



Agata Kwaśnicka-Janowicz

 <https://orcid.org/0000-0002-4827-6941>

Uniwersytet Jagielloński
Kraków, Polska

Archiwizowanie dziedzictwa językowego – w poszukiwaniu własnej tożsamości*

Archiving the Linguistic Heritage – In Search of Identity

Abstract: The aim of this article is to characterise philological research conducted today using digital tools. Its main idea is to summarise previous research approaches within the new scientific discipline of digital humanities. It is characterised as a new model of scholarly communication that uses new technologies to conduct research and is interdisciplinary and collaborative in nature. The text is of a reporting and reviewing nature and it briefly describes the research perspective of the global digital humanities as a socially open science that exploits the potential of media and digital technologies. It juxtaposes this perspective with digitisation work that aims to archive cultural heritage fixed in language in the digital humanities environment in Krakow. It presents the most important digital resources in the field of historical linguistics in Poland (digitised dictionaries and corpora of historical Polish) and postulates the creation of a diachronic corpus of the Polish language for the study of its linguistic and cultural heritage. The general conclusion of the review is that the use of digital methods in archiving linguistic and cultural heritage brings new results in the form of large, unprecedented data sets. Their processing and interpretation using the methods of computational and quantitative linguistics can help formulate answers to key questions in the humanities. The research was performed with the financial support of Priority Research Areas: POB Heritage and POB DigiWorld as part of the Strategic Excellence Initiative Programme at Jagiellonian University.

Keywords: digital humanities, archiving linguistic and cultural heritage, historical Polish language corpora

Abstrakt: Celem artykułu jest charakterystyka badań filologicznych prowadzonych współcześnie z użyciem narzędzi cyfrowych. Jego głównym założeniem jest podsumowanie dotychczasowych podejść badawczych w ramach nowej dyscypliny naukowej, jaką jest humanistyka cyfrowa. Opisuje się ją jako nowy model komunikacji naukowej, który wykorzystuje innowacyjne technologie do prowadzenia badań oraz ma charakter interdyscyplinarny i zespołowy. Tekst ma charakter sprawozdawczy i przeglądowy. Krótko opisuje perspektywę badań światowej humanistyki cyfrowej jako otwartej społecznie nauki wykorzystującej potencjał mediów oraz technologii cyfrowych. Zestawia tę perspektywę z digitalizacyjnymi pracami, które zmierzają do archiwizacji dziedzictwa kulturowego utraconego językowo w środowisku krakowskiej humanistyki cyfrowej. Prezentuje najważniejsze zasoby cyfrowe z zakresu językoznawstwa historycznego w Polsce (zdigitalizowane słowniki i korpusy historycznej polszczyzny) oraz postuluje stworzenie korpusu diachronicznego polszczyzny dla badań jej dziedzictwa językowego i kulturowego.

* Artykuł ukaże się równolegle w tomie *Filologia – od/nowa. Język, literatura, kultura w epoce cyfrowej*, red. T. Bilczewski, K. Górniak-Prasnał, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2025 [w druku].

Ogólnym wnioskiem z dokonanego przeglądu jest stwierdzenie, że wykorzystanie metod cyfrowych w archiwizowaniu dziedzictwa językowego i kulturowego przynosi nowe efekty w postaci dużych, niespotykanych dotychczas zbiorów danych. Ich przetworzenie i interpretacja z użyciem metod językoznawstwa komputerowego i kwantytatywnego mogą pomóc w formułowaniu odpowiedzi na kluczowe pytania humanistyki. Badania wykonano przy wsparciu finansowym Priorytetowych Obszarów Badawczych: POB Heritage i POB DigiWorld w ramach Programu Strategicznego Inicjatywa Doskonałości w Uniwersytecie Jagiellońskim.

Słowa kluczowe: humanistyka cyfrowa, archiwizowanie dziedzictwa językowego i kulturowego, korpusy historycznej polszczyzny

Humanistyka cyfrowa jako część nowej humanistyki

Szeroko dziś dyskutowana humanistyka cyfrowa jest częścią nowej humanistyki, o której w 2017 roku Ryszard Nycz pisał:

Humanistyka cyfrowa to najbardziej dziś rozpoznawalna, a także reprezentatywna marka nowej humanistyki oraz hasło bodaj największej rewolucji w myśleniu o jej zadaniach, o ciągle niedających się przewidzieć możliwościach rozwojowych, a zatem również ciągle podsycanych nowych nadziejach poznawczych. Choć bynajmniej nie stawia sobie celów zastąpienia dotychczasowej specyfiki badawczej humanistyki, to stwarza wrażenie, że z tymi narzędziami zdolna jest do wszystkiego (Nycz, 2017: 24).

Mówiąc o potencjale nowych ujęć, badacz zwracał uwagę na zagrożenia płynące z analiz ilościowych *big data*:

Ekstrapolując praktyki humanistyki cyfrowej do skrajnej postaci, można powiedzieć, że prowadzi ona do zastąpienia: badań tekstów (literackich i kulturowych) – opracowaniem danych (*big data*); teorii – aplikacjami technologicznych programów; interpretacji – wizualizacją, modelowaniem, symulacją (a więc formami opisu); zainteresowania tym, co jednostkowe, oryginalne, wartościowne i wartościotwórcze – statystyczną analizą makrotendencji (Nycz, 2017: 26).

Charakteryzując w tym samym 2017 roku środowisko polskich humanistów cyfrowych, Maciej Maryl wskazywał na trzy fale humanistyki cyfrowej, które dają ująć się w triadę: digitalizacja materiałów – badania ilościowe i jakościowe tak przetworzonych danych – krytyczna analiza epistemologii narzędzi cyfrowych:

pierwsza fala humanistyki cyfrowej dotyczy digitalizacji i remediacji tradycyjnego warsztatu badawczego dyscyplin (zaliczymy do niej skanowanie tekstów i obiektów, budowę cyfrowych archiwów i katalogów, tworzenie korpusów, adnotowanie tekstów itd.); fali drugiej towarzyszy rozpoznanie i wykorzystywanie możliwości dalszego przetwarzania materiału cyfrowego pozyskanego w fazie pierwszej, czyli wykorzystanie nowego medium do tworzenia nowych metod badawczych i gatunków komunikacji naukowej (narzędzia do badania i wizualizacji, badania na dużych zbiorach materiałów), oraz fala trzecia, czyli krytyczna analiza ograniczeń epistemologicznych medium i samych założeń humanistyki cyfrowej (Maryl, 2017: 291).

Maryl wskazywał również na główne problemy rozwoju i funkcjonowania badań z zakresu humanistyki cyfrowej (DH) w Polsce. Punktował rozproszenie działań i prowadzonych projektów, brak skoordynowanych działań polegających na całościowym zarządzaniu DH, obsługiwanych dziś przez resorty: nauki, kultury i cyfryzacji, i wreszcie niskie finansowanie badań (Maryl, 2017).

Światowa perspektywa humanistyki cyfrowej

Jak w tym kontekście rysuje się perspektywa światowej humanistyki cyfrowej? Podczas inauguracji Jagiellońskiego Centrum Humanistyki Cyfrowej w Krakowie 20 czerwca 2023 roku prof. Ray Siemens

z University of Victoria wygłosił wykład: *Foundations for Open Social Scholarship: The Implementing New Knowledge Environments Partnership*. Mówił w nim o budowaniu otwartej sieci badawczej jako nowej formule uprawiania nauki, wykorzystującej nowoczesne możliwości publikacji i komunikacji naukowej: otwarty dostęp, metody i narzędzia humanistyki cyfrowej (bazy danych, potencjał mediów audialnych i wizualnych). Tę nową usieciowioną humanistykę cechują: duży wpływ (*high impact*), zaangażowanie społeczne, szeroki zakres oraz długo-trwałość działań i efektów (*long term impact*).

Otwarta społecznie nauka obejmuje tworzenie i rozpowszechnianie badań naukowych oraz technologii badawczych wśród szerokiego grona specjalistów i zaangażowanych niespecjalistów w sposób przystępny i odpowiadający potrzebom różnych grup odbiorców, co znajduje zastosowanie w praktyce badawczej i edukacyjnej. Przykładem tego są m.in. zorientowane na praktykę takie inicjatywy jak: *Implementing New Knowledge Environments Partnership* (inke.ca), *Digital Humanities Summer Institute* (dhsi.org) czy *Electronic Textual Cultures Lab* (etcl.uvic.ca). Ta ostatnia stanowi centrum współpracy dla otwartych cyfrowych praktyk naukowych, angażując partnerów przez inicjatywy zorientowane na społeczeństwo. Tematyka proponowanych wykładów dotyczy m.in. bieżących problemów związanych z „ucyfrowieniem” rzeczywistości, takich jak szybki rozwój AI i jej wpływ na działania literackie. Proponowany wśród wielu wykład *Navigating Digital Writing and AI: An Optimistic and Slightly Apocalyptic Examination* porusza etyczne implikacje wykorzystania sztucznej inteligencji w literaturze, w tym kwestie praw autorskich i potencjalnego zniknięcia *human writers*. Stawiane są w nim pytania o możliwości kreowania nowej kategorii literackiej: fikcji generowanej przez sztuczną inteligencję, stanowiącej produkt marketingowy obok ludzkiej fikcji, oraz możliwości istnienia fikcji hybrydowej (*human + AI*). Pytania te wydają się niezbędne do zrozumienia, jak działa kreatywność, oraz do ustalenia, czy powinny istnieć granice tego, co technologia cyfrowa może wykonać lub stworzyć w literackim łańcuchu produkcji i konsumpcji.

Perspektywa (krakowskiej) humanistyki cyfrowej – archiwizowanie dziedzictwa językowego

W kontekście tak postawionych pytań tym istotniejsze stają się działania z zakresu humanistyki cyfrowej, których celem jest archiwizacja dziedzictwa kulturowego, utrwalonego przez medium/narzędzie językowe. Rejestrowanie i dokumentacja wytworów ludzkiej działalności tworzą językowe archiwum pamięci wspólnot komunikatywnych i narodów, umożliwiające filologiczną interpretację dokumentów oraz badanie tego, co kreujące je wspólnoty i narody uznawały za prawdziwe i dla siebie tożsame.

Wykorzystując klasyczną metodę filologiczną (badanie języka i jego wytworów), w której głównym przedmiotem zainteresowania jest tekst oraz jego komentarz i interpretacja, środowisko krakowskie prowadzi prace nad archiwizowaniem i anotacją tekstów polskiego dziedzictwa językowego, z prezentacją danych w standardzie TEI. Takim przedsięwzięciem jest Projekt Flagowy Digital Humanities Lab (PF DHL): <https://dhlab.id.uj.edu.pl/>, zainicjowany w ramach Programu Strategicznego Inicjatywa Doskonałości w Uniwersytecie Jagiellońskim, którego głównym celem stało się stworzenie środowiska do opracowania cyfrowych edycji naukowych i przestrzeni do ich prezentacji. Tym zadaniom służy utworzona dzięki współpracy krakowskich edytorów cyfrowych, informatyków UJ oraz twórców i światowego środowiska TEI – Jagiellońska Platforma Cyfrowa: <https://labedyt.dhlab.uj.edu.pl/index.html>. Na obecnym etapie prac zostały przygotowane blueprinty edycji cyfrowej Bibliotheca Neolatina Literatury Staropolskiej i Nowołacińskiej (projekt Pracowni Literatury Renesansu, Wydział Polonistyki UJ), prezentujące twórczość łacińską Jana Kochanowskiego: <https://neolatina.dhlab.uj.edu.pl/home>. W przygotowaniu pozostaje także pierwsza naukowa edycja cyfrowa *Moralionów* Wacława Potockiego, ważna nie tylko ze względu na swoją wartość naukową. Wydanie to jest okazją do opracowania i przetestowania narzędzi oraz rozwiązań niespotykanych powszechnie w edycjach cyfrowych. W toku prac nad edycją powstał

model do optycznego rozpoznawania polskiego tekstu rękopiśmianego z XVII wieku. Planowane jest również opracowanie rozwiązania pozwalającego na częściową automatyzację modernizacji ortografii (transkrypcji), a także dostosowanie poziomu objaśnień w zależności od preferencji odbiorcy edycji. Edycja *Moralistów* posłuży do stworzenia aplikacji wzorcowej dla edycji cyfrowych tekstów dawnych, uwzględniających faksymile, transliterację i transkrypcję oraz wielopoziomowy komentarz.

W ramach prac zespołu Laboratorium Edytorstwa Cyfrowego planowane jest opracowanie modelowej edycji korespondencji z uwzględnieniem procedury konwersji edycji pierwotnie przygotowanej do druku (edycja korespondencji Szymona Laksa z Krystyną i Czesławem Bednarczykami), stworzenie wzorcowej edycji powiązanych ze sobą dokumentów z wykorzystaniem kodowania TEI oraz narzędzia TEI Publisher na przykładzie źródeł historycznych projektu „IURA. Źródła prawa dawnego”, działającego dotychczas w środowisku dLibra: <https://iura.uj.edu.pl/dlibra>, oraz prezentacja projektów studenckich edycji cyfrowych realizowanych na Wydziale Polonistyki UJ, por. <https://editions.dhlab.uj.edu.pl/index.html>.

Drugą składową Projektu Flagowego Digital Humanities Lab jest Elektroniczny korpus literatury w języku polskim UJ: <https://dhlab.id.uj.edu.pl/czytanie>. Stanowi on zestaw pełnotekstowych, elektronicznych wersji tekstów literackich w języku polskim od początków piśmiennictwa polskiego do chwili obecnej i służy jako materiał do badań ilościowych nad językiem oryginałów i przekładów (połączenie stylo-metrii danych tekstowych ze statystyczną analizą metadanych: autor, tłumacz, rok pierwszego wydania oryginału, rok pierwszego wydania przekładu, język oryginału i in.).

W ramach prac zespołów badawczych PF DHLab realizowany jest również projekt cyfrowej archiwizacji, klasyfikacji i anotacji tekstów tłumaczonych na język polski od roku 1918 do dziś. Projekt „Ukryta kolekcja: Internetowa baza literatury tłumaczonej w Polsce od 1918 roku” ma na celu stworzenie innowacyjnego narzędzia dla badaczy przekładu w Polsce – dostępnej online bazy metadanych publikacji książkowych będących

tłumaczeniami na język polski od 1918 roku do dziś. Projekt jest realizowany przez Centrum Badań Przekładoznawczych we współpracy z Biblioteką Narodową: <https://dhlab.id.uj.edu.pl/ukryta-kolekcja/>.

Archiwizowanie dorobku badań polskiej kultury ludowej, w tym jej tekstów, umożliwiała baza relacyjna Cyfrowa bibliografia dialektologii polskiej (1986–1995 wraz z uzupełnieniami za lata poprzednie; wydanie elektroniczne [zmienione, uzupełnione i rozszerzone po roku 1996]), będąca projektem rozwojowym realizowanym na Wydziale Polonistyki UJ: <https://cbdp.uj.edu.pl/>.

Metody korpusowe i kwantytatywne w językoznawstwie historycznym

Wobec prowadzonych w pierwszej fali działań DH, polegających na digitalizacji danych (skanowanie tekstów i obiektów, budowa cyfrowych archiwów i katalogów, tworzenie korpusów, anotowanie tekstów), remediacji podlega także tradycyjny warsztat badawczy historyka języka. Wobec współczesnego zaawansowania prac digitalizacyjnych tym pilniejsza wydaje się potrzeba budowy korpusu diachronicznego¹ polszczyzny². W językoznawstwie polonistycznym dysponujemy dotychczas jedynie (rozproszonymi) korpusami historycznymi języka polskiego, na które składają się, w układzie chronologicznym:

- Korpus tekstów staropolskich: ijp.pan.pl/publikacje-i-materialy/zasoby/korpus-tekstow-staropolskich/ (transkrypcje zabytków [pliki PDF] + korpus w postaci plików XML), projekt Pracowni Języka Staropolskiego IJP PAN;

¹ Korpus diachroniczny to „seria podkorpusów reprezentujących poszczególne epoki”, złożona ze zbioru korpusów historycznych języka polskiego („korpusy reprezentujące stan języka w danej epoce historycznej”; Górski, Król, Eder, 2019: 14).

² O Projekcie Narodowego Korpusu Diachronicznego Polszczyzny: Pastuch i in., 2018; Król i in., 2019.

- Korpus polszczyzny XVI wieku (w opracowaniu): spxvi.edu.pl/korpus/ (transliterowane teksty kanonu SP XVI), próba korpusu z wyszukiwarką: <https://spxvi.edu.pl/korpus/probka/szukaj> (transliterowane i transkrybowane teksty kanonu; lematyzacja, metadane, lokalizacja w źródle z dokładnością strony, hiperlink do hasła w *Słowniku polszczyzny XVI w.*), projekt Pracowni Słownika Polszczyzny XVI wieku IBL PAN;
- Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772 r.)³ [KorBa]: korba.edu.pl/query_corpus/ (transkrypcja i transliteracja tekstów, metadane, znakowanie strukturalne i językowe oraz anotacja morfosyntaktyczna i lematyzacja, składnia zapytań, filtrowanie wyników oraz ich lokalizacja w źródle z dokładnością do numeru strony), projekt Pracowni Historii Języka Polskiego XVII i XVIII wieku IJP PAN;
- Korpus tekstów polskich z lat 1830–1918: korpus19.nlp.ipipan.waw.pl/overview, por. szukajwslownikach.uw.edu.pl/ (transliterowany i transkrybowany korpus tekstów, w tym podkorpus polszczyzny Józefa Ignacego Kraszewskiego, z anotacją morfosyntaktyczną i lematyzacją, składnią zapytań, metadanymi), projekt UW i IPI PAN.

Próbami konsolidacji tych zasobów są inicjatywy Instytutu Podstaw Informatyki PAN, mieszczące wymienione korpusy pod jednym adresem: Written Corpora of Historical Polish: clip.ipipan.waw.pl/LRT, oraz projekt grupy CLARIN-PL Platforma leksykalna: lexp.clarin-pl.eu/. Ta ostatnia oferuje wspólny format prezentacji zasobów leksykalnych ze źródeł: Słowosieć, Walenty, SGJP, HASK, SPXVI, ESPXVII, eSWił, a także Open Multilingual WordNet. Zasób ten pozwala na wyszukiwanie, przeglądanie i nawigowanie po materiałach na podstawie wspólnego formularza słownego i identyfikatora składni.

³ Dla lat 1570–1756 funkcjonują dwa dodatkowe, anotowane korpusy polszczyzny: Korpus GT projektu IMPACT (w-1): szukajwslownikach.uw.edu.pl/IMPACT_GT_1/ oraz Korpus GT projektu IMPACT (w-2): szukajwslownikach.uw.edu.pl/IMPACT_GT_2/ (Pałka, Kwaśnicka-Janowicz, 2024).

Jak pokazuje praktyka badawcza, potencjał stworzonego na potrzeby konkretnego projektu (Górski, Król, Eder, 2019) korpusu diachronicznego polszczyzny pozwala na obserwację dynamiki zmiany językowej. Wyznacza tym samym cezury czasowe w dziejach języka danej wspólnoty, co może zostać użyte do opracowania nowej (zweryfikowanej kwantytatywnie) periodyzacji dziejów języka polskiego. O nowych możliwościach wykorzystania korpusów w pracy historyka języka, także do śledzenia zmian językowych w układzie chronologicznym oraz procesów normalizacyjnych w języku, pisał Włodzimierz Gruszczyński (2020).

Metody kwantytatywne w światowym językoznawstwie porównawczym

Kiedy językoznawstwo polonistyczne wciąż cierpi na brak korpusu diachronicznego polszczyzny, w światowym językoznawstwie porównawczym prowadzone są projekty cyfrowe z wykorzystaniem metod kwantytatywnych, które mają pierwszorzędne znaczenie dla zdefiniowania tożsamości językowo-kulturowej wspólnot komunikatywnych.

W latach 2004–2008 zrealizowany został Loanword Typology Project (LWT, Max Planck Institute for Evolutionary Anthropology), którego efektem jest World Loanword Database: wold.clld.org/ (Haspelmath, Tadmor, eds., 2009). W projekcie wykorzystano dotychczasowe osiągnięcia językoznawstwa historyczno-porównawczego i typologicznego, których celem jest m.in. ustalanie pokrewieństwa języków oraz wizualizacja ich relacji genealogicznych. Wspierano się także metodami językoznawstwa komputerowego i kwantytatywnego. Dla potrzeb projektu zbadano zapożyczenia w 41 językach świata (wizualizacja różnorodności genealogicznej, geograficznej, typologicznej i socjolingwistycznej). Głównym celem prowadzonych prac stało się zbadanie stopnia zapożyczalności leksykalnej w perspektywie porównawczej,

przez identyfikację znaczeń i grup znaczeń, które są odporne na zapożyczenia. Aby zrealizować to zadanie, przeanalizowano 1460 znaczeń leksykalnych tłumaczonych na każdy język, czego efektem stała się Lista Lipsk–Dżakarta 100 słów (*100-Item List of Basic Vocabulary, Leipzig–Jakarta List*) o najmniejszym stopniu zapożyczalności, określanych jako tzw. słownictwo podstawowe.

Badanie zapożyczeń i zapożyczalności w językach świata daje możliwości śledzenia kontaktów językowo-kulturowych, ale oprócz tego stanowi narzędzie do odróżniania jednostek rodzimych od elementów obcych, wykorzystywane do typologii języków:

Przewidywanie przebiegu zapożyczania jest w językoznawstwie historycznym i porównawczym ważne z wielu powodów. Przede wszystkim zapożyczenie okazuje się często czynnikiem mylącym w ocenie genealogicznego powiązania języków. Jeśli na podstawie znaczeń jesteśmy w stanie określić, które słowa z mniejszym lub większym prawdopodobieństwem mogą zostać zapożyczone, łatwiej nam będzie odróżnić podobieństwa o charakterze dyfuzyjnym od podobieństw wynikających ze wspólnego pochodzenia (Tadmor, Haspelmath, Taylor, 2010: 226; przekład – red.; podkreślenia – autor).

Baza stanowiąca efekt projektu (World Loanword Database) pokazuje czynniki stabilizacji znaczeń i odpowiadających im wyrażen (typ morfemu, część mowy, frekwencja użycia, intensywność kontaktu językowego, zgodność strukturalna języka dawcy i biorcy, bliskość genealogiczna). Dzieli je także na pola leksykalno-semantyczne o mniejszym lub większym stopniu zapożyczalności⁴. Baza daje tym samym kognitywny ogląd zjawisk językowych warunkowanych z jednej strony

⁴ Por. pola o najniższym stopniu zapożyczalności: Percepcja zmysłowa [*Sense perception*], Relacje przestrzenne [*Spatial relations*], Ciało [*The body*] oraz pola o zapożyczalności najwyższej: Religia i wierzenia [*Religion and belief*], Odzież i pielęgnacja [*Clothing and grooming*], Dom [*The house*] (Tadmor, Haspelmath, Taylor, 2010).

naszym (antropocentrycznym) procesem poznawczym, z drugiej – doświadczeniem kulturowym.

Omawiana baza i jej wyniki dały asumpt do prac weryfikujących prezentowane rezultaty badań, z wykorzystaniem metod kwantytatywnych opartych zwłaszcza na bogatych korpusach historycznych języka angielskiego. Weryfikując stopień i czynniki zapożyczalności zarówno w najstarszym angielskim słownictwie religijnym (Timofeeva, 2017, 2018), jak i w zasobie słownictwa podstawowego doby średnioangielskiej (Durkin, 2018), badacze stawiali pytania o regularność zmiany reprezentowanej na poziomie językowym (leksykalnym i pojęciowym). Poszukiwali tym samym (językowych) wykładników własnej tożsamości, śledząc drogi komunikacji w pytaniu: skąd przyszliśmy i dokąd zmierzamy? Prowadzone metodami językoznawstwa komputerowego i kwantytatywnego badania opierały się na zarchiwizowanych do tego celu danych językowych. Na ich podstawie możliwe stało się ustalenie relacji pomiędzy analizowanymi językami, w tym rozróżnienie ich pokrewieństwa genetycznego od podobieństwa nabytego poprzez kontakty. Loanword Typology Project wskazuje na kierunki wędrówek zapożyczeń, stanowiących świadectwo kontaktów pomiędzy językami i używającymi ich wspólnotami kulturowymi. Umożliwia tym samym odpowiedź na pytanie o językowe znaki własnej tożsamości – co podlega zapożyczeniu, jest zmienne i wspólne, a co stałe i partykularne? Co może stanowić przedmiot wymiany, co decyduje o jej braku? Gdzie wreszcie przebiega granica pomiędzy tym, co wrodzone, uniwersalne, a tym, co nabyte i podlegające fluktuacji?

Mając na uwadze sformułowane na wstępie zagrożenia humanistyki cyfrowej operującej analizami *big data*, w świetle prezentowanych kierunków badań wykorzystujących osiągnięcia nowych technologii możemy żywić nadzieję, że pragnienie Franca Morettiego: „Byłoby wspaniale, gdyby pewnego dnia wielkie zbiory danych przywiodły nas z powrotem do wielkich pytań” (Moretti, 2017: 219) ma szansę na urzeczywistnienie.

Literatura

- Durkin P., 2018, *Exploring the Penetration of Loanwords in the Core Vocabulary of Middle English: Carry as a Test Case*, „English Language and Linguistics”, vol. 22.2, s. 265–282, <https://doi.org/10.1017/S1360674318000138>.
- Górski R.L., Król M., Eder M., 2019, *Zmiana w języku. Studia kwantytatywno-korpusowe*, Instytut Języka Polskiego PAN, Kraków.
- Gruszczyński W., 2020, *Korpusy językowe narzędziem pracy historyka języka*, w: *Pol-szczyzna w dobie cyfryzacji*, red. A. Hącia, K. Kłosińska, P. Zbróg, PAN, Warszawa, s. 255–265.
- Haspelmath M., Tadmor U., eds., 2009, *World Loanword Database*, Max Planck Digital Library Munich, <https://wold.clld.org/> [dostęp: 10.07.2023].
- Król M. i in., 2019, *Projekt Narodowego Korpusu Diachronicznego Polszczyzny*, „Język Polski”, R. 99, z. 1, s. 92–101, <https://doi.org/10.31286/JP.99.1.8>.
- Maryl M., 2017, *Kim są polscy humaniści cyfrowi?*, „Teksty Drugie”, nr 1, s. 286–300, <https://doi.org/10.18318/td.2017.1.24>.
- Moretti F., 2017, *Literatura – zmierzona* (F. Moretti, *Literature, Measured*, „Pamphlets of Stanford Literary Lab”, no. 12, April 2016), przeł. M. Maryl, „Teksty Drugie”, nr 1, s. 210–219, <https://doi.org/10.18318/td.2017.1.17>.
- Nycz R., 2017, *Nowa humanistyka w Polsce: kilka bardzo subiektywnych obserwacji, koniektur, refutacji*, „Teksty Drugie”, nr 1, s. 18–40, <https://doi.org/10.18318/td.2017.1.2>.
- Pałka P., Kwaśnicka-Janowicz A., 2024, *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia)*, wydanie drugie elektroniczne, poprawione i uzupełnione, Towarzystwo Miłośników Języka Polskiego, Kraków, <https://przewodnik.tmjp.pl/> [dostęp: 20.12.2024].
- Pastuch M. i in., 2018, *Digital Humanities in Poland from the Perspective of the Historical Linguist of the Polish Language: Achievements, Needs, Demands*, „Digital Scholarship in the Humanities”, vol. 33 (4), s. 857–873.
- Tadmor U., Haspelmath M., Taylor B., 2010, *Borrowability and the Notion of Basic Vocabulary*, „Diachronica”, vol. 27 (2), s. 226–246.
- Timofeeva O., 2017, *Lexical Loans and Their Diffusion in Old English: of ‘gospels’, ‘martys’, and ‘teachers’*, „Studia Neophilologica”, vol. 89, s. 215–237, <https://doi.org/10.1080/00393274.2017.1297208>.

Timofeeva O., 2018, *Survival and Loss of Old English Religious Vocabulary between 1150 and 1350*, „English Language and Linguistics”, vol. 22.2, s. 225–247, <https://doi.org/10.1017/S1360674318000114>.

Źródła internetowe

CLARIN-PL Platforma leksykalna: <https://lexp.clarin-pl.eu/>.

Cyfrowa bibliografia dialektologii polskiej: <https://cbdp.uj.edu.pl/>.

Digital Humanities Summer Institute: <https://dhsi.org/>.

Electronic Textual Cultures Lab: <https://etcl.uvic.ca/>.

Elektroniczny korpus literatury w języku polskim UJ: <https://dhlab.id.uj.edu.pl/czytanie>.

Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772 r.) [KorBa]: https://korba.edu.pl/query_corpus/.

Implementing New Knowledge Environments Partnership: <https://inke.ca/>.

Jagiellońska Platforma Cyfrowa: <https://labedyt.dhlab.uj.edu.pl/index.html>.

Korpus GT projektu IMPACT (w-1): https://szukajwslownikach.uw.edu.pl/IMPACT_GT_1/.

Korpus GT projektu IMPACT (w-2): https://szukajwslownikach.uw.edu.pl/IMPACT_GT_2/.

Korpus polszczyzny XVI wieku [w opracowaniu]: <https://spxvi.edu.pl/korpus/>.

Korpus tekstów polskich z lat 1830–1918: <https://korpus19.nlp.ipipan.waw.pl/overview>.

Korpus tekstów staropolskich: <https://ijp.pan.pl/publikacje-i-materialy/zasoby/korpus-tekstow-staropolskich/>.

Written corpora of historical Polish: <https://clip.ipipan.waw.pl/LRT>.

World Loanword Database: <https://wold.clld.org/>.

AGATA KWAŚNICKA-JANOWICZ – Associate Professor, Department of Language History and Dialectology, Faculty of Polish Studies, Jagiellonian University, Kraków, Poland / dr hab., prof. UJ, Katedra Historii Języka i Dialektologii, Wydział Polonistyki, Uniwersytet Jagielloński, Kraków, Polska.

Her research interests focus on historical linguistics, the history of the Polish language, lexical semantics, as well as digital humanities and computer and corpus linguistics. She is the author of the publication *Staropolska terminologia bartnicza*

(*na tle porównawczym*) [Old Polish Apicultural Terminology (in a Comparative Perspective)] (Kraków 2018) and co-author of *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia)* [Guide to Electronic Language Resources for Polonists (Dictionaries, Cartographies, Corpora, Compendia)] (co-authored with P. Pałka; second electronic edition, revised and supplemented, Kraków 2024, <https://przewodnik.tmjp.pl/>).

Jej zainteresowania naukowe skupiają się wokół językoznawstwa historycznego, historii języka polskiego, semantyki leksykalnej, jak również humanistyki cyfrowej oraz językoznawstwa komputerowego i korpusowego. Jest autorką publikacji *Staropolska terminologia bartnicza (na tle porównawczym)* (Kraków 2018) oraz współautorką *Przewodnika po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia)* (współautorstwo: P. Pałka; wydanie drugie elektroniczne, poprawione i uzupełnione, Kraków 2024, <https://przewodnik.tmjp.pl/>).

E-mail: agata.kwasnicka-janowicz@uj.edu.pl