

# SEMANTIC CLASSIFICATION OF RUSSIAN HETERONOMINATIVE NOUN PHRASES (ON THE MATERIAL OF RUCOR CORPUS)

Azerkovich I.L. ([ilazerkovich@edu.hse.ru](mailto:ilazerkovich@edu.hse.ru)) National Research  
University Higher School of Economics, Moscow, Russia

# Variety of coreference relations

*[Facebook]<sub>i</sub> уже несколько лет делает ставку на видео. [Компания]<sub>i</sub> наращивает объем роликов в лентах пользователей и нанимает людей, которые могут двигать это направление. К февралю 2017 года наконец стало понятно, насколько амбициозны планы [компании]<sub>i</sub>, – [она]<sub>i</sub> хочет заменить собой телевизор.*

## ► Three types of reference to the same NP:

1. Repeating the NP - mostly easily resolved
2. Anaphoric pronoun - most obvious resolution targets, yet may cause difficulties
3. Semantically related words (*Heteronominations*) - *not always considered in analysis*; also cause difficulties

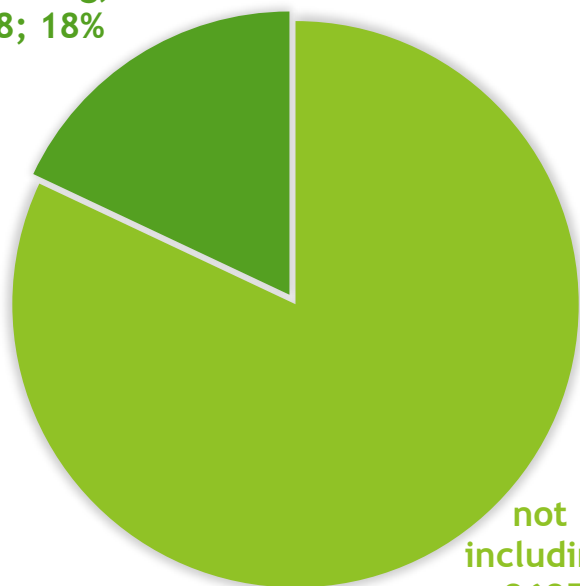
# Research layout

- ▶ **Goals:**
  - ▶ Classify heteronominative groups
  - ▶ Understand, what information is required to resolve
- ▶ **Material: RuCor, the reference corpus of Russian**
  - ▶ 185 texts
  - ▶ Various genres: news, fiction, blog posts
  - ▶ 16557 coreferent groups

# Heteronominations in RuCor

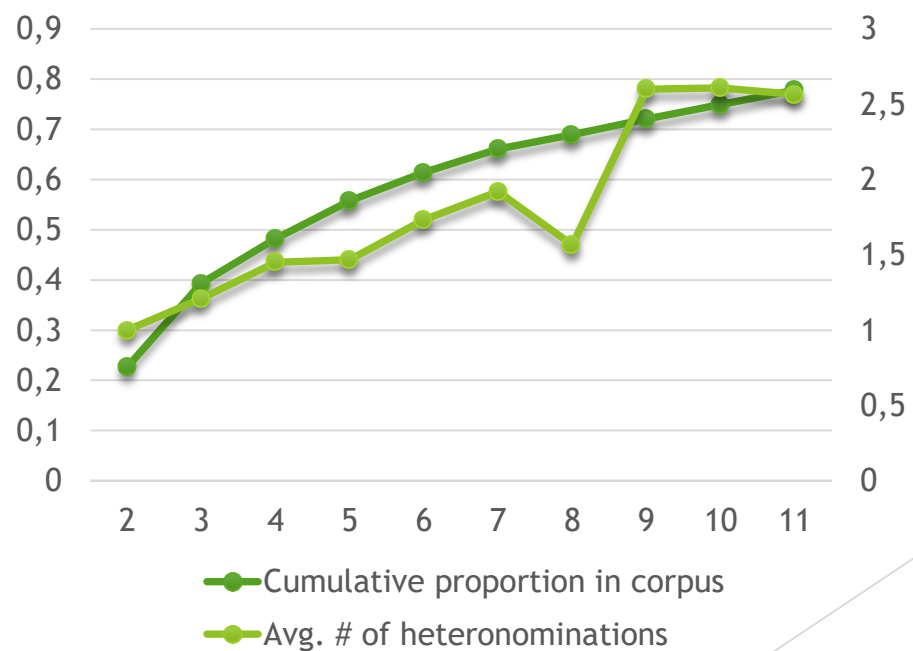
## HETERONOMINATIVE NPS IN CORPUS

including;  
798; 18%



not  
including  
; 3637;  
82%

## Amount of heteronominations by NP length

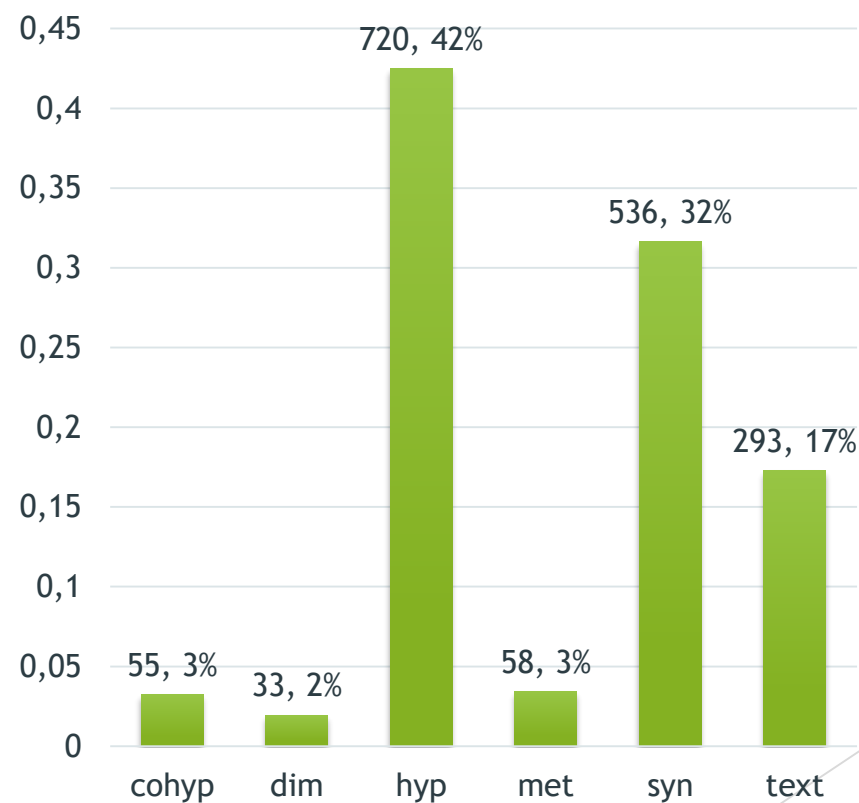


# Classification

Type of information:

- ▶ Encyclopedic
  - ▶ Hypo/hypernymic
  - ▶ Synonymic
  - ▶ Diminutive
  - ▶ Cohyponymic
  - ▶ Metonymic
- ▶ Textual

Distribution in corpus



# Hypo/hypernymy

- ▶ Relation “Type — Example”

*Protivostojaniye [stolichnogo gradonachal'nika]<sub>i</sub> i  
Federal'nogo Orgkomiteta po podgotovke prazdnovanija 65-  
letija Pobedy zakonchilos' absolutnoj pobedoj [Yurija  
Luzhkova]<sub>i</sub>.*

# Synonymy

- ▶ Words, interchangeable in most contexts

*V ramach festivalja, posvjashchennogo [legendarnomu spektaklju “Dobryj chelovek iz Sezuana”]<sub>i</sub>, teatr pokazal [znamenituju postanovku]<sub>i</sub> v dvuh cheshskih teatrah*

- ▶ Diminutives

*[Van’ka]<sub>i</sub> s [Tan’koj]<sub>j</sub>, tochnee skazat’, [Ivan Tihonovich]<sub>i</sub> i [Tat’jana Finogenovna]<sub>j</sub> Zaplatiny*

# Metonymy

- ▶ Relations “Part - Whole” or “Item - Material”
- ▶ *Nezavisimost' [Suhuma]<sub>i</sub> priznala tol'ko RF (i chastichno Nikaragua, ne ustanoviv s [Abhazijej]<sub>i</sub> diplomaticheskikh otnoshenij <....>)*



# Cohyponimy

- ▶ Words that would have a common parent node in an ontology
  - *V takoj-to temnote ne ugljadish', - skazala [pozhilaja kolhoznitsa]<sub>i</sub> - Hot' by skorej svet dali.*
  - *Nichego, [mamasha]<sub>i</sub>, v tesnote da ne v obide.*

# Textual relation

- ▶ “Situational” or relating to the content of the text information

*Vyhodec is [Nigerii]<sub>j</sub> reshil ostat'sja na PMZH v Izraile, poskol'ku na [rodine]<sub>i</sub> ego jakoby presledujet opasnyj prizrak*

# Results

- ▶ A classification of cases of heteronomination in Russian texts was built.
- ▶ Frequency of heteronominations in texts was confirmed
- ▶ Importance of encyclopedic information for coreference resolution of heteronominative groups was observed