

Korpus dla słownika, słownik dla korpusu

Włodzimierz Gruszczyński*

Maciej Ogrodniczuk**

Marcin Woliński**

*IJP PAN

**IPI PAN

Wystąpienie przygotowane w ramach projektu
Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do roku 1772)
finansowanego przez NPRH w latach 2012-2018

SŁOWNIK:

e-SXVII, czyli *Elektroniczny słownik języka polskiego XVII i XVIII wieku*, opracowywany przez zespół w IJP PAN i udostępniany na bieżąco pod adresem:
<http://sxvii.pl>

KORPUS:

KORBA, czyli *Elektroniczny korpus tekstów polskich z XVII i XVIII wieku (do roku 1772)*, powstający w ramach projektu realizowanego w latach 2013-18 przez zespół złożony z pracowników IJP PAN i IPI PAN

SXVII

1954-1993:

- ustalenie kanonu źródeł
- ekscerpca, w wyniku której powstało ok. 2,8 mln kart z cytataми

1996:

- wydanie zeszytu próbnego

1999-2004:

- wydanie I tomu (5 zeszytów)

od 2004:

- komputeryzacja prac
- przejście na formę elektroniczną
- stworzenie rozbudowanej bazy danych słownikowych, biograficznych i bibliograficznych
- udostępnianie wyników prac na bieżąco za pomocą programu przekształcającego dane z bazy na tekst artykułów hasłowych

POLSKA AKADEMIA NAUK | INSTYTUT JĘZYKA POLSKIEGO ELEKTRONICZNY SŁOWNIK JĘZYKA POLSKIEGO XVII I XVIII WIEKU

o słowniku | **lista haseł** | **zaawansowane wyszukiwanie** | **kwerendy** | szerokość: 600
pl

znajdź hasła zaczynające się od

☐ przeszukaj także hasła w indeksie ☐ a fronte ☐ a tergo

wtyczki wyszukiwania dla przeglądarek Firefox i Microsoft Internet Explorer

A B C D E F G H I J K L M N O P R S T U V W X Y Z Z Z

Aktualnie słownik opracowują:

Redaktor naukowy: Włodzimierz Gruszczyński

Autorzy haseł: Dorota Adamiec (DA – sekretarz redakcji), Renata Bronikowska (RB),
Włodzimierz Gruszczyński (WG), Paweł Kupiszewski (PK), Magdalena Majdak (MM),
Małgorzata B. Majewska (MBM), Wiesław Morawski (WM), Aldona Przyborska-Szulc (AP-S)

Prace zapoczątkowali:

Halina Koneczna | Stanisław Skorupka | Salomea Szlifarsztejnowa

Komitet Redakcyjny:

Mirosław Bańko, Ireneusz Bobrowski, Adam Karpinski, Maria Karpluk, Jadwiga Puzynina,
Wojciech Ryszard Rzepka, Krystyna Siekierska, Teresa Smółkowa, Stanisław Urbańczyk,
Wacław Twardzik, Bogdan Walczak, Halina Wiśniewska

Konsultanci naukowci haseł opublikowanych w I tomie:

Alicja Zemanek (botanika), Antoni Mączak (historia), Irena Turnau (kultura materialna),
Krystyna Weyssenhoff-Brożkowska (łacina), Tadeusz Majda (orientalistyka), ks. Józef Kulisz (religia)

Redaktor I tomu drukowanego:

Krystyna Siekierska

Hasła opracowali:

Dorota Adamiec (DA), Renata Bronikowska (RB), Jadwiga Chotkowska (JCh),
Włodzimierz Gruszczyński (WG), Marek Kunicki-Goldfinger (MK-G), Katarzyna Kupiszewska-
Grzybowska (KK-G), Paweł Kupiszewski (PK), Danuta Lankiewicz (DL),
Małgorzata B. Majewska (MBM), Magdalena Majdak (MM), Ewa Małachowska-Pasek (EM-P),
Wiesław Morawski (WM), Aldona Przyborska-Szulc (AP-S), Barbara Rykiel-Kempf (BR-K),
Krystyna Siekierska (KS), Teresa Sokółowska (TS), Kazimierz Żelazko (KŻ)

Materiały ekscerpowali:

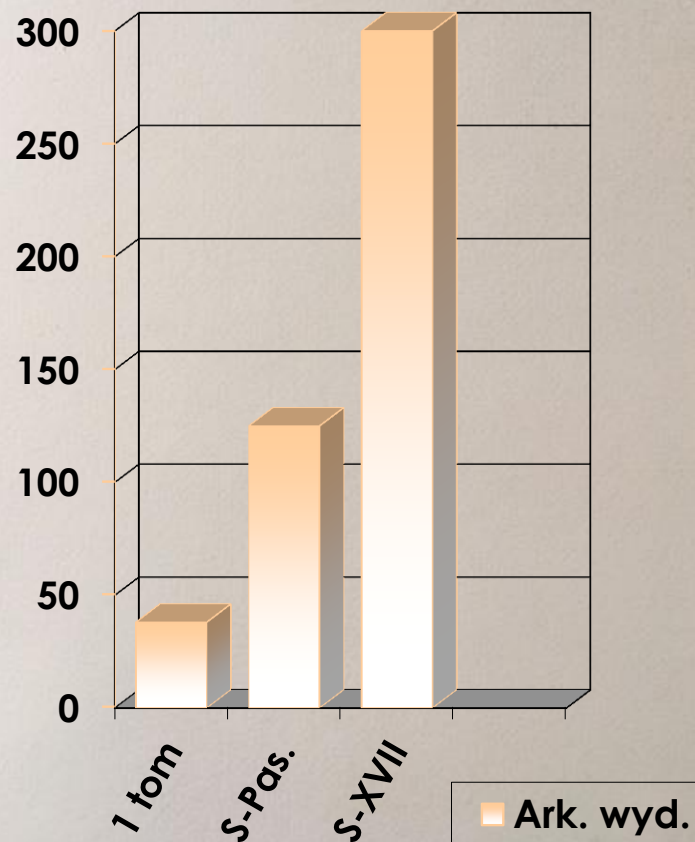
Jadwiga Chotkowska, Teresa Iglikowska, Katarzyna Kupiszewska-Grzybowska, Danuta Lankiewicz,
Andrzej Maria Lewicki, Małgorzata B. Majewska, Wiesław Morawski, Czesław Pankowski,
Aldona Przyborska-Szulc, Barbara Rykiel-Kempf, Alina Ściebora, Krystyna Siekierska,
Teresa Sokółowska, Kazimierz Żelazko, Halina Rybicka, Aleksandra Tokarska

Podstawowe dane

Słownik notuje:

- 19227 znaczeń,
- 2342 utrwalone połączenia wyrazowe,
- 485 związków frazeologicznych,
- 254 przysłowia,
- 1295 przenośni,
- 57826 cytatów z 862 źródeł,
- 35694 formy fleksyjne,
- 2952 informacje o etymologii.

Hasła				
Opr.	W opr.	Zaląż.	Odn.	Razem
390	11 294	3 442	690	15 816
2%	71%	21%	5%	100%



BABA I rzecz. ż

Słowniki notują

Formy gramatyczne:

lp

M. **baba**
D. **baby**
C. **babie**
B. **babę**
N. **babą**
Ms. **babie**
W. **babo**

lpdw

M. **babie**
B. **babie**

lm

M. **baby**
D. **bab || babów**
C. **babom || babam**
B. **baby**
N. **babami || baby**
Ms. **babach**

1. » o kobiecie, zwłaszcza starszej; czasem z niechęcią; indyw. stara panna «:

- A toć panie Moje zdanie, Staroletnej **Baby** szpetnej. *GrochWiersze* 335.
- Stárká/ vide **Bábá** 2. *Kn* 1062.
- Nigdy się ja po proźnicy przy **babach** nie bawię: Gdzie czego sprostać nie mogę, to przez **babę** sprawię. *SejmPiek* 44.
- Pozwalam iak tak młodszym ale tobie **Babo** Pieszczotkami się bawic, pfe brzydki szpeciaku. *OpalKSat* 28.
- Coli zá zábawa w tym domu okopciáły **Bábow** [w piekle]. *AndPiekBoh* 40.
- Białogłowy nie chcą żadną miarą, aby Choć im lata dokuczają, miano ich za **baby**. *MorszaUtwKuk* 329.
- Iuz nas o tę Woynę y **baby** konfuduią. *PasPam* 199.
- Już ztąd niebogo miej wrózkę niesłabą, Że rutkę siejąc pewnie umrzesz **babą**. *ZbierDrużWir I* 147.
- Das alte Weib/ **Bábá** f. *MalczWok A* 3.
- Epithalamium Korwinowi [...] Zostałeś Mężem **Babie**, mogąc iey być wnukiem. *PotPocz* 148.
- Y iuz na zwiędłe [kwiaty], choć bezzębney **babie**, Pożyczą zębów swoich twarde grabie. *DrużZbiór* 312.
- Woysko oboygá narodów Ordynansami Królewskimi wzgardziło, obróciwszy Rycerskie swoje odwagi na Dobra Duchowne, i na woynę z **babami**. *LubHist* 102.

▲ ustabilizowane połączenia wyrazowe:

▲ stara baba:

- Pieniądze lat dziesiątków sześci nie pozbawiają, Ani uczynią młodej panny z **starej baby**. *SzymSiel* 36.
- Kászleć jáko **stára Bába**. *ErnHand* 520.

● **przysłowia, sentencje, skrzydlate słowa:**

- *od przysł.: Baby sobie troje niewidy pod kądzielą bają.*
 - Taki to testament y taki wagi, iaka Ewangelia bab, kiedy sobie troie niewidy pod kądzielą baia. *WisCzar 90.*

- - *Babá z wozu/ kołom lzej.* *RysProv 12.*
 -
 - *Bábá lecąc ze wschodu [= ze schodów] wołálá/ co dalej to gorzej.* *RysProv 12.*
 -
 - *Długo ten pokuka co babę oszuka.* *T III 15.*

(variant I):

- *Gdy sam szatan sprawić niemoże/ tám oszustá/ álbo bábę posle.* *SyrZiel 1079.*

(variant II):

- *Gdzie diabeł nie może/ tám Bábę posle.* *RysProv B3v.*

- - *Iedná bábá kámień do studnie w wáli/ a dziesięć go chłopow nie dobędzie.* *RysProv Cv.*

- - *Stroi Bábá firleie/ kiedy sobie podleie.* *RysProv I.*

(variant I):

- *Taż baba na tychże kołach.* *VerdBlażSet 92.*

(variant II):

- *Taż przecię baba na tych kolech bywa.* *VerdBlażSet 93.*

(variant III):

- *Táž Bábá/ też kolá.* *RysProv Iv.*

- - *Trzebá Bábie ná boty.* *RysProv Iv.*

■ **przenośnie:**

- Ja [Merkury], wiesz mój zwyczaj, chytre-m te sylaby przejął wnet potym od stoickiej *baby* [upersonifikowanej Cnoty w dialogu z Fortuną]. *MiasKZbiór 216*

2. » matka matki a. ojca «:

- Ta Gierałtowska była *baba* matki pana Władysława Wolińskiego. *AktaKrak II/1 529.*

Problemy materiałowe i warsztatowe

Konieczność ułatwienia dostępu do materiałów, bo:

- papierowe fiszki dostępne tylko w pracowni
- wiele fiszek zawiera tylko adres wystąpienia formy, a nie przepisany cytat

Konieczność powiększenia podstawy materiałowej:

- 2,8 mln kart w SXVII
- ok. 7 mln kart w SXVI



ABUNDANCYJA

woody tam jest abun-
dancya z kmpic
wzrostu

B. CHMIEL. N. ATE NY 1746

+ . II s 180

Zmiany warsztatowe

2011-2013:

- Zeskanowanie kartoteki w ramach projektu RCIN (Repozytorium Cyfrowe Instytutów Naukowych)

2013:

- Udostępnienie fiszek w RCIN na platformie dLibra



<http://rcin.org.pl/dlibra/publication?id=20029&tab=3>

Powiększenie podstawy materiałowej

- 2013: początek realizacji projektu *Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772)*
- Planowana wielkość korpusu: 12 mln segmentów (w rozumieniu NKJP)
- Anotacja morfosyntaktyczna:
 - A. manualna 0,5 mln segmentów
 - B. automatyczna – całość
- Bogate oznakowanie strukturalne (m.in. lokalizacja cytatu z dokładnością do strony)
- Oznakowanie obcojęzycznych fragmentów tekstów (z dokładnością do języka)
- Teksty transliterowane zgodnie z zasadami przyjętymi w e-SXVII
- Możliwość przeszukiwania zarówno tekstu transliterowanego, jak transkrybowanego

KORBA - wczytywacz tekstów

włodek (Redaktor)

[Strona główna](#)

[Nowy tekst](#)

[Wyszukiwarka](#)

Pracuj jako:

[skryptor](#)

[korektor](#)

[Zgłoś błąd](#)

[Zmień hasło](#)

[Wyloguj](#)

Wczytanych tekstów: **149**,
zawierających w sumie
4133241 (4054753 wg TEI)
słów.

Zakończonych tekstów: **69**,
zawierających w sumie
1014015 (994135 wg TEI)
słów.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Identyfikator	Metadane	Stan	Redaktor	Skryptor	Korektor	Słów w doc (w TEI)	Znaków	Tekst	Akcje
PetrSInst	Sebastian Petrycy <i>Instrukcja albo nauka jak się sprawować czasu moru</i> Mikołaj Łob Kraków 1613	Gotowy	renata			11077 (10942)	72526 (40,2str)	nagłówek źródło podgląd XML	
BirkEgz	Fabian Birkowski O <i>Egzorbitancjach kazania dwoje</i> Andrzej Piotrkowczyk Kraków 1632	Gotowy	renata	renata		8657 (8340)	57010 (31,6str)	nagłówek źródło podgląd XML	
fdas (IMPACT)	Anonim TEST	Usunięty	mleart					nagłówek	
RysProv	Salomon Rysiński <i>Proverbium polonicorum</i> Piotr Blastus Kmita Lubeka 1618	Gotowy	renata	renata		16616 (16358)	103500 (57,5str)	nagłówek źródło podgląd XML	
NowZwyc	Anonim Nowina <i>o zwycięstwie cesarza z Wiednia do Lublina</i>	Gotowy	renata	renata		206 (201)	1512 (0,8str)	nagłówek źródło podgląd XML	

Korpus → słownik

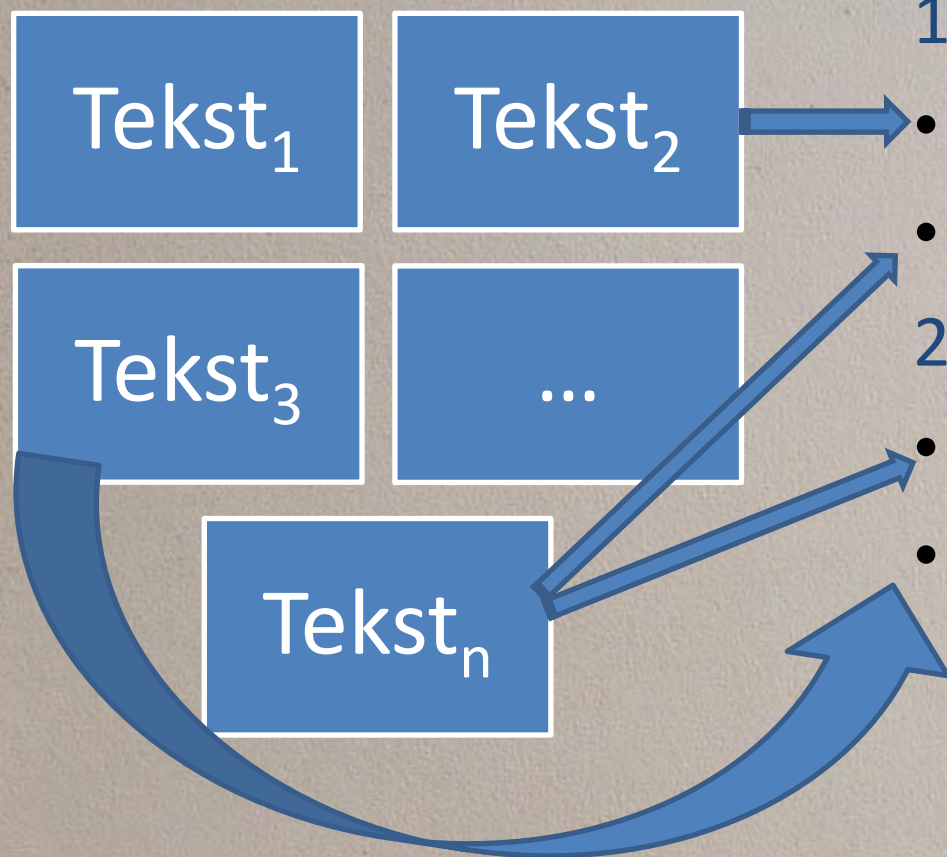
ARTYKUŁ HASŁOWY

1. znaczenie pierwsze

- cytat(y) w formie x
- cytat(y) w formie y

2. znaczenie drugie

- cytat(y) w formie x
- cytat(y) w formie z
- itd.



Korpus dla słownika

- Możliwość szybkiego wyszukania żadanego segmentu lub segmentów wraz z kontekstem.
- Możliwość sprawdzenia frekwencji, chronologii, geografii itp. występowania danej jednostki języka.
- Możliwość określenia gatunków tekstów, w których występuje dana jednostka języka, wskazania autorów, którzy jej najczęściej używają.
- Pomoc w odszukaniu ewentualnych wariantów ortograficznych, fonetycznych, morfologicznych.
- Pomoc w ustaleniu łączliwości syntaktycznej i leksykalnej.
- Pomoc w ustaleniu znaczenia lub znaczeń.

Słownik dla korpusu

Słownik może być (choć rzadko bywa) włączony do korpusu lub traktowany jako korpus:

- słownik dwujęzyczny/wielojęzyczny w korpusie dostarcza zamiast kontekstów/konkordancji – ekwiwalentów obcojęzycznych; w wypadku słowników historycznych może mieć to wartość (np. *Thesaurus* Knapskiego lub *Orbis sensualium pictus* Komenskiego mogłyby się zostać włączone do KORBY);
- słownik jednojęzyczny, zawierający przykłady użycia, jest/może być swego rodzaju (pod)korpusem (por. artykuł P. Żmigrodzkiego).

Słownik dla korpusu

Słownik może stanowić pomoc w budowie narzędzi do obsługi korpusu:

- *Słownik gramatyczny języka polskiego* stanowi podstawę dla analizatora fleksyjnego Morfeusz, który może służyć jako tzw. lematyzator;
- informacja fleksyjna zawarta w dotychczas opracowanych hasłach e-SXVII posłuży jako podstawa do tworzenia narzędzi obsługujących KORBE.

e-XVII → KORBA

Informacje fleksyjne i chronologiczne zostaną wykorzystane do zbudowania automatycznego analizatora fleksyjnego:

- obecnie w *e-SXVII* zinterpretowane zostały 35673 formy fleksyjne,
- większość tych form przytoczonych jest następnie w dokładnie datowanych cytatach, co ułatwi w procesie automatycznej analizy fleksyjnej przypisanie prawdopodobieństwa poszczególnym wynikom.

	transliteracja	transkrypcja
forma	abiudykacją	abiudykacją
lemat	abiudykacya	abiudykacja

Informacja fleksyjna w e-SXVII

POLSKA AKADEMIA NAUK | INSTYTUT JĘZYKA POLSKIEGO
ELEKTRONICZNY SŁOWNIK JĘZYKA POLSKIEGO XVII I XVIII WIEKU

o słowniku | lista haseł | zaawansowane wyszukiwanie | kwerendy | szerokość: 1000
pl

znajdź hasła zaczynające się od szukaj
☐ przeszukaj także hasła w indeksie ☒ a fronte ☐ a tergo
wtyczki wyszukiwania dla przeglądarek Firefox i Microsoft Internet Explorer
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Ź Ż

sposób prezentacji hasła ☒ obszerny ☐ zwięzły

poprzednie hasło -Ż ☐ obszerny ☐ zwięzły następne hasło ŻADEN

ARTYKUŁ HASŁOWY W TRAKCIE OPRACOWANIA
Data ostatniej modyfikacji: 20.03.2014

ŻABA rzecz. ż

Słowniki notują

Formy gramatyczne:

lp	lm
M. żaba	M. żaby
D. żaby	D. żab
	C. żabom
	B. żaby
	N. żabami

lpdw
M. żabie

1. » gatunek pospolitego płaza «:

- Niech przy swym winie Węgrzyn je z czostkiem słoninę [...] Anglik baraninę, Włoch żaby do sałaty, cytryny i figi. *PotFrasz5Kuk III 371.*
- Gdy [...] bagniska, rzeki, sadzawki wyschnęły, żabie upragnione po polach biegały. *VerdBlażSet 63-64.*
- Żaba/ Rana [...]. *Kn 1355.*
- Wnijdź do Faraona, a mów do niego: Tak mówi Pan: Wypuść lud mój, aby mi służył. Ale jeżeli ty nie będziesz chciał puścić, oto, Ja zarazę wszystkie granice twoje żabami. I wyda rzeka żaby, które wyleżą i wnijdą do domu twego, i do komory łoża twego, i na pościel twoją, i do domu sług twoich, i między lud twój, i do pieców twoich, i w dzieże twoje. *BG Jk 2 maj 8, 1-3.*
- Także i te za nieczyste mieć będziecie między płazami, które się wloczą po ziemi, łasica, i mysz, i żaba według rodzajów swoich; I jeź, i jaszczurka, i tchórz, i ślimak, i kret. *BG Jk 29-30.*
- Wydała ziemia ich mnóstwo żab/ y były w pałacach krolow ich. *BG Ps 105, 30.*

Data publikacji:
1608

