

SAMUEL BECKETT

DIGITAL MANUSCRIPT PROJECT

[ABOUT](#) ▾[GENETIC EDITIONS](#)[NOTEBOOKS](#)[LIBRARY](#)[BIBLIOGRAPHY](#)[DOCUMENTATION](#) ▾[FREE FEATURES](#) ▾

EBOOKS



MS-OSU-RARE-115

Holding library: **Ohio State University**
Contains draft material of:

Fin de partie / Endgame
En attendant Godot / Waiting for Godot
Molloy



MS-TCD-4662

Holding library: **Trinity College Dublin**
Contains draft material of:

Fin de partie / Endgame
Malone meurt / Malone Dies
Molloy



MS-UoR-1227-7-7-1

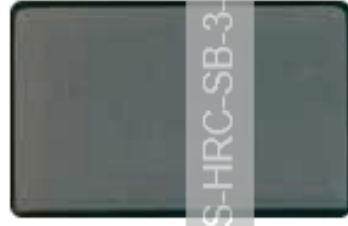
Holding library: **University of Reading**
Contains draft material of:

Fin de partie / Endgame
Krapp's Last Tape / La Dernière Bande



MS-BNF-MY-440

Holding library: **Bibliothèque Nationale de France**



MS-HRC-SB-3-10

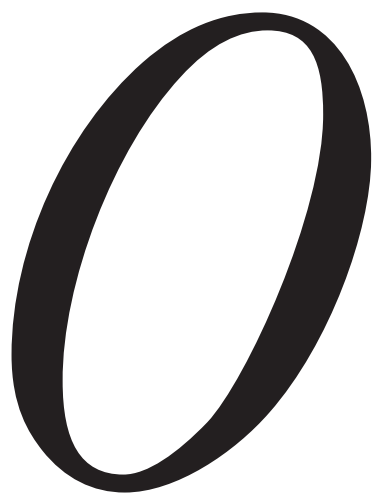
Holding library: **Harry Ransom Center**
Contains draft material of:



MS-HRC-SB-3-5

Holding library: **Harry Ransom Center**
Contains draft material of:

Dirk Van Hulle, Vincent Neyt



projekcie
digitalizacji rękopisów
Samuela Becketta:
*Beckett Digital
Manuscript Project**

Wprowadzenie

Autografy Samuela Becketta, np. *Krapp's Last Tape* (*Ostatnia taśma Krappa*), znajdują się w różnych bibliotekach i kilkunastu archiwach po obu stronach Atlantyku. Z tego powodu staramy się połączyć ze sobą cyfrowo wszystkie rękopisy tego autora. *The Beckett Digital Manuscript Project* (dalej: BDMP) oferuje genetyczne cyfrowe wydania dzieł Becketta wraz z rekonstrukcją jego osobistej biblioteki. Ze względu na ograniczenia związane z prawem autorskim zasoby te wymagają pobierania opłaty abonamentowej od użytkowników indywidualnych i instytucji. W projekcie jest też dostępna „strefa darmowa” (*free zone*), gdzie znajdziemy egzemplarz pokazowy jednej z edycji genetycznych oraz *Beckett Digital Library* (dalej: BDL), wyczerpującą dokumentację oraz kilka zapowiedzi (*sneak peeks*) udostępnianego materiału

w formie miniaturowych zdjęć notatników, statystyk dotyczących liczby dodanych lub usuniętych elementów w obrębie wszystkich genetycznych dokumentów przechowywanych w naszych wydaniach, a od niedawna również pierwszą część dzieła Breona Mitchella *Samuel Beckett: A Bibliography. Part I. The Early Years: 1929–1950*.

Edycje genetyczne

W momencie pisania tego artykułu BDMP prezentuje osiem edycji genetycznych:

- *Molloy*
- *Malone Meurt / Malone Dies*
- *L'Innommable / The Unnamable*
- *En Attendant Godot / Waiting for Godot*
- *Fin de partie / Endgame*
- *Krapp's Last Tape / La dernière bande*
- *Stirrings Still / Soubresauts*
- *Comment dire / what is the word*.

Punktem wyjścia dla każdej z tych edycji jest katalog wszystkich dokumentów wchodzących w skład genezy dzieła, który został ułożony chronologicznie. Materiały pochodzą z archiwów z całego świata, m.in. z University of Reading (Wielka Brytania), Harry Ransom Center w Austin (Texas), Trinity College w Dublinie oraz Bibliothèque nationale de France w Paryżu. Moduł *Fin de partie / Endgame* zawiera np. dokumenty z pięciu różnych instytucji: z University of Reading, Trinity College, Harry Ransom Center, Ohio State University i Harvard University.

Przy każdym z dokumentów podano nazwę, opis katalogowy z archiwum oraz miniaturę okładki lub zdjęcie pierwszej strony. Nawet bez wykupienia prenumeraty można przeglądać katalog edycji genetycznych Becketta. Materiał na osi czasu uporządkowano według dat powstania rękopisów, co pozwala dostrzec relacje między ich wersjami.

Przeglądanie

Każdy dokument został zeskanowany w wysokiej rozdzielczości i przepisany do XML-u zgodnie z wytycznymi P5 Text Encoding Initiative (dalej: TEI). Materiały można przeglądać na kilka sposobów. Zakładka *About* oferuje więcej metadanych, miniatury każdej strony dzieła i linki do różnych „widoków” BDMP: widoku tekstu (*the text view*), widoku

obrazu (*the image view*), widoku obrazu/tekstu (*the image/text view*) i widoku TEI/XML (*the TEI/XML view*).

Widok tekstu prezentuje transkrypcję w formie tekstu bieżącego. W przypadku notatników ich strony *verso* i *recto* są umieszczone obok siebie. Ten widok jest przeznaczony dla użytkowników, którzy chcą poznać treść rękopisów bez odszyfrowywania czasami prawie nieczytelnego pisma Becketta, pismo zostało także zlinearyzowane: dopiski autora na marginesach i na stronie tytułowej wstawiono w odpowiednim miejscu tekstu.

Obrazy można oglądać na wiele sposobów: jeden obraz na każdą stronę, z powiększeniem lub bez niego, albo dwie strony w przypadku zeszytów (*double page view*). Mamy wreszcie widok obraz/tekst, w którym obraz został podzielony na obszary, a transkrypcja może być wyświetlana w dowolnym momencie przez kliknięcie w wybrany obszar.

Porównaj zdania / CollateX

Wszystkie zdania w transkrypcjach zostały otagowane i nadano im numery na podstawie systemu z pierwszego wydania dzieła. Gdy użytkownik znajduje się w widoku obrazu/tekstu lub w widoku tekstu, na pasku nawigacyjnym pojawia się opcja „porównaj zdania” (*compare sentences*). Włączenie tej opcji powoduje, że numery są wyświetlane na początku każdego z nich. Linki prowadzą do „widoku synoptycznego zdań” (*synoptic sentence view*): wszystkie jego wersje są ułożone chronologicznie. Luźne notatki (*paralipomena*), które odnoszą się do danego zdania, ale nie są jego wersjami w ścisłym sensie, również są tu uwzględniane.

W BDMP nie zakodowano aparatu krytycznego i nie uwzględniono różnic między wersjami zdań. Zamiast tego w projekcie wprowadzono oprogramowanie do automatycznego zestawiania zdań CollateX². Zdania są wysyłane do instancji CollateX dzięki jej usłudze REST, a dane wyjściowe z programu są wizualizowane jako tabela wyrównania. Tekst, niezmienny przez wersje, jest umieszczony w kolumnach z zielonym obramowaniem, a różnice zaprezentowano w kolumnach z czarnym obramowaniem i szarym kolorem tła.

CollateX pozwala na przekazywanie dodatkowych informacji, takich jak wskazanie, czy coś jest „usunięciem” lub „dodaniem”, w procesie porównywania, a to umożliwi wizualizację usunięć i dodań w tabeli wyrównania w taki sam sposób, jak w zwykłych transkrypcjach. Tekst usunięty jest przekreślony, a tekst dodawany pojawia się w superskrypcie.

Doodles

W swoich notatkach Beckett zazwyczaj pozostawiał strony *verso* puste i pisał tylko na stronach *recto*. Strony *verso* miały zawierać uzupełnienia i notatki. Na wielu z nich pojawiają się rysunki. W XML-u są one oznaczone elementem <figure> (rysunek). W momencie pisania tego tekstu w korpusie znajduje się 1526 takich rysunków. By je opisać, redakcja BDMP stworzyła „kategoryzację rysunków” (*doodle categorization*). Cztery główne kategorie to: „obiekt” (*object*), „organizm” (*organism*), „kształt” (*shape*) i „symbol” (*symbol*). Można do nich przejść, klikając rysunek w widoku tekstowym i obrazkowo-tekstowym, można je także przeszukiwać i sortować za pomocą wyszukiwarki (patrz niżej: *Wyszukiwarka i sugerowane wyszukiwania*). Ponadto wszystkie zawierają kilka podkategorii. Jeśli istnieje wyraźny związek między rysunkiem a tekstem na sąsiedniej stronie, redaktor dodał opis wyjaśniający to powiązanie.

Wyszukiwarka i sugerowane wyszukiwania

Pole wyszukiwania jest stale widoczne w pasku nawigacyjnym. Użytkownicy mogą dokonywać wyszukiwania pełnotekstowego we wszystkich dokumentach i we wszystkich wydaniach genetycznych oraz w BDL. Przeszukiwaniu podlegają zarówno transkrypcje, jak i informacje redakcyjne. Może być ono również ograniczone do konkretnego dzieła, do transkrypcji lub kategorii rysunków; może być także filtrowane według dzieła i dokumentu.

Aby zwrócić uwagę na niektóre ciekawe zjawiska w tekstach, stworzyliśmy listę sugerowanych wyszukiwań. Dzięki temu użytkownicy mogą odnaleźć „rysunki i diagramy” (*doodles and diagrams*), a także udoskonalać wyniki wyszukiwania, wybierając poszczególne kategorie rysunków. Jako że ich liczba stale rośnie, funkcja ta umożliwia odnalezienie danego rysunku na podstawie tego, co on przedstawia, jego kształtu lub kombinacji kształtów. Inne sugerowane wyszukiwania skupiają się na „obliczeniach” (*calculations*), „adresach” (*addresses*) i „numerach telefonów” (*phone numbers*), „rysunkach scenicznych” (*stage drawings*) oraz „partyturach muzycznych” (*musical scores*). W transkrypcjach oznaczyliśmy „odniesienia intertekstualne” (*intertextual references*), mogą być one również wywołane razem dzięki menu sugerowanych wyszukiwań. Ponieważ Beckett często wstawiał datę na

początku sesji pisarskiej, menu pozwala też na wyszukiwanie tych dat we wszystkich dokumentach. Ostatnią opcją na liście są „luki” (*gaps*): fragmenty, w których Beckett pozostawił puste miejsce do późniejszego wypełnienia.

Beckett Digital Library

W 2016 roku włączyliśmy do BDMP *Beckett Digital Library*, cyfrową rekonstrukcję osobistej biblioteki Samuela Becketta opartą na woluminach zachowanych w jego mieszkaniu w Paryżu, w archiwach (takich jak Special Collections of the University of Reading) i kolekcjach prywatnych (James and Elizabeth Knowlson Collection, Anne Atik, Noga Arikha, Terence Killeen...). Obecnie składa się ona z 762 tomów, a także 247 wirtualnych haseł zawierających opisy niezachowanych fizycznych egzemplarzy książek. Zaznaczyliśmy i przepisaliśmy wszystkie ślady lektury, zaznaczone słowa i marginalia. Beckett, pisząc swoje dzieła, często korzystał z czytanych przez siebie książek. Gdy w szkicach jego utworów pojawia się nawiązanie do jakiegoś fragmentu, w opisie bibliograficznym danego tomu jest to zaznaczone „odnośnikiem do rękopisu” (*Manuscript Link*). Link pojawia się też w transkrypcji danego fragmentu w szkicu. Klikalna ikona „L” dostarcza pewnych informacji i zawiera link do BDL.

Po tym ogólnym wprowadzeniu do BDMP przejdziemy teraz do bardziej zaawansowanych funkcji projektu.

Kodowanie TEI. Tekst a dokument

Kiedy w 2011 roku TEI wydało wersję 2.0.0 wytycznych dotyczących kodowania P5, wprowadzono w niej ważny dodatek. Historycznie TEI zawsze był „zorientowany na tekst”, dostarczając elementów do kodowania struktur i cech tekstu. Ponieważ internet okazał się tak dobrze przystosowany do publikowania widoku poszczególnych stron manuskryptów, pojawiła się potrzeba kodowania tekstu „zorientowanego na dokument”. Wersję 2.0.0 P5 uzupełniono o nowe narzędzia do kodowania dokumentu jako serii elementów <powierzchnia> (<surface>), przy czym każdy element <powierzchnia> (<surface>) zawiera elementy <obszar> (<zone>).

BDMP, uruchomione w 2008 roku, było od początku „zorientowane na tekst”. Mimo to zakodowaliśmy w nim osobną funkcję <div> (<division>, <podział>) dla każdej strony, co również sprzyja „zorientowaniu na dokument”, który może być wizualizowany w postaci widoku tekstowego strona

po stronie (*page-per-page text view*) i w formie wizualizacji „obszar do obszaru” (*zone-per-zone visualization*) w widoku obrazu/tekstu. Takie rozwiązanie ma tę zaletę, że wygląda jak podejście dokumentacyjne, ale jest także zorientowane na tekst. Uzasadnieniem dla takiego rozwiązania okazuje się integralność zdań, która jest tu bardzo istotna. Potrzebujemy też „orientacji na tekst”, aby dopilnować, by zdania, które zostały urwane (np. na końcu strony), zostały ponownie złożone w całość; dopiero to umożliwi skorzystanie z opcji, takich jak porównywanie zdań, wyszukiwanie i CollateX.

Distánt reading

BDMP oferuje dwie funkcje, które można zaklasyfikować jako „czytanie na odległość” (*distant reading*). Po pierwsze, w sekcji *Darmowe funkcje (Free Features)* użytkownicy mogą uzyskać dostęp do strony o nazwie *Kolejność pisania „Nienazywalnego” (Writing Sequence of “L’Innommable”)*. Wizualizuje ona to, jak Beckett wypełniał strony dwóch notatników, które wykorzystał do stworzenia pierwszego szkicu powieści *L’Innommable*, zdanie po zdaniu. Używając numerów zdań w BDMP jako punktu wyjścia, strona (1) pokazuje, gdzie każde zdanie zostało zapisane (tj. na której stronie i w którym notatniku), (2) wskazuje, w którym miejscu dane zdanie ostatecznie się znalazło (w opublikowanym tekście), oraz (3) wizualizuje hipotezę edytorską dotyczącą kolejności, w jakiej zdania zostały napisane. Cele te są osiągnięte przez sekwencję pięciu różnych wizualizacji, z których ostatnia – „krok 5” (*step 5*) – oferuje animowaną interpretację kolejności pisania.

Po drugie, przygotowaliśmy statystyki dla każdej z edycji genetycznych. Wykresy kołowe dla dokumentów we wszystkich edycjach pokazują liczbę słów dodanych, usuniętych i zmodyfikowanych. Na przykład pierwszy szkic *Czekając na Godota (En Attendant Godot)* liczy 20972 słowa, które pozostały niezmienione w trakcie pisania szkicu, 1719 słów wykreślonych, 565 słów dodanych i 19 zmodyfikowanych. Oprócz wykresów kołowych dla każdej edycji genetycznej przedstawiamy także dwa zestawienia zawierające: (1) porównanie pierwszych wersji roboczych wszystkich utworów zamieszczonych w BDMP, (2) porównanie łącznej liczby słów w pełnej dokumentacji genetycznej utworów znajdujących się w BDMP.

Udostępniając plik źródłowy TEI/XML dla każdego transkrybowanego dokumentu, chcemy zachęcić innych badaczy do praktykowania technik „czytania na odległość”.

Materiały BDMP mają stosunkowo niewielki rozmiar. Większość technik „czytania na odległość” stosuje się do korpusów liczących setki, jeśli nie tysiące tekstów. Jednak wykorzystywane do tego materiały są zazwyczaj nieoznaczone (jest to czysty tekst) i oparte na nieskorygowanym OCR pochodzącym z projektów związanych z masową digitalizacją. Transkrypcje BDMP mogłyby stanowić ciekawe wyzwanie w tej dziedzinie, ponieważ są bardzo precyzyjne i wzbogacone o dodatkowe informacje odnoszące się do dodanych i usuniętych słów, postaci mówiących w tekście i ich wystąpienia (w przypadku tekstów performatywnych), rozgraniczenia zdań itd.

ModNets

Modernistyczne Sieci (ModNets) to federacja projektów cyfrowych działająca w zakresie studiów literackich i kulturowych nad modernizmem³. Ma ona na celu zgromadzenie „obiektów cyfrowych” (*digital objects*) – obecnie zawiera 84 901 takich recenzowanych „obiektów” zebranych z 69 stron internetowych. „Obiektem cyfrowym” mogą być: obraz, transkrypcja, metadane lub kombinacje tych elementów. Ograniczenia związane z prawem autorskim uniemożliwiają nam jednak przygotowanie wszystkich obrazów i transkrypcji, ale udostępniamy metadane dla każdego dokumentu w wydaniach genetycznych.

Rozpoznawanie tekstu pisanego odręcznie

Ważnym osiągnięciem w dziedzinie humanistyki cyfrowej z ostatnich pięciu lat jest Handwritten Text Recognition (dalej: HTR), czyli rodzaj OCR dla materiałów pisanych odręcznie. Obecnie wiodącym projektem w tej sferze jest Transkribus⁴. Oprogramowanie to pozwala użytkownikom na załadowanie obrazów i uruchomienie narzędzia do rozszyfrowania tekstu. Aby dane wyjściowe były użyteczne, program musi najpierw nauczyć się „czytać” pismo wybranego autora. Algorytm autorski trzeba więc „trenować” przez dostarczenie odpowiedniego materiału: obrazów i odpowiadających im transkrypcji. We współpracy z zespołem Transkribus algorytm Becketta został „przećwiczony” na 400 stronach transkrybowanego tekstu.

W przypadku mało czytelnego charakteru pisma, takiego jak Becketta, wymagana jest duża liczba stron testowych,

aby osiągnąć „współczynnik błędów” (*character error rate*), który musi być wystarczająco niski, by wynik był użyteczny. Odkryliśmy również, że algorytm działa lepiej, jeśli „trenuje się” oddzielny algorytm dla każdego języka: dzisiaj istnieje francuski i angielski algorytm Becketta.

Transkribus oferuje wiele formatów eksportu plików, jednym z nich jest TEI/XML. Możemy sobie wyobrazić, że Transkribus stanie się częścią przepływu pracy (*workflow*) dla przyszłych edycji genetycznych udostępnianych przez BDMP. Edytor przesyła obrazy dokumentu, używa narzędzia HTR do stworzenia pierwszej wersji transkrypcji, wprowadza niezbędne poprawki i dodaje znaczniki inline, takie jak `<add>` i ``, a potem eksportuje je jako TEI/XML. Obecnie zespół BDMP pracuje nad projektem badawczym o nazwie *CATCH 2020: Computer Assisted Transcription of Complex Handwriting* (Uniwersytet w Antwerpii), a naszym celem jest poprawa sposobu, w jaki Transkribus obsługuje supralinearne i marginalne dodatki oraz teksty performatywne. Rozszyfrowanie pisma mogłoby być również usprawnione przez dostarczenie informacji stylometrycznych związanych z autorem i innych wersji tego samego fragmentu tekstu.

Plany na przyszłość

W przyszłości BDMP będzie rozwijał się w dwóch kierunkach: (1) horyzontalnie, ponieważ będzie zawierał coraz więcej modułów, oraz (2) wertykalnie, gdyż będzie rozwijał więcej narzędzi, aby użytkownicy mogli eksplorować go w ramach różnych modułów. Ważne jest jednak, by doskonałość techniczna nie przyćmiła dużego potencjału hermeneutycznego BDMP wykorzystywanego przez literaturoznawców. Temu niebezpieczeństwu przynajmniej częściowo mogą zaradzić towarzyszące BDMP drukowane książki, które dostarczają tego, co Hans Zeller nazwał *Deutung* (interpretacją), do *Befund* (zapisu tekstów) zwizualizowanego w formacie cyfrowym. Komplementarny charakter projektu BDMP (moduły cyfrowe plus drukowane tomy) może przyczynić się do tak potrzebnej synergii między „starymi” a „nowymi” technologiami edycji naukowej.

Tłumaczenie z języka angielskiego
Jacek Wetniak

Weryfikacja merytoryczna tłumaczenia
Anna Mędrzecka-Stefańska
Kajetan Mojsak

Key Words: Samuel Beckett, digital library, *Beckett Digital Manuscript Project*, TEI, Transkribus

Abstract: The authors report on the *Beckett Digital Manuscript Project*. Its purpose is to reunite the manuscripts of Samuel Beckett's works in a digital way, and to facilitate genetic research: the project brings together digital facsimiles of documents that are now preserved in different holding libraries, and adds transcriptions of Beckett's manuscripts, tools for bilingual and genetic version comparison, a search engine, and an analysis of the textual genesis of his works.

.....
¹ Pierwodruk artykułu: D. Van Hulle, V. Neyt, *Developing the Beckett Digital Manuscript Project*, „Wiener Digitale Revue” 2020, Nr. 1: *Tagebuch*, DOI: 10.25365/wdr-01-03-01, <https://journals.univie.ac.at/index.php/wdr/article/view/3122/4069> (dostęp: 22.03.2023).

² Zob. <https://collatex.net>. BDMP korzysta z wersji 1.7.1 (dostęp: 22.03.2023).

³ Zob. <https://www.modnets.org/about/what-is-modnets/> (dostęp: 22.03.2023).

⁴ Zob. <https://transkribus.eu/Transkribus/> (dostęp: 22.03.2023).