

Wiktor J. Darasz

ORCID 0000-0003-4262-0665

Kraków

Od Jana Gutenberga do gutenberga.org – Patrycja Pałka, Agata Kwaśnicka-Janowicz: Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia), Biblioteczka Towarzystwa Miłośników Języka Polskiego nr 28, Kraków 2017, ss. 206.

Punktem przełomowym w dziejach kultury europejskiej było wynalezienie druku przez Johanna Gutenberga w XV wieku. Ten fakt jest niepodważalny. Możliwość mechanicznego powielania tekstu przyczyniła się w ogromnym stopniu do demokratyzacji nauki i legła w XVI stuleciu u podstaw dwóch wielkich fenomenów kulturowych, renesansu i reformacji. Wpłynęła też wydatnie na rozwój języka polskiego. Kolejną rewolucją było po pięciuset latach upowszechnienie dostępu do słowa drukowanego w postaci cyfrowej. Wprowadzenie do użytku komputerów osobistych i przyjaznego, czyli łatwego w obsłudze oprogramowania, oraz stworzenie Internetu dało każdemu możliwość publikowania własnych tekstów. Utworzenie internetowych baz danych otworzyło nowe możliwości zarówno przed osobami prywatnymi, jak też uczonymi. Symbolem zmian stała się Wikipedia, czyli sieciowa encyklopedia, którą – przynajmniej w teorii – każdy może edytować. Ważną rolę odegrał również Projekt Gutenberg, zapoczątkowany w 1971 roku przez Michaela Harta, będący pierwszą ogólnodostępną biblioteką cyfrową. Współczesny badacz dzięki dostępowi do globalnej sieci pokonuje bariery przestrzenne i czasowe, mając dostęp do książek wydanych przed wiekami. Także uczeń szkoły podstawowej lub średniej ma do dyspozycji, zwłaszcza na stronie *wolnelektury.pl*, pełne teksty wymaganych pozycji, do niedawna trudno osiągalnych.

Sytuację, jaka obecnie panuje w Internecie, można określić mianem kłęski urodzaju. Wielość dostępnych baz danych, zwłaszcza początkującym użytkownikom globalnej sieci, sprawia kłopoty. Rodzi się potrzeba usystematyzowania informacji o witrynach oferujących w Internecie, najczęściej bezpłatnie, swoje zbiory. Potrzebie tej wychodzi naprzeciw nowa publikacja przygotowana przez dwie krakowskie autorki, reprezentujące Wydział Filologii Polskiej Uniwersytetu Jagiellońskiego, Patrycję Pałkę i Agatę Kwaśnicką-Janowicz. *Przewodnik po elektronicznych zasobach językowych dla polonistów* jest pozycją adresowaną, jak zaznaczono już w tytule, do filologów polskich, przede wszystkim językoznawców. W podtytule doprecyzowano, że w monografii zostały przedstawione słowniki, kartoteki, korpusy i kompendia. Monografia ukazała się w redagowanej przez Piotra Żmigrodzkiego serii Biblioteczka Towarzystwa Miłośników Języka Polskiego. Z tego powodu można

tę pozycję uważać za swoiste nawiązanie do zrealizowanego przez TMJP pomysłu Stanisława Urbańczyka, jakim było niewielkich rozmiarów wydawnictwo *Słowniki, ich rodzaje i użyteczność*, którego pierwsze wydanie ukazało się ponad pięćdziesiąt lat temu (por. Urbańczyk 1964; też Urbańczyk, Sieradzka-Baziur 2000; Karaś 2011). O potrzebie takich publikacji świadczą nie tylko kolejne, aktualizowane wznowienia, ale i poszerzanie wiadomości o ukazujących się kompendiach typu encyklopedycznego. Przedstawiana tu dwustustronicowa książka, rozpowszechniana zarówno w postaci papierowej, jak i elektronicznej, zawiera informacje o dostępnych w Internecie źródłach, niezbędnych w pracy każdego badacza. Publikację rekomendowali do druku profesorowie Mirosław Bańko i Włodzimierz Gruszczyński.

Zadanie, jakiego się podjęły autorki opracowania, wymagało – ze względu na rozległość chronologiczną materiałów oraz ich różnorodność merytoryczną – nie tylko dobrej znajomości dzieł, ale i wypracowania właściwej strategii ich prezentacji oraz dyscypliny w realizacji projektu.

W intencji autorek „Praca jest przeznaczona dla studentów kierunków humanistycznych, magistrantów, doktorantów, pracowników naukowych, a także miłośników języka polskiego, którzy od podstaw poznają zasoby elektroniczne oraz sposoby i możliwości ich eksploataowania” (s. 14). Z uwagi na tak określonego adresata autorki zdefiniowały we *Wstępie* niezbędne terminy specjalistyczne (takie jak digitalizacja, komputeryzacja, słownik elektroniczny, korpus językowy, znakowanie), przedstawiły własną, autorską klasyfikację zasobów cyfrowych i techniczne aspekty korzystania z nich. W trosce o wysoką wartość merytoryczną rekomendowanych materiałów zdecydowały się na pewne ograniczenia, zawiązujące prezentowane publikacje do tekstów ściśle naukowych: „Nie uwzględniliśmy także żadnych amatorskich opracowań leksykograficznych, portali prowadzonych przez nieprofesjonalistów (osoby niepowiązane z jednostkami naukowymi, niezajmujące się zawodowo językoznawstwem, leksykografią, korpusologią czy szerzej – humanistyką cyfrową) ani amatorskich kompendiów udostępniających zbiory cyfrowe lub podających linki do nich” (s. 14).

Omówiły natomiast (w rozdziale *Techniczne aspekty korzystania z zasobów*, s. 28–30) kwestie praktyczne związane z wykorzystywaniem zdigitalizowanych pozycji. Wychodząc od podstawowej obserwacji, że „możliwości korzystania z najnowszych zasobów cyfrowych są uzależnione od posiadanego systemu operacyjnego, rodzaju przeglądarki internetowej oraz oprogramowania lub aplikacji (wtyczek, czyli pluginów, i innych dodatków do aplikacji)” (s. 28), przypominają najpopularniejsze wyszukiwarki, jak Internet Explorer czy Mozilla Firefox. Trzeba wyraźnie zaznaczyć, że powyższa konstatacja, oczywista dla wszystkich informatyków, dla przeciętnego użytkownika nie jest bynajmniej trywialna. Zazwyczaj dopiero po kilku miesiącach korzystania z Internetu dowiaduje się on, co to jest wtyczka i do czego służy. Informacje są podawane w sposób rzeczowy, na przykład: „Do przeglądania odnośników do części wydań oraz podobizn rękopisów potrzebne jest przede wszystkim zainstalowanie oprogramowania DjVu lub odblokowanie wymaganych wtyczek DjVu na pasku poleceń [...]” (s. 29), albo „Pliki dźwiękowe mogą być prezentowane w formacie mp3 lub mp4, audiowizualne w formacie avi, wmv, mpg, mp4 lub flv. Aby je przeglądać, nie trzeba instalować dodatkowych wtyczek, gdyż do ich

obsługi jest wykorzystywany program Windows Media Player, zwykle instalowany razem z systemem operacyjnym Windows” (s. 29).

Po wprowadzeniu następuje *Część I. Najważniejsze ogólne portale i projekty cyfrowe*. W pracy zastosowano jednolitą numerację rozdziałów, toteż *Część I* zawiera rozdziały trzeci, czwarty i piąty. Na otwierający tę część rozdział trzeci *Biblioteki cyfrowe* (s. 33–37) składają się informacje o podstawowych zbiorach źródeł, to znaczy FBC: Federacja Bibliotek Cyfrowych (s. 33–34), RCIN: Repozytorium Cyfrowe Instytutów Naukowych (s. 34), CBN Polona: Cyfrowa Biblioteka Narodowa Polona (s. 35), PBC: Polonijna Biblioteka Cyfrowa (s. 35–36), KSD BIH UW: Kolekcja Starych Druków Biblioteki Instytutu Historii Uniwersytetu Warszawskiego (s. 36), BC KLF UW: Biblioteka Cyfrowa Katedry Lingwistyki Formalnej Uniwersytetu Warszawskiego (s. 36) i VLLPHD: Virtual Lexicographical Library for Polish Historical Dictionaries (s. 37).

W następnym, czwartym rozdziale *Cyfrowe narzędzia i zasoby językowe* (s. 38–41) zostały przywołane profesjonalne instrumenty nowoczesnej, wspomaganej komputerowo lingwistyki, CLARIN-PL: Common Language Resources and Technology Infrastructure, CLIP IPI PAN: Computational Linguistics in Poland, Language Tools and Resources for Polish i GTJ PWr: Grupa Technologii Językowych G4.19 Politechniki Wrocławskiej.

Rozdział piąty, zaledwie półtorastronicowy (s. 42–43), został poświęcony inicjatywie Google Books (Google Książki). Wydaje się, że w tej kwestii można było napisać nieco więcej. Nieuwzględniony został projekt Google Play, w którym część materiałów jest dostępna bezpłatnie.

W *Części II*, zatytułowanej *Zasoby cyfrowe – rodzaje i charakterystyka* (s. 47–57) autorki przechodzą do bardziej szczegółowego omówienia dzieł dostępnych w formie cyfrowej. Zatem rozdział 6 zawiera *Najdawniejsze zabytki języka polskiego* (s. 47–57), w sumie omówienie piętnastu pozycji, na które się składają materiały ze skanami dostępnymi online (*Kazania świętokrzyskie, Psalterz floriański, Bogurodzica, Biblia Królowej Zofii...*) i offline (*Kazania gnieźnieńskie, Psalterz floriański, Kodeks Działyńskich...*).

Zakres informacji oraz ich redakcyjne uporządkowanie dobrze ilustruje przytoczona poniżej *in extenso* notka o jednym z najszlachetniejszych zabytków polskiego średniowiecza.

6.1.2. Bg: Bogurodzica

(przekaz aa, rps 1619a ii bj)

Forma dostępu: online, <http://jbc.bj.uj.edu.pl/dlibra/doccontent?id=287313>

Wymogi techniczne: pełny dostęp bez konieczności zakładania konta i logowania się, plik w formacie image/x.djvu.

Rodzaj wydania elektronicznego: wtórne wydanie elektroniczne.

Zawartość: skan oryginału rękopisu tekstu Bogurodzicy (ok. 1408).

Możliwości i ograniczenia korzystania z zasobów: brak możliwości przeszukiwania zasobów i kopiowania tekstu.

Inne wydania elektroniczne w wersji online: plik PDF, transkrypcja tekstu: Pieśni:

https://www.ijp-pan.krakow.pl/images/publikacje_elektroniczne/korpus_tekstow_staropolskich/piesni.pdf; transkrypcja wersji A, B, C, D i E pieśni wykonana w pracowni

języka staropolskiego IJP PAN, plik PDF z możliwością ograniczonego, prostego przeszukiwania (według ciągu znaków) i kopiowania tekstu. Do obsługi wymagany program Adobe Reader. Dodatkowa wersja w postaci pliku Xml.

Inne wydania elektroniczne w wersji offline: DVD, skany, transliteracje i transkrypcje tekstu (aa, ab, ac): BZPPŚ: Biblioteka zabytków polskiego piśmiennictwa średniowiecznego, red. W. Twardzik, IJP PAN, Kraków 2006; możliwość prostego przeszukiwania (wynajdywanie ciągu znaków w tekście jak w pliku PDF) oraz kopiowania tekstu.

Najdłuższy rozdział siódmy (*Słowniki*, s. 58–145) podzielony został na podrozdziały zgodnie z zaproponowanymi przez autorki cezurami czasowymi: słowniki dawne (do 1950 roku) oraz współczesne. W pierwszej grupie znalazły się oddzielne podgrupy słowników najdawniejszych (do końca XVIII wieku – 25 pozycji) oraz słowniki dziewiętnastowieczne ogólne i inne do połowy XX w. (15 pozycji). Spośród zdigitalizowanych słowników dawnych omówionych znalazły się tu wszystkie ważne z tego zakresu dzieła, poczynając od wydrukowanego w oficynie Hieronima Wietora *Dictionarius... variarum rerum* Jana Murmeliusza (1526), *Lexicon Latino-Polonicum* Jana Mączyńskiego (1564), po *Nowy dykcjonarz, to jest mownik polsko-niemiecko-francuski* Michała Abrahama Trotza (1764) i *Dokładny słownik polsko-niemiecki. Krytycznie wypracowany* Krzysztofa Celestyna Mrongowiusza (1835). I dalej słowniki polszczyzny ogólnej – Lindego, wileński i warszawski oraz słowniki specjalistyczne.

Na rozdział poświęcony słownikom współczesnym (s. 85–145), rozumianym jako pozycje powstałe w drugiej połowie XX wieku i później, składają się podrozdziały, w których znalazło się omówienie pozycji pogrupowanych zgodnie z obiegową typologią tych wydawnictw, a więc w kolejności przedstawione zostały słowniki ogólnodefinicyjne, poprawnościowe, frazeologiczne, gramatyczne, semantyczne, wyrazów obcych, polszczyzny historycznej, łaciny średniowiecznej, słowniki i atlasy dialektologiczne, słowniki etymologiczne, onomastyczne, języka autorów i polszczyzny potocznej, łącznie prawie 50 pozycji, w tym najnowsze, jak przygotowywany w Instytucie Języka Polskiego Polskiej Akademii Nauk wyłącznie w wersji elektronicznej *Wielki słownik języka polskiego PAN* (WSJP) pod redakcją Piotra Źmigrodzkiego, *Słownik pojęciowy języka staropolskiego* opracowany pod kierunkiem Bożeny Sieradzkiej-Baziur czy *Elektroniczny słownik łaciny średniowiecznej w Polsce*. Oddzielne podrozdziały zawierają dane dotyczące kartotek i korpusów języka polskiego (wraz z wyszukiwarkami), w tym tekstów współczesnych i dawnej polszczyzny, języka mówionego, gwar i odmian językowych. Całość zamyka podrozdział o dostępnych kompendiach elektronicznych (Staropolska.pl w opracowaniu Romana Mazurkiewicza, *Słowniki dawne i współczesne*. Internetowy przewodnik edukacyjny; *Dialekty i gwary polskie*. Kompendium internetowe; *Frazeologia.pl...*).

Odrębny rozdział poświęcony został *Kartotekom* (s. 146–150), w którym autorki zaprezentowały „zdigitalizowane (wtórnie wydane elektronicznie) materiały i fiszki (rękopisy lub maszynopisy), które zostały udostępnione w formie plików graficznych w federacji bibliotek cyfrowych” (s. 146).

Kolejny rozdział dziewiąty (*Korpusy*, s. 151–170) zawiera omówienie korpusów lingwistycznych, czyli obszernych zbiorów tekstów stanowiących nieodzowną podstawę wszelkich prac leksykograficznych, w tym Narodowego Korpusu Języka

Polskiego (NKJP), wykorzystywanego między innymi przez zespół przywołanego już *Wielkiego słownika języka polskiego PAN*. W tej części monografii opisane zostały wyszukiwarki służące do pozyskiwania danych z tego korpusu, Wyszukiwarka IPI PAN (s. 152–154) i Wyszukiwarka PELCRA (s. 154–155). Ponieważ nie każdy miał okazję zetknąć się bezpośrednio z badaniami korpusowymi i ma wiedzę o zawartości oraz roli korpusów w naukach filologicznych, a w słownictwie w szczególności, może warto dopowiedzieć, że korpus danego języka jest zbiorem napisanych w nim tekstów, odpowiednio dużym i zrównoważonym pod względem tematyki, pochodzenia i czasu powstania. Odpowiednia dla korpusu wyszukiwarka pozwala na przeglądanie zasobu pod zadanym kątem, umożliwiając równocześnie ustalenie liczby dokumentów (wraz z ich lokalizacją w tekście źródłowym) i sortowanie rezultatów. Najczęściej wyszukuje się formy wyrazowe w określonym kontekście, którego długość można ustalać według własnych potrzeb. Można też wynajdywać połączenia wyrazowe, przy czym niektóre wyszukiwarki (np. Poliqarp wykorzystywana w NKJP) rozpoznają wyrażenia przyimkowe, połączenia ze spójnikami czy formy czasowników zwrotnych bez konieczności wprowadzania dodatkowych technik zapisu. Badacze struktury słowa doceniają możliwość wyszukiwania dowolnych ciągów znaków tak a fronte, jak i a tergo.

Użyteczność publikacji podnosi *Alfabetyczny indeks zasobów cyfrowych i drukowanych* (s. 181–200), zawierający między innymi rozwinięcie zastosowanych w niej skrótów. Zamykająca przewodnik *Literatura* (s. 201–206), w której znalazło się ponad 60 prac teoretycznych, przy czym większość to publikacje wydane po roku 2000, m.in. Mirosława Bańko, Janusza S. Bienia, Magdaleny Derwojedowej, Krzysztofa Nowaka, Piotra Pęzika, Adama Przepiórkowskiego, Zygmunta Saloniego i Piotra Żmigrodzkiego, dokumentuje nie tylko aktualny stan wiedzy z zakresu przedstawianej problematyki, ale stanowi też odpowiedź dostępnej lektury uzupełniającej.

Autorki zrezygnowały z zamieszczania ilustracji (zrzutów ekranu) witryn internetowych, motywując swoją decyzję dużą zmiennością w czasie ich graficznej formy. Decyzja wydaje się słuszna. Choć skądinąd wiadomo, że uczniowie i studenci nierzadko rozpoznają podręczniki nie po nazwisku autora i tytule, ale po kolorze okładki. Takich odbiorców zmiana grafiki strony może istotnie dezorientować.

Niewątpliwie książka Patrycji Pałki i Agaty Kwaśnickiej-Janowicz jest pozycją cenną i bardzo użyteczną, przede wszystkim jako przewodnik dla każdego badacza – zarówno tego początkującego, jak i doświadczonego. Tempo przemian w technice informatycznej sprawia, że wiedza wyniesiona ze studiów szybko się dezaktualizuje. Dotyczy to zwłaszcza systemów i programów, które niemal co roku pojawiają się w nowej wersji. Także oferta źródeł dla humanistów staje się coraz bogatsza. Należy podkreślić, że sposób publikacji w postaci ogólnodostępnego pliku PDF na stronie Towarzystwa Miłośników Języka Polskiego jest przyjaznym gestem wobec studentów, którzy z reguły dysponują niewielkimi funduszami na zakup podręczników. Warto zauważyć również aktualność opracowania, które prezentuje „stan językowych zasobów elektronicznych na dzień 15 marca 2017 r.” (s. 13). Można się zresztą domyślać, że w przyszłości kompendium krakowskich autorek będzie poszerzane i aktualizowane w miarę pojawiania się nowych inicjatyw cyfrowych, tak jak uaktualniona została w III wydaniu (przez Bożenę Sieradzką-Baziur) wspomniana wyżej

książka Stanisława Urbańczyka. Można również postulować zredagowanie podobnych przewodników dla filologów reprezentujących inne dyscypliny, na przykład slawistów czy anglistów. Warto zresztą zauważyć, że wśród słowników historycznych omówionych w książce znalazł się *Orbis Pictus* czeskiego biskupa i pedagoga Jana Amosa Komenskigo, będący pierwowzorem ilustrowanych podręczników i popularnych obecnie słowników obrazkowych.

Bibliografia

- Karaś H., 2011, *Polska leksykografia gwarowa*, Warszawa.
- Urbańczyk S., 1964 (wyd. II – 1967), *Słowniki, ich rodzaje i użyteczność*, Wrocław.
- Urbańczyk S. (część I–II), Sieradzka-Baziur B. (część III), 2000, *Słowniki i encyklopedie, ich rodzaje i użyteczność*, Kraków.
- Żmigrodzki P. i in. (red.), 2018, *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, Kraków.