



# Limba romana in diferite regiuni din diaspora





**Este limba romana scrisa in  
diaspora diferita fata de ce  
vedem in Romania ?**



By Olivia Wilson

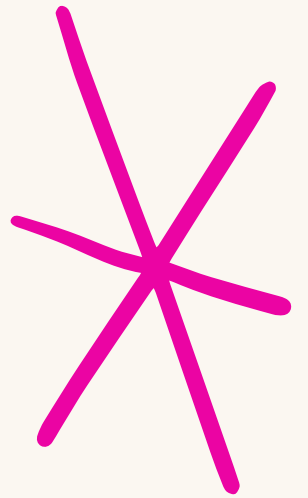
Urmeaza sa aflam ...

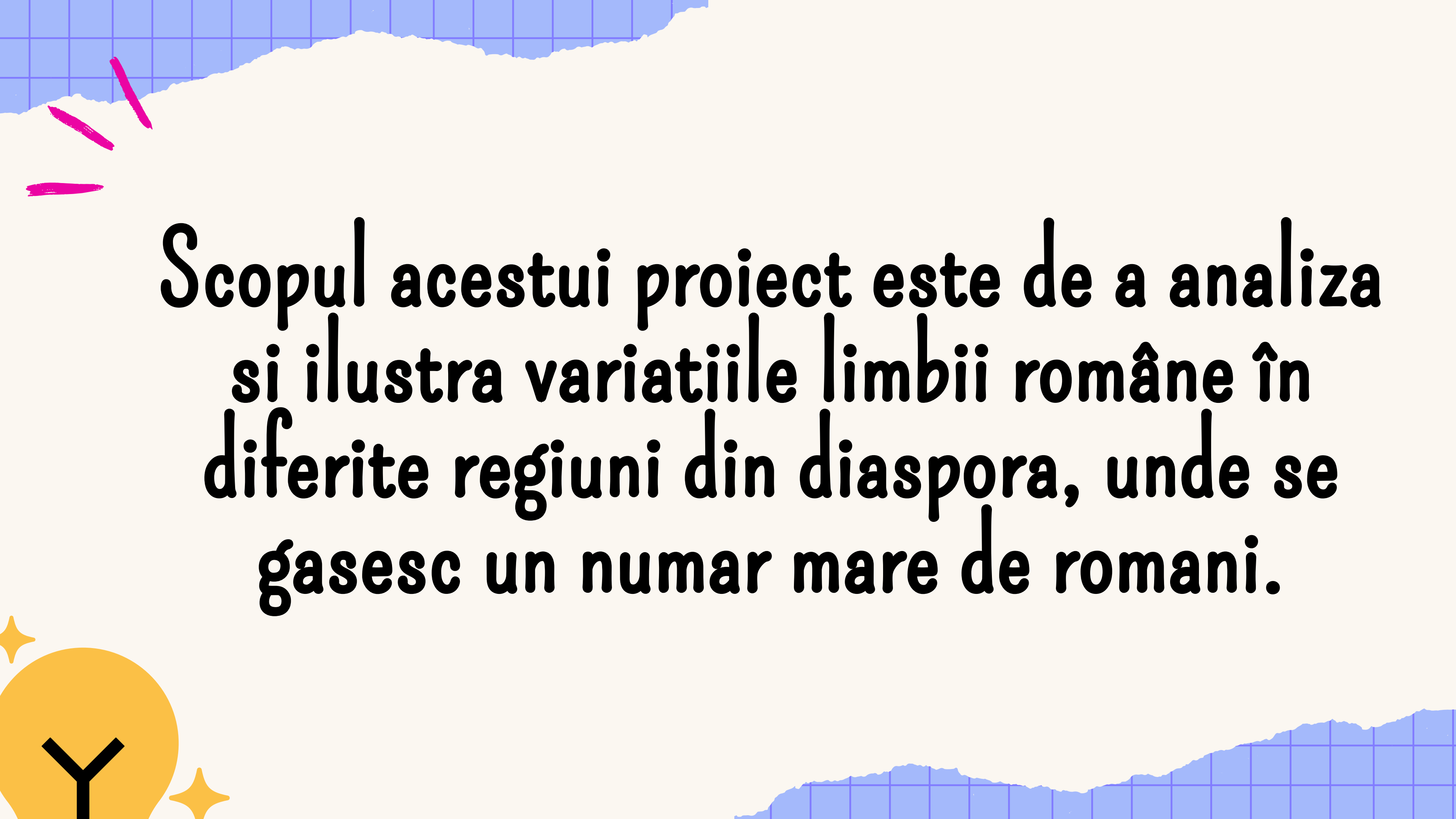


By Olivia Wilson



In primul rand, despre ce este  
proiectul ..





**Scopul acestui proiect este de a analiza  
si ilustra variatiile limbii române în  
diferite regiuni din diaspora, unde se  
gasesc un numar mare de romani.**



# De ce am ales asta ?

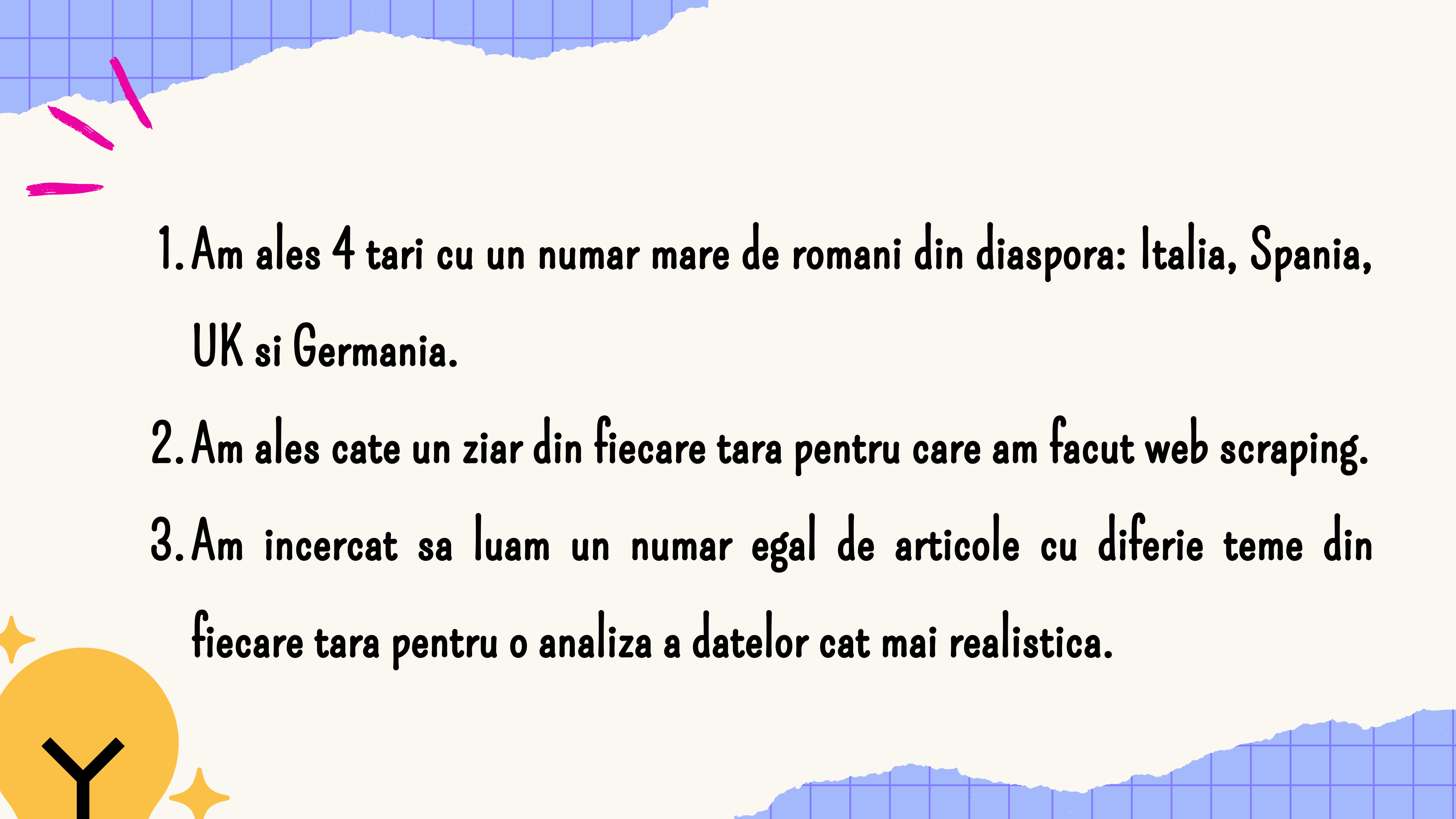
Am fost intrigati de alegerile din Romania ce au avut loc in acest an si am vrut sa vedem cum isi iau romanii din diaspora informatii si daca acestea sunt in vreun fel diferite.





Cum am facut asta ?



- 
1. Am ales 4 tari cu un numar mare de romani din diaspora: Italia, Spania, UK si Germania.
  2. Am ales cate un ziar din fiecare tara pentru care am facut web scraping.
  3. Am incercat sa luam un numar egal de articole cu diferite teme din fiecare tara pentru o analiza a datelor cat mai realistica.

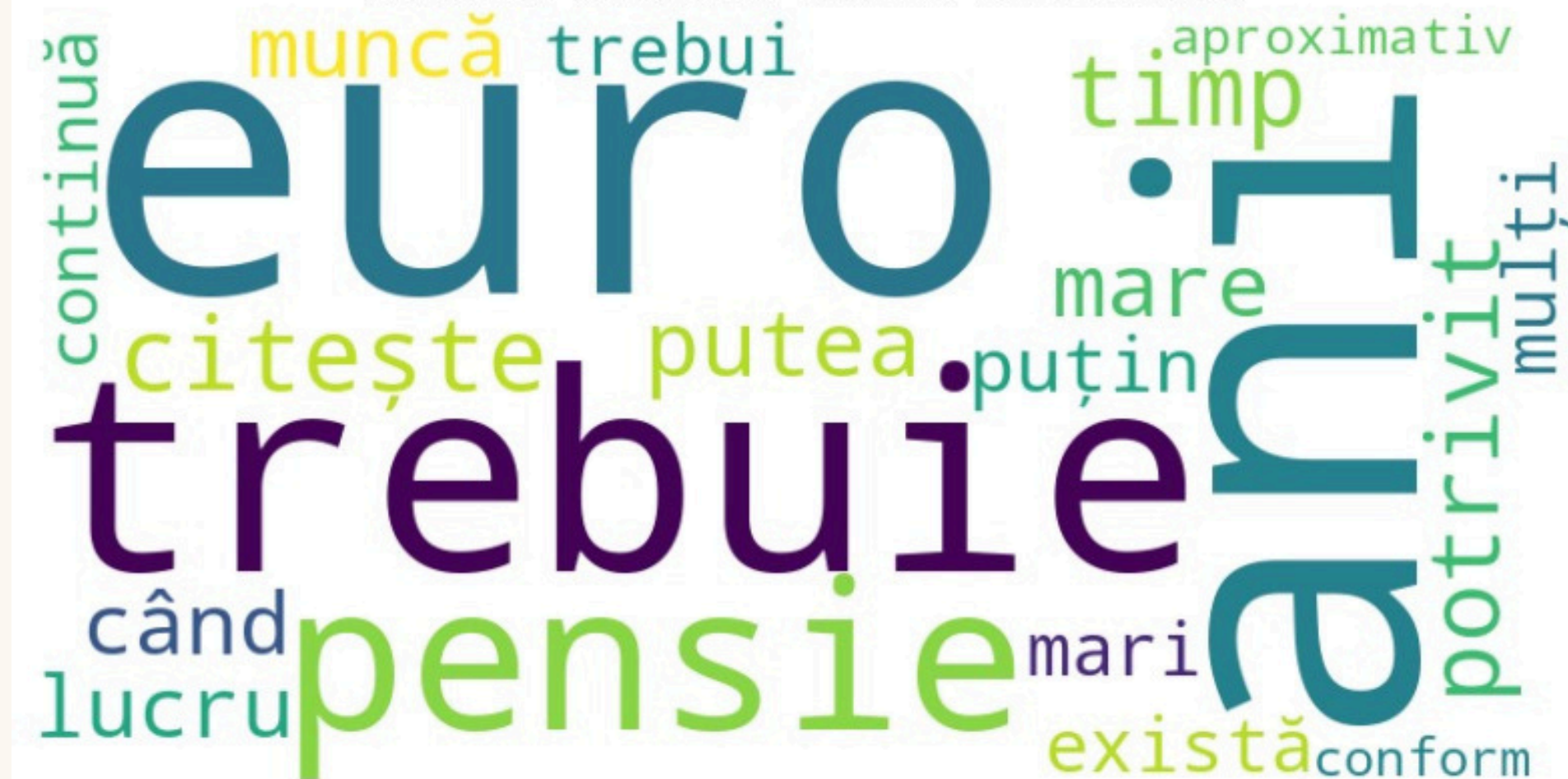


Cateva grafice obtinute ..



# Cele mai frecvente cuvinte din fiecare tara - Word Cloud

Cele mai frecvente cuvinte în Germania



Cele mai frecvente cuvinte în Italia





# Cele mai frecvente din fiecare tara - Word Cloud

## Cele mai frecvente cuvinte în UK

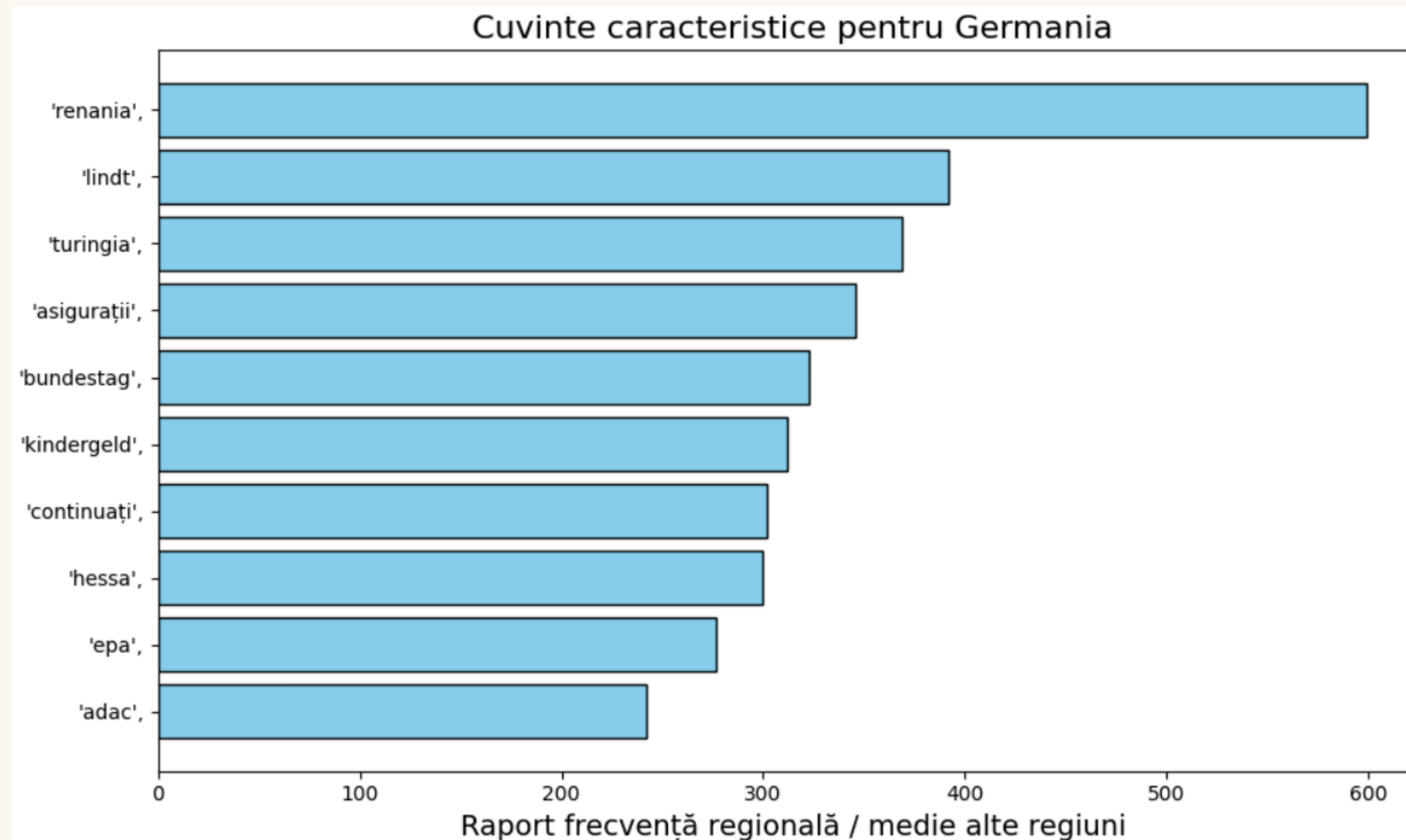
medie spus putea  
ani trebuie decât lire  
asemenea timp declarat unit  
ster milioane  
an faceți total mare  
londra creștere

## Cele mai frecvente cuvinte în Spania

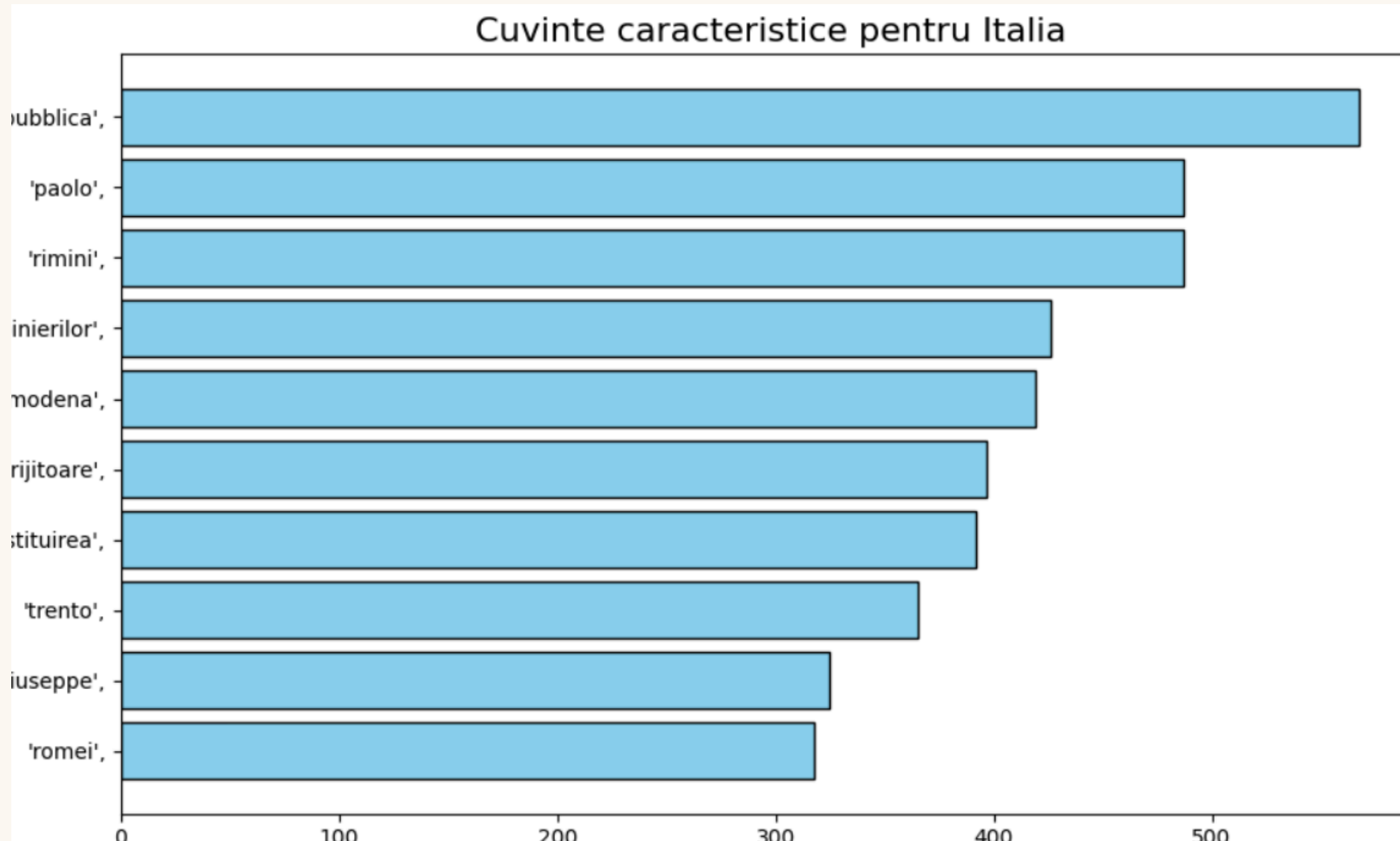
Word cloud visualization of the text, featuring the word "madrîd" (Madrid) prominently in the center, surrounded by other words like "ani" (years), "trebuie" (must), "când" (when), "precum" (as), "cadru" (framework), "muncă" (work), "regatul" (the kingdom), "mare" (big), "timp" (time), "își" (has), "anul" (the year), "general" (general), "foto" (photo), "cazul" (the case), "asemenea" (also), "loc" (place), and "euro".

# Cuvinte caracteristice pentru fiecare tara

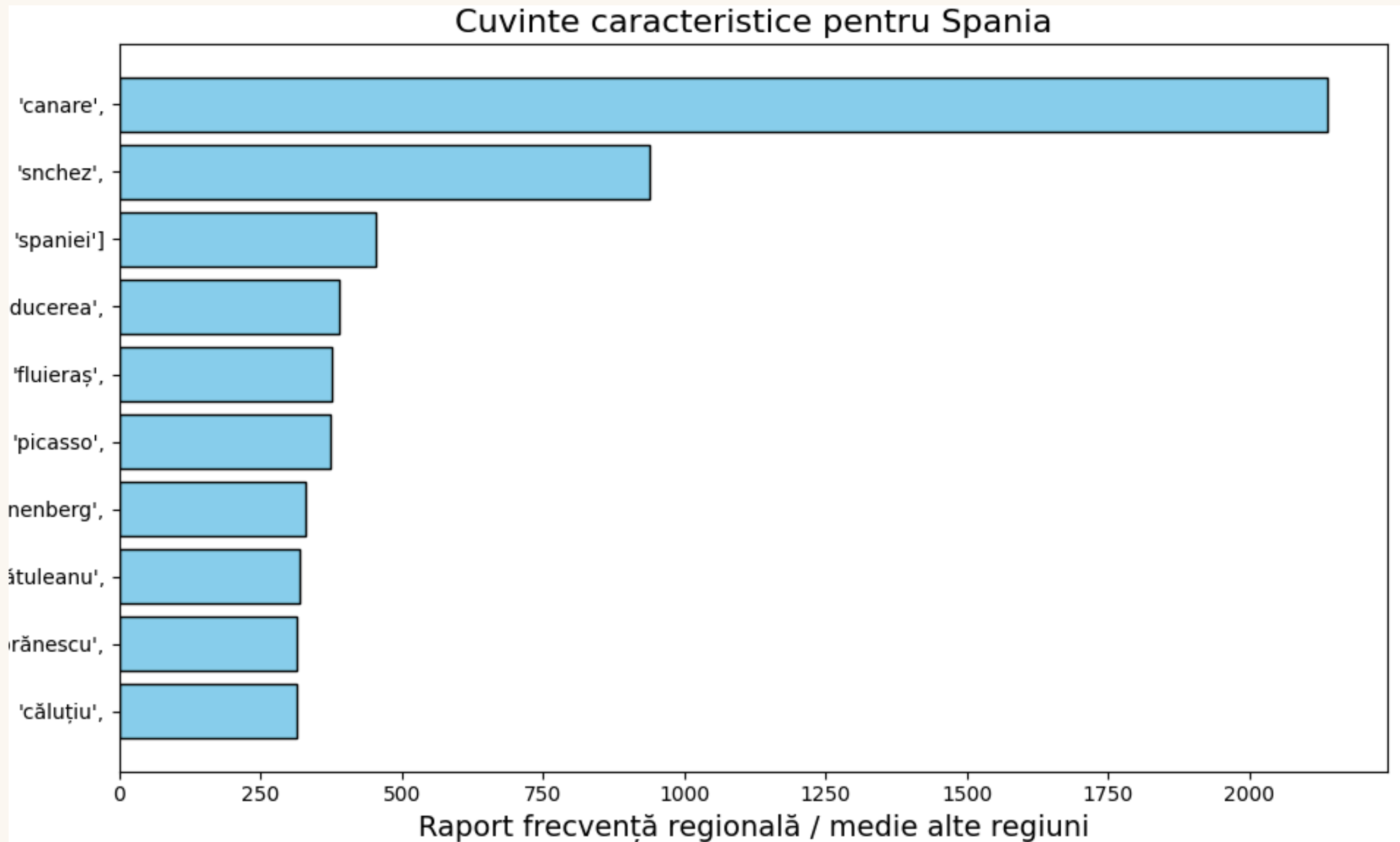
Calculam frecventa fiecarui cuvant în fiecare regiune si determinam ce cuvinte sunt caracteristice comparând frecventa lor cu alte regiuni.



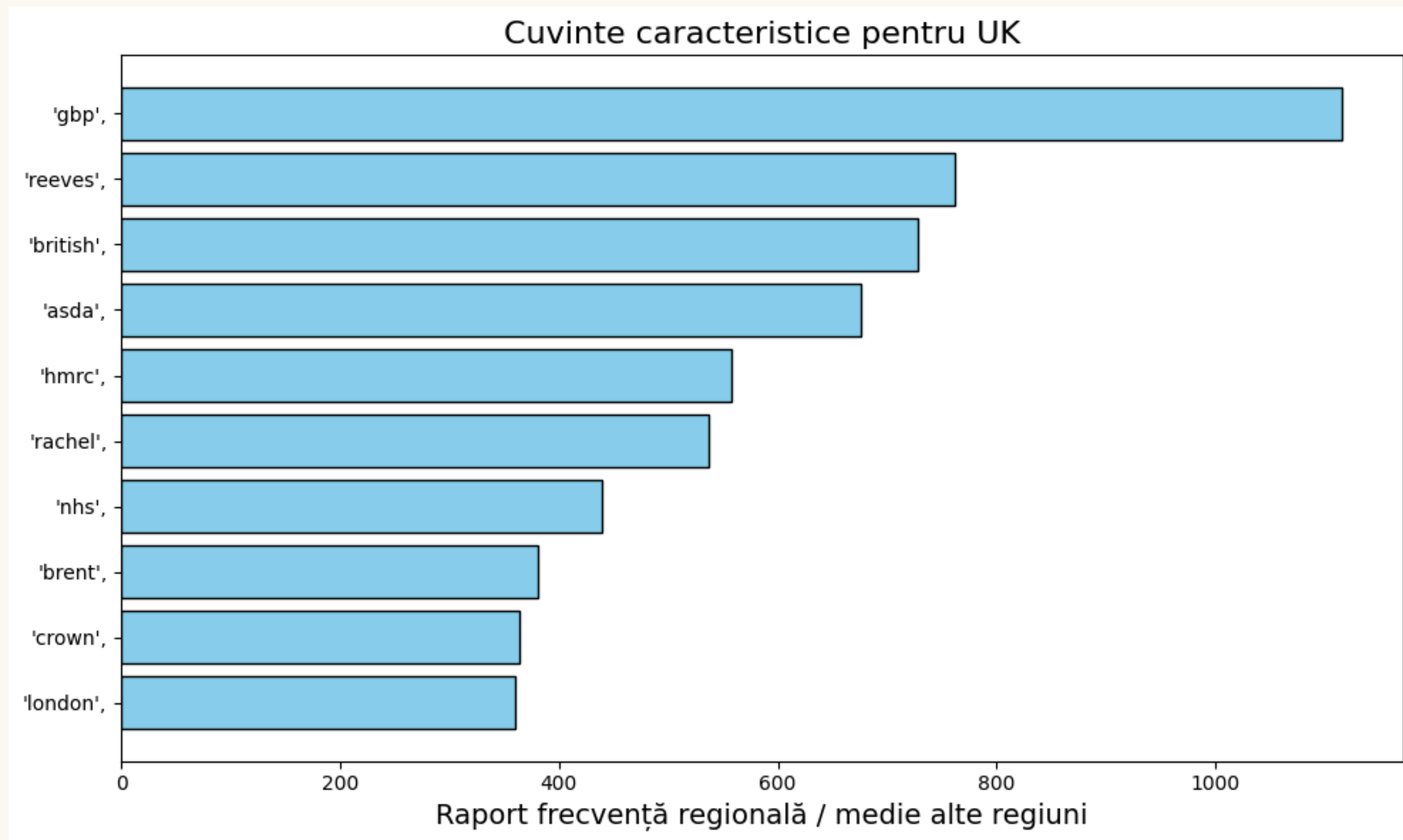
# Cuvinte caracteristice pentru fiecare tara



# Cuvinte caracteristice pentru fiecare tara

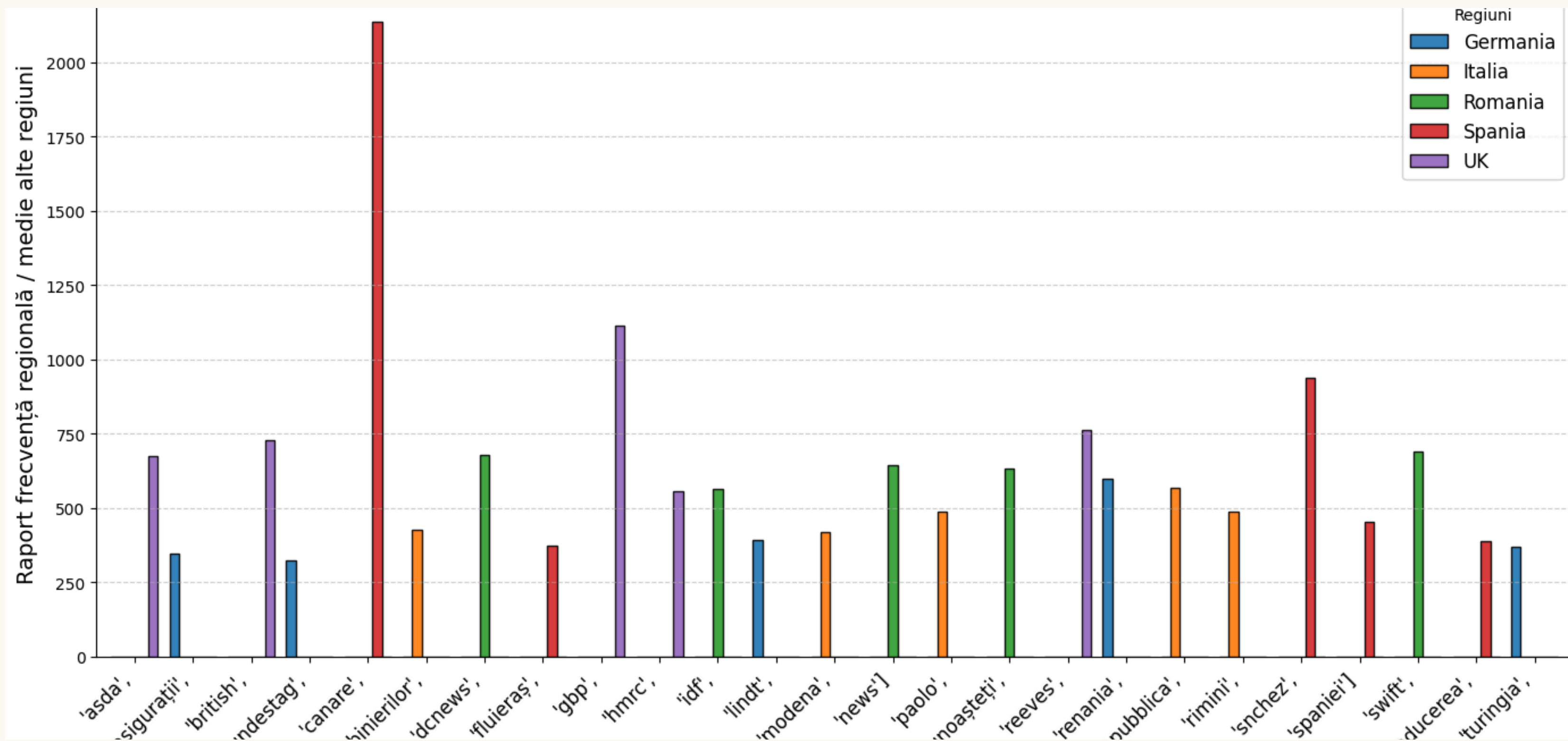


# Cuvinte caracteristice pentru fiecare tara





# Cuvinte caracteristice pentru fiecare tara



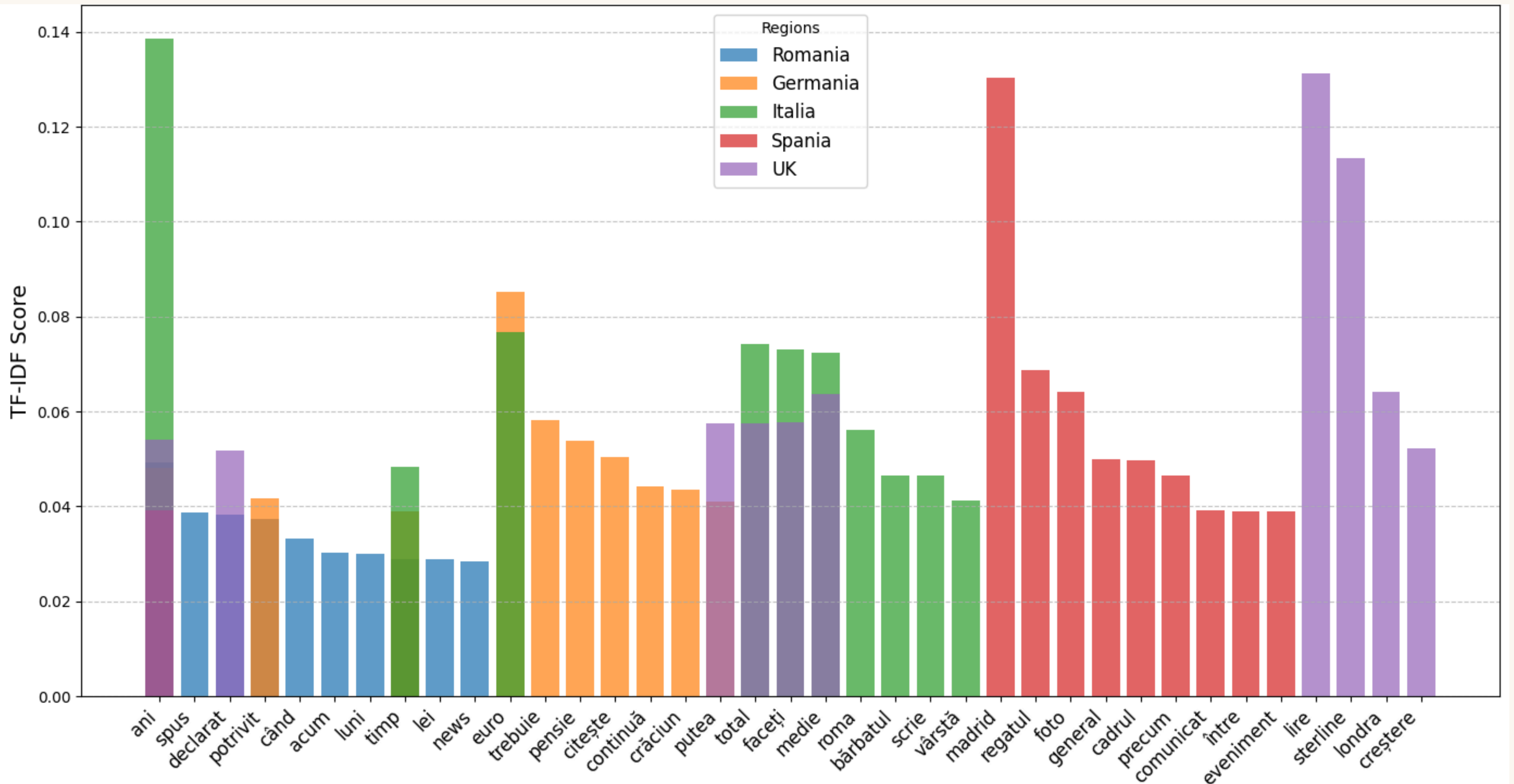


# TF - IDF

TF-IDF pune accent pe importanta relativa a cuvintelor, penalizand cele care sunt prea comune in toate articolele.

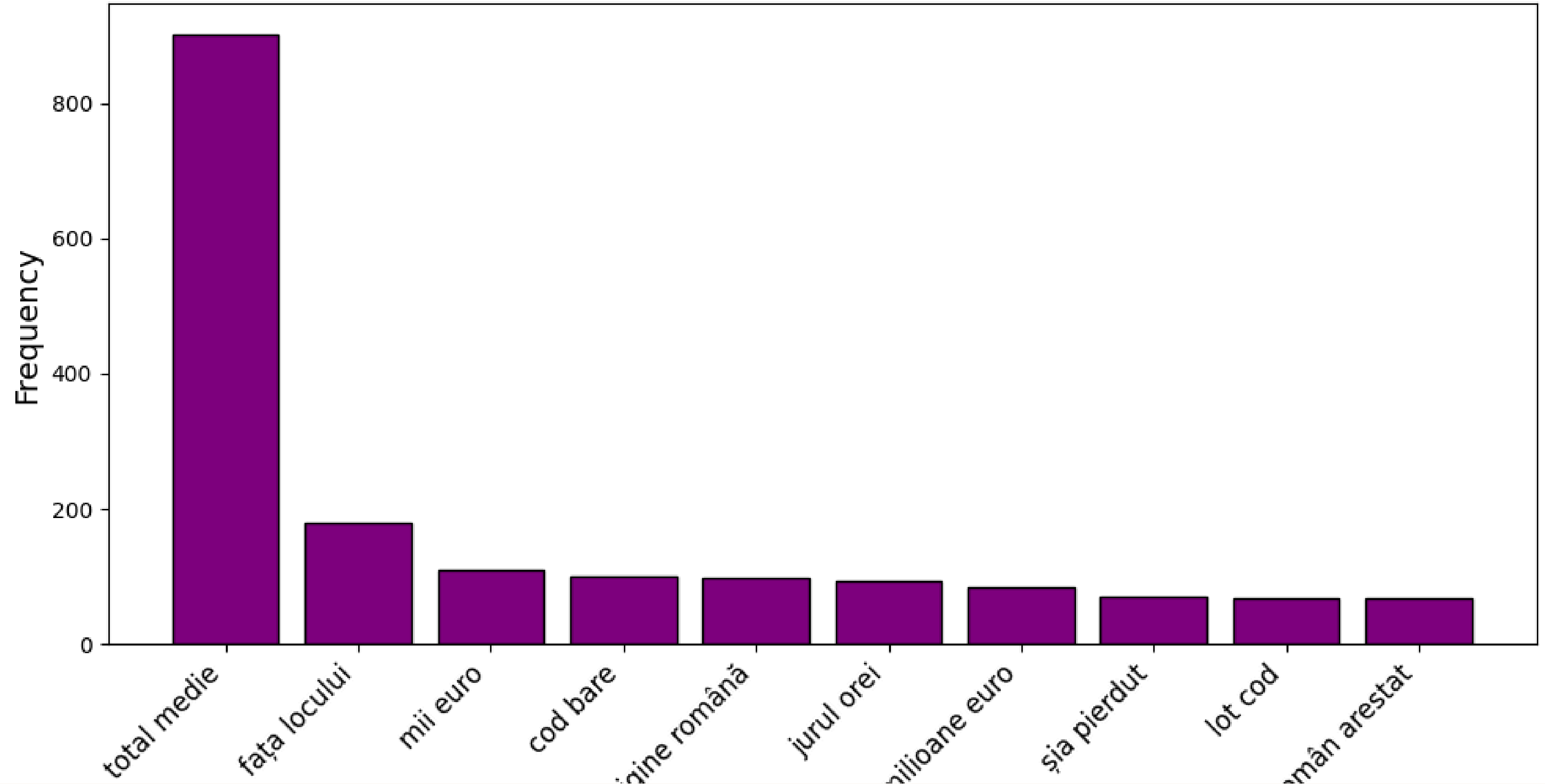
1. Calculeaza TF (cat de des apare un cuvant intr-un articol) si IDF (cat de rar este cuvantul in toate articolele).
2. Penalizeaza cuvintele foarte frecvente (ex: „si”, „de”, „in”), eliminandu-le ca fiind nerelevante.
3. Compara intre articole din toate regiunile, nu doar in cadrul unei singure regiuni.

# TF - IDF

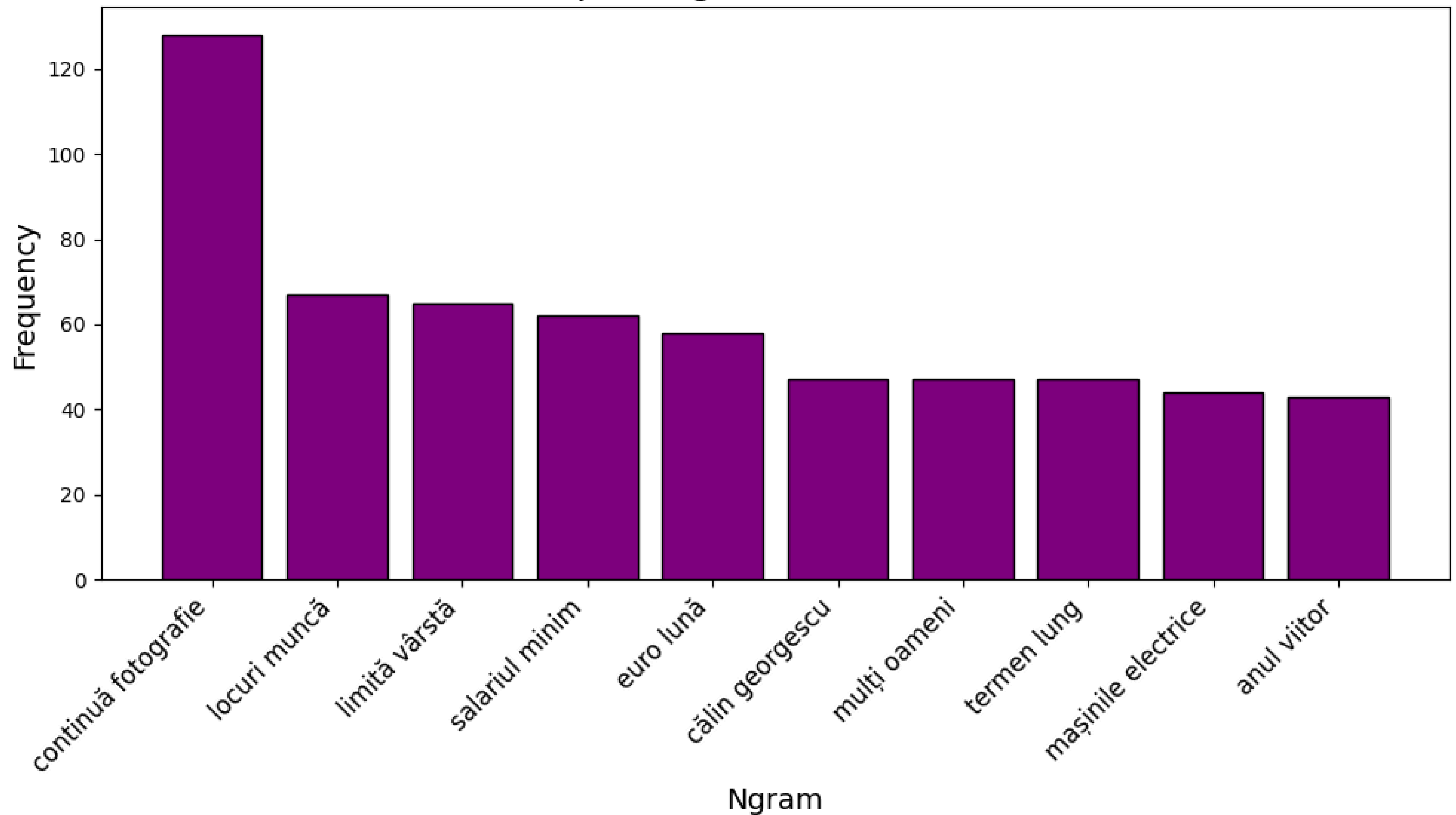


# BIGRAME

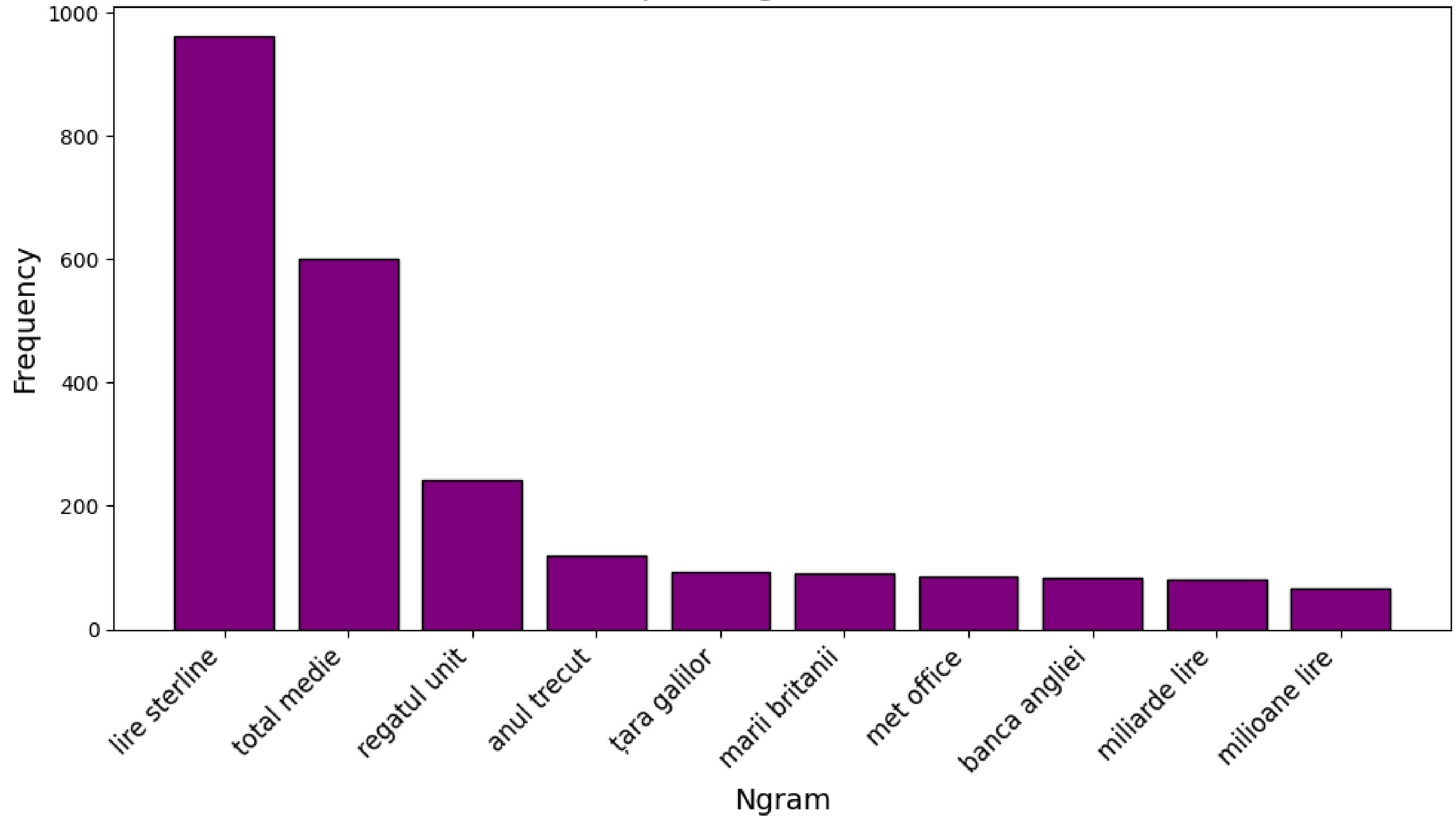
Top 10 bigrams în Italia



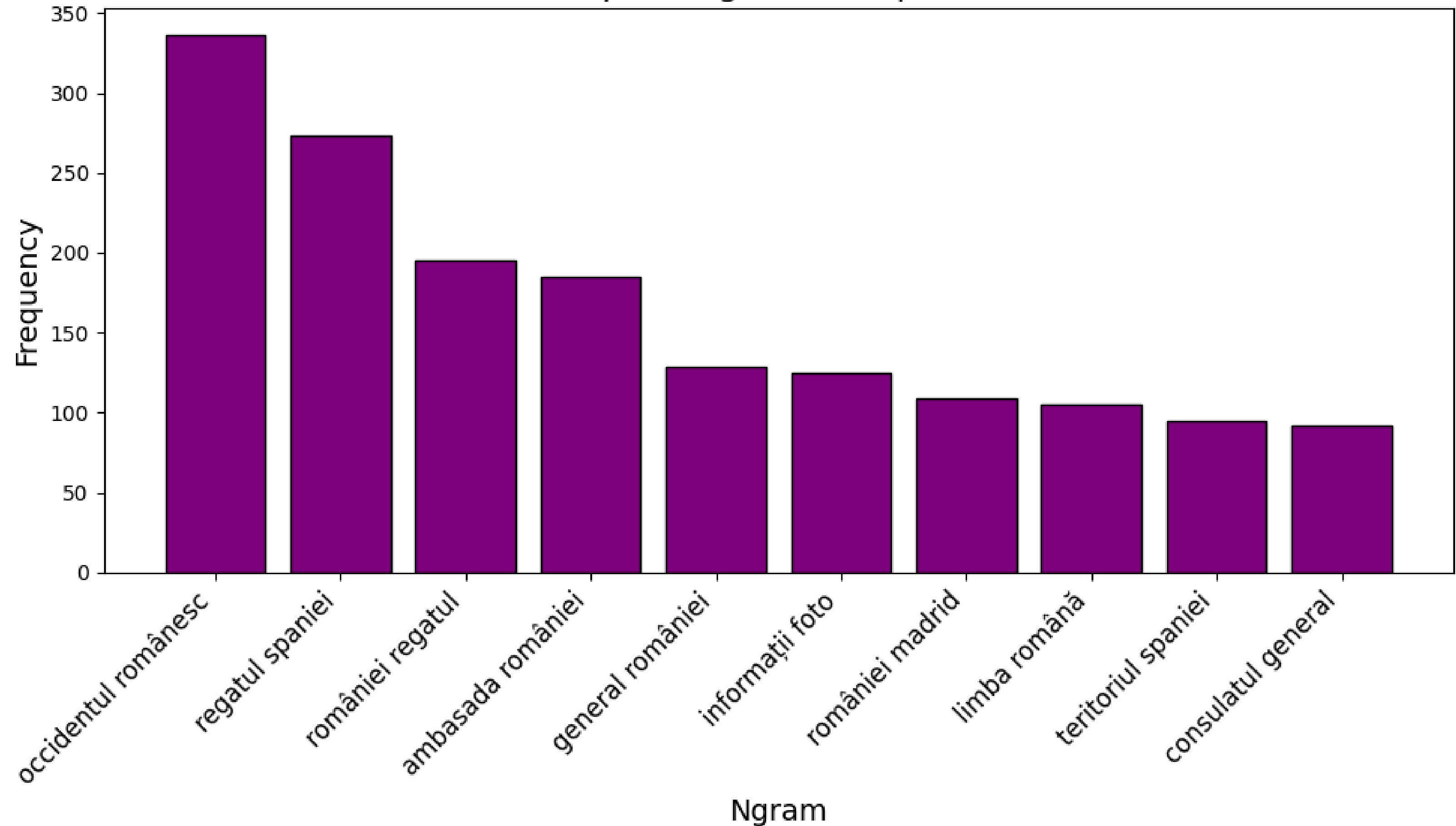
Top 10 bigrams în Germania



Top 10 bigrams în UK

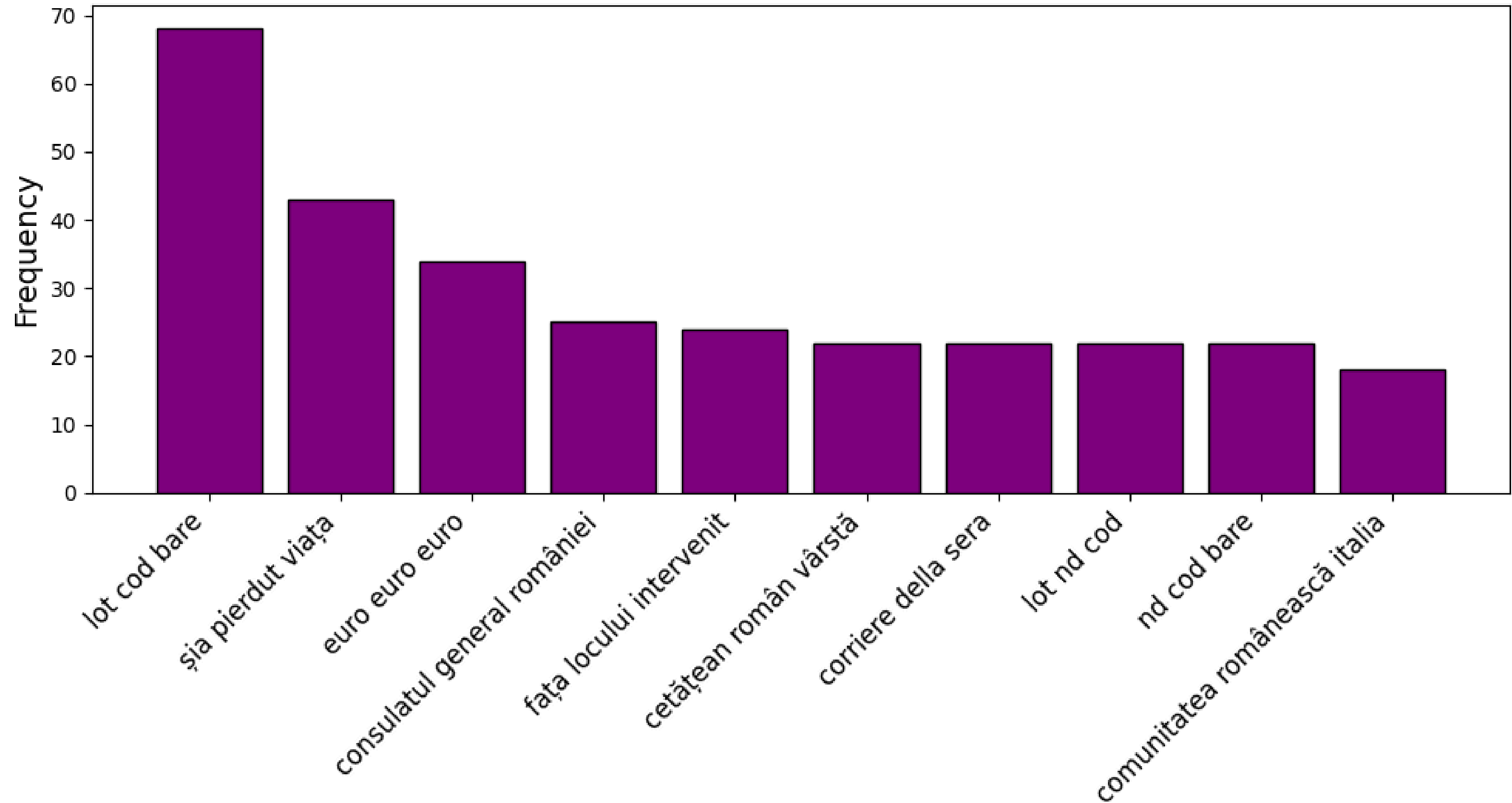


Top 10 bigrams în Spania

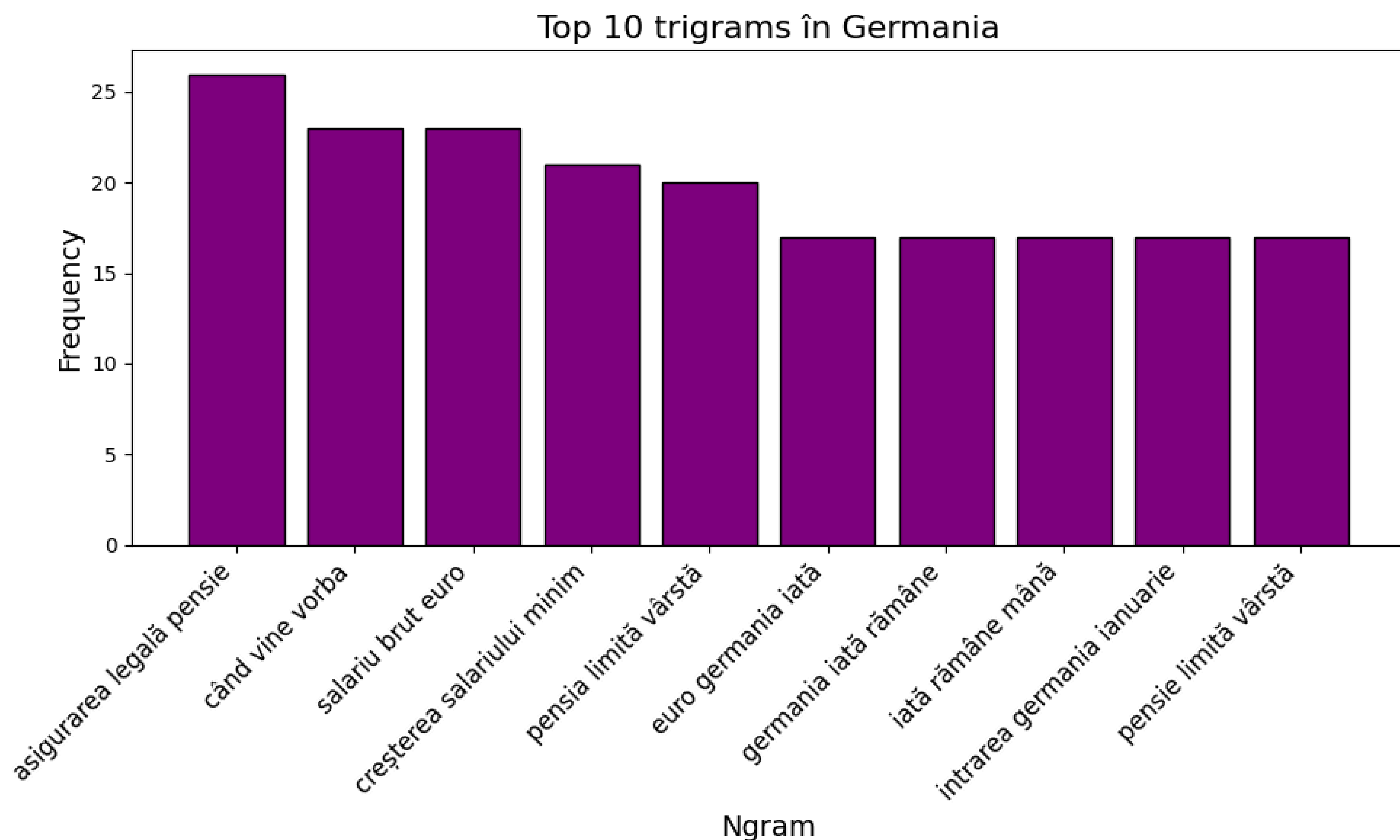


# TRIGRAME

Top 10 trigrams în Italia

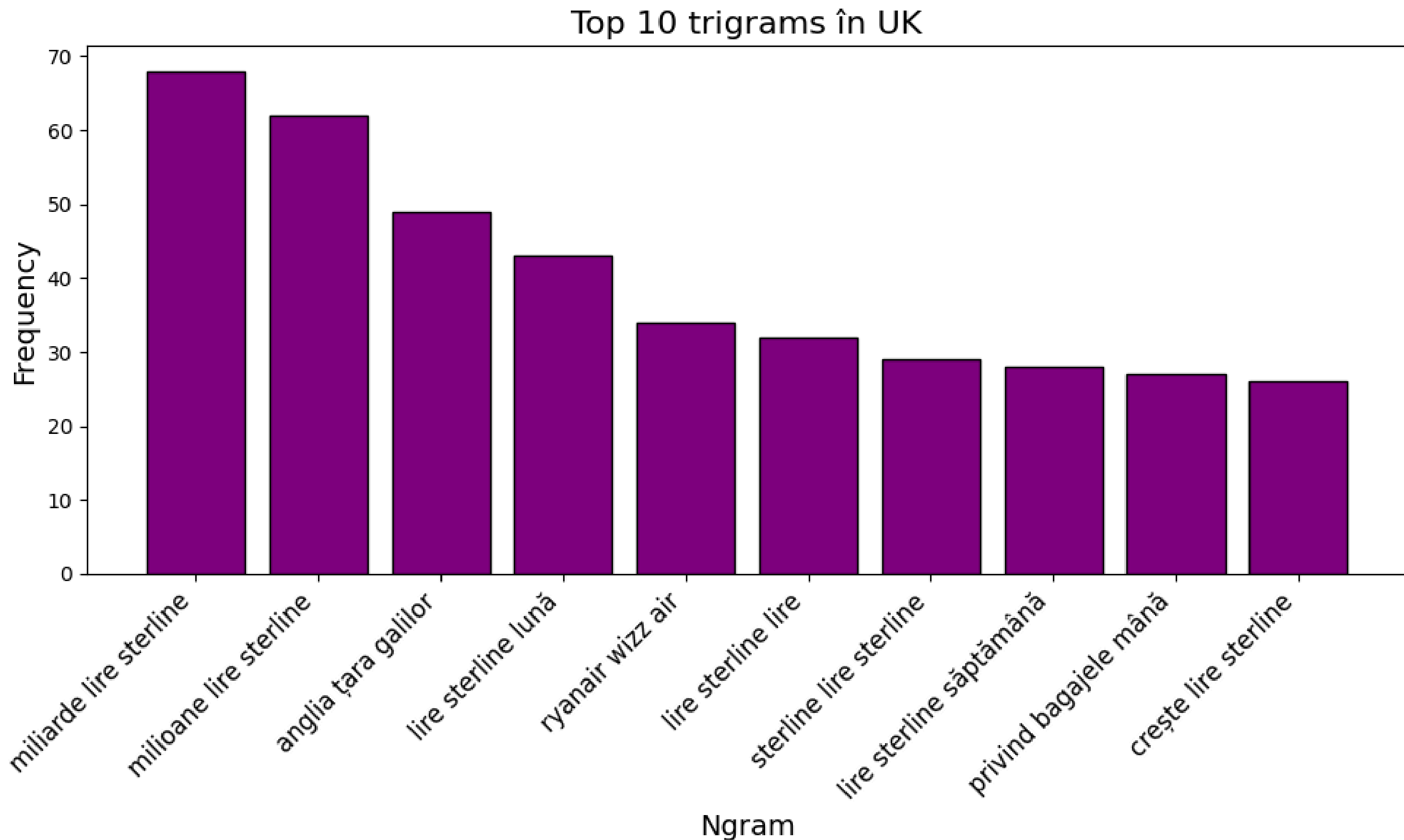


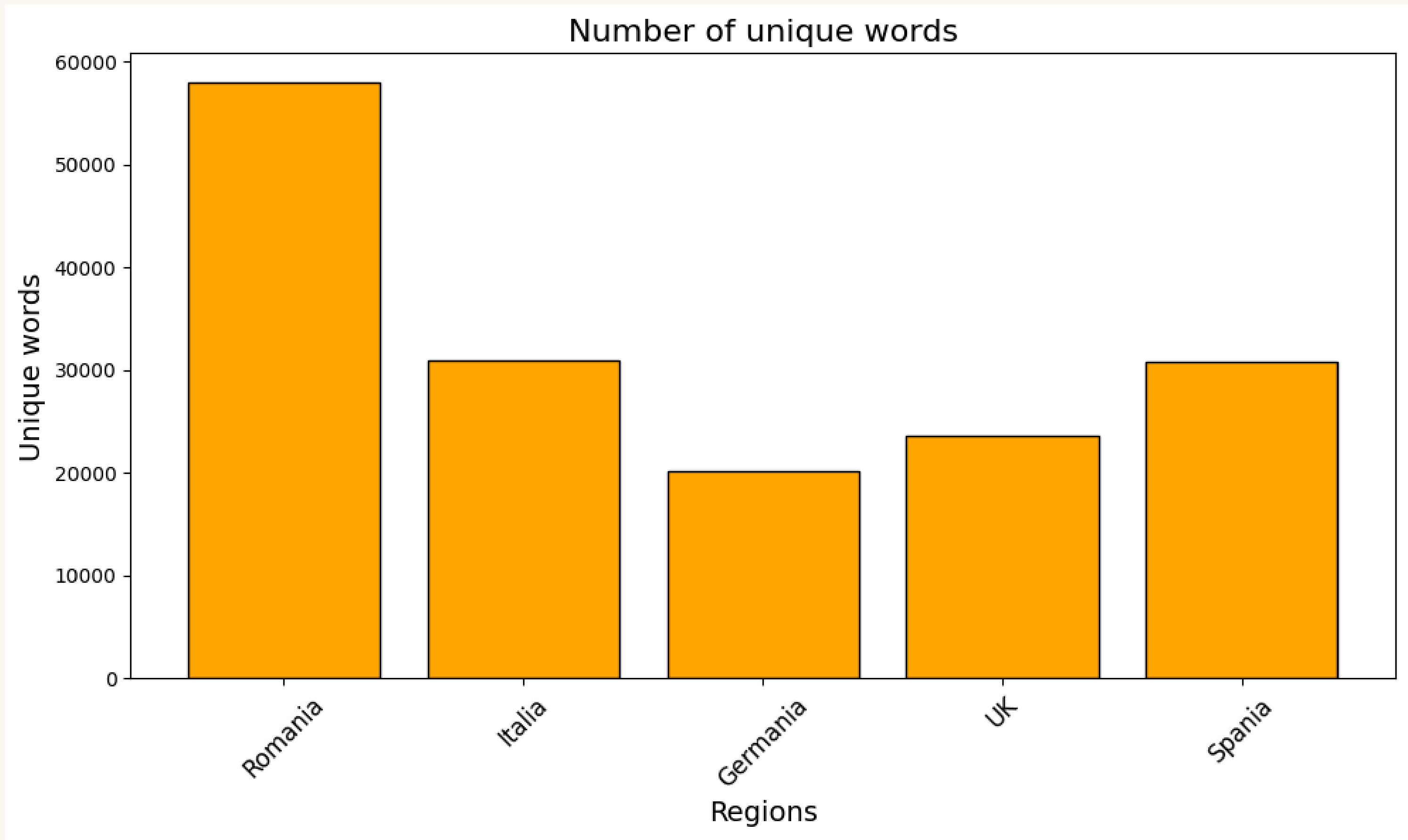
# TRIGRAME



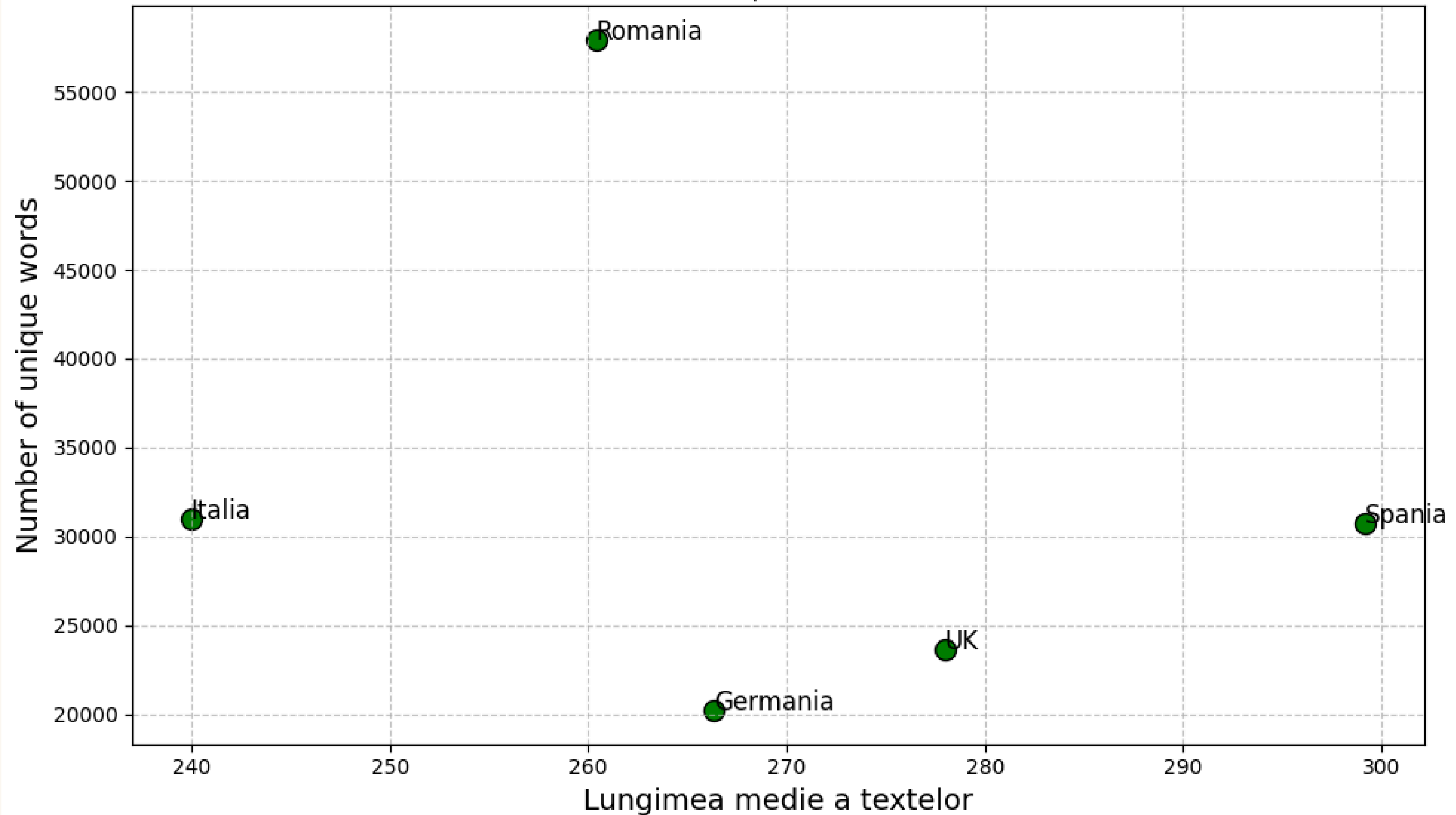


# TRIGRAME





Compararea lungimii medii și a numărului de cuvinte unice



## GERMANIA

- Influenta puternica in geografie, cultura si politica ( Renania, Bundestang)
- Subiectele administrative sunt proeminente (Kindergeld - beneficii pentru copii)

## SPANIA

- Conexiuni culturale si geografice (Canare, Picasso)
- Multe articole cu topic social.

# CONCLUZII

## ITALIA

- Influenta in cultura locala si administrare (Modena, Carabinieri)
- Mai multe topicuri sociale si de munca

## UK

- Focus administrativ (GBP, HMRC)
- Referinte la servicii de sanatate (NHS) si geografie.

- **GERMANIA & UK**

Mai mult focus pentru  
domeniul economiei

- **SPANIA & ITALIA**

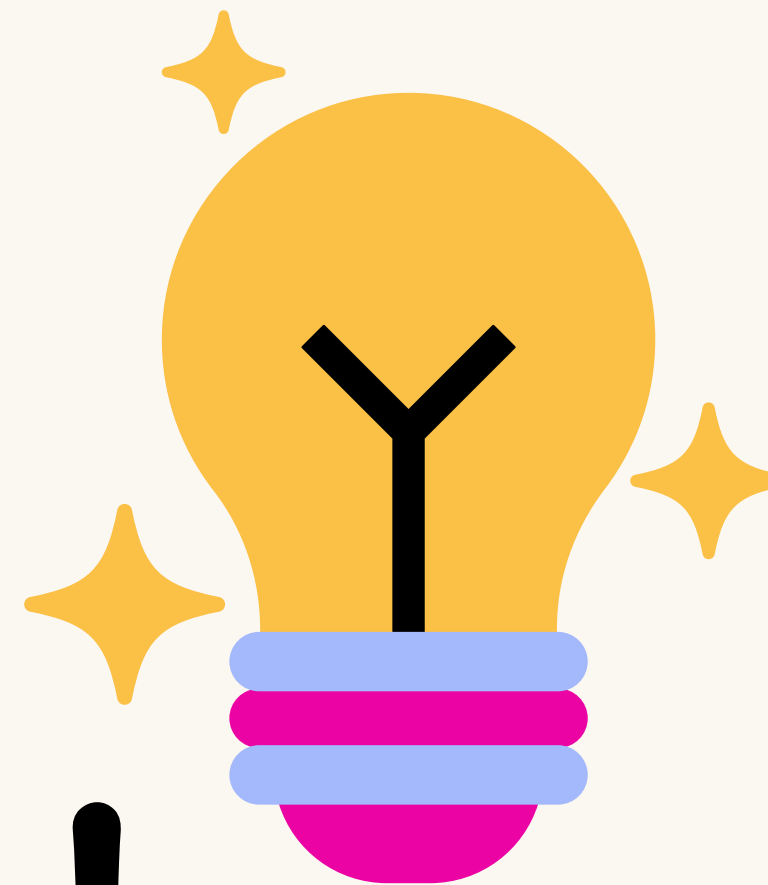
Mai mult focus pe cultura si pe  
integrarea sociala a romanilor.

- **CONCLUZII GENERALE**

SE OBSERVA O OARECARE INFLUENTA LINGVISTICA PENTRU FIECARE  
TARA, DAR IDENTITATEA ROMANEASCA SE PASTREAZA PUTERNIC IN  
FIECARE ARTICOL.

## **LIMITARI**

- **UN NUMAR LIMITAT DE ARTICOLE  
DIN FIECARE TARA**
- **NU AM AVUT SUFICIENTE  
ARTICOLE CU MAI MULTE TOPICE  
DIFERITE**



# MULTUMIM!

