

**КОРПУСНАЯ ЛИНГВИСТИКА:
УЧЕБНЫЙ СЛОВАРЬ
КЛЮЧЕВЫХ ТЕРМИНОВ
И ПОНЯТИЙ**

Алматы
«Қазақ университеті»
2017

УДК 81'1
ББК 81.2стд1-5
К 68

*Рекомендовано к изданию Ученым советом
факультета филологии и мировых языков
и РИСО КазНУ им. аль-Фараби
(протокол №4 от 26.05.2017 г.)*

Рецензенты:

доктор филологических наук **Б.К. Мурзалина**
доктор филологических наук **Б.А. Ахатова**
доктор филологических наук **Ж.К. Ибраева**

Составители:

Г.Б. Мадиева, С. Бектемирова, Н. Исмаилова

К 68 Корпусная лингвистика: учебный словарь ключевых терминов и понятий / сост.: Г.Б. Мадиева, С.Б. Бектемирова, Н.А. Исмаилова. – Алматы: Қазақ университеті, 2017. – 48 с.

ISBN 978-601-04-2656-6

Настоящий словарь представляет собой опыт создания учебного словаря по корпусной лингвистике на русском языке, подготовленного в учебных целях для студентов бакалавриата, магистрантов, докторантов Ph.D и всех, кто интересуется проблемами корпусной лингвистики. В словаре представлены базовые термины и понятия, используемые в целях исследования, изучения проблем корпусной лингвистики, создания корпусов, а также для выполнения самостоятельной работы студентами, магистрантами, докторантами по вопросам соответствующих дисциплин. Изучение ключевых терминов и понятий по курсам, посвященным проблемам корпусной лингвистики, позволяет сформировать компетенции по владению и использованию понятийно-терминологического аппарата нового для казахстанских вузов направления лингвистики, занимающегося теоретико-прикладными вопросами составления национальных корпусов различных языков.

Издается в авторской редакции.

**УДК 81'1
ББК 81.2стд1-5**

ISBN 978-601-04-2656-6

© Сост.: Мадиева Г.Б., Бектемирова С.Б.,
Исмаилова Н.А., 2017
© КазНУ имени аль-Фараби, 2017

ПРЕДИСЛОВИЕ

На протяжении ряда лет на филологических специальностях бакалавриата, магистратуры и докторантуры Казахского национального университета им. аль-Фараби читаются дисциплины *Корпусная лингвистика и инновационные технологии в лингвистике*, *Корпусная лингвистика и компьютерные инструменты в обучении языку*, *От структурализма до корпусной лингвистики*, *Национальный корпус казахского языка*, *British National corpora*, *Теория и практика корпусной лингвистики*, *Теоретические вопросы корпусной лингвистики и технология создания корпусов* и др., в которых рассматриваются проблемы корпусной лингвистики как активно развивающегося направления современного языкознания. Основополагающая цель этих дисциплин – ознакомить студентов, магистрантов, докторантов с основами корпусной лингвистики в рамках эмпирического подхода к изучению языка, с концепциями корпусной лингвистики и построением корпусов различных языков. В задачи этих дисциплин входит формирование различных компетенций: компетенций по знанию основ корпусных технологий, навыков работы с различными национальными, специализированными и специальными корпусами, умение определять место дисциплин, рассматривающих вопросы корпусной лингвистики в ряду дисциплин лингвистического цикла, выделять теоретическое и практическое значение корпусов для лингвистических исследований в рамках доказательной лингвистики, владеть метаязыком корпусной лингвистики. В связи с этим возникает необходимость представить системный характер терминологического аппарата корпусной лингвистики как составляющего компонента ее метаязыка, что возможно на материале учебного словаря. Фундаментальной теоретической основой для составления предлагаемого учебного словаря послужили работы зарубежных и казахстанских авторов: Atkins S., Clear J., Ostler N., Гвишиани Н.Б.,

Захарова В.П., Даниэля М.А., Ludeling A., Kyto M., Плунгяна В.В., Рахилиной Е., Сичинава Д.В., Тукеева А.У. и мн. др.

Функциональная направленность современных исследований породила необходимость изучения языка/языков в различных коммуникативных ситуациях с целью презентации реального речевого действия, использования различных языковых единиц в определенном контексте. Эта довольно сложная проблема должна решаться при наличии объемного иллюстративного материала, который можно получить при помощи новых информационных технологий на основе Интернет-ресурсов, в том числе корпусов различных языков. Этот факт обусловлен мощным развитием компьютерных технологий, их широким использованием практически во всех сферах социума, в том числе в обучении и специальных исследованиях казахского языка. В связи с этим в настоящее время назрела острая необходимость в разработке Национального корпуса казахского языка. Значимость корпуса, как отмечает В.М. Плунгян¹ – один из разработчиков корпуса русского языка, можно определить тем, что: «это очень эффективный и полезный инструмент (которым могут пользоваться далеко не только узкие специалисты), особенно в том случае, когда корпус является большим по объему и полным по охвату материала, т.е. представляет собой так называемый Национальный корпус языка.... Корпус некоторого языка – это, в первом приближении, собрание текстов на данном языке, представленное в электронной форме и снабженное научным аппаратом. Но это не просто собрание текстов наподобие электронной библиотеки, это текстовая база данных, которая дает возможность быстрого и эффективного представления различной информации о слове и его

¹ Плунгян В.М. Общество в зеркале языка /ж. Отечественные записки. – М., 2005. – № 2; Зачем нужен Национальный корпус русского языка. Неформальное введение.

реализации в контексте на материале текстов различного жанра, стиля, хронологического периода, определенного автора».²

Эта проблема успешно решается в зарубежной лингвистике (Великобритании, США, Франции, Германии, Японии, Румынии, Чехии и т.д.): начиная с 60-х гг., в этих странах стали, соответственно, создаваться национальные корпуса: Корпус английского языка Университета Браун, Британский национальный корпус, Национальный американский корпус, Корпус французского языка и мн. др. Приоритетной задачей является создание полномасштабных и многофункциональных фондов национальных языков. В результате свое развитие получило такое научно-прикладное направление, как корпусная лингвистика (60-е гг. XX в.), занимающаяся разработкой, созданием и использованием текстовых корпусов. Появление корпусной лингвистики обусловлено задачами создания объемного, представительного фонда текстов национального языка, который мог бы коллекционировать практически весь конкретный язык, проявленный в устной и

² Наука о корпусах (а таковая давно уже существует и называется «корпусная лингвистика») – это, прежде всего, наука о том, как сделать хорошую разметку корпуса». По его практическому замечанию, изменения в русском языке наглядно прослеживаются на основе корпуса. Так, развитие языка, рано или поздно должно сказаться на слабых зонах русской грамматической системы. Грамматика пока держится... наиболее уязвимо в русском языке склонение. Тексты Корпуса показывают заметный рост – особенно в устной речи – именительного падежа во всех позициях. Так, числительные уже практически перестали склоняться (совсем не трудно услышать что-нибудь вроде «остался без триста двадцать пять тысяч рублей»). На наших глазах лишаются формы склонения и географические названия типа Бородино: в повседневной речи фактически единственным вариантом становится «рядом с Бородино, живу в Бородино» и т. п. Не желают склоняться, что характерно, и новые заимствования, а заимствования никогда не подчиняются тем законам, которые в самом принимающем языке действуют слабо. Ждет ли русский язык судьба болгарского – потерявшего все падежи, функции которых взяли на себя предлоги, – конечно, ни один серьезный специалист предсказывать не берется. Но «что-то может произойти». Теперь у нас есть Корпус. Так что получить ответ, может быть, будет проще [www.strana-oz.ru/].

письменной форме. Задача достаточно сложная, трудоемкая, выполнению которой, в свою очередь, способствовало развитие вычислительной техники (80-е гг. XX в.). Созданные компьютерные фонды включают от 600 тыс. до более 100 млн словоупотреблений. Однако это только начальный этап в накоплении и расширении корпусов.

Казахстанская лингвистика в этом направлении также предпринимает определенные шаги: на базе машинного фонда текстов казахского языка ИЯ НАН РК созданы различные лексикографические источники. Кроме того, современное наполнение интернет-сайтов текстами на казахском языке дают уникальную возможность для парсинга – автоматизированного процесса «копирования материалов с одного сайта (или нескольких), на другой сайт (или в базу данных, для последующей заливки на собственные сайты или продажи)» [<https://ruseo.net/chto-takoe-parsing-ili-sparsit-t9447.html>]. Тем не менее развитие корпусной лингвистики и создание национального корпуса (корпусов) – перспектива современного казахстанского языкознания. В свое время в одном из интервью было отмечено, что «в Комитете по языкам будет создан аппаратно-программный комплекс на базе современной промышленной интеграционной платформы, на основе которого предполагается создать мощный единый республиканский центр, призванный обеспечить для пользователей принцип «единой точки входа» по предоставлению широкого ряда услуг или сервисов по поддержке как государственного языка, так и других языков народов Казахстана» [<http://www.zakon.kz/87167-nacionalnyjj-korpus-nacii.-intervju.html>].

Одним из основных направлений деятельности Комитета является создание системы анализа современного состояния государственного языка на базе создания национального корпуса казахского языка. Благодаря поддержке Правительства Комитет по языкам приступил к реализации трехлетней программы «Создание информационных систем поддержки

и развития государственного языка Республики Казахстан». Планируется разработать комплекс компьютерных программ, иначе говоря, софт на казахском языке, широкомасштабную инфраструктуру, внедрить казахский язык в интернет-пространство, создать серии онлайн-сервисов, целостную интегрированную систему развития государственного языка Казахстана с учетом опыта прошлых лет и современного развития информационных технологий» [www.bimash.kz].

Отрадно отметить, что эти задачи успешно внедряются в жизнь: появились многочисленные интернет-словари, созданы различные веб-сайты по изучению казахского языка и предоставлению различной информации на казахском языке. В настоящее время, например, разработаны и функционируют следующие сайты, порталы, платформы, блоги: образовательные сайты: *Уикипедия* (по аналогии с Википедия), *Викибілім*, *Сөздік.кз*, *Тұған тіл* и др.; развлекательные сайты: *Мәссаған*, *Қазақша КВН*, *ІргеТас* и др.; познавательные сайты: www.tanym.kz и др.; сайты религиозного направления: *Дін ислам*, *діни сайты* и др. Мы можем проследить как появление, так развитие интернет-ресурсов, предназначенных для носителей казахского языка, интересующихся, изучающих казахский язык и культуру, языки и культуру других народов, поскольку эти ресурсы подают различную информацию на казахском языке.

Наполняемость и характер этих сайтов, веб-страниц, платформ, блогов отличаются своими характеристиками: от простого обозначения до содержательного, информативного наполнения с разработкой своеобразного дизайна. Созданы и различного рода корпуса казахского языка, что, безусловно, показывает положительную динамику их развития. В связи с этим важно изучение терминологии корпусной лингвистики. Настоящий словарь, созданный в учебных целях в проектном режиме, преследует именно задачу системной подачи соответствующих терминов корпусной лингвистики, составляю-

щих ядерный фонд ее терминологического аппарата, необходимый для усвоения.

Материал в учебном словаре представлен в соответствии с традиционными принципами лексикографирования терминов в словарях, относящихся к обучающему типу. Все, кто интересуется вопросами корпусной лингвистики, будут иметь возможность найти искомый термин, узнать его дефиницию, пополнить свои знания, сформировать терминологический аппарат, необходимый для использования в исследовательской работе. В конце словаря дается список терминов и терминосочетаний, использующихся при описании корпусов, а также перечень сформированных корпусов и список литературы – научных источников, из которых выбраны предлагаемые в словаре термины.

СОДЕРЖАНИЕ

Предисловие	3
Ключевые термины и понятия по корпусной лингвистике	9
Список основных терминологических словосочетаний корпусной лингвистики	34
Список существующих национальных корпусов	36
Библиографический список	42