybucji malała wraz ze wzrostem n. Zgodnie z najbardziej oczywistymi oczekiwaniami zmniejszanie liczby autorów ułatwiało atrybucję; przy okazji okazało się, że skuteczność atrybucji dla liczby autorów porównywalnej z tą, którą stosowano w poprzedniej z opisywanych tu prac dla tekstów polskich, była tym razem wyraźnie wyższa. Mogły mieć na to wpływ dwa czynniki: po pierwsze, tym razem korpus był znacznie bardziej zróżnicowany historycznie; po drugie, w oprogramowaniu używano ulepszonych algorytmów statystycznych. Mniej wyraźnie skuteczność atrybucji wzrastała wraz z liczbą tekstów na jednego autora.

Ostatnia z wymienionych prac metodologicznych¹⁵ ma nieco inny charakter. To podsumowanie prac nad pakietem oprogramowania dla środowiska oprogramowania stylometrycznego R o nazwie *stylo*, który od kilku lat jest jednym z najbardziej rozpoznawalnych narzędzi w tej dyscyplinie wiedzy, stosowanym przez badaczy na całym świecie. O jego popularności świadczy choćby liczba warsztatów, którą wszyscy trzej twórcy prowadzili już w ośrodkach naukowych na kilku kontynentach – i fakt opublikowania opisu narzędzia w czasopiśmie informatycznym o tak wysokim *impact factor*. Główna część oprogramowania i stałe jego rozwijanie to owoc pracy Macieja Edera, Mike Kestemont zaś dodał kilka interesujących funkcji. Mój własny, nieco skromniejszy udział w tym przedsięwzięciu – oczywiście poza ciągłym testowaniem oprogramowania – polegał na sporządzeniu interfejsu graficznego, dzięki któremu stosowanie *stylo* nie wymaga znajomości

¹⁵ M. Eder, J. Rybicki, M. Kestemont, *Stylometry with R: A Package for Computational Text Analysis*, “R Journal” 2016 nr 8 (1), s. 107-121.

dość skomplikowanych zasad funkcjonowania środowiska R i pozwala na prowadzenie analiz stylometrycznych nawet osobom nie posiadającym umiejętności programowania – których przecież nie brakuje wśród humanistów (również cyfrowych). Artykuł w „R Journal” prezentuje podstawowe zasady działania i obsługi programu i przykłady jego wykorzystania w najróżniejszych analizach języka naturalnego. Jest też dość często cytowany (51 razy do 1.08.2018), choć nie tak często, jak wcześniejszy (i nie wliczany do prezentowanego cyklu) plakat konferencyjny naszego zespołu (52 razy).¹⁶

Wśród prac drugiej grupy, które nazwałem interpretacyjno-klasyfikującymi, wyróżnia się seria publikacji poświęconych stylometrycznemu sygnałowi tłumacza. W najwcześniejszej z nich¹⁷ powraca jeden z głównych bohaterów mojego doktoratu, tłumacz Sienkiewicza (a także Orzeszkowej i Prusa), Jeremiaś Curtin, tym razem jednak reprezentowany nie tylko przez jego niesławne przekłady Trylogii, ale całość dorobku tłumaczeniowego i własne prace podróżniczo-etnograficzne i wspomnieniowe; w dodatku pojawia się też Alma Cardell Curtin, żona i sekretarka tłumacza. Publikacja ta jest również jedną z pierwszych, w której pakiet *stylo* umożliwił prowadzenie analiz opartych na tzw. uczeniu nienadzorowanym – co likwidowało niebezpieczeństwo zaburzenia wyników przez niefortunny wybór „wzorcowych” tekstów badanych autorów; zresztą w przypadku tłumacza trudno mówić o tekście „typowym”. Kolejną innowacją było znaczne zwiększenie wiarygodności wyników przez zastosowanie tzw. pseudo-bootstrapu konsensusowego. Do tej pory pakiet przeprowadzał osobne analizy skupień macierzy miar odległości między tekstami dla każdej długości listy słów; teraz wyniki ostateczne opierały się na konsensusie między jednostkowymi analizami skupień dla całego przedziału wartości długości listy referencyjnej.¹⁸

Wyniki analiz stylometrycznych przyniosły przede wszystkim potwierdzenie podejrzeń, że tak *Memoirs* Curtina, jak i znaczna część jego opisów podróży i studiów etnograficznych, wydanych po jego śmierci, wyszły spod pióra wdowy po nim, do końca własnego życia pielęgnującej pamięć po mężu. Co więcej, pewne wyniki analiz niektórych sienkiewiczowskich przekładów Jeremiaha mogą świadczyć o tym, że Alma nie ograniczała

¹⁶ M. Eder, M. Kestemont, J. Rybicki, *Stylometry with R: a suite of tools*, Digital Humanities 2013 Conference Abstracts, s. 487-489.

¹⁷ J. Rybicki, *Ślady żony tłumacza. Alma Cardell Curtin i Jeremiaś Curtin*, „Przekładaniec” 2011, nr 24, s. 90-110.

¹⁸ R. H. Baayen, *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*, Cambridge University Press, Cambridge, 2008.

się do wiernego przepisywania na czysto mężowskich brudnopisów i mimo nieznamości polszczyzny mogła mieć wpływ na ostateczny kształt angielskiej wersji *Za chlebem i Przez stępy*. Obie wersje (polska i angielska) tej publikacji były łącznie cytowane tylko 13 razy.

Prawdopodobnie najważniejszą a na pewno najczęściej cytowaną (42 razy) spośród prac o odcieniu translatorycznym jest pierwsza dokonana na taką skalę analiza widoczności wpływu tłumacza na stylometryczne cechy przekładu literackiego.¹⁹ Podsumowanie istniejącej literatury tematu dawało obraz bardzo fragmentaryczny – były to raczej studia pojedynczych przypadków.²⁰ W mojej publikacji poszukiwałem sygnału tłumacza w kilku zestawach tekstów w przekładzie. Zacząłem od siebie: okazało się, że moje przekłady (wówczas tylko sześciu) powieści Johna le Carré i (wówczas tylko trzech) Douglasa Couplanda łączą się w wyraźne skupiska autorskie, sugerując, że ich tłumacz umie wystarczająco różnicować styl(ometrię) tekstów różnych autorów – albo że jest to jakaś bardziej ogólna cecha stylometrii przekładu. Już następny eksperyment wskazał – niestety – na tę drugą hipotezę: teksty 65 powieści 11 autorów, przetłumaczone na polski przez różnych tłumaczy i z różnych języków, zostały rozdzielone na grupy autorskie ze stuprocentową skutecznością. Podobny efekt pojawił się przy analizie trzech różnych przekładów innej trylogii – Tolkiena – i tekstów innych autorów przełożonych przez polskich tłumaczy *Władcy pierścieni*. Angielskie przekłady powieści francuskich również gromadziły się głównie autorami oryginałów; francuskie przekłady książek angielskich zachowywały się podobnie, choć najstarsze, anonimowe skróty dwóch powieści Jane Austen trzymały się razem sfrancuszczeń Richardsona pióra Prevosta; co ciekawe, dodana do tego korpusu *Histoire du chevalier Des Grieux et de Manon Lescaut* umiejscowiła się w najbliższym sąsiedztwie dwóch przekładów francuskich powieści autora *Klarysy*.

Wyniki tego zestawu doświadczeń były dość zastanawiające. Oto bowiem okazało się, że sygnał autora jest znacznie silniejszy od sygnału tłumacza. Byłoby to bardziej zgodne z intuicją, gdyby efekt taki pojawił się w analizie częstości słów „znaczących”; tymczasem oryginały i przekłady układają się najczęściej podobnie – „po autorach” – mimo tego, że klasyfikacja dokonywana jest na podstawie słów najczęstszych i „funkcyjnych”, a więc takich, które nie posiadają jednoznacznego odwzorowania w listach frekwencyjnych dwóch

¹⁹ J. Rybicki *The Great Mystery of the (Almost) Invisible Translator: Stylometry in Translation*. W: *Quantitative Methods in Corpus-Based Translation Studies*. Red. M. Oakley, M. Ji. John Benjamins, Amsterdam, 2012, s. 231-248.

²⁰ Por. M. Baker, *Towards a methodology for investigating the style of a literary translator*. „Target” 2000, nr 12(2), s. 241-266.

różnych języków (choćby dlatego, że wśród stu najczęściej używanych słów w języku francuskim zawsze znajdują się wszystkie rodzajniki; w angielszczyźnie będą to tylko dwa przedimki, określony i nieokreślony, w polszczyźnie nie będzie ich w ogóle lub zostaną zastąpione przez różne wyrażenia zaimkowe). W ten sposób opisywane studium wykazało istnienie jeszcze jednego i dość niespodziewanego aspektu tego, co Venuti nazywa niewidzialnością tłumacza;²¹ można odtąd mówić o „niewidzialności stylometrycznej”.

To samo zjawisko było wkrótce potem tematem kolejnej publikacji, tym razem polskojęzycznej, więc bardziej skoncentrowanej na polskim przekładzie literackim.²²

Powtórzyłem analizę na korpusach powieści angielskich w polskich i francuskich przekładach i powieści francuskich w przekładach polskich i angielskich, zwracając większą uwagę na zrównoważenie i zwiększenie zawartości wszystkich tych zestawów; dochowanie pełnej symetrii i przeprowadzenie analogicznych doświadczeń z angielskimi i francuskimi przekładami literatury polskiej było niemożliwe – te dwie kategorie były nie dość dostępne w wersji elektronicznej, by sporządzić z nich obszerny korpus. Niewidzialność tłumacza i/lub widoczność sygnału autora była znów zjawiskiem dominującym, ale rozszerzenie analizy powieści angielskich i francuskich na przekłady polskie przyniosło kilka dodatkowych obserwacji. Zapewne najciekawszą z nich był fakt, że mimo wspomnianej dominacji sygnały autorskiego pojawili się jednak tłumacze o znacznej jednolitości stylometrycznej. Oto bowiem teksty Tadeusza Boya Żeleńskiego gromadziły się na wspólnej „gałązce” drzewka bootstrapu konsensusowego, niezależnie od tego, czy chodziło o *W stronę Swanna*, czy o *Gobsecka*, czy wreszcie o *Marysienkę Sobieską*. Wygląda na to, że przynajmniej pod względem stylometrycznym Boy albo tłumaczył „sobą”, albo pisał własne teksty „Proustem” lub „Balzakiem”.

Trzecia praca z tej serii była kolejną próbą określenia wzajemnych relacji między sygnałem autorskim a sygnałem tłumacza.²³ Wyciągnięcie wniosków z omawianej już powyżej własnej pracy²⁴ sprawiło, że doświadczenia zostały przeprowadzone na bardziej zrównoważonych zestawach tekstów – takich, w których każdemu oryginałowi odpowiadał dokładnie jeden

²¹ L. Venuti, *The Translator's Invisibility: A History of Translation*, Routledge, London, 1995, s. 8.

²² J. Rybicki, *Stylometryczna niewidzialność tłumacza*, „Przekładaniec” 2013, nr 27, s. 61-87.

²³ J. Rybicki, *Stylometric translator attribution: Do translators leave lexical traces? W: The Translator and the Computer*. Red. T. Piotrowski, Ł. Grabowski. Wydawnictwo Wyższej Szkoły Filologicznej, Wrocław, 2013, s. 193-204.

²⁴ M. Eder, J. Rybicki, *Do birds of a feather really flock together, or how to choose training samples for authorship attribution*, „Literary and Linguistic Computing” 2013, nr 28 (2), s. 229-236.

przekład. Nie zmieniło to jednak zasadniczo zjawisk obserwowanych w atrybucji – również i w tych bardziej rygorystycznych testach dominującym sygnałem było autorstwo oryginału. W korpusie 23 polskich przekładów powieści angielskich jedynym odstępstwem od tej reguły – jedynym przypadkiem, w którym przekłady łączyły się zgodnie z autorstwem przekładu – były trącające już nieco myszką tłumaczenia *Villette* Charlotte Brontë i *Posiadacza* Johna Galsworthy’ego, dokonane przez Różę Centnerszwerową. Trudno orzec, czy to właśnie nie nieco staroświecki język tych akurat spolszczeń nie stał się głównym czynnikiem sygnału tłumaczkici.

W tej samej publikacji rozszerzyłem również prowadzone już wcześniej badania nad własnymi przekładami²⁵ przez porównanie ich z efektem pracy kilku innych tłumaczy, którzy tłumaczyli tych samych co ja autorów. Chodziło o sprawdzenie, czy – przynajmniej przy tak ograniczonym zestawie tekstów – widać stylometryczny podział na literaturę oryginalną i tłumaczoną; i czy tłumaczenia autorów obcojęzycznych potrafią znaleźć własny sygnał wśród dzieł oryginalnych, wśród których znalazła się i polska klasyka, i polska powieść współczesna. Okazało się, że np. przekłady twórczości Johna le Carrégo (dokonane przez Radosława Januszewskiego, Jerzego Kozłowskiego, Jana Kraśkę, Jana Rybickiego, Roberta Sudóła i Macieja Świerkockiego) wykazują własną, odrębną stylometrię. Po raz pierwszy jednak objawił się też sygnał tłumacza: oto bowiem polskie wersje dzieł autora *Szpiega doskonałego* grupowały się najpierw zgodnie z autorstwem przekładu i dopiero „gałązki” osobnych tłumaczy łączyły się w większe skupisko le Carrégo. Jedynym źle przypisanym przekładem tegoż autora było *Z przejmującego zimna* Kraśki, które okazało się stylometrycznym sąsiadem dwóch kryminałów Joe Aleksa – tych pseudoprzekładów z angielskiego znanego polskiego tłumacza. Z kolei z twórczość bardzo współczesnego Douglasa Couplanda, reprezentowana w tym doświadczeniu dwie powieści i jeden zbiór esejów (wszystkie trzy pozycje w moim przekładzie), umieszczała się w najbliższym sąsiedztwie polskich współczesnych, kolejno: Masłowskiej, Gretkowskiej, Grocholi i Dehnela. Czy to znaczy, że mojego Couplanda tłumaczyłem wzorując się na pisarstwie kogoś z tej listy? Nie sędzę – wtedy jeszcze nie czytałem żadnej z ich książek. Trudno nie zauważyć, że w ten sposób poszukiwania miejsca sygnału stylometrycznego literatury tłumaczonej w ramach oryginalnej literatury w tym samym języku zaczęły interesująco

²⁵ Lista przekładów literackich mojego autorstwa znajduje się na końcu wykazu publikacji – osobnego załącznika do niniejszego wniosku.

zbliżać się do zagadnień, które w tradycyjnym przekładoznawstwie rozpatruje teoria polisystemów Itamara Even-Zohara.²⁶

Poszukiwania sygnału tłumacza kontynuowałem jednak również poza ścisłym kręgiem literatury pięknej. W pisanej niemal równolegle pracy o stylometrii polskich przekładów Normana Daviesa²⁷ sprawdzałem, czy można rozpoznać w nich sygnały polskich tłumaczy: Elżbiety Tabakowskiej, Andrzeja Pawelca oraz Joanny Rumińskiej-Pietrzyk i Bartłomieja Pietrzyka. Okazało się, że jest to możliwe tylko w przypadku Pawelca i częściowo w przypadku małżeństwa Pietrzyków, natomiast znacznie trudniej znaleźć wspólną stylometrię w przekładach i fragmentach przekładów angielskiego historyka, które wyszły spod pióra jego najbardziej znanej tłumaczki. Nie oznacza to jednak, że sygnały te są całkowicie niewidzialne. W dodatkowym eksperymencie, w którym za pomocą metody podobieństwa logarytmicznego²⁸ badałem słowa kluczowe Tabakowskiej i Pawelca, udało się odnaleźć kilka par alternatywnych wyrazów z grupy tych o najwyższej frekwencji. Tabakowska woli „mimo”, „aby”, „ale”, „czyli”, „przeciwko” i „ów”. W tych samych kontekstach Pawelec używa raczej „pomimo”, „by”, „lecz”, „więc” i „ten”. Tabakowska zdecydowanie preferuje „lub” – jej uczeń stosuje „lub” i „albo” w sposób dość równomierny.

Jednak moim największym zwycięstwem nad stylometryczną niewidzialnością tłumacza była napisana wspólnie z Magdą Heydel praca o wspólnym autorstwie polskiego przekładu *Night and Day* Virginii Woolf, opublikowana w czasopiśmie z „listy A” i jak dotąd dość często cytowana (31 razy). W tym jakże ciekawym przypadku mieliśmy do czynienia z tłumaczeniem rozpoczętym przez Annę Kołyszko, a po jej śmierci dokończonym i ujednoliconym stylistycznie przez moją Współautorkę.²⁹ Otrzymały od niej ostateczny tekst spolszczenia tej wczesnej powieści autorki *Pani Dalloway* został potraktowany w ten sposób, że jego pierwszy rozdział został potraktowany jako „wzorzec” stylometrycznych cech Kołyszko, zaś ostatni miał reprezentować stylometrię Heydel. Współautorka nie wyjawiała mi,

²⁶ I. Even-Zohar, *The Position of Translated Literature within the Literary Polysystem*. “Polysystem Studies” 1990 nr 11(1), s. 45-51.

²⁷ J. Rybicki, *Tabakowska kontra Pawelec? Stylometria polskich przekładów twórczości N. Daviesa*. W: *Kompetencje tłumacza. Tom dedykowany prof. dr hab. Elżbiecie Tabakowskiej*. Red. M. Piotrowska, A. Czesak, A. Gomola, S. Tyupa. Tertium, Kraków, 2012, s. 315-336.

²⁸ P. Rayson, D. Berridge, B. Francis, *Extending the Cochran rule for the comparison of word frequencies between corpora*. 7th International Conference on Statistical Analysis of Textual Data (JADT 2004), Louvain-la-Neuve 2004, s. 926-936.

²⁹ J. Rybicki, M. Heydel, *The Stylistics and Stylometry of Collaborative Translation: Woolf's 'Night and Day' in Polish*, “Literary and Linguistic Computing” 2013, nr 28 (4), s. 708-717.

w którym miejscu przejęła przekład po zmarłej koleżance. Konsensusowy pseudo-bootstrap ciągów najczęstszych słów całości dzieła podzielonego na równe próbki wskazał na początek rozdziału XXVI, co okazało się poprawną odpowiedzią na podstawowe pytanie pracy. Był to tym cenniejszy wynik, że Magda Heydel jest wybitną tłumaczką, która w dodatku bardzo sumiennie spełniła prośbę wydawcy o ujednoczenie stylu tego przekładu.

Nic dziwnego, że uzyskany wynik stał się doskonałą okazją, by w naszym artykule pokusić się o próbę nie tyle znalezienia, co poszukiwania miejsc wspólnych między „stylem literackim” w jego tradycyjnym rozumieniu a właśnie „stylometrią”. Bowiem ani tłumaczka, ani stylometra nie uważali stylometrii za bezpośrednie narzędzie do badania stylu w tradycyjnie literaturoznawczym znaczeniu. Chodziło nam raczej o wskazanie dalszych kierunków wspólnych badań, które mogłyby rzucić światło na wzajemne relacje dwóch bardzo różnych perspektyw na to, co dzieje się, gdy tłumacz tłumaczy dzieło literackie. Publikacja kończy się dość optymistycznym stwierdzeniem, że

...stylistyka języka przekładu literackiego powinna nauczyć się stosować odkrycia dokonane przez analizę stylometryczną nie tylko po to, by poszukiwać nowych odpowiedzi, lecz również by nauczyć się zadawać nowe pytania; z kolei stylometrii wyjdzie na dobre dokładniejsze zrozumienie mechanizmów, które sprawiają, że jest tak skutecznym narzędziem atrybucji autorskiej – i to nawet w sytuacji, gdy mogłoby się wydawać, że autor [oryginału] ukrywa się skutecznie za wszelkimi możliwymi deformacjami, które tłumaczenie zdaje się wprowadzać do listy najczęstszych słów. I o ile stylometria i stylistyka – mimo podobieństw w nazwie – jak dotąd nawołują się nawzajem z dwóch stron głębokiej przepaści, nie trzeba wiele, by był to „początek pięknej przyjaźni”.³⁰

Sukces atrybucji w przekładzie dokonany przez dwie różne osoby stał się pretekstem do wypróbowania skuteczności stylometrii na innych wspólnie pisanych utworach literackich – tym razem w oryginale.³¹ Trudno nie było pokusić się o zbadanie jednej z najsławniejszych literackich kooperacji, czyli o analizę autorstwa trzech powieści, napisanych wspólnie przez Josepha Conrada i Forda Madoxa Forda: *The Inheritors* (1901), *Romance* (1903) i *The Nature of a Crime* (1909, 1924). Dodatkowym powodem atrakcyjności tego przedsięwzięcia był fakt, że Mike Kestemont właśnie dodał do pakietu *stylo* nową funkcję, „rolling.delta”, umożliwiającą podział tekstu wspólnego autorstwa na próbki równej długości i zbadanie w

³⁰ Tamże, s. 716.

³¹ J. Rybicki, D. Hoover, M. Kestemont. *Collaborative authorship: Conrad, Ford and rolling delta*, „Literary and Linguistic Computing” 2014, nr 29 (3), s. 422-431.

każdej z nich siły stylometrycznych sygnałów współautorów wzorem wyprowadzonym z podstawowej formuły Deltę Burrowsa (patrz wyżej) na podstawie cech innych ich tekstów – ale tych pisanych samodzielnie. W ten sposób można wskazać na konkretne fragmenty, które posiadają silniejsze cechy stylometryczne któregoś ze współautorów. W przypadku Conrada i Forda materiału porównawczego było pod dostatkiem. Problemem utrudniającym weryfikację uzyskanych wyników był bardzo chwiejny konsensus co do rzeczywistego wkładu pracy obu pisarzy, przyjaźni między którymi nie przysłużyła się współpraca artystyczna; dość obszerna tradycyjna literatura przedmiotu również nie jest wolna od stronniczości.

Jeżeli można wierzyć wynikom tak wówczas nowatorskiej analizy, pierwsze wspólne dzieło Conrada i Forda, *The Inheritors*, wykazuje przewagę stylometrii fordowskiej nad conradowską. Autor *Jądra ciemności* dochodzi do głosu tylko w nielicznych fragmentach tej wczesnej powieści s-f. Jest to o tyle ciekawe, że w tym przypadku pierwsza wersja powieści jest dziełem Forda; Conrad natomiast dokonał na niej licznych i daleko idących przeróbek. Podobną sytuację zaobserwowano w przypadku *The Nature of a Crime*. Natomiast w *Romance* dominuje sygnał conradowski z wyjątkiem pojedynczego fragmentu o stylometrii Forda.

Z metodologicznego punktu widzenia – a więc z perspektywy przydatności nowej metody stylometrycznej zastosowanej w opisywanym studium – spore znaczenie miał dodatkowy eksperyment. Znane są twierdzenia Forda – wygłaszane zresztą już po śmierci Conrada – jakoby ten pierwszy miał być autorem znacznej części jednej z najważniejszych dzieł swojego niegdysiejszego współpracownika, *Nostromo*. To fakt, że biblioteka Uniwersytetu Princeton posiada brulion powieści z całą partią pisaną ręką autora *The Good Soldier*, ale większość badaczy uznaje wersję Conrada o dyktowaniu tego tekstu Fordowi. Analiza stylometryczna za pomocą „rolling.delta” pozwoliła usunąć wszelkie wątpliwości – *Nostromo* wykazuje czysty sygnał Conrada od pierwszej do ostatniej strony. Nasza praca była jak dotąd cytowana 24 razy – niestety, ani razu przez tradycyjnych specjalistów od Conrada czy Forda.

Stylometria jednak nie zajmuje się tylko sygnałem autorskim. Pozostałe prace wchodzące w skład prezentowanego osiągnięcia naukowego, choć wciąż stosują podobne metody, poszukują śladów innych zmiennych w skomplikowanym równaniu produkcji literackiej. Pierwsza z tych prac zajmuje się sygnałem kulturowej płci autora/autorki w literaturze i

została opublikowana w czasopiśmie z „listy A”.³² Można powiedzieć, że był to projekt zlecony przez grupę badaczek zrzeszonych w europejskiej sieci COST Action IS0901 “Women Writers in History: Toward a New Understanding of European Literary Culture” (2009–2013), którym przewodziła prof. Suzan van Dijk z holenderskiego Instytutu im. Huyghensa, specjalistka od pisarstwa kobiet w XVIII i XIX w. Chodziło o to, czy w grupie kilkudziesięciu powieści mało znanych pisarek angielskich z przełomu tychże wieków, zebranych w tzw. „korpuse z Chawton House”, metody stylometryczne potrafią wskazać – wśród kilku anonimowych powieści – te, które mogły zostać napisane przez mężczyznę. Drugim elementem zlecenia było sprawdzenie, czy powieści tych zapoznanych obecnie autorek różnią się stylometrycznie od dzieł ich bardziej „kanonicznych” rywalek, takich jak Jane Austen czy Mary Shelley.

Co dość oczywiste, pierwszym krokiem musiało być stworzenie korpusu porównawczego złożonego z tekstów kobiet i mężczyzn czynnych literacko w podobnej epoce. Pierwsze analizy, dokonane jak w poprzednio opisywanych badaniach na najczęstszym słownictwie, nie wykazały istnienia sygnału genderowego. W tej sytuacji należało wykorzystać inną metodę stylometryczną, która zamiast prostych częstości najczęstszych słów poszukuje wyrazów nie tyle najczęstszych, co występujących konsekwentnie i równomiernie we wszystkich tekstach jednej grupy utworów i równocześnie nie wykazujących takiego zachowania w tekstach drugiej grupy. W ten sposób otrzymałem dwie listy słów charakterystycznych dla – w tym przypadku – kobiet i dla mężczyzn. Ponieważ algorytm preferuje słowa z niższych pozycji na rangowej liście frekwencyjnej, przeważają w nich słowa „znaczące”. Metoda ta, tzw. „Zeta”, została opracowana i opisana przez Burrowsa³³ z późniejszymi modyfikacjami Craiga i Kinneya³⁴ oraz zaimplementowana do pakietu *stylo* jako funkcja „oppose”.

Do zbudowania „modelu słów ulubionych” pisarek i pisarzy w Anglii przełomu XVIII i XIX wieku wykorzystałem teksty autorów i autorek „kanonicznych”. Badanie wykazało bardzo wyraźną tendencję do konsekwentnego użycia przez kobiety słów związanych z rodziną i

³² J. Rybicki, *Vive la différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies*, “Digital Scholarship in the Humanities” 2016, nr 31 (4), s. 746-761.

³³ J. Burrows, *All the Way Through: Testing for Authorship in Different Frequency Strata*, “Literary and Linguistic Computing”, 2006 nr 22, s. 27-47.

³⁴ H. Craig, A. F. Kinney, *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge University Press, Cambridge, 2009.

uczuciami a z honorem i ogólniej z metaforyką rycerską u mężczyzn. Następnie wykonałem procedurę Delta (opisaną powyżej wielowymiarową analizę skupień) z tą modyfikacją, że zamiast listy słów najczęstszych wykorzystałem listę najbardziej „kobiecych” i najbardziej „męskich” słów uzyskanych za pomocą Zety. Metodą prób i błędów dokonałem takiego wyboru wśród słów „zetowych”, która najskuteczniej wykrywała kulturową płęć autorów (2-3 błędy na kilkadziesiąt tekstów). Po uzyskaniu optymalnego klasyfikatora zastosowałem go do zbadania w ten sam sposób anonimowych i znanych z autorstwa tekstów z Chawton House. Skuteczność atrybucji genderowej tekstów znanych autorek z tego korpusu była porównywalna z korpusem kontrolnym; za to spośród kilku tekstów anonimowych – stanowiących rzeczywisty materiał badawczy – tylko dwie powieści wykazywały cechy „męskich” preferencji leksykalnych. Szczególnie jedna z nich, *The Imposters detected: or, the Life of a Portuguese* [Oszuści zdemaskowani, czyli żywot pewnego Portugalczyka] wydaje się strzałem w dziesiątkę, bo spośród wszystkich powieści z badanego korpusu tekstów mniej znanych, wśród których przeważają tendencje sentymentalistyczne, wyróżnia się jako ostra satyra na katolicyzm.

Odpowiedź na drugie ze „zleconych” pytań otrzymałem stosując podobną procedurę, z tą różnicą, że zestaw kontrolny tekstów stanowiły teksty pisarek z korpusu Chawton House i współczesnych im sławnych mężczyzn, podczas gdy testowałem teksty sławnych czy też „kanonicznych” pisarek. Okazało się, że poszukiwany sygnał sławnych kobiet to przede wszystkim odejście od słownictwa ich mniej znanych, sentymentalistycznych autorek; można zaryzykować stwierdzenie, że do powszechnego kanonu literackiego trafiają te kobiety, które piszą trochę bardziej jak... mężczyźni.

Jednak kiedy te same słowa „zetowe” zostały użyte do przebadania większego zestawu tekstów, bo zawierającego ok. 1000 powieści od XVIII do XXI wieku, okazało się, że skuteczna atrybucja genderowa udaje się tylko w omawianych zestawie tekstów wczesnych. Co więcej, poszukiwanie podobnych słów dla powieści z przełomu wieków XIX i XX wyraźnie zmienia ich zakres pojęciowy, szczególnie u mężczyzn, gdzie zaczynają dominować pojęcia związane z wojną i pieniądzem. Z kolei analogiczna analiza dla powieści polskiej XIX i XX w. wykazała bardzo podobne preferencje genderowe do tych, którymi charakteryzowały się teksty angielskie z wieków XVIII i XIX, a w dodatku – dzięki większej fleksyjności polszczyzny – można było stwierdzić, że charakterystyczne czasowniki dla prozy kobiecej występują bez wyjątku w rodzaju żeńskim, podczas gdy w prozie męskiej wyraźnie dominuje rodzaj męski. Wszystkie te analizy wskazują na 1) zmienność cech

sygnału genderowego w czasie i 2) dominację perspektywy płci autora. Choć nie jest to wynik szczególnie zaskakujący, wyłaniający się z mojego studium stopień tej dominacji jest a) potwierdzony statystycznie i b) zaznaczony zapewne znacznie silniej, niż mogłaby sugerować tradycyjna lektura testu literackiego.

Warto zaznaczyć, że w opisywanym studium posłużyłem się nowym sposobem wizualizacji wyników stylometrycznych. Chodzi mianowicie o tak zwaną analizę sieciową, chętnie ostatnio stosowaną w naukach ścisłych, przyrodniczych, społecznych i wreszcie humanistycznych. Analiza ta z jednej strony ułatwia graficzne przedstawienie układu podobieństw i różnic między elementami danego zbioru danych, ale równocześnie umożliwia znacznie czytelniejsze rozmieszczenie na płaszczyźnie wyników analizy wielowymiarowej – szczególnie w przypadku bardzo obszernych zestawów tekstów.

I rzeczywiście ten sposób wizualizacji wyników sprawdził się przy próbach „czytania na dystans” obszerniejszej kolekcji literatury polskiej. W pierwszej publikacji poświęconej temu zagadnieniu wykorzystałem korpus około pół tysiąca polskich tekstów literackich (przeważnie utworów powieściowych).³⁵ Tytułowa „stylometryczna mapa”, znów oparta na porównaniu ciągów częstości najczęstszych słów, ułożyła się – jak można było się spodziewać – w porządku chronologicznym (od Krasickiego po Dukaję), ale wewnątrz tej makrostruktury panował przede wszystkim porządek autorski. Były i zrozumiałe wyjątki: twórczość Sienkiewicza rozpadła się na dwie, bardzo słabo powiązane części: na powieści przygodowo-historyczne i powieści obyczajowe – z tym, że te pierwsze gromadziły wokół siebie sporo wcześniejszych, współczesnych i późniejszych powieści o tematyce historycznej od Krasieńskiego przez Kraszewskiego i Deotymę po Bunscha. Tak silna odrębność użycia słów funkcyjnych w tekstach umieszczonych w dziejach dawnych jest najprawdopodobniej wynikiem słabszej lub silniejszej archaizacji.

Mistrzem stylizacji językowej – przynajmniej na podstawie moich badań ilościowych – okazał się Gombrowicz, którego *Trans-Atlantyk* wyraźnie ciąży ku Trylogii i *Krzyżakom*, i którego *Opętani* zbliżają się do *Znachora*. Można twierdzić, że z kolei skuteczność parodii udało mi się wykazać dla *Na ustach grzechu* Magdaleny Samozwaniec, której prześmiewcze *opus magnum* pojawia się bardzo blisko parodiowanych Mniszkówny i Grot-Bęczkowskiej. Osobiste związki między pisarzami też wydają się przechodzić do analizy stylometrycznej,

³⁵ J. Rybicki, „Pierwszy rzut oka na stylometryczną mapę literatury polskiej. „Teksty drugie” 2014, nr 2, s. 106-128.

jak to ma miejsce np. w trójkącie Kisielewski/Stalińska-Tyrmand-Iredyński. Powstała mapa literatury polskiej z jednej strony wielokrotnie zawiera powinowactwa między tekstami i pisarzami zgodnymi z wiedzą historyczno-literacką, ale równocześnie wskazuje ciekawe i mniej oczywiste powiązania. Wydaje mi się, że tego rodzaju analiza może wskazywać drogę dla nowych analiz literackich, łączących w sobie *close* z *distant reading*.

Rozszerzeniem artykułu z roku 2014 jest „drugi rzut oka” na literaturę polską, opublikowany trzy lata później.³⁶ Tym razem jest to spojrzenie znacznie obszerniejsze, bo analizujące dwa i pół tysiąca tekstów w języku polskim: ok. 1300 oryginałów i 1200 polskich przekładów z kilku języków, reprezentujących powieść i nowelę, dramaty i prozę epicką. Drugą zasadniczą różnicą było zastosowanie zmodyfikowanej miary odległości międzytekstowej, wciąż opartej na Delcie Burrowsa, ale w wersji opartej na kosinusie kąta zawartego między wektorami częstości słów par tekstów. Ta tzw. „Delta kosinusowa” ($\Delta\angle$), wprowadzona do stylometrii przez Smitha i Aldridge’a³⁷ dla dwóch tekstów T i T_1 mierzy kąt α (im większy, tym większa różnica między tymi tekstami):

$$\Delta\angle(T, T_1) = \alpha,$$

wyliczany na podstawie podobieństwa kosinusowego tzw. „standaryzacji Z ” dwóch wektorów $x = z(T)$ i $y = z(T_1)$:

$$\cos \alpha = \frac{\sum_{i=1}^{n_s} x_i y_i}{\sqrt{(\sum_{i=1}^{n_s} x_i^2)} \sqrt{(\sum_{i=1}^{n_s} y_i^2)}},$$

gdzie n_s to liczba badanych słów, a $z(T)$ to wartość standaryzacji Z częstości słów w tekście T , liczona zwykłym wzorem:

$$z(T) = \frac{f_s(T) - \mu_s}{\sigma_s},$$

gdzie z kolei $f_s(T)$ to częstość bezwzględna słowa s w tekście T , μ_s to średnia częstość słowa s w zbiorze tekstów, do którego należy tekst T , a σ_s to odchylenie standardowe częstości słowa s w tym samym zbiorze tekstów. Według co najmniej jednego, obszernego stadium ta wersja wzoru poprawia skuteczność atrybucji autorskiej nawet o 10%.³⁸

³⁶ J. Rybicki, *Drugi rzut oka na stylometryczną mapę literatury polskiej*. „Forum poetyki” 2017 nr 8, s. 6-21.

³⁷ P. W. H. Smith, W. Aldridge, *Improving Authorship Attribution: Optimizing Burrows' Delta Method*, „Journal of Quantitative Linguistics” 2011, nr 18 (1), s. 63-88.

³⁸ S. Evert, T. Proisl, F. Jannidis, I. Reger, S. Pielström, C. Schöch, T. Vitt, *Understanding and explaining Delta measures for authorship attribution*, “Digital Scholarship in the Humanities” 2017 supl. 2: 4-16.

Przeprowadzona następnie analiza sieciowa znów stworzyła chronologiczny, ewolucyjny porządek tekstów powieściowych i równocześnie wskazała na odrębność utworów dramatycznych, reprezentowanych w badanym korpusie przez obszerny zestaw polskich przekładów Szekspira i sztuki rdzennie polskie. Szczególnie interesującym zjawiskiem było silne podobieństwo stylometryczne polskich przekładów „Łabędzia z Awonu” z XVIII, XIX i XX wieku do polskiego dramatu romantycznego, który przecież tegoż Łabędzia umieścił na swych sztandarach.

Może jeszcze ciekawszym wynikiem było rozmieszczenie dwóch najliczniej reprezentowanych literatur w przekładzie, angielskiej i francuskiej, względem ewolucyjnego łańcucha literatury polskiej. Po pierwsze znaczna część tych przekładów odsuwała się na pewną odległość od oryginałów; po drugie, oba importowane „polisystemy” zbliżały się jednak do tekstów rdzennych pisanych w okresach najsilniejszego oddziaływania obu tych literatur: równoległe do literatury polskiej XIX wieku rozwijała się sieć tłumaczeń z francuskiego (nie tylko z tego samego wieku); przekładom z angielskiego było znacznie bliżej do polskich dzieł wieków następnych. Kiedy dodać do tego, że inna licznie reprezentowana literatura – rosyjska – znacznie mocniej wiązała się z polskimi oryginałami, można mówić o nowych, bo ilościowych dowodach na istnienie stylometrycznego *translationese* – charakterystycznego języka przekładu (w tym przypadku literackiego) którego odległość – czyli odmienność – od języka twórczości oryginalnej wzrasta wraz ze stopniem zróżnicowania języków – a może i kultur – porównywanych literatur.

Kolejna z publikacji z przedstawianego cyklu, choć nadal może być nazwana stylometryczną, wprowadza pewne nowe wartości.³⁹ Jest efektem kierowanego przeze mnie i niedawno zakończonego projektu grantowego NCN z serii Opus. Po pierwsze wprowadza metody stylometryczne w tematykę dotąd mało w ten sposób badaną: homiletykę purytańskich kaznodziejów czasów Wielkiego Przebudzenia w kolonialnej jeszcze Ameryce Północnej. Po drugie – obrawszy za swój temat centralne dla tego zjawiska kulturowo-religijne pojęcie Boga – używa metod ilościowych do zbadania, czy rzeczywiście Bóg amerykańskich protestantów był wtedy tak groźny, jak sądzi się powszechnie. W tym celu zbadano słowa będące kolokatami wyrazu „Bóg” w pismach głównego reprezentanta epoki, Jonathana Edwardsa. Okazało się, że wśród wspomnianych kolokatów dominują pojęcia pozytywne; w

³⁹ M. Choiński, J. Rybicki, *Is God Really Angry at Sinners? A Stylometric Study of Jonathan Edwards's Representations of God*. W: *The Global Edwards*. Red. R. S. Bezzant. Wipf and Stock, Eugene, 2017, s. 349-359.

analizie ilościowej Bóg Edwardsa jest dobry i litościwy. Natomiast zbadanie chronologicznych zmian w stosowaniu przez kaznodzieję boskich kolokatów ujawniło wyraźną ewolucję od prostych zabiegów retorycznych po podejście bardziej filozoficzne.

W najnowszym z prezentowanych tekstów powracam do najczęstszego zastosowania stylometrii, czyli stylometrii autorskiej, ale bazuję w tej pracy również na opisanych powyżej doświadczeniach z sygnałem tłumacza.⁴⁰ Chodzi o próbę rozwiązania jednej z najbardziej sensacyjnych zagadek literackich (przynajmniej w współczesnej literaturze włoskiej): o to, kto ukrywa się za pseudonimem „Elena Ferrante” – tłumaczka literatury niemieckiej Anita Raja (na którą wskazują ślady finansowe), jej mąż, Domenico Starnone, czy jeszcze kto inny.

Analiza sieciowa wyników analizy skupień ciągów frekwencji najczęstszych słów w dużych i całkiem małych korpusach literatury włoskiej wskazują jednoznacznie, że sygnał stylometryczny Ferrante jest identyczny z sygnałem Starnone. Szczególnie ważne jest to, że do tych samych wniosków doszło kilku innych stylometrów, których prace – stosujące różne metody – zostały wydane w tym samym tomie. Pozostają drobne wątpliwości: Anita Raja była reprezentowana w analizie wyłącznie przez swoje przekłady z niemieckiego (bo innych jej tekstów nie ma), trudno więc wykluczyć, że w rzeczywistości był to raczej włoski sygnał Christy Wolf (bo w przekładach tej pisarki specjalizuje się żona Domenico Starnone). Poza tym trudno wykluczyć jakiś rodzaj współpracy między małżonkami – w takich „spółkach” sygnał bywa zdominowany przez tego z współautorów, który jest autorem prac samodzielnych (*vide* wspomniany w pracy *casus* polskiej „Szymiczkowej” Dehnela i Tarczyńskiego). W każdym razie ta najnowsza z moich publikacji rzuca ciekawe światło na kolaboracje autorskie.

Prezentowana kolekcja publikacji rządzi się pewną wspólną logiką. Opierając się na podstawowych założeniach modelu *bag of words* i statystycznej analizy tekstów literackich poszukuje różnych sygnałów, które składają się na całościowy wydźwięk poszczególnych dzieł. Widzi siłę sygnału autora, dominującą w literaturze wszystkich epok – pomimo modnych doniesień o jego śmierci. Poszukuje jednak innych sygnałów: tłumacza, współautora, chronologii, rodzaju i gatunku, płci. W miarę pojawiania się nowych algorytmów i nowych metod wizualizacji wyników powoli buduje leksykalno-

⁴⁰ J. Rybicki, *Partners in Life, Partners in Crime? W: Drawing Elena Ferrante's Profile*. Red. A. Tuzzi, M. Cortelazzo, Padova University Press, Padova, 2018, s. 109-119.

językoznawczy obraz dzieła literackiego. I wreszcie - ujawnia potencjał badań ilościowych tekstu literackiego w oryginale i przekładzie jako interesujące uzupełnienie – i w pewnym stopniu wyzwanie – dla tradycyjnej analizy jakościowej.

5. Omówienie pozostałych osiągnięć naukowo-badawczych i artystycznych.

a) inne publikacje

1. Stylometria (prace przeglądowe i studia przypadków)

W tej kategorii znajdują się prace, które były kontynuacją stylometryczno-sienkiewiczowskich wątków mojego doktoratu, stanowiły mniej ważne analizy szczególnych przypadków stylometrycznych lub – czynione głównie z myślą o mniej stylometrycznie zorientowanym czytelniku – ogólne wprowadzenia do tej dziedziny wiedzy. I tak warto wspomnieć o analizie prostych danych liczbowych Trylogii Sienkiewicza: długości rozdziałów i zdań, rozkładach słów kluczowych, rozmiarów partii dialogowych poszczególnych postaci i bogactwa słownictwa.⁴¹ Pokłosem rozprawy doktorskiej jest też jedna z moich najczęściej cytowanych prac (36 razy), opisująca zagadkowe zbieżności w analizach wielowymiarowych częstości najczęstszych słów w oryginale i dwóch angielskich przekładach sienkiewiczowskiej serii.⁴² Muszę zaznaczyć, że nieco później wykryłem, że mimo zastosowania sprawdzonej – wydawałoby się – metody Burrowsa zbieżności te były w pewnym stopniu – choć nie wyłącznie – efektem różnic w rozmiarach zastosowanych próbek, o czym doniosłem w referacie konferencyjnym.⁴³ Nowy ten trop rozpoczął się zresztą w opublikowanym w tym samym roku badaniu częstości słownictwa w partiach postaci trzech angielskich wersjach *Hamleta* i dziewięciu przekładach tego dzieła na różne języki.⁴⁴ Wątek szekspirolologiczny pojawił się u mnie również w dwóch analizach innych sztuk

⁴¹ J. Rybicki, *Policzmy Trylogię Sienkiewicza*, „Annales Universitatis Paedagogicae Cracoviensis. Studia Historicolitteraria” 2010, nr 10, s. 85-110.

⁴² J. Rybicki, *Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz's Trilogy and its Two English Translations*, „Literary and Linguistic Computing” 2006, nr 21, s. 91-103.

⁴³ J. Rybicki, *Does Size Matter? A Re-evaluation of a Time-proven Method*. Digital Humanities 2008, Oulun Yliopisto, Oulu, 2008

⁴⁴ J. Rybicki, *Stylistics in Translation: Differences between Character Idiolects in Original and Translated Plays by Shakespeare*. W: *New Trends in English Teacher Education (Linguistics, Literature and Culture)*, red. I. Ramos Gay, Ediciones de la Universidad de Castilla-La Mancha, Cuenca 2008, s. 277-290.

„awońskiego łabędzia” – w oryginale i przekładzie – dokonanych w dużej i międzynarodowej grupie współautorów.^{45,46}

Innym ciekawym przykładem interdyscyplinarnej współpracy naukowej w dużej grupie jest fraktalna analiza rozkładu częstości zdań w obszernym korpusie literackim, opublikowana w czasopiśmie o bardzo wysokim *impact factor*.⁴⁷ Grupa fizyków z krakowskiego Instytutu Fizyki Jądrowej PAN, kierowana przez prof. dr hab. Stanisława Drożdża, zaprosiła do współpracy dwoje anglistów (dr hab. Katarzynę Bazarnik i mnie) w badaniu, z którego wynikało, że rozkład ten jest monofrakalny w przypadku tradycyjnych powieści a multifrakalny w dziełach pisanych w technice strumienia świadomości.

Strumień świadomości był ważnym elementem kolejnej zbiorowej publikacji, tym razem poszukującej stylometrycznych śladów wielkiego wpływu, jaki na twórczość irlandzkiego pisarza Flanna O’Briena wywarł jego wielki rodak poprzednik i guru, James Joyce.⁴⁸

Na tym nie koniec udziału w wieloautorskich publikacjach, bo pomagałem też młodemu, belgijsko-mołdawsko-niemiecko-polsko-rumuńsko-włoskiemu zespołowi w próbach ustalenia, czy metody stylometryczne – zwykle wymagające dużych tekstów – sprawdzą się w atrybucji obszernej, lecz złożonej na ogół z krótkich listów korespondencji między braćmi Grimm.⁴⁹ Jednym z ciekawszych wniosków z tej publikacji było stwierdzenie, że jakość takiej atrybucji nie zmienia się w żaden znaczący sposób w zależności od jakości badanej wersji elektronicznej.

Prace stylometryczne będące analizami szczególnych przypadków stylometrycznych to głównie prace pisane w ramach projektów grantowych, w których uczestniczyłem jako kierownik lub główny wykonawca. I tak wspomniany już powyżej projekt „purytański” zaowocował też przyjętą już do druku pracą o stylometrycznych śladach Thomasa Foxcrofta

⁴⁵ T. Cheesman, K. Flanagan, S. Thiel, J. Rybicki, R. Laramée, J. Hope, A. Roos, *Multi-Retranslation Corpora: Visibility, Variation, Value and Virtue*, “Digital Scholarship in the Humanities” 2017, nr 32 (4), s. 739-760.

⁴⁶ T. Cheesman, K. Flanagan, J. Rybicki, S. Thiel, *Six Maps of Translations of Shakespeare*. W: *Un/Translatable: New Maps for Germanic Literatures*, red. B. Wiggin, C. MacLeod, Northwestern University Press, Evanston 2016, s. 253-278.

⁴⁷ S. Drożdż, P. Oświecimka, A. Kulig, L. Kwapien, K. Bazarnik, I. Grabska-Gradzińska, J. Rybicki, M. Stanuszek, *Quantifying origin and character of long-range correlations in narrative texts*, “Information Sciences” 2016 nr 331, s. 32-44.

⁴⁸ J. O’Sullivan, K. Bazarnik, M. Eder, J. Rybicki, *Measuring Joycean Influences on Flann O’Brien*, „Digital Studies/Le champ numérique” 2018 (8/1), 6, doi: 10.16995/dscn.288.

⁴⁹ G. Franzini, M. Kestemont, G. Rotari, M. Jander, J.K. Ochab, E. Franzini, J. Byszuk, J. Rybicki, *Attributing authorship in the noisy digitized correspondence of Jacob and Wilhelm Grimm*, „Frontiers in Digital Humanities” 2018 nr 5 (4), doi: 10.3389/fdigh.2018.00004.

w edytowanych przez niego tekstach kazań Jonathana Edwardsa.⁵⁰ W ramach innego projektu – kierowanego przez dr Agatę Hołobut grantu z serii Opus (NCN) “Gatunek filmowy a strategię w tłumaczeniu audiowizualnym na przykładzie filmu historycznego” – badaliśmy zjawiska stylometryczne w adaptacjach filmowych literatury⁵¹ oraz siłę sygnału autorskiego, chronologicznego i gatunkowego w listach dialogowych seriali i filmów historycznych.⁵² W głównym *deliverable* tegoż projektu, monografii Agaty Hołobut i Moniki Woźniak, byłem wykonawcą części stylometrycznej pracy oraz autorem posłowania, w którym przybliżałem czytelnikowi filmo- i przekładoznawczemu podstawowe założenia i metody analizy ilościowej.⁵³

Pozostałe prace w tej grupie to inne ogólne wprowadzenia do stylometrii, z których niektóre są teraz często polecane studentom kierunków nauczania związanych z szeroko pojętą humanistyką cyfrową w wielu ośrodkach światowych.⁵⁴ Podobną publikacją – choć skierowaną do statystyków – jest tekst o metodach statystycznych w literaturoznawstwie, napisany na zaproszenie do pracy zbiorowej ku czci ś.p. prof. Jamesa Thompsona.⁵⁵

2. Zagadnienia przekładoznawcze i komparatystyczne

W tej grupie znalazły się przede wszystkim prace kontynuujące wątek szeroko rozumianej recepcji twórczości Sienkiewicza w krajach kultury anglosaskiej. Są to i ogólne prezentacje

⁵⁰ M. Choiński, J. Rybicki, *Jonathan Edwards and Thomas Foxcroft: Pursuing Stylometric Traces of the Editor*, „Amerikastudien” 2018 (w druku).

⁵¹ A. Hołobut, J. Rybicki, *Pride and Prejudice and Programming*. W: *Value in Adaptation*, red. L. Raw, McFarland, Jefferson 2018 (w druku).

⁵² Hołobut, A., Woźniak, M., J. Rybicki, *Old Questions, New Answers: Computational Stylistics in Audiovisual Translation Research*. W: *Audiovisual translation: research and use*, red. M. Deckert, Peter Lang, Frankfurt a/Main 2017, s. 203-216.

⁵³ J. Rybicki, *Posłowie: metodologia stylometryczna*. W: A. Hołobut, M. Woźniak, *Historia na ekranie. Gatunek filmowy a przekład audiowizualny*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2017, s. 405-416.

⁵⁴ J. Rybicki, *Authorship Attribution and Associated Analyses: Counting Characteristics With Computers*. W: *Current Issues in English Studies*, red. M. Misztal, M. Trawiński. Wydawnictwo Naukowe Uniwersytetu Pedagogicznego, Kraków 2009, s. 401-410; J. Rybicki, *Stylometria komputerowa w służbie tłumacza. Na przykładzie własnych przekładów*, „Rocznik przekładoznawczy” 2007-2008, nr 3-4, s. 169-179 ; J. Rybicki, *Signals in Stylometry: What Numbers (May) Tell Us about Literature*, “Atti dell’Accademia Polacca” 2014-2015, nr 4, s. 232-247; M. Eder, J. Rybicki, M. Kestemont, *Measuring Style W: Text Technologies: Computation for Literary Analysis*, red. J. O’Sullivan, Penn State University Press, University Park, 2016 (w druku); J. Rybicki, M. Eder, D. Hoover, *Computational Stylistics and Text Analysis*. W: *Doing Digital Humanities. Practice, Training, Research*, red. C. Crompton, R. J. Lane, R. Siemens, Routledge, Oxford 2016, s. 123-144.

⁵⁵ J. Rybicki, *Reading Novels with Statistics: What Numbers of Words Tell Us about Authorship, Genre, or Chronology*. W: *Models and Reality: Festschrift For James Robert Thompson*, red. J. A. Dobelman. T&NO Company, Chicago 2017, s. 207-224.

stanu wiedzy o tej recepcji w Stanach Zjednoczonych,⁵⁶ i artykuły, które oprócz tego przynoszą nowe informacje o angielskojęzycznych tłumaczach autora Trylogii.⁵⁷ Nowe informacje pojawiają się też w tekście o angielskich przekładach jedynej sienkiewiczowskiej powieści dla młodzieży.⁵⁸ W artykule wydanym w tym samym tomie umieszczam Sienkiewicza w szerszej perspektywie (kolonialnej) literatury angielskiej kontrastując afrykańskie przygody i autora *W pustyni i w puszczy*, i autora *Jądra ciemności*, które miały miejsce w ciągu jednego roku, i omawiając wpływ, jaki wywarły na twórczości obu urodzonych w Polsce pisarzy.⁵⁹ Ten sam tom zawiera zresztą moje komparatystyczne studium postaci chłopięcych bohaterów Sienkiewicza (Staś Tarkowski) i Szklarskiego (Tomek Wilmowski).⁶⁰ Jeszcze szerszą perspektywę komparatystyczną, bo nie tylko literacką, ale i filmoznawczą, przedstawia mój tekst o obrazie Hiszpanii w *Rękopisie znalezionym w Saragossie* Potockiego i w filmowej adaptacji Hasa.⁶¹ W ostatnim tekście z tej grupy przedstawiam sylwetkę Uriego Orleva, urodzonego w Warszawie izraelskiego pisarza i tłumacza literatury polskiej.⁶²

3. Literaturoznawstwo angielskie

Znacznie skromniej przejawia się mój dorobek literaturoznawczy, w którym nie ma mowy ani o stylometrii, ani o przekładzie. Na dobrą sprawę można tu zaliczyć tylko obszerny zarys legend arturiańskich⁶³ i analizę pojęć piękna i zła w poezji Gerarda Manleya Hopkinsa.⁶⁴

⁵⁶ J. Rybicki, *Trylogia po amerykańsku*, „Konspekt” 2003, nr 14/15, s. 36-41.

⁵⁷ J. Rybicki, *Sienkiewicz po angielsku*, „Przekładaniec” 2006, nr 15, s. 101-126.

⁵⁸ J. Rybicki, *Angielskie przekłady 'W pustyni i w puszczy'*. W: *Wokół „W pustyni i w puszczy”*, red. T. Bujnicki, J. Axer. Universitas, Kraków 2012, s. 555-570.

⁵⁹ J. Rybicki, *Conrad i Sienkiewicz: polska zmiana warty na Czarnym Łądzie*. W: *Wokół „W pustyni i w puszczy*, red. T. Bujnicki, J. Axer, Universitas, Kraków 2012, s. 107-125.

⁶⁰ J. Rybicki, *Sienkiewicz i Szklarski. Staś Tarkowski i Tomek Wilmowski. Rozpoznanie*. W: *Wokół „W pustyni i w puszczy*, red. T. Bujnicki, J. Axer, Universitas, Kraków 2012, s. 363-376.

⁶¹ J. Rybicki, *Polish Recreations of Spain in French: Jan Potocki's and Wojciech Has's Manuscripts Found in Saragossa*. W: *New Trends in English Teacher Education (Linguistics, Literature and Culture)*, red. I. Ramos Gay. Ediciones de la Universidad de Castilla-La Mancha, Cuenca 2008, s. 405-412.

⁶² J. Rybicki, *Uri Orlev: the Man from Both Sides*. W: *Poles and Jews. History, Culture, Education*, red. M. Misztal, M. Trawiński. Wydawnictwo Naukowe Uniwersytetu Pedagogicznego, Kraków 2011, s. 122-130.

⁶³ J. Rybicki, *Legends arturiańskie*. W: *Wielka historia literatury anglojęzycznej*, t. 2: *Literatura średniowieczna. Od najazdu normandzkiego do 1485*, red. M. Misztal, Wydawnictwo Naukowe Uniwersytetu Pedagogicznego, Kraków 2012, s. 223-285.

⁶⁴ J. Rybicki, *'To What Serves Mortal Beauty' in 'The Windhover'*. W: *Fictions and Metafictions of Evil. Essays in Literary Criticism, Comparative Literature and Interdisciplinary Studies*, red. G. Branny, J. Gill Holland, Peter Lang, Frankfurt a/Main 2013, s. 39-50.

4. Przekład literacki

Choć zwykle przekładu literackiego nie zalicza się do osiągnięć naukowych, uważam, że muszę o nim wspomnieć, ponieważ działalność ta była kiedyś w centrum moich zainteresowań i na pewno miała wpływ na moje badania przekładoznawcze. Wśród tłumaczonych przeze mnie autorów jest kilkoro noblistów (Buck, Golding, Gordimer, Ishiguro, Oe); mogę też pochwalić się najwyższą wśród polskich tłumaczy liczbą przekładów tekstów Johna le Carrégo. Warto dodać, że jeden z moich przekładów (*Pokolenie X* Douglasa Couplanda) był cytowany w pracach naukowych 26 razy. Łączna liczba moich przekładów literackich to 28 powieści.

b) działalność organizacyjna i międzynarodowa

Moja działalność w światowym środowisku humanistyki cyfrowej rozpoczyna się jeszcze w ubiegłym stuleciu. Przez pierwszą dekadę obecnego wieku byłem zazwyczaj jedynym polskim uczestnikiem światowych kongresów tej dyscypliny; w latach 2009-2012 byłem też członkiem komitetu programowego. W roku 2010 wszedłem w skład komitetu wykonawczego Association for Literary and Linguistic Computing, zrzeszającej europejskich humanistów cyfrowych i wkrótce przemianowanej na European Association for Digital Humanities, jednej z głównych organizacji zrzeszonych w ogólnoswiatowej Association of Digital Humanities Organizations. Członkiem komitetu wykonawczego EADH byłem przez dwie kadencje (do roku 2016); w tych samych latach byłem też członkiem Rady Naukowej najważniejszego i „*impact-factorowego*” czasopisma naukowego związanego z humanistyką cyfrową, *Literary and Linguistic Computing/Digital Scholarship in the Humanities*.

Niejako ukoronowaniem tej działalności i równocześnie największym osiągnięciem organizacyjnym było doprowadzenie do wyboru Krakowa (w konkurencji z Oksfordem, Rzymem, Los Angeles i Utrechtem) na miejsce dorocznego Światowego Kongresu ADHO (Digital Humanities 2016), współorganizowanego z dr hab. Maciejem Ederem, reprezentującym wówczas Uniwersytet Pedagogiczny. Obrady odbywały się w terminie 10-15.07.2016 w obiektach Uniwersytetu Jagiellońskiego i zgromadziły 902 zarejestrowanych uczestników z 42 krajów. Odbyły się łącznie 3 wykłady plenarne i 17 paneli a w 10 sesjach równoległych wygłoszono łącznie 260 referatów. Przedstawiono ok. 130 plakatów. Odbyło się 35 warsztatów przedkonferencyjnych. Pod każdym względem był to jak dotąd najliczniejszy kongres z serii Digital Humanities, który umieścił Kraków wśród najbardziej rozpoznawalnych centrów humanistyki cyfrowej na świecie mimo nieposiadania formalno-

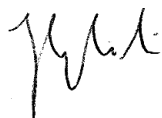
instytucjonalnego zaplecza. Współpraca między reprezentowanym przeze mnie Instytutem Filologii Angielskiej UJ a Instytutem Języka Polskiego PAN, kierowanym przez dr hab. Macieja Edera, określana jest czasem na świecie mianem „krakowskiej szkoły stylometrycznej” – co roku przyjeżdża do Krakowa kilkoro doktorantów lub pracowników naukowych instytucji zagranicznych, którzy uczestniczą w prowadzonych przez nas seminariach i projektach badawczych.

Inne organizowane przeze mnie konferencje to zebranie komitetu wykonawczego EADH w Krakowie (2010), seminarium Stylometry@Kraków (2012) i konferencja Stylometry Workshop w Żabnicy (2016). W ramach Kwietniowej Konferencji Anglistów w 2014 i 2017 r. organizowałem obszerne panele stylometryczne.

Od chwili powstania konsorcjum DARIAH-PL ds. humanistyki cyfrowej (2014) jestem członkiem jej rady jako pełnomocnik Rektora UJ. W europejskim związku krajowych konsorcjów, DARIAH-EU, pełnię funkcję członka Rady Naukowej (od 2016). Jestem również jednym z inicjatorów i przedstawicielem Polski w Komitecie kierującym COST Action CA16204 „Distant Reading for European Literary History” (2017-2021).

Regularnie uczestniczę w przedsięwzięciach popularyzujących naukę, m.in. w ostatnich Małopolskich Nocach Naukowców. Jestem częstym gościem popularno-naukowych audycji w Radiu Kraków, Programie 2 Polskiego Radia i w Radio TOKFM. Moje wykłady gościnne na kilku konferencjach krajowych i międzynarodowych dostępne są w całości w serwisie Youtube.

Kraków, 8.08.2018

A handwritten signature in black ink, appearing to be 'J. M.' or similar, written in a cursive style.