

ARTYKULY, KULY, DO KU MEN TY, MANU SKRYP TY

Instytut Badań Literackich PAN, kontakt: agnieszka.szulinska@ibl.waw.pl, ORCID ID: 0000-0001-5778-6006

Pomiędzy światami

Anotacja semantyczna w naukowej edycji cyfrowej

W prezentowanym artykule chcę pokazać i przybliżyć zagadnienia praktyczne związane z naukowym edytorstwem cyfrowym, szczególnie te opierające się na ręcznej anotacji semantycznej w komentarzu edytorskim. Materiałem badawczym będzie korespondencja Mieczysława Grydzewskiego z Haliną i Kazimierzem Wierzyńskimi, pierwsza w pełni gotowa naukowa edycja cyfrowa przygotowana w Instytucie Badań Literackich PAN¹ i jedna z pierwszych ukończonych edycji cyfrowych w kraju. Zanim jednak przejdę do szczegółowej analizy materiału i kategoryzacji wyzwań napotykanych w trakcie anotowania semantycznego, proponuję krótkie wprowadzenie w sytuację (polskiego) edytorstwa cyfrowego i w stosowaną terminologię.

Status edytorstwa cyfrowego w Polsce

Liczba (z)realizowanych naukowych edycji cyfrowych (*digital scholarly editions*, dalej: NEC) polskich tekstów literackich wciąż jest stosunkowo mała, zważywszy na długie tradycje rodzimego edytorstwa naukowego i w porównaniu z podejmowanymi w ostatnich latach większymi projektami edyorskimi przygotowywanymi przede wszystkim w formie książkowej i finansowanymi ze środków ministerialnych (granty)². Na czym dokładnie polega ich cyfrowość?

¹ Równoległe z cyfrową edycją listów Jana Lechonia i Kazimierza Wierzyńskiego. Oba projekty były realizowane w ramach grantu „Skamandrycka triada na emigracji. Edycja listów Jana Lechonia, Kazimierza Wierzyńskiego i Mieczysława Grydzewskiego (publikacja książkowa i cyfrowa)”, nr 1aH 15 0173 83.

² Jako przykłady można podać projekty finansowane ze środków Narodowego Programu Rozwoju Humanistyki, takie jak: „*Dzieła zebrane* Zygmunta Krasińskiego”, 2012–2017, nr 11H 11 008580; „Nowe wydanie krytyczne *Edycja Krytyczna Pism Wszystkich Bolesława Prusa*”, 2012–2018, nr 11H 11 024680 (oraz „Kontynuacja *Edycji Krytycznej Pism Wszystkich Bolesława Prusa*”, 2022–, nr 11H 20 0355 88); „Kontynuacja wydania krytycznego *Dzieł zebranych* Gustawa Herlinga-Grudzińskiego, t. 4–23”, 2015–2021, nr 11H 13 0403 82; „Edycja krytyczna *Dzieł* Stanisława Przybyszewskiego (utwory literackie)”, 2018–2023,

Często w tym kontekście przywołuje się definicję niemieckiego badacza Patricka Sahle'a, autora katalogu naukowych edycji cyfrowych³. Sahle konstatuje, że czynności naukowe podejmowane w trakcie tworzenia edycji cyfrowej prawie nie różnią się od tych realizowanych w edytorstwie tradycyjnym, wymagane jest natomiast podążanie za cyfrowym paradygmatem na poziomie teoretycznym, metodologicznym i praktycznym⁴. Za tą pozornie zawiłą (choć efektowną) frazą kryje się postulat tworzenia edycji od początku naznaczonej refleksją nad cyfrowością, to w tym medium zostanie bowiem opublikowana i do niego właśnie dostosowana. Interpretuję to jako brak przeszkód w tym, by teorie i metodologie znane w tradycyjnym edytorstwie naukowym dotychczas *a priori* odrzucać bez uprzedniego sprawdzenia możliwości ich wykorzystania w przestrzeni cyfrowej, gdyż może okazać się, że znajdują odzwierciedlenie w jej funkcjach i możliwościach. W powstającej naukowej edycji cyfrowej współczesnych dramatów „patronem” teoretycznym jest np. Pierre-Marc de Biasi⁶, którego propozycje metodologiczne w zakresie genetyki tekstów powstały w odniesieniu do konkretnych (drukowanych) studiów przypadku. Praktykę należałoby zaś tutaj powiązać z kompetencjami cyfrowymi wykonawców edycji – czy to ze znajomością budowy użytecznych stron internetowych i programowania, czy z aktywnościami typowymi dla dyscypliny, takimi jak: wzbogacanie danych, transkrypcja za pomocą cyfrowych narzędzi, analiza ilościowa lub tytułowa anotacja semantyczna. Sahle dodatkowo zauważa, że tak przygotowanej NEC nie sposób „przełożyć” na medium drukowane (czyli po prostu jej wydrukować) bez znaczącej utraty funkcjonalności, co ponownie wydaje się jasne i efektowne, jednak w trakcie analizy konkretnych edycji może domagać się dookreślenia i dopowiedzenia, co rozumiemy pod pojęciem „znacząca utrata”. Oczywista pozostaje obserwacja niemieckiego badacza, że jeżeli między edycją cyfrową a drukowaną nie daje się określić różnic w funkcjonalnościach, to wtedy nie mamy do czynienia z wydaniem cyfrowym, a jedynie zdigitalizowanym, choć zasoby zdigitalizowane są bez wątpienia równie potrzebne w dzisiejszej pracy i komunikacji badawczej.

Do NEC polskich tekstów literackich i okołoliterackich w rozumieniu Patricka Sahle'a zaliczyłabym przede wszystkim *Korpus Tekstów i Korespondencji Jana Dantyszka*⁷ oraz *Korespondencję Jana Lechonia i Kazimierza i Haliny Wierzyńskich*⁸. Pewne cechy

nr 0133/NPPH6/H11/85/2018; „Pisma Henryka Sienkiewicza. Edycja krytyczna pierwsza. Z okazji 100. Rocznicy Odzyskania Niepodległości”, 2022–, nr 11H 20 0153 88. Nawet jeśli w granicie przewidziano środki na działania cyfrowe, przygotowanie edycji papierowej nadal jest głównym celem, i to stworzeniu tej wersji zapewne będzie poświęconych najwięcej prac – nie wykluczam jednak w tych wypadkach zaistnienia jakiejś formy „edycji hybrydowej” (pojęcie to wyjaśniam później).

³ A *Catalog of Digital Scholarly Editions*, v.4.085 2020ff, eds. P. Sahle et al., ostatnie zmiany: 17.04.2023. Podobnym projektem jest katalog: G. Franzini, *Catalogue of Digital Editions*, 2012–, <https://doi.org/10.5281/zenodo.1161425> (dostęp: 13.01.2023). Należy jednak zwrócić uwagę, że znajdują się w nim nie tylko edycje naukowe. Niewątpliwie efekty działań polskich edytorów cyfrowych mogą znaleźć się w obu tych wykazach stanowiących swoiste narzędzie badawcze przy zastosowaniu metod ilościowych. Jeśli chodzi o wskazane edycje korespondencji skamandrytów, to trwają właśnie ostatnie prace związane ze zgłoszeniem ich do wyżej wymienionych katalogów. Warto zaznaczyć, że edycja nie musi być ukończona, aby to zrobić.

⁴ P. Sahle, *What Is a Scholarly Digital Edition?*, w: *Digital Scholarly Editing: Theories and Practices*, eds. M. J. Driscoll, E. Pierazzo, Cambridge 2016, <https://doi.org/10.11647/OBP.0095> (dostęp: 13.01.2023).

⁵ J. Kociński (kierownik projektu), *Cyfrowa edycja współczesnego dramatu polskiego*, koncepcja merytoryczna A. Kramkowska-Dąbrowska, koncepcja cyfrowa A. Kramkowska-Dąbrowska, B. Szleszyński, K. Niciński, K. Mojsak i A. Szulińska, tei.nplp.pl.

⁶ P.-M. Biasi, *Genetyka tekstów*, tłum. M. Prussak i F. Kwiatek, Warszawa 2015.

⁷ *Corpus of Ioannes Dantiscus' Texts & Correspondence / Korpus Tekstów i Korespondencji Jana Dantyszka*, red. A. Skolimowska, M. Turska i K. Jasińska-Zdun, dantiscus.al.uw.edu.pl, publikacja w trakcie realizacji, pierwsza publikacja w internecie: 1.07.2010.

⁸ B. Dorosz, B. Szleszyński (kierownicy naukowi edycji), *Korespondencja Jana Lechonia i Kazimierza i Haliny Wierzyńskich. Naukowa edycja cyfrowa*, <https://tei.nplp.pl/documents/korespondencja-jana-lechonia-i-kazimierza-wierzynskiego/>, także: B. Dorosz, B. Szleszyński (kierownicy naukowi edycji), *Korespondencja Mieczysława Grydzewskiego i Kazimierza i Haliny Wierzyńskich. Naukowa edycja cyfrowa*, <https://tei.nplp.pl/documents/korespondencja-mieczyslawa-grydzewskiego-i-kazimierza-wierzynskiego/>.

NEC, mimo nazwy „archiwum”⁹, przejawia też *Cyfrowe Archiwum Romana Ingardena*¹⁰, które jest ważne głównie dlatego, że przywraca i upowszechnia korespondencję filozofa i jego nieznane wcześniej prace naukowe. Zespół merytoryczny opracował również cyfrowe metadane związane z Ingardenem (m.in. nawiązujące do problematyki poruszanej w tekstach), podał dokładną lokalizację materiałów znajdujących się w archiwum tradycyjnym oraz zwrócił uwagę na aspekt przestrzenny korespondencji i wymienił miejsca powiązane z danym tekstem¹¹. Podobne wymogi będzie spełniał zapewne także projekt *Archiwum Filomatów – edycja cyfrowa*¹² realizowany pod kierownictwem Wojciecha Kruszewskiego¹³.

Nadto na platformie TEI Panorama (dawniej: TEI NPLP) prowadzone są zaawansowane prace nad NEC współczesnych dramatów, wczesnych powieści Elizy Orzeszkowej, korespondencji Marii Dąbrowskiej z Anną Kowalską oraz *Samuela Zborowskiego* Juliusza Słowackiego¹⁴. Stopień realizacji prac można śledzić na Tei.nplp.pl – wszystkie edycje cyfrowe znajdują się w otwartym dostępie.

Nikt nie opisał jeszcze dokładnie, dlaczego w Polsce powstaje i realizuje się tak mało edycji cyfrowych¹⁵. Uważam, że dzielenie się doświadczeniem zdobytym w trakcie pracy nad tworzeniem NEC pomoże niektórym osobom i zespołom w podjęciu decyzji o przygotowaniu takiej właśnie edycji. Ponadto tego typu projekty edytorskie w humanistyce cyfrowej wspierają zwrot ku wielojęzyczności badań. Chodzi tutaj o podkreślenie wartości danych i tekstów naukowych powstałych w innych językach niż język angielski, przy czym dopuszczalne jest pisanie we współczesnej *lingua franca* (zresztą rezygnacja z niej nie byłaby korzystna dla świata nauki). Trzeba też apelować o inny sposób ewaluacji badań i dążyć do zmiany podejścia do publikowania w językach rodzimych. NEC przygotowana w języku polskim wspiera te postulaty humanistyczno-cyfrowe¹⁶.

Nikt nie opisał
jeszcze dokładnie,
dlaczego w Polsce
powstaje
i realizuje się
tak mało
edycji cyfrowych

⁹ To zresztą powtarzające się zjawisko nazewnictwa w humanistyce cyfrowej, gdzie projekt z podtytułem „archiwum” czy „biblioteka” może pełnić funkcje edycji krytycznej.

¹⁰ Zob. <http://ingarden.archive.uj.edu.pl/home> (dostęp: 13.01.2023).

¹¹ Zob. tekst Barbary Góry zamieszczony w prezentowanym numerze „Sztuki Edycji”.

¹² Strona informacyjna projektu: https://wiki.kul.pl/lhdb/Archiwum_Filomat%C3%B3w_-_edycja_cyfrowa (dostęp: 13.01.2023).

¹³ Warto zauważyć, że na stronie danego projektu należałoby zaproponować konkretny zapis bibliograficzny związany z cytowaniem NEC. Dobrą praktyką jest tworzenie specjalnych podstron poświęconych temu zagadnieniu lub w sekcji *O projekcie* dodanie zakładki *Jak cytować / Cytowanie etc.* Wcześniej jednak zespół musi podjąć następujące decyzje: 1) jakie dane bibliograficzne powinny być zawarte w opisie bez względu na wybierane potem style cytowania oraz 2) jak cytować cały projekt, a jak poszczególne jego elementy (w przypadku korespondencji: jak cytować cały korpus lub pojedynczy list). Należy jednocześnie pamiętać, że zawsze mogą pojawić się nowe dane, nieodnotowywane dotąd w opisach bibliograficznych, takie jak „edytor cyfrowy” czy numer DOI będący numerem-identyfikatorem dla zasobów cyfrowych.

¹⁴ Z uwagi na dużą liczbę projektów i wielu wykonawców zaangażowanych w prace wszelkie informacje związane z naszą działalnością zespół Nowej Panoramy Literatury Polskiej (do którego należą) publikuje na profilu facebookowym: <https://www.facebook.com/nowa.panorama.literatury.polskiej> (dostęp: 13.01.2023).

¹⁵ Jeśli chodzi o badanie potrzeb użytkowników NEC na arenie międzynarodowej, to można tu przywołać ankiety dotyczące: ogólnych oczekiwań względem projektów NEC tekstowych (G. Franzini, M. Terras, S. Mahony, *Digital Editions of Text: Surveying User Requirements in the Digital Humanities*, „Journal on Computing and Cultural Heritage” 2019, Vol. 12, No. 1, 13 February 2019, s. 1:1–1:23, <https://doi.org/10.1145/3230671>, dostęp: 13.01.2023) oraz koncepcji otwartości i dostępności w NEC (M. Martinez et al., *Refining Our Conceptions of “Access” in Digital Scholarly Editing: Reflections on a Qualitative Survey on Inclusive Design and Dissemination. Variants*, „The Journal of the European Society for Textual Scholarship” 2019, No. 14, 20 March 2019, s. 41–74, <https://doi.org/10.4000/variants.1070>, dostęp: 13.01.2023). Niewątpliwie jednak przydatne byłoby zbadanie potrzeb edytorów pracujących na konkretnej grupie tekstów, np. napisanych w języku polskim czy powieści dziewiętnastowiecznych.

¹⁶ Więcej o wielojęzyczności i bibliodiversity w komunikacji naukowej w humanistyce i naukach społecznych w: A. Balula et al., *Innovative Models of Bibliodiversity in Scholarly Publications: OPERAS Special Interest Group Multilingualism White Paper*, Zenodo 2021, <https://doi.org/10.5281/zenodo.5653084> (dostęp: 13.01.2023).

Anotacja semantyczna wśród kompetencji humanistyczno-cyfrowych

Poniższe przykłady obrazują wyzwania związane z anotacją semantyczną, jedną z podstawowych aktywności w humanistyce cyfrowej. Według *Taxonomy of Digital Research Activities in the Humanities* (TaDiRAH)¹⁷ opisującej działania badawcze w humanistyce cyfrowej anotowanie jest częścią szerszego procesu „wzbogacania” danych i jest to:

Aktywność polegająca na uwypukleniu informacji o cyfrowym obiekcie przez dodanie komentarzy, metadanych, słów kluczowych, tagów lub linków do cyfrowej wersji bądź do osobnego pliku z anotacjami powiązanego z obiektem cyfrowym. Może być formą anotacji wyjaśniających komentujących albo kontekstualizujących dany fragment, anotacji dodających informacje o charakterze strukturalnym czy lingwistycznym, anotacji w formie otwartych danych powiązanych i tworzących odczytywane maszynowo relacje między obiektami lub – w przypadku ogólnych metadanych – informacji o całym obiekcie¹⁸.

Do powyższej definicji należałoby jeszcze dodać, że anotacja semantyczna będzie łączyła się z innymi aktywnościami z kategorii „interpretacji/kontekstualizacji”, szczególnie semantyfikacji:

Proces identyfikowania i łączenia ze sobą informacji bezpośrednich przez mierzenie się z wyzwaniami niejednoznaczności i formalizacji wiedzy wyrażonej nie wprost służy temu, aby poprawić dostępność, automatyczną analizę danych i ich ponowne wykorzystanie¹⁹.

Biorąc pod uwagę przyjęty model wprowadzania (czy „przekładu”) danych z edycji papierowej do cyfrowej, proponuję określenie „anotacja semantyczna” z zaznaczeniem, że definiowane wyżej działania badawcze nie są przeprowadzane automatycznie (np. za pomocą oprogramowania rozpoznającego nazwy własne, tzw. *Name Entity Recognition*), lecz ręcznie przez edytorów cyfrowych. Zdaję sobie sprawę z licznych zalet zastosowania anotacji automatycznej, przede wszystkim z możliwości opracowania dużych objętościowo korpusów danych w konkretnym czasie, jednak w projekcie dotyczącym korespondencji skamandrytów, który łączy ze sobą dwa światy – papierowy i cyfrowy – anotacja ręczna przynosi wiele ważnych rozpoznań w zakresie strukturyzacji i wzbogacania cyfrowego komentarza edytorskiego. Być może część omawianych tu przykładów i tak musiałaby zostać dodana manualnie, nawet przy założeniu częściowej automatyzacji, ponieważ określenia typu „Żanulka potworna baba” czy „kobiecie, która

Anotacja
ręczna przynosi
wiele ważnych
rozpoznań
w zakresie
strukturyzacji

¹⁷ L. Borek et al., *TaDiRAH: a Case Study in Pragmatic Classification*, „Digital Humanities Quarterly” 2016, Vol. 10, No. 1, <http://digitalhumanities.org/dhq/vol/10/1/000235/000235.html> (dostęp: 13.01.2023).

¹⁸ Tłumaczenie własne hasła *Annotating* z: L. Borek et al., *Taxonomy of Digital Research Activities in the Humanities*, <https://vocabs.dariah.eu/tadirah/en/page/annotating> (dostęp: 13.01.2023).

¹⁹ Tłumaczenie własne hasła *Semantification* z: ibidem, <https://vocabs.dariah.eu/tadirah/en/page/semantification> (dostęp: 13.01.2023).

ma serce z głazu” mogłyby nie być prawidłowo oznaczone przez oprogramowanie jako „Zofia Floyar-Rajchmanowa”²⁰.

Jak już wspomniałam, omawiane przykłady dotyczą korespondencji skamandrytów w dwóch równoległych wydaniach: papierowym i cyfrowym. Dlatego stosuję w przypadku tych listów określenie „edycja hybrydowa”, oznaczające wydanie tych samych tekstów w wersji zarówno drukowanej, jak i cyfrowej z poszanowaniem ich odmienności oraz wynikającą z niego akceptacją różnic między tymi wydaniem przy zachowaniu współpracy zespołów tworzących obie edycje. Każda z realizacji – cyfrowa i papierowa – to inny świat, wymagający innego podejścia i trybu pracy, odmiennych kompetencji, mający też własne cele – choć nadal naukowe – takie jak zwiększenie przetwarzalności danych za pomocą innych narzędzi cyfrowych po zakończeniu projektu. W konsekwencji to świat potrzebujący dostosowanych doń metod ewaluacji i rozpowszechniania²¹.

Opisywane podejście może być zaskakujące dla zwolenników projektów typu „edycja pierwotnie cyfrowa”²². Jednak warto zdać sobie sprawę, że „edycja hybrydowa” w przypadku społeczności edytorskich, w których NEC nie są jeszcze powszechne, jest ciekawą próbą ich wprowadzenia i porównania z wersją papierową, aby uczestnicy tej społeczności mogli lepiej określić swoje potrzeby i zainteresowania badawcze względem cyfrowego krytycznego wydawania tekstów.

Zanim przejdę do zaprezentowania konkretnych przykładów anotacji semantycznej w cyfrowej korespondencji Grydzewskiego z Wierzyńskimi na podstawie komentarza edytorskiego zawartego w przypisach w wydaniu drukowanym, chcę wyjaśnić kilka pojęć.

Bytami w omawianej NEC nazywam jednostki informacyjne przekształcone i przeniesione z komentarza (przypisów) do jednej z pięciu kategorii: *Osoby, Miejsca, Utwory, Czasopisma, Organizacje*²³. Częściowo pełnią one funkcje jednostek indeksowych, zawierają jednak takie cechy i elementy, których nie daje się przenieść do medium drukowanego, jest to np. gromadzenie wszystkich wystąpień i określeń w jednym miejscu, w przypadku *Osób*: autorstwo dokumentów (np. listów), powiązanie semantyczne między bytami (np. autor–dzieło, funkcja testowa edycji). Szczegółowe zasady ich anotacji znajdują się w *Zasadach wydania Korespondencji Mieczysława Grydzewskiego i Kazimierza i Haliny Wierzyńskich. Naukowa Edycja Cyfrowa*²⁴.

²⁰ Określenia zebrane przy hasle indeksowym *Floyar-Rajchmanowa Zofia z domu Malecka primo voto Bagniewska tertio voto Kowalska* na platformie TEI Panorama, <https://tei.nplp.pl/entity/307>.

²¹ Więcej o „edycjach hybrydowych” w: E. Oléron, S. Müller, C. Giannaccini, *Cultural Constellations: Burckhardtsource.org*, „Open Library of Humanities” 2017, Vol. 3, Issue 1, s. 7, <https://doi.org/10.16995/olh.158> (dostęp: 13.01.2023); A.-M. Sichani, *The Business Logic of Digital Scholarly Editing and the Economics of Scholarly Publishing*, w: *Advances in Digital Scholarly Editing: Papers Presented at the Dixit Conferences in the Hague, Cologne, and Antwerp*, 2017, s. 450–451, <https://www.sidestone.com/books/advances-in-digital-scholarly-editing> (dostęp: 13.01.2023); J. Flaišman, M. Kosák, J. Říha, *Hybrid Scholarly Edition and the Visualization of Textual Variants*, w: *ibidem*, s. 290 i 292 (dostęp: 13.01.2023).

²² Tłumaczenie własne angielskiego terminu *digital born* pojawiającego się w tekstach humanistyczno-cyfrowych, a oznaczającego wszelkie efekty pracy badawczej tworzone z myślą o wyłączonej ich publikacji cyfrowej z wykorzystaniem możliwości tego medium. Publikacja PDF-u książki papierowej nie jest więc *digital born*. Przykładem pierwotnie cyfrowego projektu jest np. KORBA – *Elektroniczny Korpus Tekstów Polskich z XVII i XVIII w. (do 1772 r.)*, https://korba.edu.pl/query_corpus/ (dostęp: 13.01.2023) służący do analizowania polszczyzny barokowej i oświeceniowej. Ze względu na jego objętość oraz różnorodność danych publikacja papierowa jest niemożliwa, środowisko cyfrowe pozwala także na zaawansowane badania metodami ilościowymi.

²³ W innych NEC na platformie TEI Panorama rozbudowano byty o dodatkowe kategorie: *Leksykon*, zawierający informacje językowe, np. wyjaśnienia archaizmów, oraz *Wydarzenie*, zbierający dane o wydarzeniach przede wszystkim historycznych, np. „Bitwa Warszawska 1920 roku”. W zależności od założeń edycji możliwe są jednak rozszerzenia zasad włączania danych w tę strukturę – przykładowo w powieści *W klatce* Elizy Orzeszkowej wyobrażam sobie wyodrębnienie bytu typu *Wydarzenie*, „Odpust w miasteczku N.”, który byłby nie odniesieniem do konkretnego wydarzenia historycznego, lecz ważnym fabularnie i badawczo bytem służącym zestawieniu różnic między wielkim miastem a miasteczkiem/wsią.

²⁴ Zob. <https://tei.nplp.pl/categories/1#editorial> (dostęp: 13.01.2023).

Informacje z komentarza edytorskiego są przenoszone do struktury bytów i dzielone na **opisy stałe** i **opisy kontekstowe**. Opisy stałe to informacje podstawowe, „encyklopedyczne” o danym bycie, w przypadku utworu to najczęściej adres bibliograficzny, a w przypadku osób: lata życia, profesja, informacje o rodzinie itd. Opisy kontekstowe najczęściej są powiązane z konkretnym fragmentem danej edycji, w korespondencji skamandrytów może to być ustęp, w którym nadawca listu pisze o okolicznościach śmierci jakiejś osoby, w rzeczywistości jednak pomylił się i edytor(ka) to prostuje. Przykłady rozdzielania tych dwóch typów informacji omówię później.

Ze względu na hybrydowość pracy, jaką przedstawię, poniższe przypadki będą opisem działań przeprowadzonych przez edytora cyfrowego, a nie po prostu przywołaniem fragmentów edycji cyfrowej (co zresztą nie jest w pełni możliwe w medium papierowym, ponieważ NEC nie jest, jak już podkreślałam, zdigitalizowaną formą edycji papierowej).

Zaznaczam również, że pierwsze omówienie wyzwań związanych z „translacją” danych z przypisu w edycji papierowej do struktury cyfrowej w polskich tekstach literackich pojawiło się w artykule Konrada Nicińskiego (także członka zespołu Nowej Panoramy Literatury Polskiej)²⁵. Moja propozycja zmierza do unaocznienia pracy edytora cyfrowego oraz wprowadza klasyfikację zjawisk napotykanych podczas anotacji semantycznej ręcznej (na przykładzie korespondencji Kazimierza i Haliny Wierzyńskich z Mieczysławem Grydzewskim).

Kilka bytów w jednym przypisie. Przypadek „rodziny” Matuszewskich

Opisy stałe
to informacje
podstawowe

W liście z 24 września 1945 roku Kazimierz Wierzyński pisze do Mieczysława Grydzewskiego:

Ignac powiedział mi, że jest teraz zajęty, że nie może podjąć się niczego pośpiesznego. Miał on list od Rayskiej, która przyjechała do Kairo, że córka ich Ewa zginęła w powst[aniu].

W wydaniu papierowym tej korespondencji (s. 125) po słowie „Ewa” zamieszczono przypis 14 o następującej treści:

EWA MATUSZEWSKA, pseud. Mewa (1919–1944). Absolwentka Gimnazjum i Liceum im. J. Słowackiego w Warszawie, w 1938 r. rozpoczęła studia medyczne na UW, była łączniczką w batalionie „Parasol”, poległa w powstaniu warszawskim 26 września w Al. Niepodległości 217. Była córką Ignacego Matuszewskiego i STANISŁAWY KUSZELEWSKIEJ, 1° voto Matuszewskiej, 2° voto Rayskiej (1894–1966), działaczki społecznej, tłumaczki, pisarki (zob. PSB, t. 16 oraz WPPiBL, t. 4). Kuszelewska była żoną Matuszewskiego od 1917 r., po unieważnieniu ich małżeństwa poślubiła w 1928 r. gen. LUDOMIŁA RAYSKIEGO (1892–1977; zob. PSB, t. 30). W 1945 r. opuściła nielegalnie Polskę i wyjechała do Anglii, gdzie przebywał jej drugi mąż. Publikowała artykuły w prasie emigracyjnej.

²⁵K. Niciński, *Obecność przypisu w edycji cyfrowej: rekonesans*, „Napis” 2019, t. 25, s. 196–208, https://rcin.org.pl/Content/112785/WA248_134022_P-I-2795_nicinski-obecnosc_o.pdf (dostęp: 13.01.2023).

Zob. W. Pełczyńska, *Wspomnienie o Stanisławie Kuszelewskiej*, „Wiadomości” 1966, nr 10 (1040) z 6 marca; E.M. Gieniusz, *Ewa Matuszewska „Mewa”*, Warszawa 2021.

Przypis zawiera znacznie więcej danych – niż o samej Ewie – dotyczących w istocie dwóch zacytowanych na początku zdań. Odnosząc go do struktury NEC, znajdują się w nim informacje związane z trzema bytami osobowymi: *Matuszewska Ewa*, *Kuszelewska Stanisława* (matka Ewy Matuszewskiej) oraz *Ludomił Rayski* (drugi mąż Stanisławy Kuszelewskiej). „Ignac” to z kolei Ignacy Matuszewski, pierwszy mąż Stanisławy i ojciec Ewy. Obie kobiety pojawiają się w omawianej korespondencji po raz pierwszy, stąd edytorka przytoczyła informacje podstawowe – trafiające w strukturze TEI Panorama do pola „stałego” opisu bytów. Warto jednak zwrócić uwagę, że fragment „poległa w powstaniu warszawskim 26 września w Al. Niepodległości 217” ma w sobie cechy kontekstowe w odniesieniu do omawianego listu, w którym Wierzyński informuje „Grydza” o śmierci Ewy Matuszewskiej. Niemniej informacje o okolicznościach śmierci są często przedstawiane w różnego rodzaju opisach encyklopedycznych, danych podstawowych²⁶, stąd decyzja edytorów cyfrowych o niewłączeniu tego fragmentu do części kontekstowej. Nie można jednak wykluczyć (ani z całą stanowczością uznać tego za błędne), że inny zespół edytorski uwzględniłby te dane tylko w części kontekstowej. Interpretacji nie oddzielimy tu od edytorskiego opracowania tekstu.

Przypis w NEC został podzielony na następujące byty osobowe.

1. Kuszelewska Stanisława:

1° voto Matuszewska, 2° voto Rayska (1894–1966), działaczka społeczna, tłumaczka, pisarka (zob. PSB, t. 16 oraz WPPiBL, t. 4). Kuszelewska była żoną Matuszewskiego od 1917 r., po unieważnieniu ich małżeństwa poślubiła w 1928 r. generała Ludomiła Rayskiego (1892–1977; zob. biogram: PSB, t. 30). W 1945 r. opuściła nielegalnie Polskę i wyjechała do Anglii, gdzie przebywał jej drugi mąż. Publikowała artykuły w prasie emigracyjnej. Zob. W. Pełczyńska, *Wspomnienie o Stanisławie Kuszelewskiej*, „Wiadomości” 1966, nr 10 (1040) z 6 marca; E.M. Gieniusz, *Ewa Matuszewska „Mewa”*, Warszawa 2021 [w miejscu: „Rayskiej” – A. Sz.].

2. Matuszewska Ewa:

Ps. Mewa (1919–1944). Absolwentka Gimnazjum i Liceum im. J. Słowackiego w Warszawie, w 1938 r. rozpoczęła studia medyczne na Uniwersytecie Warszawskim, była łączniczką w batalionie „Parasol”, poległa w powstaniu warszawskim 26 września w Al. Niepodległości 217. Była córką Ignacego Matuszewskiego i Stanisławy z d. Kuszelewskiej, primo voto Matuszewskiej, secundo voto Rayskiej (1894–1966)²⁷, działaczki społecznej, tłumaczki, pisarki. Zob. E.M. Gieniusz, *Ewa Matuszewska „Mewa”*, Warszawa 2021 [w miejscu „Ewa” – A. Sz.].

²⁶ W serwisie Wikidata (bezpłatnej i otwartej bazie danych zawierającej informacje o charakterze „encyklopedycznym” odczytywane zarówno przez ludzi, jak i maszyny [*machine-readable*]) pojawiają się pola, poza datą i miejscem, sposobu i przyczyną śmierci – na przykładzie danych dla Jan Lechoń: Wikidata contributors, „Q1399280”, Wikidata, <https://www.wikidata.org/w/index.php?title=Q1399280&oldid=1750834485> (dostęp: 8.11.2022).

²⁷ W tym miejscu zrezygnowałam z dodatku (zob. PSB, t. 16 oraz WPPiBL, t. 4) w biogramie Ewy Matuszewskiej, który został umieszczony w biogramie jej matki – choć pierwotnie edytor cyfrowy tę informację zachował. Pozostawienie tego fragmentu w biogramie Ewy mogłoby sugerować, że to jej dotyczą uwagi zawarte od „zob.”

Nieprzywołany bezpośrednio w tym liście Ludomił Rayski otrzymał zaś w swoim stałym opisie bytu dopisek na początku: „(1892–1977; zob. biogram: PSB, t. 30)”.

Dodatkowo w pierwszym zdaniu „Ignac” połączono z bytem *Matuszewski Ignacy*. Można zastanowić się, dlaczego jego „opis stały” bytu nie zawiera informacji o żonie i córce i ewentualnie które fragmenty z przypisu 14 należałoby dodać. Jako że edytorzy merytoryczni wydania papierowego nie zamieścili tych danych w jego biogramie (s. 58), to edytorzy cyfrowi również nie zdecydowali się zmieniać tego elementu struktury. Powiązania osobowe oddano jednak za pomocą testowej funkcji linkowania między bytami – w ten sposób byty *Matuszewski Ignacy*, *Kuszelewska Stanisława* i *Matuszewska Ewa* zostały ze sobą połączone.



Matuszewski Ignacy

(1881–1946), polityk, publicysta, redaktor (zob. PSB, t. 20). Odsunięty od służby państwowej przez środowisko gen. Władysława Sikorskiego, po kapitulacji Francji w czerwcu 1940 r. opuścił Francję i przez Hiszpanię, Portugalię i Brazylię dotarł w 1941 r. do Stanów Zjednoczonych. Pozostał przeciwnikiem polityki Sikorskiego wobec ZSRS, będącej następstwem układu Sikorski–Majski, którą konsekwentnie zwalczał w swojej publicystyce. W 1942 r. wraz z Wacławem Jędrzejewiczem i Henrykiem Floyarem-Rajchmanem współtworzył Komitet Narodowy Amerykanów Polskiego Pochodzenia, a w 1943 r. należał do twórców Instytutu Józefa Piłsudskiego w Ameryce. Odgrywał ważną rolę organizatora polskiej opinii publicznej w Stanach Zjednoczonych przeciw polityce ustępstw wobec Stalina i sterowanych przez niego rządów komunistów w Polsce. Zmarł nagle w Nowym Jorku 3 sierpnia 1946 r. Zob. I. Matuszewski, *Wybór pism*, New York 1952; tenże, *Nie ma wolności bez wielkości. Pisma wybrane*, t. 1: 1912–1942, wpraw., wyb. i oprac. S. Cenckiewicz, Warszawa 2019 oraz tenże, *O Polskę całą, wielką i wolną. Pisma wybrane*, t. 2: 1943–1946, wpraw., wyb. i oprac. S. Cenckiewicz, Warszawa 2019.

Linki:

- Kuszelewska Stanisława (żona)
- Matuszewska Ewa (córka)

Il. 1. Fragment bytu osobowego *Ignacy Matuszewski*, <https://tei.nplp.pl/entity/751>

Powiązania osobowe oddano za pomocą testowej funkcji linkowania między bytami

Granice anotacji. Przypadek „studni opieszałości”

W omawianej korespondencji wyznaczenie granic anotacji, tj. podjęcie decyzji o tym, ile w tekście zaznaczyć jako odniesienie się do poszczególnego bytu, pozostawiono edytorom cyfrowym. To, zdawałoby się, dość liberalne rozwiązanie ma swoje uzasadnienie w założeniu, aby dostarczać użytkownikom-czytelnikom jak najwięcej informacji kontekstowych także na poziomie cyfrowego indeksu. W praktyce oznacza to wprowadzenie „szablonów użycia” na karcie poszczególnego bytu w indeksie. Szablony są tworzone na podstawie anotacji ręcznych, a ich forma jest związana z tym, ile tekstu edytor cyfrowy oznaczył w listach i przypisał konkretnemu bytowi.

Poniżej prezentuję zebrane szablony użycia dla bytu osobowego *Kossowska Stefania*, z d. *Szurlej*:

Szablony użycia		
Stefa	Stefe	Kossowska
Kossowskiej	Stefy Kossowskiej	Stefy Szurlej
Twoja przyjaciółka	Stefa K.	Stefcią Szurlej
Stefe Kossowskiej	Stefa Kossowska	Mój Cook
Big-Benowi	Ben	Pani Stefe
Stefy K.	Big Bena	Stefę
Big Ben	Stefci Kos.[sowskiej]	jest ona Twoją przyjaciółką

Il. 2. Szablony użycia dla bytu osobowego *Kossowska Stefania*, <https://tei.nplp.pl/entity/1755>

W tym samym liście (z 24 września 1945 roku) toczą się rozmowy na temat przyjazdu Mieczysława Grydzewskiego do USA i ewentualnej pracy dla niego²⁸.

Teraz co do przyjazdu. Gotowego tu nie ma nic, ale może być wiele, jeśli nie wszystko. Co do tego różnię się bardzo z Leszkiem i nie przyznaję mu racji. Może być „Tyg[odnik]” – o czym nie chcę pisać i co odkładam do osobistej rozmowy. [...] – Dalej – może być „N[owy] Św[iat]”, choć to niewiarogodna studnia opieszałości, niedbalstwa, arogancji, chamstwa i czego jeszcze chcesz. Mogą być organizacje polityczne lub prace specjalne dla nich (vide Henius i Ig[nacy] – bez których nic by tu nie było, ani Kongresu, ani też jakiegos innego wystąpienia).

Omawianym przykładem będzie tutaj „Nowy Świat”. W edycji papierowej nie ma w tym miejscu przypisu, lecz w edycji cyfrowej oznaczono wystąpienie w tym zdaniu bytu czasopisma „Nowy Świat”. Zaznaczenie jego wystąpienia w liście spowoduje połączenie z bytem mającym opis stały, który w edycji papierowej pojawia się tylko w wypowiedzi z 7 września 1945 roku jako przypis drugi. Ale edytor cyfrowy musi zdecydować o zaznaczeniu miejsca wystąpienia bytu. Można byłoby anotować wyłącznie „N[owy] Św[iat]”, czyli bezpośrednio, skrótowe przywołanie nazwy czasopisma. Dodatkowo wybór frazy „choć to niewiarogodna studnia opieszałości, niedbalstwa, arogancji, chamstwa i czego jeszcze chcesz” zwiększa kontekstowość danych w cyfrowym indeksie i pokazuje nastawienie autora listu do periodyku. Jak wspomniałam, przyjęcie dość szerokich granic anotacji ręcznej i dodatkowe analizy niektórych przypadków mogą spowalniać tego typu prace (wskutek nieustalenia sztywnych reguł anotacji) i powodować pozorny brak spójności, jednak za cenę bogatszych danych na wielu poziomach struktury NEC (tj. dokumentu-listu i cyfrowego indeksu).

Miejsce danych kontekstowych

W liście z 24 września 1945 roku Wierzyński opisuje swoje przemyślenia związane z objęciem przez Jana Lechonia redakcji „Tygodnika Polskiego”:

Może być „Tyg[odnik]” – o czym nie chcę pisać i co odkładam do osobistej rozmowy. Leszek dokonał iście cyrkowego popisu, a, jeśli wolisz, to po prostu cudu, że

²⁸ Ponieważ list z 15 września 1945 roku od Grydzewskiego do Wierzyńskiego nie zachował się, wnioskuję o tym na podstawie odpowiedzi z 24 września tego roku.

przetrzywał okres zdawałoby się beznadziejny. Niemniej jego gospodarka w piśmie jest fantastyczna i jeśli nie zostanie opanowana, pismo znajdzie się znów nad przepaścią. Ratunek jest: p. Obierek jako administrator (już działa) i p. Grydzewski w redakcji – oto moja opinia.

Interesujące nas tu dane kontekstowe stanowiące wyzwanie odnoszą się do Jana Lechonia, ale znajdują się w przypisie... dotyczącym Leopolda J. Obierka. W wydaniu papierowym edytorka po słowach „Ratunek jest: p. Obierek” dodała przypis o następującej treści:

LEOPOLD J. OBIEREK (1902–1986), doktor praw, działacz społeczny, przed II wojną światową dyrektor Funduszu Pracy w Województwie Śląskim. 1 sierpnia 1945 r. objął obowiązki administratora „Tygodnika Polskiego”; od numeru 32 z 19 sierpnia 1945 r. Lechoń występował jako jedyny wydawca i redaktor („publisher and editor”) pisma.

Większość informacji zawartych w tym przypisie edytor cyfrowy przeniósł do części stałej bytu *Obierek Leopold J.*²⁹ Wątpliwości budzi ostatnie zdanie odnoszące się do bycia redaktorem i wydawcą „Tygodnika Polskiego” przez inną osobę. Moim zdaniem jest to informacja kontekstowa dla sytuacji, o której pisze Kazimierz Wierzyński, tj. działań Lechonia mogących doprowadzić do kolejnej trudnej sytuacji finansowej czasopisma. Biorąc to pod uwagę, ostatnie zdanie może zostać kontekstowo dodane do jednego z bytów:

1. „Tygodnik Polski” w zdaniu „Może być »**Tyg[odnik]**«...” – ponieważ dotyczy to bycia wydawcą i redaktorem tego czasopisma.
2. Lechoń Jan w zdaniu „**Leszek** dokonał iście cyrkowego popisu...” – gdyż wiąże się to z działaniami Lechonia jako jedynego wydawcy i redaktora.

Edytor cyfrowy podjął jeszcze inną decyzję – utworzył przypis w zdaniu „Ratunek jest: p. Obierek jako administrator (już działa) i p. Grydzewski w redakcji – oto moja opinia” po wyrażeniu „(już działa)”. Nie jest to jednak decyzja zasadna z punktu widzenia użytkownika, ponieważ nie ma tu już bezpośrednio nic o działaniach Lechonia jako redaktora/wydawcy. Stąd jako osoba nadzorująca anotację semantyczną na TEI Panorama (*TEI Encoding Supervisor*) zdecydowałam po konsultacji z edytorem cyfrowym, aby przypisać to zdanie jako opis kontekstowy w miejscu omawianym w punkcie 2. Wydaje się, że właśnie wtedy czytelnik może otrzymać jasną informację o roli Lechonia w „Tygodniku Polskim”, gdyż rozpoczyna się wówczas ocena jego działań wystawiona przez Wierzyńskiego. Nie byłoby jednak błędem przypisanie tych danych do bytu „Tygodnika” w zdaniu wcześniejszym – druga możliwość wydaje się bardziej zasadna z punktu widzenia użyteczności i czytelności w środowisku cyfrowym.

Powyższa propozycja kategoryzacji ma charakter wstępny. W artykule ograniczyłam się do wybranych problemów, tych szczególnie wyrazistych, z korespondencji skamandrytów, choć problem wyznaczania granic anotacji jest problemem uniwersalnym i z pewnością niejedyn edytor cyfrowy będzie musiał się z nim zmagać.

Opisanie anotowania semantycznego w przestrzeni cyfrowej w „tradycyjnej” formie artykułu naukowego nie jest łatwe, wymaga szczegółowych wyjaśnień i może być trudne

²⁹ Zob. <http://tei.nplp.pl/entity/867>.

Problem
wyznaczania
granic anotacji
jest problemem
uniwersalnym

do wizualizacji bez obejrzenia konkretnych przykładów na stronie NEC (do czego zachęcam). Postrzegam to w kategoriach wskazywanej przez Sahle'a nieprzekładalności środowiska cyfrowego na obecne, znane od lat środki komunikacji naukowej – ważne dla większości osób w niej uczestniczących, a przede wszystkim poddawane ewaluacji, co jest nie bez znaczenia. W przyszłości będą potrzebne w tym aspekcie jeszcze inne formy komunikacji, takie jak: szkolenia, warsztaty, hackathony z użyciem cyfrowych narzędzi oraz cyfrowe instrukcje (*workflows*), living booki czy różnego rodzaju raporty. Ale bez wsparcia finansowego i infrastrukturalnego to wszystko jest znacznie utrudnione i, co gorsza, takie działania nie są zazwyczaj liczone do dorobku badawczego. Dla osób chcących związać się z humanistyką cyfrową jest to nadal ogromne wyzwanie, co pośrednio wpływa na sytuację naukowego edytorstwa cyfrowego w Polsce.

Nie można mieć jednak wątpliwości, że w naukowym edytorstwie cyfrowym trzeba wykazać się nie tylko konkretnymi kompetencjami/umiejętnościami, ale i wiedzą z zakresu edytorstwa tradycyjnego. Podane przeze mnie przykłady bytów, które układają się w bazę danych (częściowo przejmującą też funkcje indeksu) i zawierają szablony użycia, będą pomocne zarówno humanistom cyfrowym przetwarzającym dane ilościowo i poszukującym np. gotowych opisów osobowych czy organizacji i instytucji, jak i badaczom zajmującym się językiem i twórczością pisarzy w perspektywie lingwistycznej.

Key Words: digital scholarly edition, semantic annotation, Mieczysław Grydzewski, Kazimierz Wierzyński, Halina Wierzyńska, correspondence, digital humanities

Abstract: The text presents the practical challenges of creating digital scholarly editions of Polish literary correspondence, specifically, manual semantic annotation. The research material consists of letters between Mieczysław Grydzewski and Halina and Kazimierz Wierzyński, compiled by Beata Dorosz and published in paper and digital versions. The analysis and a proposal for a preliminary classification of the most common problems are preceded by an outline of the situation of Polish digital editing in the context of current phenomena in the digital humanities, such as multilingualism, the need for user research or the creation of bibliographic data of digital research results.