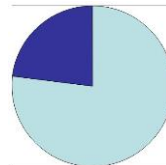
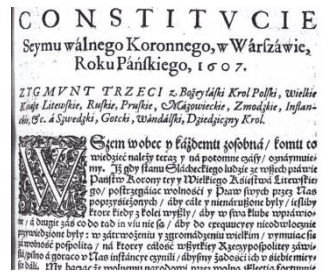


Informacje o projekcie:

- **Kryptonim:** KorBa (KORpus BArokowy)
- **Cel:** stworzenie korpusu tekstów z epoki baroku – historycznego podkorpusu Narodowego Korpusu Języka Polskiego (<http://nkjp.pl>)
- **Planowany rozmiar korpusu:** 12 mln segmentów
- **Czas trwania projektu:** 2013-2018
- **Jednostka koordynująca:** IJP PAN
- **Współpraca:** IPI PAN
- **Kierownik:** Włodzimierz Gruszczyński
- **Finansowanie:** Narodowy Program Rozwoju Humanistyki (grant nr 0036/NPRH2/H11/81/2012)

Typy tekstów:



Udział w korpusie:

- Rękopisy i starodruki: 77%
- Wydania współczesne: 23%

Metadane:

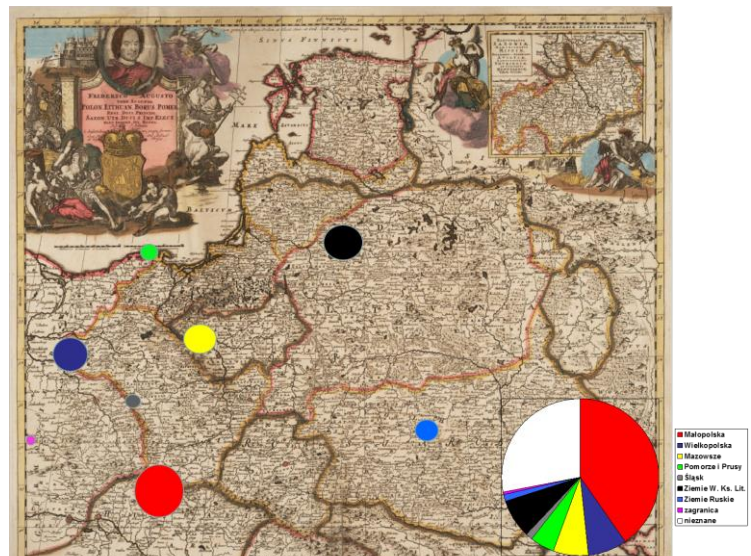
Typy anotacji:

- **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;
- **Konwerter¹** – przekształca teksty transliterowane na transkrybowane;
- **Anotatornia** – wspomaga ręczne znakowanie morfosyntaktyczne;
- **Morfeusz barokowy** – 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.

Reprezentacja geograficzna tekstów:



¹ Narzędzie stworzone przez zespół Janusza Bienia na potrzeby projektu IMPACT, <https://bitbucket.org/jsbien/pol>