

Dealing with polysemy in Russian National Corpus: the case of adjectives

Rakhilina E.V., Reznikova T.I.,
Shemanaeva O.Yu.

Moscow

Russian National Corpus

www.ruscorpora.ru

- POS (parts of speech) and morphological tagging
- accentuation (partly)
- syntactic annotation (partly)
- meta-textual annotation
- semantic annotation

Semantic annotation in Russian National Corpus

- Taxonomic class (*kinžal* 'dagger': «arms», *pečal'* 'sorrow' «emotion», *krivljatsja*: 'to make faces' «behavior»)
- Mereology (*kabluk* 'heel': «parts of clothes and shoes»)
- Topology (*čša* 'cup': «container»)
- Evaluation (*blagouhanie* 'odor': «positive», *presmykat'sja* 'lick the boots': «negative»)

The problem of polysemy

- luk 'onion' = «plant» / «food»
- luk 'bow' = «arms»
- luk = «plant» / «food» / «arms»
- ❑ tihij 'low (about sounds)' «sound»
- ❑ tihij 'quiet' «human quality»
- ❑ tihij 'faint' «degree»
- ❑ tihij 'slow' «speed»
- ❑ tihij = «sound» / «human quality» / «degree» / «speed»

Adjectives in RNC

10,7% tokens in RNC are adjectives, and the half of them is ambiguous.

Semantic filters cover 500 000 occurrences of the most widely spread adjectives.

Disambiguation efficiency is about 70%.

The scope of our research is adjectives of:

- Physical quality

(temperature, color, size, form, time, speed, etc.)

- Human quality

Construction Grammar

[Fillmore et al. 1987], [Goldberg 1995]

- Constructions rather than words
- Constructions lead to the meaning shift
- Meanings inappropriate for the construction are deleted

The rule-based approach to WSD

Information needed for the rules:

- Grammatical form of the adjective ('full'-
'short'; case, number, comparative,
superlative)
- Grammatical form of the neighbour noun
(animate vs. non-animate, case, number)
- Taxonomic class of the neighbour noun
- Stable collocations

Grammatical form of the adjective

- celyj & brev: celyj 'safe, intact' (cf. *cel, cela, celo, cely*) *ovcy cely* 'the sheep are safe'
- celyj + «time»: celyj 'whole, entire' (cf. *celyj den* 'the whole day'),
- ostat'sja + celyj: celyj 'safe, intact' (cf. *ostat'sja celym* 'remain safe')

Grammatical form of the adjective

- *skoryj* 'fast'

skoryj&comparative + ves'&GEN 'whole':

skoree vsego 'most likely' «parenthetical construction»

Grammatical form of the noun

Tolstyj 'fat' / 'thick'

- tolstyj + «animate» 'fat' (cf. *tolstyj chelovek* 'fat man')
- tolstyj + S: tolstyj 'thick' (cf. *tolstaja kniga* 'thick book')

Pokojnyj 'calm' / 'deceased'

- Pokojnyj + «animate»: pokojnyj 'deceased, dead' (cf. *pokojnyj otec* 'deceased father')
- Pokojnyj + «non-animate»: pokojnyj 'comfortable' (cf. *pokojnyj divan* 'comfortable sofa')

Grammatical form of the noun

- *Polnyj* ‘full of’ / ‘fat’ / ‘complete’ / ‘absolute’
polnyj + S&GEN: polnyj ‘full of’ (cf. *glaza*,
polnye slez ‘eyes full of tears’)

Taxonomic class of the noun

- 'motion', 'time', 'sound', 'color', 'place', 'emotions', 'parts of the body', 'hair', 'animals', 'plants', 'texts', 'groups of people', 'relatives', 'professions', 'stuff', etc.

Taxonomic class of the noun

Melkij 'fine' / 'shallow'

- Melkij + «stuff»: melkij 'fine' (cf. *melkij pesok* 'fine sand')
- Melkij + «non-animate&concrete»:
melkij 'shallow' (cf. *melkaja reka* 'shallow river')

Tjzhelyj 'heavy' / 'serious'

- tjzhelyj + «non-animate&concrete»: tjzhelyj 'heavy' (a physical quality) (cf. *tjzhelyj čemodan* 'heavy bag')
- tjzhelyj + «diseases»: tjzhelyj 'painful, serious' (a degree) (cf. *tjzheloe otravlenie* 'serious poisoning')

Stable collocations

- *černyj* 'black' «color»

na černyj den' 'for a rainy day' «negative»

- *Gluhoj* 'deaf, indifferent'

Gluhoj + *k* (*k* is a preposition meaning 'to'): *gluhoj* 'indifferent'

Gluhoj + «people»: *gluhoj* 'deaf'

- *Krugly* 'round'

Krugly + *stol* is either a 'round table' or a 'discussion'.

V (*v* is a preposition meaning 'in') + *kruglyj* + *stol* (*v kruglom stole*) means only 'in discussion', and not 'in the round table'.

Syntactic patterns for filters

- A + S (agreement)
slaby veter
- A + S (government without preposition)
slaby glazami
- A + PR + S (government with preposition)
slaby na golovu
- PR + A + S (prepositional phrase)
v slaboy stepeni

Elaborating the system of taxonomic classes

- staršij 'elder' / 'senior';
- mladšij 'younger' / 'junior';

staršij + «professions»: staršij 'senior';

mladšij + «professions»: mladšij 'junior';

(cf. staršij/ mladšij naučnyj sotrudnik 'senior/junior researcher')

staršij + «human»: staršij 'elder';

mladšij + «human»: mladšij 'younger';

(cf. staršij/ mladšij brat 'elder/younger brother')

Possible types of the meaning shifts

Physical quality + «human» – human quality

- *mjagkij divan* ‘soft sofa’ – *mjagkij čelovek* ‘tender person’,
- *legkij čemodan* ‘light bag’ – *legkij čelovek* ‘easy person to get along with’

Metaphor and metonymy

«Human» – «parts of the body»

- *dobryj čelovek* ‘kind person’ – *dobrye glaza* ‘kind eyes’

Adverbs are bound metonymically with adjectives

- *redkiy* ‘thin-growing, sparse’:

redkaya boroda ‘thin beard’ –

derev’ya redko rastut ‘the trees grow sparsely’

Conclusion

Theoretical extensions of disambiguation:

- Elaborating the system of taxonomic classes
- Systematization of metaphorical and metonymical shifts
- Evolution of the meanings traced by corpus