

Instrukcja korzystania z wyszukiwarki do *Elektronicznego Korpusu Tekstów Polskich* z XVII i XVIII wieku

styczeń 2024

Spis treści

Wprowadzenie	1
1. Dwie wersje korpusu	2
2. Teksty w korpusie — warstwa transliteracji, transkrypcji i anotacji	2
3. Zestaw znaczników morfosyntaktycznych	3
3.1. Klasy gramatyczne	3
3.2. Kategorie gramatyczne	4
4. Interfejs wyszukiwarki MTAS	7
5. Język zapytań wyszukiwarki MTAS	9
5.1. Proste zapytania o segmenty i formy podstawowe	9
5.2. Wyrażenia regularne	10
5.3. Łączenie atrybutów	11
5.4. Zapytania o kilka pozycji	11
5.5. Zapytania o znaczniki morfosyntaktyczne	11
5.6. Ograniczenie zapytania do zdania lub akapitu	12
5.7. Zapytania o interpretacje Concrafta w KorBie 1.0	13
6. Ograniczenie zapytania za pomocą metadanych	13
7. Logowanie	15

Wprowadzenie

Niniejsza instrukcja opisuje sposoby korzystania z zasobów zgromadzonych w *Elektronicznym Korpusie Tekstów Polskich z XVII i XVIII w.* za pomocą wyszukiwarki MTAS. Dokument ten stanowi zmodyfikowaną wersję instrukcji opisującej sposób działania wyszukiwarki MTAS, opracowanej przez Witolda Kierasia na potrzeby *Korpusu tekstów polskich z lat 1830–1918*. Z kolei podstawą tej instrukcji była *Ściągawka do Narodowego Korpusu Języka Polskiego*, czyli instrukcja użytkownika wyszukiwarki Poliqarp, wykorzystującej, podobnie jak MTAS, język zapytań znany pod nazwą *Corpus Query Language (CQL)*. *Ściągawka* została opracowana przez Adama Przepiórkowskiego, a następnie poprawiona i rozszerzona przez Jakuba Wilka i Aleksandra Buczyńskiego. Jej pełna wersja znajduje się w repozytorium wyszukiwarki Poliqarp. Modyfikacje wprowadzone do pierwotnej wersji instrukcji uwzględniają różnice w języku zapytań oraz specyfikę korzystania z korpusu historycznego. Za zgodą wszystkich wyżej wymienionych autorów niniejsza wersja dokumentu zostaje udostępniona na zasadach licencji Creative Commons BY-SA.

1. Dwie wersje korpusu

Elektroniczny Korpus Tekstów Polskich z XVII i XVIII w. dostępny jest w dwóch wersjach: pierwotnej, oddanej do użytku w 2018 r. (KorBa 1.0) oraz rozbudowanej, udostępnionej w 2023 r. (KorBa 2.0). Pierwsza z nich jest pozostawiona, aby umożliwić dokończenie prowadzonych na niej badań; po upływie dwóch lat zostanie wycofana z użytku.

Obie wersje korpusu różnią się przede wszystkim rozmiarem i zasięgiem czasowym zgromadzonych w nich tekstów. KorBa 1.0 ma objętość 13,5 miliona segmentów i obejmuje teksty powstałe w latach 1601-1772. KorBa 2.0 to korpus prawie 27-milionowy, w którym zawarte są teksty z całych dwóch stuleci: XVII i XVIII. Wszystkie teksty zgromadzone w KorBie 1.0 znajdują się również w KorBie 2.0

Oprócz tego pomiędzy obiema wersjami korpusu występują różnice wynikające z zastosowania różnych narzędzi do automatycznej anotacji morfosyntaktycznej (tagerów). W KorBie 1.0 anotacja została przeprowadzona niezależnie przez dwa tagery: Concraft i Toygger. Na skutek tego ta wersja korpusu występuje w dwóch wariantach, które mogą się od siebie różnić na poziomie lematyzacji i przypisanych poszczególnym formom znaczników morfosyntaktycznych (więcej o tych wariantach w p. 5.7). Do automatycznej anotacji tekstów w KorBie 2.0 zastosowano tager KFTT.

Część różnic w znacznikach morfosyntaktycznych między KorBą 1.0 a 2.0 wynika również z wprowadzenia w drugiej wersji niewielkich zmian w tagsecie (patrz: p. 3) oraz z przyjęcia dodatkowych założeń znakowania.

2. Teksty w korpusie — warstwa transliteracji, transkrypcji i anotacji

Teksty w korpusie dostępne są w dwóch formach ortograficznych — w transliteracji i transkrypcji. W wypadku tekstów pozyskanych bezpośrednio z XVII- i XVIII-wiecznych druków lub rękopisów transliteracja oddaje ortografię podstawy. Drobne różnice dotyczą jedynie niektórych cech (typo)graficznych, np. nie uwzględniamy różnych form grafemów *s* i *r*, a ligatury oddajemy za pomocą osobnych liter (np. *β* jako *sz*). W tekstach pozyskanych z wydań późniejszych (filologicznych) wersja transliterowana odpowiada dokładnie temu, co znajduje się w wydaniu. Skutkiem tego jest np. występowanie w pewnych tekstach (pochodzących z wydań dziewiętnastowiecznych) grafemu *é*, który w zasadzie nie pojawiał się w drukach barokowych.

Wszystkie teksty poddano automatycznej transkrypcji, tzn. nadano im postać zgodną z dzisiejszą ortografią. W transkrypcji zostały jednak zachowane niektóre cechy fonetyczne i morfologiczne charakterystyczne dla średniopolszczyzny, np. forma imiesłowu przysłówkowego uprzedniego zapisana w oryginale *donioższy* pojawia się w transkrypcji w postaci *doniószy* (a nie *doniósłszy*). Transkrypcji nie podlegają fragmenty obcojęzyczne. Dotyczy to także pojedynczych wyrazów, ale użytych na prawach cytatu, np. *bom znał Qualitates tego*. Wyrazy pochodzenia obcego włączone w polski paradygmat odmiany traktujemy jak rodzime (np. transliteracja *Novembra*, transkrypcja *Nowembra*). W tekstach zarówno transliterowanych, jak i transkrybowanych zachowana została pisownia wielką lub małą literą zgodnie z oryginałem (lub podstawą).

Wszystkie teksty zostały poddane automatycznej anotacji morfosyntaktycznej, tzn. poszczególnym jednostkom tekstu (segmentom) zostały przypisane znaczniki określające ich formę podstawową (lemat), klasę gramatyczną oraz odpowiednie wartości kategorii gramatycznych. Przez segmenty rozumiemy zazwyczaj słowa tekstowe (ciągi liter ograniczone spacjami lub znakami interpunkcyjnymi), ale w niektórych wypadkach w obrębie słowa wyróżniamy kilka segmentów. Przede wszystkim wyszukiwarka ignoruje oryginalną segmentację tekstów barokowych, tzn. słowa zapisane niezgodnie ze współczesną ortografią mają segmentację uwspółcześioną, np. w ciągu *niefrasować* zostaną wyróżnione dwa segmenty: *[nie]* i *[frasować]*, a w ciągu *nioczym ‘o niczym’* — trzy segmenty: *[ni]*, *[o]*, *[czym]*.

Ponadto, analogicznie jak w NKJP, jako odrębne segmenty traktowane są tzw. ruchome końcówki osobowe czasowników, np. [*łgał*][*eś*], [*długo*][*śmy*], partykuły *by*, *-ż(e)*, *-li* i *-ć (-ci)*, np. [*przyszedeł*][*by*], [*pragnę*][*ć*], poprzyimkowa nieakcentowana forma zaimka ON, np. [*do*][*ń*] oraz w niektórych wypadkach poszczególne człony wyrazu zawierającego łącznik, np. [*melancholiczno*][*-*][*choleryczny*], [*puku*][*-*][*huku*]. Za segmenty uznajemy także znaki interpunkcyjne.

3. Zestaw znaczników morfosyntaktycznych

Każdy znacznik morfosyntaktyczny jest ciągiem wartości rozdzielonych dwukropkami. Przykładowo segmentowi *textem* (postać transliterowana) zostały przypisane następujące wartości: *tekst:subst:sg:inst:m*. Pierwsza wartość (*tekst*) to forma podstawowa (lemat) zapisana w transkrypcji, druga (*subst*) określa klasę gramatyczną, następnie zaś to wartości odpowiednich dla tej klasy kategorii gramatycznych. Repertuar klas i kategorii gramatycznych oparty jest na zbiorze klas i kategorii (tagsecie) stosowanym w analizatorze morfologicznym Morfeusz, został on jednak dostosowany do języka średniopolskiego.

3.1. Klasy gramatyczne

W *Elektronicznym Korpusie Tekstów Polskich z XVII i XVIII w.*, podobnie jak w NKJP, klasy gramatyczne są oparte na pojęciu fleksemu, będącym pojęciem węższym od terminu leksem. O ile tradycyjnie ujmowane leksemy mogą zawierać formy, którym przysługują różne kategorie gramatyczne¹, o tyle fleksemy stanowią zbiory tylko tych form, które można scharakteryzować za pomocą tych samych kategorii gramatycznych.

Tabela 1 zawiera listę wszystkich klas gramatycznych przyjętych w niniejszym tagsecie (w zestawieniu z tradycyjnymi częściami mowy) wraz z informacją o ich formach podstawowych oraz o ich symbolach używanych w korpusie². W stosunku do zestawu klas gramatycznych wyróżnionych w NKJP zostały tu zastosowane następujące modyfikacje:

- rezygnacja z traktowania form deprecjatywnych rzeczownika jako odrębnych fleksemów (różnica między formami typu *królowie* i *króle* została oddana za pomocą różnych wartości kategorii rodzaju — por. p. 3.2);
- wyróżnienie (w oparciu o kryterium semantyczne) dwóch dodatkowych fleksemów liczebnikowych: liczebnika przymiotnikowego (*adjnum*), np. *dwojaki*, pełniącego funkcję przymiotnika, i liczebnika przysłówkowego (*advnum*), np. *dwojako*, pełniącego funkcję przysłówka;
- zastąpienie dwóch fleksemów — przymiotnika poprzyimkowego (*adjc*) i przymiotnika predykatywnego (*adjp*) — fleksemem o nazwie *przymiotnik w odmianie niezłożonej* (*adjb*), zawierającym formy typu *polsku*, *zdrów*, *gwałtownę*;
- wyróżnienie użyc czasownika BYĆ w funkcji słowa posiłkowego czasów złożonych: przyszłego (*fut*), np. *będę* (*jechał*), i zaprzęsłego (*plusq*), np. *był* (*jechał*);
- wyróżnienie aglutynantu aorystycznego (*agltaor*), obejmującego dawne końcówki osobowe czasownika, typu *-(e)ch* i *-(e)chmy*;
- powiększenie zbioru imiesłów przymiotnikowych o imiesłów przeszły (*ppraet*), np. *osiwiały*, a także o imiesłowy w odmianie niezłożonej: czynny (*pacrb*), np. *będący*, i bierny (*ppasb*), np. *umęczon*.

¹ Przykładowo w obrębie leksemu czasownikowego można wyróżnić m.in. osobowe formy czasownika (odmienne przez osoby i liczby), bezokolicznik (nieodmienny), a w niektórych ujęciach także odsłownik (odmienny przez przypadki i liczby).

² Dwie ostatnie klasy: *ignndm* i *ign* mogą pojawić się wyłącznie w korpusie anotowanym ręcznie. W taki sposób zostały oznaczone formy, których funkcji i/lub postaci lematu anotatorzy nie byli w stanie rozpoznać.

Tabela 2 zawiera przybliżoną charakterystykę morfosyntaktyczną wyróżnionych w korpusie klas gramatycznych (z pominięciem tych klas, którym nie przysługują żadne kategorie gramatyczne). Symbol \oplus oznacza, że dla danej klasy gramatycznej dana kategoria gramatyczna jest morfologiczna (fleksemy należące do tej klasy „odmieniają się” przez tę kategorię), zaś symbol \odot oznacza, że dana kategoria jest słownikowa (wszystkie formy fleksemu należącego do tej klasy mają tę samą wartość tej kategorii³).

3.2. Kategorie gramatyczne

Tabela 3 przedstawia repertuar kategorii gramatycznych oraz ich wartości używanych w korpusie. W porównaniu z tagsetem NKJP występują tu następujące zmiany:

- rezygnacja z kategorii akomodacyjności form liczebników;
- dodatkowa wartość kategorii liczby — liczba podwójna (du)⁴;
- dodatkowa wartość kategorii aspektu — aspekt podwójny (bi_{asp}), przypisywany czasownikom dwuaspektowym (np. *abdykować*) oraz tym, których aspekt jest niemożliwy do ustalenia ze względu na niewystępowanie w korpusie kontekstów diagnostycznych;
- dodatkowa kategoria przyrodzaju dla rzeczowników *plurale tantum*, którym przypisujemy wartość pt (pozostałe rzeczowniki mają dla tej kategorii pustą wartość);
- trzy wartości w obrębie rodzaju męskiego — męski uogólniony (m), męski żywotny 1 (manim1), męski żywotny 2 (manim2)⁵ — patrz: komentarz poniżej.

W KorBie, odmiennie niż w NKJP, przypisujemy poszczególne wartości rodzaju męskiego formom wyrazowym, a nie całym leksemom. Ma to na celu oddanie zmian, jakim podlegała ta kategoria. W XVII i XVIII w. dwa istniejące współcześnie podrodzaje męskie — żywotny osobowy i żywotny nieosobowy (w NKJP odpowiednio m1 i m2) — dopiero się kształtowały, co przejawiało się w chwiejnej postaci M., B. i W. lm. rzeczowników żywotnych. Z tego względu przyjmujemy, że podział na podrodzaje (manim1 i manim2⁶.) zaznaczamy tylko w tych formach rzeczowników rodzaju męskiego, których końcówki pozwalają na wyróżnienie tych podrodzajów. Pozostałym formom przypisujemy wartość rodzaju męskiego uogólnionego (m). Ta sama zasada odnosi się do form przymiotników i liczebników uzgadniających swój rodzaj z rodzajem rzeczownika.

Wartość manim1 otrzymują więc formy M. i W. lm., takie jak: *panowie, ptacy, ważni, pięciu* w opozycji do form *pany, ptaki, stoły, ważne, pięć*, którym zostanie przypisana wartość m, oraz formy B. lm. takie jak: *panów, ptaków, nauczycieli, ważnych, pięciu* w opozycji do form *pany, ptaki, nauczyciele, stoły, ważne, pięć*, które będą miały wartość m. Natomiast wartość manim2 otrzymują formy B.lp. takie jak *pana, ptaka* w opozycji do formy typu *stół*, której przypiszemy wartość m. Jeśli w którymś z tych przypadków brak jest opozycji i możliwa jest tylko jedna forma (np. M. lm. *ludzie, kamienie*), to forma ta otrzymuje zawsze wartość rodzaju m. Rodzaj męski uogólniony (m) będą również miały formy takie jak *pan, ptakowi, stołom*.

³ Wyjątkiem jest tutaj kategoria rodzaju w odniesieniu do rzeczowników — poszczególne formy rzeczowników rodzaju męskiego otrzymują różne szczegółowe wartości tej kategorii (np. m i manim1) — patrz p. 3.2.

⁴ Jak wiadomo, niektóre formy liczby podwójnej zachowały się we współczesnej polszczyźnie i pełnią funkcję jedynych lub wariantywnych form liczby mnogiej (np. *ręce, rękoma*). Proces zmiany funkcji tych form z *dualis* na *pluralis* zaczął się już w XVI w. i trwał przez cały wiek XVII. Za graniczną datę zakończenia tej zmiany uznajemy arbitralnie rok 1740. A zatem w tekstach datowanych najpóźniej na 1740 r. formom typu *oczyrna, rękoma* przypisujemy wartość du, a w tekstach datowanych na 1741 r i późniejszych – wartość pl.

⁵ Opis rodzajów męskich w KorBie 2.0 jest nieco zmieniony w stosunku do KorBy 1.0.

⁶ Wartości manim1 i manim2 odpowiadają wyróżnianym w tagsecie NKJP wartościom m1 (męski osobowy) i m2 (męski żywotny). Zmiana tych oznaczeń jest celowa, ponieważ w XVII-XVIII w. formy manim1 nie musiały wskazywać na osobowe znaczenie rzeczownika (por. *ptacy, tygrysowie*).

Tradycyjne części mowy	Fleksem	Przykład	Symbol	Charakterystyka formy podstawowej	Forma podstawowa
rzeczowniki	rzeczownik	<i>woda, drzwi</i>	subst	M. l. poj. (<i>pl. tant. — l. mn.</i>)	WODA, DRZWI
liczebniki	liczebnik główny	<i>dwa, pięciu</i>	num	M. rodz. m	DWA, PIĘĆ
	liczebnik zbiorowy	<i>dwoje</i>	numcol	M. rodz. m	DWA
	liczebnik przymiotnikowy	<i>drugi, dwojaki</i>	adajnum	M. l. poj. rodz. m	DRUGI, DWOJAKI
	liczebnik przysłówkowy	<i>dwakroć, dwojako</i>	advnum	jedyna forma fleksemu	DWAKROĆ, DWOJAKO
przymiotniki	przymiotnik	<i>polski, dobry</i>	adj	M. l. poj. rodz. m st. równego odm. złoż.	POLSKI, DOBRY
	przymiotnik odm. niezłożona	<i>zdrow, polsku</i>	adjb	M. l. poj. rodz. m st. równego odm. złoż.	ZDROWY, POLSKI
	przymiotnik przyprzymiotnikowy	<i>polsko-</i>	adja	M. l. poj. rodz. m st. równego odm. złoż.	POLSKI
przysłówki	przysłówek	<i>dobrze, barzo</i>	adv	forma stopnia równego	DOBRCZE, BARZO
zaimki	zaimek nietrzecioos.	<i>ja, ty, my, wy</i>	ppron12	mianownik	JA, TY, MY, WY
	zaimek trzecioos.	<i>on</i>	ppron3	mianownik l. poj.	ON
	zaimek SIEBIE	<i>siebie</i>	siebie	biernik	SIEBIE
czasowniki	forma nieprzeszła	<i>czytam</i>	fin	bezokolicznik	CZYTAĆ
	forma przyszła BYĆ	<i>będę (tam)</i>	bedzie	bezokolicznik	BYĆ
	pseudoimiesłów	<i>czytał</i>	praet	bezokolicznik	CZYTAĆ
	rozkaźnik	<i>czytaj</i>	impt	bezokolicznik	CZYTAĆ
	bezosobnik	<i>czytano</i>	imps	bezokolicznik	CZYTAĆ
	bezokolicznik	<i>czytać</i>	inf	bezokolicznik	CZYTAĆ
	odśownik	<i>czytanie</i>	ger	bezokolicznik	CZYTAĆ
	imiesłów przysłówk. współczesny	<i>czytając</i>	pcon	bezokolicznik	CZYTAĆ
	imiesłów przysłówk. uprzedni	<i>(prze)czytawszy</i>	pant	bezokolicznik	(PRZE)CZYTAĆ
	imiesłów przymiot. czynny	<i>czytający, bołatszy</i>	pact	bezokolicznik	CZYTAĆ, BOLEĆ
	imiesłów przymiot. czynny odm. niezł.	<i>będący, jadący</i>	pactb	bezokolicznik	BYĆ, JECHAĆ
	imiesłów przymiot. bierny	<i>czytany</i>	ppas	bezokolicznik	CZYTAĆ
	imiesłów przymiot. bierny odm. niezł.	<i>umęczon</i>	ppasb	bezokolicznik	UMĘCZYĆ
	imiesłów przeszły	<i>osiwiał</i>	ppraet	bezokolicznik	OSIWIEĆ
	forma BYĆ — wykładnik cz. przyszłego	<i>będę (czytać)</i>	fut	bezokolicznik	BYĆ
	forma BYĆ — wykładnik cz. zaprzeszłego	<i>był (czytał)</i>	plusq	bezokolicznik	BYĆ
	aglutynant BYĆ	<i>-(e)m, -(e)śmy</i>	aglt	bezokolicznik	BYĆ
	aglutynant aoryst. BYĆ	<i>-(e)ch, -(e)chmy</i>	agлтаor	bezokolicznik	BYĆ
	czasownik typu WINIEN	<i>winiem, powinien</i>	winien	forma l. poj. rodz. m	WINIEN, POWINIEN
	predykatyw	<i>trzeba</i>	pred	jedyna forma fleksemu	TRZEBA
przyimki	przyimek	<i>na, przeze</i>	prep	forma niewokaliczna	NA, PRZEZ
spójniki	spójnik współrz.	<i>oraz, i, więc</i>	conj	jedyna forma fleksemu	ORAZ, I, WIĘC
	spójnik podrz.	<i>że, ponieważ</i>	comp	jedyna forma fleksemu	ŻE, PONIEWAŻ
partykuły	partykuła	<i>nie, -li, -że</i>	part	niewokaliczna forma fleksemu	NIE, -LI, -Ż
wykrzykniki	wykrzyknik	<i>ach, dlaboga</i>	interj	jedyna forma fleksemu	ACH, DLABOGA
inne	skrót	<i>r.</i>	brev	forma hasłowa rozwinięcia skrótu	ROK
	człon wyrażenia	<i>kolwiek</i>	frag	jedyna forma fleksemu	KOLWIEK
	znak interpunkcyjny	<i>;, !, ?</i>	interp	jedyna forma fleksemu	;, !, ?
	ciało obce	<i>item</i>	xxx	forma fleksemu występująca w tekście	ITEM
	zapis cyfr. (notacja arabska)	<i>1, 225</i>	dig	jedyna forma fleksemu	1, 225
	zapis cyfr. (notacja rzymska)	<i>MDCCLIII</i>	romandig	jedyna forma fleksemu	MDCCLIII
	wyraz nieodmienny o niejasnej funkcji		ignndm	jedyna forma fleksemu	
	wyraz o niejasnej funkcji i niejasnej lematyzacji		ign	forma fleksemu występująca w tekście	

Tabela 1. Klasy gramatyczne

	liczba	przypadek	rodzaj	przyrodz.	osoba	stopień	aspekt	zaneg.	akcent.	poprzyim.	aglutyn.	wokal.	kropk.
rzeczownik	⊕	⊕	⊙	⊙									
liczebnik główny	⊙	⊕	⊕										
liczebnik zbiorowy	⊙	⊕	⊙										
liczebnik przymiotnikowy	⊕	⊕	⊕			⊕							
liczebnik przysłówkowy						⊕							
przymiotnik	⊕	⊕	⊕			⊕							
przymiotnik odm. niezłożona	⊙	⊕	⊕			⊙							
przysłówek						⊕							
zaimek nietrzecioosobowy	⊙	⊕	⊕		⊙				⊕				
zaimek trzecioosobowy	⊙	⊕	⊕		⊙				⊕	⊕			
zaimek SIEBIE		⊕											
forma nieprzeszła	⊕				⊕		⊙						
forma przyszła BYĆ	⊕				⊕		⊙						
pseudoimiesłów	⊕		⊕				⊙				⊕		
rozkaznik	⊕				⊕		⊙						
bezosobnik							⊙						
bezokolicznik							⊙						
odśłownik	⊕	⊕	⊙				⊙	⊕					
im. przys. współczesny							⊙						
im. przys. uprzedni							⊙						
imiesłów przymiotnikowy czynny odm. złożona	⊕	⊕	⊕			⊕	⊙	⊕					
imiesłów przymiotnikowy czynny odm. niezłożona	⊙	⊙	⊙			⊙	⊙	⊕					
imiesłów przymiotnikowy bierny odm. złożona	⊕	⊕	⊕			⊕	⊙	⊕					
imiesłów przymiotnikowy bierny odm. niezłożona	⊙	⊕	⊕			⊙	⊙	⊕					
imiesłów przeszły	⊕	⊕	⊕			⊕	⊙	⊕					
forma BYĆ cz. przyszłego	⊕				⊕		⊙						
forma BYĆ cz. zaprzeczonego	⊕		⊕				⊙					⊕	
aglutynant BYĆ	⊕				⊕		⊙					⊕	
aglutynant aorystyczny BYĆ	⊕				⊕		⊙					⊕	
czasownik typu WINIEN	⊕		⊕				⊙						
przyimek		⊙										⊕	
partykuła												⊕	
skrót													⊕

Tabela 2. Charakterystyka morfosyntaktyczna klas gramatycznych

4. Interfejs wyszukiwarki MTAS

Interfejs wyszukiwarki zawiera pasek wyszukiwania, w którym wpisuje się zapytanie skonstruowane zgodnie ze składnią zapytań programu MTAS (por. p. 5). Po prawej stronie paska znajdują się przyciski umożliwiające wpisanie liter, które nie występują we współczesnej polskiej ortografii, ale mogą się pojawiać w warstwie transliteracyjnej przeszukiwanych tekstów.

Powyżej paska wyszukiwania znajdują się dwa okna. Okno **KORPUS** pozwala określić, który z dostępnych korpusów chcemy przeszukiwać. Do wyboru są cztery możliwości — dwa korpusy ukończone w 2018 r. (wersja 1.0) oraz ich rozbudowane edycje oddane do użytku w 2023 r. (wersja 2.0):

- **KORBA RĘCZNA 1.0** — 500-tysięczny podkorpus znakowany ręcznie, obejmujący teksty z lat 1601-1772;
- **KORBA AUTOMATYCZNA 1.0** — 13,5-milionowy korpus znakowany automatycznie tagami Concraft i Toygger, obejmujący teksty z lat 1601-1772;
- **KORBA RĘCZNA 2.0** — 800-tysięczny podkorpus znakowany ręcznie, obejmujący teksty z lat 1601-1800;
- **KORBA AUTOMATYCZNA 2.0** — pełny, 27-milionowy korpus znakowany automatycznie tagerem KFTT, obejmujący teksty z lat 1601-1800.

Domyślnie przeszukiwana jest najnowsza, pełna wersja korpusu.

Okno **RAMY CZASOWE** umożliwia ograniczenie przeszukiwania do tekstów powstałych w określonej epoce literackiej: w Baroku lub w Oświeceniu. Przyjmujemy, że podkorpus barokowy obejmuje teksty z lat 1601-1740, a podkorpus oświeceniowy — teksty z lat 1741-1800. Domyślnie przeszukiwane są teksty należące do obu tych epok (**CAŁOŚĆ**).

Poniżej paska wyszukiwania znajdują się trzy przyciski dające dodatkowe możliwości pracy z wyszukiwarką. Przycisk **KONSTRUKTOR ZAPYTAŃ** daje dostęp do prostej nakładki na wyszukiwarkę, która ułatwia zdefiniowanie warunków określających cechy segmentu lub sekwencji segmentów występujących w zapytaniu, np. postać segmentu, formę hasłową, klasę gramatyczną, wartości kategorii gramatycznych. Poszczególne warunki w obrębie segmentu mogą być łączone operatorami *oraz* (koniunkcja) i *lub* (alternatywa). Zbudowane za pomocą konstruktora zapytanie pojawi się w pasku wyszukiwania, dzięki czemu można zweryfikować jego poprawność.

Przycisk **ODRZUĆ OBCE SEGMENTY** spowoduje dodanie do zapytania ograniczenia, dzięki któremu wśród wyników wyszukiwania nie pojawią się segmenty obcojęzyczne równo kształtne z polskimi, np. łaciński wyraz *do* ‘daję’.

Przycisk **METADANE** umożliwia wprowadzanie do zapytania warunków ograniczających jego zasięg, na przykład do tekstów danego autora lub tekstów powstałych w danym przedziale czasowym (por. p. 6). Przycisk **DODAJ OGRANICZENIE** pozwala na łączenie poszczególnych warunków, można więc ograniczyć badanie np. do pism politycznych powstałych na terenie Małopolski w pierwszej połowie XVII w.

Wyniki wyszukiwania mogą wyświetlać się w postaci albo transliterowanej, albo transkrybowanej (uwspółcześnionej). Sposób prezentacji wyników wybieramy w oknie **WARSTWA WYŚWIETLANIA**. Domyślnie wyświetlana jest postać transliterowana. Na liście wyników, podobnie jak w NKJP, wyszukiwana sekwencja segmentów opatrzonych znacznikami zostaje wyróżniona kolorem, a po nakierowaniu na nią kursora pokazuje się dymek z podstawowymi informacjami o lokalizacji wyszukanego cytatu, z dokładnością do numeru strony książki będącej podstawą transliteracji.

Liczba		
pojedyncza	sg	<i>rzeka</i>
podwójna	du	M. B. W. (<i>dwa</i>) <i>szczyta, męża; (dwie)</i> <i>ręce, świecy; (dwie)</i> <i>ście, plecy;</i> D. Msc. (<i>dwu</i>) <i>panu, królu; kopu, niedzielu; latu, pokoleniu;</i> C.N. (<i>dwiema</i>) <i>mężoma, zakonoma; niewiastama, rzeczoma;</i> <i>latoma, plecoma</i>
mnoga	pl	<i>rzeki</i>
Przypadek		
mianownik	nom	<i>woda</i>
dopełniacz	gen	<i>wody</i>
celownik	dat	<i>wodzie</i>
biernik	acc	<i>wodę</i>
narzędnik	inst	<i>wodą</i>
miejsownik	loc	<i>wodzie</i>
wołacz	voc	<i>wodo</i>
Rodzaj		
męski (uogólniony)	m	<i>aktor, baran, dzban, aktory, barany, dzbany, nauczyciele;</i>
męski żywotny 1	manim1	<i>aktorzy, wilcy, aktorów (B.lm.), wilków (B. lm.);</i>
męski żywotny 2	manim2	<i>aktora (B. lp.), baranka (B. lp.)</i>
żeński	f	<i>stula</i>
nijaki	n	<i>dziecko, okno, co</i>
Przyrodzaj		
przymnogi	pt	<i>królestwo 'król i królowa', arcaby</i>
Osoba		
pierwsza	pri	<i>piszę, piszewa, piszemy</i>
druga	sec	<i>piszesz, piszeta, piszecie</i>
trzecia	ter	<i>pisze, piszeta, piszą</i>
Stopień		
równy	pos	<i>cudny</i>
wyższy	com	<i>cudniejszy</i>
najwyższy	sup	<i>nacudniejszy</i>
Aspekt		
niedokonany	imperf	<i>iść</i>
dokonany	perf	<i>zajść</i>
podwójny	biasp	<i>abdykować</i>
Zanegowanie		
niezanegowana	aff	<i>pisanie, czytanego</i>
zanegowana	neg	<i>niepisanie, nieczytanego</i>
Akcentowość		
akcentowana	akc	<i>niego, jego, tobie</i>
nieakcentowana	nakc	<i>go, -ń, ci</i>
zneutralizowana	zneut	<i>one, im, je</i>
Poprzyimkowość		
poprzyimkowa	praep	<i>niego, -ń</i>
niepoprzyimkowa	npraep	<i>jego, go</i>
Aglutynacyjność		
nieaglutynacyjna	nagl	<i>niósł</i>
aglutynacyjna	agl	<i>niósł-</i>
Wokaliczność		
wokaliczna	wok	<i>-em</i>
niewokaliczna	nwok	<i>-m</i>
Kropkwalność		
z następującą kropką	pun	<i>r</i>
bez następującej kropki	npun	<i>źł</i>

Tabela 3. Kategorie gramatyczne

Po kliknięciu na wybrany wynik na dole strony pokazuje się szerszy kontekst wraz ze szczegółowymi metadanymi tekstu, w tym z adresem strony internetowej, na której są umieszczone skany tekstu (o ile taka istnieje). Poniżej pojawia się informacja, czy dane słowo ma swój artykuł hasłowy w Elektronicznym Słowniku Języka Polskiego XVII i XVIII W. (e-SXVII). Jeśli tak, to wyświetla się skrócona wersja tego artykułu (część mowy oraz definicja znaczenia) oraz odnośnik do pełnego hasła w słowniku.

W oknie LICZBA WYNIKÓW NA STRONĘ można określić, ile wyszukiwanych pozycji ma być widocznych jednocześnie.

5. Język zapytań wyszukiwarki MTAS

Jak już było powiedziane, składnia zapytań w programie MTAS została oparta na języku zapytań o nazwie *Corpus Query Language* (CQL) wykorzystywanym m.in. w znanym z NKJP Poliqarpie. Należy jednak zwrócić uwagę na drobne różnice, ponieważ mogą one wpływać na poprawność formułowanych zapytań.

Niniejsza instrukcja nie uwzględnia wszystkich możliwości wyszukiwarki, częściowo ze względu na jej skrótowy charakter, a częściowo dlatego, że nie wszystkie zapytania będą miały sens w odniesieniu do korpusu XVII i XVIII wieku (mogą być natomiast użyteczne w tych korpusach, w których uwzględniono znakowanie warstw innego typu, np. nazw własnych, nadrzędników składniowych, sensu słów itp.). Podstawowa dokumentacja wyszukiwarki znajduje się na jej stronie internetowej.

5.1. Proste zapytania o segmenty i formy podstawowe

Zapytania wpisujemy w nawiasach kwadratowych według schematu:

```
[atrybut="wartość atrybutu"]
```

W najprostszycy pytaniach o kształt tekstowy segmentu atrybutem będzie `translit` — jeśli chcemy wyszukać segment w postaci transliterowanej, lub `orth` — jeśli poszukujemy segmentu w postaci transkrybowanej (uwspółcześnionej). Wartością obu tych atrybutów będzie poszukiwany ciąg liter, np.

```
[translit="sejm"]  
[orth="sejm"]
```

W wypadku zapytań o kształt segmentów w warstwie współczesnionej można pominąć nawiasy kwadratowe oraz nazwę atrybutu. Zatem poniższe zapytania:

```
[orth="sejm"]  
sejm
```

zwrócą takie same wyniki.

Domyślnie rozróżniana jest wielkość liter, a więc poniższe trzy zapytania:

```
[orth="sejm"]  
[orth="Sejm"]  
[orth="SEJM"]
```

dadzą różne wyniki. Aby znaleźć podane słowo niezależnie od wielkości poszczególnych liter, należy użyć atrybutu `orth` z rozszerzeniem `_lc` (ang. *lower case*):

```
[orth_lc="sejm"]
```

Wynikiem wszystkich powyższych zapytań będzie jedynie ciąg liter podany jako wartość atrybutu. Aby znaleźć wszystkie formy rzeczownika SEJM, należy użyć następującego zapytania:

[base="sejm"]

Wartością atrybutu base jest forma podstawowa szukanego fleksemu, a zatem mianownik liczby pojedynczej dla rzeczowników czy bezokolicznik dla wszystkich fleksemów czasownikowych. Formy podstawowe zapisane są wyłącznie w postaci współczesnionej, zatem zapytanie:

[base="sejm"]

zwróci wystąpienia różnych form fleksyjnych rzeczownika SEJM zapisane według różnych konwencji ortograficznych (np. *Sejm, seymu, sejmem*). Z kolei zapytanie:

[base="seym"]

nie zwróci żadnego wyniku. Aby znaleźć wszystkie formy rzeczownika SEJM zapisane przez y, należy dołączyć dodatkowy warunek na segment opisujący jego postać w warstwie transliterowanej (por. p. 5.4).

5.2. Wyrażenia regularne

W zapytaniach o segmenty i formy podstawowe (a także w innych zapytaniach, które zostaną opisane niżej) można używać standardowych wyrażeń regularnych wykorzystujących znaki specjalne, takie jak: !, ?, *, ., |, [,], (,) oraz liczby naturalne pisane cyframi arabskimi, np. 0 czy 21. Poniżej kilka przykładów zastosowania takich wyrażeń:

1. [translit="(sejm|seym)"]
znak | oznacza alternatywę dwóch wyrażeń (całość należy dodatkowo ująć w nawiasy okrągłe), a zatem zapytanie to może zostać użyte do znalezienia wszystkich wystąpień segmentów, które w warstwie transliteracyjnej mają postać *sejm* lub *seym*;
2. [orth="[Ss]ejm"]
nawiasy kwadratowe oznaczają alternatywę znaków, a zatem zapytanie to może zostać użyte do znalezienia wszystkich segmentów o postaci *sejm* niezależnie od wielkości litery początkowej;
3. [translit="komm?endant"]
znak zapytania oznacza opcjonalność znaku lub ujętego w nawiasy okrągłe wyrażenia bezpośrednio przed nim, a zatem w wyniku zadania tego zapytania znalezione zostaną segmenty *komendant* i *kommendant*;
4. [orth="k.ždy"]
kropka oznacza dowolny znak, a zatem wynikiem tego zapytania będą segmenty *každy* i *koždy*;
5. [orth="owa.*"]
gwiazdka oznacza dowolną liczbę wystąpień znaku lub ujętego w nawiasy okrągłe wyrażenia bezpośrednio przed nią, a zatem zapytanie to pozwoli znaleźć segmenty zaczynające się na *owa*, np. *owa, owaką, owakimi, owad, owada, owalną, owalaszyc*;
6. [orth="do{1,2}koła"]
konstrukcja typu *n,m* oznacza od *n* do *m* wystąpień znaku lub ujętego w nawiasy okrągłe wyrażenia bezpośrednio przed nią, a zatem zapytanie to pomoże znaleźć segmenty zaczynające się od litery *d*, po której następuje ciąg od 1 do 2 liter *o*, a następnie ciąg *koła*, a więc zarówno *dokoła*, jak i *dookoła*;
7. [orth="(pra){2,}.*"]
konstrukcja typu *n*, oznacza co najmniej *n* wystąpień znaku lub ujętego w nawiasy okrągłe wyrażenia bezpośrednio przed nią, a zatem zapytanie to może posłużyć do znalezienia segmentów, w których ciąg *pra* występuje przynajmniej 2 razy z rzędu, np. *praprababa, praprapradziad*. Specyfikacje segmentów muszą pasować do całych segmentów,

stąd konieczność umieszczenia po prawej stronie ciągu (pra){2,} wyrażenia .*, pasującego do dowolnego ciągu znaków.

UWAGA: Jeśli zadajemy pytanie o znak interpunkcyjny, który jest jednocześnie znakiem specjalnym używanym w wyrażeniach regularnych, należy poprzedzić go znakiem \, np.

```
[base="\." ]
```

5.3. Łączenie atrybutów

W zapytaniach o segmenty lub formy podstawowe segmentów (a także w innych zapytaniach, które zostaną opisane niżej) można określić kilka atrybutów, łącząc je operatorem koniunkcji & lub alternatywy |. W takich zapytaniach może się także pojawić operator negacji !. Poniżej kilka przykładów takich zapytań:

1. [base="sejm" & translit="seym.*"]
jako wynik otrzymamy wystąpienia wszystkich form rzeczownika SEJM zapisane przez y;
2. [orth="minę" & base="mina"]
w wyniku tego zapytania znalezione zostaną te segmenty, które są jednocześnie segmentem *minę* i formą rzeczownika MINA, a więc wyłącznie te wystąpienia segmentu *minę*, które są interpretowane jako formy rzeczownika MINA (a nie czasownika MINĄĆ);
3. [orth="minę" & !base="minąć"]
w wyniku tego zapytania znalezione zostaną te segmenty, które są jednocześnie segmentem *minę* i nie są formą czasownika MINĄĆ, a zatem zapytanie to jest równoważne poprzedniemu;
4. [orth="minę" | base="mina"]
w tym przypadku znalezione zostaną te segmenty, które są albo segmentem *minę* (niezależnie od tego, czy jest to forma fleksy MINA czy MINĄĆ), albo dowolną formą rzeczownika MINA, np. *mina*, *miny*, *minami*.

5.4. Zapytania o kilka pozycji

Jeśli chcemy wyszukać kilka sąsiadujących ze sobą segmentów lub form podstawowych, każdą pozycję zapisujemy w oddzielnych nawiasach kwadratowych, np.:

```
[base="sejm" ] [base="walny" ]
```

Oznaczając za pomocą pustych nawiasów kwadratowych dowolny segment, możemy znaleźć formy, które nie sąsiadują ze sobą bezpośrednio, np.:

```
[orth="się" ] [ ] [base="bać" ]
```

W wyniku tego zapytania zostaną znalezione takie ciągi, jak *się nic nie boi* czy *się o to boję*. Użycie wyrażen regularnych pozwala na określenie minimalnej i maksymalnej odległości pomiędzy szukanymi formami, np.:

```
[orth="się" ] [ ] {2,4} [base="bać" ]
```

W wyniku tego zapytania zostaną znalezione segmenty *się* oraz segmenty odpowiadające formom czasownika BAĆ przedzielone dwoma, trzema lub czterema segmentami, a zatem oba ciągi uzyskane w wyniku poprzedniego zapytania, a także na przykład ciąg *się już niczego nie boją*.

5.5. Zapytania o znaczniki morfosyntaktyczne

W zapytaniach można również określać wartości klasy gramatycznej (za pomocą atrybutu pos, ang. *part of speech*), a także wartości poszczególnych kategorii gramatycznych,

np. przypadku czy rodzaju. Wartościami atrybutu `pos` są symbole nazw klas gramatycznych znajdujące się w tabeli 1. Atrybuty poszczególnych kategorii gramatycznych i ich wartości wymienione są w tabeli 4.

Kategoria	Atrybut	Możliwe wartości
liczba	number	sg du pl
przypadek	case	nom gen dat acc inst loc voc
rodzaj	gender	m manim1 manim2 f n
przyrodzaj	subgender	pt
osoba	person	pri sec ter
stopień	degree	pos com sup
aspekt	aspect	imperf perf biasp
zanegowanie	negation	aff neg
akcentowość	accentability	akc nakc zneut
poprzyimkowość	post-prepositionality	npraep praep
aglutynacyjność	agglutination	agl nagl
wokaliczność	vocalicity	nwok wok
kropkwalność	fullstoppedness	pun npun

Tabela 4. Atrybuty kategorii gramatycznych i ich wartości

W zapytaniach tego typu również można używać wyrażeń regularnych. Możliwe jest zadanie na przykład następujących zapytań:

1. `[number="sg"]`
znalezione zostaną wszystkie formy w liczbie pojedynczej,
2. `[pos="subst" & number="sg"]`
znalezione zostaną formy rzeczownikowe w liczbie pojedynczej,
3. `[pos="subst" & !gender="f"]`
znalezione zostaną formy rzeczownikowe rodzaju męskiego, nijakiego lub przymnogiego,
4. `[number="sg" & case="nom|acc" & gender="n"]`
znalezione zostaną pojedyncze mianownikowe lub biernikowe formy nijakie.

O klasy gramatyczne i kategorie gramatyczne można także pytać łącznie, używając do tego atrybutu `tag`. Na przykład, aby znaleźć wszystkie rzeczowniki żeńskie w mianowniku liczby pojedynczej, można zadać następujące zapytanie:

```
[tag="subst:sg:nom:f"]
```

Aby zadać pytanie o pełną interpretację, czyli łącznie o formę podstawową, klasę gramatyczną i przysługujące jej kategorie gramatyczne, należy określić wartości dwóch atrybutów: `base` i `tag`, np.:

```
[base="baba" & tag="subst:sg:nom:f"]
```

W wyniku zadania tego zapytania znalezione zostaną wszystkie wystąpienia rzeczownika BABA w mianowniku liczby pojedynczej.

5.6. Ograniczenie zapytania do zdania lub akapitu

Teksty zawarte w korpusie zostały podzielone na zdania i akapity. Informację tę można wykorzystać w zapytaniach, na przykład ograniczając dopasowanie do jednego zdania. Aby ograniczyć zasięg zapytania, należy dopisać do zapytania słowo kluczowe `within`, a po nim `<s/>` lub `<p/>`, w zależności od tego, czy zasięg ma być ograniczony do zdania (ang. *sentence*) czy do akapitu (ang. *paragraph*). Ilustruje to następujący przykład zapytania o zdania,

w których forma *się* występuje za formą leksemu БАĆ, w odległości co najmniej jednego i nie więcej niż dziesięciu segmentów:

```
[base="bać"] [!orth="się"]{1,10}[orth="się"] within <s/>
```

Dodatkowo można również na elementy <s/> i <p/> nałożyć pewne warunki dotyczące tego, czy zawierają segmenty innego typu. Przykładowo, za pomocą następującego zapytania można znaleźć wszystkie wystąpienia czasownika *być* występującego w funkcji słowa posiłkowego czasu przeszłego złożonego ograniczone do zdań zawierających formę bezokolicznika:

```
[pos="fut"] within (<s/> containing [pos="inf"])
```

Wśród wyników będą oczywiście również takie zdania, w których czas przeszły został utworzony z formy pseudoimiesłowu, a bezokolicznik pełni w zdaniu inną funkcję gramatyczną.

5.7. Zapytania o interpretacje Concrafta w KorBie 1.0

Jak zaznaczono w p. 1., w KorBie 1.0 anotacja została przeprowadzona niezależnie przez dwa tagery: Concraft i Toygger. Domyślnie w wyszukanych wynikach pojawiają się interpretacje Toyggera, zaś interpretacje Concrafta wyświetlają się w dymku zawierającym podstawowe informacje o szukanej formie. Jeśli chcemy wyszukać formy z interpretacją Concrafta, używamy nazw atrybutów poszerzonych o *_c*, a więc *base_c*, *pos_c*, *tag_c*, *number_c*, *case_c* itd. Istnieje także możliwość porównywania interpretacji obu tagerów. W tym celu w jednym zapytaniu należy określić atrybuty odnoszące się do obu interpretacji z osobna. Przykładowo, jeśli chcemy wyszukać wszystkie segmenty, którym Concraft przypisał interpretację rzeczownikową, a Toygger dowolną inną, zadamy następujące pytanie:

```
[pos_c="subst" & !pos="subst"]
```

Podobnie jak w wypadku innych wyszukiwań interpretacje Toyggera pojawią się przy znalezionych formach, a interpretacje Concrafta będą widoczne w dymku po nakierowaniu kursora na daną formę.

6. Ograniczenie zapytania za pomocą metadanych

Wszystkie teksty w korpusie zostały opatrzone bogatymi metadanymi, czyli informacjami o autorstwie, tytule, miejscu wydania, gatunku, temacie itd. Wyszukiwarka umożliwia ograniczanie wyników do wybranych tekstów za pomocą tych metadanych. W oknie META-DANE można wybrać następujące pola:

- Skrót tekstu (identyfikator tekstu)
- Data (wydania lub powstania tekstu)
- Autor
- Tytuł
- Miejsce (wydania lub powstania tekstu)
- Region
- Typ mowy (proza lub wiersz)
- Rodzaj
- Gatunek
- Tematyka
- Poetyka żartu
- Drukarnia

- Tłumacz
- Tekst uwspółcześniony (teksty oryginalne — pochodzące z rękopisów i wydań XVII-XVIII-wiecznych, lub teksty uwspółcześnione — pochodzące z późniejszych wydań filologicznych)
- Typ źródła (rękopis, starodruk lub tekst uwspółcześniony)⁷
- Projekt (KorBa 1 — teksty włączone w pierwszej edycji projektu, lub KorBa 2 — teksty włączone w drugiej edycji projektu)⁸

W oknie OGRANICZENIE pojawiają się różne sposoby określania metadanych, zdefiniowane odmiennie dla poszczególnych pól. Ograniczenia można wprowadzać za pomocą następujących wyrażeń:

1. Wyrażenie „zawiera” — dostępne dla pól SKRÓT TEKSTU, AUTOR, TYTUŁ, MIEJSCE, DRUKARNIA, TŁUMACZ. W tym przypadku wielkość liter nie ma znaczenia. Można korzystać z wyrażeń regularnych. Wyszukiwarka znajduje teksty zawierające w określonym polu podany ciąg liter, np.:
 - TYTUŁ zawiera *akademia* — otrzymujemy tekst o tytule *Akademia dziecinna*.
 - TYTUŁ zawiera *akad dziec* — wynik jest taki sam, jak powyżej.
 - TYTUŁ zawiera *akad* — wynikiem będą teksty *Akademia dziecinna* i *Notyfikacja o niniejszym królewskiej akademii rycerskiej...*
2. Wyrażenie „nie zawiera” — działa analogicznie do wyrażenia „zawiera”. Wyszukiwarka pomija teksty zawierające w określonym polu podany ciąg liter, np.:
 - TYTUŁ nie zawiera *akademia* — otrzymujemy wszystkie teksty z wyjątkiem tych, które w tytule mają słowo *akademia*.
3. Wyrażenie „=” — dostępne dla pól SKRÓT TEKSTU, DATA, AUTOR, TYTUŁ. W tym przypadku istotna jest wielkość liter. Można korzystać z wyrażeń regularnych, np.:
 - SKRÓT TEKSTU = *AkDziec* — otrzymujemy tekst oznaczony skrótem *AkDziec*.
 - SKRÓT TEKSTU = *.*Dzie.** — otrzymujemy teksty oznaczone skrótami zawierającymi ciąg liter *Dzie*, a więc *AkDziec*, *KwiatDzieje* itp.
 - AUTOR = *Zbigniew Morsztyn* — otrzymujemy teksty autorstwa Zbigniewa Morsztyna.
 - AUTOR = *.*Morsztyn* — otrzymujemy teksty, których autorami byli Hieronim Morsztyn, Jan Andrzej Morsztyn, Stanisław Morsztyn lub Zbigniew Morsztyn.
 - AUTOR = *(Zbigniew Morsztyn)|(Stanisław Morsztyn)* — otrzymujemy teksty, których autorami byli albo Zbigniew Morsztyn, albo Stanisław Morsztyn. W przypadku gdy łączymy znakiem alternatywy kilka członów wielowyrazowych, każdy z nich należy wziąć w nawias okrągły.
 - DATA = *1633* — otrzymujemy teksty, które były wydane (lub napisane, w wypadku rękopisów) w roku *1633*.
 - DATA = *16.** — otrzymujemy teksty, które były wydane (lub napisane, w wypadku rękopisów) w latach *1601–1699*.
4. Wyrażenie „≠” — działa analogicznie do wyrażenia „=” . Wyszukiwarka pomija teksty, których odpowiednie pola zostały dokładnie określone, np.:
 - SKRÓT TEKSTU \neq *AkDziec* — otrzymujemy wszystkie teksty z wyjątkiem *Akademii Dziecinnej*.
5. Wyrażenie „<” — dostępne tylko dla pola DATA. Wyszukiwarka znajduje teksty, które zostały wydane (lub napisane, w wypadku rękopisów) przed określoną datą, np.:

⁷ Jak widać, to pole jest bardziej precyzyjną wersją poprzedniego. Zostało wprowadzone w Korbie 2.0, w której starodruki i rękopisy zostały oznakowane osobno. Nie stosuje się go do przeszukiwania KorBy 1.0.

⁸ To pole zostało dodane czasowo na potrzeby dokończenia badań rozpoczętych w pierwszej edycji projektu.

- DATA < 1635 — otrzymujemy teksty o datach określonych dokładnie (np. 1630) i w przybliżeniu (np. przed 1630).
6. Wyrażenie „>” — dostępne tylko dla pola DATA. Wyszukiwarka znajduje teksty, które zostały wydane (lub napisane, w wypadku rękopisów) po określonej dacie, np.:
- DATA > 1670 — otrzymujemy teksty o datach określonych dokładnie (np. 1680) i w przybliżeniu (np. po 1690).
7. Wyrażenie „z przedziału” — dostępne tylko dla pola DATA. Wyszukiwarka znajduje teksty, które zostały wydane (lub napisane, w wypadku rękopisów) w określonym przedziale czasowym. Wymaga podania dwóch dat oddzielonych myślnikiem, np.:
- DATA z przedziału 1633–1670 — otrzymujemy teksty o datach określonych dokładnie (np. 1654) i w przybliżeniu (np. między 1635 a 1644).

Dla pól, które mają ograniczoną liczbę wyborów (np. REGION), wystarczy wybrać pożądane wartości z rozwijanej listy.

Uwaga: istnieją pewne różnice w metadanych przypisanych tekstom w pierwszej i drugiej wersji korpusu. W drugiej edycji projektu zostały dodane nowe wartości np. w polu GATUNEK czy TEMATYKA. Zmieniono też kwalifikację niektórych tekstów z pierwszej edycji. W rzadkich wypadkach wartości istniejące w KorBie 1.0 nie mają zastosowania do tekstów z KorBy 2.0, np. gatunek *Umowy* z KorBy 1.0 odpowiada kilku bardziej precyzyjnie określonym gatunkom w KorBie 2.0.

Poszczególne ograniczenia można łączyć za pomocą przycisku DODAJ OGRANICZENIE.

7. Logowanie

Użytkownicy zalogowani zyskują trzy udogodnienia w korzystaniu z korpusu. Po pierwsze, ich wyniki wyszukiwania nie są ograniczone do 500. Po drugie, mogą pobrać otrzymane wyniki do pliku XLS lub CSV za pomocą przycisku PÓBIERZ WYNIKI znajdującego się pod wynikami wyszukiwania. Po trzecie, mogą pobrać pełną listę frekwencyjną lematów (zakładka „Listy frekwencyjne”). Prośbę o login i hasło należy wysłać na adres: korba@ijp.pan.pl.

Opracowanie: Włodzimierz Gruszczyński i Renata Bronikowska