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

Azerkovich I.L. (ilazerkovich@edu.hse.ru) National Research University Higher School of Economics, Moscow, Russia
Variety of coreference relations

[Facebook] already several years making a bet on video. [Company] is expanding the volume of videos in feed users and hiring people who can move this direction. By February 2017 it finally became clear how ambitious plans [company], — [she] wants to replace itself by a television.

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
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

- Cumulative proportion in corpus
- Avg. # of heteronominations
Classification

Type of information:
- Encyclopedic
  - Hypo/hypernymic
  - Synonymic
  - Diminutive
  - Cohyponymic
  - Metonymic
- Textual

Distribution in corpus

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohyponymic (cohyp)</td>
<td>55</td>
<td>3%</td>
</tr>
<tr>
<td>Diminutive (dim)</td>
<td>33</td>
<td>2%</td>
</tr>
<tr>
<td>Hypo/hypernymic (hyp)</td>
<td>720</td>
<td>42%</td>
</tr>
<tr>
<td>Metonymic (met)</td>
<td>58</td>
<td>3%</td>
</tr>
<tr>
<td>Synonymic (syn)</td>
<td>536</td>
<td>32%</td>
</tr>
<tr>
<td>Textual (text)</td>
<td>293</td>
<td>17%</td>
</tr>
</tbody>
</table>
Hypo/hypernymy

- Relation “Type — Example”

Protivostojanije [stolichnogo gradonachal’nika], i Federal’nogo Orgkomiteta po podgotovke prazdnovaniija 65-letija Pobedy zakonchilos’ absolutnoj pobedoj [Yuriya Luzhkova].
Synonymy

- Words, interchangeable in most contexts

*V ramkah festivalja, posvjashchennogo [legendarnomu spektaklju “Dobryj chelovek iz Sezuana”]*, teatr pokazal [znamenituyu postanovku] v dvuh cheshskih teatrah

- Diminutives

*Van’ka* s [Tan’koj], tochnee skazat’, [Ivan Tihonovich] i [Tat’jana Finogenovna] Zaplatiny
Metonymy

- Relations “Part - Whole” or “Item - Material”

- Nezavisimost’ [Suhuma] _i_ priznala toл’ko RF (i chastichno Nicaragua, ne ustanoviv s [Abhazijej] _i_ diplomaticheskih otnoshenij <....>
Cohyponimy

- Words that would have a common parent node in an ontology

- V takoj-to temnote ne ugljadish’, - skazala [pozhilaja kolhoznitsa], - Hot’ by skorej svet dali.
- Nichego, [mamasha], v tesnote da ne v obide.
Textual relation

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

Vyhodec is [Nigerii], reshil ostat'sja na PMZH v Izraile, poskol'ku na [rodine], ego jakoby presledujet opasnuyj 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.