

# Metaphor Shifts in Constructions: the Russian Metaphor Corpus

Yulia Badryzlova<sup>1</sup>, Olga Lyashevskaya<sup>2</sup>

National Research University Higher School of Economics (HSE)

Russia, Moscow, Staraya Basmannaya 21/4 office 518

E-mail: yuliya.badryzlova@gmail.com<sup>1</sup>, olesar@yandex.ru<sup>2</sup>

## Abstract

In the ongoing corpus project we annotate Russian constructions that have a metaphoric potential. Indirect linguistic metaphors are defined according to a customized version of the metaphor identification procedure MIPVU as the contrast between the basic and the contextual meaning of the lemmas participating in a construction. Direct Metaphors are defined as linguistic metaphors whose contextual meaning has two referents simultaneously, or, in terms of conceptual metaphor, there is a cross-domain mapping. Personification is a subtype of Indirect Metaphor where slots that require only animate participants are filled with non-animate arguments. The annotation of metaphor-related constructions is added as a new layer to SynTagRus, the Russian syntactical dependencies treebank. The paper focuses on the procedure of metaphor identification and the types of linguistic metaphors annotated.

## Introduction

Construction recognition and word-sense disambiguation are areas of Natural Language Understanding that attract much attention and effort of researchers and resource developers. Nevertheless, either a comprehensive theory of CxG or a functional technology of WSD still have a long way to go. Both of them are massive tasks that require large-scale language-specific resources with extensive coverage, which are not always available for under-resourced languages.

Identification of linguistic metaphor (LM) is a challenging area of research which recently received much attention (e.g. Proceedings of the Fourth Workshop on Metaphor in NLP 2016). Contemporary theory of metaphor studies LMs as material manifestations of conceptual metaphors (Lakoff 1993) in language and discourse. The practical demands of enhancing the quality of NLU and NLP also call for advances in LM identification and interpretation. This task can be viewed as a specific type of WSD which, broadly speaking, boils down to the task of classifying words in a text into metaphoric and non-metaphoric occurrences. Such reduction of the task to a binary classification, in its turn, may increase the efficiency of the state-of-the art WSD.

Corpus-based identification of LMs can provide interesting insights into constructions of a given language. We can study constructions in which LMs participate; thus we can identify metaphor not only on the level of lexemes but also on the level of entire constructions. We can describe constructions that occur exclusively as non-metaphors while other constructions will always be metaphoric. Metaphoric usage is particularly characteristic of non-compositional constructions, as in (a) or in the the ‘way’ constructions (b) investigated in (Goldberg 1995)

a) He drank his problems away.

b) She whistled her way down the lane to the silo. (Sato 2012)

Of much interest are the ambivalent cases, when the same construction can be used both metaphorically and non-metaphorically in various contexts; i.e. when metaphoric shifts occur. Cf. e.g.:

- The floor was covered in mud. (non-metaphoric)
- The attack was covered in media. (metaphoric)

## The Corpus and the Annotation Protocol

### SynTagRus, the Russian Treebank

We use as a base the SynTagRus treebank, a 1-million syntactic dependencies annotated subcorpus of the Russian National Corpus (Boguslavsky et al. 2009). In total, there are 69 dependency relations in SynTagRus, including predicative, agentive, quasi-agentive, first complement, modifier, and prepositional (see more at <http://ruscorpora.ru/instruction-syntax.html>). Such detailed annotation makes it comparatively easy to identify the argument structure of the predicate words and some lexicalized syntactic idioms.

In addition, each lemma SynTagRus carries full morphological information: part of speech, gender, number, case, degree of comparison, short form (of adjectives and participles), representation (of verbs), aspect, tense, mood, person, voice, composite form, and attenuation and animacy.

	predic	adv_mod	delim	1-compl	
	Skandal	zaydet	tak	daleko	chto...
		phrasemic unit			
	A,sg,m,nom,inan	V,perf,ind,nonpast,sg,3pers	Adv	Adv	Conj
	Scandal	go	so	far	that...
MRW type		INDIR_MET		INDIR_MET	
Basic Meaning		'To have walked beyond a certain point'		'A long distance away'	
Contextual Meaning		'To have progressed beyond a certain limit' <b>NONBASIC</b>		'To have progressed beyond a certain limit' <b>NONBASIC</b>	

Example 1. Indirect Metaphor: 'The scandal will go so far that...'

	predic	adv_mod	1-compl	prep	appos
	On	yastrebom	vzglivanu	na	Ilyu Ilyicha.
	S,sg,m,nom,anim	S,sg,m,instr,anim	V,perf,ind,past,sg,m	Prep	S,sg,m,acc,anim S,sg,m,acc,anim
	He	hawk	glance	at	Ilya Ilyich
MRW type		DIR_MET:MCue_Morph			
Basic Meaning		'Bird of prey'			
Contextual Meaning		'Bird of prey' <b>EXTRALING</b>			

Example 2. Direct Metaphor with MCue: 'He hawkishly (=fiercely) glanced at Ilya Ilyich.'

The latter is of high relevance for the study of constructions; although (non-)animacy is an inherent grammatical feature of Russian nouns and their dependent adjectives, it quite often correlates with the semantic characteristic of (non-)animacy and, presumably, may be included into analysis of constructions and their metaphoric behaviors.

Beside syntactic relations, SynTagRus is annotated for collocates (referred to as lexical functions, LFs); there are 84 types of LFs in SynTagRus A prototypical LF is a triple of elements {R, X, Y}, where R is a certain general semantic relation obtaining between the argument lexeme X (the keyword) and some other lexeme Y which is the value of R with regard to X. Y is often represented by a set of synonymous lexemes Y1, Y2, ..., Yn, all of them being the values of the given LF R with regard to X. E.g., for English, MAGN is a LF for which the semantic relation is 'high degree': MAGN (desire) = strong / keen / intense / fervent / ardent / overwhelming. LF OPER 1 is a semantically empty verb such that the first actant of a certain situation functions as the subject of this verb and the name of the situation itself is the verb's first object. E.g. in Russian, OPER 1 ('kontrol') = 'osuschestvlyat' (cf. to exercise control) (Boguslavsky et al. 2009).

The annotation of SynTagRus is recognized as a standard of high quality and reliability; it was initially produced by an advanced multi-purpose parser and then edited manually by linguists who corrected errors made by the parser and handled cases of ambiguity that couldn't be reliably resolved without extralinguistic knowledge (ibid).

We identify several types of linguistic metaphors (see below, Metaphor Identification Procedure: a Custom Application of MIPVU) and add a new layer to the existing annotation of SynTagRus.

Originally, the metaphor identification procedure MIPVU was designed to identify metaphoricity in single-word lexical units. However, metaphoric potential is often realized through constructions associated with a word, so the meaning of the construction can be considered a relevant factor of linguistic metaphoricity.

The unit of annotation in our corpus is a construction licensed by a verb, noun, adjective or adverb. Technically, we define the construction as specific types of relations coming in and out of the focal word of the construction and some other secondary projections.

### Metaphor Identification Procedure: a Custom Application of MIPVU

Our metaphoric annotation is predominantly guided by the MIPVU procedure proposed by (Steen et al. 2010) which we have customized to our project.

In MIPVU, metaphor-related words (MRWs) are the words whose contextual meanings are not basic. The basic meaning of a word is:

- more concrete; what it evokes is easier to imagine, see, hear, feel, smell and taste;
- related to bodily action;
- more precise (as opposed to vague) (ibid).

The basic and the contextual meanings for each word in the corpus are established by annotators using dictionary definitions.

It is important to stress that the MRWs identified by MIPVU are not metaphors proper but words that carry a certain metaphoric potential (therefore they are referred to as metaphor-related words, not metaphors).

In our project we currently annotate four parts of speech: verbs, nouns, adjectives and adverbs.

	Starik	ischez	kak	dym
	S,sg,m,nom,anim	V,perf,indic,past,sg,m	Conj	S,sg,m,nom,inan
	Old man	disappear	like	smoke.
MRW type			MFlag	DIR_MET
Basic Meaning				'Small hard particles and gases released in the process of burning'
Contextual Meaning				'Small hard particles and gases released in the process of burning' + EXTRALING

Example 3. Direct Metaphor with MFlag: 'The old man disappeared like smoke.'

	U	menya	srazu	zhe	rozhdetsya	zamysel,	chto...
	Prp	S,sg,m,gen,anim	Adv	Part	V,imperf,indic,nonpast,sg,3pers	S,sg,m,nom,inan	Conj
	At	me	at once	is born		idea	that...
MRW type				S PERS		T PERS	
Basic Meaning				'To be born, come to life'			
Contextual Meaning				'To emerge':	NONBASIC		

Example 4. Personification: 'The idea is immediately born in my mind that...'

We employ the workflow where the annotation process is subdivided into the two parallel activities: 1) deciding about basic meanings and recording them in a retrievable database; 2) deciding about contextual meanings of the words in the corpus, i.e. the WSD processing.

### Deciding about Basic Meanings

Basic meanings are defined by linguists with expertise in metaphor identification. They examine all the meanings of a given lexeme; the meanings that comply with MIPVU's criteria of basic meaning are recorded in the database with the tag 'BASIC'. The meanings that fail these criteria are tagged as either 'NON-BASIC' or WIDLII (acronym inherited from MIPVU to indicate borderline or ambiguous cases).

For example, the meanings of the adverb 'daleko' (far) will receive the following tags:

1. 'A long distance away': BASIC;
2. 'Not soon': NONBASIC;
3. 'Inferior to': NONBASIC.

Beside the meanings of the headword, the dictionary entries also feature phrasemic patterns in which it participates. Some phrasemic patterns are associated with more than one construction depending on the semantics of the lexical items that fill the non-fixed slot(s). E.g. #5 in the list of phrasemic patterns for 'daleko' (below) corresponds to the two constructions:

- DALEKO\_ADV ZA\_PREP TIME\_N,inan / ADV
- DALEKO\_ADV ZA\_PREP AGE\_NUMcard.

Just as isolated word meanings, the phrasemic patterns will also be tagged in terms of BASIC vs. NONBASIC. WIDLII is used when a construction allows both metaphoric

and non-metaphoric usage in Russian, even though statistically such patterns may be predominantly biased towards metaphor, as in #6 and #7 below (see also Example 1).

4. 'Daleko za...' (*lit.* far beyond...) – 'a long time after'; much older than...: NONBASIC;
5. 'Daleko ne...' (*lit.* far not...) – 'by no means': NON-BASIC
6. 'Daleko zayti' (*lit.* to have walked far) – 'to have progressed beyond a certain limit': WIDLII;
7. Daleko poyti (*lit.* to go far) – to become successful, to make a career: WIDLII.

If every meaning of a lexeme is tagged as BASIC, this word (and consequently, the constructions associated with it) are discarded from further word-sense disambiguation; we presume that these constructions are void of internal metaphoric potential because their meanings cannot be contrasted with each other on the basis of concreteness, body-relatedness and preciseness, as required by MIPVU. The same reasoning applies if all the dictionary meanings of a word are tagged as NONBASIC; such words and their constructions are excluded from the subsequent WSD.

However, they are not removed from the overall metaphor analysis because they may still display external metaphoric potential – this type of MRW is called Direct Metaphor (see Types of MRWs).

A construction may be restored for further WSD if a new construction is encountered in the corpus, whose meaning is not recorded in the dictionary, and the new meaning belongs to the opposite class (BASIC vs. NONBASIC) than the previously tagged meanings. For instance, the verb 'zavisnut' has only one meaning in the dictionary: 'to be hanging, to

hover’. However, at least three other constructions are frequently attested in modern Russian usage:

- COMPUTER EQUIPMENT<sub>S,inan</sub> ZAVISNUT’\_V – ‘to freeze’;
- PERSON<sub>S,anim</sub> ZAVISNUT’\_V – ‘to interrupt a current activity and start thinking about smth’;
- PERSON<sub>S,anim</sub> ZAVISNUT’\_V (PREP) PLACE<sub>N,inan</sub> / ADV – ‘to hang around’.

### Deciding about Contextual Meanings

Defining contextual meanings, or WSD, does not require specialized linguistic expertise and can be carried out by any competent native speaker of the language – and thus can be crowdsourced.

Contextual meanings are selected from the dictionary and recorded in the database for every token of the corpus that is valid for WSD. Special treatment is given to Direct Metaphors and Personifications (see Types of MRWs).

If the attested contextual meaning is BASIC, the token receives the tag NONMET. If the contextual meaning is NONBASIC, it is tagged as Indirect Metaphor, INDIR\_MET.

### Types of MRWs

Indirect Metaphors are MRWs that are based on the juxtaposition of basic and non-basic meanings. For instance, see Example 1 where the construction<sub>S,nom</sub> ZAYTI\_V DALEKO\_Adv is used in its metaphoric meaning.

Beside Indirect Metaphors, MIPVU also differentiates other types of MRWs: Direct Metaphor, Personification and Implicit Metaphor.

Direct Metaphor is created when the contextual meaning simultaneously has two referents, the prototypical referent and the referent defined by the extralinguistic frame of reference defined by the context or, in terms of the conceptual metaphor theory, the meaning is involved in cross-domain mapping. Such words are tagged as DIR\_MET in our corpus, and their contextual meanings receive the additional notation + EXTRALING, as in Example 2. In this sentence the instrumental verbal construction has the meaning of the manner of action (here of the verb ‘vzglyanut’ – ‘to glance’):<sub>S,nom</sub> VZGLYANUT’\_V<sub>S,instr</sub>. The instrumental case of the second noun contributes to the creation of metaphor and can be used as its predictor; we tag it as the morphological Metaphor Cue: MCue\_Morph.

Direct Metaphors can also be signaled explicitly by specific lexical units, as in Example 3. Such lexical units are tagged as Metaphor Flags, MFlag.

Personification is a subtype of Indirect Metaphor when a construction whose basic meaning(s) requires only animate participants in certain slots features an inanimate argument instead. Such lexemes are labeled in the dictionary with the tag <PERS>, as in the verb ‘rodit’sya’ – ‘to be born’.

‘rodit’sya’<PERS>

1. ‘To be born, come to life’: BASIC;
2. ‘To emerge’: NONBASIC;

3. ‘To bear fruit’: NONBASIC.

Such lexemes are tagged in the corpus as Source of Personification, S\_PERS, while the inanimate arguments that fill the animate slots are tagged as the Target of Personification, T\_PERS, as in Example 4.

We also annotate Implicit Metaphor – these are pronominal referents of MRW antecedents.

The presented multilevel architecture of the annotation scheme will enable users of the corpus to obtain the data about Russian constructions that will allow them to capture correlations between the presence of metaphoric potential in a construction (metaphor-related vs. nonmetaphor-related), its type (Indirect, Direct, Implicit or Personification) and the morpho-syntactic characteristics of the construction and its participants.

## Conclusion

The presented annotation scheme provides a tool for assessing the metaphoric potential of Russian constructions in discourse. It also affords in-depth analysis of those components of the semantics of constructions that give rise to their metaphoric potential. The corpus will demonstrate how metaphors are created through interactions within the semantics of constructions and through interaction with the surrounding context.

## Acknowledgments

This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE).

## References

- Boguslavsky, I., Iomdin, L., Frolova, T. and Timoshenko, S. 2009. Development of a Russian Tagged Corpus with Lexical and Functional Annotation. In *Proceedings of MONDILEX Third Open Workshop*, 83-90. Bratislava, Slovakia.
- Goldberg, A.E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*: University of Chicago Press.  
<http://ruscorpora.ru/instruction-syntax.html>. Syntactically Annotated Corpus of Russian Language: User Guide: *Russian National Corpus*.
- Lakoff, G. 1993. The Contemporary Theory of Metaphor. In Ortony, A. (Ed.), *Metaphor and Thought*, 2nd ed. Cambridge: Cambridge University Press.
- Proceedings of the Fourth Workshop on Metaphor in NLP*. 2016. San Diego, CA, USA: The Association for Computational Linguistics.
- Sato, H. 2012. A Search Tool for FrameNet Construction. In *Proceedings of LREC 2012*, 1655–1658.
- Steen, G.J., Dorst, A.G., Herrmann, J.B., Kaal, A., Krennmayr, T. and Pasma, T. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*, Philadelphia, PA, Amsterdam: John Benjamins Publishing Company