

# Comparative Analysis of Anglicism Distribution in Russian Social Network Texts

Fenogenova A., Karpov I., Kazorin V., Lebedev I.  
NRU Higher School of Economics, RDI KVANT

*Dialogue-2017, RSUH*



# Problem

The phenomenon of anglicisms in Russian language

The growing number of loanwords in Russian:

- problem for NLP tasks like spell-checking, taggers, sentiment detection, etc.
- interest for theoretical researches, studying language contact processes

# Task

We propose the method for detecting English loanwords (Anglicisms) in Russian Social Network texts

- unsupervised
- fully automated
- can be applied to any domain-specific area

The comparative study of anglicisms from different sources

Dictionary growth strategy

# Methodology

Idea: scripting, phonetics and semantics similarity of the original Latin word and its Cyrillic analogue

Hypotheses generation:

- Set of transliteration, phonetic transcribing methods to find possible hypotheses

Hypotheses filtering:

- Distributional semantic models
- Linear mapping

# Material

- VK
- Habr
- Twitter
- LiveJournal

# Results

- Presented an effective dictionary collection method
  - The method does detect Anglicisms!
  - Linear mapping improves the total recall to 0.74%
- Compared Anglicisms from different sources
- Compared Anglicisms from different social, gender and age groups
- Proposed the best corpus formation strategy

Thank you! See  
you around our  
poster!