

Keywords in religious literature of the 17th and 18th centuries
in light of the data from
the Electronic Corpus of 17th- and 18th-century Polish Texts

Magdalena Majdak
Institute of Polish Language
Polish Academy of Sciences

Plan

1. Introduction: the questions and the dreams
2. Assumptions and goals
3. Material
4. Methods
5. Results
6. Analysis (keywords in subcorpora, adequate thematic divisions)
7. Conclusions

Introduction: the questions and the dreams

Keyword of the day, keyword of the year, keyword of the decade...

Is it possible to find the word of the century (or two)?

Introduction: the questions and the dreams

Keyword of the day, keyword of the year, keyword of the decade...

Is it possible to find the word of the century (or two)?

**Is it possible to find the word of the century (or two)
in the selected thematic category?**

What is a *keyword*? Many meanings...

- author (e.g. Mickiewicz, Przyboś, Zembaty-Michalakowa 1982)
- works (e.g. *Pan Tadeusz*, *Dziady*, Sambor 1969, Rudnicka-Fira 1986)
- epoche (e.g. romantism, Stachurski 1989)
- corpora of texts (e.g. corpus of prime ministers' exposé, Kieraś, Zawadzka-Palucktau 2023)
- words important for a specific culture and its understanding (Wierzbicka 1997, Nagórko 2021)

Many meanings of *keywords*

- flagship words, perceived as determinants of the value system of a given community (Pisarek 2002)
- thematic centers that are ideologically and politically important for a specific time or community (e.g. included in the dictionary Williams 1976 et seq.)
- culturemes - units used to describe the relationship between language and culture (Nagórko 2021, Rak 2017, Bartmiński 2016), important for the self-identification of a community, characterizing its attitude to tradition; thanks to them, it is possible to understand the specificity of a given national, ethnic or regional community (Rak 2015)

Meaning of *keywords*

In linguistic analyses, the main goal is essentially to extract words whose frequency of occurrence is significantly higher compared to the reference set

In the case of cultural research, it is the importance or specificity of the word whose meaning is based on an important cultural category

Goals

- Application of the keyword method to old Polish
- Select keywords for religious texts from the Electronic Corpus of 17th- and 18th-century Polish Texts
- Select keywords also for non-religious texts from the Electronic Corpus of 17th- and 18th-century Polish Texts

KorBa - Korpus Barokowy - Baroque Corpus

- The Corpus of Polish Texts of the 17th and 18th centuries (until 1772) was created at the Institute of the Polish Language of the Polish Academy of Sciences with the participation of the Linguistic Engineering Group of the Institute of Fundamentals of Computer Science of the Polish Academy of Sciences under the supervision of Włodzimierz Gruszczyński
- In 2013 - 2018 a collection of 13.5 million tokens was extracted from texts published between 1601 and 1772
- The second part of the project (2019-2023) is coming to an end, thanks to which the resource will double (to approximately 25 million text words) and the time range will cover materials until the end of the 18th century

Material

- Corpus of religious texts - it covered 91 texts tagged as *religion* and contained a total of 2,064,623 tokens
- Reference corpus - all texts in the corpus apart from those tagged as *religion*. It covered 639 texts and contained a total of 9,166,861 tokens

Sample search options - material selection criteria

The search options at various levels can be used to extract sub-corpora.
It is possible to search by:

1. thematic (category *religion*)
2. genologic - search was conducted by category (*factual literature, persuasive texts, scientific-didactic texts or information and handbooks*)
 - type (*prose, poetry, drama*)
 - genre (*sermons, religious writings, nativity plays, prayer books, catechisms, letters*)
3. authors

What is category *religion*?

The category *religion* was chosen as one of the most represented in the corpus and featuring in all subperiods

Inside: spirituality, liturgy, rites, polemics, element of social and political life

We talk about category *religion*, specifically, about vocabulary from texts assigned in Corpus KorBa the *religion* label

Method

The log-likelihood method was used to extract the keywords.

The cut-off point was established at approximately $p = 0.0000000001$ corresponding to a very high statistical significance (high keyness) to ensure that the list of keywords was neither too long nor cut off arbitrarily.

Words that occurred in less than five different texts in entire KorBa were not included

Method

The log-likelihood method allows to assign the positive or negative keyness value:

- **positive**(significantly more occurrences than in the reference corpus)
- **negative** (significantly fewer occurrences than in the reference corpus) keyness

Results

Managed to extract precisely religious vocabulary (and not, for example, mining or geographical vocabulary), which means that the choice of method was accurate

2153 records were obtained, including 952 units with positive keyness, of which approximately: 430 nouns, 215 verbs, 90 adjectives, 5 adverbs

These are approximate results for the entire 17-th and 18-th centuries. In the next steps, it's necessary to assign and specify meanings, e.g. *złoty* color or coin, *święty* noun or adjective

Another problem is the separation of thematic groups

Results

The top 30 nouns: *God, Jesus, sin, Christ, soul, (Orthodox) church, Satan, heaven, spirit, man, faith, sacrament, word, cross, witch, life, angel, father, world, saviour, salvation, glory, body, love, child, scripture, the Passion, apostle, prayer, creation*

The top 30 adjectives: *saint/holy, God's, divine, heavenly, eternal, Christlike, human, of Jesus, blessed, satanic, monastic, Christian, everlasting, beloved, Christ's, sinful, salutary, catholic, carnal, apostolic, angelic, clerical, true, pious, Lord's, eastern, worldly, just, spiritual, infernal*

The top 30 verbs: *speak, create, love, suffer, believe, confess, say, eat, save, convert, desire, curse, punish, receive, do, pray, say, be born, teach, have mercy, resurrect, imitate, lie carnally, understand, sin, condemn, celebrate, pardon, praise, crucify*

The analysis allowed to identify the following groups

- names of divine persons (*God, Jesus, Christ, Saviour, Messiah, Spirit, Trinity*)
- designations of the Mother of God (*mother, Mother of God, Virgin Mary*)
- names of celestial beings and places (*angel, Satan, devil, hell, heaven, Heavens*)
- humans and faith (*man, creation, creature, being, sinner, sin, error, act, deed, conscience, life, death, heart, gift, grace, consolation, prayer, fasting, alms, exercise, punishment, suffering, cross, the Passion, sacrament, baptism, confession, penance, repentance, communion, faith, will, blessing, conversion, salvation*)
- names of positive actions, qualities and values (*love, truth, mercy, goodness, perfection, reason, humility, earnestness, holiness, purity, piety, lovingkindness, wisdom, joy, gladness*)

names of negative actions, qualities and values (*lie, ugliness, wickedness, blasphemy, heresy, anger, obscenity, contempt, foolishness, imperfection, vanity, wrath, fear*)

characteristic religious imagery (*shepherd, sheep, reed, bridegroom, bride, light, glory, reverence, cross, thorn, supper, star, neighbour, vanity, majesty, eternity, revelation, vision, child, infant, baby, swaddling cloths, son, miracle*)

names describing believers (*Christian, rabbi, Muslim, Arian, Catholic, heretic*)

church functions (*pope, priest, Jesuit, clergyman, parish priest, monk, minister, preacher*)

places of prayer (church, Orthodox church)

biblical figures (*apostle, prophet, patriarch, Job, Peter, Thomas, Cain*)

Bigrams

- Bóg mój
- duch święty
- w niebo
- syn boży
- z krzyż
- chwała twój
- święty sakrament
- święty i
- pan Bóg
- miłość twój
- dusza mój
- żywot Bóg
- Bóg i
- pan mój
- wola twój
- Chrystus pan
- na wiek
- pismo święty
- wszytek świat
- w cerkiew
- Bóg nasz
- ciało i
- do niebo
- być Bóg
- ty Bóg
- łaska twój
- pan Chrystus
- Bóg ojciec
- zbawiciel nasz
- wieczny chwała
- ten świat
- z niebo
- pan Jezus
- Chrystus Jezus

Trigrams

- ciało i krew
- nad my jako
- się nad my
- ty Bóg mój
- zmiłować się nad
- pan Bóg nasz
- i duch święty
- syn i duch
- i od syn
- chleb i wino
- ojciec i syn
- i krew pan
- w trójca jedyny
- na ten świat
- ciało i dusza

- Bóg w trójca
- niebo i ziemia
- osoba chleb i
- być w niebo
- zakon Dominik święty
- pod osoba chleb
- cześć i chwala
- Bóg mój i
- twój Bóg mój
- tu na ziemia
- w imię ojciec
- imię ojciec i
- i krew pański
- męka i śmierć
- pan i Bóg

Tetragrams

- zmiłować się nad my
- syn i duch święty
- ciało i krew pan
- Bóg w trójca jedyny
- osoba chleb i wino
- ojciec i syn i
- i syn i duch
- pod osoba chleb i
- imię ojciec i syn
- w imię ojciec i
- ciało i krew swój
- sakrament ciało i krew
- ciało i krew pański
- święty sakrament ciało i
- od ojciec i od
- ojciec i od syn
- na niebo i na
- syn i duch z
- niebo i na ziemia
- pan i zbawiciel nasz
- chleb i wino w
- i z duch święty

Pentagrams

- pod osoba chleb i wino
- ojciec i syn i duch
- imię ojciec i syn i
- w imię ojciec i syn
- i syn i duch święty
- święty sakrament ciało i krew
- otec syn i duch święty
- od ojciec i od syn
- ojciec syn i duch święty
- na niebo i na ziemia

Hexagrams

- w imię ojciec i syn i
- imię ojciec i syn i duch
- ojciec i syn i duch święty

<https://korba.edu.pl>

[base="pod"][base="osoba"][base="chleb"][base="i"][base="wino"]
[base="w"][base="imię"][base="ojciec"][base="i"][base="syn"][base="i"]

Negative keyness

- The negative keyness can serve as *à rebours* testimony to the importance of words in fields other than religion
- It highlights important categories of everyday life
- It reveals a much higher prevalence of nouns denoting values, states, emotions and other abstract concepts in the corpus of religious texts compared to the reference corpus where are words describing what lies within sight

Thematic groups selected on the basis of *negative* keywords (1)

- political and state-related vocabulary (*army, sejm, sejmik, election, law, constitution*)
- military vocabulary (*camp, fortress, cannon*)
- functions (*hetman, marshal, deputy, king, prince, voievode, starosta, chancellor, castellan*)
- time division - seasons and months (*winter, March, September*)
- food products and spices (*vodka, liquor, wine, beer, alcoholic beverage, juice, salt, pepper, spices, sugar, olive oil, vinegar, peas*)
- animals (*horse, cow, hare*)
- cereals (*grain, barley, oats*)
- rivers (*Vistula*)
- vocabulary related to houses and farmyards (*cottage, room, barn, hallway, chamber, chimney, door, porch, castle*)
- measurement units - mass and distance (*mile, lot, pound, ell*)

Thematic groups selected on the basis of *negative* keywords (2)

- monetary units (*money, zloty , grosz, sterling, florin, thaler*)
- specialised vocabulary in mathematics (*line, square, half, triangle*)
- cities, countries and continents (*Warsaw, Poland, Republic of Poland, Kamenets, Lviv, Vienna, Moscow, Lithuania, Prussia, Italy, Europe, Asia*);
- nationalities (*Pole, Tatar, Cossack, German, Swede, French*)
- administrative division (*country, voivodeship, city, poviat, village*)
- vocabulary related to terrain (*sea, land, river, island, mountain, field*)
- kinship, affinity and other relations (*brother, wife, bachelor*)
- materials, metals and minerals (*iron, coal, sulphur*)
- values (*honour*)
- weather conditions (*heat, humidity*)
- body parts (*stomach, liver*)
- surnames (*Potocki, Lubomirski, Sapieha, Chmielnicki, Czartoryski*)

Analysis of sub-corpus *Nativity plays*

The corpus covers 15 texts and featured a total of 41,684 tokens

Groups of the nouns: principal figures (*Jesus, Mary, Joseph, mother, Blessed Virgin, boy, child*); shepherds' names (*Kuba, Matys, Bartos, Wojtek, Bartek, Janek, Wawrzek*); local flavour elements (*flock, cattle, lamb, wolf, shepherd, shed, bag, punnet, sausage, hut, innkeeper, farmhand, master of the house, inn*); circumstances of the birth (*manger, hay, swaddling cloths, misery, night, birth, spend the night*); witnesses' reactions (*bow, knee, gift, gospel, singing, playing*); emotions (*joy, gladness*); elements of the supernatural order (*heaven, glory, angel*); and biblical realities (*Bethlehem, Herod*).

Groups of the verbs started with the activities of shepherds (*sleep, talk, quarrel, chatter, trouble, play [cards and instruments]*), followed by a sequence of actions related to the birth of Jesus from the perspective of shepherds (*be born, give birth, lie, sleep*), their reactions (*hear, rise up, go, follow, come, greet, welcome, watch, play, chant, sing, long, feel*) and references to the spiritual plan (*remit, deliver, please*)

Groups of the adjectives included the typical folklore references to the birth of Jesus (*beloved, poor, heavenly, little [baby]*)

Analysis of sub-corpus *Nativity plays*

- The use of the log-likelihood method allowed us to extract a typically traditional vocabulary related to Christmas and embedded in the local, Polish reality
- Identified with regards to the reference corpus composed of other religious texts, negative key words are: *Christ, faith, cross, word, sin, soul, saint, church*. It possible to point the lack of theological reflection and vocabulary
- What emerges from the extracted keywords is the description of the 'stop-motion' nativity scene (*angel, kneeling shepherds, crib*), typical of the nativity play genre

Analysis of sub-corpus „Birkowski”

Consisted of four texts by one author - Fabian Birkowski (47,311 tokens):

- **nouns** (*exorbitance, tyrant, dissenter, work, crime, Luther, heretic, cross, hell, cry, wound, war, tombstone, Lucifer, [wrap-over] vest, prophet, justice, janissary, sin, paganism, weight, Mehmed, freedom, apostate, condemnation, death, rope*)
- **verbs** (*curse, suffer, be harmed, bellow, rebuke, reign, murder, perish, liberate, disinherit, create, rout*)
- **adjectives** (*disgraceful, cruel, heavenly, impertinent, infernal, ungodly, devilish, double-edged, pagan, executioner's*)

Particularly interesting were negative key nouns identified based on the reference corpus of other religious texts: *Jesus, Satan, [Orthodox] church, person*

Analysis of sub-corpus „Counter-reformation”

- Consists of the texts from *Counter-Reformation moral satire in 17th century Poland (1623)*
- Vocabulary: minister, church, devil, Saxon, Lutheran, Luther, pope, faith, pludry, diver, joke, wife, synod, predicate, old, faith, gospel, leather, patron, weaver, Calvin, confession, wolf, Bible, wife, Matys, congregation, luter, ministerial, tribunal, priest, negotiate, head, bies, roraty, papist, playpen, pluder
- Particularly interesting were negative key nouns identified based on the reference corpus of other religious texts: way, father, man, divine, holy/Saint, God, king

Conclusions

- The presented work fits into both the research on keyword extraction methods and the analysis of religious lexis
- Statistical research is used in the analysis of old Polish
- It was a pilot studies will be repeated on larger material
- The are the estimations, but these results provide a substantial material for future interpretations