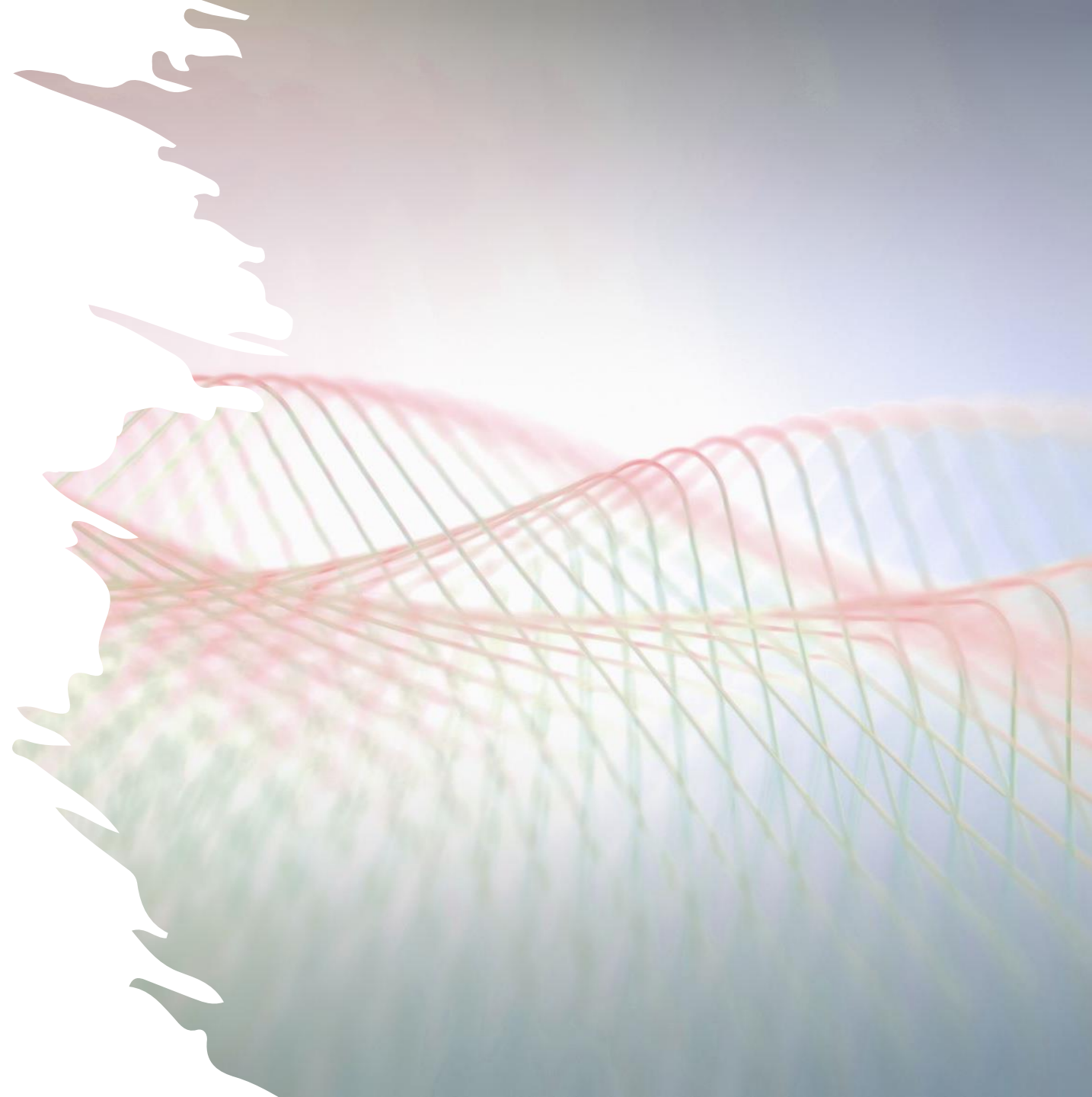


# Relationship graph of online dictionary

Authors: Karol Bieńkowski, Jakub  
Łuczka, Kacper Paluch

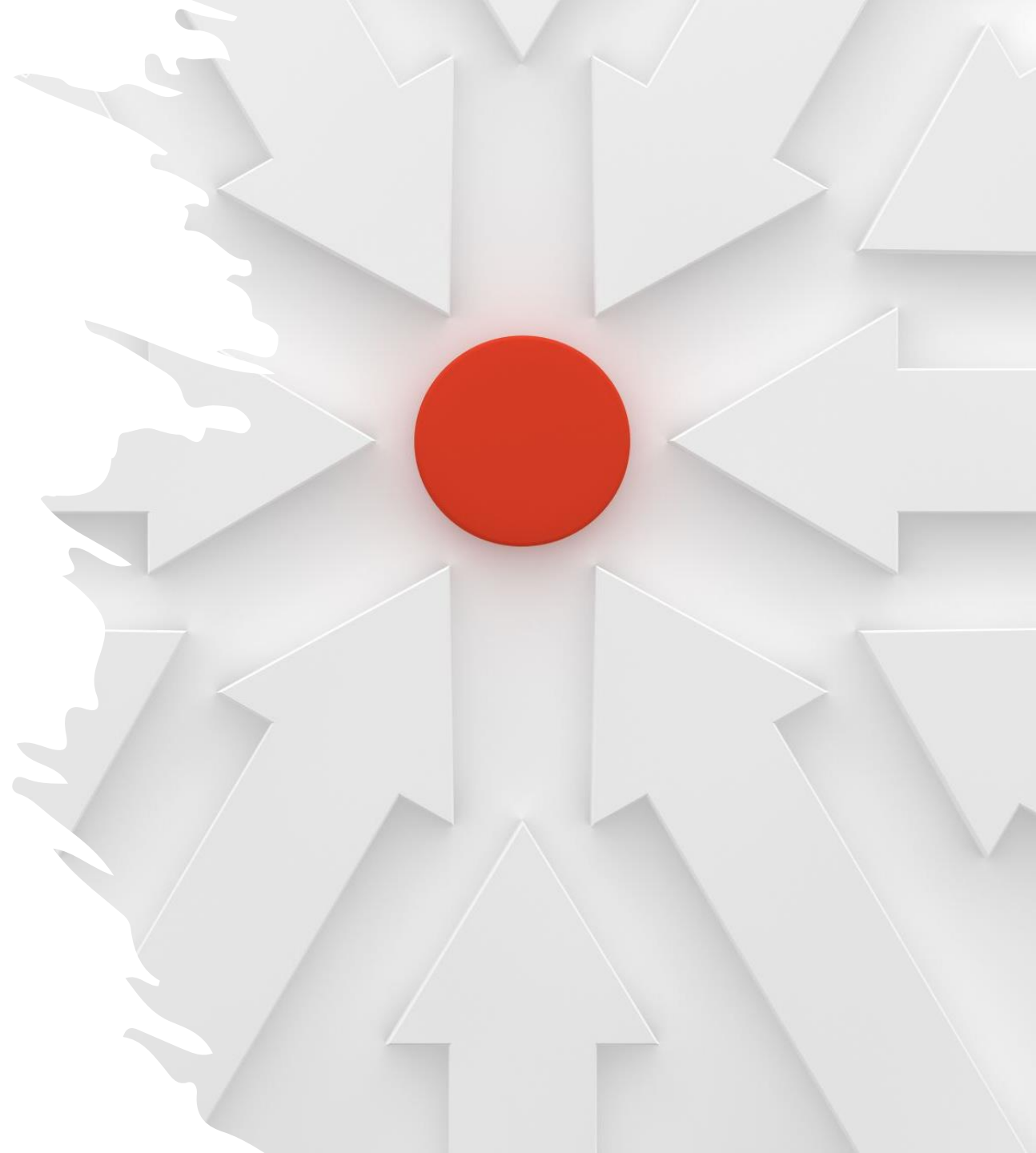


# Table of Contents

1. Motivation.
2. Few words about used Wiktionary.
3. Graph visualization.
4. Statistics of words and their definitions.
5. Transitive closure of words.
6. Crawled graph.
7. Loops in definitions and uniqueness of nodes.
8. Method to find unique nodes.
9. Summury.

# 1. Motivation

- Meaning of words – *How can we define words by other words?*
- Relationships between words (and their definitions) – *Could be these relationships potentially represented as a graph structure?*
- Popularity of the words – *Which words are commonly used to describe others?*



politechnika (język polski) [ edytuj ]

wymowa:

IPA: [ .poliˈtɛxɲika ], AS: [ polˈiteχˈnika ], zjawiska fonetyczne: *zmięk* • *akc. na 3 syl.* • *akc. pob.*



znaczenia:

rzeczownik, rodzaj żeński

(1.1) uczelnia wyższa o statusie uniwersyteckim i profilu technicznym; zob. też *politechnika* w Wikipedii

odmiana:

(1.1) [ukryj▲]

przypadek	liczba pojedyncza	liczba mnoga
mianownik	<b>politechnika</b>	<b>politechniki</b>
dopełniacz	politechniki	politechnik
celownik	politechnice	politechnikom
biernik	politechnikę	politechniki
narzędnik	politechniką	politechnikami
miejscownik	politechnice	politechnikach
wołacz	politechniko	politechniki

przykłady:

(1.1) ***Politechnika** Świętokrzyska znajduje się w Kielcach.*

składnia:

(1.1) *na + Ms.*

kolokacje:

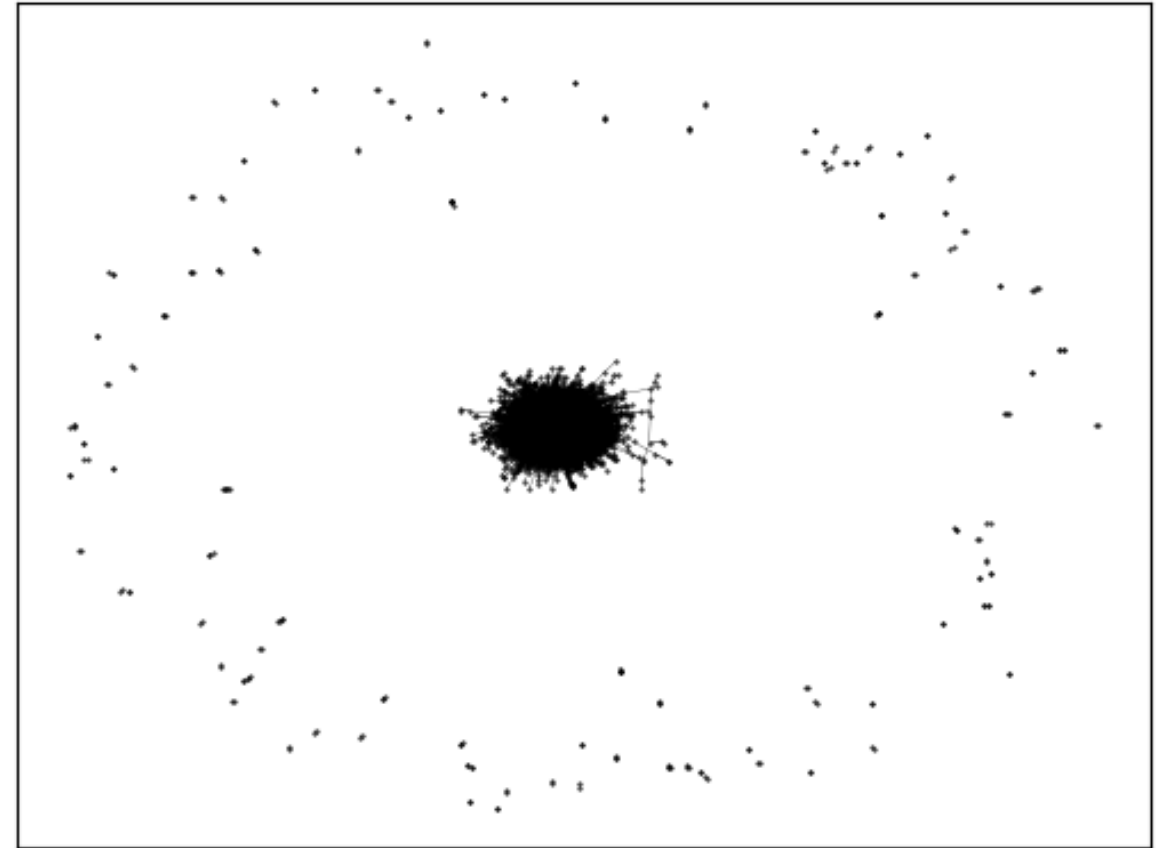
(1.1) *studiować / uczyć się na politechnice*



## 2. Few words about Wiktionary

- In our project we used **Polish Wiktionary**, created by Wikimedia Foundation.
- It explains the etymology and meaning of words, idioms and some abbreviations.
- The dictionary currently contains **99 751** entries, **145 536** meanings, including **950** proverbs.

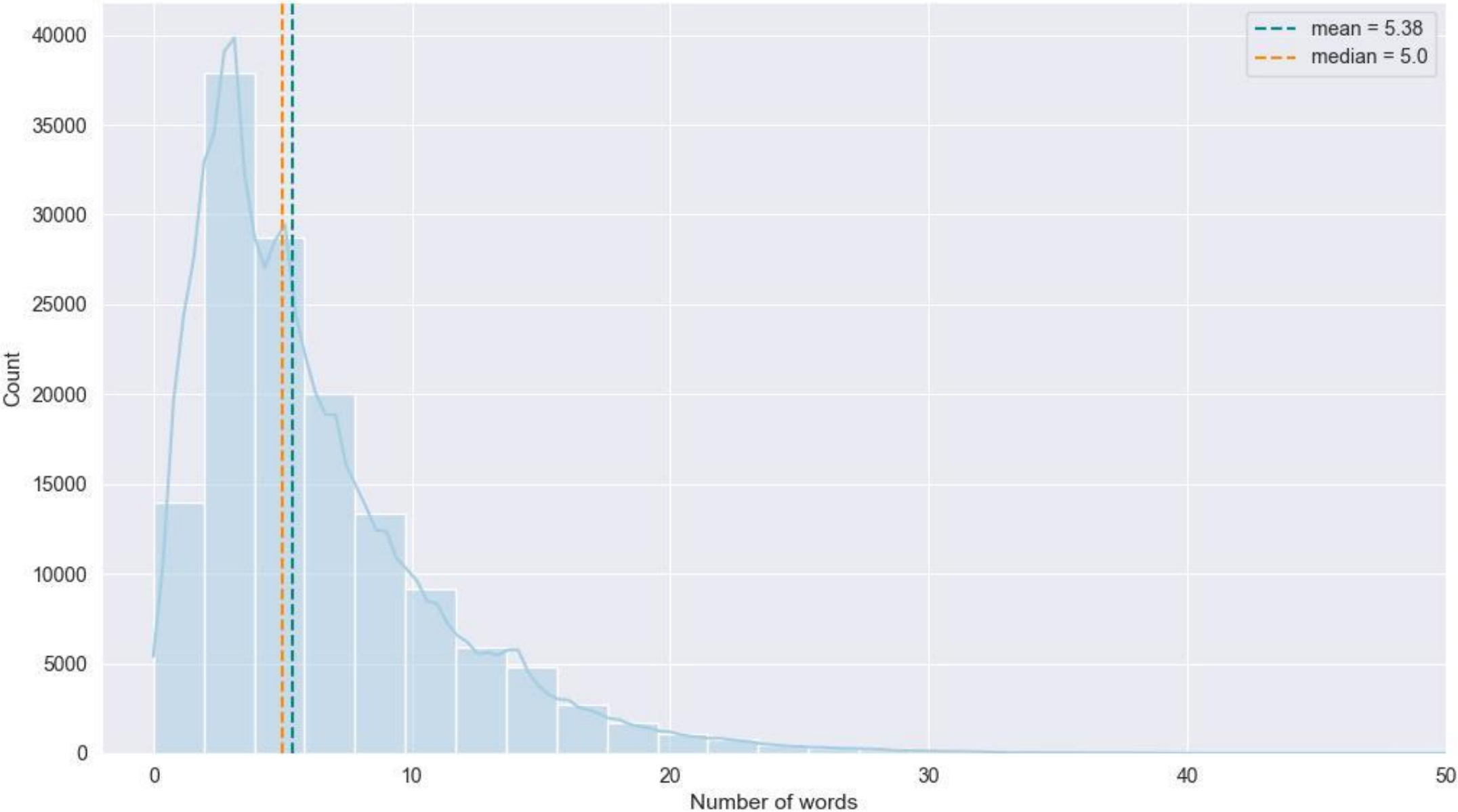
### 3. Graph visualization



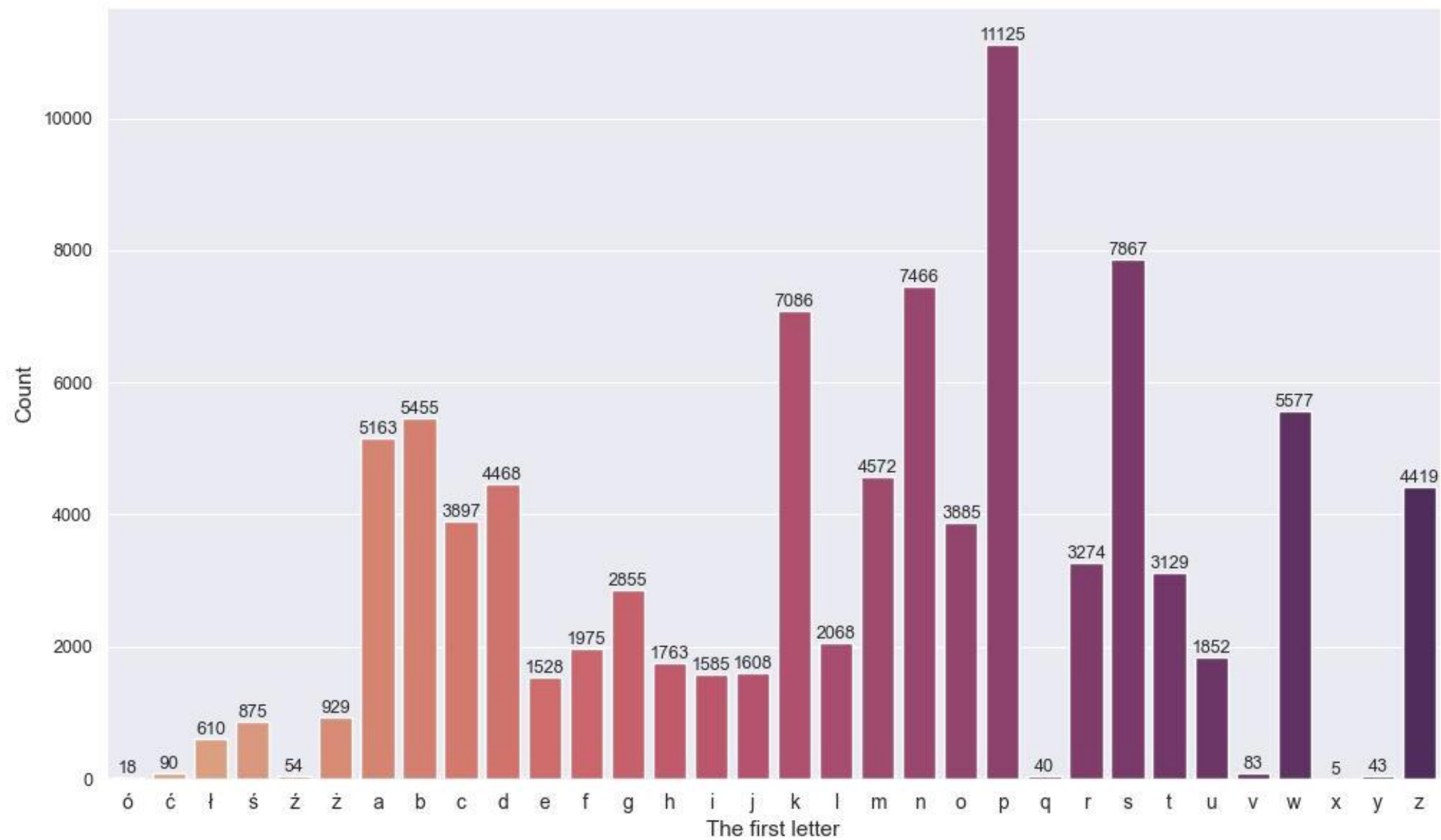
## 4. Statistics of headwords and their definitions

- The longest headwords in the dictionary are for *proverbs*. The TOP 3 are the following ones:
  - *W dzień świętej Agaty, jeśli słonko przez okienko zajrzy do chaty, to wiosenka na świat podgląda zza zimowej kraty.* (94 letters)
  - *Na świętego Walentego bywa zwykle mróz niczego, lecz jak Walek się rozdeszcze, pewne mrozy wrócą jeszcze.* (87 letters)
  - *Czasem luty się zlituje, że człek niby wiosnę czuje; ale czasem tak się zżyma, że człek prawie nie wytrzyma.* (86 letters)
- The longest definition of the headwords is for *gomukhasana* – this is the name for one of the yoga positions. It is defined by the following expression:
  - *pozycja siedząca, w której nogi są skrzyżowane w taki sposób, że lewe kolano znajduje się na szczycie prawego (lub odwrotnie), a lewa stopa i łydka wzdłuż prawego biodra, natomiast prawa stopa i łydka znajdują się na zewnątrz lewego biodra; plecy są wyprostowane; dłonie są splecione na plecach przy jednoczesnym unoszeniu prawego łokcia w kierunku sufitu i wyciąganiu lewego w kierunku podłogi (lub odwrotnie, przy odwrotnym splocie nóg).* (439 letters, 67 words)

Histogram of number of words in definitions

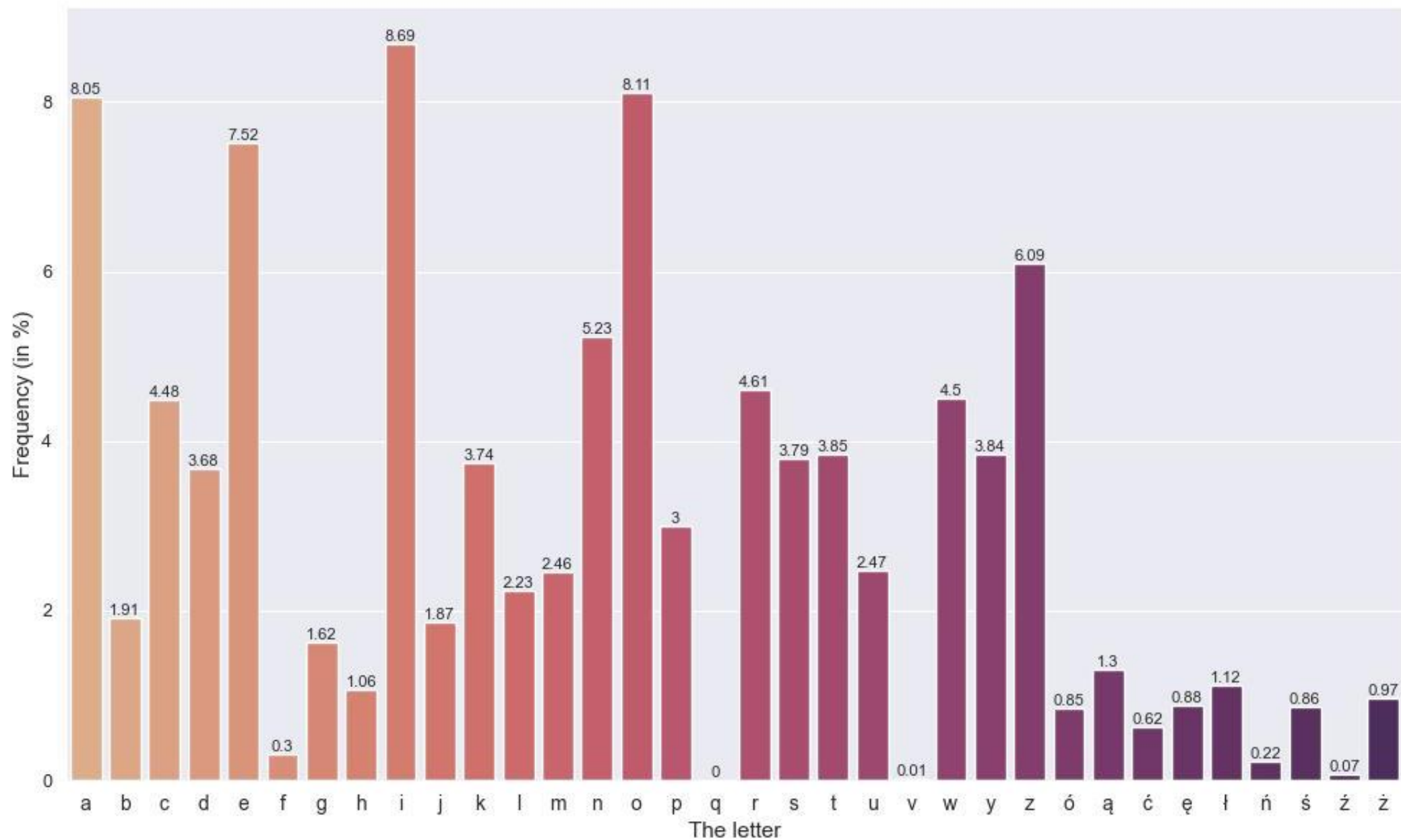


Most popular first letters of the words



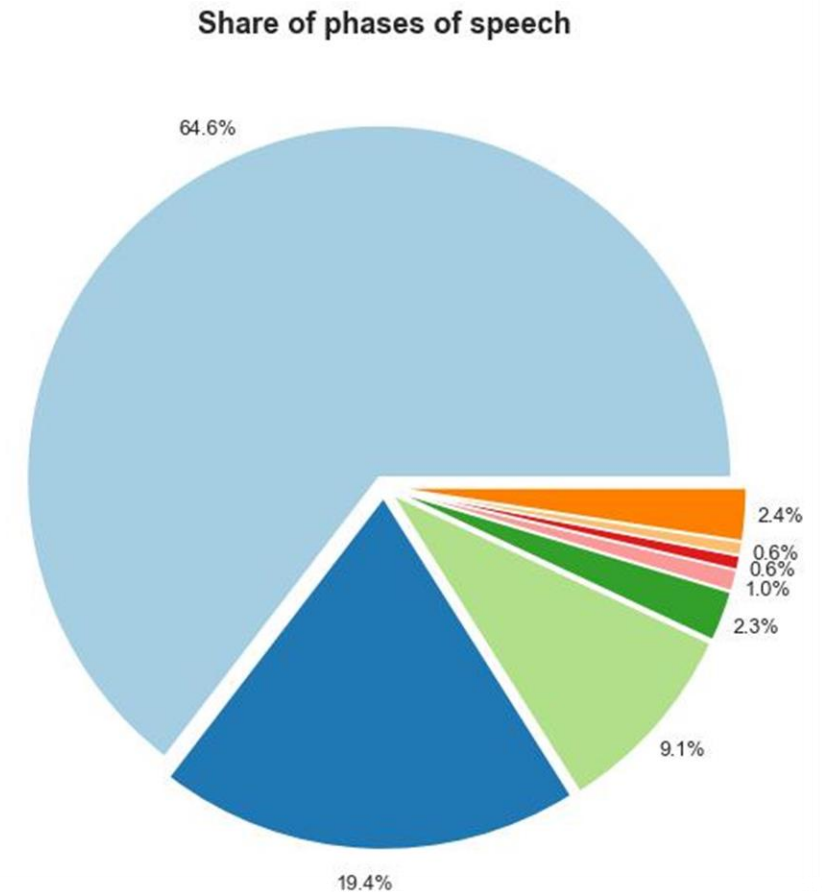


Most popular letters across the definitions of the words

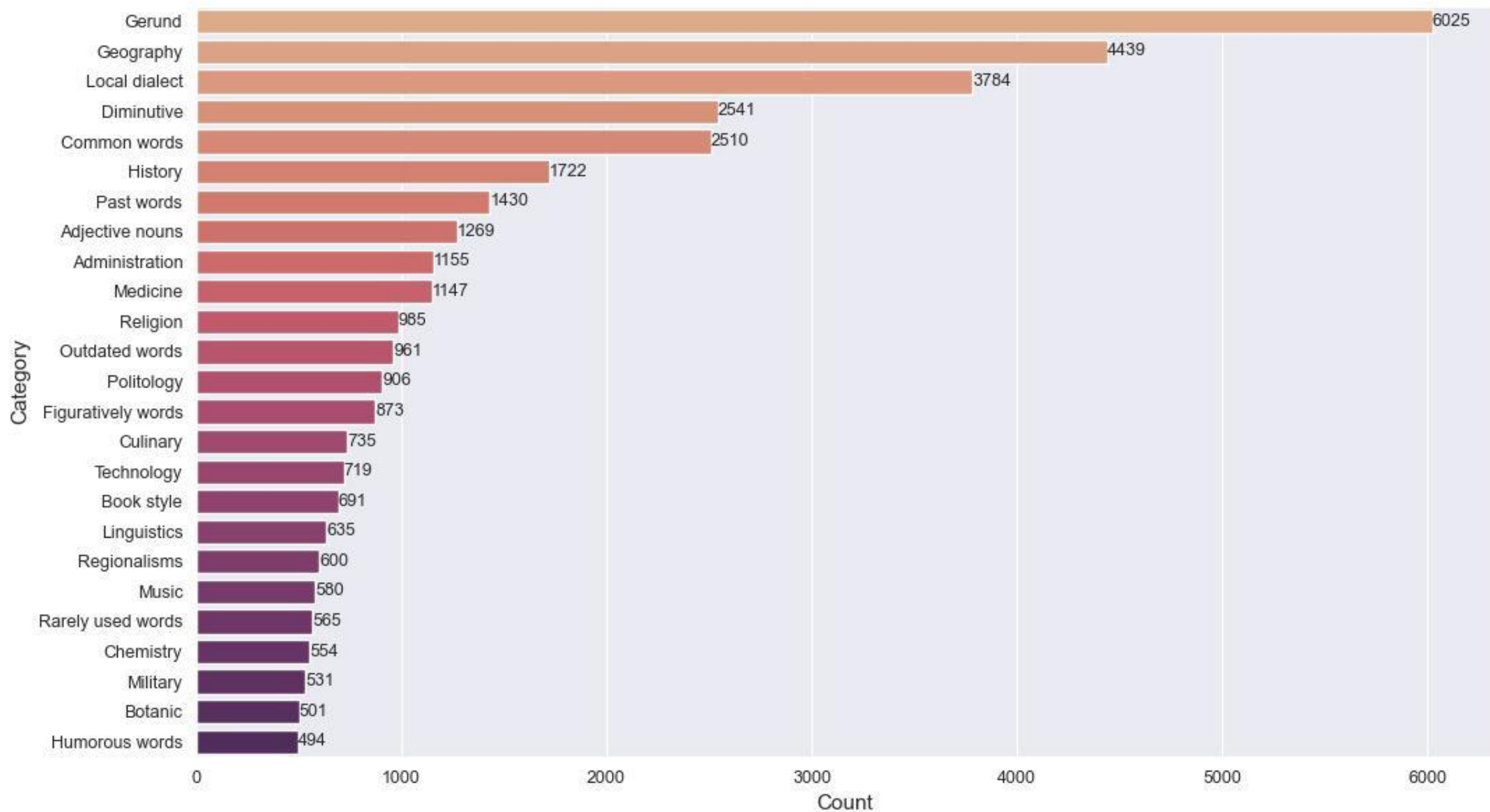


# Phases of speech

- The most part of all headwords are **nouns**, **adjectives** and **verbs**. The least part are **expressive words**, **lower forms** and **interfixes**.
- There are a few headwords from other languages mainly are descended from **English**, **Polabian** (West Slavic language) and **French**.
- The headword with the biggest number of definitions is verb – **tluc**, which has **21** different meanings.



## Types, categories and stylistic devices of words



## 75 most popular Polish nouns in Wiktionary



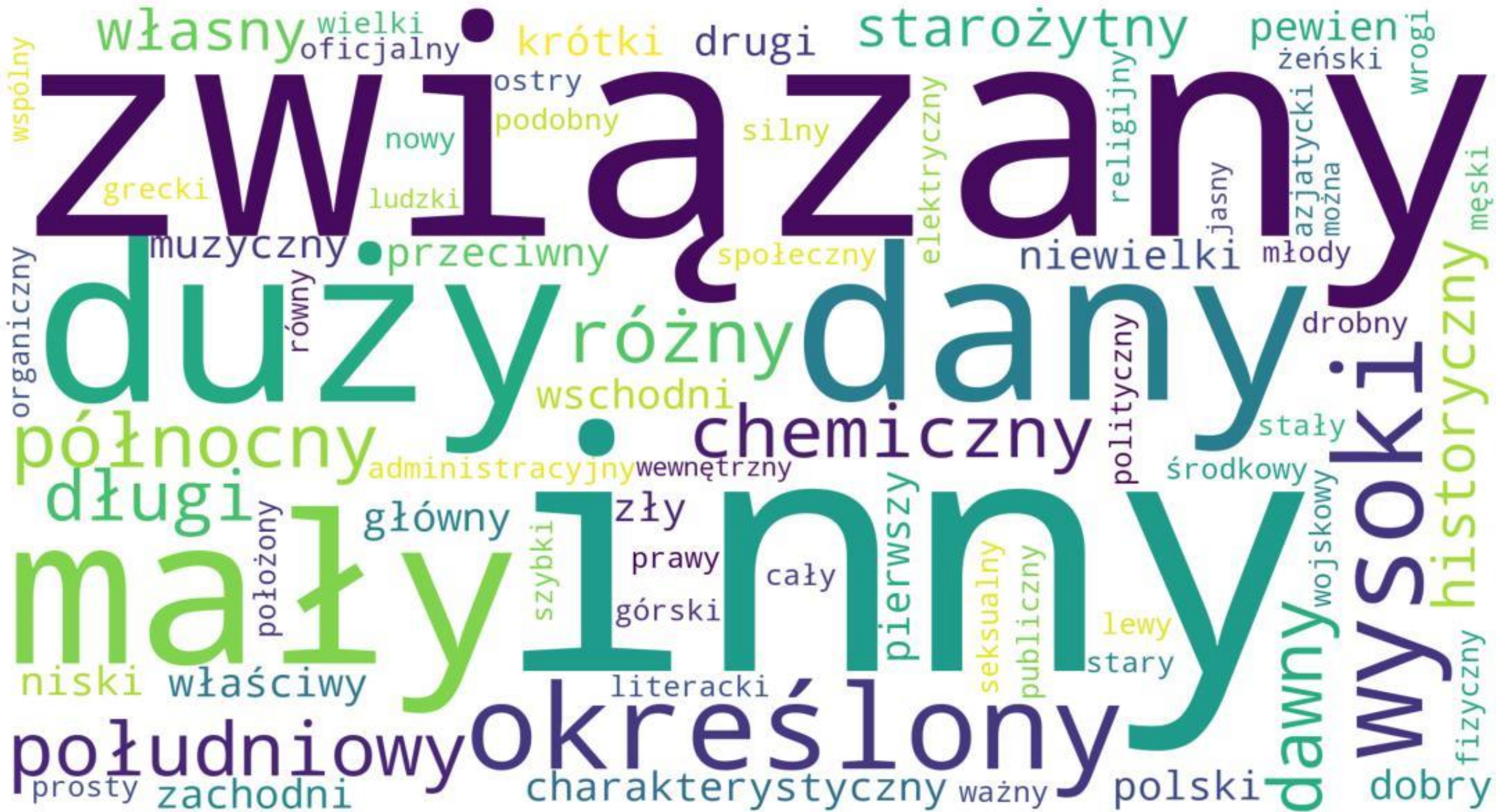


## 75 most popular Polish verbs in Wiktionary





## 75 most popular Polish adjectives in Wiktionary





## 75 most popular Polish adverbs in Wiktionary

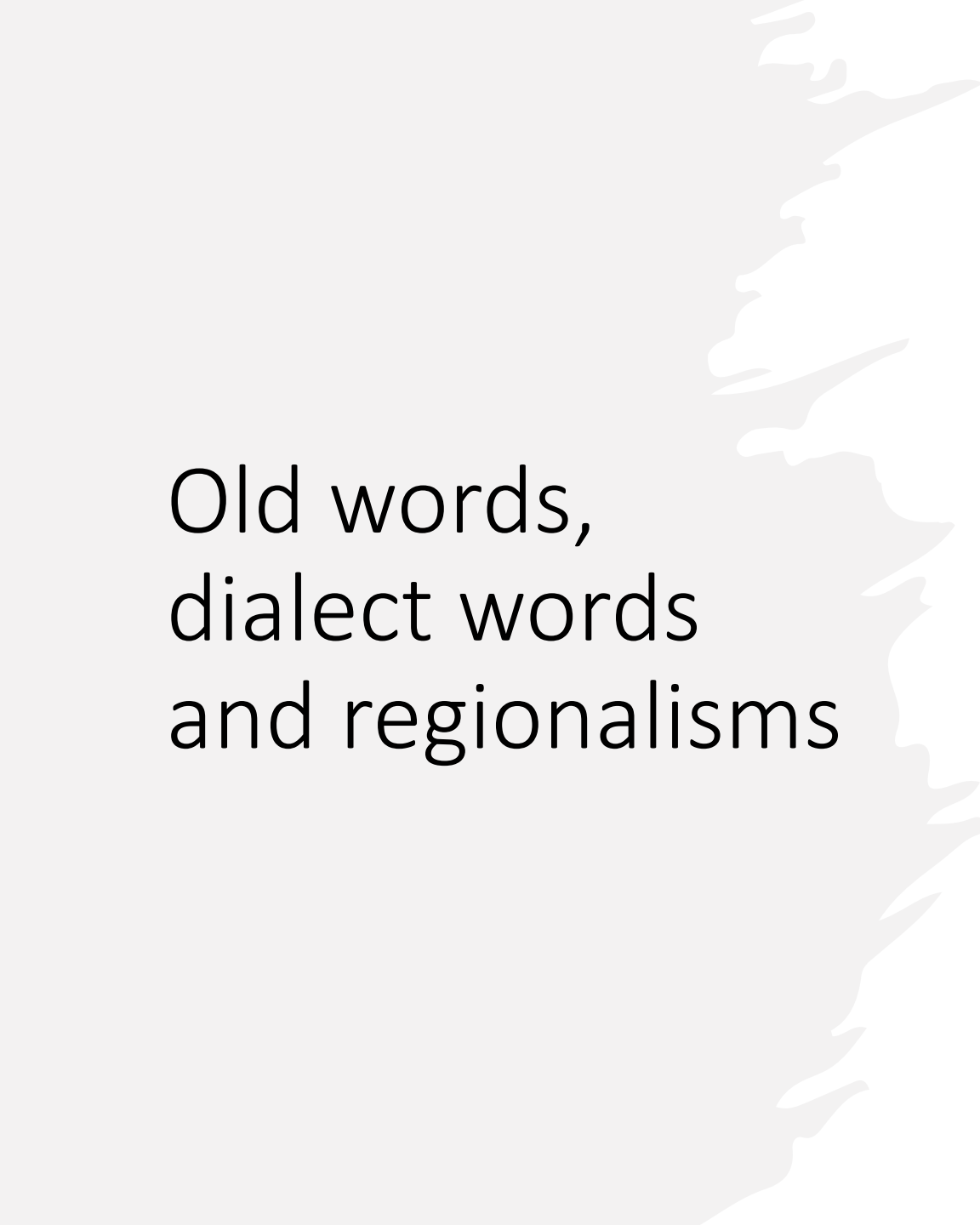




## 150 most popular Polish words in Wiktionary



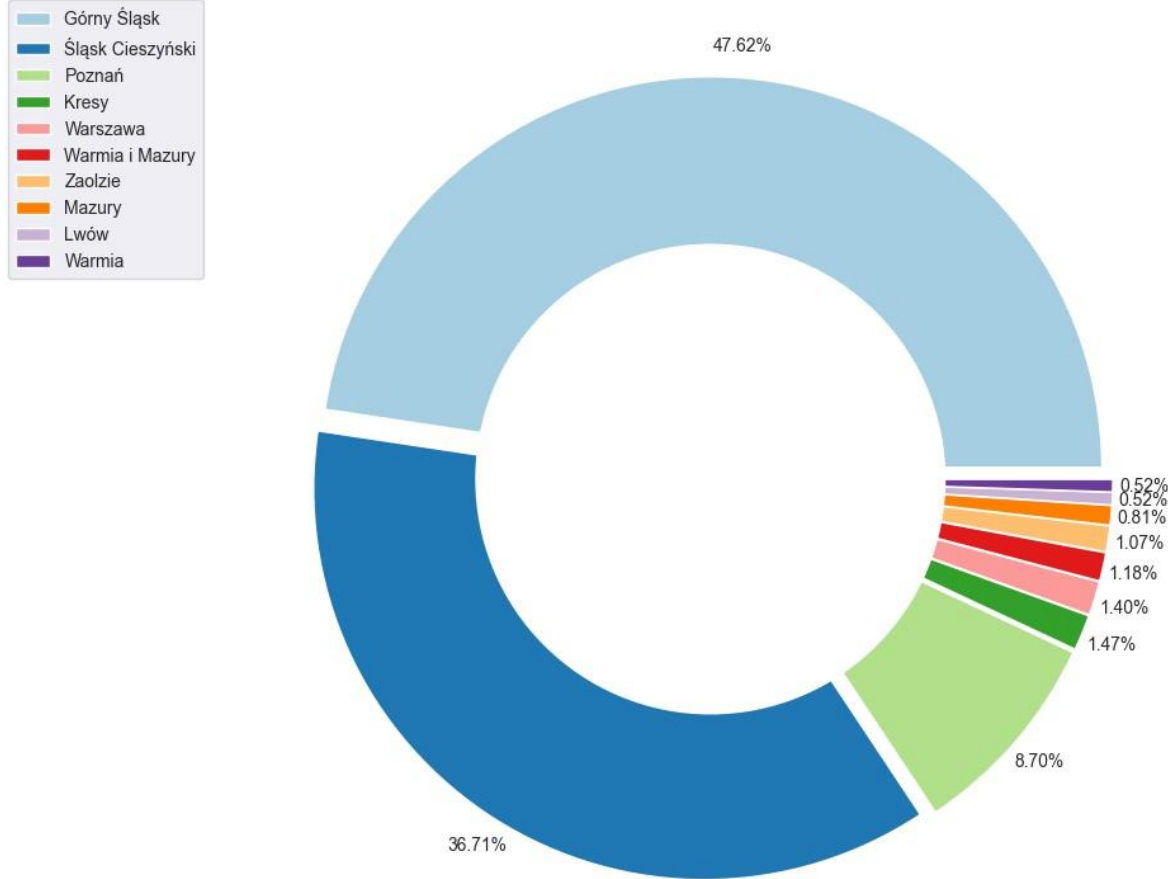




# Old words, dialect words and regionalisms

- There are 406 headwords descended from Old Polish, 28 headwords descended from Middle Polish, 600 regionalisms and 3784 dialect words.
- Most of dialect words are descended from Silesia, when most of the regionalisms come from Wielkopolska.
- Among the Wiktionary there are a few very odd headwords as ćpnięcie, łożgol, ófnol, praśny or zozula...

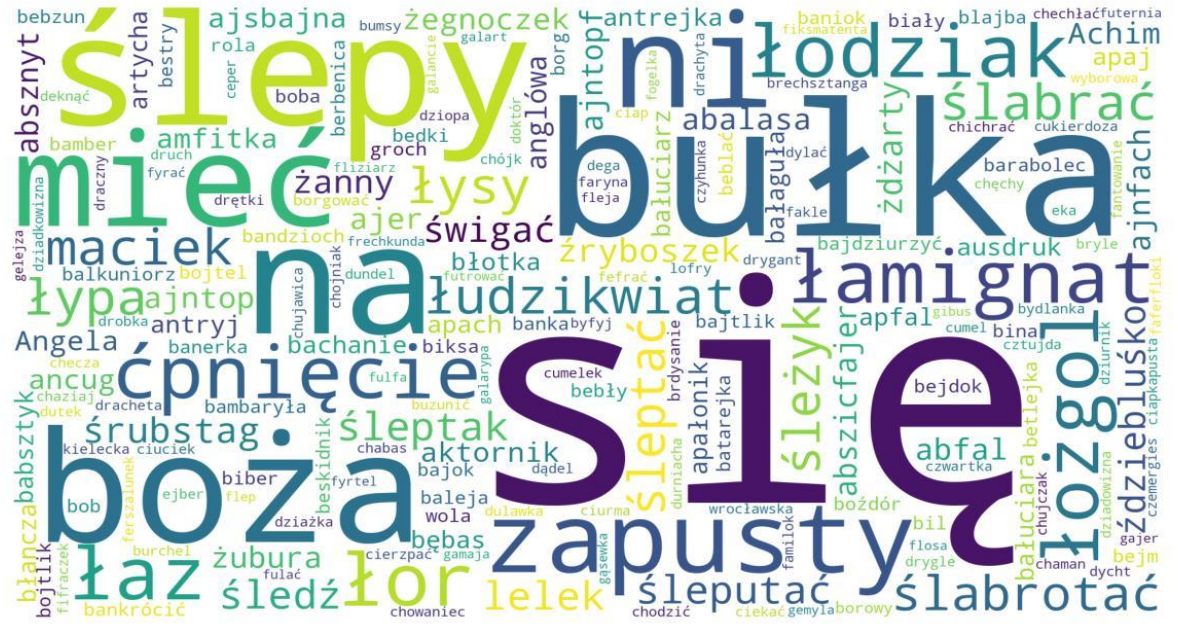
Frequency of destination of local distinct words in Polish language



Popular local dialect Polish words

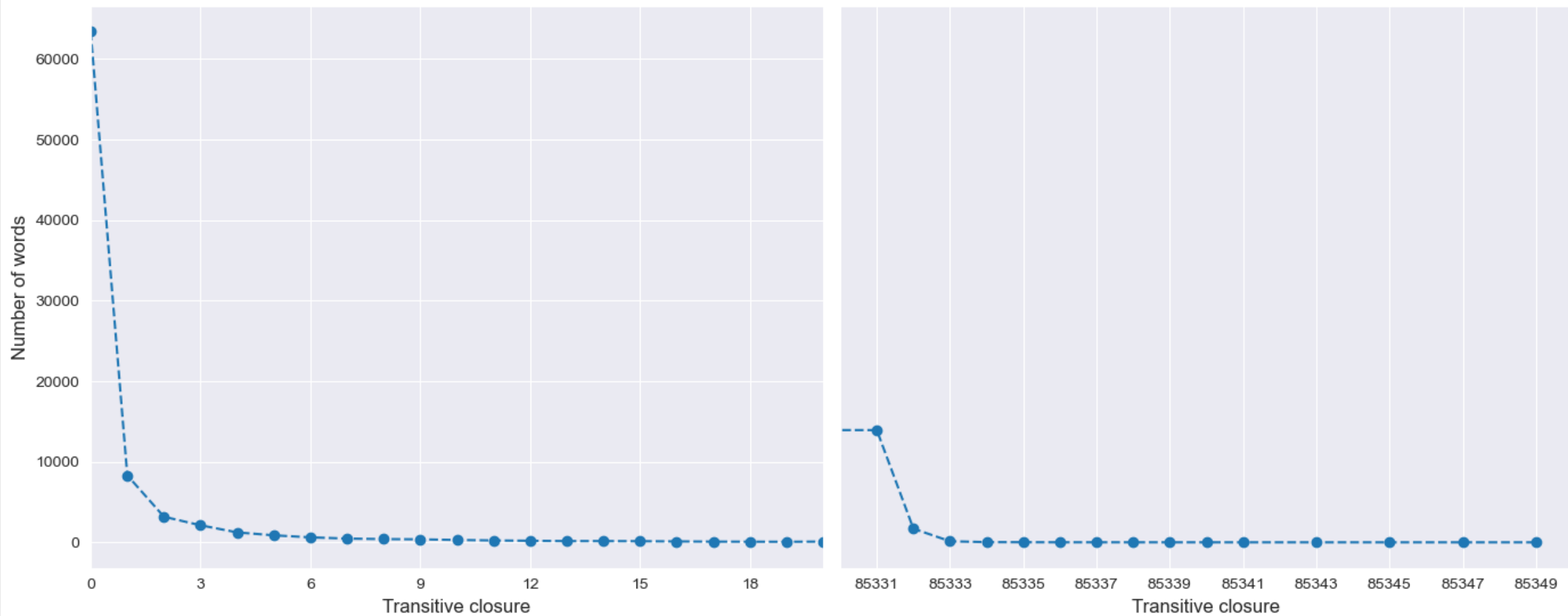


Popular Polish regionalisms



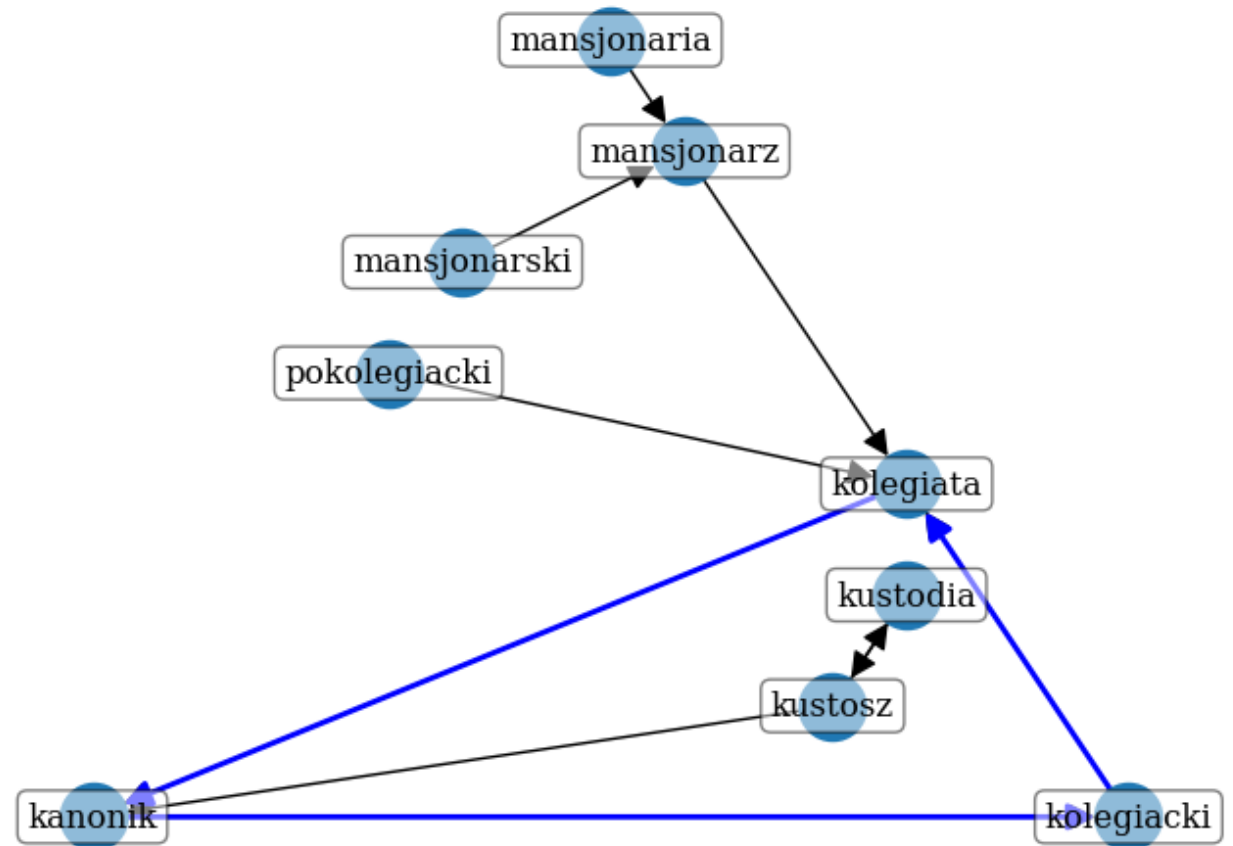
# Transitive closure of words

Transitive closure of words

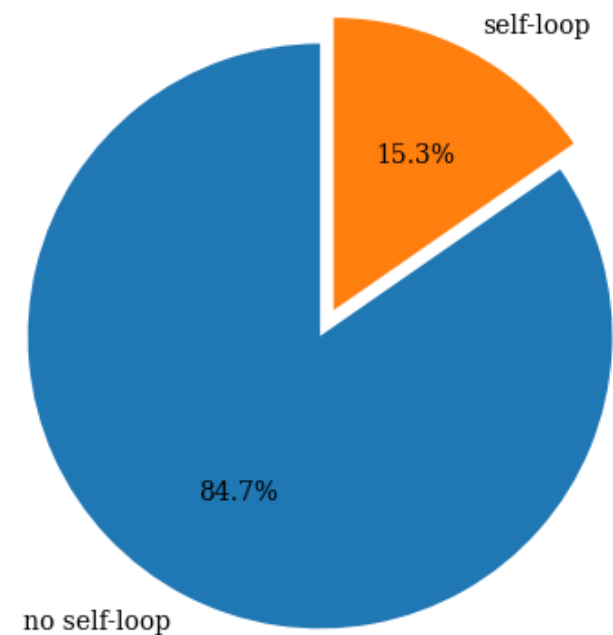
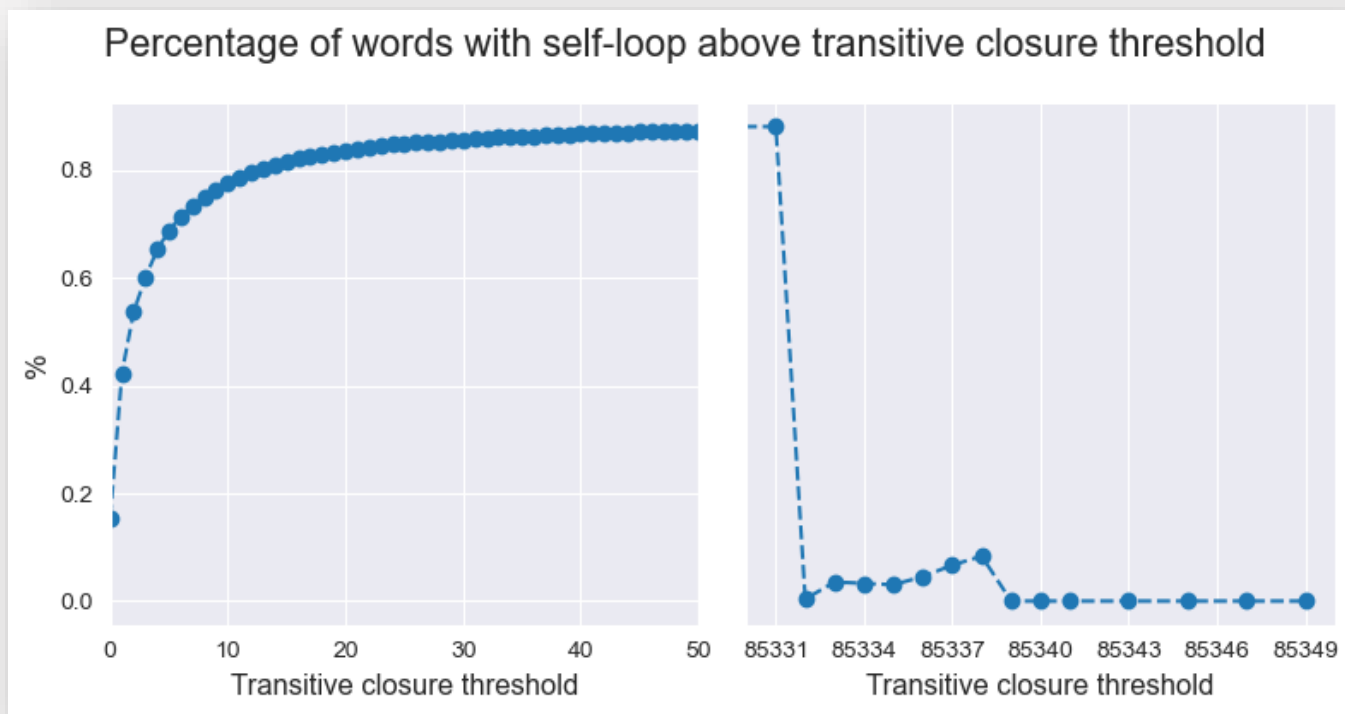


# Crawled graph

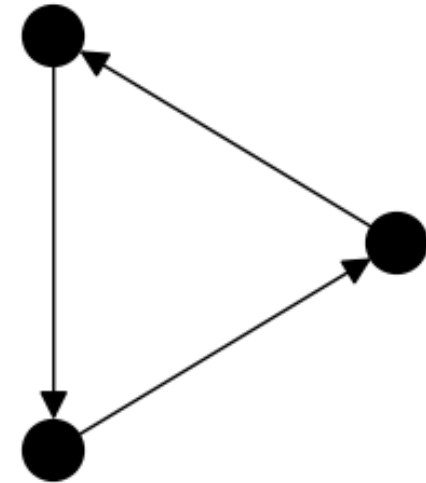
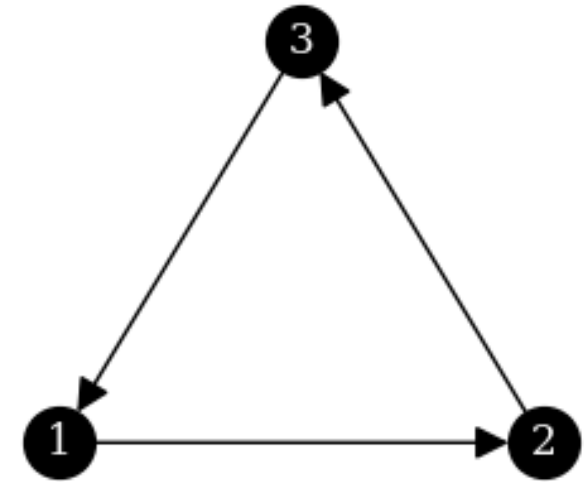
- We create graph which starts from word „kanonik” and then goes to all words which are defined by it even indirectly.
- Transitive closure of „kanonik” is 11.



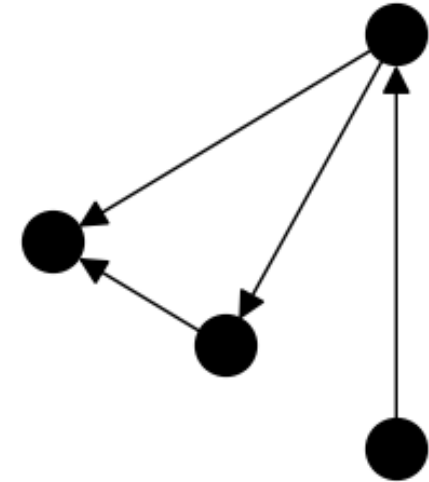
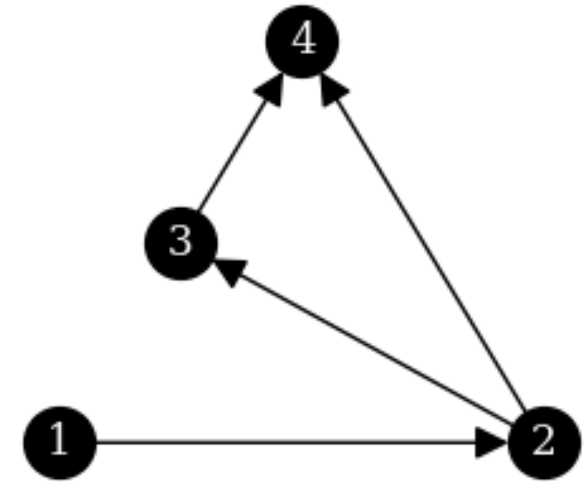
# Loops in the definitions



Uniqueness of nodes  
in graph - Example 1



# Uniqueness of nodes in graph - Example 2





## 8. Method to find unique nodes

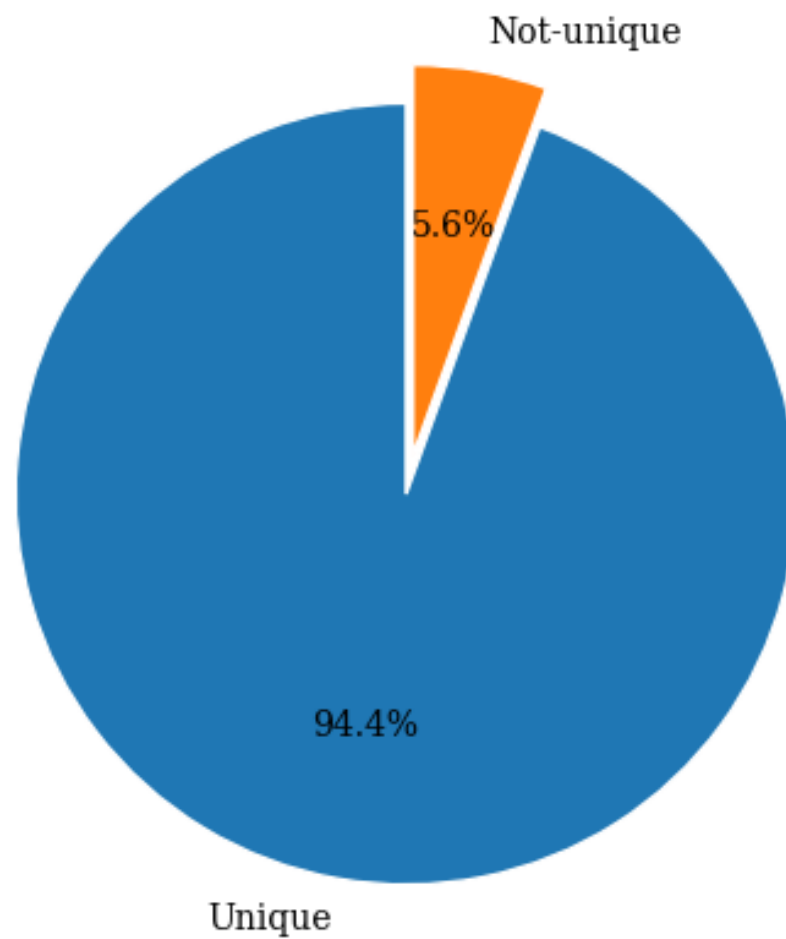
For every node we construct vector of:

- In-degree,
- Out-degree,
- Transitive closure,
- Self-loop (0, 1),
- Sum of in/out-degrees of out/in neighbors of node,
- Number of in/out-neighbors of out/in neighbors of node.





# Result of the method



## 9. Summary

- Hypothesis that headwords are well defined by their relationships between their definitions deserves at least some plus.
- We found the most important words and headwords which are used to define others.

Thank You  
for Your  
attention!

