

ББК 81.1-923
УДК 811.111

А. В. Уланов,
*доктор филологических наук,
Сибирский институт бизнеса и
информационных технологий
ulanov.andrey.2014@mail.ru*

КОРПУСНЫЙ МЕТОД В ИССЛЕДОВАНИИ ЛЕКСИКИ РУССКОГО ЯЗЫКА

Статья посвящена особенностям применения корпусного подхода к исследованию лексики русского языка. Автор останавливается на важности применения корпусного подхода, позволяющего исследовать конкретные единицы русского языка с формально-семантической и количественно-статистической точек зрения. Приводятся мнения ученых на природу лингвистического, или языкового, корпуса, устанавливаются общие особенности корпусного подхода, которые терминологически определяются представителями научной лингвистической мысли. Особую роль в корпусной лингвистике современного русского языка играет исследование лексики русского языка с позиции словесной единицы / словосочетания; производной / непроизводной лексической единицы; словесной единицы в прямом значении / в переносном значении.

Ключевые слова: корпусная лингвистика, лингвистический корпус, корпусные методы, лексический фонд русского языка

Корпусная лингвистика является относительно новым разделом прикладной лингвистики. В центре внимания корпусной лингвистики – общие принципы и подходы к построению и изучению лингвистических корпусов, иначе в корпусной лингвистике называемых корпусами текста. Превосходством корпусной лингвистики является возможность применения корпусных методов в целях изучения одного лингвистического явления на материале массива (корпуса) языковых единиц. Развитию корпусной лингвистики способствовало интенсивное внедрение корпусных методов обработки текста, а также появление в языкознании компьютерных технологий, которые позволили отобрать большой массив языковых средств разных уровней, а уже впоследствии применить методы лингвистических исследований к отдельным таким единицам.

Корпусная лингвистика обеспечила не только возможность формально-семантического исследования конкретных единиц по различным принципам, но и возможность количественно-статистического исследования, позволившего сделать выводы о широте и объеме представленности языковых единиц в текстах конкретной стилистической природы.

Информационные технологии позволяют различать в корпусной лингвистике, с одной стороны, поиск и обработку ценных в языковом отношении контекстов, с другой стороны, обеспечивать экспериментальные исследования внутри данного корпуса.

Методы корпусной лингвистики активно используются в компьютеризированном языковом обучении [2], в общелингвистических исследованиях – в рамках изучения паремиологии [3]. Особую роль корпусная лингвистика играет в изучении иностранных языков [15].

В научных работах даются разные дефиниции корпуса текстов.

«Корпус текстов – это вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, определенные отрывки текстов данной проблемной области» (А. Н. Баранов) [1].

«Лингвистический или языковой, корпус – большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» (В. П. Захаров, С. Ю. Богданова) [4, 5, 6].

«Корпус – репрезентативное собрание текстов, обычно в машиночитаемом формате, включающее информацию о ситуации, в которой текст был произведен, такую, как информация о говорящем, авторе, адресате или аудитории» (Э. Финеган) [22].

«Корпус – это не только мощный инструмент исследования языка, но и новая идеология, ориентирующая исследователя на текст как главный объект теоретической рефлексии. Можно сказать, что корпус в каком-то смысле вернул лингвистам их подлинный объект – тексты на естественном языке в максимально полном объеме, и последствия этого неожиданного приобретения, на наш взгляд, еще скажутся в ближайшем будущем не только на лингвистической практике, но и на лингвистической теории» (В. А. Плунгян) [14].

М. Ю. Колокольникова пишет, что «корпус часто рассматривается как некая уменьшенная модель языка или подъязыка». Автор устанавливает соотношение между корпусным методом и дискурс-анализом. «На первоначальном этапе дискурс-

анализа необходимо, прежде всего, выделить корпус текстов, в котором бы нашли отражение специфические черты изучаемой культурно-речевой сферы. При этом электронная обработка текстов имеет, естественно, неоспоримые преимущества перед ручной, крайне трудоемкой и требующей большой затраты времени, которая далеко не всегда оправдывает себя» [8].

В прикладных исследованиях под корпусом часто понимается электронная база, которая включает два вида электронных ресурсов – корпусы текстов и словарь [8].

Е. А. Нахимова лингвистическим корпусом называет совокупность текстов, прошедших необходимую компьютерную обработку, они размечены «по определенной схеме и обеспечены поисковой системой, которая позволяет выявлять те или иные лингвистические феномены» [11].

Н. В. Козлова указывает, что корпусом считается собрание текстов одного или нескольких языков, связанных между собой определенными параметрами. Н. В. Козлова верно определяет особенности лингвистического корпуса: «множество текстов представлены в электронном виде; языковые данные размечены для анализа в лингвистических целях»; возможно дальнейшее распределение текстового материала по различным параметрам [7]. Таким образом, внимание к дефиниции понятия «корпус» находит отражение в различных направлениях языкознания [7].

Наукометрический анализ корпусной лингвистики, проведенный на сайте НЭБ (elibrary.ru) – Российского индекса научного цитирования (РИНЦ), позволяет судить о чрезвычайной распространенности данного понятия в современной лингвистике в целом. Путем ввода синтагматически зафиксированных словосочетаний мы обнаружили виды анализа таких видов лингвистических корпусов, как:

а) корпус по материальной форме представления: *электронный корпус, размеченный корпус текстов;*

б) корпус по характеру включенных документов: *корпус аутентичных устных текстов, корпус текстов, лингвистический корпус, Корпус Гиппократа (медицинские тексты), корпус публицистических текстов, электронный корпус публикаций по технической эксплуатации, корпус текстов интервью, корпус параллельных текстов, корпус сказок, корпус PR-текстов, корпус современных церковнославянских богослужебных текстов, корпус текстов Платона.*

в) корпус по языковой принадлежности:

Национальный корпус (калмыцкого языка, русского языка, корпус современного американского английского языка (COCA), корпус английских текстов, электронный корпус текстов тувинского языка, русско-белорусский корпус учебных текстов, корпус прозаических текстов башкирского языка).

Отметим что в «Лингвистическом энциклопедическом словаре русского языка» [9] понятие «корпус» отсутствует, что свидетельствует о развитии данного понятия в лингвистике постсоветского периода.

Появление корпусной лингвистики предопределило и закрепление терминологического аппарата этой прикладной дисциплины: лингвистический корпус, корпусные методы исследования и т. д. Корпусные методы могут быть применены в рамках исследования отдельных областей лексики русского языка, однако их специфика может варьироваться в соответствии с особенностями данных смысловых областей, возможностей языковых единиц, их употребления, синтагматических и парадигматических возможностей и т. п. Особенности словоупотребления лексем военной сферы почерпнуты нами в работах, посвященных русскому военному дискурсу [16–20]. Мы предприняли корпусное исследование военной лексики русского языка и пришли к выводу, что оно возможно в следующих аспектах:

1. С позиции словесной единицы / словосочетания;
2. С позиции производной / непроизводной лексической единицы;
3. С позиции словесной единицы в прямом значении / в переносном значении.

Отметим, что в «Национальном корпусе русского языка» поиск предлагается в акцентологическом / газетном / диалектном / мультимедийном / обучающем / параллельном / поэтическом / синтаксическом / устном корпусах. В исследованиях, построенных на корпусном методе возможно изучение лексем, отличающихся продуктивностью употребления. Корпусное исследование лексем предпринято нами и в работе [21] о присутствии / соотношении в «Национальном корпусе русского языка» (ruscorpora.ru) терминов «предприниматель» и «бизнесмен».

При количественном подсчете использования экономических терминов *предприниматель* и *бизнесмен* в русском медийном дискурсе удалось установить, что термин *предприниматель* зафиксирован в 689 документах 956 раз, а термин *бизнесмен* в 574 документах 886 раз. Таким образом, можно сделать вывод о том, что термин *предприниматель* чаще используется в языке и сфера его

функционирования намного шире. В сфере публицистики нам удалось выявить 745 случаев употребления словоформ термина *предприниматель*, и 540 случаев употребления термина *бизнесмен*. Единственная сфера, в которой нам удалось выявить преимущество в употреблении термина *бизнесмен*, это художественная, здесь количественное соотношение выглядит следующим образом: *предприниматель* – 109 случаев употребления словоформ в 71 документе, и 325 случаев употребления словоформ термина *бизнесмен* в 136 документах. Например, в сфере электронной коммуникации соотношение употребления данных терминов составляет 13 словоформ к 1 в пользу термина *предприниматель*. В таких нехудожественных сферах, как рекламная, церковно-богословская, бытовая и производственно-техническая, использования данного термина не выявлено вообще, в то время, как термин *предприниматель* активно используется в них.

Отметим, что важность корпусного метода предопределяет важность исследования лексического материала в массиве лингвистических единиц.

Литература

1. Баранов, А. Н. Введение в прикладную лингвистику: учебное пособие / А. Н. Баранов. – М.: Эдиториал УРСС, 2001. – 360 с.
2. Бубнова, Н. В. Оним Смоленск в составе общенациональных фоновых знаний (на материале национального корпуса русского языка) / Н. В. Бубнова [Электронный ресурс] // Язык и культура. – 2014. – № 3 (27). – Режим доступа: <http://cyberleninka.ru/article/n/onim-smolensk-v-sostave-obshchenatsionalnykh-fonovykh-znaniy-na-materiale-natsionalnogo-korpusa-russkogo-yazyka>
3. Дюрчо, П. Паремология и корпусная лингвистика / П. Дюрчо [Электронный ресурс] // Вестник НовГУ. – 2014. – № 77. – Режим доступа: <http://cyberleninka.ru/article/n/paremiologiya-i-korpusnaya-lingvistika>
4. Захаров, В. П. Поисковые системы Интернета как инструмент лингвистических исследований / В. П. Захаров // Русский язык в Интернете. – Казань, 2003.
5. Захаров, В. П., Богданова С. Ю. Корпусная лингвистика / В. П. Захаров, С. Ю. Богданова: учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн. – СПб.: СПбГУ, 2013. – 148 с.
6. Захаров, В. П., Богданова С. Ю. Корпусная лингвистика / В. П. Захаров, С. Ю. Богданова: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.

7. Козлова, Н. В. Лингвистические корпуса: Определение основных понятий и типология / Н. В. Козлова [Электронный ресурс]. Режим доступа: <http://www.nsu.ru/rs/mw/link/Media/34759>

8. Колокольникова, М. Ю. Дискурс-анализ и корпусный анализ в исследованиях в области исторической лексикологии / М. Ю. Колокольникова [Электронный ресурс] // Известия Саратовского университета. Новая серия. – Серия Филология. Журналистика. – Т. 10. – Вып. 2. – С. 3–6. – Режим доступа: <http://www.philology.ru/linguistics1/kolokolnikova-10.htm>

9. Лингвистический энциклопедический словарь / Под ред. В. Н. Ярцевой. – Москва: Советская энциклопедия, 1990. – 682 с.

10. Нагель, О. В. Корпусная лингвистика и ее использование в компьютеризированном языковом обучении / О. В. Нагель [Электронный ресурс] // Язык и культура. 2008. – № 4. – Режим доступа: <http://cyberleninka.ru/article/n/korpusnaya-lingvistika-i-ee-ispolzovanie-v-kompyuterizirovannom-yazykovom-obuchenii>

11. Нахимова, Е. А. Использование корпусной методологии при сопоставительном изучении прецедентных имен / Е. А. Нахимова // Политическая лингвистика. – Вып. 3 (45). – Екатеринбург, 2013. – С. 48–56.

12. Некрасова, Г. А. Электронный диалектный корпус как ресурс сохранения и изучения коми диалектов / Г. А. Некрасова // Финно-угорский мир. № 1 (6). – Саранск, 2010. – С. 13–16.

13. Пахолкова, И. А. Применение методов корпусной лингвистики в традиционном языкознании / И. А. Пахолкова [Электронный ресурс] // Вестник МГЛУ. 2012. – №13 (646). – Режим доступа: <http://cyberleninka.ru/article/n/primenenie-metodov-korpusnoy-lingvistiki-v-traditsionnom-yazykoznanii>

14. Плунгян, В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики / В. А. Плунгян // Русский язык в научном освещении. – 2008. – № 2 (16). – С. 7–20.

15. Сысоев, П. В. Лингвистический корпус в методике обучения иностранным языкам / П. В. Сысоев [Электронный ресурс] // Язык и культура. – 2010. – №1 (9). – Режим доступа: <http://cyberleninka.ru/article/n/lingvisticheskiy-korpus-v-metodike-obucheniya-inostrannym-yazykam>

16. Уланов, А. В. Лексические особенности русского военного дискурса XIX века (региональный аспект) / А. В. Уланов [Электронный ресурс] // Гуманитарные исследования в Восточной Сибири и на Дальнем Востоке. – 2013. – № 4 (24). – Режим доступа:

<http://cyberleninka.ru/article/n/leksicheskie-osobennosti-russkogo-voennogo-diskursa-xix-veka-regionalnyy-aspekt>

17. Уланов, А. В. Региональные особенности русского военного дискурса XIX начала XX веков / А. В. Уланов [Электронный ресурс] // ОНВ. – 2014. – №3 (129). – Режим доступа: <http://cyberleninka.ru/article/n/regionalnye-osobennosti-russkogo-voennogo-diskursa-xix-nachala-xx-vekov-1>

18. Уланов, А. В. Диахроническое развитие русского военного дискурса XIX века / А. В. Уланов [Электронный ресурс] // Вестник Северо-Восточного федерального университета им. М. К. Аммосова. – 2014. – № 4. – Режим доступа: <http://cyberleninka.ru/article/n/diahronicheskoe-razvitiie-russkogo-voennogo-diskursa-hih-veka>

19. Уланов, А. В. Военная структура в рамках русского военного дискурса и военной языковой картине мира XIX в. / А. В. Уланов [Электронный ресурс] // Вестник ОмГУ. – 2013. – №1 (67). – Режим доступа: <http://cyberleninka.ru/article/n/voennaya-struktura-v-ramkah-russkogo-voennogo-diskursa-i-voennoy-yazykovoy-kartine-mira-xix-v>

20. Уланов, А. В. Лексема солдат в военном дискурсе русского языка / А. В. Уланов // Вестник Северо-Восточного федерального университета им. М.К. Аммосова. – 2016. – № 6 (56). – С. 110–118.

21. Уланов А. В., Просвирнина Л. Г. Предприниматель vs. Businessman в русском медийном дискурсе / А. В. Уланов, Л. Г. Просвирнина // Социально-экономические и правовые системы: современное видение : материалы Междунар. науч.-практ. конф. (Омск, 3 марта 2017 г.) / АНОО ВО «СИБИТ»; Северо-Казахстан. гос. ун-т им. М. Козыбаева; Межд-унар. ин-т Ун-та внеш. экон. связей и внеш. торговли; Ун-т «Туран-Астана»; [редкол.: В. А. Ковалев, М. Г. Родионов, А. В. Уланов]. – Омск : Изд-во ОмГТУ, 2017. – С. 235–239.

22. Finegan E. LANGUAGE: its structure and use / E. Finegan. – N. Y.: Harcourt Brace College Publishers, 2004.

**CORPUS METHOD IN THE STUDY
OF RUSSIAN VOCABULARY**

A.V. Ulanov

Abstract

The article is devoted to peculiarities of application of the corpus approach to the study of military vocabulary of the Russian language. The author dwells on the importance of using corpus approach to the Russian language, which allows you to explore specific units not only from the formal semantic, but also from a quantitative and statistical point of view. Opinions of scientists on the nature of linguistic or language, corpus are presented in the article. The general characteristics of the corpus approach, which define terminology within the linguistic thought, are established. The study of Russian vocabulary from the position of the word units / collocations, derivative / non-derivative lexical units, verbal units in the literal / figurative sense plays an important role in corpus linguistics of the modern Russian language.

Key words: corpus linguistics, linguistic corpus, corpus-based methods, lexical fund of the Russian language