NORDSCI INTERNATIONAL CONFERENCE

12-14 October 2020
www.nordsci.org

STATISTICAL ANALYSIS OF RUSSIAN MULTIWORD PREPOSITIONS

Assoc. Prof. Dr. Victor Zakharov[1]
Mgr. Anastasia Golovina[2]
Assoc. Prof. Dr. Irina Azarova[3]
[1, 2, 3] St Petersburg University, Russia

Introduction

The present study deals with Russian multiword prepositions (MWPs) – multiword units of content words (nouns, adverbs, verbs) combined with primary prepositions. Multiword prepositions perform the grammatical function of a preposition in some contexts and can be a free combination in others. Although a large subclass, it remains poorly defined in Russian grammar books and dictionaries. As a matter of fact, there exists no exhaustive list of all Russian MWPs.

We have tackled this issue by forming our own list of MWPs identified as such in various reputable sources. The main characteristic features of multiword prepositions have been formulated in order to define the borders of this subclass.

Presented here is the statistical analysis of the use of multiword prepositions in the Araneum Russicum III Maius corpus (1.25 billion tokens) created by Vladimír Benko (www.unesco.uniba.sk).

Our current research is part of a larger study that aims to create the first quantitative grammar of the Russian prepositional system.

Main characteristic features of a multiword preposition:

- **governs** a **noun**, a **nominalised** word or an **infinitive**;
- **inherits the semantics** of the notion word (noun, verb) it derives from as well as its **valency;**
- contains **one or two primary prepositions;**
- **nominal components** tend to have **abstract** semantics;
- has a relatively **high frequency** among multiword units of the **same structural type**;
- has a **primary preposition as a synonym**;
- does **not allow** for **separation** or **determiner insertion**;
- its **nominal** component **loses its lexical meaning** to an extent;
- the **grammatical number** of the noun **cannot be changed**;

All of these features are characterised by significant **statistical regularity**.

Research

Figure 1. Percentage of MWPs vs. free combinations within 25 top frequency n-grams of the same structural type. MWPs are found to be highly common multiword units with stable bonds between components.
Table 1. Shared frequency counts of top frequency MWPs and some common vocabulary notion words (as per the "Frequency Dictionary of the Modern Russian Language"). MWPs appear to be regular and essential elements of the Russian syntactic sentence structure.
Table 2. Frequency counts of 4 nouns used almost exclusively as MWP components. Loss of independent use signifies partial transition to multiword unit component status.

Prepositional vs. free uses of some multiword units in random corpus samples (100 contexts):
- **exclusively prepositional** use (MWP subclass core): *в виде, по сравнению с/со, через посредство*;
- **mixed use** (MWP subclass periphery; predominant group): *смотря по* (91% MWP), *по части* (68%), *что до* (18%).

Figure 1.

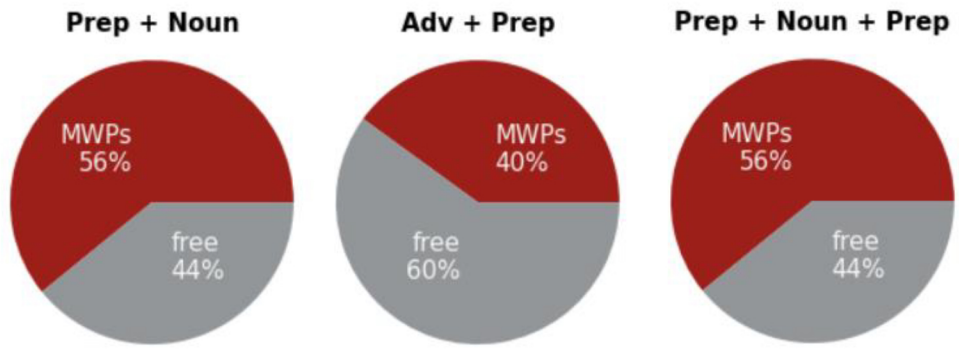

Table 1.

MWP	ipm	Notion word
во время *('during')*	268	положение *('position')*
с помощью *('with the help of')*	240	красный *('red')*
в качестве *('as')*	215	вспомнить *('to remember')*
в течение *('over the course of')*	207	принимать *('to accept')*

Table 2.

Base word	MWPs	Base word ipm	As MWP component	% of MWP use
довершение	в довершение, к довершению	0.63	0.61	97
неимение	за неимением, по неимению	1.49	1.36	91
рамка	в рамках, в рамки, вне рамок, за рамками, за рамки, из рамок	204.30	175.84	86
зависимость	в зависимости от(о), вне зависимости от(о)	171.10	126.68	74

Conclusion

The data presented signify that MWPs as a class are widely represented in Russian. It is a **large and diverse subclass** that is nonetheless characterised by a number of common features and, therefore, **lends itself to description, definition and measurement**.

Further stages of our research:

- calibration and extension of described experiments, statistical data on the entirety of the MWP subclass;
- detection of clusters of conditional synonymy;
- study of preposition stranding and separability of the components of MWPs;
- description of prepositional use in fixed phrases and idioms.

Acknowledgements

This work is supported by the Russian Foundation for Basic Research [grant No. 17-29-09159 **"Quantitative grammar of Russian prepositional constructions"**].