FrameBank: A Database of Russian Lexical Constructions

Olga Lyashevskaya[1,2] and Egor Kashkin[2(✉)]

[1] National Research University Higher School of Economics, Moscow, Russia
olesar@yandex.ru
[2] V.V. Vinogradov Russian Language Institute of RAS, Moscow, Russia
egorkashkin@rambler.ru

Abstract. Russian FrameBank is a bank of annotated samples from the Russian National Corpus which documents the use of lexical constructions (e.g. argument constructions of verbs and nouns). FrameBank belongs to FrameNet-oriented resources, but unlike Berkeley FrameNet it focuses more on the morphosyntactic and semantic features of individual lexemes rather than the generalized frames, following the theoretical approaches of Construction Grammar (C. Fillmore, A. Goldberg, etc.) and of Moscow Semantic School (J.D. Apresjan, E.V. Paducheva, etc.).

Keywords: Russian · Construction Grammar · Frames · Corpus linguistics · Morphosyntax · Semantic roles · Polysemy

1 Background

FrameBank[1] is an open access database which consists of a dictionary of Russian lexical constructions and a corpus of their uses tagged with a FrameNet-like annotation scheme [1–3]. The examples are randomly taken from the Russian National Corpus [4]. At present the dictionary provides data for ca. 4000 target verbs, adjectives, and nouns, and the corpus part includes ca. 50000 annotated examples.

The project under discussion started in 2011. The ideology of FrameBank has obviously been inspired by Berkeley FrameNet [5], but there are some crucial differences in how these two resources are organized. Firstly, FrameBank is more focused on morphosyntactic patterns than FrameNet. This is determined by the grammatical properties of Russian (which are not relevant in English), where different case structures often help to profile the situation differently. Secondly, the target entries in FrameNet are extralinguistic situations – frames, which are further linked to a list of semantically related verbs (e.g., the frame of Motion embraces such lexical units as *to come, to go, to fly, to float, to glide, to blow*, etc.). On the contrary, FrameBank has

The work was partly supported by the Russian Foundation for the Humanities, grant No.13-04-12020, and by the Russian Basic Research Foundation, grant No. 15-07-09306.
[1] www.framebank.ru.

M.Y. Khachay et al. (Eds): AIST 2015, CCIS 542, pp. 1–11, 2015.
DOI: 10.1007/978-3-319-26123-2_34

particular lexical items as target entries, providing data on their morphosyntactic patterns and on the frames corresponding to different meanings of a lexeme.

The theoretical basis of FrameBank includes Construction Grammar (C. Fillmore, A. Goldberg, etc.) as well as some approaches developed in the Moscow Semantic School (J.D. Apresjan, E.V. Paducheva et al.) with its attention to the differences between close synonyms and to the interaction between lexical and grammatical features of lexical items. There is another resource developed within the Moscow Semantic School – namely, the Lexicographer database [6]. However, it does not seem to equally embrace all the main semantic classes of Russian verbs and all the possible constructions of the verbs it includes. Neither is it directly linked to a set of corpus examples, which is one of the main features of FrameBank.

The paper is structured as follows. After outlining how the dictionary of constructions is designed, we discuss the annotation scheme and some theoretical issues it raises. Further, we consider two databases included in FrameBank: the graph of semantic roles and the graph of formal and semantic shifts between constructions. The graph of semantic roles presents our own inventory, which correlates with the semantic classification of verbs and forms a hierarchy in order to support flexible search options. The other graph shows both formal changes of verbal constructions (omission of a participant, change of a morphosyntactic pattern, diathetic alternations etc.) and their semantic changes (metaphor, metonymy, and also some shifts which have not been discussed so widely, like specialization or rebranding). FrameBank also provides quantitative data on the frequency of semantic roles and semantic shifts, which could be used in the automatic annotation of texts (e.g. for the tasks of semantic role labelling). Finally, we outline some future steps in developing FrameBank.

2 Dictionary of Construction Patterns

We will discuss the architecture of FrameBank using the example of verbs, which form the core of the database. Information about each lexical construction is stored as a construction template, which includes:

1. the syntactic rank of the element (Subject, Object, Predicate, Peripheral, Clause);
2. the morphosyntactic features of the element[2] (including POS, case and preposition marking);
3. its status: lexical constant vs. variable;
4. the semantic roles of the argument (e.g., Agent, Patient, Instrument);
5. the lexical-semantic class of the element (e.g., human, animate, abstract entity, means of transport, etc.);
6. the morphosyntactic features of the target lexical unit itself (e.g. impersonal, passive participle, etc.);
7. one or several examples.

Figure 1 shows a sample pattern in the dictionary.

[2] This part was originally based on [7].

ID230. Cx name: *Pjatno vystupilo na rubaške* ['a stain appeared on the short']. Cx Pattern: Snom V na + Sloc.

Cx Item ID	Pl	Letter	Head	Phrase	Explication	Syntactic Rank	Lexico-semantic constraints	Status [obligatory / optional]
2077	1	**X**	**Snom** [Nominative case]	**NPnom**	**Theme**	**Subject**	**natural object**	**Oblig.**
2078	2	–	***vystupit'*** ['to appear; lit. to step forward']	–	**to appear**	**Predicate**	–	**Oblig.**
2079	3	**Y**	***na* + Sloc** [preposition *na* 'on' + Locative case]	***na* + NPloc**	**Location**	**Peripheral**	**space and place**	**Oblig.**

Lexical Index of target words

Index of Morphosyntactic Items

Fig. 1. The template of the construction *Pjatno*[Noun.Nom] *vystupilo*[Verb] *na rubaške* [PREP + Noun.Loc] 'a stain appeared on the short'.

Each verb is followed in the database by a list of lexical constructions in which it serves as a target word (each construction is named by a mnemonic sentence label). Lexical constructions are grouped in clusters usually corresponding to a particular lexical meaning; the constructions belonging to one cluster differ in the number of explicit arguments and in their morphosyntactic marking. Figure 2 shows two groups of LexCxs of the verb vystupit' 'to step forward' which correspond to the frame of motion and the frame of coming into existence, respectively.

Target Lexeme: *vystupit'*
1. 'to step forward'
ID220. <Snom V> *Vystupilo srazu pjat' soldat* 'Five soldiers stepped forward at once'
ID221. <Snom V PR_from+S>. *Iz stroja vystupil čelovek* 'A man stepped forward from the line'
ID222. <Snom V PR_to+S> *On vystupil na seredinu komnaty* 'He stepped forward to the center of the room'
...
5. 'to appear (about blood, tears, stains, etc.)'
ID 230. <Snom V *na*.PR+Sloc> *Pjatno vystupilo na rubaške* 'A stain appeared on the short'
ID 231. <Snom V *na*.PR+Sloc *u*.PR+Sgen> *Sljozy vystupili u nee na glazax* lit. 'Tears appeared on the eyes at her'
ID 232. <Snom V *u*.PR+Sgen *ot*.PR+Sgen> *U nee ot smexa vystupili sljozy* lit. 'Tears appeared at her from laughing'

Fig. 2. The passport of the lexeme *vystupit'*

3 Corpus Annotation

The dictionary of constructions is supplemented by examples tagged manually. The examples are randomly selected from the Russian National Corpus, each target lexical unit is illustrated by up to 100 sentences with their pre- and post-context. Each example is annotated by one of the annotators in the online FrameBank Markup environment, and then is checked and corrected by the editor. An example is matched to a suitable construction pattern, which includes establishing correspondences between their elements and assigning morphosyntactic and semantic features of the arguments in a particular example. If an example does not fit any of the existing patterns, an annotator should add a new item into the dictionary of constructions (this is often the case for colloquial constructions, for the on-going changes in the semantics of verbs, and for idiomatic expressions). Note that the participants of a frame are annotated irrespective of their syntactic relation to the predicate (this distinguishes FrameBank from the treebanks like SynTagRus or Prague Dependency Treebank). For example, if we annotate the verb *vyslušat'* 'listen to somebody' and come across sentence (1), we will mark the NPs 'Andropov' and 'the marshal' as the participants of the frame referred to by the verb *vyslušal* 'listened' (the fact that they are not syntactically related to the predicate will also be mentioned in the annotation).

(1) *Andropov prin'al maršala v svojem rabočem kabin'et'e,* ***vyslušal*** *i ob'eščal razobrat's'a v etoj probl'em'e* 'Andropov received the marshal in his office, **listened** to him and promised to examine the problem'

The annotators of FrameBank also mark non-standard types of constructions or non-standard variants of argument realization, such as passive, imperative, participial or converbal constructions, constructions with infinitives, control, genitive of negation. The annotation takes into account not only construction arguments and the properties of the predicate, but also adjuncts and modal particles. More details on the annotation procedure can be found in the full version of the manual for annotators, which is available online[3].

4 Semantic Roles

As has already been mentioned, construction patterns in FrameBank contain information on the semantic roles of the participants. The inventory of semantic roles may have quite different volume and structure depending on the particular research task and theoretical framework (see, for example, [8: 587–588, 9, 10, 11: 125–126, 12: 370–377]). The most important principles governing the inventory of semantic roles in FrameBank are as follows:

- the inventory should be hierarchical in order to support flexible search options (it may be reduced to 5–10 basic roles, or enlarged to several dozen labels);

[3] http://framebank.wikispaces.com/.

- the roles should correlate with the semantic classification of verbs (what follows from this is that traditionally "broad" roles such as Agent or Patient should get different labels in different semantic classes, cf. Agent in destruction vs. speech vs. motion);
- the scope of a semantic role is defined in accordance with the Prototype Theory: for instance, the prototype of Patient is a participant changing under the physical influence of an Agent; peripheral examples (Patient of a non-physical process, Patient which is not changing, Patient created as a result of a physical action) get specific labels (Theme, Result, etc.) and are considered as specific types of Patient.

The detailed list of semantic roles currently contains 91 items classified into seven domains (those of Agent, Possessives, Patient, Addressee, Experiencer, Instrument, Settings), which are further subdivided into smaller units. Initially, we intended to use a list of semantic roles suggested in [12: 370–377]. However, we had to work out some of its parts in further detail in order to be in line with our theoretical principles. For instance, the inventory suggested by J.D. Apresjan includes the role of Experiencer without any further semantic specification. To achieve our goals, we considered Experiencer not as a single semantic role, but as a domain including Subject of Perception ('see', 'hear'), Subject of Mental State ('think', 'understand'), Subject of Psychological State ('love', 'be afraid'), Subject of Physiological State ('feel pain', 'have a buzzing in one's ears'), Subject of Physiological Response ('tremble with cold', 'feel sick'), and Subject of Psychological Response ('laugh', 'cry (burst into tears)'). Similarly, the role of Agent is defined in our inventory as an active (prototypically animate) participant of a situation, intentionally changing something in the world. This role is typically assigned to verbs of physical impact, eating and drinking, creation, causation of motion, while more specific verbs which are less closer to the prototype of Agent receive their own semantic roles (Speaker, Subject of motion, Subject of social relationship, etc.).

It should also be noted that the principles of FrameBank annotation allow marking double roles (following the ideas of [11: 140]). Thus, examples like ***kormit'*** *r'eb'enka s ložečki* '**to feed** a child with a spoon' or ***myt's'a*** *pod kranom* '**to wash oneself** under a tap' contain instrumental participants, which at the same time have locative properties (which influences their morphosyntactic marking). Therefore, these participants receive a double role Instrument and Place in our annotation scheme.

FrameBank also provides frequency data about semantic roles in lexical constructions. Table 1 shows the top-15 roles (the calculation is based on the number of construction patterns with this role; the data on the other roles are left out of this paper due to size limits). These data supplemented with the morphosyntactic patterns may be useful for the tasks of semantic role labelling [13, 14], see [15] for a case study based on FrameBank.

5 Non-core Elements

Along with marking the arguments of target lexical units, the annotation of examples in FrameBank covers their adjuncts (non-obligatory valencies), see, for instance, [8: 72–79] on the theoretical foundations of the distinction between arguments and adjuncts.

Table 1. Frequency of semantic roles in FrameBank (top-15).

Semantic role	Number of construction patterns	Example	Number of predicates in the dictionary
Agent	4787	*Prodav'ec r'ežet syr* 'The seller is cutting cheese'	1824
Patient	3086	*Prodav'ec r'ežet syr* 'The seller is cutting cheese'	1498
Theme	1591	*Na polu l'ežal č'elovek* 'There was a man lying on the floor'	1004
Subject of motion	1520	*My jed'em v Moskvu* 'We are going to Moscow'	515
Speaker	1304	*On govorit pravdu* 'He is telling the truth'	749
Patient of motion	1049	*Mal'čik v'el sl'epogo za ruku* 'The boy led a blind man by the hand'	358
Point of destination	921	*My jed'em v Moskvu* 'We are going to Moscow'	657
Place	903	*Na polu l'ežal č'elovek* 'There was a man lying on the floor'	738
Message	776	*On skazal, čto rabotajet nad knigoj* 'He said that he was working on a book'	454
Effector	643	*V'et'er povalil d'er'evo* 'The wind threw down a tree'	565
Subject of psychological state	643	*On toskujet po druz'jam* 'He misses his friends'	526
Mental content	637	*My sčitali jego opasnym č'elov'ekom* 'We considered him a dangerous person'	438
Content of action	634	*Potrudit'es' vstat', požalujsta!* 'Be so kind to stand up, please!'	526
Result	633	*Mama svarila sup* 'Mother has cooked soup'	445
Reason	616	*Komandira b'espokoilo, jesli razv'edčiki dolgo n'e vozvraščalis'* 'The commander was worried if the scouts didn't return long'	501

This provides large amounts of empirical evidence for discussing the restrictions imposed on the combinability of adjuncts with different types of predicates (cf. a traditional view touched upon in [8: 75] and stating that arguments are specific for each verb, while adjuncts are compatible with various verbs). Table 2 contains statistical data on co-occurrence of verbs and adjuncts depending on the semantic classes of both.

Table 2. Co-occurrence of verbs and adjuncts.

verb class	time	place	degree	manner	usualness	reason	duration	simultaneity	sequence	purpose	precision	frequency	comparison	speed
motion	**426**	**172**	**133**	219	148	93	107	62	58	64	26	38	**74**	**82**
speech	120	**131**	**44**	**139**	68	56	31	18	24	18	22	38	11	9
physical impact	70	89	44	80	36	28	33	20	21	19	4	24	12	7
emotion	69	44	**224**	21	41	39	21	11	11	6	29	3	8	
mental	79	79	40	36	31	28	32	18	17	12	21	5	6	
social interaction	71	64	58	41	43	35	14	17	9	11	13	10	6	4
start of existence	90	**104**	5	24	27	29	8	28	17	18	4	13	5	2
possessive	62	**69**	10	20	12	14	7	14	2	25	4	5	2	3
psychical	33	22	**71**	8	16	22	14	16	6	5	9	2	3	6
sound	17	42	2	16	9	2	2	5	6	2	7	2	2	
physiology	23	21	18	12	10	6	6	4	5		2	2		2
change of state	17	8	12	15	15	10	8	6	5	4		6	4	4
end of existence	19	6	10	2	10	8	7	8	10		4	5	1	8
SUM	1096	851	671	633	466	370	290	227	191	184	145	153	134	127

As can be seen in Table 2, the ratio of co-occurrence is much higher than average for verbs of emotion and the psychical sphere with adjuncts of degree, for verbs of motion with adjuncts of time, speed and comparison, for verbs of speech with adjuncts of manner and place, for verbs expressing start of existence or possessive relations with adjuncts of place (the overrepresented combinations are marked in bold). On the contrary, the ratio of co-occurrence is lower than average for verbs of motion and adjuncts of place, degree, reason and precision, for verbs of speech and adjuncts of degree and time, for verbs of physical impact and adjuncts of time, etc. (the underrepresented combinations are marked with a gray background; the least represented cases are on a dark-gray background). Interestingly, adjuncts referring to usualness, frequency, simultaneity and sequence do not tend to favor any particular verb class. Nevertheless, the data of FrameBank show that the combinability of adjuncts is not arbitrary: the choice of an adjunct with a particular semantics is to some extent predetermined by the semantic class of a verb.

6 Construction Grapher

Another component of FrameBank is the graph of lexical constructions. It documents the systematic relations between constructions. First, it systematizes semantic shifts in verbal lexemes (metaphor, metonymy and some more complex relations). Second, the graph represents formal changes in argument structure, such as omission of a participant, diathetic alternations (cf. [8]), the inheritance of a pattern from another verb etc. The semantic part of the project is inspired by FrameNet grapher as well as by E. Rakhilina's research database on Russian polysemous adjectives and adverbs (see [16]

and references therein). The formal part is guided by E. Paducheva and G. Kustova's theoretical and empirical analysis of polysemy in Russian verbs [6, 8, 17].

The types of formal and semantic changes are represented below (for the previous stage of its discussion see [3]). The figures in brackets after the name of a shift indicate the number of its occurrences in the database. Sometimes a construction undergoes more than one formal or semantic change, in such cases all changes are counted. For each verb in the database we construct a graph showing the formal and semantic changes undergone by its constructions. These graphs are tied into a larger graph of lexical constructions, since some edges of the latter establish linkages between different verbs, consider "Inheritance of a pattern" below. A case study of how the construction grapher works can be found in [18].

6.1 Formal Changes

1. Morphosyntactic alternation (1796): *Vy **govorit'e** pravdu* 'You **are telling** the truth' ↔ *Papa **govorit**, čto bojat's'a n'ečego* 'Father **says** that there is nothing to be afraid of' ↔ *"Moemu drugu groz'at n'eprijatnosti", – **govoril** on* '"My friend is facing troubles", – he **said**'. This formal change is bidirectional (as well as all the changes marked with the left-right arrow), as we assume all the morphosyntactic variants to have equal status in the graph, instead of choosing the primary one, which would often be not quite evident.
2. Focus shift between participants (1230): *Žuravli **l'et'at** s vostoka* 'The cranes **are flying** from the east' ↔ *Lastočki **l'et'at** na jug* 'The swallows **are flying** to the south' ↔ *Nad gorami **letit** or'el* 'An eagle **is flying** over the mountains' In particular, this change is typical of motion verbs. We treat all the constructions with a mover + one locative participant as basic and formally interrelated by means of a focus shift, instead of deriving them from constructions like *Pticy **l'et'at** s vostoka na jug nad gorami* 'The birds **are flying** from the east to the south over the mountains', as the latter ones are quite rare in our corpus data and do not seem to be natural for human language.
3. Diathetic alternation (407): *Korma lodki **ušla** v vodu* 'The stern of the boat **plunged (lit.: went)** into water' → *Lodka **ušla** v vodu kormoj* 'lit.: The boat **went** into water with its stern'.
4. Omission of a participant belonging to a definite class (335): *On **rastvor'aet** sahar v vod'e* 'He is dissolving sugar in water' → *On **rastvor'aet** sahar* 'He is dissolving sugar'.
5. Omission of a participant which is deictically or situationally defined (875): *Avtobus **prišel** na stanciju* 'The bus **arrived** at the station' → *Begite, avtobus **prišel**!* 'Hurry up, the bus has arrived!'
6. Omission of an indefinite (or unimportant) participant (1152): *Korabl' **plyv'et** iz gavani* 'The ship **is sailing** from the harbour' → *Korabl' m'edl'enno **plyv'et*** 'The ship **is sailing** slowly'.
7. Addition of a participant (2269): *Lastočki **l'et'at*** 'The swallows **are flying**' → *Lastočki **l'et'at** za kormom* 'The swallows **are flying** to find some food' This formal shift usually involves adding peripheral participants like Goal, Reason,

Method, etc. Omission is in its turn marked when there is a core participant of a frame missing in a derived construction (e.g., Instrument in the frames of destruction or any kind of locative participant in the frames of motion).

8. Hybrid of two constructions (91): *Ptica* ***prygala*** *po trav'e* 'A bird **jumped** on the grass', *Ptica* ***prygala*** *p'er'ed domom* 'A bird **jumped** in front of the house' → *Ptica* ***prygala*** *po trav'e p'er'ed domom* 'A bird **jumped** on the grass in front of the house'.
9. Inheritance of a pattern (706): *'"Sl'edujt'e za mnoj"*, – ***skazal*** *oficiant* '"Follow me", – **said** the waiter' → *'"Sl'edujt'e za mnoj"*, – ***brosil*** *oficiant* '"Follow me", – **dropped** the waiter' The annotation of such examples sheds light on the most productive sources of inherited morphosyntactic patterns. These are the verbs *govorit'* 'to speak, to say' (66 constructions acquiring its pattern), *nakazat'* 'to punish' (32 cases), *bol'et'* 'to be ill' (21 cases), *bit'* 'to beat' and *udarit'* 'to hit once' (total 20 cases), *dat'* 'to give' (14 cases), *byt'* 'to be' (12 cases). The position of *govorit'* at the top of the list can be explained by the high productivity of metaphors referring to speech, as well as by the frequent occurrence of metonymic contexts which describe expressing emotions, cf. *"Vot eto fokus!"* – ***udivils'a*** *on* 'lit.: "What a trick!", he **was surprised**' In this example the verb *udivit's'a* 'to be surprised' not only denotes the emotional state of the experiencer, but also indicates that he is saying something. The latter part of meaning is supported by the use of direct speech inherited from verbs like 'to say'. In the case of *bol'et'* 'to be ill', the number of inherited patterns is high, as this semantic domain is inherently metaphorical: according to the cross-linguistic data analyzed in [19], most pain sensations are described with verbs borrowed from other domains (burning, cutting and breaking, sound, etc.), rather than with specific pain expressions. This semantic shift tends to be accompanied with morphosyntactic changes which make source verbs more "similar" to verbs of pain in their construction patterns (see [19] for details). The case of the verb *nakazat'* 'to punish' is a bit different. Many verbs become embedded into a construction with the preposition *za* + NPacc describing Motivation. This argument is typical of *nakazat'* and occurs with other verbs when they denote an action evaluated as punishment, cf. ***ar'estovat'*** *za ubijstvo* '**to arrest** for murder', ***iskl'učit'*** *iz komandy za opozdanije* 'to expel from the team *for being late*', *S'erg'ej* ***byl ostanovl'en*** *policijej za to, čto projehal na krasnyj signal sv'etofora* 'Sergej **was stopped** by the police for running a red light'.

6.2 Semantic Changes

1. Metonymy: an associated participant (517): *Voda* ***zam'erzla*** 'The water **has frozen**' → *Prud* ***zam'erz*** 'The pond **has frozen up**'.
2. Metonymy caused by diathetic alternations (432): *Pojezd* ***jed'et*** *v gorod* 'The train **is going** to the city' → *Ja* ***jedu*** *v gorod pojezdom* 'I **am going** to the city by train'.
3. Metonymy: an associated domain (726): *Vasilij* ***int'er'esujets'a*** *russkoj lit'eraturoj* 'Vasilij **is interested** in Russian literature' → *Vasilij* ***int'er'esujets'a***, *vo skol'ko prihodit pojezd* 'Vasilij **wonders (lit.: is interested)** when the train arrives' Here in the first example the verb *int'er'esovat's'a* 'to be interested in sth.' describes the

mental state of the experiencer, while in the second example it shifts to expressing the speech of a person aiming at find something out.

4. Metaphor (5498): *Mat' **budit** syna* 'Mother **is waking** her son' → *Tišina **budit** vospominanija* 'Silence **evokes (lit.: wakes)** memories'.
5. Rebranding (146): a semantic shift where the derived meaning is an implicature from the source meaning [16], e.g. *Smotri: zv'er' **podhodit*** 'Look: a beast **is approaching**' → *Eto pal'to t'eb'e **podhodit*** 'This coat **suits (lit.: approaches)** you' In this example the idea of something approaching, conveyed in the direct use, implies meeting some standard as a figurative meaning. However, these two domains are not adjacent and therefore are not related metonymically. Neither is there a direct metaphoric relation which could be established between these two meanings.
6. Idiomatization (89): *On ulybnuls'a i **prot'anul** ruku* 'He smiled and stretched his hand' → *Vy tak nogi **prot'an'et'e*** 'You'll turn up your toes (lit.: stretch your legs)'.
7. Specialization (94): *Po utram on **pjet** čaj* 'He drinks tea in the morning' → *On **pjet*** 'He drinks (abuses alcohol)'.
8. Semantic bleaching (46): *javl'at's'a* 'to be (lit.: to come, to appear)'; ***obratit'** vnimanije* '**to pay (lit.: to turn)** attention'.

7 Future Prospects

In the previous sections we have discussed the main parts of FrameBank: the dictionary of construction patterns, the annotation of constructions in corpus examples, the graphs of semantic roles and of shifts between constructions. Since FrameBank is an ongoing project, its development entails many further goals and challenges. The first task is to work out a graph of frames which could tie the constructions from the dictionary to the ontological classification of the lexicon. Although this graph may be to a great extent based on the broad inventory of semantic roles already existing in the database, it will sometimes require a more fine-grained semantic specification of the verbal ontology. The second task is to enlarge the database with constructions of nouns, adjectives, and adverbs which are now on the periphery of our research. It will also be promising to add full-text annotation, as this would allow studying the distribution and interaction of constructions in paragraphs and large texts.

References

1. Lyashevskaya, O.N., Kuznetsova, J.L.: Russian FrameNet: constructing a corpus–based dictionary of constructions [Russkij Frejmnet: k zadache sozdanija korpusnogo slovarja konstruktsij] (in Russian). In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog" [Komp'juternaja lingvistika i intelleltual'nye tehnologii: po materialam ezhegodnoj Mezhdunarodnoj konferentsii "Dialog"], vol. 8, pp. 306–312. RSUH, Moscow (2009)
2. Lyashevskaya, O.: Bank of Russian Constructions and Valencies. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010), pp. 1802–1805. ELRA, Valletta (2010)

3. Kashkin, E.V., Lyashevskaya, O.N.: Semantic roles and construction net in Russian FrameBank [Semanticheskie roli i set' konstrukcij v sisteme FrameBank] (in Russian). In: Computational Linguistics and Intellectual Technologies. Proceedings of International Conference "Dialog", vol. 12-1, pp. 297–311. RSUH, Moscow (2013)
4. Russian National Corpus. http://ruscorpora.ru
5. FrameNet. http://framenet.icsi.berkeley.edu
6. Lexicographer database. http://lexicograph.ruslang.ru
7. Apresjan, J.D., Pall, E.: Russian verb – Hungarian verb. Government and combinability [Russkij glagol – vengerskij glagol. Upravlenie i sochetaemost'] (in Russian). Tankyonvkiado, Budapest (1982)
8. Paducheva, E.V.: Dynamic patterns in lexical semantics [Dinamicheskie modeli v semantike leksiki] (in Russian). Jazyki slavjanskoj kul'tury, Moscow (2004)
9. Fillmore, C.J.: The case for case. In: Bach, E., Harms, R.T. (eds.) Universals in Linguistic Theory, pp. 1–88. Holt, Rinehart and Winston, New York (1968)
10. Dowty, D.R.: Thematic proto roles and argument selection. Language **67**, 547–619 (1991)
11. Apresjan, J.D.: Selected papers, vol. 1, Lexical Semantics [Izbrannye trudy, tom I. Leksicheskaja semantika] (in Russian). Jazyki Russkoj Kul'tury, Vostochnaja Literatura, Moscow (1995)
12. Apresjan, J.D., Boguslavskij, I.M., Iomdin, L.L., Sannikov, V.Z.: Theoretical issues of Russian syntax: the interrelation between grammar and vocabulary [Teoreticheskie problemy russkogo sintaksisa: vzaimodejstvie grammatiki i slovar'a] (in Russian). Jazyki slavjanskih kul'tur, Moscow (2010)
13. Màrquez, L., Carreras, X., Litkowski, K.C., Stevenson, S.: Semantic role labeling: an introduction to the special issue. Comput. Linguist. **34–2**, 145–159 (2008)
14. Palmer, M.S., Wu, S., Titov, I.: Semantic role labeling tutorial. NAACL 2013 tutorials. http://naacl2013.naacl.org/Documents/semantic-role-labeling-part-3-naacl-2013-tutorial.pdf
15. Lyashevskaya, O.N., Kashkin, E.V.: Evaluation of frame-semantic role labeling in a case-marking language. In: Computational linguistics and intellectual technologies. Proceedings of International Conference "Dialog", vol. 13-1, pp. 362–379. RSUH, Moscow (2014)
16. Rakhilina, E.V., Reznikova, T.I., Karpova, O.S.: Semantic shifts in attributive constructions: metaphor, metonymy, and rebranding [Semanticheskie perehody v atribultivnyh konstrukcijah: metafora, metonimija i rebrending] (in Russian). In: Rakhilina, E.V. (ed.) Linguistics of Constructions [Lingvistika konstrukcij], pp. 398–455. Azbukovnik, Moscow (2010)
17. Kustova, G.I.: Types of figurative meanings and mechanisms of linguistic broadening [Tipy proizvodnyh zhachenij i mehanizmy jazykovogo rasshirenija] (in Russian). Jazyki slavjanskoj kul'tury, Moscow (2004)
18. Kashkin, E.V., Lyashevskaya, O.N.: Construction Grapher in Russian FrameBank: semantic and formal kinship map [Semanticheskie i formal'nye sv'azi russkih glagol'nyh konstrukcij: grafovoe predstavl'enije v sisteme FrameBank] (in Russian). In: Proceedings of the Kazan School on computational and cognitive linguistics TEL-2014 [Trudy Kazanskoj shkoly po kompjuternoj i kognitivnoj lingvistike TEL-2014], pp. 165–170. Fen, Kazan (2014)
19. Bricyn, V.M., Rakhilina, E.V., Reznikova, T.I., Yavorska, G.M. (eds.): The concept of pain from a typological point of view [Kontsept bol' v tipologicheskom osveschenii] (in Russian). Vidavnicjij Dim Dmitra Burago, Kiev (2009)

Contents

Pattern Recognition and Machine Learning

http://www.springer.com/978-3-319-26122-5

Analysis of Images, Social Networks and Texts
4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9-11, 2015, Revised Selected Papers
Khachay, M.Y.; Konstantinova, N.; Panchenko, A.; Ignatov, D.I.; Labunets, V.G. (Eds.)
2015, XVIII, 460 p. 145 illus. in color., Softcover
ISBN: 978-3-319-26122-5

Preface

This volume contains the proceedings of the 4th Conference on Analysis of Images, Social Networks, and Texts (AIST 2015)[1]. The first three conferences during 2012–2014 attracted a significant number of students, researchers, academics, and engineers working on interdisciplinary data analysis of images, texts, and social networks.

The broad scope of AIST makes it an event where researchers from different domains, such as image and text processing, exploiting various data analysis techniques, can meet and exchange ideas. We strongly believe that this may lead to cross-fertilisation of ideas between researchers relying on modern data analysis machinery. Therefore, AIST brings together all kinds of applications of data mining and machine learning techniques. The conference allows specialists from different fields to meet each other, present their work, and discuss both theoretical and practical aspects of their data analysis problems. Another important aim of the conference is to stimulate scientists and people from industry to benefit from the knowledge exchange and identify possible grounds for fruitful collaboration.

The conference was held during April 9–11, 2015. Following an already established tradition, the conference was organised in Yekaterinburg, a cross roads between European and Asian parts of Russia, the capital of the Urals region. The key topics of AIST are analysis of images and videos; natural language processing and computational linguistics; social network analysis; pattern recognition, machine learning, and data mining; recommender systems and collaborative technologies; Semantic Web, ontologies, and their applications.

The Program Committee and the reviewers of the conference included well-known experts in data mining and machine learning, natural language processing, image processing, social network analysis, and related areas from leading institutions of 22 countries including Australia, Bangladesh, Belgium, Brazil, Cyprus, Egypt, Finland, France, Germany, Greece, India, Ireland, Italy, Luxembourg, Poland, Qatar, Russia, Spain, The Netherlands, UK, USA, and Ukraine.

This year the number of submissions doubled and we received 140 submissions mostly from Russia but also from Algeria, Bangladesh, Belgium, India, Kazakhstan, Mexico, Norway, Tunisia, Ukraine, and USA. Out of 140, only 32 papers were accepted as regular oral papers (24 long and eight short). Thus, the acceptance rate was around 23 %. In order to encourage young practitioners and researchers we included five industry papers in the main volume and 26 papers in the supplementary proceedings. Each submission was reviewed by at least three reviewers, experts in their fields, who supplied detailed and helpful comments.

The conference also featured several invited talks and tutorials, as well as an industry session dedicated to current trends and challenges.

[1] http://aistconf.org/

The conference featured the following invited talks:

- Pavel Braslavski (Ural Federal University, Yekaterinburg, Russia), Questions Online: What, Where, and Why Should We Care?
- Mikhail Yu. Khachay (Krasovsky Institute of Mathematics and Mechanics UB RAS & Ural Federal University, Yekaterinburg, Russia), Machine Learning in Combinatorial Optimisation: Boosting of Polynomial Time Approximation Algorithms
- Valeri G. Labunets (Ural Federal University, Yekaterinburg, Russia), Is the Human Brain a Quantum Computer?
- Sergey Nikolenko (National Research University Higher School of Economics and Steklov Mathematical Institute, St. Petersburg, Russia), Probabilistic Rating Systems
- Andrey Savchenko (National Research University Higher School of Economics, Nizhny Novgorod, Russia), Sequential Hierarchical Image Recognition Based on the Pyramid Histograms of Oriented Gradients with Small Samples
- Alexander Semenov (International Laboratory for Applied Network Research at HSE, Moscow, Russia), Attributive and Network Features of the Users of Suicide and Depression Groups of Vk.com

The were also two tutorials:

- Alexander Panchenko (Technische Universität Darmstadt, Germany), Computational Lexical Semantics: Methods and Applications
- Artem Lukanin (South Ural State University, Chelyabinsk, Russia), Text Processing with Finite State Transducers in Unitex

The industry speakers also covered a wide variety of topics:

- Dmitry Bugaichenko (OK.ru), Does Size Matter? Smart Data at OK.ru
- Mikhail Dubov (National Research University Higher School of Economics, Moscow, Russia), Text Analysis with Enhanced Annotated Suffix Trees: Algorithmic Base and Industrial Usage
- Nikita Kazeev (Yandex Data Factory), Role of Machine Learning in High-Energy Physics Research at LHC
- Artem Kuznetsov (SKB Kontur), Family Businesses: Relation Extraction Between Companies by Means of Wikipedia
- Alexey Natekin (Data Mining Labs), ATM Maintenance Cost Optimisation with Machine Learning Techniques
- Konstantin Obukhov (Clever Data), Customer Experience Technologies: Problems of Feedback Modeling and Client Churn Control
- Alexandra Shilova (Centre IT), Centre of Information Technologies: Data Analysis and Processing for Large-Scale Information Systems

We would also like to mention the best conference paper selected by the Program Committee. It was written by Oleg Ivanov and Sergey Bartunov and is entitled "Learning Representations in Directed Networks."

We would like to thank the authors for submitting their papers and the members of the Program Committee for their efforts in providing exhaustive reviews. We would also like to express special gratitude to all the invited speakers and industry representatives.

We deeply thank all the partners and sponsors, and owe our gratitude to the Ural Federal University for substantial financial support of the whole conference, namely, the Center of Excellence in Quantum and Video Information Technologies. We would like to acknowledge the Scientific Fund of Higher School of Economics for providing AIST participants with travel grants. Our special thanks goes to Springer's editorial team, who helped us, starting from the first conference call to the final version of the proceedings. Last but not least, we are grateful to all organisers, especially Eugeniya Vlasova and Dmitry Ustalov, and the volunteers, whose endless energy saved us at the most critical stages of the conference preparation.

We would like to mention the Russian word "aist" is more than just a simple abbreviation (in Cyrillic), it means a "stork." Since it is a wonderful free bird, a symbol of happiness and peace, this stork inspired us to organise the AIST conference. So we believe that this young and rapidly growing conference will be bringing inspiration to data scientists around the world!

April 2015

Mikhail Yu. Khachay
Natalia Konstantinova
Alexander Panchenko
Dmitry I. Ignatov
Valeri G. Labunets

http://www.springer.com/978-3-319-26122-5

Analysis of Images, Social Networks and Texts
4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised Selected Papers

Khachay, M.Y.; Konstantinova, N.; Panchenko, A.; Ignatov, D.I.; Labunets, V.G. (Eds.)

2015, XVIII, 460 p. 145 illus. in color., Softcover
ISBN: 978-3-319-26122-5