*Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 3198–3206
Marseille, 11–16 May 2020


# Collocations in Russian Lexicography and Russian Collocations Database

**Maria Khokhlova**
St Petersburg State University
Universitetskaya emb., 11, 199034 St. Petersburg, Russia
m.khokhlova@spbu.ru

**Abstract**

The paper presents the issue of collocability and collocations in Russian and gives a survey of a wide range of dictionaries both printed and online ones that describe collocations. Our project deals with building a database that will include dictionary and statistical collocations. The former can be described in various lexicographic resources whereas the latter can be extracted automatically from corpora. Dictionaries differ among themselves, the information is given in various ways, making it hard for language learners and researchers to acquire data. A number of dictionaries were analyzed and processed to retrieve verified collocations, however the overlap between the lists of collocations extracted from them is still rather small. This fact indicates there is a need to create a unified resource which takes into account collocability and more examples. The proposed resource will also be useful for linguists and for studying Russian as a foreign language. The obtained results can be important for machine learning and for other NLP tasks, for instance, automatic clustering of word combinations and disambiguation.

**Keywords:** collocations, Russian dictionaries, lexical database

## 1. Introduction

The study of collocability has not lost its relevance over the past decades. The identification of lexical constructions and their further analysis are crucial for various issues in modern applied linguistics: creating dictionaries for sentiment analysis, search queries expansion, machine translation, language learning etc.

A powerful surge of interest to collecting and analyzing data about joint occurrence of lexical units can be explained by the increased role of corpus linguistics in recent years, in which the study of set expressions is associated both with the solution of applied problems and with the theoretical interpretation of the acquired material. Existing methods for collocation extraction cannot be considered perfect, because, firstly, they distinguish phrases of a different nature, and secondly, there is not enough data to evaluate them. The latter fact is the reason that there is a need in a variety of dictionaries and other resources which will reflect verified collocations for the subsequent verification of data obtained automatically. This was the motivation of our project on building a database of Russian collocations that will comprise both dictionary and empirical collocations, i.e. ones extracted from lexicographic resources and corpora (Khokhlova, 2018). The paper focuses on giving a survey of different dictionaries and online systems for Russian that can be used for compiling an integrated database.

A movement from printed dictionaries to electronic ones accessible via web, desktop or mobile applications was a basic trend of lexicography within the last decades. Russian dictionaries have a long tradition but when it comes to computational lexicography it is long overdue. It can be partially explained by the historical events of the Soviet and post-Soviet era that had so much influence on the selection of lexis and its definitions and also by the approach of Russian lexicographers to describing language that differs from the one of their Western colleagues. Russian language does not belong to less resourced ones but there is still a need in up-to-date tools and hence every effort in this direction should be welcomed. In this case we would strive to highlight different projects in order to draw attention to them and to make clear what our database can benefit from them.

The paper has the following structure. The Introduction explains the motivation of the project. Section 2 describes the notion of collocation in Russian linguistics and how it is interpreted in the database. The next section gives a deep overview of the printed explanatory and collocations dictionaries, part of them were used as a source for the database. Section 4 exemplifies online resources that show Russian collocations. Section 5 discusses the processing of the dictionaries for the database. The last section concludes the paper and proposes plans for future work.

## 2. Collocations in Russian

Our project deals with the process of building a database that will represent information on collocability in Russian from dictionaries and corpora (Khokhlova, 2018). When building a database we had to answer the following question: what kind of collocation should be considered as an appropriate item for the database. And here we need to analyse existing lexicographic resources.

Within our approach we will consider collocations extremely broadly that is motivated by the practical purpose of our project. Following Testelets (2001) we interpret collocability as the ability to connect with other lexical units. Thus, we will take into consideration phrases of different degrees of stability (from collocations to idioms and phraseological units with non-compositionality). For example, according to Teliya's (1996) classification: idioms (*rabochaja loshad'* 'working horse'), phraseological units (*teljachij vostorg* 'foolish enthusiasm'), fixed expressions (*vsego khoroshego* 'all the best'), cliche (*minutu vnimanija* 'minute of attention'). Also here we add terms (*kontrol'naja palata* 'control chamber') and set phrases (*podvergnut'sja deformatsii* 'undergo deformation').

It could be a certain problem to find an appropriate dictionary that describes collocations and is large enough in order to both list high frequency word combinations and give a sufficient number of examples for them. Strictly speaking, on the one hand there is no collocations dictionary for Russian that could have been compared to its Western couterparts (for example, Oxford Collocations Dictionary). On the other hand since the majority of Russian dictionaries (if can ever be found in digital form) represent scanned copies of printed versions without

OCR. Their recognition is then should be followed by further division of entries into a structured format that takes into account possible grammatical information, examples, quotations etc.

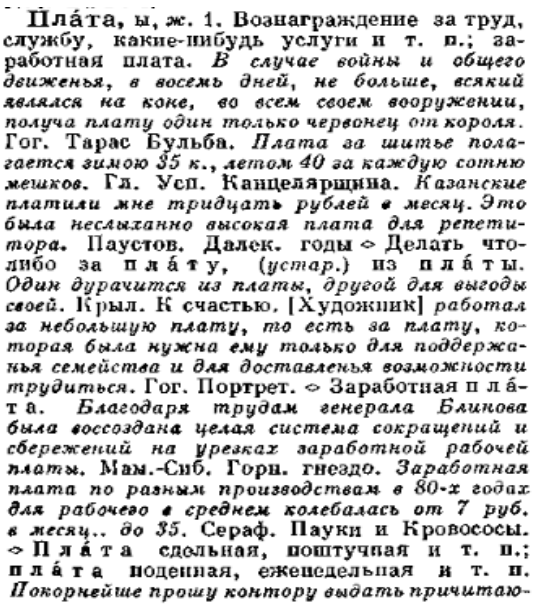
For our project we made a survey of various dictionaries that can be used as sources of collocations. In the present paper we will dwell on those of them that have been already processed during the project and also on those that can be found interesting for further work.

## 3. Printed Dictionaries

### 3.1 Russian Explanatory Dictionaries

Russian explanatory dictionaries play a significant role in Russian lexicography and studies being based on the results of fundamental work describing lexis. They can implement various approaches, i.e. set phrases, multiword expressions and collocations can be described not only in special sections of the entries but also in the examples, sayings and quotations. Below we will present three main dictionaries of the type, two of them, however, exist only in printed version.

The *Dictionary of Contemporary Literary Russian Language* (1948–1965) was one of the most important projects in Soviet lexicography started before the Second World War but the first volume appeared only in 1950. In total it comprises 17 volumes describing 120,480 words.

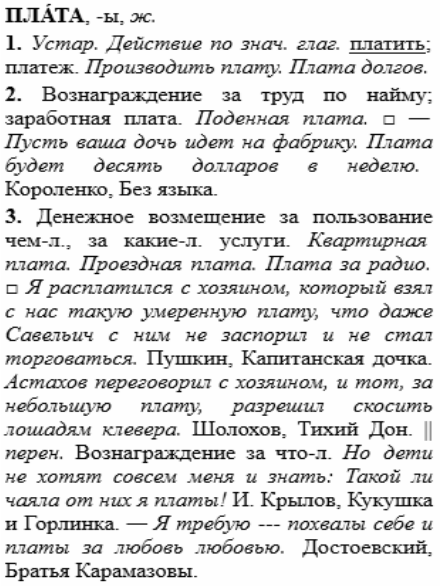


**Пла́та**, ы, ж. 1. Вознаграждение за труд, службу, какие-нибудь услуги и т. п.; заработная плата. *В случае войны и общего движенья, в восемь дней, не больше, всякий являлся на коне, во всем своем вооруженьи, получа плату один только червонец от короля.* Гог. Тарас Бульба. *Плата за шитье полагается зимою 35 к., летом 40 за каждую сотню мешков.* Гл. Усп. Канцелярщина. *Казанские платили мне тридцать рублей в месяц. Это была неслыханно высокая плата для репетитора.* Паустов. Далек. годы ◇ Делать что-либо за пла́ту, (*устар.*) из пла́ты. *Один дурачится из платы, другой для выгоды своей.* Крыл. К счастью, [Художник] *работал за небольшую плату, то есть за плату, которая была нужна ему только для поддержанья семейства и для доставленья возможности трудиться.* Гог. Портрет. ◇ Заработная пла́та. *Благодаря трудам генерала Блинова была воссоздана целая система сокращений и сбережений на урезках заработной рабочей платы.* Мам.-Сиб. Горн. гнездо. *Заработная плата по разным производствам в 80-х годах для рабочего в среднем колебалась от 7 руб. в месяц.. до 35.* Сераф. Пауки и Кровососы. ◇ Пла́та сдельная, поштучная и т. п.; пла́та поденная, еженедельная и т. п. *Покорнейше прошу контору выдать причитаю-*

Figure 1. Headword *plata* ‘fee’, part of the entry in the Dictionary of Contemporary Literary Russian Language (1948–1965).

Figure 1 demonstrates a part of the entry for the headword *plata* that corresponds to its first meaning. After the diamond symbol ‘◊’ there are examples of verbal and attributive collocations for the headword supplied with the citations, e.g. *zarabotnaya plata* ‘wage’, *plata sdel’naya* ‘accord loan’, *plata podennaya* ‘day rate payment’ etc.

The *Dictionary of the Russian Language* (1981–1984) followed the previously mentioned project and comprises more than 80,000 lexical items resulting in 4 volumes. Its first edition was published in 1957–1961, the revised version came out in 1981–1984 and soon became popular among linguists and other scholars. This explanatory dictionary is the only one existing as an online system that enabled its further processing for our project.



**ПЛА́ТА**, -ы, *ж.*

**1.** *Устар. Действие по знач. глаг.* платить; платеж. *Производить плату. Плата долгов.*

**2.** Вознаграждение за труд по найму; заработная плата. *Поденная плата.* □ — *Пусть ваша дочь идет на фабрику. Плата будет десять долларов в неделю.* Короленко, Без языка.

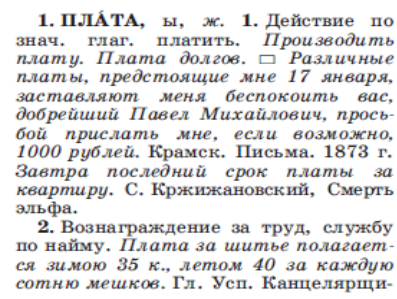
**3.** Денежное возмещение за пользование чем-л., за какие-л. услуги. *Квартирная плата. Проездная плата. Плата за радио.* □ *Я расплатился с хозяином, который взял с нас такую умеренную плату, что даже Савельич с ним не заспорил и не стал торговаться.* Пушкин, Капитанская дочка. *Астахов переговорил с хозяином, и тот, за небольшую плату, разрешил скосить лошадям клевера.* Шолохов, Тихий Дон. ‖ *перен.* Вознаграждение за что-л. *Но дети не хотят совсем меня и знать: Такой ли чаяла от них я платы!* И. Крылов, Кукушка и Горлинка. — *Я требую --- похвалы себе и платы за любовь любовью.* Достоевский, Братья Карамазовы.

Figure 2: Headword *plata* ‘fee’ in the Dictionary of the Russian Language (1981–1984).

Compared to the Dictionary of Contemporary Literary Russian Language (1948–1965) the given dictionary is more concise and we can see it on the excerpt from its online version (Figure 2). Here collocations do not have a special mark-up and they are merely given as examples in italics. E.g. *proizvodit’ platu* ‘to pay the fee’, *kvartirnaya plata* ‘rent’, *proyezdnaya plata* ‘fare’.

The *Big Academic Dictionary of Russian* (2004–2019) is being compiled now. At the beginning at was aimed at 150,000 words and 25 volumes, however it seems to outperform initial plans (the 26th volume was published in 2019).

The diamond symbol ‘◊’ introduces the most typical collocates for the headwords, while the tilde symbol ‘~’ corresponds to the phraseological units (Figures 3 and 4). The dictionary shares much in common with its “ancestor” (Dictionary of Contemporary Literary Russian Language, 1948–1965), however, it is more comprehensive.



**1. ПЛА́ТА,** ы, *ж.* **1.** Действие по знач. глаг. платить. *Производить плату. Плата долгов.* □ *Различные платы, предстоящие мне 17 января, заставляют меня беспокоить вас, добрейший Павел Михайлович, просьбой прислать мне, если возможно, 1000 рублей.* Крамск. Письма. 1873 г. *Завтра последний срок платы за квартиру.* С. Кржижановский, Смерть эльфа.

**2.** Вознаграждение за труд, службу по найму. *Плата за шитье полагается зимою 35 к., летом 40 за каждую сотню мешков.* Гл. Усп. Канцелярщи-

Figure 3: Headword *plata* ‘fee’, part of the entry in the Big Academic Dictionary of Russian (2004–2019).

Explanatory dictionaries can be used as sources of information on collocability only to a certain degree as they do not focus on collocations in their entries and hence the amount of such data is still small and leaves much to be desired. The majority of phrases extracted from the dictionaries are phraseological units.

*своей.* Крыл. К счастью. [Художник] *работал за небольшую плату, то есть за плату, которая была нужна ему только для поддержания семейства и для доставленья возможности трудиться.* Гог. Портрет. ◇ Заработная пла́та. *Заработная плата по разным производствам в 80-х годах для рабочего в среднем колебалась от 7 руб. в месяц.. до 35.* Сераф. Пауки и Кровососы. *Размеры заработной платы в Японии зависят от числа лет, проработанных на данном предприятии.* Ю. Семенов, На «козле» за волком. ◇ Пла́та сдельная, поштучная и т. п.; пла́та поденная, еженедельная и т. п. — *Они будут отделывать набережную: каждый месяц я буду сам рассчитывать и, кроме задельной платы, пойдет еще сумма на улучшение пищи.* Писем. Тысяча душ. *Повар получал от некоторых учеников еженедельную плату за то, что кормил их утром и вечером кашею.* Помял. Оч. бурсы. *Гейша вовсе не непременно продажная женщина..; скорее всего это артистка, которую приглашают за известную часовую плату для развлечения и удовольствия художественного.* Овчинник. Ветка сакуры.

Figure 4: Headword *plata* ‘fee’, part of the entry in the Big Academic Dictionary of Russian (2004–2019).

### 3.2 Russian Collocations Dictionaries

Language learners and teachers are usually the main target audience of collocation dictionaries. Below we will discuss Russian dictionaries that represent collocability and can be to a certain degree thus called collocations dictionaries.

*Set Verb-Noun Phrases in Russian* (Deribas, 1983) is a dictionary intended for students of Russian and in total comprises 5,197 collocations for 744 verbs and 1,345 nouns. The authors put collocations between free phrases and phraseological units: *opravdyvat’ (opravdat’) ozhidaniya* ‘to confirm expectations’; *pitat’ uvazheniye* ‘to respect’; *chitat’ lektsiyu* ‘to hold a lecture’ etc. The majority of phrases consist of bigrams including verbs and nouns as direct or indirect objects. The authors emphasize the dictionary is focused on language learners and does not provide any definitions or explanations merely listing the collocations of literary language. There are two lists sorted alphabetically according to verbs and nouns. The verbal entry (Figure 5) includes the infinitive of main imperfective form and if it has the corresponding perfective form gives it in bold italics in parentheses. The collocates are used in their word forms followed by the interrogative pronouns (with or without prepositions) and cases that indicate possible distributions of collocations (a kind of a valency frame).

**принимать *(принять)*** **ультиматум** кого (Р), чей, какой
**принимать *(принять)*** **условия** кого-чего (Р), чьи, какие
**принимать *(принять)*** **устав** чего (Р), какой
* **принимать *(принять)*** **участие** в чём (П); в ком (П), какое
**принимать *(принять)*** **форму** чего (Р), какую
**принимать *(принять)*** **характер** чего (Р), какой
**принимать *(принять)*** **экзамен** у кого (Р), по чему (Д), какой
**принимать *(принять)*** **эстафету** у кого (Р), от кого (Р), какую
**приниматься *(приняться)*** **за дело** когда, за какое

Figure 5: Headword *prinimat’* ‘take’ in (Debibas, 1983).

The nominal entry (Figure 6) lists all the collocations that include the noun. Verbs with possible close or opposite meanings are marked as synonyms and antonyms and are given in parentheses. As we can see the nominal part of the dictionary was not as elaborated as the verbal one.

**победа** добиваться победы, завоёвывать победу, закреплять победу, ковать победу, одерживать победу *(син.* ***побеждать)***, праздновать победу, предвкушать победу, приводить к победе, приносить победу, присуждать победу, упускать победу
**победитель** выходить победителем *(син.* ***побеждать)***
**поблажка** делать поблажку

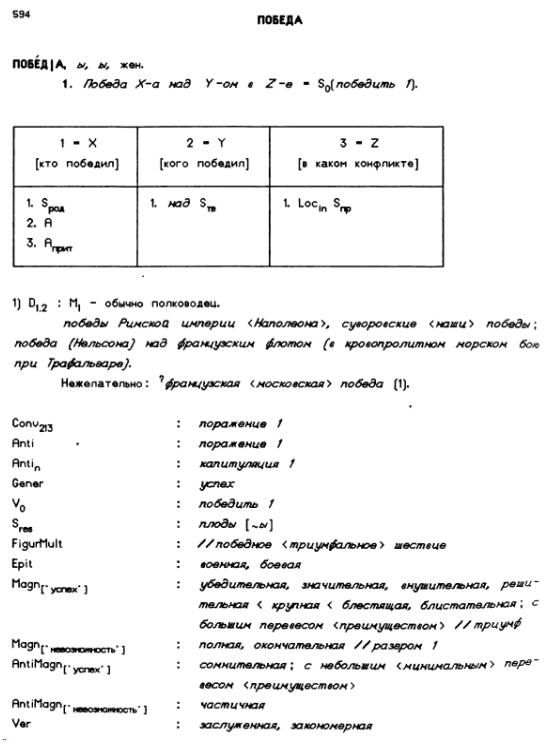
Figure 6: Headwords *pobeda* ‘victory’, *pobeditel’* ‘winner’, *poblazhka* ‘indulgence’ in (Deribas, 1983).

The *Explanatory Combinatorial Dictionary of Modern Russian* (Mel’čuk and Zholkovsky, 1984) came out in Vienna in 1984 and follows a unique approach to formal description of collocability. It was developed as one of the obligatory components to be used for the implementation of the “Meaning ↔ Text” (Mel’chuk, 1974) model. The printed edition has a limited vocabulary that counts about 250 headwords. The formal description method developed within the “Meaning ↔ Text” model allows for presenting information in a unified form (in particular, using lexical functions). Lexical function is a core notion of the model associating a word (or an argument) with a set of words and phrases expressing the meaning or role which correspond to the function. These lexical functions helped to make a formal description of phrases and their meanings. The most famous function Magn can be translated as “very” or the highest degree of something, e.g. Magn (*dozhd’* ‘rain’) = *prolivnoj* ‘heavy’, *livnevyj* ‘torrential’, *liven’* ‘shower’. Ten zones are distinguished inside a dictionary entry: morphological information, stylistic labels, definition (by constants and variables), a government pattern which uses variables from the meaning, restrictions on the government pattern, examples of the government pattern, lexical functions (several dozen of them are entered), illustrations, encyclopaedic information and idioms (Figure 7). Opposed to other projects this dictionary was made by linguists and for linguists and had to be used for automatic text processing. Despite its small volume the dictionary is very different from modern explanatory and collocations dictionaries and is a unique work, the implementation of one of the complex linguistic theories in the lexicographic work.

Figure 7 exemplifies lexical functions applied to the headword *pobeda* ‘victory’. Here we find the following values of Magn function: *ubeditel’naya* ‘convincing’, *znachitel’naya* ‘superb’, *vnushitel’naya* ‘tremendous’, *razitel’naya* ‘notable’, *krupnaya* ‘significant’, *blestyascaya* ‘overwhelming’ etc.

The *Dictionary of Russian Collocations with English-Russian Dictionary of Keywords* (Borisova, 1995) was devoted to Russian collocations and the first one to use this notion. Students and teachers of Russian were the main audience of the reference book. In the dictionary collocations are structured according to their semantics with numbers that correspond to lexical functions and are represented graphically in capital letters (Figure 8). For example, number 8 in the dictionary entry denotes the above-mentioned Magn function (*ubeditel’naya pobeda*‘convincing victory’). Nouns are the most frequent headwords in the dictionary followed by verbs, adjectives and phrases. The dictionary is rather concise; its word list counts only 512 items.

The *Set Phrases in Russian: Reference Book for Foreign Students* (Reginina et al., 1980) comprises 3,000 collocations that are typical for journalistic, colloquial and scientific text types.



594 ПОБЕДА

**ПОБЕ́Д|А**, ы, ы, жен.

1. *Победа X-а над Y-ом в Z-е* = $S_0$(*победить 1*).

| 1 = X [кто победил] | 2 = Y [кого победил] | 3 = Z [в каком конфликте] |
|---|---|---|
| 1. $S_{род}$<br>2. A<br>3. $A_{прит}$ | 1. *над* $S_{тв}$ | 1. $Loc_{in}$ $S_{пр}$ |

1) $D_{1,2}$ : $M_1$ – обычно полководец.

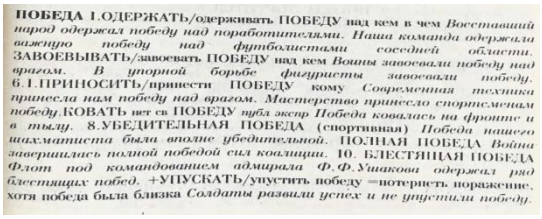
*победы Римской империи* ‹*Наполеона*›, *суворовские* ‹*наши*› *победы*; *победа (Нельсона) над французским флотом (в кровопролитном морском бою при Трафальгаре).*

Нежелательно: ?*французская* ‹*московская*› *победа* (1).

| | | |
|---|---|---|
| $Conv_{213}$ | : | *поражение 1* |
| Anti | : | *поражение 1* |
| $Anti_n$ | : | *капитуляция 1* |
| Gener | : | *успех* |
| $V_0$ | : | *победить 1* |
| $S_{res}$ | : | *плоды* [*~ы*] |
| FigurMult | : | *// победное* ‹*триумфальное*› *шествие* |
| Epit | : | *военная, боевая* |
| $Magn_{[\text{'успех'}]}$ | : | *убедительная, значительная, внушительная, решительная* ‹ *крупная* ‹ *блестящая, блистательная*; *с большим перевесом* ‹*преимуществом*› *// триумф* |
| $Magn_{[\text{'невозможность'}]}$ | : | *полная, окончательная // разгром 1* |
| $AntiMagn_{[\text{'успех'}]}$ | : | *сомнительная*; *с небольшим* ‹*минимальным*› *перевесом* ‹*преимуществом*› |
| $AntiMagn_{[\text{'невозможность'}]}$ | : | *частичная* |
| Ver | : | *заслуженная, закономерная* |

Figure 7: Headword *pobeda* ‘victory’, part of the entry in (Mel’čuk, Zholkovskiy, 1984).

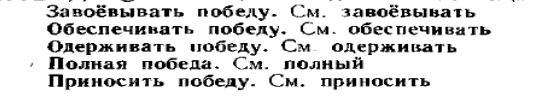
The selection of the lexis to be described was limited to the topics approved in the official teaching plans for Russian as a second language in the Soviet Union.



**ПОБЕДА** 1.ОДЕРЖАТЬ/одерживать ПОБЕДУ над кем в чем *Восставший народ одержал победу над поработителями. Наша команда одержала важную победу над футболистами соседней области.* ЗАВОЕВЫВАТЬ/завоевать ПОБЕДУ над кем *Воины завоевали победу над врагом. В упорной борьбе фигуристы завоевали победу.* 6.1.ПРИНОСИТЬ/принести ПОБЕДУ кому *Современная техника принесла нам победу над врагом. Мастерство принесло спортсмену победу.*КОВАТЬ нет св ПОБЕДУ *публ экспр Победа ковалась на фронте и в тылу.* 8.УБЕДИТЕЛЬНАЯ ПОБЕДА (спортивная) *Победа нашего шахматиста была вполне убедительной.* ПОЛНАЯ ПОБЕДА *Война завершилась полной победой сил коалиции.* 10. БЛЕСТЯЩАЯ ПОБЕДА *Флот под командованием адмирала Ф.Ф.Ушакова одержал ряд блестящих побед.* +УПУСКАТЬ/упустить победу =потерпеть поражение, хотя победа была близка *Солдаты развили успех и не упустили победу.*

Figure 8: Headword *pobeda* ‘victory’ in (Borisova, 1995).

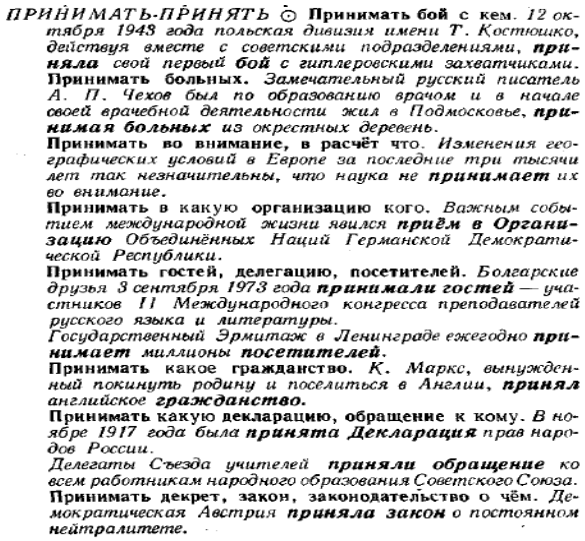
The authors presented only high frequency collocations and thus the entries seem to be quite concise (Figure 9). For example, the headword *pobeda* ‘victory’ has only two attributive collocates that are linked to other entries.



***ПОБЕДА*** ⊙ **Блестящая победа.** См. **блестящий**
**Завоёвывать победу.** См. **завоёвывать**
**Обеспечивать победу.** См. **обеспечивать**
**Одерживать победу.** См. **одерживать**
**Полная победа.** См. **полный**
**Приносить победу.** См. **приносить**

Figure 9: Headword *pobeda* ‘victory’ in (Reginina et al., 1980).

The compilation of the dictionary was affected by the historical period and the events that took place. The language of the official newspapers used as corpus data influenced the citations, hence we find many examples that mention *partija* “party”, *pjatiletka* “five-year plan”, *kommunist* “communist”, *terpet’ lishenija* “to suffer hardship”, *revoljutsionnyj narod* “revolutionary nation” etc. Verbal collocations gained more attention from the authors and were described in detail (Figure 10).



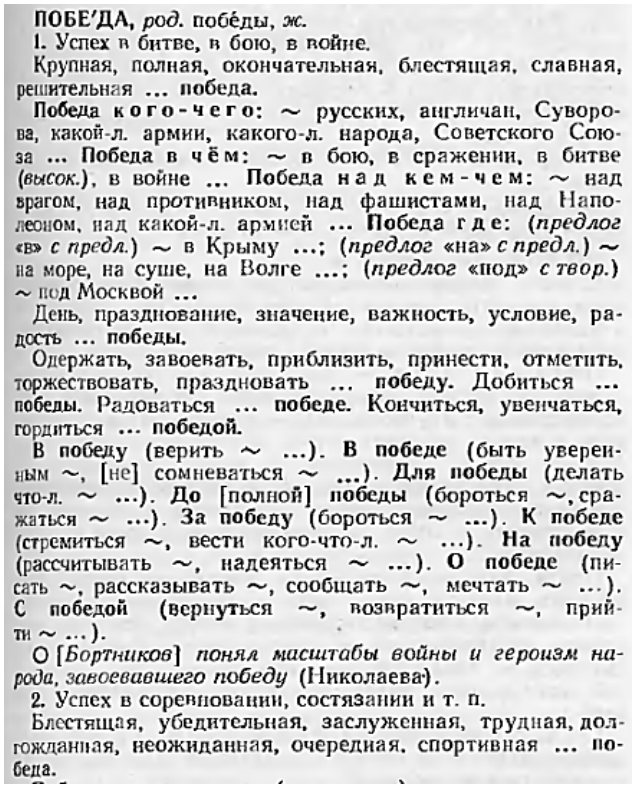
***ПРИНИ́МАТЬ-ПРИ́НЯТЬ*** ⊙ **Принимать бой с кем.** *12 октября 1943 года польская дивизия имени Т. Костюшко, действуя вместе с советскими подразделениями,* ***приняла*** *свой первый* ***бой*** *с гитлеровскими захватчиками.*
**Принимать больных.** *Замечательный русский писатель А. П. Чехов был по образованию врачом и в начале своей врачебной деятельности жил в Подмосковье,* ***принимая больных*** *из окрестных деревень.*
**Принимать во внимание, в расчёт что.** *Изменения географических условий в Европе за последние три тысячи лет так незначительны, что наука не* ***принимает*** *их во внимание.*
**Принимать в какую организацию кого.** *Важным событием международной жизни явился* ***приём в Организацию*** *Объединённых Наций Германской Демократической Республики.*
**Принимать гостей, делегацию, посетителей.** *Болгарские друзья 3 сентября 1973 года* ***принимали гостей*** *— участников 11 Международного конгресса преподавателей русского языка и литературы.*
*Государственный Эрмитаж в Ленинграде ежегодно* ***принимает*** *миллионы* ***посетителей.***
**Принимать какое гражданство.** *К. Маркс, вынужденный покинуть родину и поселиться в Англии,* ***принял*** *английское* ***гражданство.***
**Принимать какую декларацию, обращение к кому.** *В ноябре 1917 года была* ***принята Декларация*** *прав народов России.*
*Делегаты Съезда учителей* ***приняли обращение*** *ко всем работникам народного образования Советского Союза.*
**Принимать декрет, закон, законодательство о чём.** *Демократическая Австрия* ***приняла закон*** *о постоянном нейтралитете.*

Figure 10: Headword *prinimat’* ‘take’, part of the entry in (Reginina et al., 1980).

The *Dictionary of the Collocability of the Words of the Russian Language* (Denisov, Morkovkin, 1983) aims at teachers of Russian and philologists and presents 2,500 entries for nouns, verbs and adjectives. It is the most famous and comprehensive collocations dictionary of Russian. The authors distinguish between lexical and semantic collocability (according to Yu. D. Apresyan) and also defines syntactic collocability as a set of semantic and syntactic positions available for a word, i.e. its valency frame. The main task of the collocations dictionary is to identify these semantic and syntactic positions for each word and to describe their filling. Probably the structure of its entries is the most similar to the one we find in collocations dictionaries published in Europe or USA. The basic unit of the dictionary, thus, is the phrase, i.e. the representation of the valencies of a keyword. It can be described in three ways: 1) complete lists of words that fill a given valency (*katat’sja na kon’kakh, na lyzhakh* ‘run on skates, on skis’); 2) a selective enumeration of words that are typical for a given position (*nachalo chego: sorevnovanij, predstavlenija* ‘the beginning of what: competitions, performance’ ...); 3) an indication of the lexical characteristics and enumeration of the most typical words (*fotografirovat’ kogo: (o cheloveke) druga, syna, doch, pamjatnik* ‘to photograph someone: (about a person) a friend, son, daughter, monument’ ...).

The dictionary entry (Figure 11) however do not distinguish between its parts with special labels (except for the circle ‘○’ in this example that is used for illustrative sentences) merely listing collocates that belong to different parts-of-speech in separate paragraphs and using numbers for different meanings. Here we find the following adjectives and verbs collocating with the headword: 1) *krupnaya* ‘significant’, *polnaya* ‘complete’, *okonchatel’naya* ‘final-round’ etc; 2) *oderzhat’* ‘to win’, *zavoevat’* ‘to gain’, *priblizit’* ‘to bring’ etc.

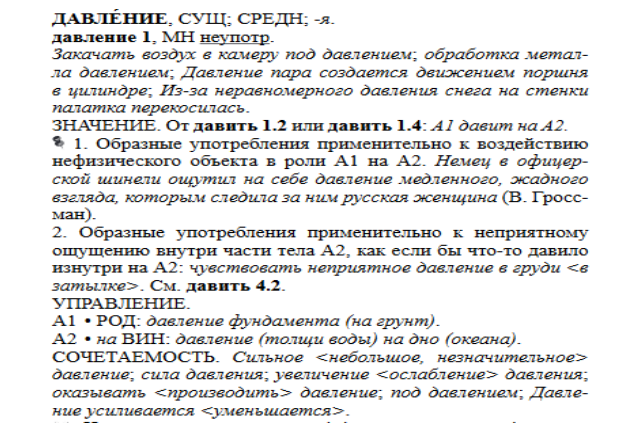
The boundary between collocations and other non-free word combinations is quite vague and thus lexical collocations can be described in the dictionaries of other types.



**ПОБЕ́ДА**, *род.* побе́ды, *ж.*
1. Успех в битве, в бою, в войне.
Крупная, полная, окончательная, блестящая, славная, решительная ... **победа.**
**Победа кого-чего:** ~ русских, англичан, Суворова, какой-л. армии, какого-л. народа, Советского Союза ... **Победа в чём:** ~ в бою, в сражении, в битве (*высок.*), в войне ... **Победа над кем-чем:** ~ над врагом, над противником, над фашистами, над Наполеоном, над какой-л. армией ... **Победа где:** (*предлог* «в» *с предл.*) ~ в Крыму ...; (*предлог* «на» *с предл.*) ~ на море, на суше, на Волге ...; (*предлог* «под» *с твор.*) ~ под Москвой ...
День, празднование, значение, важность, условие, радость ... **победы.**
Одержать, завоевать, приблизить, принести, отметить, торжествовать, праздновать ... **победу.** Добиться ... **победы.** Радоваться ... **победе.** Кончиться, увенчаться, гордиться ... **победой.**
**В победу** (верить ~ ...). **В победе** (быть уверенным ~, [не] сомневаться ~ ...). **Для победы** (делать что-л. ~ ...). **До** [полной] **победы** (бороться ~, сражаться ~ ...). **За победу** (бороться ~ ...). **К победе** (стремиться ~, вести кого-что-л. ~ ...). **На победу** (рассчитывать ~, надеяться ~ ...). **О победе** (писать ~, рассказывать ~, сообщать ~, мечтать ~ ...). **С победой** (вернуться ~, возвратиться ~, прийти ~ ...).
О [*Бортников*] *понял масштабы войны и героизм народа, завоевавшего победу* (Николаева).
2. Успех в соревновании, состязании и т. п.
Блестящая, убедительная, заслуженная, трудная, долгожданная, неожиданная, очередная, спортивная ... **победа.**

Figure 11: Headword *pobeda* 'victory', part of the entry in (Denisov, Morkovkin, 1983).

And here we can name a unique lexicographic project under the direction of Yu. D. Apresyan, i.e. *Active Dictionary of the Russian Language* (Apresyan, 2014-2017) which includes extensive information on collocability. Since 2014 the authors have published 3 volumes and continue the compilation that makes this dictionary one of the most ambitious projects of the last decades. Each zone is labelled separately in the dictionary entries, e.g. meaning, collocability, government model, synonyms etc. Figure 12 presents a part of the dictionary entry for the headword *davleniye* 'pressure'.



**ДАВЛЕ́НИЕ**, СУЩ; СРЕДН; *-я.*
**давление 1**, МН неупотр.
*Закачать воздух в камеру под давлением; обработка металла давлением; Давление пара создается движением поршня в цилиндре; Из-за неравномерного давления снега на стенки палатка перекосилась.*
ЗНАЧЕНИЕ. От **давить 1.2** или **давить 1.4**: *A1 давит на A2.*
1. Образные употребления применительно к воздействию нефизического объекта в роли A1 на A2. *Немец в офицерской шинели ощутил на себе давление медленного, жадного взгляда, которым следила за ним русская женщина* (В. Гроссман).
2. Образные употребления применительно к неприятному ощущению внутри части тела A2, как если бы что-то давило изнутри на A2: *чувствовать неприятное давление в груди <в затылке>*. См. **давить 4.2**.
УПРАВЛЕНИЕ.
A1 • РОД: *давление фундамента (на грунт).*
A2 • *на* ВИН: *давление (толщи воды) на дно (океана).*
СОЧЕТАЕМОСТЬ. *Сильное <небольшое, незначительное> давление; сила давления; увеличение <ослабление> давления; оказывать <производить> давление; под давлением; Давление усиливается <уменьшается>.*

Figure 12: Headword *davleniye* 'pressure' in (Apresyan, 2014–2017).

The data is well-structured and includes information about syntactic actants, collocations and constructions that are given explicitly in separate paragraphs. The authors describe valency frames of the lexical items and restrictions that can be put on actants. Collocations lists are quite long and include possible variants and examples as well.

## 4. Online Russian Dictionaries and Databases

At the moment, to the best of our knowledge there are no online dictionaries available for Russian that would have traditional lexicographic structure (headword, grammatical characteristics, senses, citations and quotes, collocations, phraseological units etc). At the same time, there are a number of unique and valuable lexicographic projects that describe collocational nature and valencies of lexical units, although in different ways. And thus we can speak about so called dictionaries that are presented in form of web sites (even though they are called "dictionaries"). They represent themselves lexicographic resources of a new type being not dictionaries in its proper sense but advanced online systems.

The Russian National Corpus (RNC, 2003–2019) is an excellent source of data and was used for building a number of dictionaries.

Here we can name the *Dictionary of Russian Abstract Nouns' Verbal Collocability* (Biryuk et al., 2008). According to its title the dictionary focuses on nouns and presents information for over 10,000 phrases of the following structures: 1) noun+verb; 2) verb+noun; 3) verb+adjective+noun. The authors use the notion of a lexical function (Mel'čuk and Zholkovsky, 1984) for describing and classifying collocations and their senses. The nouns were extracted from the syntactically parsed subcorpus of RNC and occupied one of the syntactic positions: 1) direct object of a transitive verb; 2) indirect object of a transitive verb; 3) subject of an intransitive verb. A user can search by part-of-speech (nouns, adjectives and verbs), sense, syntactic relation (object, indirect object, passive structure), negation etc.

The *Dictionary of Russian Idiomatic Expressions* (Kustova, 2008) presents information about 10,000 high frequency intensifiers found in RNC. Such linguistic units are characterized by restricted collocability, hence they should be learned by non-native speakers. The initial word list was based on RNC and printed dictionaries and represented by the examples from RNC. One can find phraseological units (*kruglyj sirota* 'orphan'), collocations (*plakat' navzryd* 'to sob violently'), idiomatic expressions (*gluboko blagodarny* 'deeply grateful') and semantically motivated free phrases (*chrezvychajno malen'kiy* 'extremely small).

The *FrameBank* database is the Russian prototype of FrameNet (Baker et al., 1998), being an online open resource (Lyashevskaya, 2010). It includes descriptions of valency frames for 2,200 verbs and constructions and has features of both a dictionary and a corpus. Lexical constructions are represented in form of patterns and list sematic roles of the participants and collocates from RNC.

The *Collocations, Colligations, Constructions* project was initially focused on the extraction of bigrams from Russian corpora (Kopotev et al., 2015). The authors used statistical measures in order to find the best examples of collocations suitable for language learners. Now the database provides information about collocations on the basis of RNC, Taiga and the ruWac corpora. The results are ranked by statistical values that enable a user to understand the significance of a collocation (e.g. whether it should be learned or is a merely free phrase).

## 5. Russian Collocations Database

The database includes two kinds of collocations, i.e. dictionary and statistical ones. The former present in various lexicographic resources whereas the latter can be extracted automatically from text corpora.

A number of dictionaries mentioned in the previous section were used as a source for collocations of the first type and were limited to: Dictionary of the Russian Language (1981–1984), Dictionary of Collocations (Borisova, 1995), Dictionary of Russian Abstract Nouns' Verbal Collocability (Biryuk et al., 2008), Dictionary of Russian Idiomatic Expressions (Kustova, 2008) and Dictionary of Set Verb-Noun Phrases in Russian (Deribas, 1983). Based on this dataset we created a prototype of a gold standard for collocability, i.e. dictionary collocations. Statistical collocations can be retrieved automatically from texts. In order to obtain data on co-occurrences in the Russian language, we process Araneum Russicum Maximum corpus (about 15 billion words), which was created automatically and is based on web texts of different genres being one of the largest collection of Russian texts (Benko, 2014). We use a statistical approach for automatic extraction of word combinations from corpora that implied several association measures (t-score, MI, log-likelihood).

Below we present the pipeline for data processing and discuss the principles of the database focusing on its part dealing with dictionary collocations (statistical collocations and their interpretation need to be described separately).

We examined dictionary entries and extracted collocation candidates either from phraseological sections or as separate items written with special fonts. The analysis of the dictionaries suggests that there is a need in a unified format that can be used for describing data. An entry has the following characteristics:

- a collocation;
- a syntactic model;
- a dictionary index (if applicable);
- references to the dictionaries (if applicable);
- frequencies in corpora (in ipm);
- values of the association measures;
- visualization.

The database includes information about 20,000 collocations. At the present stage of the project we processed in detail noun and verbal collocations focusing on the following models:

- adjective + noun.

Examples : *bolotnyj gaz* 'marsh gas', *gomericheskij khokhot* 'Homeric laughter', *zolotaja svad'ba* 'golden wedding', *zhivoj um* 'nimble mind'.

- verb + noun / verb + preposition + noun.

Examples : *bit' kartu* 'to cover a card', *nesti otvetstvennost'* 'to be responsible', *oblech doverijem* 'to trust', *stavit' tochku* 'to end'.

### 5.1 Extraction of Dictionary Collocations

We developed tools for extracting collocations from resources (with respect to the dictionaries that are not presented in the form of an electronic database), since the structure of dictionary entries is different and, accordingly, the collocation candidates are marked up in them in different ways (either by special fonts or symbols). Preprocessing involved also morphological analysis in order to present collocations in their canonical form but grammatical information was also preserved.

#### 5.1.1 Dictionary of the Russian Language

Altogether we extracted 11,210 phraseological units that were marked with a special diamond '◊' symbol; the total number of headwords was 5,955 (which is more than 7% of the total word list of the dictionary). The length of the extracted phrases varies from bigrams (*lomat' golovu* 'to puzzle') and trigrams *igrat' pervuju skripku* 'to play first fiddle') to 6 grams (*makovoj rosinki v rot ne brat'* 'starving'). Among the extracted phrases, the following models are presented: 1) adjective + noun (*morskaya milja* 'nautical mile'); 2) verb + noun / verb + preposition + noun (*boltat' jazykom* 'to jabber away'); 3) noun + noun (*kniga pocheta* 'book of honorable guests'); 4) preposition + noun (*bez umolku* 'nonstop'); 5) pronoun + noun (*nechego skazat'* 'nothing to say').

The phraseological units found in the dictionary often enumerate certain semantic groups and their lexical items collocate with a keyword. Examples: *v pylu (srazhen'ja, bitvy, spora etc)* 'heat of the (fight, battle, debate)', *v rassrochku (kupit', prodat')* 'in instalments (to buy, to sell)', *v storonu (skazat', proiznesti)* 'aside (to say, to utter)'. We considered each unit separately. Such an approach enabled us to enrich the data and also in future we plan to use special semantic tags to make the information clear for language learners.

#### 5.1.2 Dictionary of Russian Collocations

The initial preprocessing involved digitization of the dictionary and further OCR procedure. The extraction of the collocations was focused on the phrases written in capital letters. This resulted in a total sum of 3,058 collocation candidates. We also analyzed and extracted collocations from quotations that were not highlighted in the entries and marked them with the asterisk '*' symbol. These phrases vary in their fixedness but they can be an important source of information. Hence additional 232 lexical constructions enriched the database. Examples: *tesnoye sotrudnichestvo* 'close collaboration', *golovnaya bol'* 'headache' etc. Polysemic headwords marked in the dictionary with digits were preserved in the database as separate entries. Examples: *dobrozhelatel'nyy vzglyad 1* 'benevolent look' vs *original'nyy vzglyad 2* 'ingenious view'.

In the dictionary 2,044 verb pairs (imperfective and perfective aspects) are given via slash (e.g. *vnosit' /vnesti jasnost'* 'to clarify something'). They were separated in order to make two records resulting in total 4,088 verb phrases.

The dictionary lists a large number of collocations that slightly differ in their meaning and can be called synonyms to a certain degree. They vary either in wordforms (case or number) or in prepositions. Examples: *nakhodit'sja vo vlasti* vs *nakhodit'sja pod vlast'ju* 'to be in smb's power', *ostavit' pamjat' o sebe* vs *ostavit' pamjat' po sebe* 'to leave memories'.

#### 5.1.3 Dictionary of Russian Idiomatic Expressions

The preprocessing of the data extracted from the given electronic dictionary was not so elaborated as it was the case with other dictionaries. The following models were retrieved from the dictionary:

- adjective + noun (*redkostnyy talant'* 'exceptional talent') ;
- adverb + verb (*bezgranichno verit'* 'to trust implicitly') ;
- adverb + adverb (*sovsem nedavno* 'just recently') ;
- adverb + predicative (*iskljuchitel'no vazhno* 'exceptionally important');
- adverb + adjective (*gluboko porjadochnyj* 'totally honest') ;
- particle + noun (*prjamo chudo* 'really badly').

We confined ourselves to the first type of phrases and lemmatized them. The total amout was about 7,000 collocations.

#### 5.1.4 Dictionary of Russian Abstract Nouns' Verbal Collocability

The structure of the online dictionary enabled us to retrieve collocations straightforward. Altogether we extracted more than 8,000 verb phrases : *poluchit' vygodu* 'to get benefit', *chas probil* 'clock struck' etc. As next step we will analyze adjective + noun collocations that are embedded into longer ones, e.g. *khranit' glubokoye molchaniye* 'to keep absolute silence', *proizvesti blagopriyatnoye vpechatleniye* 'to create an impression' etc.

#### 5.1.5 Dictionary of Set Verb-Noun Phrases in Russian

The dictionary comprises more than 3,770 collocations with perfective and imperfective verb pairs and 383 with only imperfective verb forms. After processing the dictionary data, excluding prepositional phrases and representing phrases with perfective and imperfective verbal forms as different pairs we had a list of 7,923 items.

### 5.2 Dictionaries: Results

The volume of the verified (dictionary) collocations depends on the volume of the dictionaries that are used. As it was mentioned above dictionaries' volume is not sufficient enough to describe vaster groups of lexis and hence to give a broader coverage that could be comparable to a word list of an explanatory dictionary (it often counts several thousand). Altogether we extracted more than 35,000 collocation candidates from the above mentioned dictionaries, part of the gathered data overlapped. These collocations received the corresponding index, i.e. the number of the dictionaries they were presented in (so called dictionary index). It indicates the given items are highly reproducible in speech and can be used by language learners.

Table 1 demonstrates quantity of collocation candidates (noun and verb phrases) extracted from five dictionaries.

The central core of the merged lists is rather small and this can be explained by the following reasons. Firstly, we have processed a small number of lexicographic resources. Secondly, the dictionaries describe different lexis that hardly overlaps.

For example, the Dictionary of the Russian Language (1981–1984) aims at a comprehensive representation of the lexicon while other dictionaries focus on restricted collocability of certain semantic groups.

| Dictionary | Number of collocations | |
|---|---|---|
| | adj + noun | verb + noun |
| Dictionary of the Russian Language | 3,243 | 5,230 |
| Dictionary of Russian Collocations | 613 | 4,177 |
| Dictionary of Russian Abstract Nouns' Verbal Collocability | 0 | 8,140 |
| Dictionary of Russian Idiomatic Expressions | 6,962 | 0 |
| Dictionary of Set Verb-Noun Phrases in Russian | 0 | 7,923 |

Table 1: Statistics of collocations in the dictionaries

Table 2 shows the overlap between the dictionaries. As one can see overwhelming majority of collocations presents only in one dictionary and to a certain degree could be called unique.

| Number of dictionaries | Number of collocations |
|---|---|
| 1 | 22,090 |
| 2 | 1,981 |
| 3 | 41 |
| 4 | 10 |

Table 2: Statistics of collocations found in dictionaries

Four of the five examined dictionaries listed the same 10 phrases, for example: *vesti delo* 'to carry on business', *imet' tsel'* 'to have a goal', *imet' zadachu* 'to have a task', *stavit' tsel'* 'to set goal', *stavit' zadachu* 'to set a task', *stavit' usloviye* 'to set a condition', *zaronyat' podozreniye* 'to inspire suspicion', *podozreniye zakralos'* 'suspicion arises', *predstavlyat' interes* 'to be of interest'.

Result Grid  Filter Rows: Search

| w1 | p1 | w2 | p2 | coll_type | dicts |
|---|---|---|---|---|---|
| включать | VERB | свет | NOUN | 2 | der |
| выключать | VERB | свет | NOUN | 2 | der |
| свет | NOUN | белый | ADJF | 1 | mas |
| свет | NOUN | ближний | ADJF | 1 | mas |
| свет | NOUN | близкий | ADJF | 1 | mas |
| свет | NOUN | божий | ADJF | 1 | mas |
| свет | NOUN | большой | ADJF | 1 | mas |
| свет | NOUN | выгодный | ADJF | 1 | mas |
| свет | NOUN | дисперсия | ADJF | 1 | mas |
| свет | NOUN | дневной | ADJF | 1 | mas |
| свет | NOUN | ложный | ADJF | 1 | mas |
| свет | NOUN | необыкно... | ADJF | 1 | kust |
| свет | NOUN | новый | ADJF | 1 | mas |
| свет | NOUN | розовый | ADJF | 1 | mas |
| свет | NOUN | сильный | ADJF | 1 | kust |
| свет | NOUN | старый | ADJF | 1 | mas |

Figure 13: Headword *svet* 'light'.

Figure 13 presents results from the database for the headword *svet* 'light'. The columns *w1* and *w2* show the headword and its collocates, *p1* and *p2* indicate part-of-speech tags, *coll_type* shows the collocation type ('1' for adjective+noun collocations and '2' for verb+noun collocations), the last column *dicts* refers to the

dictionaries. For examples the collocations *blizhniy svet* 'near' or *bozhiy svet* 'world' can be found in the Dictionary of the Russian Language (1981–1984).

Below we show an example of an entry for the collocation *oderzhat' pobedu* 'to win' from the database:

*collocation:* oderzhat' pobedu;
*model:* VN;
*dictionary index:* 4;
*references to the dictionaries:* Dictionary of the Russian Language, Dictionary of Collocations, Dictionary of Russian Abstract Nouns' Verbal Collocability, Dictionary of Set Verb-Noun Phrases in Russian;
*frequency:* 2.59 ipm.

For the given collocation the dictionary index is quite high and this fact can indicate that this construction is important for language learners.

## 6. Conclusion

We made an overview of the dictionaries that describe collocations and of the database that includes data on collocability from various lexicographic resources. It is already available online (http://collocations.spbu.ru[1]).

The aim of the theoretical part of our work was to show the variety of dictionaries and differences of approaches to presenting collocations in entries. The low overlap between the dictionaries suggests that they describe different lexical units, e.g. free phrases, phraseological units or collocations with certain semantics. Hence we need to process other resources in future. There are also other syntactic models that should be taken into account as well as examples from quotations.

An open database of more than 20,000 Russian collocations can be used for evaluating the performance of various machine learning algorithms dealing with automatic text processing, as today there is no single system that includes sufficient amount of such information about collocations. The system will help researchers of the Russian language, and can also give valuable results in the process of dictionaries and grammars compiling. The system will be used by a wide range of users, which is not limited exclusively to specialists, e.g. it is interesting for users to independently study corpus examples and draw their own conclusions. The proposed system fully complies with this requirement and at the same time provides access to verified linguistic data.

It is planned to open free access to the system for a possible assessment by the users of the degree of stability of the phrases found in corpora. This will allow getting an interpretation by the speakers of the language and can be used in further improving the resource and other tasks, for example, when creating a specialized dictionary or in systems using machine learning. Also, the amount of illustrative material and information on collocability presented in lexicographic resources (and, accordingly, in the "gold standard") may indicate the frequency of the unit and correlate with it. This may be required when developing teaching and learning materials for students of the Russian language.

[1] For registration please contact the author.

## 7. Acknowledgements

This work was supported by the grant of the Russian Science Foundation (Project No. 19-78-00091).

## 8. Bibliographical References

Apresyan, Yu. D. (ed.) (2014-2017). Active Dictionary of the Russian Language [Aktivnyy slovar' russkogo yazyka]. Vol. 1-3. Yazyki slavyanskoy kul'tury, Moscow.

Baker, C. F., Fillmore, Ch. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *COLING-ACL'98: Proceedings of the Conference*, Montreal, Canada, pp. 86–90.

Benko, V. (2014). Aranea Yet Another Family of (Comparable) Web Corpora. *Text, Speech and Dialogue. 17th International Conference, TSD 2014*, Brno, Czech Republic, September 8-12, 2014. Proceedings. LNCS 8655. Springer International Publishing Switzerland, pp. 257-264.

Big Academic Dictionary of Russian [Bolshoy akademicheskiy slovar v 30 tomakh]. (2004–2016). Nauka, Moscow, St. Petersburg.

Biriuk O. L., Gusev V. Yu., and Kalinina E. Yu. (2008). Dictionary of Russian Abstract Nouns' Verbal Collocability. A Dictionary based on the Russian National Corpus [Slovar' Glagol'noj Sochetaemosti Nepredmetnykh Imen Russkogo Yazyka. Slovar' na osnove Natsional'nogo Korpusa Russkogo Yazyka], http://dict.ruslang.ru.

Borisova, E. G. (1995). A Word in a Text. A Dictionary of Russian Collocations with English-Russian Dictionary of Keywords [Slovo v tekste. Slovar' kollokatsiy (ustoychivykh sochetaniy) russkogo yazyka s anglo-russkim slovarem klyuchevykh slov]. Filologiya, Moscow.

Denisov, P. N. and Morkovkin, V. V. (1983). An Academic Collocation Dictionary of Russian [Uchebnyy slovar' sochetaemosti slov russkogo yazyka]. Russkiy yazyk, Moscow.

Deribas, V. M. (1983). Verb–Noun Collocations in Russian [Ustoychivye glagol'no-imennye slovosochetaniya russkogo yazyka]. Russkiy yazyk, Moscow.

Dictionary of Contemporary Literary Russian Language [Slovar' sovremennogo russkogo literaturnogo yazyka v 17 tomakh]. (1948–1965). Chernyshev, V.I. (ed.). Izd-vo Akademii nauk SSSR, Moscow, Leningrad.

Dictionary of the Russian Language [Slovar' russkogo yazyka v 4 tomakh]. (1981–1984). Yevgen'yeva, A. P. (ed.-in-chief). Vol. 1–4, 2nd edition, revised and supplemented. Russkiy yazyk, Moscow.

Khokhlova, M. (2018). Building a Gold Standard for a Russian Collocations Database. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana, pp. 863–869.

Kopotev, M., Escoter, L., Kormacheva, D., Pierce, M., Pivovarova, L., and Yangarber, R. (2015). CoCoCo: Online Extraction of Russian Multiword Expressions. *The 5th Workshop on Balto-Slavic Natural Language Processing (10–11 September 2015, Hissar, Bulgaria).* INCOMA Ltd, Sofia, pp. 43-45.

Kustova, G. I. (2008). Dictionary of Russian Idiomatic Expressions [Slovar' russkoyj idiomatiki. Sochetaniya

slov so znacheniyem vysokoy stepeni], http://dict.ruslang.ru.
Lyashevskaya, O. (2010). Bank of Russian Constructions and Valencies, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, pp. 1802–1805.
Mel'chuk, I. A. (1974). The experience of the theory of linguistic models "Meaning-Text" [Opyt teorii lingvisticheskikh modelej "Smysl-Tekst"]. Moscow.
Mel'čuk, I. and Zholkovsky, A. (1984). Explanatory Combinatorial Dictionary of Modern Russian [Tolkovo-kombinatornyy slovar russkogo yazyka]. Vienna.
Reginina, K. V., Tjurina, G. P., and Shirokova, L. I. (1980). Set Expressions of the Russian Language. A Reference Book for Foreign Students [Ustoychivye slovosochetaniya russkogo yazyka: Uchebnoye posobiye dlya studentov-inostrantsev]. Shirokova, L. I. (ed.). Moscow.
Teliya, V. N. (1996). Russian Phraseology: Semantic, Pragmatic and Cultural Aspects [Russkaya frazeologiya: semanticheskiy, pragmaticheskiy i lingvokul'torologicheskiy aspekty]. Moscow.
Testelets, Y. (2001). Introduction to the General Syntax [Vvedenie v obshchiy sintaksis]. RGGU, Moscow.

## 9. Language Resource References

RNC. (2003-2019). Russian National Corpus. URL: http://ruscorpora.ru.
Russian Collocations Database. (2019). URL: htpp://collocations.spbu.ru.