

# Łacina w KorBie

Co ciekawego znajdzie latynista w korpusie polskich tekstów barokowych?

Renata Bronikowska  
Instytut Języka Polskiego PAN

[renata.bronikowska@ijp.pan.pl](mailto:renata.bronikowska@ijp.pan.pl)

Katarzyna Kryńska  
Instytut Języka Polskiego PAN

[katarzyna.krynska@ijp.pan.pl](mailto:katarzyna.krynska@ijp.pan.pl)



# Finansowanie projektu

Prezentacja powstała w ramach projektu pt.

## **Rozbudowa Elektronicznego Korpusu Tekstów Polskich XVII i XVIII w. i jego integracja z *Elektronicznym słownikiem języka polskiego XVII i XVIII w.***

finansowanego z funduszu Narodowego Programu Rozwoju Humanistyki MNiSW i realizowanego w latach 2019–2023 pod kierownictwem W. Gruszczyńskiego w IJP PAN we współpracy z Zespołem Inżynierii Lingwistycznej IPI PAN.

# Elektroniczny Korpus Tekstów Polskich z XVII i XVIII Wieku

- ◆ KorBa (Korpus Barokowy)
- ◆ dostępny od 2018 r.: <http://korba.edu.pl/>
- ◆ zawartość:
  - obecnie 13,5 mln segmentów (ok. 700 tekstów z lat 1601–1772)
  - docelowo 25 mln segmentów (w tym ok. 4 mln z lat 1773–1800)
- ◆ teksty zróżnicowane chronologicznie, geograficznie, gatunkowo i tematycznie
- ◆ bogate metadane
- ◆ teksty dostępne w postaci transliteracji i transkrypcji
- ◆ warstwy anotacji:
  - strukturalna: odwzorowanie struktury dokumentu
  - językowa: oznaczenie fragmentów obcojęzycznych
  - morfosyntaktyczna: lematyzacja + wartości kategorii gramatycznych
- ◆ wyszukiwarka MTAS: język zapytań CQL + konstruktor zapytań

# Łacina w KorBie - możliwości i ograniczenia badawcze

- ◆ Ok. 0,5 mln segmentów łacińskich (4% objętości korpusu)
- ◆ Możliwość wyszukiwania segmentów łacińskich
- ◆ Wyrażenia regularne
- ◆ Metadane
- ◆ Dokładna lokalizacja cytatu w źródle
  
- ◆ Pominięto dłuższe fragmenty łacińskie
- ◆ Brak lematyzacji i anotacji gramatycznej form łacińskich
- ◆ Nerozwiazane skróty, np. `quiq<quique>`
- ◆ Niekonsekwencja w anotacji nazw własnych i cyfr arabskich

# Błędy na różnych poziomach anotacji

## ◆ Błędy w znakowaniu ręcznym

- **Item:** czego ludzie lubo najdroższego y najwspánialszego nie zazdroszczą?
- a pamiętasz, kontuszu, **quando feresia fuisti, nuncmodo żupan eris, postea curta brevis**

## ◆ Błędy w znakowaniu automatycznym

- Co náder życzliwa dostoieństwu Páńskiemu wiará, y cáłości Oyczyzny miłość, y **resolutia** Wm. M. M. Páná ná ten czas robiła
- Tym niepozwálam **foventur dyssensye**, zawziętości miedzy Synámi Oyczyzny

# Jak interpretować wyniki wyszukiwania?

- ◆ Forma “item” – 2791 wyników
- ◆ Forma “item” + ograniczenie wyszukiwania do segmentów łacińskich – 1744 wyniki
- ◆ Forma “item” + wyłączenie segmentów łacińskich – 1047 wyników
- ◆ Forma “item” + odrzucenie obcych segmentów – 0 wyników

Tagery zawsze traktują formę “item” jako obcojęzyczną (znacznik “xxx”), ale nie przypisują jej znacznika łaciny (“la”).

# Ograniczenie przeszukiwania do segmentów łacińskich

Korpus  
Korba automatyczna ▼

Zapytanie  
[orth="armata"&foreign="la"]

á é Ā Ę

KONSTRUKTOR ZAPYTAŃ ODRZUĆ OBCE SEGMENTY METADANE ▼

Warstwa wyświetlania  
transliterowana ▼

Liczba wyników na stronie  
10 ▼

Wyszukaj

Znaleziono 14 wyników.

Lp	Lewy kontekst	Rezultat	Prawy kontekst	Skrót	Data
1	się zbił i pas jeden axamitny zerwał. Cechy manu	<a href="#">armata [armata:xxx]</a>	koło mar szy, których pięć przystojnymi całónami okryto,	DiarPogKoniec	1646
2	nad słabszą, ieżeli nie pluralitas votorum, á zwłaszcza	<a href="#">armata [armata:xxx]</a>	, ktorey słabsza nie może resistere, lubo przy rowney	LeszczStGłos	1733
3	z świecami assistują, w tyle koniuszy jedzie, potym	<a href="#">armata [armata:xxx]</a>	sequitur. Kiedy zaś Ojciec ś. w Watykanie umiera	KronZakBarącz	między 1703 a 1756
4	/ y zástąpienie ná dobrowolney drodze we szrod białego dnia	<a href="#">armata [armata:xxx]</a>	manu/ y vderzenie ono ták ná lezuity wyżey pomienione	SzemGrat	1627
5	do tak gwałtownego środka? I jużż temu gwoli zjazdy	<a href="#">armata [armata:xxx]</a>	manu stroić, szkalowac na nich pana, inauditum,	ZdanieSzlachCz_II	1606
6	mnie albo Sądowi memu Grodzkiemu Nowomiejskiemu admota manu forti et	<a href="#">armata [armata:xxx]</a>	motae Nobilitatis Powiatów, do Jurisdictey moeiy należących. Zaczym	ReyWKoniec	1663
7	miesiąca terazniejszego. A przedtym wprzód zjechać się mamy manu	<a href="#">armata [armata:xxx]</a>	i z żołnierzem 21 tegoż miesiąca pod Sieradz i	AktaPozn_I_1	między 1601 a 1616

# Wyszukanie wszystkich segmentów łacińskich w danym tekście

Korpus  
Korba automatyczna

Zapytanie  
[foreign="la"]

á é Ā Ę

KONSTRUKTOR ZAPYTAŃ ODRZUĆ OBCE SEGMENTY METADANE

Metadane  
Skrót tekstu

Ograniczenie  
=

Zapytanie o metadane  
NiewiesKom

DODAJ OGRANICZENIE

Warstwa wyświetlania  
transliterowana

Liczba wyników na stronę  
10

Wyszukaj

Znaleziono 393 wyników.

Lp	Lewy kontekst	Rezultat	Prawy kontekst	Skrót	Data
1	Drukowano w ZAMOSCIVRoku 1681 / w Mieście Lutym.	<a href="#">Omni [Omni:xxx]</a>	saeculo experiētiā comprobatum, Cometas, aliasq; apparentias de	NiewiesKom	1681
2	w ZAMOSCIVRoku 1681 / w Mieście Lutym. Omni	<a href="#">saeculo [saeculo:xxx]</a>	experiētiā comprobatum, Cometas, aliasq; apparentias de rarò	NiewiesKom	1681
3	ZAMOSCIVRoku 1681 / w Mieście Lutym. Omni saeculo	<a href="#">experiētiā [experiētiā:xxx]</a>	comprobatum, Cometas, aliasq; apparentias de rarò contingentes	NiewiesKom	1681
4	Roku 1681 / w Mieście Lutym. Omni saeculo experiētiā	<a href="#">comprobatum [comprobatum:xxx]</a>	Cometas, aliasq; apparentias de rarò contingentes	NiewiesKom	1681



# Wyszukiwanie kilkuwyrazowych fraz łacińskich

Korpus

Korba automatyczna

Zapytanie

[foreign="la"][foreign="la"][foreign="la"][foreign="la"][foreign="la"]

á é Á É

KONSTRUKTOR ZAPYTAŃ

ODRZUC OBCE SEGMENTY

METADANE

Warstwa wyświetlania

transliterowana

Liczba wyników na stronie

10

Wyszukaj

Znaleziono 13 wyników.

Lp	Lewy kontekst	Rezultat	Prawy kontekst	Skrót	Data
1	ku Pułnocy: gdzie dnia 28. Grud: wieczorem	<a href="#">/[:interp] distabat [distabat:xxx] ab [ab:xxx] Ecliptica [Ecliptica:xxx] ná [na:prep:acc]</a>	15. niemal stopni/ także ku Pułnocy etc.	NiewiesKom	1681
2	. Musiała tedy zginąć całe około dnia 10. Lutego	<a href="#">/[:interp] penes [penes:xxx] Caput [Caput:xxx] Medusae [Medusae:xxx] :[:interp]</a>	bo iey tak bieg przypadał/ y ku ostatkowi/	NiewiesKom	1681
3	albo blisko nich głowę swoje prowadził/ te są:	<a href="#">gwiazda [gwiazda:subst:sg:nom:f] in [in:xxx] manu [manu:xxx] Antinoi [Antinoi:xxx] ,[:interp]</a>	natury =, y =. Gwiazdy Delfinowe/ natury	NiewiesKom	1681
4	okazałe/ jedná in Capite, drugá in humero,	<a href="#">trzecia [trzeci:adnum:sg:nom:f:pos] in [in:xxx] cingulo [cingulo:xxx] Andromedae [Andromedae:xxx] ,[:interp]</a>	natury =, y =. Ogon iey zakrywał te	NiewiesKom	1681
5	przynosi effectá. A zaś gdy w przeciwnym znaku/	<a href="#">albo [albo:coni] in [in:xxx] trigono [trigono:xxx] signi [signi:xxx] ktorego [który:adj:sg:gen:n:pos]</a>	Páństwa świeci; temu iuz nie tak bárzo szkodzi:	NiewiesKom	1681
6	/ aby ná daremne od Bogá y Natury (ktore	<a href="#">oboie [oba:numcol:pl:nom:p1] nihil [nihil:xxx] faciant [faciant:xxx] frustra [frustra:xxx] )[:interp]</a>	były sporządzone; niechay nikt tego niebędzie rozumienia.	NiewiesKom	1681
7	Krolestwá rozpierác zwykły. Do tey Materyey/ kwadruie sie	<a href="#">y [i:coni] Oraculum [Oraculum:xxx] Sybillae [Sybillae:xxx] Tiburtinae [Tiburtinae:xxx] ,[:interp]</a>	ktorey są te słowa: Pokáże się Gwiazda w Ewropie	NiewiesKom	1681

# Wyszukiwanie użyć *ad + accusativus gerundium*

Korpus

Korba automatyczna

Zapytanie

[orth="[Aa]d"][]?{orth=".\*(andum|endum)"}]

á

é

Á

É

KONSTRUKTOR ZAPYTAŃ

ODRZUĆ OBCE SEGMENTY

METADANE

Warstwa wyświetlania

transliterowana

Liczba wyników na stronie

10

Wyszukaj

Znaleziono 260 wyników.

Lp	Lewy kontekst	Rezultat	Prawy kontekst	Skrót	Data
1	/ jáko jeden Pers mowi: Non opus est tormentis	<a href="#">ad [ad:xxx] exquirendum [exquirendum:xxx]</a>	veritatem: Vinum prodit secreta: Nie trzebá/ pry	GdacKon	1681
2	hetmana chorągwie tatarskie aggressus pobrał i powiązał. Designati Commissarii	<a href="#">ad [ad:xxx] tractandum [tractandum:xxx]</a>	od króla imć tak z jpanem chorążym i z wojskiem	ZawiszaPam	między 1715 a 1717
3	boku z województw i ziem, aby deputaci przysłani byli	<a href="#">ad [ad:xxx] consulendum [consulendum:xxx]</a>	et decidendum. Te uniwersaly commoverunt konfederatów, pisali do	ZawiszaPam	między 1715 a 1717
4	monarchom jakoby contra Baal flectere genu), dedit alas	<a href="#">ad [ad:xxx] festinandum [festinandum:xxx]</a>	iter, do jak najprędszego złączenia armorum, et consiliorum	ZawiszaPam	między 1715 a 1717
5	tacendum ori, ten chyba zaniedbać może tak pogodnej okazyi	<a href="#">ad [ad:xxx] eluctandum [eluctandum:xxx]</a>	z gwałtownej niebezpieczeństwa toni, panowania najjaśniejszego króla imci Augusta	ZawiszaPam	między 1715 a 1717
6	, wzruszona prowincya wielkiego księstwa litewskiego tantis extremis malorum,	<a href="#">ad [ad:xxx] succursum [succursum:xxx] dandum [dandum:xxx]</a>	braci, początkiem czego powiat oszmiański, czy to vero	ZawiszaPam	między 1715 a 1717

# Lista frekwencyjna - informacje ogólne

- ◆ lista form tekstowych w postaci transliteracji
- ◆ formy zapisywane małymi i wielkimi literami liczone łącznie
- ◆ ignorowane znaki diakrytyczne nad samogłoskami
- ◆ bez cyfr arabskich, chyba że forma składa się z cyfr i liter, np. *2do*
- ◆ bez znaków interpunkcyjnych
- ◆ 77 256 form

# Lista frekwencyjna - statystyka użycia form łacińskich

1	<b>in</b>	13366
2	<b>et</b>	12002
3	<b>ad</b>	4842
4	<b>non</b>	4105
5	<b>de</b>	3700
6	<b>ex</b>	2852
7	<b>est</b>	2847
8	<b>etc</b>	2617
9	<b>cum</b>	2412
10	<b>alias</b>	2100

11	<b>pro</b>	1858
12	<b>per</b>	1753
13	<b>item</b>	1746
14	<b>a</b>	1745
15	<b>lib</b>	1266
16	<b>qui</b>	1265
17	<b>cap</b>	1210
18	<b>sub</b>	1088
19	<b>ut</b>	1051
20	<b>anno</b>	995

*Anno et die ut supra* - 27  
wystąpień

21	<b>ab</b>	991
22	<b>die</b>	945

## Frekwencja łacińskich form rzeczowników

44	<b>omnia</b>	537
45	<b>deus</b>	491
46	<b>dei</b>	481
47	<b>teste</b>	479
53	<b>domini</b>	434
57	<b>summa</b>	396
58	<b>vulgo</b>	389
60	<b>nihil</b>	385
67	<b>omnes</b>	358
74	<b>staus</b>	338

81	<b>parte</b>	323
85	<b>dominus</b>	304
87	<b>consilium</b>	300
89	<b>nomine</b>	294
91	<b>titulo</b>	291
92	<b>rerum</b>	291
93	<b>ratione</b>	291
97	<b>vita</b>	283
99	<b>deo</b>	282
119	<b>rex</b>	247
121	<b>deum</b>	246

## Zależność między ilością łaciny w tekście a rodzajem tekstu

### Literatura piękna i Biblia

Biblia – 0%

Epika – 1%

Liryka -1%

Dramat – 1%

Synkretyczne – 2%

### Teksty użytkowe

Faktograficzne – 2%

Kancelaryjne – 3%

Listy – 3%

Dydaktyczne – 5%

Prasa – 5%

Perswazyjne – 8%

## Zmiana stopnia nasycenia tekstów łacina na przestrzeni lat (na przykładzie tekstów perswazyjnych)

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

## Łacina w tekstach różnych autorów

Teksty	Stosunek segmentów łacińskich do wszystkich segmentów
B. Chmielowski, Nowe Ateny... (1745-1756)	10%
W. Bystrzonowski, Informacja matematyczna... (1743)	1%
W. Bystrzonowski, Polak sensat... (1733)	20%



## Występowanie łaciny na przykładzie tekstów o kometach

Identyfikator tekstu	Stosunek segmentów łac. do wszystkich segm. (w przybliżeniu)	Rok wydania
ŻędzKom	8,6%	1618
BemKom	1,5%	1619
FurUważ	0,4%	1664
<b>CiekAbryz</b>	<b>7,5%</b>	<b>1681</b>
<b>NiewiesKom</b>	<b>7,7%</b>	<b>1681</b>
DuńCiek	9,7%	1744
BohJProg	0,6%	1770

## Przykłady problemów badawczych

- ◆ badania porównawcze nad utartymi zwrotami łacińskimi i polskimi
  - *in casu* i *in casum* - łącznie 124 wystąpień
  - *w przypadku, w wypadku, w sytuacji, w okoliczności, w okolicznościach* - łącznie 27 wystąpień
- ◆ analiza porównawcza charakterystycznych zwrotów łacińskich w obrębie tekstów różnych autorów lub tekstów tego samego rodzaju, np. ***Anno et die ut supra*** - 27 wystąpień w korpusie - wszystkie frazy występują jedynie w tekstach urzędowo-kancelaryjnych.

**Dziękujemy za uwagę**