Academic Journal of Modern Philology

ISSN 2299–7164
Vol. 13 (2021)
Special Issue
s. 123–132

Łukasz Halida
Instytut Języka Polskiego PAN, Kraków

Kościelne słownictwo prawno-administracyjne w polskiej łacinie średniowiecznej – analiza z wykorzystaniem metod korpusowych[1]

The Vocabulary of the Church's Law and Administration in Polish Medieval Latin: An Analysis Using the Methods of Corpus Linguistics

Abstract

This paper concerns the vocabulary of the Church's law and administration in Latin texts written during the Middle Ages in Poland and its automatic extraction using the methods of corpus linguistics. The first part of this article considers the basic theoretical assumptions of the automatic extraction of this specialized vocabulary and the main characteristics of the Electronic Corpus of Polish Medieval Latin. In the second part presents the methods and results of term extraction. For the purpose of this research, a specialized subcorpus, including synodal statues and documents of ecclesiastical chapters, was created and then compared with the reference corpus. As a result, a list of lexemes, which appeared relatively frequently in the subcorpus and rarely in the reference corpus, was obtained. This difference in relative frequency was the main criterion for the recognition of potential terminological units. Verification on the basis of lexicographic data demonstrated the effectiveness of the adopted methods. The aim of this paper was to present the usefulness of the Electronic Corpus of Polish Medieval Latin for the research and analysis of specialized vocabulary.

Keywords: medieval Latin, electronic corpus, automatic term extraction, specialized vocabulary, ecclesiastical law and administration

doi:10.34616/ajmp.2021.13

1 Publikacja finansowana w ramach programu Ministra Nauki i Szkolnictwa Wyższego pod nazwą „Narodowy Program Rozwoju Humanistyki" w latach 2018–2023, nr projektu 0116/NPRH6/H11/85/2018 (Elektroniczny korpus polskiej łaciny średniowiecznej. Kontynuacja prac), kwota finansowania 1,8 mln zł.

1. Wstęp

Łacińska leksyka prawno-administracyjna ewoluowała wraz z rozwojem prawa i struktur urzędowych. Wzmożonego tempa proces ten nabrał w średniowieczu – pojawienie się nowych instytucji, funkcji i urzędów, a także wyjście łaciny z przestrzeni języka mówionego skutkowało istotnymi przeobrażeniami w leksyce prawno-administracyjnej – z jednej strony było bodźcem do tworzenia nowych leksemów, z drugiej natomiast do używania wyrazów odziedziczonych w nowych, specjalistycznych znaczeniach (Sondel 2001: XVIII–XXII). Analiza leksyki specjalistycznej przysparza jednak pewnych trudności. Jednym z podstawowych problemów, z jakimi musi zmierzyć się badacz tego zagadnienia, jest bowiem ustalenie kryteriów pozwalających określić, które wyrazy można uznać za specjalistyczne, oraz wybór metod ich ekscerpcji. Szczególnie istotne okazuje się to w kontekście języka historycznego, takiego jak łacina średniowieczna, w wypadku którego niemożliwa jest weryfikacja uzyskanych wyników przez specjalistów będących rodzimymi użytkownikami języka. Z pomocą przychodzą jednak osiągnięcia dynamicznie rozwijającej się w ostatnich latach humanistyki cyfrowej i językoznawstwa kwantytatywnego. W niniejszym artykule zaprezentowana zostanie jedna z automatycznych metod pozyskiwania leksemów specjalistycznych z wykorzystaniem Elektronicznego korpusu polskiej łaciny średniowiecznej (https://www.scriptores.pl/efontes), stanowiącego główne narzędzie analizy. W pierwszej części artykułu przedstawione zostaną podstawowe założenia metody automatycznego pozyskiwania wyrazów specjalistycznych oraz ogólna charakterystyka wykorzystywanego w pracy korpusu, w drugiej zaś przebieg ekstrakcji kościelnego słownictwa prawno-administracyjnego wraz z podaniem krótkiego opisu analizowanych tekstów i wyłonionych z nich leksemów. Głównym celem artykułu jest ukazanie funkcjonalności i przydatności Elektronicznego korpusu polskiej łaciny średniowiecznej w badaniach leksyki specjalistycznej oraz możliwości jego wykorzystania.

2. Podstawowe założenia automatycznej ekstrakcji leksyki specjalistycznej

Zasadniczym celem automatycznej ekstrakcji leksemów specjalistycznych jest wytypowanie i rozpoznanie słownictwa określonej dziedziny wiedzy (Heylen & De Hertog 2015: 203). Jej wyniki wykorzystywać można między innymi w pracy terminograficznej (tworzenie słowników terminologicznych lub bazy słownictwa specjalistycznego) oraz translatorskiej. W porównaniu z ekstrakcją ręczną metody automatyczne znacznie usprawniają wydobywanie leksemów z tekstów, pozwalając zaoszczędzić czas. Automatyczna ekstrakcja umożliwia wstępną identyfikację i preselekcję kandydatów na wyrazy specjalistyczne – konieczna jest zatem późniejsza weryfikacja i obiektywizacja uzyskanych danych. W wypadku języka historycznego (tutaj łaciny średniowiecznej) owa obiektywizacja opiera się przede wszystkim na danych leksykograficznych oraz analizie kontekstów źródłowych. Bazą automatycznej ekstrakcji są elektroniczne korpusy tekstów specjalistycznych. W podejściu kontrastywnym oprócz korpusów specjalistycznie sprofilowanych obejmujących teksty wybranej dziedziny wiedzy ważne są również korpusy ogólne (tzn. w sposób zrównoważony prezentujące teksty o różnych rejestrach), a ekstrakcja kandydatów na leksemy specjalistyczne opiera się między innymi na porównaniu obydwu zbiorów (Heylen & De Hertog 2015: 204).

3. Elektroniczny korpus polskiej łaciny średniowiecznej

Elektroniczny korpus polskiej łaciny średniowiecznej powstaje w Pracowni Łaciny Średniowiecznej Instytutu Języka Polskiego PAN w ramach projektu *„eFontes*. Elektroniczny korpus polskiej łaciny średniowiecznej" finansowanego ze środków Narodowego Programu Rozwoju Humanistyki w latach 2018–2023. Jest kontynuacją pilotażowej wersji korpusu opracowanej w latach 2012–2017. Docelowo ma liczyć 15 milionów tokenów. Korpus obejmuje teksty spisane w języku łacińskim między początkiem XI a połową XVI wieku. Ma zatem charakter synchroniczny i pozwala na badanie przede wszystkim łaciny polskiego średniowiecza, chociaż włączone do korpusu teksty późniejsze pochodzące z XVI wieku umożliwiają również analizę zmian językowych zachodzących w początkach renesansu (Nowak 2014: 106). W korpusie znalazły się zasadniczo źródła pochodzące z ziem polskich. Ze względu jednak na zmieniające się na przestrzeni kilku wieków granice Polski i trwający proces kształtowania się świadomości narodowej precyzyjna atrybucja kraju powstania danego źródła często okazywała się niemożliwa. Dlatego też do korpusu postanowiono włączyć również teksty z ziem pogranicznych, co spowodowało brak konieczności jednoznacznych rozstrzygnięć i dodatkowo poszerzyło możliwości badawcze omawianego narzędzia między innymi o analizę interferencji między łaciną a innymi niż polski językami narodowymi (np. czeskim i niemieckim) (Nowak 2014: 106–107). Ponadto prezentowany korpus jest korpusem ogólnym, tzn. przedstawia całokształt piśmiennictwa łacińskiego średniowiecznej Polski i okolic, przy czym należy zaznaczyć, że nie zbiera on wszystkich dzieł ze wskazanego obszaru i epoki. Korpus nie jest bowiem zwykłą, nieustrukturyzowaną kolekcją przypadkowych tekstów, lecz w założeniu ma być reprezentacją języka (Biber, Conrad & Reppen 1998: 246). Ponadto stworzenie takiego zbioru w praktyce okazałoby się nie do zrealizowania – nie tylko ze względu na ilość materiału, który należałoby uwzględnić, ale także na fakt, że wciąż bardzo duża liczba źródeł pozostaje niewydana (a w korpusie uwzględnianie są tylko teksty wydane). Zadaniem Elektronicznego korpusu polskiej łaciny średniowiecznej jest zatem przedstawienie możliwie najbardziej reprezentatywnej i zrównoważonej próby tekstów odzwierciedlającej obraz polskiej łaciny średniowiecznej (Nowak 2014: 107). Możliwie najbardziej reprezentatywnej, gdyż w dążeniu do reprezentatywności i zrównoważenia korpusu napotyka się kilka trudności – nie tylko w kwestii wskazanego problemu niewydania tekstów, ale także stanu zachowania źródeł, możliwości ich pozyskania oraz ich chronologicznej, geograficznej i gatunkowej dysproporcji (Nowak 2014: 111–112). Mimo to w korpusie znalazła się szeroka gama tekstów stanowiąca przekrój piśmiennej produkcji polskiego średniowiecza: od dyplomatyki i zapisków sądowych, przez teksty historiograficzne i hagiograficzne, aż po poezję. Mimo że prace nad projektem wciąż trwają, obecny stan korpusu umożliwia już przeprowadzenie miarodajnych badań – kolejne działania nie zmienią bowiem znacząco architektury korpusu, lecz zwiększając jego zawartość, pozwolą na prowadzenie jeszcze bardziej zaawansowanych analiz. Elektroniczny korpus polskiej łaciny średniowiecznej – podobnie jak inne korpusy elektroniczne – pozwala między innymi na tworzenie konkordancji wystąpień, badanie kolokacji oraz frekwencji wyrazów w tekstach – zarówno pod względem fleksyjnym (co możliwe jest dzięki anotacji morfosyntaktycznej), jak i dystrybucyjnym (np. pod względem gatunku, autora i czasu powstania wskutek wprowadzonych metadanych). Korpus z powodzeniem wyzyskać można w pracach leksykograficznych, a także w analizach leksykalno-semantycznych i kwantytatywnych.

4. Kościelne słownictwo prawno-administracyjne w polskiej łacinie średniowiecznej

4.1. Podkorpus tekstów związanych z prawem i administracją kościelną

Elektroniczny korpus polskiej łaciny średniowiecznej pozwala również na tworzenie podkorpusów z dowolnie wybranych tekstów wchodzących w jego skład, a następnie porównywanie ich bądź to z całym zasobem korpusu, bądź innymi podkorpusami. Porównywanie podkorpusu specjalistycznie sprofilowanego z całością danych korpusowych (podejście kontrastywne) jest – jak wspomniano wyżej – jedną z metod umożliwiających wskazanie kandydatów na leksemy specjalistyczne. Metodę tę przyjęto także w próbie analizy kościelnego słownictwa prawno-administracyjnego w tekstach polskiej łaciny średniowiecznej. Na jej potrzeby stworzono liczący 346 234 tokenów specjalistyczny podkorpus tekstów związanych z prawem i administracją kościelną, wyodrębniony z Elektronicznego korpusu polskiej łaciny średniowiecznej i stanowiący 10,4% jego całości. Niewielki, specjalistyczny korpus – tworzony w celu analizy określonego, specyficznego problemu badawczego – obejmujący wybrane rejestry i gatunki umożliwia w większym stopniu niż duży, ogólny korpus dostrzeżenie regularnych wzorców w typowym dla tekstów słownictwie i strukturach (Koester 2010: 71; 69). Jest to możliwe między innymi dzięki temu, że dane w nim zawarte nie są zdekontekstualizowane, co uważa się za jedną z głównych zalet niedużych, wyspecjalizowanych korpusów (Koester 2010: 67; 74). W skład omawianego podkorpusu weszły wszystkie – ręcznie wyselekcjonowane z aktualnej listy źródeł korpusu – najważniejsze dotychczas włączone doń zbiory tekstów kościelnych o charakterze prawno-administracyjnym: akta kapituł gnieźnieńskiej, poznańskiej i włocławskiej pochodzące z lat 1408–1530, a ponadto najstarsze statuty synodalne krakowskie biskupa Nankera z 1320 roku, najdawniejsze statuty synodalne archidiecezji gnieźnieńskiej z początku XV wieku, a także statuty synodalne krakowskie Zbigniewa Oleśnickiego z 1436 roku oraz 1446 roku.

4.2. Charakterystyka tekstów podkorpusu

Statuty synodalne były normatywnymi aktami prawa Kościoła partykularnego uchwalanymi w czasie zgromadzeń duchowieństwa danej prowincji kościelnej lub diecezji i dla niej przeznaczonymi. Stanowiły efekt dostosowania powszechnego prawa kanonicznego do prawodawstwa poszczególnych jednostek kościelnych, ich warunków oraz potrzeb (Pontal 1975: 17–51; Zygner 2013: 423). I chociaż najistotniejszą rolę w kształtowaniu prawa kościelnego w Polsce odgrywały synody prowincjonalne (przede wszystkim synody prowincji gnieźnieńskiej obejmującej przez długi czas całe terytorium ówczesnej Polski), to jednak postanowienia powzięte na synodach diecezjalnych także nie pozostawały w tym względzie bez znaczenia (Grochowski 1977: 334–335). Statuty synodalne Nankera (ok. 1270–1341), biskupa krakowskiego, a następnie wrocławskiego, ogłoszone na synodzie diecezjalnym w Krakowie w 1320 roku, zasługują na szczególną uwagę. Wyszły one spod ręki samego biskupa (Wójcik 1969: 454) i świadczyły o jego doskonałej znajomości prawa. Powstały jako odpowiedź na potrzebę odnowy duchowej diecezji krakowskiej, na długo normując działalność duszpasterską (Kumor & Obertyński 1974: 234; Kłoczowski 1966: 198–199). Statuty ogłoszone przez Nankera (jak również działania innych synodów partykularnych) odgrywały bowiem ważną rolę w dziele reformy Kościoła. Stanowiły pierwsze tego rodzaju kompendium pastoralne zawierające zalecenia natury prawniczej, teologicznej i moralnej (Zygner 2013:

424–425; 438). Charakter reformatorski miały także statuty ogłoszone na synodach diecezjalnych przez Zbigniewa Oleśnickiego (1389–1455) – jednego z najznamienitszych następców Nankera na stolicy biskupów krakowskich. Dotyczyły one spraw duszpasterskich i duchowieństwa parafialnego. W tym samym nurcie sytuują się statuty diecezji gnieźnieńskiej obejmujące nakazy dotyczące sprawowania liturgii i postępowania kleru.

Oprócz synodów ważną funkcję w średniowiecznym Kościele pełniła kapituła. Ta szczególnie eksponowana w administracji kościelnej instytucja była kolegium duchownych stanowiącym rodzaj senatu biskupa (Kłoczowski 1966: 199). Jej rola nie ograniczała się jednak wyłącznie do głosu doradczego. Gremium to – mające własny autonomiczny ustrój – cieszyło się bowiem wieloma daleko sięgającymi uprawnieniami, uczestnicząc w sprawowaniu władzy w diecezji (Kumor & Obertyński 1974: 401). Przebieg posiedzeń danej kapituły i powzięte w ich trakcie postanowienia spisywano w aktach kapitulnych.

W podkorpusie znalazły się zatem przede wszystkim teksty pragmatyczne (akta kapitulne) oraz normatywne (statuty synodalne). Powstawały one na przestrzeni trzech wieków, od I poł. XIII do I poł. XVI, w różnych częściach Polski, w ważnych ośrodkach kościelnych: Krakowie, Poznaniu, Gnieźnie i Włocławku. Można zatem mówić – pomimo selektywnego i próbnego charakteru stworzonego zbioru – o pewnej jego reprezentatywności, ważnej w wypadku każdego korpusu, a zwłaszcza korpusu niedużego i specjalistycznego (Koester 2010: 69; Heylen & De Hertog 2015: 205). Istotne również, że teksty podkorpusu mają wyraźnie specjalistyczny profil – można więc przypuszczać, że dzięki tak dobranemu zbiorowi przyjęta w niniejszej analizie metoda okaże się skuteczna nie tylko w badaniu słów kluczowych, lecz również słownictwa specjalistycznego.

4.3. Porównywanie korpusów

Pierwszy etap analizy stanowi wygenerowanie listy frekwencyjnej leksemów występujących w stworzonym podkorpusie. Najczęściej pojawiające się wyrazy wstępnie umożliwiają bowiem wskazanie typowej leksyki wchodzących w jego skład tekstów – lista frekwencyjna może w ten sposób służyć za narzędzie wstępnej analizy słownictwa specjalistycznego. Bardziej szczegółowe wyniki uzyskać można jednak dopiero przez porównywanie między sobą dwóch zbiorów – całości danych korpusowych i utworzonego na ich podstawie podkorpusu. Analiza porównawcza pozwala bowiem na wyłonienie wyrazów, które z dużym prawdopodobieństwem można wykluczyć z danej leksyki specjalistycznej lub które można do niej zaliczyć.

Należy podkreślić, że dane prezentowane będą (poza frekwencją bezwzględną pokazującą rzeczywistą liczbę wystąpień) również w formie wyników znormalizowanych według frekwencji względnej (na milion wyrazów), uwzględniającej rozmiary porównywanych zbiorów i określającej liczbę występowania danej jednostki w próbce na milion tokenów. Wprawdzie posługiwanie się frekwencją znormalizowaną również obciążone jest pewnym ryzykiem błędu – ta bowiem nie zawsze w sposób uprawniony zakłada jednorodność danych językowych (Piotrowski & Grabowski 2013: 62–65) – umożliwia jednak przeprowadzenie analizy z większą precyzją i wiarygodnością.

W wyniku porównania całości danych korpusowych z podkorpusem generowana jest lista leksemów względnie często występujących w całym korpusie, a rzadko w podkorpusie. Takie jednostki z dużym prawdopodobieństwem można wykluczyć z grona kandydatów na leksemy specjalistyczne danej dziedziny – ich wysoka frekwencja względna w całym korpusie, a niska w podkorpusie specjalistycznym wskazywałaby bowiem na ich niespecjalistyczny charakter. Dzięki natomiast porównaniu podkorpusu

specjalistycznego z całością korpusu uzyskuje się listę wyrazów względnie często występujących w podkorpusie, rzadko zaś w całym korpusie. Jeśli zatem dany leksem relatywnie często pojawia się w sprofilowanych specjalistycznie tekstach, a jednocześnie rzadko w pozostałych źródłach, z dużą dozą prawdopodobieństwa będzie można uznać go za kandydata na jednostkę słownictwa specjalistycznego.

4.4. Porównanie podkorpusu specjalistycznego z całością korpusu

Przed prezentacją wyników porównania specjalistycznego podkorpusu kościelnych tekstów prawno-administracyjnych z całością Elektronicznego korpusu polskiej łaciny średniowiecznej należy zaznaczyć, że oprogramowanie korpusowe daje możliwość określenia parametru, który pozwala skoncentrować poszukiwania na wyrazach bądź to rzadko, bądź też często występujących w stanowiącym punkt odniesienia korpusie. W niniejszej analizie dla tego parametru ustalono wartość 50 (możliwa skala od 0,00001 do 999 999 999), dzięki czemu na liście leksemów będącej efektem porównania podkorpusu z całym korpusem znalazły się wyrazy raczej – ale nie wyjątkowo – rzadkie w całości danych korpusowych, a zatem takie, których występowanie nie ograniczało się wyłącznie do tekstów podkorpusu specjalistycznego (Kilgarriff 2009). Wskazanie takiej właśnie wartości uzasadnione było architekturą korpusu i charakterem analizowanej leksyki. Słownictwo specjalistyczne związane z prawem i administracją kościelną występowało bowiem nie tylko w tekstach wchodzących w skład omawianego podkorpusu, lecz pojawiało się także między innymi w dokumentach traktujących o sprawach Kościoła spoza zbioru statutów synodalnych i akt kapitulnych. Gdyby wskaźnik ten był niższy, niektóre istotne dla analizy leksemy mogłyby zostać pominięte – nie mogłyby bowiem zostać wykorzystane do celów porównawczych, gdyż potwierdzone zostałyby tylko w tekstach podkorpusu. Co więcej, liczne wyrazy będące elementami leksyki specjalistycznej to wyrazy polisemiczne, które jako leksemy języka ogólnego używane były również w nowym, specjalistycznym znaczeniu (Buttler 1979). Z uwagi na wysoki stopień polisemiczności leksemów łacińskich konieczne zatem było uwzględnienie również tego czynnika. Ustalona więc – co wypada podkreślić – empirycznie – wartość 50 okazała się dla potrzeb niniejszej analizy optymalna: lista kandydatów uwzględnia leksemy często występujące w podkorpusie, ale jednocześnie obecne także w innych źródłach spoza jego zbioru wchodzących w skład Elektronicznego korpusu polskiej łaciny średniowiecznej. Poniżej przedstawiono listę 25 leksemów, w wypadku których różnica między względną frekwencją w analizowanym podkorpusie specjalistycznym (*podkorpus*) i całym Elektronicznym korpusie polskiej łaciny średniowiecznej (*korpus*) jest największa. Uwzględniono wystąpienia wszystkich form fleksyjnych danego leksemu. Zestawienie uporządkowano według frekwencji względnej.

Tab. 1

	Leksem	Frekwencja bezwzględna		Frekwencja względna		Punkty
		podkorpus	*korpus*	*podkorpus*	*korpus*	
1.	*venor*	1 068	1 112	3 084,619	336,592	8,11
2.	capitulum	3 963	4 948	11 446,016	1 484,366	7,49
3.	canonicus	1 498	2 168	4 326,554	650,385	6,25
4.	archidiaconus	686	941	1 981,319	282,294	6,11
5.	deputo	627	886	1 810,914	265,794	5,89
6.	*honor*	265	334	765,378	100,198	5,43
7.	administrator	210	245	606,526	73,498	5,32

	Leksem	Frekwencja bezwzględna		Frekwencja względna		Punkty
		podkorpus	*korpus*	*podkorpus*	*korpus*	
8.	decanus	636	1 024	1 836,908	307,193	5,28
9.	cantor	343	495	990,659	148,497	5,24
10.	*IV*	506	795	1 461,439	238,495	5,24
11.	*Hector*	177	191	511,215	57,299	5,23
12.	capitularis	183	230	528,544	68,998	4,86
13.	generalis	1 376	2 629	3 974,191	788,682	4,8
14.	custos	390	663	1 126,406	198,895	4,73
15.	congrego	321	553	927,119	165,896	4,53
16.	scolasticus	243	404	701,837	121,197	4,39
17.	tractatus	205	329	592,085	98,698	4,32
18.	*o*	518	1 035	1 496,098	310,493	4,29
19.	*Aprilis*	305	560	880,907	167,996	4,27
20.	decerno	826	1 755	2 385,67	526,488	4,23
21.	provincialis	173	269	499,662	80,698	4,21
22.	*II*	842	1 813	2 431,881	543,888	4,18
23.	*Iulius*	223	395	644,073	118,497	4,12
24.	archiepiscopalis	141	206	407,239	61,799	4,09
25.	*October*	284	545	820,255	163,496	4,08

Powyższe wyniki należało od razu poddać wstępnej selekcji[2]. Niektóre wskazania są bowiem efektem błędnego przyporządkowania przez oprogramowanie danego wyrazu określonej lemmie. Jest to spowodowane faktem, że – jak wskazano – Elektroniczny korpus polskiej łaciny średniowiecznej to wciąż *opus imperfectum*, a trzeba zaznaczyć, że lemmatyzacja tekstów korpusu należy do jednych z najbardziej wymagających procesów, zwłaszcza w wypadku korpusów języków historycznych ze skomplikowaną morfologicznie leksyką, takich jak łacina średniowieczna (Manjavacas, Kádár & Kestemont 2019: 1493). Ponadto owe błędy wynikają pośrednio z rozwiązań edytorskich zastosowanych w uwzględnianych źródłach. Dla przykładu w edycji akt kapitulnych będących przedmiotem niniejszej analizy zachowano występujące w rękopisach liczne skróty. W takiej skróconej formie wyrazy znalazły się w korpusie, a następnie zostały uznane za formy innej niż w rzeczywistości lemmy. W konsekwencji na liście pojawiły się leksemy *VENOR* (poz. 1) 'poluję' oraz *HONOR* (poz. 6) 'zaszczyt, honor'. Konkordancja ich wystąpień ujawnia jednak, że faktycznie chodzi o leksemy VENERABILIS 'czcigodny' (skrót *vener.*) oraz HONORABILIS 'szanowny' (skrót *honor.*). Są to wyrazy pełniące funkcję honoratywną, licznie występujące także poza analizowanymi tekstami – wydaje się zatem, że można je wyeliminować bez większego uszczerbku dla wiarygodności prowadzonych badań. Ponadto z analizy wykluczyć trzeba imiona (*HECTOR* – poz. 11) oraz nazwy miesięcy (*APRILIS* – poz. 19, *IULIUS* – poz. 23, *OCTOBER* – poz. 25), a także liczebniki w zapisie rzymskim (*IV* – poz. 10, *II* – poz. 22). Pojawienie się na liście leksemów niebędących wyrazami specjalistycznymi, np. nazw miesięcy, spowodowane jest charakterem wykorzystanych źródeł. Przykładowo akta kapitulne opatrywane były dokładną datacją z podaniem dnia, miesiąca i roku, przy czym miesiąc

2 Leksemy, które powinny zostać wyeliminowane, zaznaczono w tabeli kursywą.

zapisywano słownie. Datacja ta stanowiła integralną część tekstu łacińskiego i z tego powodu znalazła się zarówno w edycji źródła, jak i w korpusie. Względnie wysoka liczba posiedzeń kapituł w kwietniu, lipcu i październiku sprawiła, że nazwy tych właśnie miesięcy – dość rzadko występujące w innych tekstach spoza podkorpusu – znalazły się w powyższej tabeli.

Ze względu na to, że badaniu poddawane są leksemy pojedyncze, a nie wielosegmentowe, w poniższym opisie szczególną uwagę zwrócono na rzeczowniki. Te natomiast tworzą w pewien spójny zbiór wyrazów specjalistycznych dotyczących administracji kościelnej. Skuteczność zastosowanej metody i specjalistyczny charakter wskazanych w jej wyniku leksemów potwierdzają ustalenia zawarte w *Słowniku łaciny średniowiecznej w Polsce* (SŁŚ).

CAPITULUM (poz. 2) to wyraz polisemiczny o proweniencji starożytnej, który w okresie średniowiecza oznaczał między innymi 'kapitułę, kolegium duchownych oraz ich posiedzenie'. Fakt, że występuje on względnie często w podkorpusie, a jednocześnie rzadko poza nim, pozwala wskazać go jako kandydata na wyraz specjalistyczny. Warto zaznaczyć, że wyraz ten w SŁŚ opatrzony został kwalifikatorem *eccl. t.t.* (kościelny termin techniczny).

Kolejne rzeczowniki to nazwy urzędników i dostojników kościelnych. Dwa leksemy – CANONICUS (poz. 3) i SCOLASTICUS (poz. 16) – mogły funkcjonować zarówno w wariancie rzeczownikowym (w znaczeniach odpowiednio 'kanonika, członka kapituły' i 'scholastyka, przełożonego szkoły katedralnej lub kolegiackiej', także 'członka kapituły'), jak i przymiotnikowym (w znaczeniach odpowiednio 'kanoniczny, obowiązujący na mocy prawa kościelnego' oraz 'szkolny, związany ze szkołą' i 'scholastyczny'). Analiza wystąpień tych wyrazów pokazuje, że w tekstach podkorpusu w przeważającej mierze funkcjonują one jako rzeczowniki. W SŁŚ leksemy te w wariantach rzeczownikowych opatrzone zostały kwalifikatorami odpowiednio *eccl. t.t.* oraz *eccl.* (termin kościelny).

ARCHIDIACONUS (poz. 4) – wyraz ten wywodzi się ze starożytnej łaciny chrześcijańskiej, w której prymarnie oznaczał 'zwierzchnika diakonów'. W wiekach średnich zaczął zaś funkcjonować przede wszystkim w znaczeniu 'dostojnika kościelnego wykonującego władzę jurysdykcyjną w pewnej części diecezji'. DECANUS (poz. 8) – wyraz również znany w starożytności – oznaczał przełożonego grupy dziesięciu i odnosił się zwłaszcza do żołnierzy. W łacinie średniowiecznej natomiast używany był zwłaszcza w kościelnym znaczeniu 'dziekana' (w SŁŚ z kwalifikatorem *eccl.*), tzn. 'prałata stojącego na czele kapituły', 'duchownego sprawującego nadzór nad kilkoma parafiami' lub 'przełożonego kolegium kardynalskiego'.

ADMINISTRATOR (poz. 7) – leksem pochodzenia starożytnego – występował w ogólnym znaczeniu 'kierownika, zarządcy'. W średniowieczu jednak kontekst jego użycia coraz bardziej się zawężał i – jak wskazuje SŁŚ – używano go zarówno w odniesieniu do prawa świeckiego, jak i kościelnego – między innymi w znaczeniu 'duchownego sprawującego rządy w zastępstwie niewybranego jeszcze lub nieobecnego biskupa' (w SŁŚ znaczenie z kwalifikatorem *eccl.*). Z podobnym mechanizmem specjalistycznego użycia leksemu ogólnego mamy do czynienia w wypadku wyrazów CANTOR (poz. 9) oraz CUSTOS (poz. 14) – oba pochodzenia starożytnego. Leksem CANTOR, ogólnie oznaczający 'śpiewaka', w kościelnej łacinie średniowiecznej używany był również na określenie 'kapłana prowadzącego chóry w kościele' (była to także godność w kapitule). Wyraz CUSTOS (ogólnie 'stróż, strażnik') w znaczeniu specjalistycznym służył zaś określeniu 'kustosza, prałata kapituły'. Obydwa leksemy we wskazanych specjalistycznych znaczeniach kościelnych otrzymały w SŁŚ kwalifikator *eccl. t.t.*

Wśród omawianych rzeczowników znalazł się także wyraz TRACTATUS (poz. 17) 'rada, opinia, umowa, uzgadnianie'. Analiza jego wystąpień w podkorpusie i korpusie ogólnym wskazuje jednak, że bę-

dzie to leksem niespecjalistyczny. Nie oznacza on bowiem jakiejś specjalnej rady kościelnej lub związanej ze sferą kościelną umowy. Uznać go można raczej za wyraz kluczowy – charakterystyczny dla badanego typu tekstów.

Poza omówionymi szerzej rzeczownikami na uwagę zasługują również wskazane w wyniku porównania przymiotniki: CAPITULARIS (poz. 12) 'kapitulny, należący do kapituły', GENERALIS (poz. 13) 'ogólny, powszechny', PROVINCIALIS (poz. 21) 'prowincjalny, dotyczący prowincji kościelnej' oraz ARCHIEPISCOPALIS (poz. 24) 'arcybiskupi, należący do arcybiskupa'. Analiza wyrazów najczęściej z nimi występujących (przeprowadzona z użyciem funkcji kolokacji aplikacji korpusowej) mogłaby pomóc w wyłonieniu złożonych jednostek specjalistycznych. Zagadnienie to powinno być jednak przedmiotem osobnego opracowania.

5. Podsumowanie

Celem niniejszego artykułu było przedstawienie korpusowej metody ekstrakcji słownictwa specjalistycznego na przykładzie kościelnej leksyki prawno-administracyjnej zawartej w średniowiecznych tekstach łacińskich. W wyniku porównania stworzonego dla potrzeb badań specjalistycznego podkorpusu tekstów związanych z prawem i administracją kościelną z całością Elektronicznego korpusu polskiej łaciny średniowiecznej udało się wyłonić leksemy, które uznać można za jednostki słownictwa specjalistycznego. Szczególną uwagę skupiono na jednostkach prostych, poddając analizie rzeczowniki. Wśród nich dominowały wyrazy znane ze starożytności, które w średniowieczu używane były zarówno w znaczeniu ogólnym, jak i specjalistycznym (np. CAPITULUM, ADMINISTRATOR, CANTOR, CUSTOS). Główne kryterium ich wyodrębnienia stanowiło porównanie ich frekwencji względnej w podkorpusie specjalistycznym i całości omawianego korpusu. Wyłonienie typowego dla analizowanych tekstów kościelnego słownictwa prawno-administracyjnego może okazać się przydatne zarówno dla językoznawców, w badaniach leksykograficznych nad terminologią kościelną w łacinie średniowiecznej, jak również dla historyków i prawników. Dzięki metodom automatycznym proces selekcji materiału zostaje usprawniony, co z kolei pozwala zwiększyć tempo i efektywność prowadzonych analiz.

Pomimo pewnych istniejących jeszcze – i na obecnym etapie prac nad projektem w pełni zrozumiałych – niedostatków Elektronicznego korpusu polskiej łaciny średniowiecznej zaprezentowana analiza wykazała efektywność przyjętej metody. Wprawdzie ze względu na różne czynniki jest ona obciążona pewnym ryzykiem błędu, niemniej jednak z powodzeniem – jak się zdaje – można wykorzystywać ją w badaniach nad słownictwem specjalistycznym.

Należy również podkreślić, że w wypadku analizy danych frekwencyjnych szczególnie istotna jest umiejętność ich właściwej interpretacji z zachowaniem zasady ograniczonego zaufania i późniejszą weryfikacją (McEnery & Hardie 2012: 48–49; Piotrowski & Grabowski 2013: 70). Owa weryfikacja przeprowadzona została na podstawie analizy kontekstów źródłowych oraz danych *Słownika łaciny średniowiecznej w Polsce*. Przedstawione wyniki w przyszłości zapewne jednak będą musiały zostać poddane dalszej weryfikacji. Do Elektronicznego korpusu polskiej łaciny średniowiecznej sukcesywnie dodawane są bowiem nowe teksty – planowane jest również poszerzenie bazy tekstów prawa kościelnego o kolejne wydania statutów synodalnych (między innymi diecezji chełmskiej, łuckiej, płockiej, poznańskiej, przemyskiej, wileńskiej i wrocławskiej). Tym samym także objętość podkorpusu specjalistycznego będzie mogła się

odpowiednio zwiększyć. Ważna dla prowadzonych analiz okaże się także poprawa funkcjonalności korpusu, w tym jego lemmatyzacji, co pozwoli uniknąć pewnej nieprecyzyjności otrzymywanych danych. Można jednak przypuszczać, że owe zmiany potwierdzą dotychczasowe ustalenia, umożliwiając prowadzenie dalszych, bardziej zaawansowanych analiz z wykorzystaniem przedstawionej w artykule metody.

Bibliografia

Biber, Douglas, Susan Conrad, Randi Reppen (1998) *Corpus Linguistics. Investigating Language Structure and Use*. New York: Cambridge University Press.

Buttler, Danuta (1979) „O wzajemnym oddziaływaniu terminologii i słownictwa ogólnego. I. Terminologizacja wyrazów potocznych". [W:] *Poradnik Językowy*. 2; 58–66.

Grochowski, Leonard (1977) „Dzieje ustawodawstwa kościelnego w średniowieczu (XII-XV w.)". [W:] Józef Keller (red.) *Katolicyzm średniowieczny*. Warszawa: PWN; 293–337.

Heylen, Kris, Dirk De Hertog (2015) „Automatic Term Extraction". [W:] Hendrik J. Kockaert, Frieda Steurs (red.) *Handbook of Terminology. Volume 1*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Kilgarriff, Adam (2009) „Simple Maths for Keywords. https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf [data dostępu: 29.9.2020].

Kłoczowski, Jerzy (red.) (1966) *Kościół w Polsce. T. 1. Średniowiecze*. Kraków: Znak.

Koester, Almut (2010) „Building Small Specialised Corpora". [W:] Anne O'Keeffe, Michael McCarthy (red.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge; 66–79.

Kumor, Bolesław, Zdzisław Obertyński (red.) (1974) *Historia Kościoła w Polsce. T. 1. Do roku 1764. Cz. 1. Do roku 1506*. Poznań, Warszawa: Pallotinum.

Manjavacas, Enrique, Ákos Kádár, Mike Kestemont (2019) „Improving Lemmatization of Non-Standard Languages with Joint Learning". [W:] *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Volume 1. Minneapolis: Association for Computational Linguistics; 1493–1503.

McEnery, Tony, Andrew Hardie (2012) *Corpus Linguistics. Method, Theory and Practice*. New York: Cambridge University Press.

Nowak, Krzysztof (2014) *„Fontes Mediae et Infimae Latinitatis Polonorum*. Z prac nad korpusem polskiej łaciny średniowiecznej". [W:] *Polonica*. XXXIV; 105–114.

Piotrowski, Tadeusz, Łukasz Grabowski (2013) „Interpretacja danych frekwencyjnych z korpusów językowych: opis pewnych problemów (na kilku przykładach z życia wziętych)". [W:] Wojciech Chlebda (red.) *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Opole: Wydawnictwo Uniwersytetu Opolskiego; 59–71.

Pontal, Odette (1975) *Les statuts synodaux*. Turnhout: Brepols.

Sondel, Janusz (2001) *Słownik łacińsko-polski dla prawników i historyków*. Kraków: Universitas.

Wójcik, Walenty (1969) „Kościelne ustawodawstwo partykularne w Polsce przedrozbiorowej na tle powszechnego prawodawstwa kościelnego". [W:] Piotr Kałwa, Marian Rechowicz (red.) *Księga tysiąclecia katolicyzmu w Polsce. Cz. 1.* Lublin: Wydawnictwo Towarzystwa Naukowego Katolickiego Uniwersytetu Lubelskiego; 423–502.

Zygner, Leszek (2013) „Późnośredniowieczne synody narzędziem reformy Kościoła". [W:] Tomasz Gałuszka, Tomasz Graff, Grzegorz Ryś (red.) *Ecclesia semper reformanda. Kryzysy i reformy średniowiecznego Kościoła*. Kraków: Societas Vistulana; 423–441.