G. Kustova, O. Lashevskaja, E. Rakhilina, E. Paducheva
(Moscow)

## ON TAXONOMY IN COGNITIVE SEMANTICS AND CORPUS LINGUISTICS: PARTS OF BODY

**2.** The theme of this paper is Russian National Corpus, which has been developed in Moscow by Russian Academy of Science. The corpus manager is freely accessible online at http://www.ruscorpora.ru.

What is unique for the RNC is a semantic annotation layer, which make it possible to search in the corpus not only for individual lexemes and grammatical features but also for certain traditional lexico-semantic groups. By adding and combining lexical, morphological and semantic criteria in the query window, we can check our linguistic hypotheses about lexical co-occurence and construction grammar. For this purpose the RNC is annotated according to a branching network of taxonomic classes which include

***main ontological classes*** *of nouns, verbs, adjectives and adverbs*, such as 'person', 'space', 'text', 'motion', 'location', 'emotion', 'speed', among others;

***mereological classes*** *of nouns***:** 'parts', 'sets' etc., which reveal partonomic and set-element hierarchy between objects;

***topological (geometrical) classes*** *of nouns***:** 'containers', 'horizontal surfaces' etc.;

***evaluation:*** 'positive' and 'negative'.

**3.** In this picture you can see a list of semantic classes of concrete nouns (nouns which refer to physical objects), and this is a **body parts** class that will be in the focus of my talk.

**4.** There are about 350 items in this class. They can be split further into narrower groups, such as

(**5**) active parts (*golova* 'head', *ruki* 'hands & arms', *nogi* 'foots & legs', *kryl'ja* 'wings', kopyta 'hooves', etc.);

(**6**) inner parts (*serdce* 'heart', *počki* 'kidneys', *legkije* 'lungs', *myšcy* 'muscles', etc.);

(**7**) organs of perception and feelings (*glaza* 'eyes', *nos* 'nose', *ushi* 'ears', *serdce* 'heart', etc.);

(**8**) parts of animal body (*kryl'ja* 'wings', *lapy* 'paws', *kopyta* 'hooves', *žabry* 'gills', etc.);

(**9**) paired parts (*ruki* 'hands & arms', *nogi* 'foots & legs', *glaza* 'eyes', *kryl'ja* 'wings').

**10.** From the point of view of RNC taxonomy, "parts of body" is a mereological class, along with 'parts of artefacts', 'sets', and 'quanta of substance'. Some names of BP are tagged with other semantic features. For example, names for 'back' and 'side' are tagged as 'space' (main ontology); nouns *šejka* 'little neck' and *glazki* 'little eyes'are tagged as 'diminutive' (derivational class); and Russian *roža* 'face (mug)' bears the 'negative' feature.

**11.** BP class is one of the most prominent classes in the cognitive linguistics. There is a lot of verbs that imply body part in their lexical semantics, e. g. verbs of motion (*vz'at'* 'take'*, dat'* 'give'*, brosit'* 'throw' – apparently 'by hand'); verbs of physical impact (*pnut'* 'kick' – 'by foot'), physiological verbs (*zhevat'* 'chew', *kusat'* 'bite' – 'with teeth/jaws'). Gestures can be done by various kinds of BP: *kivat'* 'nod' – 'by head', *klan'at's'a* 'bow' – 'by head or body'*, maxat'* 'wave; wag' – 'by hand', and so on.

**12.** Our preliminary investigation of 400 most frequent Russian verbs demonstrates that almost all of them have BP as an obligatory semantic argument which can be expressed as a syntactical argument of the verb at least in some peculiar syntactical conditions. The portion

of these verbs is displayed by red line. As you can see only verbs of possession and verbs of existence are not cognitively connected immediately with body parts.

**13.** Among Russian verbs there are a lot of those with incorporated argument (Jackendoff 1990, Paducheva 2004). For example, *celovat'* 'to kiss' means "to cause *lips* to come into contact with somebody/something", *videt'* 'see' implies the help of *eyes*. The overt expression of the BP participant may be excessive. It is interesting that verbs with incorporated argument include verbs derived from BP names: *glaz'et'* 'stare', *vručit'* 'hand over', *oblokotit's'a* 'lean one's elbows', etc.

**14.** Many nouns naming parts of the body are highly polysemous in Russian, and this poses a problem when running a program of semantic annotation. The noun *lico* 'face', for example, is highly ambiguous, with up to ten meanings clustered round different lexical domains: parts of the body ('face'); humans ('person'); properties ('identity': cf. *to change the face of the office*), abstract terms ('grammatical person'), etc (**15-18**).

**19.** The semi-automated method of filtering allows for disambiguation of polysemous nouns. The filtering approach is based on the principles of the Construction Grammar (cf. Fried, Östman 2004); it demonstrates that basic usages of top frequent words are strongly entrenched (Tomasello 2006) by a number of reasonable constructions and collocations.

**20.** We have 4 levels of filtering:
- filters on idiomatic expressions;
- 'shallow' filters (based on lexical items);
- general filters (based on morphological and semantical features of the context);
- filters 'by default'.

**21.** So, for the Russian name for face, first we have to sift out <u>idiomatic expressions</u>. It can be

1. Compound prepositions:

    *pered licom + GEN* 'in the face of', lit. 'in front of the face'

    *ot lica + GEN* 'on behalf of', lit. 'from the face of'

2. Adverbial phrases

    v pote lica '(to work) by the sweat of one's face'

    na odno lico '(to be) the same'; lit. 'on the same face'

3. Other idioms

    *steret' s lica zemli* 'to wipe smb. off the face of the earth'

**22.** The 2nd stage is the stage of <u>'shallow' filters</u>. Most frequent collocations, i.e. lexical items and word forms in the context, are taken into account, if they can not be generalized as a lexical group (semantc class).

    *fizičeskij + lico* 'natural person' $\Rightarrow$ PERSON (1000)

    *častnyj + lico* 'private person' $\Rightarrow$ PERSON (600)

    *tretij* &PL + *lico* &PL $\Rightarrow$ PERSON (400)

        *v interesax tret'ix lic* 'in the interests of a third party'

    *čerta* &PL + *lico* 'features of the face' $\Rightarrow$ BODY_PART (800)

**23.** On the 3rd stage <u>general filters</u> should be applied. They include

1. Semantic and POS conditions:

    *lico* + CONJ + BODY_PART $\Rightarrow$ BODY_PART

        *lico i ruki* 'face and hands', not 'person and handwritng'

    A&derived_from(PERSON) + *lico* $\Rightarrow$ BODY_PART

        *Marusino lico* 'Marusia's face'

    PERSON + *s licom* + GEN $\Rightarrow$ BODY_PART

*čelovek s licom d'javola* 'a man with the devil's face'

**24.** 2. Semantic conditions which unmask the fact that the face is a means of emotional expression (Iordanskaya&Paperno 1995):

A&EMOTION + *lico* ⇒ BODY_PART

*sčastlivoje lico* 'happy face'

A&COLOR + *lico* ⇒ BODY_PART

*blednoje lico* 'pale face'

PR + *lico* +S&EMOTION ⇒ BODY_PART

*na lice razočarovanije* 'disappointment on the face'

**25.** 3. Semantic conditions which expose the face as a physical object:

A&FORM + lico ⇒ BODY_PART

*krugloje lico* 'round face'; *ploskoje lico* 'plain face'

A&SIZE + lico ⇒ BODY_PART

uzkoje lico 'narrow face'.

**26.** 4. The following semantic conditions are connected with the idea that 'person' is a social being:

S&SPEECH + *lico* &GEN ⇒ PERSON

*xodatajstva lic* 'persons' petitons'

A&HIERARCHY + *lico* ⇒ PERSON

*glavnoje lico* 'the prime person'

**27.** After the 3$^{rd}$ stage of filtering the disambiguation approaches 80%. The rest of the contexts are distributed equally between BODY_PART and PERSON meanings.

For the other members of the BODY_PART class, the tag BODY_PART should be ascribed by default, if the corresponding meaning is listed first in the dictionary.

**28.** It seems, that construction of semantic filters is a purely technical task, which has nothing to do with linguistic theory. But we treat it in other way. We argue, that the mechanisms of word-sense disambiguation in the corpus should be based on the same cognitive principles as the mechanisms for understanding of polysemous items in discourse. And these mechanisms are those of Ch. Fillmore's Construction Grammar.

A new construction coerces type-shifting and therefore a new meaning. And what is a new construction? It is either a new syntactic structure or the same syntactic structure with another taxonomic feature of a certain variable. The case of syntactic construction is very clear. Let's consider the case when taxonomic feature of the construction is changed.

As we have seen, two interpretation of attributive construction are possible: one is physical, as in *narrow face*, other interpretation is social, as in *the prime person.* One can see that it is quite simple that if an adjective is from the physical domain, then the meaning is also physical, and if an adjective comes from social domain then the social meaning of the noun is chosen. So the rule displays a sort of semantic coordination within the construction. This is an example of a cognitive mechanism exploited in semantic filters.

As Lora Janda argued yesterday, applicative linguistics gives an impulse for a new stage of linguistic research. This is another example of this case. Our filters are based on Fillmore's idea of constructions. It is one of the important applications of this theory, but filters themselves push further theoretical research in lexical semantics and polysemy.

## References

Fried, Mirjam & Jan-Ola Östman (eds.). 2004. *Construction grammar in a cross-language Perspective*. Amsterdam etc.: John Benjamins.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge (MA): MIT Press.

Kemmer Susan, 1993. *The Middle Voice*. Amsterdam etc.: John Benjamins.

Podlesskaya, Vera I. & Ekaterina V. Rakhilina. 1999. "External Possession, Reflexivization and Body Parts in Russian". In Payne, Doris L. & Immanuel Barshi (eds.), *External Possession*. Amsterdam etc.: John Benjamins.

Taylor, John R. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory*. 2nd expanded edition. Oxford: Oxford University Press.

Tomasello, Michael. 2006. Acquiring linguistic constructions. In D. Kuhn & R. Siegler (eds.), *Handbook of Child Psychology*. New York: Wiley.