Lexicon-Grammar of Russian Verbal Idioms

Tetyana Fukova
University of Algarve

tatyanafukova@gmail.com

Svitlana Chornobay
Crimean Federal
V.I.Vernadsky University
svetoniy@gmail.com

Jorge Baptista
University of Algarve
L2F/INESC-ID Lisboa
jbaptis@ualg.pt

Abstract

This paper describes an on-going project to build a lexicon-grammar of Russian verbal idioms for natural language processing. The aim is to produce a language resource that can be used to automatically identify these idioms in naturally occurring texts. Such resource can also be useful to several fields of research, such as language acquisition, or language learning and teaching, natural language processing, among others.

1. Introduction

Though many definitions, and also many conceptual and terminological disputes, can be found in the literature around the term *idiom* (*idioms*, *collocations*, *phrasemes*, etc.), for this paper we adopt the definition of *verbal idioms* given by Baptista (2004), who also called it *frozen sentences*:

> "*Frozen sentences* are elementary sentences where the main verb and at least one of its argument are distributionally constraint, and usually the global meaning of the expression cannot be calculated from the individual meaning of its component elements when they are used independently."

For example, in Russian, an idiom like держать язык за зубами (deržat' jazyk za zubami)[1], literally 'to hold one's tongue behind one's teeth', and corresponding to the English idiom 'to keep one's tongue between one's teeth', has nothing to do with the relative position of the *tongue* and the *teeth*, but rather 'to be prudent when speaking, not saying things that should not be said'. The body-part nouns are frozen with the verb and the locative preposition. However, the nouns establish a part-whole relation with a free human noun (in the form of a possessive), as the translation illustrates, and this have referential value – possessives are obligatorily co-referent with the free human subject of the verb. Therefore, there is some syntactic structure in the idiom, and some lexical and structural variation, which must be adequately captured, and not just try to match the sequence as a whole (Baptista *et al*., 2004; Cowie, 1998). The idiom, however, can

---

[1] In this paper, the Passport (2013), ICAO romanization norm was adopted.

be parsed like any ordinary, that is, semantically compositional, sentence. In fact, this expression could be uttered, and literally interpreted, in an appropriate context (for example, by a dentist to his patient). Ambiguous idioms like this also need to be specifically marked, so that the adequate reading be found from context, if possible.

The automatic identification of the meaning units in texts involves the correct delimitation and tagging of idioms in texts. Using available linguistic resources, such as phraseological dictionaries (Molotkov, 1986; Fedosov and Lapisky, 2003) and the linguistic development platform UNITEX (Paumier 2003, 2014)[2], along with the machine readable dictionary distributed with this software, we intend to determine the relevant linguistic information required to process this type of expressions, and to formalize it into a database of idioms. This we will call the Lexicon-Grammar of the Russian idioms, in the sense that it contains not only the lexical entries of the idiomatic frozen sentences, but also the relevant linguistic properties that describe their behaviour in texts.

For the remainder of this paper, we describe preliminary experiments on the application of this linguistic resource to the task of identifying the Russian idioms described in the lexicon-grammar to real texts. The paper is structured as follows: In Section 2, the methods, the process of data collection and the formal classification of the idioms are presented (§2.1). Then, the corpus collection and annotation process is explained (§2.2). Next, the reference graphs building process is presented and exemplified (§2.3). Section 3 presents the evaluation of some preliminary experiments on automatic identification of Russian idioms in the corpus. Finally, in Section 4, the paper concludes with some perspectives for future work.

2. Methods

2.1. Data collection

To this date, we collected almost 1,000 Russian verbal idioms from phraseological dictionaries and other sources, and classified most of them using the Lexicon-Grammar framework (Gross, 1996). The idioms were formalized into a tabular format, aiming at computational processing and automatic identification in texts. This database is called a *Lexicon-Grammar* and it consist of a fine-grained description of the syntactic structure of those idioms, the lexical content of their frozen elements, the distributional constraints on their free syntactic slots (human/non-human), and the transformational properties of the construction, that is, the alternative, paraphrastically equivalent, forms (or alternations) they can yield (*e.g.* Passive).

For each idiom, a word-by-word English translation and the relevant morphosyntactic (part-of-speech) information is provided, along with a free translation (or gloss) or the English equivalent, when available or known. An illustrative, 'artificial' example, produced with just the essential elements of the idiom is also provided. In these examples, verbs are usually in the past tense; the human noun free slots are fulfilled by a proper noun, etc.

The classification here adopted is inspired in that proposed by M. Gross (1982; see a later synthesis in M. Gross 1996), and already adopted for several other languages or language varieties: French and the four main varieties, from France, Belgium, Switzerland and Québec (Lamiroy, 2010), Greek (Fotopoulou, 1993), Italian (Vietri, 2015), Portuguese, both European and Brazilian (Baptista, 2004, 2014; Vale, 2001).

Since Russian nominal morphology includes cases, the classification was adapted so that, instead of (or along with) prepositions, cases are used to mark the syntactic function of the verbs' arguments. Table 1 presents a gist of this classification:

Table 1. Classification of Russian verbal idioms

Class	Structure	Example	Count
C1	$N_0 V C\text{-}acc_1$	*Бить баклуши* (bit' baklushi) N_0 beat/V spoons/C_1-acc to twiddle one's thumbs, to be idle	311
CP1	$N_0 V (Prep_1) C_1$	*Влететь в копеечку* (vletet' v kopeechky) N_0 fly/V in/Prep penny/C_1-acc to cost smb. a pretty penny	220
CAN	$N_0 V (C\text{-}acc\ N\text{-}gen)_1$ $= N_0 V (C\text{-}acc\ N\text{-}dat)_1$	*Капать на мозги* (kapat' na mozgi) N_0 drop/V on/Prep brain/C_1-acc smb/N-dat/gen to assert sth. over and over to s.o	50
CPN	$N_0 V Prep (C\text{-}acc\ N\text{-}gen)_1$	*Играть на нервах* (igrat' na nervah) N_0 play/V on/Prep nerves/C_1-obliq N-gen to jangle on someone's ears/nerves	6
C1PN	$N_0 V C\text{-}acc_1 (Prep_2) N_2$	*Задать пару* (zadat' paru) N_0 set/V steam/C_1-acc smb/N_2-dat to give smb. hell	80
CNP2	$N_0 V N\text{-}acc_1 (Prep_2) C_2$	*Взять под крыло* (vzyat' pod krilo) N_0 take/V smb/N_1-acc under/prep wing/C_2-acc to take smb. under one's wing	183
C1P2	$N_0 V C\text{-}Acc_1 (Prep_2) C_2$	*Брать быка за рога* (brat' bika za roga) N_0 take/V bull/C_1-acc of/Prep horns/C_2-acc to take the bull by the horns	99
CPP	$N_0 V w (Prep_1) C_1(Prep_2) C_2$	*Лезть в душу без мыла* (lezt' v dushu bez mila) N_0 get/V into/Prep soul/C_1-acc without/Prep soap/C_2-gen to try to gain smb.'s favor or trust by cunning	13
CADV	$N_0 V Adv_1 w$	*Выходить боком* (vihodit' bokom) N_0 appear/V sideways/Adv to turn out badly	24
		Total	987

N0 stands for the subject, *N1* and *N2* for the first and second complement, respectively. C_1 and C_2 indicate the constant (frozen) element of the complement; *Prep* is a preposition and *Adv* and adverb; w represents an unspecified sequence of complements; the cases are shortened: *acc* for accusative, *dat* for dative and *gen* for genitive.

Notice that in class CP1, the first complement is either introduced by a preposition or it does not receive the accusative case (*acc*). The same happens in all the remaining classes. No cases were yet found of frozen subject.

The work of collecting the idioms from dictionaries and other sources is still going on but, based on the length of the dictionaries already consulted, a reasonable estimate would consider around 2 thousand and up to 3 thousand of frequently occurring, frozen verbal idioms to attain a reasonable lexical coverage.

2.2. Corpus collection and annotation

For this paper, we used the Russian National Corpus[3], henceforward *RNC*, which covers primarily the period from the middle of the 18th to the early 21st centuries. According to the corpus' website:

> "[t]his period represents the Russian language of both the past and the present in a wide range of sociolinguistic variants: literary, colloquial, vernacular, in part dialectal. The corpus includes original (non-translated) works of fiction (prose, drama and poetry) of cultural importance, which are interesting from a linguistic point of view. Apart from fiction, the corpus includes a large volume of other sources of written (and, for the later period, spoken) language: memoirs, essays, journalistic works, scientific and popular scientific literature, public speeches, letters, diaries, documents".

The full RNC corpus contains 85,996 documents, 19,362,746 sentences and 229,968,798 words. We chose the 10 most frequent verbs from the Lexicon-Grammar table, excluding verbs that are often support verbs, namely брать (brat) 'to take' (26 entries), давать (davat) 'to give' (19), and делать (delat) 'to do' (12). The most frequent verbs are : держать (derzhat) 'to hold', идти (idti) 'to go', играть (igrat) 'to play', бить (bit) 'to beat', смотреть (smotret) 'to look', класть (klast) 'to put', лезть (lezt) 'to climb', лежать (lezhat) 'to lie', выйти (viiti) 'to go out', жить (zhit) 'to live'.

Using the RNC web interface in its default options, several queries were conducted, extracting the sentences where each one of these verbs (the lemma and the associated inflected forms) occur (we called this **Corpus 1**). Next, we retrieved from the website the top search results (about 2,000 sentences) in a spreadsheet format. Random numbers were accorded to the sentences and, after sorting them, the first 50 were selected. We then manually annotated the idioms found, by delimiting the verb and the frozen elements with underscore, '_'. Table 1 shows the results from this data collection and annotation process.

Table 1. Corpus 1 data collection (from Russian National Corpus).

verb	translit	gloss	RNC	in sample (n=50)	diff. idioms	diff. LG entries	Total LG entries w/ V
держать	derzhat'	to hold	50 643	5	4	4	33
идти	idti	to go	241 225	1	1	1	23
играть	igrat'	to play	66 077	0	0	0	14
бить	bit'	to beat	33 393	6	5	4	12
смотреть	smotret'	to look	157 516	0	0	0	11
класть	klast'	to put	10 458	1	1	1	11
лезть	lezt'	to climb	11 273	3	3	3	9
лежать	lezhat'	to lie	80 235	2	2	2	9
выйти	viiti	to go out	165 311	1	1	1	8
жить	zhit'	to live	187 841	2	2	2	7
		Total	**1 003 972**	**21**	**19**	**18**	**137**

---

[3] http://www.ruscorpora.ru/

The 500 instances of these 10 verbs retrieved from the corpus constitute just a small sample, which cannot be considered representative of the total of occurrences of those verbs in RNC corpus. Our point was to have a glimpse of how complete the Lexicon-Grammar may already be. Still, 4.2% of the retrieved instances contained idioms, attesting the frequency of the phenomenon. The 21 instances correspond to 19 different idioms, which signals few repetitions. These idioms, in their turn, represent only 18 entries on the Lexicon-Grammar, but only one idiom, бить поклоны (bit' pokloni) 'pay smb. one's respects', had not been previously registered, thus indicating adequate lexical coverage. Still, the 18 entries found are just a fraction (13%) of all the entries involving those same verbs that had already been collected in the Lexicon-Grammar.

In a second moment, we retrieved from the RNC all the sentences containing the verbs of the previous list and the first frozen nominal element of the 137 idioms listed in the Lexicon-Grammar involving those verbs, allowing for a window from 0 up to 3 intervening words. As in the sampling procedure above, we randomly selected the top 50 sentences; if the result yields less than 50 occurrences, all of them were considered. These sentences constitute **Corpus 2**. This second sample also underwent manual classification of the idioms found therein, using the same delimitation procedure, with underscores '_'. Since some expressions can be ambiguous between a literal and a figurative (idiomatic) interpretation, literal expressions were delimited with '#'. Results of this procedure can be found in Table 2.

**Table 2.** *Corpus 2* data collection (from Russian National Corpus) (extract)

Verb	Translit	Gloss	diff. idioms	matches	idioms/50
Держать	derzhat'	'hold/keep'	29	1,048	594
Идти	idti	'go'	23	988	490
Играть	igrat'	'play'	13	468	189
Бить	bit'	'beat'	11	501	295
Смотреть	smotret'	'look'	9	296	158
Класть	klast'	'put'	9	229	71
Лезть	lezt'	'climb'	5	225	154
Лежать	lezhat'	'lie'	6	253	152
Выйти	viiti	'go out'	6	229	151
Жить	zhit'	'live'	6	193	80
			117	4,430	2,334

In the Appendix, the full list of the 117 idioms is presented, with the breakdown of matches

2.3. Building reference graphs
In order to automatically retrieve from texts the idioms represented in the Lexicon-Grammar in a tabular format, we used UNITEX (Paumier 2003, Paumier 2014), an open-source linguistic development platform. One of its functionalities is to intersect

data matrices with finite-state transducers (FST), which can then be used to match and label complex patterns in texts. This is done by first building reference graphs, which are Directed Acyclic Graphs (DAGs) where each variable (noted @X) refer to the corresponding column in the matrix. Figure 1 illustrates the reference graph for class **C1**:
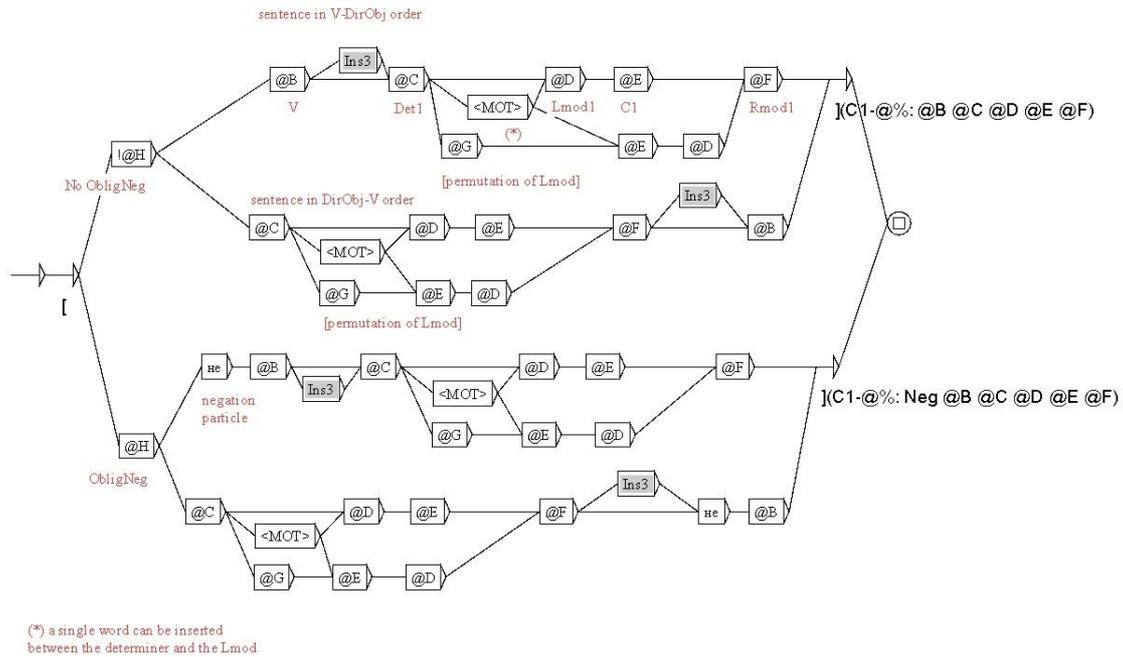
Fig. 1. Reference graph for class **C1**, e.g. Бить баклуши, N_0 *beat/V spoons/C_1-Acc* 'twiddle one's thumbs'

The system reads each line of the matrix in turn, and then builds a graph for each expression, replacing the variables by the content of the corresponding cell in the matrix. The graph in Fig. 1 reads as follows: Variable @H corresponds to the binary property 'obligatory negation', e.g. Бровью не повести (brov'u ne povesti) N_0 *not_stir/V eyebrow*/N-instrum 'do not display emotions toward smth.' This property is encoded by '+' and '-'; the graph is split in two paths, the upper path for the expressions that do not allow for this property (the symbol '!' indicates negation) and the lower path for those that do. The negative particle не (ne) explicitly appears in the lower path. When the system reads variable @H, it follows and builds a graph for the remainder of the path if this variable, at the given line in the matrix, is marked with '+' (the reverse happens if it is marked with negation: !@H); otherwise, it collapses the graph at that point.

Next, the sequence of elements constituting the idiom are represented by the variables @B for the verb, @C for the determiner, @D for the left modifier (LMod), @E for the frozen noun of the direct object (C_1), and @F for the right modifier (RMod); an insertion window from 0 to 3 words is represented by auxiliary sub-graph Ins3 (grey box).

The transformation consisting in the alternative ordering of the left modifier is encoded by the property represented by variable @G, so that in this case the intervening elements are permuted. Finally, a single word may be inserted between the determiner and the left modifier, which is represented by the node <MOT>.

The transducer's *output* is represented below the nodes. This consists in the square brackets used as delimiters for the matching strings, and the codes for the class and the line number of each idiom (provided by variable @%, indicating the current line). In this output, we also added (inside brackets) the main elements of the idiom.

A similar procedure was carried out for the remaining classes. The system generates a sub-graph for every line of the matrix Fig. 2 shows two sub-graphs from

class CP1, one with obligatory negation. All the sub-graphs of the same class are automatically gathered in a resulting graph, which can then be applied to texts.
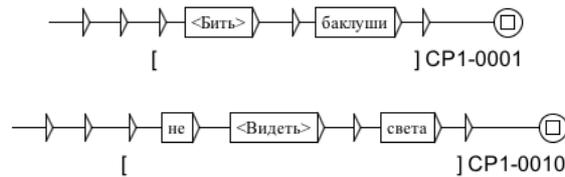
Fig. 2. Two sub-graph from class **CP1**, e.g. *Влететь в копеечку* (vletet' v kopeechky) N_0 *fly/V in/Prep penny/C_1-acc* 'to cost smb. a pretty penny'

3. Evaluation

In order to do some preliminary evaluation of the Lexicon-Grammar and the corresponding finite-state transducers built for the automatic identification of idioms in texts, we first apply them to the two lists of examples that were produced for each idiom. Two formats of examples were produced: (a) 'lexical entry' examples, where there is neither a subject nor other free arguments, just the verb in the infinitive and its frozen complements, like one would find the idiom in a dictionary; and (b) 'artificial examples', produced as described above (section §2.1), with the verb in an inflected form, a proper noun as a human subject, etc. The first list of examples is meant to test the FSTs while ignoring inflection, agreement and word order issues. To a limited extent, the second list of examples highlights the impact of these issues to the language-processing task. Table 3 shows the results of these two experiments:

Table 3. Evaluation of class **C1**

examples (N=312)	matches	precision
lexical entries	278	0.89
artificial examples	248	0.79

Precision of the results is relatively high. Actually, most word combinations in the dictionary of frozen sentences are unique, therefore, unequivocal combinations. However, some matches were not found and defined as idioms. One of the reasons for not recognizing idioms in the text is the lack of a verb or a noun (frozen component), or one of their inflected forms, in the dictionary. For instance, in the example *лежать лежнем* 'lezhat lezhnem'' (N_0 *lie*$_V$ *ledger*$_{N\text{-}Instrum}$), which means 'to be in poor health', (the inflected form of) the noun *лежень* lezhen (ledger) is not present in the dictionary, and as the Lexicon-Grammar encoded C_1 by its lemma $<$ *лежень* $>$, the system fails to recognize the inflected form *лежнем*. A solution is to consider the word forms only, but that cannot always be the case, as some constants do inflect, while the verb is always inflected.

Another reason for not recognizing idioms is the passive voice, which is expressed by the suffix *ся* 'sya' of verbs. These verb forms have been attributed to different lemmas from the active forms, in the dictionary produced with the system, and therefore they were not detected. In order to make such verbs not to have several lemmas, we made some changes in the dictionary. We extracted all the verbs that ended on the suffix *ся* (sya), added lemma without this suffix and appended this grammatical feature to those entries.

4. Conclusions and future work

To the best of our knowledge, there is no available machine-readable lexicon of Russian verbal idioms. Idioms represent a significant part of the lexicon units of any language. The identification of frozen sentences is a necessary part of the lexical analysis of texts for natural language processing, as they constitute units on meaning.

In this paper, we have presented the current state of an on-going project of building a Lexicon-Grammar of Russian verbal idioms, containing around 1,000 entries, described the information there encoded, and the finite-state tools built to identify those expressions in texts. The preliminary tests of the examples have shown that this is no small task, as a number of other language-processing issues have to be dealt with, besides the idioms' linguistic description. Among them, there is the degree of lexical coverage of the base dictionary and the quality and granularity of the information there represented, something outside the scope of the paper. Even so, adaptation and extension of the current, freely distributed, machine-readable dictionary of the Russian language is underway, like the treatment given to passive verb forms. On the other hand, refining of the reference graphs will allow to deal with several syntactic issues not yet addressed, like the interaction of obligatory negation with passive or the intrinsically reflexive constructions. The free-order syntax of sentence constituents in Russian also produces highly complex reference graphs that require a systematic verification of all possible combinatory variants. The paper also produced two annotated sub-corpora, extracted from the Russian National Corpus, where verbal idioms were manually delimited, including literal uses of ambiguous expressions. These resources can now be used as a workbench to help develop and test several procedures for automatic detection and delimitation of Russian verbal idioms.

References

BAPTISTA, J. 2004. Compositual vs. Frozen Sequences. Laporte, Eric; Ting Au-Chen, (eds).Proceedings of the Lexicon-Grammar Workshop. Beijing 14-18 de Outubro de 2004. *Journal of Applied Linguistics*, Special Issue on Lexicon-Grammar. Papers presented at the Lexicon-Grammar Workshop, pp. 81-93 (Chinese version).

BAPTISTA, J., CORREIA A. AND FERNANDES G. 2004. Frozen Sentences of Portuguese: Formal Descriptions for NLP. *Workshop on Multiword Expressions: Integrating Processing, International Conference of the European Chapter of the Association for Computational Linguistics*, Barcelona (Spain), July 26, 2004. ACL: Barcelona, pp. 72-79.

CHORNOBAY, S., BAPTISTA, J. 2014. Semantic Peculiarities of Portuguese and Russian Idioms within the Conceptual Domain "Death". International Scientific Conference Modern Philology: Paradigms, trends, problems (Міжнародна наукова конференція "Сучасна філологія: парадигми, напрямки, проблеми"), October 9, 2014, Kyiv, Kyiv National Taras Shevchenko University.

COWIE, A. 1998. *Phraseology. Theory, analysis, and applications*. Oxford: Oxford University Press.

FEDOSOV, I. and LAPITSKY, A. 2003. *Phraseological dictionary of the Russian language.* (Федосов, И. и Лапицкий, А. *Фразеологический словарь русского языка*). Moscow: Unves.

FOTOPOULOU, A. 1993. Une classification des phrases a complèments figés en grec moderne: étude morphosyntaxique des phrases figées . Ph.D. thesis, Université Paris VIII.

GROSS, M. 1982. Une classification des phrase "figées" du français. *Revue Québécoise de Linguistique* 11-2: pp. 151-185.

GROSS, M.1996. Lexicon-Grammar. *Concise Encyclopedia of Syntactic Theories*. Cambridge. Pergamon. pp.244-258.

LAMIROY, B., KLEIN, J.-R. 2010. Lamiroy, Béatrice, and Jean-René Klein. *Les expressions verbales figées de la francophonie: Belgique, France, Québec et Suisse*. Editions OPHRYS.

MOLOTKOV, A. 1986. *Phraseological dictionary of the Russian language* (Молотков, А., *Фразеологический словарь русского языка*). Moscow: ACT.

PAUMIER, S. 2003. De la reconnaissance des formes linguistiques à l'analyse syntaxique. PhD thesis, Université de Marne-la-Vallée, 2003.

PAUMIER, S. 2014. *Unitex 3.0 - User's Manual*. Paris: Université Paris-Est Marne-la-Vallée.

VALE, O. 2001. *Expressões Cristalizadas do Português do Brasil: Uma proposta de tipologia*. Araraquara, SP (Brasil): Universidade Estadual Paulista.

VIETRI, S. 2015. *Idiomatic Constructions in Italian. A Lexicon-Grammar Approach*. Amsterdam: John Benjamins.

APPENDIX

In this table, for the 137 entries of the Lexicon-Grammar of Russian Verbal Idioms that are formed with the 10 most frequent verbs (excluding some support/auxiliary verbs), we indicate the 117 that were found, their transliteration and a gloss, along with (last two columns): (a) the total number of matches extracted from the Russian National Corpus with the queries consisting of the *verb* and the head noun of the first frozen complement (C_1), allowing for a window of 0 up to 3 intervening words (V-C_1) – these matched sentences constitute *Corpus 2*; and (b) the number of idioms found therein; if the total of V-C_1 matches is larger than 50, only a random sample of 50 matches was used; if they are less than 50, the entire V-C_1 list of matches was used.

Idiom (V-C_1)	Transliteration	Gloss	# Corpus 2	idioms/50
Держать в курсе	derzhat' v kurse	keep informed	324	26
Держать в уме	derzhat' v ume	carry smth. in one's head / mind	122	44
Держать в черном теле	derzhat' v chernom tele	keep smb on short rations	88	17
Держать верх	derzhat' verkh	gain the upper hand	16	10
Держать вожжи в руках	derzhat' vozhi v rukah	concentrate power or leadership in one's hands, taking full control of sth	1	1
Держать дверь открытой	derzhat' dver' otkritoi	be hospitable	90	6
Держать камень за пазухой	derzhat' kamen' za pazukhoi	hold a stone inside one's shirt	43	19
Держать кого-то в ежовых рукавицах	derzhat v ezhovih rukavitsakh	keep or assume strict control over s.o.	35	28
Держать кого-то на прицеле	derzhat' na pritsele	keep in mind something or someone	62	12

Idiom (*V-C₁*)	Transliteration	Gloss	# Corpus 2	idioms/50
Держать кого-то под колпаком	derzhat' pod kolpakom	keep a check on smb	8	5
Держать марку	derzhat' marku	maintain one's reputation	64	46
Держать на почтительном расстоянии	derzhat' na pochtitelnom rasstoyanii	keep at a respectful distance	49	30
Держать на привязи	derzhat' na privyazi	restrict s.o. s independence, not allow s.o. to act on his own initiative	66	38
Держать нос по ветру	derzhat' nos po vetru	trim the sails to the wind	83	19
Держать пари	derzhat' pari	to bet on	250	50
Держать под каблуком	derzhat' pod kablukom	keep under thumb	3	2
Держать под своим крылом	derzhat' pod svoim krilom	have smb. under one's wing, take care of smb.	20	1
Держать порох сухим	derzhat' poroh sukhim	be prepared to defend one's country (cause etc) at any moment	7	7
Держать руки по швам	derzhat' ruki po shvam	stand to attention	59	48
Держать себя	derzhat' sebya	act in a certain manner	3633	20
Держать себя в рамках	derzhat' sebya v ramkakh	behave in a restrained, correct manner	20	15
Держать себя в руках	derzhat' sebya v rukakh	To hold/keep oneself in check; to keep oneself in hand.	7505	1
Держать себя в узде	derzhat' sebya v uzde	to refrain from emotional outbursts	95	47
Держать ухо востро	derzhat' ukho vostro	have an ear to the ground	283	29
Держать хвост трубой	derzhat' khvost truboi	keep cheerful, keep one's chin / spirits up	16	8
Держать шаг	derzhat' shag	march in step	22	14
Держать язык за зубами	derzhat' yazik za zubami	keep one's tongue between one's teeth	216	42
Держать язык на привязи	derzhat' yazik na privyazi	to remain silent	66	9
Держаться за юбку	derzhatsya za yubku	be completely under the control of, and fully dependent on, some woman (one's wife, mother etc)	8	0
Идти прахом	idti prakhom	go to rack and ruin	26	25
Идти прямой дорогой	idti pryamoi dorogoi	take the shortest route (to)	1624	1
Идти своей дорогой	idti svoei dorogoi	go one's own way	1624	3

Idiom $(V-C_l)$	Transliteration	Gloss	# Corpus 2	idioms/50
Идти против течения	idti protiv techeniya	to go against the tide/current/stream	187	22
Идти в разрез	idti v razrez	act in defiance to someone or something	66	48
Идти в ногу	idti v nogu	to progress at a pace equal to that of (another person, branch of technology etc, or the times in general)	584	22
Идти впрок	idti vprok	to be beneficial to s.o.	50	50
Идти врозь	idti vroz'	go apart	14	14
Идти (кому-то) навстречу	idti navstrechu	to satisfy s.o.'s needs, requests, desires	1880	20
Идти насмарку	idti nasmarku	to yield no positive result, end in nothing	30	29
Идти в гору	idti v goru	to improve one's status or job, gain influence, importance, succeed in one's career	574	12
Идти в дело	idti v delo	to be or start to be used	1211	8
Идти ко дну	idti ko dnu	be almost ruined	196	15
Идти на лад	idti na lad	take a turn for the better	81	49
Идти на поводу	idti na povodu	to submit fully to s.o., not act on one's own initiative	174	41
Идти на попятную	idti na popyatnuyu	to retreat from a decision made earlier, go back on an agreement	20	18
Идти по линии наименьшего сопротивления	idti po linii naimenshogo soprotivleniya	to choose the easiest course of action, avoiding difficulties, trouble	334	1
Идти по миру	idti po miru	to lead a beggarly life	288	4
Идти под венец	idti pod venets	to marry	40	36
Идти под гору	idti pod goru	to deteriorate sharply	574	1
Не идти на ум	ne idti na um	the thought of sth. does not occupy s.o.'s mind	102	25
Идти на удочку	idti na udochku	to end up being deceived, duped	8	4
Идти по стопам	idti po stopam	to follow after s.o. in doing sth., looking to him as an example	147	42
Играть первую скрипку	igrat' pervuyu skripku	to occupy the leading, predominant position in sth	449	9
Играть глазами	igrat' glazami	to glance at s.o. in a flirtatious manner, trying to gain his or her interest	139	15
Играть жизнью и смертью	igrat' zhizn'yu I smert'yu	to put o.s. in a life-threatening situation, disregarding danger	238	1

11

Idiom (V-C_l)	Transliteration	Gloss	# Corpus 2	idioms/50
Играть роль	igrat' rol'	to have a certain (as specified) meaning, significance, influence	6704	47
Играть словами	igrat' slovami	to speak evasively, using imprecise, ambiguous language	173	20
Играть на руку	igrat' na ruku	help s.o. or further sth. by one's actions, often without being aware of doing so	264	21
Играть в бирюльки	igrat' v birul'ki	to occupy o.s. with trifles	9	9
Играть в загадки	igrat' v zagadki	to speak evasively, enigmatically, in a roundabout way	7	4
Играть в молчанку	igrat' v molchanku	to keep silent, avoid conversation	17	16
Играть в прятки	igrat' v pryatki	to hide the truth by being evasive	127	18
Играть в струнах души	igrat' v strunakh dushi	to touch smb's feelings	43	5
Играть с огнём	igrat' s ognem	To play with edge-tools; to play with fire.	12	1
Играть на нервах	igrat' na nervakh	to irritate s.o	30	23
Бить баклуши	bit' baklushi	to be idle, do nothing	61	61
Бить копытом	bit' kopitom	be hot to trot	142	5
Бить тревогу	bit' trevogu	to draw attention to impending danger, trouble etc, calling for measures to suppress or fight it	196	21
Бить фонтаном	bit' fontanom	emerge rapidly	167	14
Бить наповал	bit' napoval	act in a fashion that guarantees success, rules out any possibility of failure	27	12
Бить по глазам	bit' po glazam	strike the eye	361	41
Бить по карману	bit' po karmanu	cost a pretty penny	24	14
Бить кого-то смертным боем	bit' smertnim boem	beat smb. ruthlessly	59	12
Бить в набат	bit' v nabat	to draw general attention persistently to sth. alarming, to impending danger	86	31
Бить в цель	bit' v tsel'	to be effective, attain the desired result	58	34
Бить через край	bit' cherez krai	to manifest itself, come forth with great force, in abundance	86	50
Волком смотреть	volkom smotret'	look angrily, in a hostile way	51	25
Смотреть бирюком	smotret' birukom	to look gloomy, morose	2	2

Idiom (V-C_I)	Transliteration	Gloss	# Corpus 2	idioms/50
Смотреть женихом	smotret' zhenihom	look very happy, content	26	0
Смотреть зверем	smotret' zverem	be gloomy, morose	45	6
Смотреть именинником	smotret' imeninnikom	look very happy	5	5
Смотреть правде в (глаза)	smotret' pravde v glaza	see things or evaluate facts as they really are, look at things clearly	120	34
Смотреть смерти в глаза	smotret' smerti v glaza	be exposed to mortal danger	89	46
Смотреть сквозь пальцы	smotret' skvoz' paltsi	intentionally not to react to sth., as if accepting it by allowing it to go on	296	29
Смотреть со своей колокольни	smotret' so svoei kolokol'ni	to make one-sided judgments about s.o. or sth. based entirely on one's own limited perspective	18	11
Класть голову	klast' golovu	sacrifice one's life for smb., smth.	260	3
Класть оружие	klast' oruzhie	stop fighting for sth., abandon some course of action, admit defeat	28	3
Класть деньги (в кубышку, на бочку)	klast' den'gi (v kubishku, na bochku)	accumulate money (keeping it at home, not investing it)	78	0
Класть зубы на полку	klast' zubi na polku	go hungry; tighten one's belt	16	9
Класть душу	klast' dushu	To put one's heart and soul into something	21	7
Класть на бумагу	klast' na bumagu	to be expressed in written form	37	31
Класть на лопатки	klast' na lopatki	coll to win a victory over s.o. (in an argument, contest etc)	12	7
Класть на музыку	klast' na musiku	to write music for some verses or text	7	6
Класть под сукно	klast' pod sukno	to postpone making a decision on some matter	8	5
Лезть из кожи вон	lezt' iz kozhi von	To lean/bend over backwards; to go out of one's way; to go all out.	90	23
Лезть в глаза	lezt' v glaza	to attract attention to o.s. or itself, be noticeable	188	34
Лезть в голову	lezt' v golovu	creep into one's head	312	41
Лезть в душу	lezt' v dushu	(try to) worm oneself into smb's confidence	62	38
Лезть в бутылку	lezt' v butilku	To fly off the handle; to blow one's top.	25	18
Лежать бревном	lezhat' brevnom	lie like a log	87	46

13

Idiom (V-C$_l$)	Transliteration	Gloss	# Corpus 2	idioms/50
Лежать мертвым грузом	lezhat' mertvim gruzom	to go unused, be without application	36	16
Лежать пластом	lezhat' plastom	(usu. of a person who is sick or very tired) to lie completely immobile	154	44
Лежать на печи	lezhat' na pechi	to do nothing, be idle	103	27
Лежать на поверхности	lezhat' na poverhnosti	be obvious, evident	245	12
Лежать под сукном	lezhat' pod suknom	(of an application, request, complaint etc) not to be given any attention, not be processed	17	7
Выйти сухим из воды	viiti sukhim iz vodi	to escape well-deserved punishment, remain unpunished, uncompromised	364	23
Выйти в люди	viiti v ludi	make one's way (in life)	1475	45
Выйти из возраста	viiti iz vozrasta	be past the age (for)	82	33
Выйти из пеленок	viiti iz pelenok	be no longer a kid / baby	20	17
Выйти из-под кисти	viiti iz-pod kisti	be painted (by)	9	7
Выйти из-под пера	viiti iz-pod pera	come from smb.'s pen	132	26
Жить барином	zhit' barinom	live a life of ease	96	18
Жить бирюком	zhit' biryukom	to be unsociable, live in seclusion	6	4
Жить минутой	zhit' minutoi	live in the moment	108	4
Жить нараспашку	zhit' naraspashku	live openly and widely	10	9
Жить чужим (своим)умом	zhit' chuzhim (svoim)umom	let others think for oneself	139	24
Жить на вулкане	zhit' na vulkane	to be living on the edge of a volcano; to be sitting on a volcano.	27	21