

**Prof. dr hab. B. Krzysztof Bogacki**

**Profesor emeritus Uniwersytetu Warszawskiego**

**Recenzja rozprawy doktorskiej**

**autorstwa mgr Marty Kunegundy Witkowskiej pt. „Analiza warsztatu  
pisarskiego Julii Hartwig metodami lingwistyki komputerowej”**

Przesłana mi do oceny dysertacja zawiera wyniki badań przeprowadzonych przez mgr Martę Kunegundę Witkowską pod kierownictwem prof. dra hab. Zygmunta Vetulaniego (promotor pomocniczy: prof. UAM dr hab. Marcin Telicki) i została przedłożona na Wydziale Filologii Polskiej i Klasycznej Uniwersytetu Adama Mickiewicza w Poznaniu w ramach przewodu doktorskiego. Struktura pracy jest bardzo przejrzysta, co ułatwia jej lekturę. Tekst liczy 154 strony, rozpada się na 5 rozdziałów poprzedzonych „przedśłowiem” i wstępem. Na końcu pracy znaleźć można podsumowanie i bibliografię . W chwili obecnej prace naukowe zwyczajowo zaopatruje się w streszczenie zredagowane po angielsku i w języku, w którym została napisana dysertacja. Tak też jest w recenzowanej rozprawie. Czytelnik znajdzie *abstract* na początku książki, po spisie treści. Jest też streszczenie polskie i francuskie. Indeks pozwala odnaleźć w tekście nazwiska i najważniejsze terminy. Przewidziano „Słownik terminów” (str. 133-135) wyjaśniający 37 najważniejszych użytych w pracy oraz „aneks” zawierający informacje o systemie kodowania fleksji polskiej POLEX (symbole, ich znaczenie, liczba jednostek, przykłady). Wykaz zamieszczony na str. 149-150 zawiera spis tabel i wykresów. W wersji wydrukowanej strony 115 – 120 zostały powtórzone, dla poprawy czytelności, w postaci rozkładanych uzupełnień umieszczonych na końcu rozprawy (patrz strony o numerach 151-154 w wersji pdf).

Bliższe zaznajamianie się z pracą zacznę od bibliografii (str. 139-148). To ona pozwala w dość dużym stopniu zdać sobie sprawę z zawartości pracy: przyjętych ram teoretycznych i metodologii. Otóż bibliografia jest jak na pracę doktorską obszerna i bardzo dobrze dobrana. Uwzględniła najważniejsze publikacje, w tym te świeżej daty,

w językach angielskim, francuskim i polskim. Lista obejmuje 159 pozycji z tego 128 jest po polsku, 27 po angielsku i 4 po francusku.

Korpus tekstów, na których przeprowadzono badania został utworzony na podstawie publikacji J. Hartwig, które podlegały różnym ograniczeniom usprawiedliwionym przekonująco przez Doktorantkę. Dotyczyły one np. bogatego dorobku translatorskiego, wywiadów, komentarzy, recenzji, korespondencji, twórczości adresowanej do dzieci. Wykorzystano 32 pozycje książkowe pisarki (tylko pierwsze wydania na przestrzeni lat od 1954 do 2016; 10 tomów prozy oraz 22 tomy poezji). Cała dysertacja ma bardzo staranną formę redakcyjną i edytorską, co ułatwia lekturę. Sposób cytowania literatury w tekście jest spójny, podobnie jeśli chodzi o podział na części składowe i ich numerowanie. Recenzent, do którego obowiązków obok oceny merytorycznej należy tropienie literówek czy innych potknięć redakcyjno-formalnych tym razem miał kłopot w doszukaniu się poważniejszych uchybień pod tym względem.

Praca rozpoczyna się od przedstawienia celu badawczego, oraz postaci i dorobku Julii Hartwig. Zważywszy na typ badań oraz na materiał, na którym zostały one przeprowadzone, nie ma potrzeby uzasadniania wyboru tej właśnie autorki. Doktorantka zajęła się warsztatem pisarki, której działalność przypadła w okresie powojennym pełnym doniosłych wydarzeń politycznych i która okazała się nietuzinkową postacią w historii literatury i kultury polskiej. Jej dorobek pisarski nie ogranicza się do poezji i tłumaczeń. Historycy literatury zajmujący jej twórczością piszą o dziennikach, felietonach, esejach, wstępach i komentarzach do książek innych autorów, wzmiankują też pozycje epistolograficzne, literaturę dla dzieci, tzw. „błyski” (oryginalny gatunek właściwy tylko J. Hartwig mający formę krótkich zapisków) itd.

Z wykształcenia Julia Hartwig była romanistką, jednak pisarka podczas pobytu Stanach Zjednoczonych w latach 1970-1974 zetknęła się literaturą anglojęzyczną, co zostawiło znaczący ślad w jej twórczości. Jej działalność spotkała się z szerokim uznaniem, czego dowodem były liczne wyróżnienia przyznane na świecie zarówno przez stowarzyszenia twórcze jak i instancje akademickie czy rządy: we Francji (1978 oraz 2007 [Legia Honorowa]), w USA, (1981, 1986), w Austrii (1991), we Włoszech (2011 – Premio Acerbi za całokształt twórczości) i wielokrotnie w Polsce (m. in. Nagroda Polskiego PEN Clubu (1979, 1997), Krzyż Wielki Orderu Odrodzenia Polski

(2011), Krzyż Oficerski Odrodzenia Polski (1997), Medal „Zasłużony Kulturze Gloria Artis” (2005) i wreszcie doktorat honoris causa nadany przez Uniwersytet im. Adama Mickiewicza w Poznaniu (2015).

Wszystkie te informacje czytelnik znajdzie na samym początku dysertacji (do II rozdziału włącznie), wraz ze sformułowaniem opisującym skrótowo problem badawczy w niniejszej pracy. Otóż jest nim sprawdzenie, które elementy w wypowiedziach Julii Hartwig można uznać za najważniejsze, tzn. orzekające o świecie i odsyłające do konkretnych wrażeń. Odpowiedź jest prosta i udziela jej sama pisarka za nią zaś powtarzają krytycy literaccy zajmujący się jej twórczością: funkcję tą pełni przymiotnik mający charakter predykatywny, stanowiący najważniejszą część mowy i będący wykładnikiem „literackości” dzieła. Z recenzenckiego obowiązku muszę zauważyć, że wartość predykatywną mają także inne części mowy: w pierwszym rzędzie czasownik, ale także niektóre rzeczowniki, przysłówki a nawet przyimki (np. *Jan jest przeciw*). Skoro więc wartość predykatywną mają także inne części mowy, warto by zadać sobie pytanie o zestawienie przymiotników z innymi, predykatywnymi częściami mowy i zbadanie różnych aspektów – jakościowych i ilościowych występowania ukochanych przez J. Hartwig przymiotników z nieprzymiotnikowymi predykatami: czy proporcje między tymi grupami zmieniają się w miarę upływu czasu? Czy są zależne od tematyki utworów czy nie? itd. Zagadnienia te mogłyby wydłużyć listę przyszłych badań wyszczególnionych w ostatnim rozdziale.

Tym, co stanowi o oryginalności przedstawionej pracy jest fakt, że analiza warsztatu pisarskiego Julii Hartwig odbywa się z zastosowaniem metod zaczerpniętych z lingwistyki komputerowej, co w praktyce oznacza odwołanie się w pierwszym rzędzie do korpusów elektronicznych w formatach przystosowanych do ich przeszukiwania i dalszego przetwarzania. Doktorantka ma też do swej dyspozycji bogatą gamę technik badawczych. Ilość narzędzi jest całkiem pokaźna, zwłaszcza gdy uświadomimy sobie, że nurt badawczy, w który wpisuje się Doktorantka pojawił się stosunkowo niedawno, a prawdziwa eksplozja badań nad językiem przy użyciu komputerów stała się możliwa z chwilą upowszechnienia się informatyki oraz spadku cen maszyn i oprogramowania. Tym dwóm grupom zagadnień poświęcone są kolejne rozdziały: III i IV.

Rozdział III omawia szereg zagadnień poświęconych korpusowi tekstów J. Hartwig. Doktorantka zajmuje się najpierw ogólnie pojęciem korpusu, wszak metodologia dysertacji wpisuje się w nurt *lingwistyki korpusowej*, a wbrew pozorom stworzenie korpusu nie należy do zadań prostych, słusznie więc czyni Autorka omawiając takie cechy korpusu stanowiące o jego jakości jak jego reprezentatywność, rozmiary, zrównoważenie, otwartość, referencyjność itd. Wzmiankuje oczywiście różne korpusy tekstowe wykorzystywane na świecie poczynając od korpusu Uniwersytetu Brown, szereg korpusów języka angielskiego (*American National Corpus*, *British National Corpus*, *Oxford English Corpus*, *Penn Treebank*), niemieckiego (*DeReKo*), czeskiego (*Czech National Corpus*), rosyjskiego (*Russian National Corpus*). Jeden z podrozdziałów zawiera rozbudowaną informację o korpusach języka polskiego. Jako romanista szukałem informacji o korpusach z danymi pochodzącymi z języków romańskich. Otóż wspomniany jest hiszpański (<https://www.corpusdelespanol.org>), ale nie prestiżowy korpus Królewskiej Akademii (*Corpus del diccionario histórico*). Pominęto język francuski, który mógłby zostać uwzględniony choćby przez doskonale otagowany Frantext zawierający 260 milionów wyrazów i wykorzystany do opracowania drukowanego słownika *Trésor de la Langue Française*, czy któraś ze skromniejszych pod względem rozmiarów leksykalnych baz danych rozwijanych w różnych uczelniach francuskich i kanadyjskich. Nie czynię z tego zarzutu. Analogiczne korpusy istnieją dla każdego języka romańskiego, niekiedy dla niektórych najważniejszych dialektów (prowansalski, gaskoński, pikardzki, normandzki itd.), nie sposób wymienić je wszystkie, ani nawet najważniejsze. Istotna jest odpowiedź na pytanie o to, czy Doktorantka zdaje sobie sprawę z tego, jakie warunki powinien spełniać korpus brany do badań nad warsztatem pisarskim. Otóż nie ulega wątpliwości, że rozdział III zawiera pozytywną odpowiedź na tak postawione pytanie.

Druga część rozdziału zawiera informacje na temat zawartości korpusu prozy i poezji. W pierwszym znalazły się dzienniki, reportaże, monografie literackie, zbiór esejów i zbiór felietonów. W korpusie poezji mamy 22 utwory. Prace pochodzą z okresu od 1956 do 2016 roku. Doktorantka opisała też kolejne etapy pozyskiwania materiałów do korpusu: skanowanie, dygitalizacja tekstu za pomocą OCR, wykorzystanie wersji elektronicznej (pliki .pdf), których przetwarzanie jest regulowane przepisami prawa autorskiego. Bardzo szczegółowo wyszczególniono trudności, które

pojawiły się na różnych etapach normalizacji tekstów konkludując, że „etap przygotowawczy do przeprowadzenia właściwych analiz okazał się czasochłonny, niemniej przy stosunkowo małym korpusie możliwy do samodzielnej realizacji. Ponadto normalizacja przyniosła korzyści, umożliwiając przyspieszenie prac w kolejnych etapach przetwarzania danych.” (str. 66).

Rozdział IV, najobszerniejszy, poświęcony jest przedstawieniu narzędzi i metod, które zostały użyte w automatycznym przetwarzaniu tekstu do analizy przymiotnikowej strukturalizacji predykatu w korpusie prozy i poezji Julii Hartwig. Użyte zostały 4 programy.

- Lematyzator POLEX posłużył do utworzenia listy przymiotników, która jest traktowana jako lista referencyjna i wykorzystywana dalej we wszystkich obliczeniach ilościowych przedstawianych w postaci list frekwencyjnych przymiotników i ekscerpcji kontekstów użycia przymiotników w formie konkordancji.

- Text SubCorpora Creator (TSCC1.3.) posłużył do ekstrakcji podkorpusów wyodrębnionych z większych zbiorów danych tekstowych, które były dzielone na części o określonej długości, a nadto do przeprowadzenia analizy „nasylenia” korpusu poprzez obserwowanie przyrostu nowych form w kolejnych próbkach korpusu<sup>1</sup>. Oba wyżej wymienione narzędzia pochodzą z Zakładu Lingwistyki Informatycznej i Sztucznej Inteligencji UAM.

- Normalizacji tekstu i obliczeń frekwencji wyrazów dokonano za pomocą języka programowania wysokiego poziomu ogólnego przeznaczenia Python w wersji 3.4 współpracującego z biblioteką Natural Language Toolkit.

- Użyta została wreszcie opracowana w Zespole Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN aplikacja o nazwie Korpusomat<sup>2</sup>. Wykorzystano ją m. in. do porównywania korpusów, generowania statystyk i pozyskiwania kolokacji (64 000 tysięcy dla korpusu prozy i ponad 11 000 dla korpusu poezji).

---

<sup>1</sup> Konsultantem implementacji systemu była Doktorantka nie będąc jednak pracownikiem Zakładu Lingwistyki Informatycznej i Sztucznej Inteligencji.

<sup>2</sup> Jest to część polskich zasobów infrastruktury dla humanistów w ramach projektu Clarin-pl

Kolejne operacje dokonywane przy wykorzystaniu wymienionych powyżej narzędzi są opisane w sposób wyjątkowo skrupulatny i obficie ilustrowane. Nie miałbym nic przeciwko pominięciu niektórych szczegółów technicznych niezbędnych dla laika niezbyt biegłego w posługiwaniu się narzędziami informatycznymi (por. np. wyszczególnianiu wyrażeń nierozpoznanych w plikach wyjściowych etapu lematyzacji). Godne pochwały są natomiast następujące punkty: po pierwsze, bardzo bogata prezentacja uzyskanych rezultatów, po drugie, precyzyjne omówienie predykatów przymiotnikowych sytuujących się w pierwszej dwudziestce na listach frekwencyjnych, po trzecie, wyczerpująca analiza nasycenia obu korpusów formami przymiotnika, po czwarte, przeanalizowanie przyrostu form przymiotnikowych (tokenów) dla korpusu i rozwoju zasobu leksykalnego wykorzystywanego przez pisarkę. Rozdział zawiera 17 tabel i kilka zrzutów ekranowych. Rozważania ilustrowane są bardzo porządnie wykonanymi wykresami – to bardzo ułatwia lekturę tekstu.

Rozdział V podsumowuje wykonane badania i sugeruje niektóre kierunki dalszych prac możliwych do wykonania przy wykorzystaniu dotychczas zgromadzonego materiału i niezbędnych uzupełnieniach (np. rozszerzeniu lematyzatora POLEX o moduł składniowo-semantyczny).

Oceniając ogólnie tekst mgr M. K. Witkowskiej należy podkreślić umiejętność wprowadzenia w tematykę badawczą i jasność sformułowania celu, do którego zmierza. Muszę pochwalić trafny dobór metod badawczych i poprawną analizę wyników, wreszcie jasność i poprawność wniosków. Przemyslenia Autorki są przekazane językiem eleganckim w sposób precyzyjny, bez przytłaczających technicyzmów.

Bardziej szczegółowo wartość merytoryczną pracy mgr M. K. Witkowskiej wypada ocenić w dwóch płaszczyznach. Pierwsza odnosi się do warstwy informatycznej. Odnotuję tu kilka spraw. Lektura przedłożonej dysertacji prowadzi do wniosku, że jej Autorka bardzo sprawnie posługuje się skomplikowanymi narzędziami informatycznymi, a rezultaty analiz pozwalają powiedzieć na temat języka Julii Hartwig znacznie więcej niż to wynika z obiegowych opinii o warsztacie artystycznym pisarki wygłaszanych na podstawie deklaracji samej autorki i powielanych przez wielu krytyków literackich na podstawie tradycyjnych „ręcznych”

badania. Konfrontowanie informatyki z czymś tak nieuchwytnym i subiektywnym jak „styl pisarza” czy „warsztat pisarza” wydaje się być zadaniem karkołomnym. Doktorantka pokazała jednak, że zadaniu temu można podołać i na tym polega jej osiągnięcie. Oczywiście, użyte narzędzia niejako ze swej natury nakierowane są na badanie aspektów ilościowych słownictwa i dynamiki rozwoju warsztatu twórczego pozwalając zanalizować jak zmienia się nasycenie leksykalne polegające na zatrzymaniu lub spowolnieniu przyrostu nowych leksemów w badanym korpusie. Analiza ilościowa pozwoliła zauważyć różnicę między korpusem poezji i prozy co do stopnia powiązania między tytułami i tematyką a częstotliwością przymiotników należących do obszaru leksykalnego związanego z tytułem lub tematyką dzieła. Ustalono też na podstawie list frekwencyjnych jakie obszary semantyczne są szczególnie eksploatowane przez pisarkę.

Nie ulega wątpliwości, że w realizacji zadania, jakie postawiła przed sobą Doktorantka zastosowane oprzyrządowanie informatyczne sprawdza się doskonale, Autorka dysertacji daleka jest jednak od bezkrytycznej fascynacji elektroniką. Pokazuje ona też słabe strony dostępnych aktualnie narzędzi oraz sugeruje w jakim kierunku powinny iść poprawki i uzupełnienia dostępnego oprogramowania NLP. Na przykład rozpowszechniona w językach naturalnych homonimia form wyrazowych zafałszowuje obraz, jakiego dostarczają listy frekwencyjne, na których zliczane są razem wystąpienia formy *była* należącej do paradygmatu czasownika *być* i przymiotnika *były*. Część z dostrzeżonych braków w oprogramowaniu została zresztą przez Doktorantkę usunięta, co doprowadziło do poprawienia funkcjonalności. (por. rozdz. IV). Efektem pracy wykonanej przez Doktorantkę jest stworzenie korpusu Julii Hartwig nadającego się do dalszych badań wykorzystujących metodologię dająca wyniki pewne, które trudno zakwestionować.

Druga sfera, która podlega ocenie jest natury lingwistycznej. I tutaj też moja ocena kompetencji Doktorantki jest pozytywna. Kluczowym pojęciem jest nadzwyczaj wieloznaczne pojęcie predykatu, które pojawiło się już u starożytnych Greków, m. in. u Arystotelesa, następnie zaś było obecne na przestrzeni wieków głównie w rozważaniach logików, filozofów i językoznawców. Nawet gdyby ograniczyć się do lingwistyki, okazałoby się, że pojęcie to ma różne znaczenia: kilka składniowych i kilka semantycznych. W sumie daje to całą mozaikę znaczeń, w których nie trudno się pogubić. Jeśli dodać do tego fakt, że na analizę predykatowo-argumentową powołują

się ostatnio etnologzy, kulturoznawcy, semiotycy, socjologzy i inni specjaliści posługujący się terminologią, która nie została zunifikowana, to tym bardziej docenić trzeba klarowny i precyzyjny sposób referowania przez Doktorantkę tego, co wiąże się z predykatem.

To, co dotychczas powiedziałem pozwala domyśleć się w jakim kierunku zmierzam. Sformułuję więc mój wniosek końcowy *expressis verbis*.

W świetle przytoczonych wyżej argumentów stwierdzam, że przedłożona mi do oceny praca mgr Marty K. Witkowskiej pt. *„Analiza warsztatu pisarskiego Julii Hartwig metodami lingwistyki komputerowej”*

**spełnia z naddatkiem wymagania stawiane rozprawom doktorskim**

określone w ustawie z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (patrz Dz. U. nr 65, poz. 595, z późn. zm.) Pracę uważam za wyróżniającą. Wnoszę zatem o dopuszczenie jej Autorki do dalszych etapów przewodu doktorskiego.



**Prof. dr hab. B. Krzysztof Bogacki**

Warszawa, dnia 01 września 2021