



Казакский национальный университет
имени аль-Фараби

НАЦИОНАЛЬНЫЙ КОРПУС ЯЗЫКА



ВОПРОСЫ ПРЕПОДАВАНИЯ ИНОСТРАННЫХ ЯЗЫКОВ

Сборник материалов
международного научно-методического семинара
26 ноября 2015 г.

Алматы 2015

СОДЕРЖАНИЕ

<i>Г.Б. Мадиева, С.Б. Бектемирова, М.Қ. Мәмбетова, Г.Н. Исакова</i> Методы и приемы составления корпуса казахского языка	3
<i>С.Б. Бектемирова, Д.А. Карагойшиева</i> Электронды сөздіктерді құрастырудың тиімді жолдары	12
<i>А.А. Бонч-Осмоловская</i> Национальный корпус русского языка: культуромика (к постановке задачи)	19
<i>Н.Р. Добрушина, М.А. Даниэль</i> Корпус устной диалектной речи	25
<i>Н. Р. Добрушина, М. А. Даниэль, Л.Я.Зайдельман</i> Автоматические подходы к сбору разметке текстов казахского языка	35
<i>Г..N. Iskakova, D.A. Karagoishiyeva</i> The corpus-based approach to language teaching and learning.....	41
<i>Д.А. Қарагойшиева, С.Б. Бектемирова, Г. Н. Исакова</i> Ауызша сөйлеу корпусының тілді оқытудағы тиімділігі (қазақ тілінің сөйлеу синтезаторы негізінде)	44
<i>Р. Кондыбаева, Ж. Кузембекова</i> Методы исследования интенсивности изолированных слов в казахском языке	52
<i>М.Қ. Мәмбетова, А.Қ. Мәмбетова</i> Қазақ тілінің ресми-іскери корпусын оқытудағы клишелердің рөлі	60
<i>А. Әлиакбарова, Н. Утемғалиева, Д. Рскелдиев</i> Ағылшын тілін оқытуда лексикалық интерференцияның рөлі	66
<i>A.Zh. Aksholakova, V.B. Shakenova</i> Effective time management in the process of language learning.....	71
<i>T. Dzharasova, D. Rskeldieva</i> Toastmaster's speaking lessons.....	76

П.Т. Медетбекова

«Сәлемдесу/коштасу» үлгілерінің тілді үйретудегі ерекшеліктері....80

З.А. Муканова

Методика преподавания латинского языка в системе гуманитарного образования87

А.К. Садыкова, А.Т. Аушахман

Современные подходы к обучению иностранному языку для профессиональных целей92

З.С. Советова

Қазақ тілін екінші тіл ретінде үйретуде техникалық бағытта білім алатын болашақ мамандардың коммуникативтік құзыреттілігін дамыту жолдары.....97

А.К. Таусогарова

Қақтығысты мәтіндердің прагматикалық ерекшеліктеріне лингвистикалық талдау жүргізу тәжірибесінен104

Ж.М. Уматова, А.Т. Алимбаева

Шет тілдерін оқыту үдерісінде интернет-технологияларды қолдану.....109

Н.А. Утемғалиева, Д.Б. Рскелдиева, А.Т. Әлиакбарова

Ағылшын тілін оқытуда case-study әдісін қолдану.....117

V.B. Shakenova, A.Zh. Aksholakova

Scientific and methodological culture of teachers122

*Международный научно-методический семинар и сборник материалов
выполнены в рамках проекта грантового финансирования МОН РК №4769/ГФ
«Разработка интеллектуальной информационной системы
Национального корпуса казахского языка»*

Под общей редакцией д.ф.н. профессора **Г.Б. Мадиевой**

Редакционная коллегия:

Карагойшиева Д.А., Искакова Г.Н., Мамбетова М.К.

**Национальный корпус языка & Вопросы преподавания
иностранных языков: сборник материалов международного
научно-методического семинара. Алматы, 26 ноября 2015 г.
– Алматы: Қазак университеті, 2015. – 134 с.**

ISBN 978-601-04-1604-8

Сборник материалов международного научно-методического семинара подготовлен на факультете филологии и мировых языков Казахского национального университета имени аль-Фараби. Сборник посвящен проблемам корпусной лингвистики, вопросам наполнения и реализации Национального корпуса казахского языка и проблемам преподавания иностранных языков в свете новейших достижений современной дидактики.

Все публикации даются в авторской редакции.

МЕТОДЫ И ПРИЕМЫ СОСТАВЛЕНИЯ КОРПУСА КАЗАХСКОГО ЯЗЫКА

Г.Б. Мадиева, д.ф.н., проф.
С.Б. Бектемирова, к.ф.н., доцент
М.К. Мамбетова, к.ф.н., доцент
Г.Н. Исакова, магистр филологии

*Казахский национальный университет им. аль-Фараби
Алматы, Казахстан*

Ключевые слова: национальный корпус, казахский язык, информационные технологии, проект.

С приданием казахскому языку статуса государственного языка его роль значительно усилилась, что проявляется в расширении всех его функций и использовании во всех сферах общения. Статусному положению казахского языка, законодательным мерам проводимой языковой политики должно быть соответственное лингвистическое наполнение, одним из приоритетов которого является проблема корпусного планирования – попытки стандартизировать и систематизировать язык на основе информационных технологий. В рамках Государственной программы функционирования и развития языков (2011-2020) появилась острая необходимость создания Национального корпуса казахского языка.

Формирование Корпуса – актуальная задача многих современных мировых сообществ, поскольку государственному статусу языка могут соответствовать не только кодифицированные языки, но и все его реализации во всех стилях и жанрах. Как отмечает один из создателей Национального корпуса русского языка В. Плунгян, корпус языка – это эффективный и полезный инструмент, особенно в том случае, когда корпус является большим по объему и полным по охвату материала, т.е. представляет собой Национальный корпус языка <...>. Корпус языка – это, в первом приближении, собрание текстов на данном языке, представленное в электронной форме и снабженное научным аппаратом. Аппарат, «встроенный» в корпус, называется «разметкой», или «аннотацией»; корпус тем лучше, чем полнее и совершеннее его аннотация» [1]. Создание корпуса – длительный, трудоемкий

процесс, который создается усилиями многих центров и институтов при поддержке государственных программ и информационных ресурсов.

Большинство крупных языков мира имеет свои национальные корпуса, различающиеся по полноте и уровню научной обработки текстов. Общеизвестным образцом является Британский национальный корпус (BNC), на который ориентированы многие современные корпуса. Среди корпусов славянских языков выделяется Чешский национальный корпус [2]. Первый большой компьютерный корпус – *Брауновский корпус* (США, 500 фрагментов текстов, 1 млн. слов). По модели Брауновского корпуса создан частотный словарь русского языка Л.Н. Засориной (1970), построенный на основе корпуса текстов в 1 млн. слов, а также русский корпус, созданный в Университете Уппсалы (Швеция).

Развитие информационных технологий и компьютерных мощностей, способных работать с большими объемами текстов, позволили в 80-е годы XX в. предпринять попытки создать корпуса большего размера: Банк Английского, Британский Национальный Корпус, Машинный фонд русского языка и др. [3]. Формирование фонда текстов в электронном формате значительно облегчило задачу создания представительных корпусов объемом в десятки и сотни миллионов слов. Однако проблемы по созданию корпусов остаются актуальными, т.к. необходимо решить такие задачи, как: инвентаризация большого количества текстов, снятие проблем с авторскими правами, приведение текстов в единый формат, классификация корпуса по темам, стилям, жанрам, снятие омонимии. Представительные корпуса существуют (или разрабатываются) для многих языков мира: финского, польского, лезгинского, турецкого, словенского, немецкого, армянского, японского, болгарского и др. [2]. Например, Национальный корпус русского языка содержит более 300 млн. словоупотреблений.

В Казахстане существуют попытка создания Национального корпуса [4], однако его наполнение до сих пор не достигло ожидаемого, даже минимального, результата. На настоящий момент в рамках научно-исследовательского проекта учеными Казахского национального университета им. аль-Фараби разработана пилотная версия корпуса казахского языка в 1 млн. словоупотреблений, который был назван как *Алматинский корпус казахского*

языка (АККЯ). Эта версия Корпуса создавался совместными усилиями с учеными Научно-исследовательского университета *Высшая школа экономики*. Для корпуса была адаптирована поисковая система Восточноармянского национального корпуса. Однако размер в один млн. слов достаточен только для лексикографического описания самых частотных слов [3]. В связи с этим необходимо совершенствование Корпуса, увеличения объема текстов различного жанра и стилей, улучшения его поисковой системы, качественной разработки разметок, метаразметок, снятия омонимии, расширение контекста и мн.др. Он должен быть сбалансированным и представительным по объему (сотни миллионов словоупотреблений), оснащенный всеми возможными видами полной и удобной разметки.

Создание корпуса позволит изучать историю казахского языка, обучать и обучаться казахскому языку, осуществить статистический мониторинг функционирования лексических, грамматических и стилистических языковых средств, работать по лексикографической поддержке современного казахского языка, его стандартизации, создавать словари, учебники, справочные пособия, проводить статистический анализ различных языковых единиц. АККЯ (Алматинский корпус казахского языка) способен служить современным источником кодификации и стандартизации казахского языка, поскольку в корпусе оказывается зафиксированным письменный язык в его максимально репрезентативном виде. В перспективе намечается фиксировать и звучащую речь.

Казахский язык, его особенность и уникальность, история и современность, будущее, вероятность кардинальных изменений все это является важным и актуальным не только для лингвистов, но и для специалистов многих отраслей: культуры, экономики, истории, политики и т.д., в том числе языковой. Подобные масштабные вопросы с недавнего времени обсуждаются и решаются с помощью такого механизма, как корпус языка.

Создание национальных корпусов государственных языков ведущих стран мира возведено в ранг важных историко-культурных и политических мероприятий современности. Понятие *корпус* многими отождествляется с понятием «набора текстов или языковых единиц», что не дает необходимой теоретико-методологической базы для того, чтобы рассматривать корпус не толь-

ко как феномен, обладающий определенным набором характерологических свойств и признаков, свойственных разным типам, стилям любого языка, но и как феномен идиоэтнического порядка, определяемый особенностями национальной ментальности. Эта проблема с помощью Корпуса была решена для многих хорошо изученных языков мира (английского языка, американского варианта английского языка, немецкого, русского, французского, польского и др.). Формирование Национального корпуса казахского языка — одна из важнейших, задач суверенного Казахстана. В обозримом будущем это задача

В Республике Казахстан впервые планируется создание НККЯ (Национальный корпус казахского языка) столь масштабного по объему текстов и тематике подкорпусов, который будет востребован не только отечественными потребителями, но и зарубежными с целью исследования казахского языка, его изучению и обучению.

Национальный корпус языка — неопределимый инновационный инструмент, сокращающий затраты времени на техническую работу по изучению языковых явлений и за считанные минуты дающий возможность найти справочную информацию. НККЯ — это не просто техническая поддержка лингвистических исследований. Это справочно-информационная база по современному казахскому языку, позволяющая получать ответы на многие вопросы, которые возникают перед любым потребителем, изучающим казахский язык, а также ставить новые проблемы, которые не входили в круг проблем лингвистики прошлых лет, революционизировать работу с языковым материалом. Так, создатели корпуса русского языка отмечают, что с его помощью появилась возможность «поиска по грамматическим критериям автоматически получить примеры, из которых уже вручную можно выбрать наиболее подходящие для наших целей... Корпус дает возможность отсортировать источник примеров». Еще одно преимущество Корпуса — скорость подбора примеров и источников [5]. Кроме того, «действительно, пока основная масса пользователей Корпуса — ученые-исследователи; огромный резерв здесь составляют преподаватели и учащиеся самых разных уровней — от школ до университетов, подготовительных курсов, курсов усовершенствования или второго высшего образования...», где Корпус фактически служит активным инструментом обучения грамматике,

стилистике, культуре речи и всему комплексу дисциплин, связанных с русским языком...» [6].

В рамках казахского языкознания и прикладной лингвистики исследование и разработка НККЯ представляет особый интерес, что определяется недостаточной разработанностью проблематики в данной области. Так, считая исследование корпуса казахского языка логическим продолжением традиции его изучения, тем не менее, можно апеллировать только к констатации Корпуса казахского языка, к достаточно ограниченному числу работ, посвященных описанию данного феномена, но не его наличию в полном объеме. Несмотря на достижения в этой области (попытка составления корпуса с необходимыми разметками, наличие множества научных исследований в виде монографий, диссертаций, учебников казахского языка), границы исследований не выходят за рамки традиционного языкознания, что ограничивает усилия по разработке корпуса или сводит их к механистическому выявлению отличий казахского языка. Нужен современный исследовательский механизм и практический инструмент, которым будет Корпус казахского языка. Помимо этого, обучение языку при помощи компьютерных технологий отходит от традиционных способов подачи материала и фокусируют внимание на тех видах деятельности, которые стимулируют новые подходы, например, аутентичные тексты, к которым можно получить доступ в языковом корпусе. Эти факты свидетельствуют об объективной реальности и актуальности проблемы создания Национального корпуса казахского языка. В Республике Казахстан к настоящему времени корпусная лингвистика как научное и как прикладное направления не получила до сих пор своего должного развития. Соответственно остроактуальным вопросом остается создание НККЯ.

Обозначенные выше научные лакуны, связанные с разработкой Национального корпуса казахского языка, определяют **актуальность** предпринятого научного и прикладного исследования и могут быть решены участниками рабочей группы предполагаемого Проекта с привлечением специалистов, работающих над Национальным корпусом русского языка.

В результате реализации проекта в казахстанской лингвистике впервые будет проведено планомерное исследование зарубежного опыта по корпусной лингвистике, сформируется отечествен-

ное направление по корпусной лингвистике, будет создана мощная текстовая база для наполнения контента Корпуса казахского языка, употребляемого в различных видах дискурса с использованием методов и основных принципов корпусной лингвистики. В рамках проекта предполагается обобщить практический и теоретический опыт использования различных Корпусов мира в преподавании языков в Казахстане; изучить казахский язык функционально: в науке, технике, экономике, культуре и т.п. в синхронном срезе на широком фоне социальной, культурной, политической жизни.

Настоящий проект актуален в рамках поиска путей совершенствования форм и методов образования студентов, школьников и других слоев населения на основе компьютерной компетентности и проблемно-ориентированного обучения, реализует принципы национальной политики в сфере развития государственного языка и сохранения его богатства; ориентирован на обновление содержания образовательного процесса на основе инновационно-информационных технологий, создание креативных платформ, в которые может быть инсталлирована и на которых могут быть продемонстрированы достоинства НККЯ как объемного открытого информационно-образовательного портала.

Реализация проекта активизирует новые формы использования компьютерных технологий, формирование у пользователей (специалистов, студентов, магистрантов, докторантов, школьников, учителей, т.п.) новых форм диалога «пользователь-компьютер» для различных видов работ с казахским языком; главное условие эффективности – скорость поиска, устранение механистической работы с различными текстами, поиском форм слов, подбора необходимого слова в различных реализациях и контекстах, что позволит решить ряд вопросов и достичь существенного эффекта экономии в исследовании, изучении казахского языка, а также коллекционировать его разнообразие и богатство.

Для того чтобы разработать платформу Корпуса и составить репрезентативную текстовую базу для его наполнения в ходе выполнения проекта будет использоваться совокупность лингвистических методов:

- выборка и систематизация текстов, инвентаризация текстов по хронологическим, жанровым и стилевым критериям;

- графематический анализ, позволяющий выделить синтаксические и структурные единицы входного текста (абзацы, предложения, словосочетания, отдельные слова, знаки препинания);
- морфологический анализ, предполагающий определить структуру слова, основное слово и его словоформу, отнесение к той или иной части речи для дальнейшей процедуры снятия омонимии;
- синтаксический анализ, позволяющий определить функцию слова в составе предложения, его сочетаемость с другими словами, порядок слов в предложении;
- семантический анализ, необходимый для анализа текста по смыслу, уточнения связи слов, исключающий бессмысленный набор слов.

В современных условиях в данной отрасли значительную помощь оказывает привлечение методов разработки современных корпусов:

1. морфологическая разметка: полная морфологическая характеристика каждой словоформы с возможностью определения спорных случаев, имеющих неоднозначное понимание;
2. синтаксическая разметка: выделение различных типов синтаксических единиц (предложение, словосочетание);
3. семантическая разметка: информация о семантических категориях казахского языка;
4. метаразметка (метаинформация о типе текста и его выходных данных).

Кроме того, будут использованы методы обработки естественного языка (*natural language processing* (NLP): лексикографическая обработка, токенизация, лемматизация, морфологический анализ) и другие с целью разработки *автоматизированного извлечения информации*; текстовые поиски в крупномасштабных корпусах (конкордансы).

Корпусные методы широко зарекомендовали себя в мировой практике составления корпуса языка, лингвистических исследованиях и преподавании иностранных языков как эффективные инновационные дополнения к традиционным образовательным технологиям.

Предлагаемый проект по созданию НККЯ, базируемый на пилотной версии корпуса казахского языка, должен найти широкое применение как у специалистов преподавателей казахского как родного и иностранного, так и у всех заинтересованных граждан Республики Казахстан и за рубежом.

Благодаря созданию Национального корпуса казахского языка как информационно-справочной базы появятся следующие уникальные возможности:

- 1) обеспечение создания учебников и учебных пособий по казахскому языку текстовым материалом; обеспечение в электронном виде разносторонним языковым материалом процесс обучения казахскому языку;
- 2) многократное упрощение и ускорение процедур лингвистической обработки массивов текстов на основе современных компьютерных технологий;
- 3) развитие современных знаний о казахском языке: возможность статистической обработки текстов с целью научного описания строя казахского языка на основе инновационных технологий; формирование базы знаний использования национального корпуса казахского языка;
- 4) становление теоретико-методологического подхода к организации переводческого процесса в стране на основе НККЯ;
- 5) применимость полученных научных результатов: Корпус предоставит широкие возможности для создания различного типа и жанра авторитетных академических и переводных словарей на основе НККЯ, он-лайновых отраслевых, одно-, и двуязычных толковых, терминологических, фразеологических и иных словарей; быстро и эффективно проверять особенности употребления незнакомого слова или грамматической формы у авторитетных авторов и для использования корпусных данных при многих более специальных научных исследованиях.

Предлагаемый проект по созданию Национального корпуса казахского языка, базируемый на пилотной версии корпуса казахского языка уже дает свои первые результаты.

Работа над проектом Корпуса началась при поддержке ректора КазНУ им. аль-Фараби Г. М. Мутанова. Корпус создаётся

силами кафедры общего языкознания и европейских языков факультета филологии и мировых языков Казахского национального университета им. аль-Фараби под руководством заведующей кафедрой Г. Б. Мадиевой при участии сотрудников факультета филологии НИУ ВШЭ (Москва).

Для корпуса была адаптирована поисковая система Восточноармянского национального корпуса (EANC).

http://web-corpora.net/KazakhCorpus/search/?interface_language=ru

На этом сайте размещена пилотная версия Алматинского корпуса казахского языка, находящегося на начальном этапе разработки. В настоящий момент размер корпуса составляет около 1 миллиона словоупотреблений. Тексты корпуса были размечены с помощью автоматического морфологического анализатора, 80% словоформ корпуса имеют грамматический разбор. Омонимия в корпусе не снималась, т. е. каждой словоформе приписаны все возможные варианты разбора без учёта контекста.

Это первая версия корпуса Национального корпуса казахского языка – НККЯ как справочно-информационной системы на основе обширного фонда размеченных текстов литературного казахского языка, государственного языка Республики Казахстан. Безусловно, корпус будет дополняться, обновляться как количественно, так и качественно, кроме того будет существенно улучшаться поисковая функциональность корпуса.

В перспективе основные характеристики НККЯ следующие:

- лингвистически репрезентативный корпус;
- мощный поисковый аппарат для осуществления сложных лексико-морфологических запросов;
- удобный инструмент для самостоятельного изучения казахского языка, дающий для большинства словоформ лексико-морфологические разборы и русские/английские переводные эквиваленты;
- диахронически ориентированный корпус, покрывающий различные периоды истории современного казахского языка;
- диверсифицированный корпус, включающий разножанровые письменные и устные тексты разных типов;
- аннотированный корпус, снабженный грамматической и

библиографической разметой;

- корпус, находящийся в открытом доступе.
- электронная библиотека, включающая более 100 классических произведений казахской литературы.

Литература

1. Плунгян В.А. Зачем мы делаем Национальный корпус русского языка? // Отечественные записки. 2005. № 2, 296–308.// <http://www.strana-oz.ru/2005/2/>
2. Национальный корпус русского языка <http://www.ruscorpora.ru/corpora-intro.html>
3. <https://ru.wikipedia.org/wiki>
4. <http://til.gov.kz/wps/portal>
5. Добрушина Н.Р. Как использовать Национальный корпус русского языка в образовании? // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005, 308–329.
6. Рахилина Е.В. Корпус как творческий проект // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 7–26.

ЭЛЕКТРОНДЫ СӨЗДІКТЕРДІ ҚҰРАСТЫРУДЫҢ ТИІМДІ ЖОЛДАРЫ

*С.Б. Бектемирова ф.ғ.к., доцент,
Д.А. Карагойшиева, PhD, доцент
Әл-Фараби атындағы Қазақ ұлттық университеті
Алматы, Қазақстан*

Түйін сөздер: компьютер, лексикография, сөздік, тип, электронды сөздік

Сөздіктердің типтері әртүрлі. Лексикография нысанды нақты сипаттаудан яғни тілдің күрделілігімен және көпаспектілілігімен түсіндіріледі. Тіл туралы әртүрлі ақпаратты алуда сөздіктерді күрделілендіреді және кеңейтеді. Тіл туралы барлық ақпаратты бір сөздікте беру мүмкін емес. Сол себепті де кез келген ұлттық лексикографиялардан әртүрлі типтегі ондаған, тіпті жүздеген сөздік таба аламыз.

Классификаторлардың айтуы бойынша сөздіктерді типтерге бөлу әртүрлі негіздерге байланысты болады: *сөздіктің мақсаты-*