

# Reprezentatywność i zrównoważenie – jak zrealizować te założenia podczas tworzenia korpusu historycznego?

(na podstawie doświadczeń z pracy nad Korpusem Barokowym)

1

Dorota Adamiec  
Włodzimierz Gruszczyński  
IJP PAN

# Reprezentatywność i zrównoważenie

- ▶ Terminy nie są synonimiczne.
- ▶ Reprezentatywność (ang. *representativeness*) to odnoszenie się do jakiejś rzeczywistości istniejącej poza korpusem.
- ▶ Zrównoważenie (ang. *balance*) to dbałość o taką budowę korpusu, by żaden składnik na żadnym poziomie nie dominował nad innymi.

# Różne możliwe typy reprezentatywności

- Reprezentacja typów tekstów wyróżnianych na podstawie różnych kryteriów.
- Reprezentacja percepcji języka = struktura czytelnictwa.
- W korpusie historycznym osiągnięcie reprezentatywności pod jakimkolwiek względem jest niemożliwe, gdyż:
  - nie znamy struktury zbioru tekstów powstałych w naszej epoce, ale wiemy, że były ogromne różnice
    - ilościowe pomiędzy różnymi regionami (wyraźna przewaga Małopolski),
    - jakościowe i ilościowe pomiędzy różnymi „podokresami” (wyraźna zapaść w 1. połowie XVIII w.),
    - językowe – duża część produkcji rękopiśmiennej i wydawniczej to publikacje w języku łacińskim (por. Estreicher),
    - stopnia dostępności/popularności (np. rękopisy vs. druki, a zwłaszcza druki o dużych nakładach, jak awizy, gazety ulotne itp.) .
  - nie znamy struktury czytelnictwa (najpoczytniejsze pozycje prawdopodobnie nie dotwały do naszych czasów).

# Reprezentatywność gatunkowa?

## ➤ Problem Biblii.

- Jedyne tłumaczenie Biblii z okresu XVII-XVIII w. to protestancka tzw. *Biblia Gdańska*,
- W Kościele Katolickim obowiązywało tłumaczenie ks. Wujka z 1599 r., wznowione (i zmienione) dopiero w 1740 r.
- W kościołach w czasie nabożeństw korzystano z fragmentów, tzw. czytań biblijnych zawartych w mszałach.
- Pytania metodologiczne:
  - czy tekst Biblii powinien być uwzględniony w korpusie?
  - jeśli tak, to w całości czy w części?
  - jeśli tak, to które tłumaczenie?

# Reprezentatywność gatunkowa?

- Gatunki reprezentowane w literaturze XVII-wiecznej niezwykle często:
  - Poezja okolicznościowa.
  - Rozprawy teologiczne.
  - Homiletyka.
- Gatunki reprezentowane rzadko:
  - Rozprawy naukowe.
- Gatunki użytkowe, („częste”), których realizacje dotrwały w niewielkim zakresie:
  - Umowy,
  - Listy urzędowe.
  - Przywileje, akty nadania.
  - Druki ulotne.
  - Listy.

# Problem rękopisów

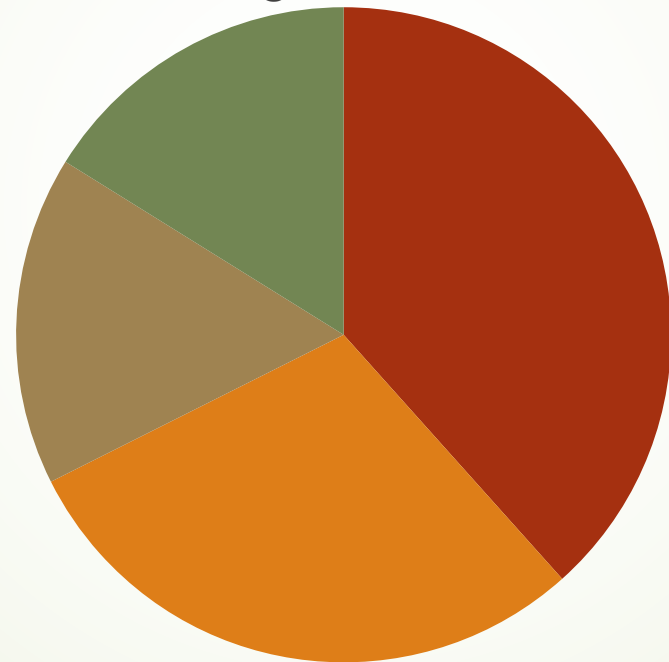
- Wiek XVII i XVIII to ciągle epoka tekstów głównie rękopiśmiennych.
- Czy powinny przeważać w korpusie?
  - Oddają perspektywę nadawcy (wariantywność, idiolekty, regionalizmy itp.).
  - Nie miały wpływu na szeroki krąg odbiorców, nie kształtowały języka wspólnoty tak, jak druki.
  - Trudności z pozyskaniem.

# Zrównoważenie

- W jakimś zakresie jest możliwe do osiągnięcia i bardzo pożądane.
- Sposoby równoważenia korpusu
  - losowanie krótkich próbek tekstów, jak w fXIX
  - dobieranie fragmentów o określonej objętości
    - KorBa 1 – 13,5 mln segmentów
      - Młodzianowski Tomasz, *Kazania i homilie ...*, Poznań 1681  
t. I – IV = 1000 s. in folio
        - 100 s. = ok. 58 tys. segmentów (całość to ponad 0,5 mln)
      - Syreński Szymon, *Zielnik*, Kraków 1613, 1540 s. folio
        - 336 s. = ok. 208 tys. segmentów (całość to ok. 1,1 mln)
      - Chmielowski Benedykt, *Nowe Ateny, IV księgi*, Lwów 1746-1756
        - 3433 s. = ponad 1,1 segmentów.

# Korpus Barokowy (etap 1)

➤ Zróżnicowanie chronologiczne

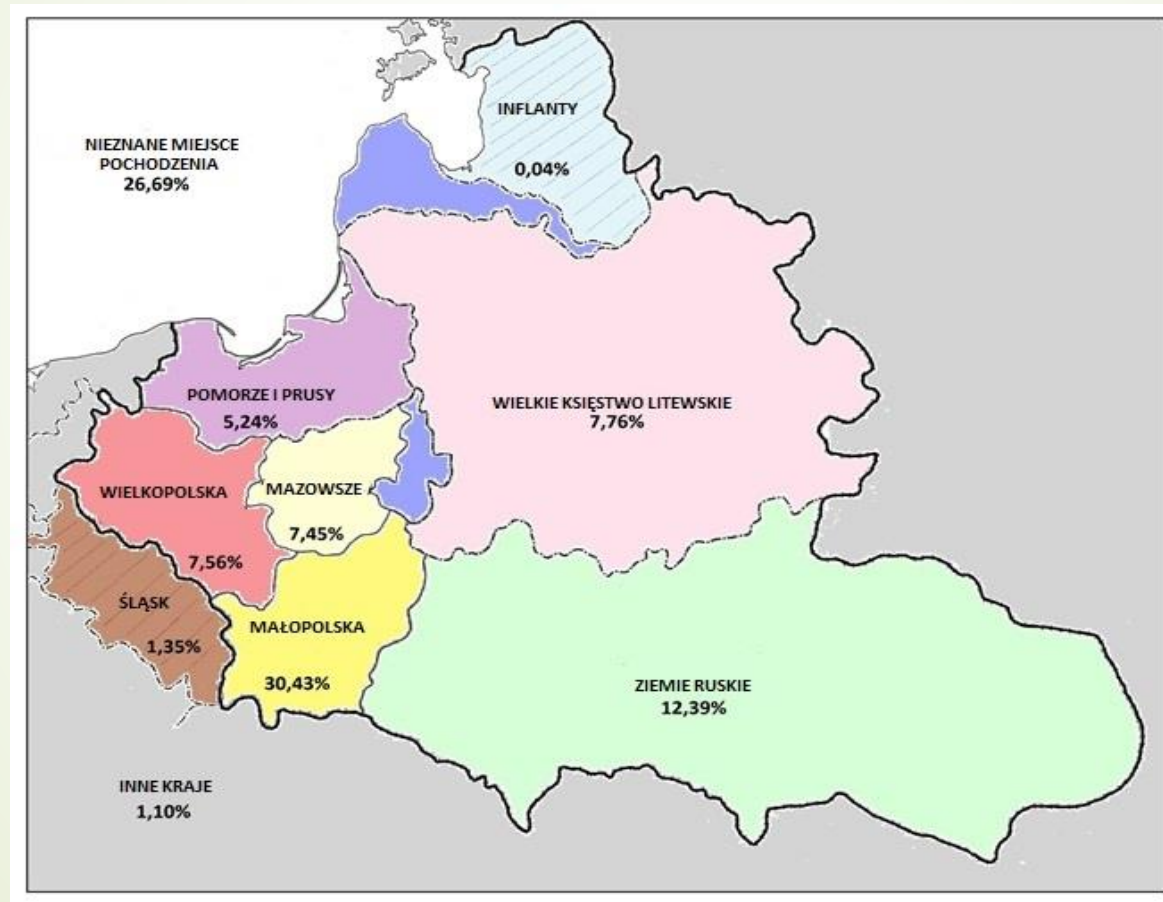


■ 1601-1650   ■ 1651-1700   ■ 1701-1750   ■ 1751-1772



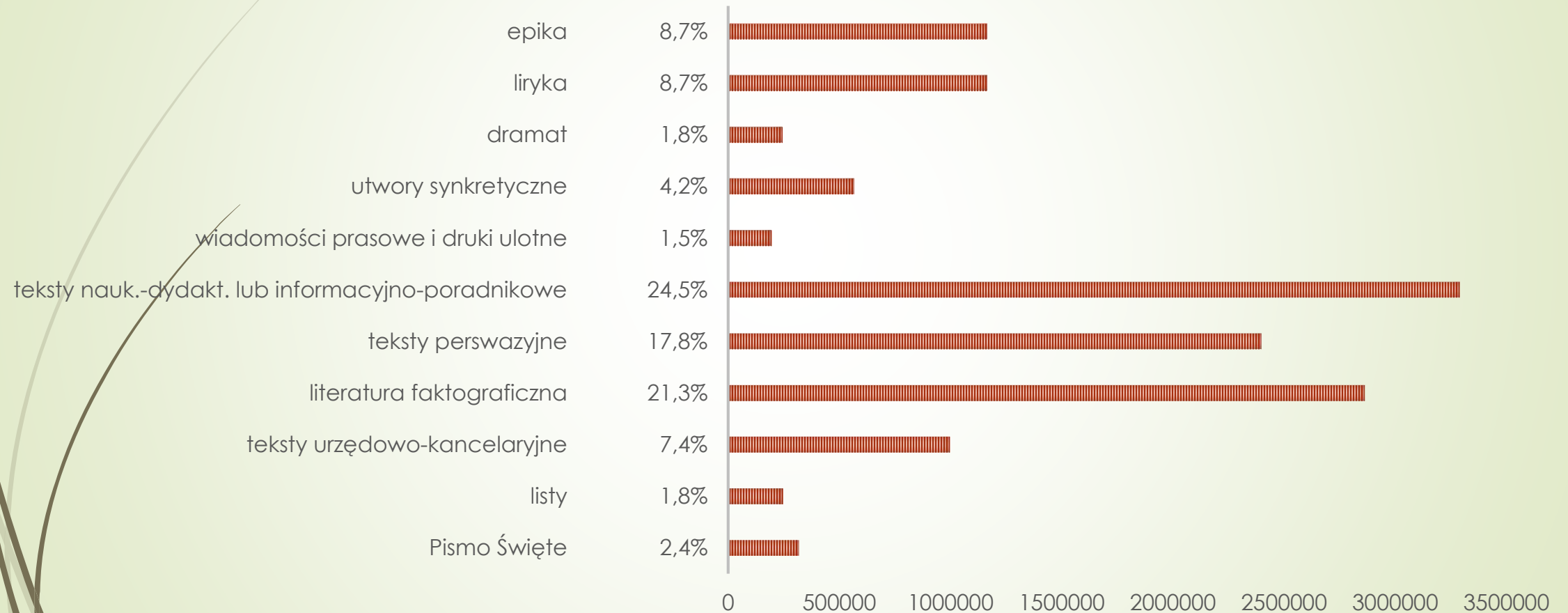
# Korpus Barokowy (etap 1)

- Zróżnicowanie geograficzne



# Korpus Barokowy (etap 1)

## Zróźnicowanie gatunkowe



# Rozbudowa Korpusu Barokowego

## KorBa 2

- Założenia ilościowe w odniesieniu do chronologii

OKRES	KORBA 1	KORBA 2 (02.12.19)	KORBA 2 DOCELOWO	KORBA 1+2 DOCELOWO
1601-1650	5 054 407	35 613	2 000 000	7 054 407
1651-1700	3 746 063	170 234	2 000 000	5 746 063
1701-1750	2 059 850	382 365	2 000 000	4 059 850
1751-1772	2 161 794	337 538	1 000 000	3 161 794
1773-1800	-----	900 650	5 000 000	5 000 000
<b>RAZEM</b>	13 022 114	1 826 400	12 000 000	25 022 114

# Konsekwencje ścisłego zrównoważenia chronologicznego

- ▶ Nadmierna reprezentacja tekstów o podobnej tematyce z okresu „ubogiego” (1. poł. XVIII w.)
  - ▶ *Beatyfikacja Wielkiego Sługi Bozego, Wincentego a Pavlo, Instytutora Kongregacji Missionis [...] Na Gorze Wawell, w Prześwietney Bazylice Katedry Krakowskiej Kaznodziejow Mową Wstawiona; A Jaśnie Oświeconemu Xiżzeciv JMCi Konstantynowi Felicjanowi Na Szaniawach Szaniawskiemu, Biskvpowi Krakowskiemu, [...] Fundatorowi y Dobrodzieiowi swemu Miościwemu, razem cum Relatione Actorum in Solennijs ejusdem Beatificationis, 1731.*
  - ▶ *Chetmicki Maciej, Błogostawiony Jan Kanty Wszystkiej Mądrości Akademia, Panegirycznym Kazaniem przy Solenney Uroczystości Jego w Kościele Kathedralnym Poznańskim, 1729.*
  - ▶ *Denhoff Mikołaj, Chwata Boska Od Świętego Jgnacego, y Towarzystwa Jezusowego y na krok nieodstąpiona Kazaniem na Uroczystosc S. Jgnacego Fundatora Societatis Jesu. Wystawiona w Kościele Lubelskim tegoż Zakonu, 1736.*
  - ▶ *Bujdecki Florian Sebastian, Chwata Nayswiętszego Grobv Chrystvsa Na Solenną Vroczystość Tegoż Nayswiętszego Grobv [...] Pod czas Intermedyalney Kapityły kazaniem Przez X. Floryana Buydeckiego Kanonika Rgularnego Hierozolymskiego Ogłoszona. Roku Pańskiego 1725.*
  - ▶ *Chwata Rzymska Z Patryotow Jego Wielka wyniesienia Onychże w komput Świętych Naywiększa w Chwale Niebieskiej Ogłoszona a ziemi zaś do promowania łatwiejszego Pokazana Przez zebranie Zycia Błogostawioney Ludowiki w Godzinkach y Litaney Od Pewnego Autora Roku 1746 Dnia 28. Kwietnia do Druku Podana, 1746.*
- ▶ Okrojona reprezentacja tekstów z innego okresu – bardzo różnorodnego i bogatego leksykalnie (oświecenie!).

# Konsekwencje wynikające z zakładanego przeznaczenia KorBy

- ▶ Korba ma służyć pomocą przy tworzeniu eSXVII
  - ▶ Celem jest udokumentowanie bogactwa leksykalnego, zróżnicowania tematyki i gatunków pisarskich – mogą więc i powinny być uwzględnione teksty będące tłumaczeniami z innych języków.
- ▶ Badanie zróżnicowania terytorialnego polszczyzny – inne byłyby wymagania wobec korpusu mającego służyć takim badaniom.

# Jak dostosować Korbę do innych zadań badawczych?

- ▶ Stworzenie kilku korpusów zrównoważonych pod różnymi względami metodą losowania dłuższych próbek z poszczególnych tekstów zgodnie z przyjętymi kryteriami, np.:
  - ▶ podkorpus zrównoważony chronologicznie – wylosowanie z każdego 50-lecia próbek o identycznej łącznej długości, np. 4 x 1 mln segmentów;
  - ▶ podkorpus zrównoważony pod względem regionalnym (utworzony analogicznie do poprzedniego);
  - ▶ podkorpus zrównoważony genologicznie.
- ▶ Problemy:
  - ▶ zrównoważenie pod jednym względem może prowadzić do bardzo silnego zaburzenia zrównoważenia pod innymi względami (np. przewaga jednego autora).
  - ▶ zrównoważenie pod pewnymi względami może okazać się bardzo trudne do realizacji z powodu braku tekstów reprezentujących określone cechy, np. w podkorpusie zrównoważonym regionalnie zabraknie tekstów z Inflant czy ze Śląska.

*Dziękujemy wam b́arzo za łaskę y chęć/  
ktoraśćie nam pokazáli.*

15

VolcDial 110v