



# **„KORBA”**

**Elektroniczny korpus tekstów  
polskich z XVII i XVIII w. (do 1772 r.)**

Pracownia Historii Języka Polskiego  
XVII i XVIII wieku IJP PAN

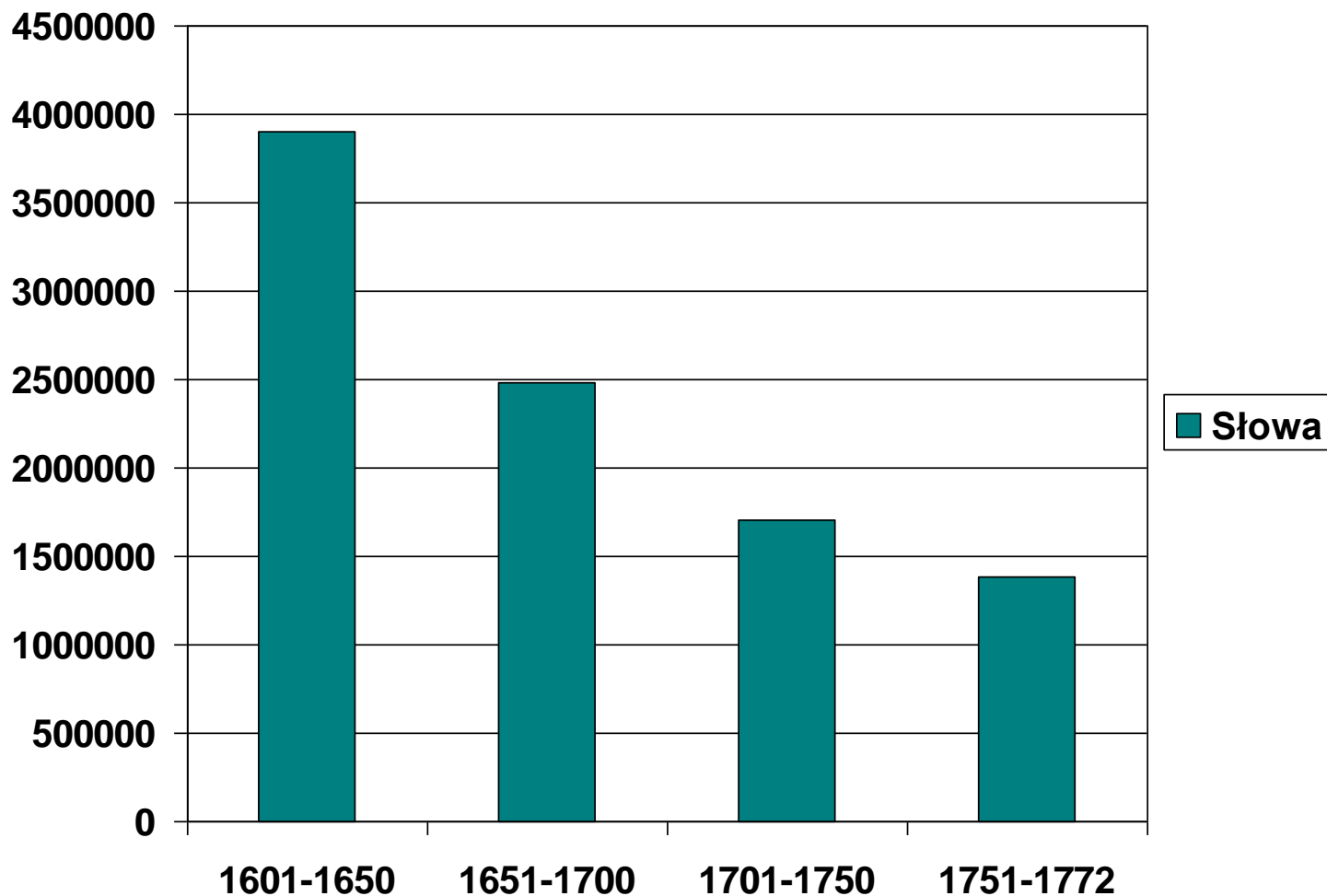
# Podstawowe informacje o projekcie

- Projekt realizowany przez IJP PAN we współpracy z IPI PAN, finansowany w ramach Narodowego Programu Rozwoju Humanistyki na lata: 2013-2017
- Kierownik projektu: prof. dr hab. Włodzimierz Gruszczyński
- Planowana objętość korpusu to 12 mln segmentów
- Korpus historyczny rozszerzający Narodowy Korpus Języka Polskiego
- Kryptonim: KORBA (KORpus BArokowy)
- <http://korba.nlp.ipipan.waw.pl/login/?next=/> (obecnie dostępny dla członków zespołu pracującego nad korpusem oraz w Pracowni Historii Języka Polskiego XVII i XVIII w. IJP PAN w Warszawie)

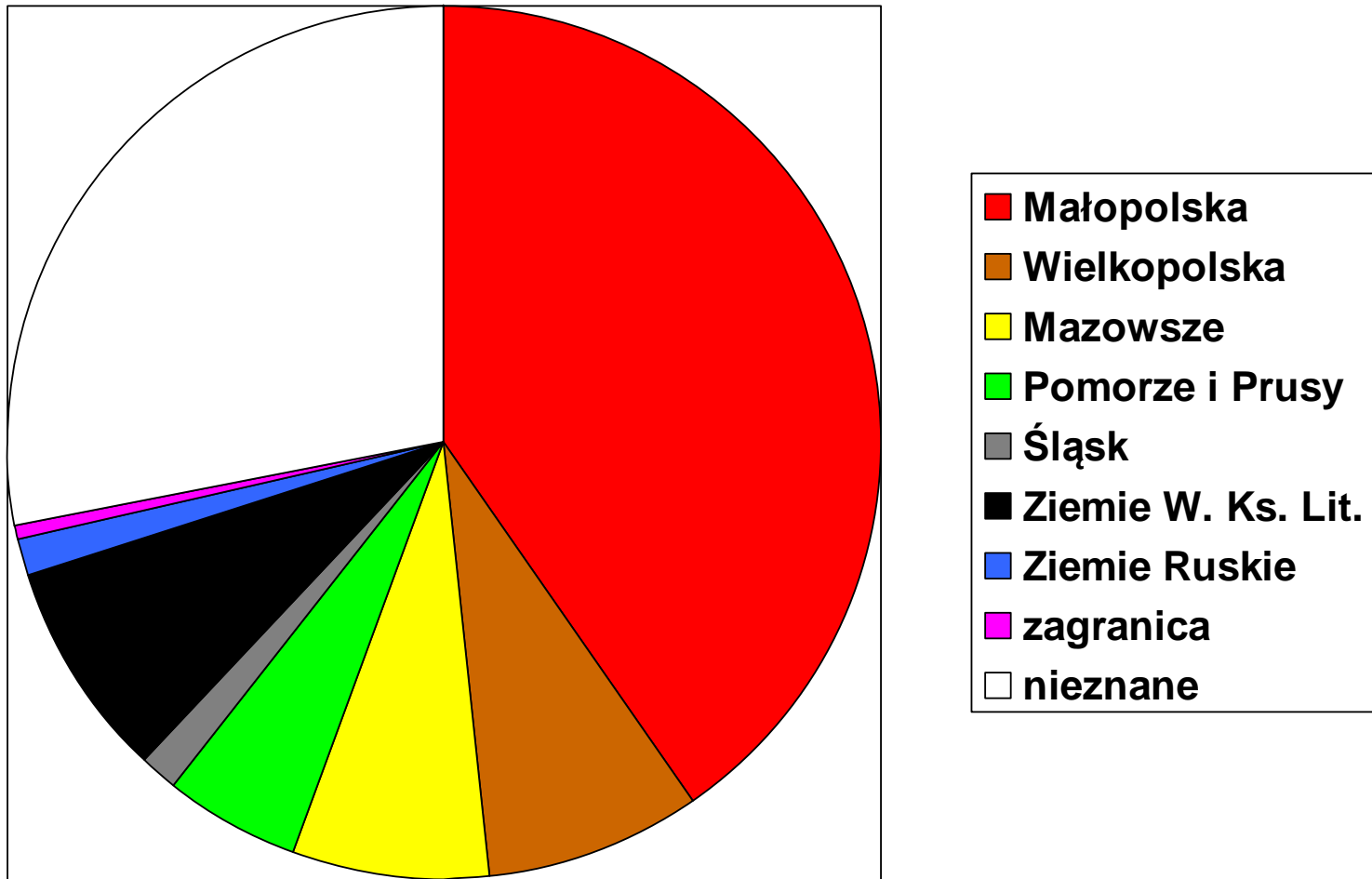
# Obecny stan korpusu w liczbach

- Ponad 650 tekstów zawierających ponad 10 mln 300 tys. słów
- Podstawy:
  - Starodruki: 75,5%
  - Rękopisy: 1,5%
  - Wydania współczesne (XIX-, XX-, XXI-wieczne): 23%
- Typ literatury:
  - Literatura piękna (w tym Biblia): 25%
  - Pozostałe: 75%
- Obecność rymów:
  - Mowa niewierszowana: 75%
  - Mowa wierszowana i mieszana: 25%
- Poetyka żartu: 2%

# Reprezentacja chronologiczna tekstów



# Reprezentacja geograficzna tekstów



# Zakres anotacji

- Metadane – informacje bibliograficzne (autor, tłumacz, wydawca, miejsce i data wydania lub napisania) oraz stylistyczno-genologiczne
- Anotacja strukturalna – oznaczanie wyodrębnionych fragmentów struktury tekstu (np. strona, rozdział, notka marginesowa), a także elementów pominiętych, np. ilustracja, dłuższy fragment w języku obcym;
- Anotacja językowa – oznaczanie obcojęzycznych fragmentów tekstu, np. łaciny;
- Anotacja morfosyntaktyczna – przypisanie informacji gramatycznej każdemu segmentowi; planowane jest ręczne oznakowanie 0,5-milionowej próbki, która będzie potem użyta do budowy tagera, służącego do automatycznego oznakowania pozostałej części korpusu.

# Narzędzia

- Wczytywacz tekstów – konwertuje pliki wordowe na format xml zgodny z TEI, wykrywa błędy w anotacji strukturalnej i językowej, pokazuje statystyki korpusu;
- Konwerter – przekształca teksty transliterowane na transkrybowane (<https://bitbucket.org/jsbien/pol>);
- Anotatornia – wspomaga ręczne znakowanie morfosyntaktyczne;
- Morfeusz XVII – analizator morfologiczny, dostosowany do analizy tekstów XVII- i XVIII-wiecznych;
- Tager – służy do automatycznego znakowania morfosyntaktycznego tekstów;
- Poliqarp 2 – wyszukiwarka dostosowana do przeszukiwania korpusu barokowego.

## KORBA — wczytywacz tekstów

## renata (Redaktor)

[Strona główna](#)[Nowy tekst](#)[Wyszukiwarka](#)[Raporty](#)

Pracuj jako:

[skryptor](#)[korektor](#)[Zgłoś błąd](#)[Zmień hasło](#)[Wyloguj](#)

[Wczytanych tekstów:](#) ,  
zawierających w sumie ( wg  
TEI) słów.

[Zakończonych tekstów:](#) ,  
zawierających w sumie ( wg  
TEI) słów.

[Powrót do strony głównej](#)

Edytuj metadane

Id: *	<input type="text" value="OvZebrMet"/>
Tytuł: *	<input type="text" value="Metamorphoseon"/>
Autor: *	<input type="text" value="Publius Ovidius Naso"/> <input type="checkbox"/> Anonimowy
Tłumacz:	<input type="text" value="Jakub Żebrowski"/> <input type="checkbox"/> Anonimowy
Drukarnia:	<input type="text" value="Franciszek Cezary"/>
Rok:	<input type="text" value="1636"/>
Rok wydania niepewny (wydano nie wcześniej, niż podany rok):	<input type="checkbox"/>
Typ mowy: *	<input type="text" value="wierszowana"/>
Rodzaj: *	<input type="text" value="epika"/>
Gatunek:	<input type="checkbox"/> pieśni <input type="checkbox"/> fraszki i epigramaty <input type="checkbox"/> epitafia <input type="checkbox"/> satyry <input type="checkbox"/> sielanki <input type="checkbox"/> kazania <input type="checkbox"/> pisma polityczne <input type="checkbox"/> polemiki religijne <input type="checkbox"/> mowy okolicznościowe <input type="checkbox"/> traktaty <input type="checkbox"/> dialogi <input type="checkbox"/> pamiętniki <input type="checkbox"/> kroniki <input type="checkbox"/> relacje <input type="checkbox"/> opisy podróży <input type="checkbox"/> herbarze <input type="checkbox"/> akta sejmikowe <input type="checkbox"/> wilkierze <input type="checkbox"/> księgi sądowe <input type="checkbox"/> inwentarze <input type="checkbox"/> rejestry <input type="checkbox"/> diariusze sejmowe <input type="checkbox"/> rozmówki do nauki języka <input type="checkbox"/> podręcznik <input type="checkbox"/> przysłowia <input type="checkbox"/> kalendarze <input type="checkbox"/> przewodniki <input type="checkbox"/> żywoty świętych <input checked="" type="checkbox"/> poematy epickie <input type="checkbox"/> przypowieści, specula (zwierciadła) <input type="checkbox"/> modlitwy <input type="checkbox"/> panegiryk <input type="checkbox"/> lamenty <input type="checkbox"/> emblematy <input type="checkbox"/> przywileje <input type="checkbox"/> konstytucje sejmowe <input type="checkbox"/> bajki <input type="checkbox"/> księgi liturgiczne
Tematyka:	<input type="checkbox"/> alchemia <input type="checkbox"/> anatomia <input type="checkbox"/> architektura <input type="checkbox"/> astrologia <input type="checkbox"/> astronomia <input type="checkbox"/> biologia <input type="checkbox"/> botanika <input type="checkbox"/> budownictwo <input type="checkbox"/> chemia <input type="checkbox"/> egzotyka <input type="checkbox"/> ekonomia <input type="checkbox"/> filozofia <input type="checkbox"/> fizyka <input type="checkbox"/> geografia <input type="checkbox"/> gospodarstwo <input type="checkbox"/> gramatyka <input type="checkbox"/> górnictwo <input type="checkbox"/> historia <input type="checkbox"/> hutnictwo <input type="checkbox"/> języki <input type="checkbox"/> kulinaria <input type="checkbox"/> matematyka <input type="checkbox"/> medycyna <input checked="" type="checkbox"/> mitologia <input type="checkbox"/> miłość <input type="checkbox"/> muzyka <input type="checkbox"/> myślistwo <input type="checkbox"/> obyczajowość <input type="checkbox"/> poetyka <input type="checkbox"/> polityka <input type="checkbox"/> prawo <input type="checkbox"/> religia <input type="checkbox"/> retoryka <input type="checkbox"/> wojskowość <input type="checkbox"/> zielarstwo <input type="checkbox"/> zoologia <input type="checkbox"/> żeglarstwo
Poetyka żartu:	<input type="checkbox"/>
Miejsce: *	<input type="text" value="Kraków"/>
Region: *	<input type="text" value="Małopolska"/>
Wydanie uwspółcześnione:	<input type="checkbox"/>
Pochodzi z korpusu IMPACT:	<input type="checkbox"/>
Tekst chroniony prawami autorskimi:	<input type="checkbox"/>



# Próbka tekstu – Word

[POCZĄTEK·DOKUMENTU]¶

[POCZĄTEK·STRONY·TYTUŁOWEJ]¶

## **PIECHOTNE·CWICZENIE¶**

Albo· WOIENNOSC· PIESZA· Ktorą· Láćinnicy·  
Pedestrem· Militiam· názywáią· Wodzom,  
 Pułkownikom,· y· wszelkiew· Woienney·  
Stárszynie·· Lubo· ktoszkolwiek· iest· Woienney·  
spráwy· Miłóśnikiem.¶

Do·wiadomości· podána.¶

**Przez·BLAZEIA·LIPOWSKIEGO¶**

Nakładem·Ierzego·Forstera,·I.K.M.·Bibli·opoli.<?>¶

**W·KRAKOWIE¶**

W·Drukárni·v·Wdowy·Lukaszá·Kupiszá,·I.K.M.¶

Typogr.·Roku·Panskiego·1660.¶

[KONIEC·STRONY·TYTUŁOWEJ]¶

# Próbka tekstu – XML

```

- <teiCorpus>
  <xi:include href="KORBA_header.xml"/>
- <TEI>
  <xi:include href="header.xml"/>
- <text>
  - <front>
    <pb/>
  - <titlePage xml:id="txt_1-titlePage">
    - <docTitle xml:id="txt_1.1-docTitle">
      <titlePart type="main" xml:id="txt_1.1.1-titlePart">PIECHOTNE CWICZENIE</titlePart>
      - <titlePart type="sub" xml:id="txt_1.1.2-titlePart">
        Albo WOIENNOSC PIESZA Ktorą Lácinnicy
        <foreign xml:lang="la" xml:id="txt_1.1.2.1-foreign">Pedestrem Militiam</foreign>
        nazywáią. Wodzom, Pułkownikom, y wszelkiej Woienney Stárszynie. Lubo ktoszkolwiek iest Woienney sprawy Miłóśnikiem.
      </titlePart>
    </docTitle>
    <note xml:id="txt_1.2-note">Do wiadomości podána.</note>
    <docAuthor xml:id="txt_1.3-docAuthor">Przez BLAZELA LIPOWSKIEGO</docAuthor>
  - <note xml:id="txt_1.4-note">
    Nakładem Ierzego Forstera, I.K.M. Bibli
    <uncertain reason="unreadable" xml:id="txt_1.4.1-uncertain">opoli.</uncertain>
  </note>
  - <docImprint xml:id="txt_1.5-docImprint">
    <pubPlace xml:id="txt_1.5.1-pubPlace">W KRAKOWIE</pubPlace>
    <publisher xml:id="txt_1.5.2-publisher">W Drukárni w Wdowy Lukaszá Kupiszá, I.K.M. Typogr.</publisher>
    <docDate xml:id="txt_1.5.3-docDate"> Roku Panskiego 1660.</docDate>
  </docImprint>
</titlePage>

```

# Konwersja do wersji transkrybowanej

transliteracja	transkrypcja
<p>B. Nie z<sup>á</sup>da mi żaden te<sup>y</sup> rzeczy, ktoraby b<sup>ár</sup>zie<sup>y</sup> owemu niżeli mnie należeć nie mi<sup>á</sup>ła. A t<sup>á</sup>k śmierć <sup>y</sup> Kupidyn poz<sup>á</sup>rszy się i<sup>á</sup>ko <b>bestie</b>, posnęli w Koś<sup>ć</sup>iele Bacchusowym, g<sup>d</sup>zie niewyszumiawszy z przepi<sup>ć</sup>ia śmierć Cupidin<sup>ow</sup>, <sup>á</sup> Cupido śmier<sup>ć</sup>i należący przypasawszy s<sup>á</sup>y<sup>d</sup>ak do boku szły, swych należących odpr<sup>á</sup>wow<sup>á</sup>ć powinnoś<sup>ć</sup>i.</p>	<p>B. Nie z<sup>a</sup>da mi żaden te<sup>j</sup> rzeczy, kt<sup>ó</sup>raby <b>barzyej</b> owemu niżeli mnie należeć nie mi<sup>a</sup>ła. A t<sup>a</sup>k śmierć <sup>i</sup> Kupidyn poz<sup>a</sup>rszy się j<sup>a</sup>ko <b>beztie</b>, posnęli w Koś<sup>c</sup>iele Bacchusowym, g<sup>d</sup>zie niewyszumiawszy z przepi<sup>c</sup>ia śmierć Cupidyn<sup>ów</sup>, <sup>a</sup> Cupido śmier<sup>c</sup>i należący przypasawszy s<sup>a</sup>j<sup>d</sup>ak do boku szły, swych należących odpr<sup>a</sup>wow<sup>a</sup>ć powinnoś<sup>c</sup>i.</p>

# Tworzenie Morfeusza XVII





Dziękuję za uwagę