

RENATA BRONIKOWSKA

Instytut Języka Polskiego PAN, Kraków

ORCID: 0000-0001-9000-5355

KATARZYNA KRYŃSKA

Instytut Języka Polskiego PAN, Kraków

ORCID: 0000-0003-3446-122X

Łacina w KorBie. Użyteczność Elektronicznego Korpusu Tekstów Polskich XVII i XVIII wieku dla filologa neolatynisty¹

I. Wstęp

W ostatnich latach w środowisku filologów klasycznych i polonistów pojawiają się postulaty stworzenia słownika łaciny używanej w Polsce w okresie od XVI do XVIII w. (Czerenkiewicz, Ryczek, 2013; Czerenkiewicz, 2015). Autorzy tej koncepcji podkreślają odmienną łaciny nowożytnej zarówno od łaciny klasycznej, jak i średniowiecznej. Zgodnie z ich opisem był to język z jednej strony czyszczony z naleciałości wieków średnich, a z drugiej – rozwijający się pod wpływem coraz bardziej ekspansywnej polszczyzny. Na jego postać wpłynęły również przemiany gospodarcze, społeczne i religijne I Rzeczypospolitej, które od ówczesnych pisarzy łacinników wymagały stworzenia lub doprecyzowania pojęć adekwatnych do opisu nowej rzeczywistości.

Potrzeby neolatynistów częściowo zaspokaja powstający w Pracowni Łaciny Średniowiecznej IJP PAN „Słownik łaciny średniowiecznej w Polsce” (Plezia i in., 1953–), którego pierwszych siedem tomów dostępnych jest również w postaci elektronicznej (Elektroniczny Słownik Łaciny Średniowiecznej w Polsce, obejmujący hasła na litery A-Q, <https://lexicon.scriptoros.pl/>), oraz tworzony jako jego uzupełnienie

¹ Artykuł powstał w ramach projektu „Rozbudowa Elektronicznego Korpusu Tekstów Polskich XVII i XVIII Wieku i jego integracja z *Elektronicznym słownikiem języka polskiego XVII i XVIII w.*” realizowanego w IJP PAN pod kierownictwem W. Gruszczyńskiego (nr projektu: 0413/NPRH7/H11/86/2018).

Elektroniczny Korpus Polskiej Łaciny Średniowiecznej (<https://scriptores.pl/efontes/>)². Oba te zasoby obejmują swoim zasięgiem oprócz średniowiecza również początek epoki nowożytnej dzięki wykorzystaniu łacińskich tekstów powstałych na ziemiach polskich do połowy XVI w.³ Zgodnie z koncepcją twórców korpusu rozszerzenie jego zakresu chronologicznego na czas uznawany za początek ery nowożytnej „pozwole zweryfikować proponowane w literaturze przedmiotu daty renesansowego przełomu, tym razem w oparciu o dane empiryczne” (Nowak, 2014, s. 106).

Nie został dotychczas stworzony korpus gromadzący teksty łacińskie powstałe w Polsce po 1550 r., który mógłby stać się podstawą słownika polskiej łaciny nowożytnej. Neolatyniści w swojej pracy muszą więc sięgać po inne pomoce, także te, które w założeniu nie były przeznaczone dla nich. W naszym artykule przedstawiamy jedno z takich pomocniczych narzędzi. Jest nim Elektroniczny Korpus Tekstów Polskich z XVII i XVIII wieku (do 1772 r.), zwany środowiskowo KorBa (akronim od *Korpus Barokowy*) (<https://korba.edu.pl/>), powstający w Pracowni Historii Języka Polskiego XVII i XVIII wieku IJP PAN. Chociaż korpus ten z założenia nie zawiera tekstów napisanych wyłącznie po łacinie, to dzięki dużemu udziałowi słów łacińskich stwarza pole do badań nad specyfiką języka łacińskiego w XVII–XVIII-wiecznej Polsce. W dalszej części artykułu pokażemy, w jaki sposób może on być wykorzystany w badaniach neolatynistycznych.

2. Łacina w KorBie – możliwości i ograniczenia badawcze

Elektroniczny Korpus Tekstów Polskich z XVII i XVIII wieku (do 1772 r.) został stworzony w latach 2013–2018 w ramach grantu NPRH⁴. Od 2019 roku trwają prace nad rozbudową korpusu (polegające m.in. na włączeniu do niego tekstów powstałych w latach 1773–1800) oraz jego integracją z innymi zasobami językowymi polszczyzny obejmującymi okres XVII–XVIII w. (Ogrodniczuk, Gruszczyński, 2019).

Obecnie KorBa liczy blisko 13,5 mln segmentów⁵ pochodzących z ok. 700 źródeł, zróżnicowanych chronologicznie, geograficznie, gatunkowo i tematycznie. W korpusie została zastosowana kilkupoziomowa anotacja, która pozwoliła odwzorować strukturę każdego dokumentu i wyodrębnić w tekście fragmenty obcojęzyczne, a każdemu segmentowi przypisać lemat i wartości właściwych dla niego kategorii gramatycznych. Korpus jest również opatrzony licznymi metadanymi, które umożliwiają filtrowanie

² Dużą pomocą dla neolatynistów w zakresie średniowiecznych terminów prawnych jest również „Słownik łacińsko-polski dla prawników i historyków” (Sondel, 1997).

³ W „Słowniku łaciny średniowiecznej w Polsce” pojawiają się także cytaty ze źródeł powstałych po 1550 r., są one jednak nieliczne.

⁴ Projekt był realizowany w IJP PAN we współpracy z Zespołem Inżynierii Lingwistycznej IPI PAN pod kierownictwem W. Gruszczyńskiego (nr projektu: 0036/NPRH2/H11/81/2012).

⁵ Pojęcie *segment* używane jest w rozumieniu przyjętym przez twórców Narodowego Korpusu Języka Polskiego (NKJP, <http://nkjp.pl/>). Segmentem jest zazwyczaj słowo tekstowe (ciąg liter ograniczony spacjami lub znakami interpunkcyjnymi), ale w niektórych wypadkach słowo może się składać z kilku segmentów, np. *zrobili+by+śmy*.

wyników wyszukiwania. Teksty zgromadzone w KorBie udostępniane są w dwóch postaciach: transliterowanej i transkrybowanej, co zwiększa możliwości ich przeszukiwania. Korpus obsługuje wyszukiwarka MTAS, wykorzystująca język zapytań CQL⁶.

Szczególne znaczenie dla filologów niepolonistów ma oznakowanie fragmentów obcojęzycznych, przy czym znaczniki nie tylko identyfikują segment jako niepolski, ale również przypisują go do konkretnego obcego języka (arabskiego, czeskiego, francuskiego, greckiego, hebrajskiego, hiszpańskiego, litewskiego, łacińskiego, niemieckiego, węgierskiego, włoskiego) lub grupy językowej⁷ (skandynawskiej, turecko-tatarskiej, południowosłowiańskiej i wschodniosłowiańskiej). Jak można się spodziewać, wśród segmentów obcojęzycznych przytłaczającą większość stanowią segmenty łacińskie (494 195 segmentów łacińskich na 514 719 wszystkich segmentów obcojęzycznych, czyli 96%). Jednocześnie stanowią one stosunkowo dużą część całego korpusu (3,7%). Daje to możliwość prowadzenia różnorodnych badań nad obecnością łaciny w polskich tekstach z XVII i XVIII w. Ponieważ jednak KorBa pomyślana była jako korpus języka polskiego, badanie łaciny oraz wzajemnych relacji między językiem polskim a łacińskim napotyka pewne ograniczenia.

Najważniejszym z nich jest to, że podczas przepisywania tekstów były pomijane dłuższe fragmenty łacińskie. Co do zasady do korpusu miały być włączone tylko te fragmenty obcojęzyczne, które były uwikłane składniowo w zdanie polskie albo stanowiły najbliższy kontekst zdania polskiego, a więc m.in.:

- przysłowia: *Albo piy/ albo się biy. Aut bibe, aut abi;*
- cytaty z Biblii: ... *przydáiie Pismo Święte: facta est coniuratio valida. stáło się z-przy-siężenie potężne;*
- cytaty z dzieł łacińskich: *Prawdźi się tu Tacyt: Falso Libertatis vocabulum ostendi ab eis, qvi privatim degeneres, in publicum exitiosi, nihil spei nisi per discordias habeant* [Fałszywie imieniem wolności ci się zaślaniają, którzy osobiście zwyrodniali, dla państwa zgubni, w niczym innym, jak tylko w konfliktach pokładają nadzieję];
- terminy prawne i urzędowe: *Administratorowie niepowinni będą Przykomorkow ustanawiać, w Dźiedzicznych dobrach sine scitu et consensu Haeredum* [bez postanowienia i zgody spadkobierców];
- terminy naukowe: *Dla tegoż z przodku nie wiele grad: vbiegał Kometá/ że in minori latitudine* [na mniejszej szerokości] *był/ kędy iest maior quantitas graduum* [większa wartość stopni], *odpráwiwszy bieg in signo Scorpii* [w znaku Skorpiona] *wstąpiel ná 30. grad: Librae* [stopień Wagi], *dnia niemal 10 Grud.;*
- utarte zwroty: *Synowców jego niesłusznie pobitych trzy głowy po stawie pływając, in praesentia* [w obecności] *żałosnego Stryja gadały;*
- daty i inne określenia czasu: *Anno 1702* [roku 1702] *urodziła się nam córka 8va 7bris* [ósmego września] *w Wilnie, in die Nativitatis B. V* [w dniu Narodzin Najświętszej Maryi Panny].

⁶ Dokładny opis struktury KorBy znajduje się w: Gruszczyński i in., 2020.

⁷ W wypadku słów, które w tekstach z tej epoki trudno jednoznacznie uznać za należące do któregoś ze spokrewnionych ze sobą języków.

W praktyce od tej zasady istnieją pewne odstępstwa. Po pierwsze, nie usunięto dłuższych fragmentów łacińskich z wydań współczesnych oraz z tekstów oryginalnych, które zostały przepisane już wcześniej, w ramach innych projektów⁸. Po drugie, osoby przepisujące teksty (zwane w projekcie skryptorami) nie zawsze umiały się powstrzymać od pominięcia łaciny, stąd również w niektórych tekstach przepisywanych w ramach projektu KorBa można napotkać dłuższe fragmenty łacińskie.

Kolejne ograniczenie związane jest z brakiem lematyzacji i anotacji gramatycznej segmentów łacińskich, przez co zakres wyszukiwania ograniczony jest tylko do form wyrazowych. Nie można także sporządzić listy frekwencyjnej dla całych leksemów.

Pewne kłopoty sprawiało skryptorom (w większości niezającym łaciny na poziomie zaawansowanym) transliterowanie łacińskich abrewiacji. W rezultacie segmenty takie jak *quoq* (*quoq*), będące skrótem słowa *quoque* ‘również’, oddawano zazwyczaj w postaci *quoq* (*quoq*:). W przeciwieństwie do polskich abrewiacji, którym zgodnie z instrukcją przypisywano również pełny odpowiednik (np. *atramētem* <atramentem>), skróty łacińskie pozostawiono nierozwiązane⁹.

Oprócz błędów w transliteracji zdarzają się również błędy w oznakowaniu segmentów znacznikiem łaciny – niektórym wyrazom łacińskim nie został przypisany taki znacznik, a z kolei niektóre wyrazy polskie zostały oznaczone jako łacińskie. W wielu wypadkach zresztą trudno było jednoznacznie określić przynależność danego segmentu do jednego z dwóch języków. Przykładowo w zdaniu: *Cosz to spráwilo? nic inszego záprawdě iedno Obedientia* wyraz *obedientia* może być potraktowany albo jako błędnie zapisany wyraz łaciński (poprawnie: *oboedientia* ‘posłuszeństwo’), albo jako wyraz polski pochodzenia łacińskiego (jak wiadomo, w XVII–XVIII-wiecznej polszczyźnie normy ortograficzne nie były ściśle ustalone, zatem słowo to mogło się pojawić zarówno z łacińskim zakończeniem *-tia*, jak i w spolszczonej postaci, z zakończeniem *-cia*).

Na decyzje podejmowane przez skryptorów nakładały się „decyzje” tagerów¹⁰ dokonujących automatycznej anotacji tekstów. Tagery mogły bowiem przypisać słowu uznanemu przez skryptorów za polskie znacznik języka obcego (ale bez określania jego przynależności do konkretnego języka). Wybory te były w pewnych wypadkach poprawne, np. wszystkie wystąpienia słowa *item*, a więc także te, które zostały potraktowane przez skryptorów jako użycia polskie, otrzymały w interpretacji tagerów znacznik „xxx” oznaczający wyraz obcy. Czasem jednak poprawki tagerów trzeba uznać za błędne. Wyraz *resolutia* przez jeden z tagerów został oznaczony jako obcy, choć spolszczenie

⁸ Chodzi tu przede wszystkim o teksty z projektu IMPACT (Bień, 2014), włączone do KorBy za zgodą kierownika tego projektu, Janusza Bienia.

⁹ W przygotowywanej obecnie rozszerzonej wersji korpusu planowane jest dodanie do abrewiacji łacińskich ich pełnych odpowiedników.

¹⁰ Automatycznej anotacji KorBy dokonywały niezależnie od siebie dwa tagery: Toygger i Concraft. Na skutek tego powstały dwie wersje korpusu, które mogą się od siebie różnić na poziomie lematyzacji i przypisanych poszczególnym formom znaczników morfosyntaktycznych. Domyślnie w wyszukanych wynikach pojawiają się interpretacje Toyggera, zaś interpretacje Concrafta wyświetlają się w dymku zawierającym podstawowe informacje o szukanej formie.

jego postaci (*resolutia* zamiast łacińskiego *resolutio* ‘rozwiązanie’) przemawia za uznaniem go za słowo istniejące w zasobie leksykalnym XVII–XVIII-wiecznej polszczyzny.

Mając w pamięci te ograniczenia, przyjrzyjmy się możliwościom przeszukiwania korpusu pod kątem wystąpień wyrazów łacińskich.

- Wyszukiwanie konkretnych form łacińskich
Najprostszym sposobem korzystania z wyszukiwarki jest wpisanie w jej okno żądanej formy łacińskiej. W wypadku, kiedy forma ta jest homonimiczna z formą polską, do zapytania należy dodać oznaczenie języka łacińskiego (`foreign="la"`). Oto przykład zapytania pozwalającego wyszukać w korpusie wszystkie wystąpienia łacińskiej formy *armata* ‘uzbrojona’: `[orth_lc="armata"&foreign="la"]`¹¹.
- Wyszukiwanie wszystkich segmentów łacińskich w wybranych tekstach
Za pomocą wyrażenia `[foreign="la"]`, nakładającego na segmenty warunek przypisania do języka łacińskiego, można wyszukać wszystkie segmenty łacińskie w korpusie. Ograniczenie wyszukiwania za pomocą rozmaitych parametrów (np. do konkretnego tekstu, do tekstów jednego autora, do tekstów pochodzących z jakiegoś przedziału czasowego) pozwala badać różnice w stopniu nasycenia łaciną między różnymi tekstami.
- Wyszukiwanie kilkuwyrazowych fraz łacińskich
Poprzez określenie w zapytaniu układu segmentów łacińskich i niełacińskich można wyszukać w tekście charakterystyczne dla tekstów średniopolskich frazy łacińskie. Przykładowo zapytanie `[!foreign="la"] [foreign="la"] [foreign="la"] [foreign="la"] [!foreign="la"]` pozwala znaleźć sekwencję trzech segmentów łacińskich otoczonych segmentami polskimi, wśród których część stanowią utarte zwroty (np. *ad absolutum dominium* ‘do władzy absolutnej’) lub znane cytaty, np. *sensim sine sensu* ‘stopniowo i niepostrzeżenie’¹².
- Wyszukiwanie łacińskich konstrukcji gramatycznych
W pewnym ograniczonym zakresie w KorBie można też wyszukiwać łacińskie konstrukcje gramatyczne, np. połączenia przyimka *ad* z formami gerundium w bierniku¹³. Zapytanie `[orth_lc="ad"] [][?][orth=".*(andum|endum)"]` pozwala znaleźć układ segmentów, z których pierwszy stanowi forma *ad*, a ostatni kończy się na *-andum* lub *-endum*¹⁴. Pomiedzy nimi może, ale nie musi, pojawić się dodatkowy segment. W wynikach wyszukiwania znajdują się takie wyrażenia, jak *ad deliberandum* ‘do rozważenia’ czy *ad bene merendum* ‘dla dobrego przysłużenia się’.

¹¹ Język zapytań wyszukiwarki MTAS jest opisany w instrukcji znajdującej się na stronie korpusu: <https://korba.edu.pl/manual>. Przy budowaniu zapytań można też skorzystać z prostej nakładki na wyszukiwarkę (konstruktora zapytań), która ułatwia zdefiniowanie warunków określających cechy segmentu lub sekwencji segmentów występujących w zapytaniu.

¹² Jest to cytat z traktatu Cyserona „O starości”: *sensim sine sensu aetas senescit* ‘stopniowo i niepostrzeżenie nadchodzi starość’ (Cyseron, „De senectute” 11,38). W tekście pochodzącym z korpusu został wpleciony w zdanie: ... *takim sposobem sensim sine sensu giniecie, a przeciw widzieć tego nie chcecie*.

¹³ W gramatyce łacińskiej gerundium użyte w bierniku w połączeniu z przyimkiem *ad* wyraża cel, np. *ad deliberandum* można przetłumaczyć ‘do rozważenia/przemyślenia’, ale też ‘aby rozważyć/przemyśleć’.

¹⁴ W zapytaniu pominięto rzadko występujące trzecie zakończenie gerundium *-undum* ze względu na możliwe błędne wyszukanie form, które gerundium nie są, np. popularne *ad secundum* ‘po drugie’ czy *ad fundum* ‘do głębi, do podstawy’.

W dalszej części artykułu pokażemy przykłady badań nad fragmentami łacińskimi w KorBie z użyciem wyszukiwarki korpusowej. Przykłady te obejmują analizy zarówno konkretnych słów i wyrażeń łacińskich, jak i wzajemnych relacji między segmentami łacińskimi i polskimi. Na koniec zaprezentujemy możliwości analizy danych znajdujących się na liście frekwencyjnej łacińskich form wyrazowych.

3. Badanie charakterystycznych zwrotów łacińskich w polskich tekstach

Jak wspomniano wyżej, wyszukiwarka KorBy pozwala na znalezienie zarówno konkretnej formy łacińskiej, jak i formy o ogólnej charakterystyce ‘segment łaciński’. Poprzez stworzenie zapytania o układ segmentów, z których tylko jeden ma postać ściśle określonej, można m.in. ustalić typową łączliwość przyimków. Przykładowo formuła [orth_lc="in"][foreign="la"] pozwala na znalezienie kontekstów zawierających przyimek *in* z następującym po nim segmentem łacińskim.

Aby określić frekwencję wystąpień poszczególnych wyrażeń, można użyć formuły zapytania dotyczącej konkretnej formy połączonej z *in*, np.: [orth_lc="in"][orth_lc="anno"], gdy chcemy zobaczyć, ile razy w korpusie występuje wyrażenie *in anno* ‘w roku’. Wyrażenie to pojawia się 99 razy, co stanowi blisko 10% wszystkich wystąpień słowa *anno* w korpusie. Wielokrotnie w tekstach występują również połączenia tego przyimka ze skróconą formą *anno – in a., in an.* – łącznie takich wyrażeń przyimkowych jest 170. W ten sposób można sprawdzić liczbę innych wyrażeń przyimkowych z *in*, m.in.: *in toto* ‘w całości’ – 99 wystąpień, *in praesentia* ‘w obecności’ – 91, *in vim* ‘w sile’ – 83, *in casu* ‘w przypadku’ – 77 oraz *in casum* ‘w przypadku’, ‘na wypadek’ – 44 (dużo ponad połowę wszystkich wystąpień słowa *casu* – 224), *in usu* ‘w użyciu’ – 74, *in genere* ‘w ogólności’ – 71, *in aeternum* ‘na zawsze’ – 68, *in medio* ‘w środku’, ‘pośrodku’ – 65, *in futurum* ‘w przyszłości’ – 58.

Często występujący przyimek *ad* można spotkać w takich popularnych łacińskich wyrażeniach, jak np.: *ad praesens* ‘obecnie’ – 87, *ad instar* ‘na kształt’, ‘na wzór’ – 40, *ad vivum* ‘do żywego’ – 34, *ad mentem* ‘w myśl’ – 33, *ad acta* ‘do akt’ – 30, a także w połączeniu z biernikiem gerundium – 260 razy, co oznacza, że w polskich tekstach badanego okresu cel często wyrażany był poprzez konstrukcję łacińską, na przykład w zdaniu: *Wziął to sobie ad deliberandum*. Na uwagę zasługuje również przyimek *sub*, którego polscy autorzy w XVII i XVIII w. chętnie używali w takich utartych wyrażeniach łacińskich, jak na przykład: *sub specie* ‘pod pozorem’ – 29 razy, *sub cura* ‘pod opieką’ – 14 razy czy *sub titulo* ‘pod tytułem’ – 127 razy, co stanowi ok. 44% wszystkich wystąpień słowa *titulo*, które pojawia się w korpusie 291 razy.

Niejednokrotnie formy łacińskie występujące w zgromadzonych w korpusie tekstach można spotkać w utartych zwrotach, jak na przykład wspomniany już wyżej rzeczownik *anno*, który występuje 48 razy w wyrażeniu *eodem anno*, natomiast w zwrocie *anno Domini* (bez względu na wielkość liter) występuje aż 152 razy. Wyszukiwarka korpusu daje ponadto możliwość badania frekwencji łacińskich zwrotów i ich polskich odpowiedników znaczeniowych. Przykładowo frazy *in casu* oraz *in casum* występują

w KorBie łącznie 124 razy, podczas gdy ich polskie odpowiedniki znaczeniowe, takie jak: *w przypadku*, *w wypadku*, *w okoliczności*, *w okolicznościach*, *na wypadek* występują łącznie 27 razy.

Za pomocą wyszukiwarki korpusu można ponadto przeprowadzać analizę porównawczą charakterystycznych zwrotów łacińskich w obrębie tekstów różnych autorów lub w zależności od stylu, tematyki lub gatunku, na przykład formuła *anno et die ut supra* ‘roku i dnia jak powyżej’ występuje 27 razy w korpusie, przy czym wszystkie frazy pochodzą jedynie z tekstów urzędowo-kancelaryjnych.

4. Badanie stopnia nasycenia łaciną różnego rodzaju tekstów

Współistnienie języka polskiego i łaciny w tekstach XVII–XVIII-wiecznych stwarza dobrą okazję do badania ich wzajemnych relacji. W tym punkcie artykułu skupiamy się na analizach stosunków ilościowych między językiem łacińskim a polskim, ponieważ w tym zakresie korpus dzięki dokładnej charakterystyce tekstów daje dość duże możliwości¹⁵. Przypomnijmy jednak, że do wyników tego typu badań należy podchodzić z dużą ostrożnością ze względu na to, że teksty włączone do korpusu nie zawierają zazwyczaj dłuższych fragmentów łacińskich. Dane korpusowe mogą jednak wskazać kierunek dalszych badań lub pomóc zweryfikować hipotezy badawcze. Poniżej przedstawiamy kilka zestawień liczbowych dotyczących obecności łaciny w polskich tekstach XVII–XVIII-wiecznych.

Dzięki przypisanym do każdego tekstu metadansom klasyfikującym teksty m.in. pod względem genologii można zbadać zależności między ilością łaciny a rodzajem tekstu. Klasyfikacja genologiczna tekstów w KorBie to dostosowana do specyfiki tekstów średniopolskich wersja typologii przyjętej w Narodowym Korpusie Języka Polskiego. Zgodnie z tym podziałem wyróżniono cztery rodzaje w obrębie literatury pięknej i sześć rodzajów wśród tekstów użytkowych. Oddzielny rodzaj stworzono dla włączonych do korpusu fragmentów Biblii mieszczących się na pograniczu beletrystyki i utworów niefikcyjnych.

Tabela 1. pokazuje zróżnicowanie pod względem ilości łaciny w tekstach polskich w zależności od charakteru danego tekstu. Część z tych różnic jest nieistotna statystycznie, zwraca jednak uwagę większe nasycenie łaciną tekstów użytkowych w porównaniu z beletrystycznymi oraz wyjątkowo duży udział łaciny w tekstach perswazyjnych (obejmujących kazania, pisma polityczne i społeczne, mowy okolicznościowe, przysłowia itp.).

¹⁵ Przeprowadzanie bardziej szczegółowych badań nad interferencjami polsko-łacińskimi oraz zjawiskiem code-switchingu wymaga od użytkownika korpusu pewnej pomysłowości w tworzeniu zapytań, a także większego wysiłku włożonego w analizę uzyskanych danych. To, że poprzez wyszukiwarkę korpusową dostajemy dostęp jedynie do fragmentów tekstu, wpływa na ograniczenie badań w tym zakresie. Z pewnością można jednak badać m.in. związki składniowe między segmentami polskimi i łacińskimi (np. realizowanie przez polskie formy łacińskich wymagań składniowych) czy polsko-łacińskie realizacje utartych zwrotów, takich jak określenia czasu (*die 28 maja vs. dnia 21 Aprilis*).

Literatura piękna i Biblia	Procentowy udział segmentów łacińskich	Teksty użytkowe	Procentowy udział segmentów łacińskich
Biblia	0%	faktograficzne	2%
epika	1%	kancelaryjne	3%
liryka	1%	listy	3%
dramat	1%	dydaktyczne	5%
utwory synkretyczne	2%	prasa	5%
		perswazyjne	8%

Tab. 1. Procentowy udział segmentów łacińskich w tekstach należących do poszczególnych rodzajów.

Możliwość ograniczenia wyszukiwania do tekstów pochodzących z pewnego przedziału czasowego pozwala na badanie zależności między ilością łaciny w tekstach a czasem powstania utworu. W tabeli 2. zestawiono wyniki wyszukiwania segmentów łacińskich znajdujących się w tekstach perswazyjnych w czterech podokresach wydzielonych w obrębie zakresu czasowego KorBy. W zestawieniu tym widać wyraźny wzrost ilości łaciny w tekstach perswazyjnych powstałych w 1. poł. XVIII w. Choć na ten wynik mogły mieć wpływ różne czynniki (np. w korpusie mógł się znaleźć duży fragment tekstu autora, który szczególnie chętnie makaronizował), to jednak pokazuje on jakąś prawidłowość, którą warto badać dalej.

1. poł. XVII w.	2. poł. XVII w.	1. poł. XVIII w.	2. poł. XVIII w. (do 1772 r.)
7%	8%	13%	6%

Tab. 2. Zmiana stopnia nasycenia łaciną tekstów perswazyjnych na przestrzeni lat 1601–1772.

Analiza liczby segmentów łacińskich w tekstach polskich może być również pomocna w badaniach porównawczych nad językiem różnych autorów piszących dzieła o tej samej tematyce. Tabela 3. pokazuje nasycenie łaciną zgromadzonych w KorBie tekstów o kometach XVII- i XVIII-wiecznych autorów. Zostały wśród nich wyróżnione dwa traktaty będące reakcją na pojawienie się komety w 1681 r.¹⁶, które mają podobny stosunek form łacińskich do całości tekstu – 7,5% i 7,7%. Dane te można wykorzystać jako punkt wyjścia do badania porównawczego języka tych dzieł (a także do badania

¹⁶ K.S. Ciekankowski, *Abryz Komety z Astronomicznej y Astrologicznej Uwagi, pod Meridianem Krakowskim, od 29. Grudnia Roku 1680, aż do pierwszych dni Lutego Roku 1681 wyrachowany [...]*, Kraków 1681; S. Niewieski, *Komety roku 1680. Widziáne, o których jest tu relacya z Prognostrykiem do Roku 1686 służącym*, Zamość 1681.

wszystkich polskich tekstów opisujących to zdarzenie¹⁷), jako tekstów tego samego typu i pisanych językiem nauki¹⁸.

Identyfikator tekstu	Stosunek segmentów łacińskich do wszystkich segmentów	Rok wydania
ŻędzKom	8,6%	1618
BemKom	1,5%	1619
FurUważ	0,4%	1664
CiekAbryz	7,5%	1681
NiewiesKom	7,7%	1681
DuńCiek	9,7%	1744
BohJProg	0,6%	1770

Tab. 3. Procentowy udział łaciny w XVII- i XVIII-wiecznych tekstach o kometach.

Badania nad stopniem nasycenia tekstów łaciną warto rozszerzyć o badania sposobu wykorzystania wyrażen łacińskich w utworach danego rodzaju. W odniesieniu do tekstów o tematyce astronomicznej można np. zobaczyć, jak kształtował się stosunek użytych wybranych terminów łacińskich do ich odpowiedników polskich. Następnym krokiem w badaniu języka tych prac pod kątem występowania łaciny może być szukanie *loci communes* – miejsc wspólnych tych dzieł. Niewykluczone, że oprócz wielu powtarzających się terminów łacińskich z dziedziny astronomii znajdziemy także te same, dłuższe fragmenty (kilkuczłonowe bądź całe zdania) z prac innych autorów, np. Arystotelesa czy Seneki często cytowanych w traktatach o kometach w XVII i XVIII w.

5. Lista frekwencyjna form łacińskich w KorBie

Wszystkie wymienione wyżej badania zostały przeprowadzone za pomocą wyszukiwarki korpusowej. Segmenty łacińskie w KorBie można jednak również analizować w inny sposób. Umożliwia to wygenerowana z korpusu lista frekwencyjna łacińskich form

¹⁷ O kometach pisali wówczas:

J. Wiśniowski, *Cometologia, Albo krótki dyskurs Astrologiczny, o Komecie i skutkach jego, który się pokazał [...] Roku [...] 1680 i trwał w Roku blisko Następującym 1681*, Kraków 1681;

S. Słowakowic, *Postliminium Cometarum, abo raczej Niebo z Dawną a niesłusznie Kometom przez Filozofow odebrane, A teraz znowu Prawem y wymiarem Geometrycznym, z Okazyjej świecącej w Roku 1680 y 1681, w Grudniu y w Styczniu Komety, za staraniem [...] Przywroczone*, Kraków 1681;

K. Krzykawski, *Krotka Astrologiczna Uwaga Komety. Na Horyzoncie Krakowskim, okolo ostatnich dni Grudnia powstającego w Roku Pańskim 1680, ginącego zaś okolo pierwszych dni Lutego w roku niniejszym 1681*, Kraków 1681.

¹⁸ Podobne badania ilościowe nad łaciną w tekstach średniopolskich zostały przeprowadzone m.in. na materiale XVII- i XVIII-wiecznych diariuszy podróży (Skolimowska, Kunicki-Goldfinger, 2004).

wyrazowych. Każdej formie została przypisana informacja o liczbie jej wystąpień w korpusie. Przy podliczaniu pominięto znaki interpunkcyjne oraz cyfry arabskie, chyba że dana forma składała się z cyfr i liter, np. *2do* (*secundo* ‘po drugie’). Zignorowana została wielkość liter oraz znaki diakrytyczne nad samogłoskami (czyli np. forma *anno* liczona była łącznie z formą *Anno*, a forma *aliàs* – z formą *alias*). Na tak skonstruowanej liście znalazło się 77 256 form. Nie jest to jednak dokładna liczba łacińskich form w KorBie, gdyż pojawiły się na niej także formy błędnie zapisane (np. *quidem*, *aanno*) lub błędnie oznaczone jako łacińskie (np. *życzę*, *śmiercią*). Wylimitowanie tych błędów byłoby bardzo pracochłonne, gdyż wymagałoby przejrzania całej listy i sprawdzenia, jaka jest rzeczywista postać podejrzanych form w tekstach źródłowych. Takie ewidentnie błędne formy mają jednak pojedyncze wystąpienia, w małym stopniu zaburzają więc frekwencję form znajdujących się na początku listy, czyli tych, których analiza przynosi najwięcej informacji o obecności łaciny w polskich tekstach z XVII i XVIII w. Poniżej przedstawiamy parę obserwacji dotyczących form łacińskich najczęściej występujących w KorBie.

Tabela 4. pokazuje pierwsze 22 pozycje¹⁹ na liście frekwencyjnej. Jak można było przypuszczać, na początku listy znalazły się wyrazy funkcyjne, takie jak przyimek *in* (13 366 wystąpień) i spójnik *et* (12 002 wystąpień), a w następnej kolejności inne formy przyimków (np. *ad*, *de*, *ex*, *cum*, *pro*), spójników (np. *item*, *cum*, *ut*), zaimków (np. *qui*) oraz partykuła *non*. Do najczęściej występujących form w polskich tekstach z XVII i XVIII w. należą również popularne łacińskie skróty: *etc.* (*et cetera* ‘i tak dalej’), *lib.* (*liber* ‘księga’), *cap.* (*caput* ‘rozdział’), a także czasownik *est* ‘jest’ oraz formy rzeczowników związanych z datacją: *anno* ‘roku’ i *die* ‘dnia’.

Lp.	Forma łacińska	Liczba wystąpień w korpusie	Lp.	Forma łacińska	Liczba wystąpień w korpusie
1.	<i>in</i>	13 366	12.	<i>per</i>	1753
2.	<i>et</i>	12 002	13.	<i>item</i>	1746
3.	<i>ad</i>	4842	14.	<i>a</i>	1745
4.	<i>non</i>	4105	15.	<i>lib</i>	1266
5.	<i>de</i>	3700	16.	<i>qui</i>	1265
6.	<i>ex</i>	2852	17.	<i>cap</i>	1210
7.	<i>est</i>	2847	18.	<i>sub</i>	1088
8.	<i>etc</i>	2617	19.	<i>ut</i>	1051
9.	<i>cum</i>	2412	20.	<i>anno</i>	995
10.	<i>alias</i>	2100	21.	<i>ab</i>	991
11.	<i>pro</i>	1858	22.	<i>die</i>	945

Tab. 4. Początek listy frekwencyjnej form łacińskich w KorBie.

¹⁹ Liczba przedstawionych w tabeli słów została ustalona przez autorów na 22, ponieważ odnotowuje się dużą różnicę między liczbą wystąpień wyrazu z pozycji 22 (*die* – 945) a użytą formą z pozycji 23 (*quod* – 829).

W tabeli 5. zawarte zostały najczęściej występujące formy rzeczowników łacińskich²⁰. Warto zauważyć, że najpopularniejsze formy rzeczowników to wyrazy oznaczające Boga/bóstwo, czyli wszystkie z wyjątkiem wołacza formy liczby pojedynczej rzeczownika *deus*: *deus*, *dei*, *deum*, *deo* – łącznie formy te występują w korpusie 1500 razy. Słowo o podobnym znaczeniu, *dominus* ‘pan’, zostało użyte 304 razy, dopełniacz tego słowa, *domini* – 434 razy, natomiast wszystkich użyć form w liczbie pojedynczej od tego słowa odnotowuje się w KorBie 1087. Dalsze badania korpusowe wykazały, że tylko w nielicznych wypadkach formy od słowa *dominus* zostały użyte w znaczeniu innym niż ‘Pan Bóg’ i zostały zapisane małą literą – w odniesieniu do człowieka używano zwykle polskiego rzeczownika *pan*, natomiast dla Boga pozostawiano terminy łacińskie jako bardziej wyszukane i wyrażające szacunek.

Inne najczęściej występujące formy rzeczowników łacińskich to głównie terminy związane ze społeczeństwem, prawem i ustrojem, np. *status* ‘stan, sytuacja’, *consilium* ‘rada, zamiar’, *rex* ‘król’.

Lp.	Forma łacińska	Liczba wystąpień w korpusie	Lp.	Forma łacińska	Liczba wystąpień w korpusie
15.	lib	1266	74.	status	338
17.	cap	1210	81.	parte	323
20.	anno	995	85.	dominus	304
22.	die	945	87.	consilium	300
44.	omnia	537	89.	nomine	294
45.	deus	491	91.	titulo	291
46.	dei	481	92.	rerum	291
47.	teste	479	93.	ratione	291
53.	domini	434	97.	vita	283
57.	summa	396	99.	deo	282
58.	vulgo	389	119.	rex	247
60.	nihil	385	121.	deum	246
67.	omnes	358			

Tab. 5. Najczęściej występujące formy łacińskich rzeczowników w KorBie.

6. Podsumowanie

Elektroniczny Korpus Tekstów Polskich z XVII i XVIII wieku (do 1772 r.) dzięki oznakowaniu segmentów obcojęzycznych umożliwia prowadzenie badań nie tylko nad XVII- i XVIII-wieczną polszczyzną, ale także nad nowożytną łaciną. Badania te z jednej strony mogą dotyczyć poszczególnych słów i zwrotów łacińskich – ich znaczeń oraz

²⁰ Lista przedstawionych najczęściej występujących form rzeczowników łacińskich została ustalona przez autorów na 25 i kończy się słowem *deum*, aby łatwiej przedstawić pokazaną zależność użyć form od słowa *deus* oraz *dominus*.

sposobów używania w tekstach średniopolskich – a z drugiej mogą być nakierowane na analizę wzajemnych związków między językiem polskim a łacińskim w epoce nowożytnej. To drugie podejście otwiera przed badaczami szereg możliwości, począwszy od określania stopnia nasycenia łaciną różnego rodzaju tekstów aż do zbadania funkcji, jakie zwroty łacińskie pełnią w konkretnym utworze. Dając łatwy dostęp do licznych fragmentów łacińskich zawartych w polskich tekstach z XVII i XVIII w., KorBa przynajmniej w części rekompensuje brak polskiego korpusu łacińskich tekstów nowożytnych oraz słownika polskiej łaciny nowożytnej.

Bibliografia

- Bień, J.S. (2014). The IMPACT project Polish Ground-Truth texts as a DjVu corpus. *Cognitive Studies | Études Cognitives*, 14, s. 75–84. Pobrane z <https://ispan.waw.pl/journals/index.php/cs-ec/article/view/cs.2014.008/174> (4.08.2020).
- Czerenkiewicz, M. (2015). Czy potrzeba nam dzisiaj polskiego słownika łaciny renesansowej i barokowej? *Polonica*, 35, s. 39–47.
- Czerenkiewicz, M., Ryczek, W. (2013). Studia (neo)latina rediviva. Kilka uwag i postulatów odnośnie neolatynistyki. *Terminus*, T. 15, 3(28), s. 335–358. Pobrane z <https://doi.org/10.4467/20843844TE.13.021.1578> (4.08.2020).
- Gruszczyński, W., Adamiec, D., Bronikowska, R., Wieczorek, A. (2020). Elektroniczny Korpus Tekstów Polskich z XVII i XVIII w. – problemy teoretyczne i warsztatowe. *Poradnik Językowy*, 8, s. 32–51.
- Nowak, K. (2014). Fontes Mediae et Infimae Latinitatis Polonorum. Z prac nad korpusem polskiej łaciny średniowiecznej. *Polonica*, 34, s. 105–114.
- Ogrodniczuk, M., Gruszczyński, W. (2019). Connecting Data for Digital Libraries: The Library, the Dictionary and the Corpus. W: A. Jatowt, A. Maeda, S. Yeon Syn (eds.), *Digital Libraries at the Crossroads of Digital Information for the Future. ICADL 2019. Lecture Notes in Computer Science*, 11853, s. 125–138. Pobrane z https://doi.org/10.1007/978-3-030-34058-2_13 (2.12.2019).
- Plezia, M., Weyssenhoff-Brożkowa, K., Rzepiela, M. (red.) (1953–). *Słownik łaciny średniowiecznej w Polsce*. Wrocław–Kraków: Zakład Narodowy im. Ossolińskich – Wydawnictwo Polskiej Akademii Nauk.
- Skolimowska, A., Kunicki-Goldfinger, M. (2004). Łacina w polskojęzycznych diariuszach podróży. W: J. Axer (red.), *Łacina jako język elit* (s. 189–220). Warszawa: Wydawnictwo DiG.
- Sondel, J. (1997). *Słownik łacińsko-polski dla prawników i historyków*. Kraków: Universitas.

STRESZCZENIE

Słowa kluczowe: łacina nowożytna, badania korpusowe, teksty średniopolskie.

Artykuł przedstawia Elektroniczny Korpus Tekstów Polskich z XVII i XVIII wieku (nieformalnie zwany KorBa) pod kątem jego użyteczności dla badań neolatynistycznych. Korpus zawiera ok. 0,5 mln segmentów oznakowanych jako słowa łacińskie, co pozwala na ich łatwe wyszukanie za pomocą wyszukiwarki korpusowej. W artykule opisano przykładowe badania, jakie można prowadzić na zgromadzonym w korpusie materiale, zarówno nad poszczególnymi zwrotami łacińskimi, jak i nad wzajemnymi relacjami między językiem polskim a łaciną w tekstach średniopolskich. Zostały także zaprezentowane niektóre dane zaczerpnięte z listy frekwencyjnej łacińskich form wyrazowych znajdujących się w korpusie.

SUMMARY

Latin in KorBa. Usefulness of the Electronic Corpus of 17th- and 18th-century Polish Texts for a Neo-Latinist philologist

Keywords: modern Latin, corpus research, Middle Polish texts.

The article presents the Electronic Corpus of 17th- and 18th-century Polish Texts (informally called KorBa) in terms of its usefulness for neo-Latinist research. The corpus contains approximately 0.5 million tokens annotated as Latin words, which makes them easy to find by means of the corpus search engine. The article describes examples of research that can be carried out on the material collected in the corpus, both on individual Latin phrases and on the mutual relations between the Polish language and Latin in Middle Polish texts. Some data taken from the frequency list of Latin word forms in the corpus were also presented.