

© Грудева Е.В., Бучилова И.А., Волкова Н.А., 2018

Грудева Елена Валерьевна

Доктор филологических наук, профессор,
Череповецкий государственный университет
(Череповец, Россия)
E-mail: egrudeva@gmail.com

Grudeva Elena Valeryevna

Doctor of Philology, Professor,
Cherepovets State University
(Cherepovets, Russia)
E-mail: egrudeva@gmail.com

Бучилова Ирина Анатольевна

Кандидат психологических наук,
аспирант, Череповецкий государственный
университет (Череповец, Россия)
E-mail: ibuchilova@yandex.ru

Buchilova Irina Anatolyevna

PhD in Psychology Sciences,
post-graduate student,
Cherepovets State University
(Cherepovets, Russia)
E-mail: ibuchilova@yandex.ru

Волкова Наталия Александровна

Доктор филологических наук, профессор,
Череповецкий государственный университет
(Череповец, Россия)
E-mail: volkova35@rambler.ru

Volkova Nataliya Aleksandrovna

Doctor of Philology, Professor,
Cherepovets State University
(Cherepovets, Russia)
E-mail: volkova35@rambler.ru

**КОРПУСЫ ОШИБОК: ЦЕЛЕВАЯ
АУДИТОРИЯ, ВОЗМОЖНАЯ
АРХИТЕКТУРА КОРПУСА¹**

**CORPORA OF ERRORS: TARGET
AUDIENCE, A POSSIBLE
ARCHITECTURE OF THE CORPUS**

Аннотация. В статье рассматривается особый тип языковых корпусов – корпусы ошибок, которые обычно относятся к категории учебных корпусов. Представлены общие вопросы создания корпусов ошибок, дан обзор наиболее известных на сегодняшний день учебных корпусов русского языка, а также предложен проект по созданию электронной базы речевых ошибок детей-инофонов, обучающихся в общеобразовательных школах города Череповца Вологодской области.

Abstract. In the article a special type of the linguistic corpora i.e. the corpora of errors which, as a rule, refer to the category of didactic corpora, is examined. One deals with the general questions about the creation of corpora of errors. The survey of today's most significant didactic corpora of the Russian language is given. Also the authors offer the project of forming an electronic database of speech errors made by the foreign children who study at the comprehensive schools of Cherepovets, Vologda region.

Ключевые слова: корпусная лингвистика, электронная база данных, учебный корпус, корпус ошибок, дети-инофоны

Keywords: corpus Linguistics, electronic database, Learner Corpus, Corpus of Errors, foreign children

¹ Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект №18-412-350003 «Создание электронной базы речевых ошибок детей-инофонов, обучающихся в общеобразовательных школах города Череповца Вологодской области»).

Введение

В последние годы в работах лингвистов-теоретиков в истории лингвистики XX века проводится достаточно четкая демаркационная линия между лингвистикой так называемой «докорпусной» эпохи и лингвистикой эпохи создания корпусов (см.: [6], [12]).

Как известно, создание языковых корпусов было связано с появившейся возможностью использования новых – компьютерных – технологий в работе лингвиста: прежде всего языковой корпус экономит время и силы лингвиста, собирающего эмпирический материал для исследования. Корпусные технологии активно внедряются в лингвистическую практику¹, практически сразу лингвисты осознали, что самые разные типы текстов могут стать материалом для создания языковых корпусов разного типа (на данный момент в типологии корпусов выделяют корпуса диахронические и синхронические, региональные и диалектологические, газетные и публицистические, поэтические, корпуса звучащей речи, монологические и полилогические корпусы, корпусы параллельных текстов, корпусы разных редакций/переводов одного памятника и т.п.). Создание национального языкового корпуса в этой связи стало восприниматься как некая сверхзадача – создание универсального инструмента для всестороннего изучения того или иного национального языка.

В последние десятилетия в мировой и отечественной корпусной лингвистике активизировалась работа по созданию корпусов языковых/речевых ошибок, которые в западной традиции относятся к разряду учебных корпусов [13]. Цель данной статьи – рассмотреть общие вопросы создания корпусов ошибок (целевая аудитория, возможная архитектура корпуса), представить обзор наиболее известных на сегодняшний день учебных корпусов русского языка, а также предложить проект по созданию электронной базы речевых ошибок детей-инофонов, обучающихся в общеобразовательных школах города Череповца Вологодской области.

Основная часть

1. Цели и задачи создания корпусов ошибок

Традиционно корпусы языковых/речевых ошибок рассматриваются в рамках учебных корпусов. Повышение интереса к созданию разного рода корпусов ошибок связано, с одной стороны, с использованием корпусных данных в образовательной деятельности. Известно, что сегодня, например, при изучении иностранного языка считается недостаточным ограничиваться только традиционными средствами обучения – словарем и грамматикой, в числе обязательного третьего средства обучения называется корпус целевого (изучаемого) языка. Работа с корпусом позволяет учащимся верифицировать лексическую и грамматическую сочетаемость слов в порожденных ими текстах, допустимость употребления той или иной языковой единицы в том или ином контексте, т.е. обращаться к корпусу с вопросами, с которыми ранее учащийся обращался к носителям изучаемого языка. Основоположник корпусной лингвистики Джон Синклер в свое время предсказывал, что «с более широким рас-

¹ Конечно, не все лингвистические вопросы могут быть решены путем обращения к корпусным данным. Об ограничениях в использовании корпусных данных в решении целого ряда лингвистических вопросов см., например: [4], [7], [11].

пространением корпуса носитель языка будет смещен с роли главной модели и верховного арбитра в вопросах языковой практики» [15].

С другой стороны, корпус ошибок нужен не только тем, кто изучает иностранный язык, но и тем, кто его преподаёт. Учителя и преподаватели иностранных языков должны быть осведомлены о сравнительной характеристике изучаемого и родного для учащегося языка, о возможной интерференции двух языковых систем (методологические основы в этой области разрабатывает контрастивная лингвистика), а также о типичных ошибках, которые допускают носители определенного языка при изучении того или иного иностранного языка. Работая с такого рода корпусом, практик-преподаватель мог бы постоянно корректировать свою педагогическую деятельность и более эффективно обучать иностранному языку.

За рубежом интерес к созданию корпусов ошибок появился в 90-е гг. XX в. Один из теоретиков этой отрасли корпусной лингвистики С. Гранже назвал появление учебных корпусов революцией в прикладной лингвистике (см. характерное название его статьи: «The learner corpus: a revolution in applied linguistics» [19]).

В последние десятилетия в отечественной лингвистике наблюдается новая волна лингвистического интереса к языковым/речевым ошибкам в теоретическом аспекте (см., например, работы А.А. Залевской, Е.В. Рахилиной). Е.В. Рахилина пионером в этой области называет А. Фрея [14] и его работу 1929 г. «Грамматика ошибок» [16]. Однако, как представляется, в отечественной теоретической лингвистике интерес к языковым/речевым ошибкам был осознан гораздо раньше. Вспомним, что именно наблюдения над речевыми ошибками-оговорками позволили И.А. Бодуэну де Куртэне обнаружить особую знаковую языковую единицу – морфему¹. Не случайно ученик И.А. Бодуэна де Куртэне по Петербургской лингвистической школе Л.В. Щерба в своей знаменитой статье 1928 г. «О тройном аспекте языковых явлений и об эксперименте в языкознании», посвященной памяти учителя, вводит понятие отрицательного языкового материала и утверждает незаменимую роль отрицательного языкового материала при изучении речевой деятельности. По сути дела, изложенные в указанной статье Щербы положения послужили на долгие десятилетия программой развития нового лингвистического направления – психолингвистики. К видам отрицательного языкового материала Щерба относит афатические расстройства, ослышки, оговорки, речевые ошибки носителей языка, детскую речь, а также материал, полученный в ходе специально организованных лингвистических экспериментов с носителями языка [17].

Дополнительным стимулирующим фактором для развития грамматики ошибок и для создания языковых корпусов, в основу которых положен отрицательный языковой материал, как представляется, выступают процессы глобализации и, в связи с этим, увеличение числа поликультурных школ, в составе которых могут находиться учащиеся с разной степенью владения языком страны пребывания / языком обучения.

¹ Бодуэн де Куртэне писал, например, о замеченных им оговорках школьников при чтении басни И.А. Крылова «Ворона и лисица», типа: «*Лиса вертом хвостит*» вместо «*Лиса хвостом вертит*», где заменам подвергаются значимые части слова – корень и окончание [1, с. 124].

Так, в отечественном образовании несколько десятилетий тому назад была предпринята попытка разграничить понятия «русский язык как иностранный» и «русский язык как неродной» [8]. Сегодня многие ученые и педагоги-практики приходят к выводу, что методически особой разницы в преподавании русского языка как иностранного и русского языка как неродного нет. В то же время в некоторых вузах (например, в Российском государственном педагогическом университете им. А.И. Герцена) ведется набор на разные бакалаврские образовательные программы – «Русский язык как иностранный» и «Русский язык как неродной», тем самым утверждаются существенные различия в этих смежных областях.

Специалисты в сфере корпусной лингвистики и в сфере разработки вопросов использования языковых корпусов в образовательной деятельности выявили разные группы носителей языка в зависимости от их статуса в области владения языком. Так, А. Мустайоки с соавторами предложил разграничивать следующие группы «нестандартных» носителей русского языка:

- студенты, изучающие русский язык как иностранный (РКИ),
- носители регионального русского,
- носители так называемого «эритажного», или унаследованного, русского (heritage Russian),
- дети и подростки,
- взрослые носители стандартного русского в определенных обстоятельствах [21].

В идеале должны быть созданы корпуса ошибок разных категорий «нестандартных» носителей русского языка.

2. Обзор русских корпусов ошибок

Подробный обзор наиболее известных в мировой корпусной лингвистике учебных корпусов представлен в статье Yukio Tono «Learner corpora: design, development and applications» [22]. Примечательным фактом является то, что большинство представленных учебных корпусов построены на текстах иностранцев, изучающих английский язык.

Интерес к созданию русских учебных корпусов возник относительно недавно. Лидером этого направления на данный момент является Лингвистическая лаборатория по корпусным технологиям НИУ «Высшая школа экономики».

Далее мы рассмотрим некоторые корпуса ошибок русского языка с точки зрения их состава и архитектуры.

Корпус русских учебных текстов (КРУТ, <https://ling.hse.ru/krut>) – это коллекция текстов на русском языке, в основу которой положены тексты, написанные студентами разных вузов [23]. Общий объем корпуса по сведениям, представленным на сайте http://web-corpora.net/CoRST/search/?interface_language=ru, составляет около 3,1 млн. слов. Как утверждают разработчики, тексты в корпусе прошли через метатекстовую, морфологическую разметку и разметку по ошибкам. Поиск по корпусу поэтому может быть осуществлен с учетом указанных видов разметки. На данный момент в корпусе КРУТ недоступен поиск по ошибкам, т.к. разметка по ошибкам проходит вручную и на данный момент эта работа еще не завершена.

Метатекстовая разметка включает информацию об авторе (пол, возраст, направление, в рамках которого студент получает образование, дисциплина, в рамках изу-

чения которой написан данный учебный текст, в ряде случаев указывается регион, в котором до 18 лет проживал автор, а также принадлежность (в случае таковой) к билингвам).

В жанровом отношении в корпусе представлены основные разновидности учебных студенческих текстов: курсовые работы, выпускные квалификационные работы, контрольные работы, эссе, письменные домашние задания, конспекты, рефераты и т.п.

Что касается морфологической разметки, то разработчики Корпуса русских учебных текстов использовали ту же систему тэгов, которая используется в Национальном корпусе русского языка (НКРЯ, www.ruscorpora.ru). Морфологическая разметка осуществлялась автоматически с помощью программы MYSTEM.

Что касается разметки по ошибкам, то, как утверждают разработчики, в корпусе будет представлена многомерная система разметки текстов по ошибкам. Планируется классифицировать тип ошибки (лексическая, грамматическая, стилистическая, дискурсивная), а также причину ошибки. В числе причин возникновения ошибок пока названы лишь две: опечатки и контаминации конструкций.

Как отмечают разработчики корпуса КРУТ, работа с ним может осуществляться для коррекции и предупреждения ошибок на занятиях со школьниками и студентами в курсах типа «Литературное редактирование», «Стилистика», на практикумах по письменной речи, а также в исследовательских целях – для изучения направления развития русского языка [13].

Русский учебный корпус RLC (<http://web-corpora.net/RLC/>) создан в 2013 г. под руководством Е.В. Рахилиной в Лингвистической лаборатории по корпусным технологиям НИУ ВШЭ. В основу корпуса положены устные и письменные тексты двух категорий носителей русского языка – иностранцев, изучающих русский язык как иностранный, и так называемых эритажных носителей русского языка. Для последней категории носителей русский язык является родным, но в силу разных причин (чаще всего – эмиграция) в качестве основного средства общения используется другой язык [2].

Метатекстовая разметка включает информацию об авторе (пол, возраст, тип носителя (эритажный/иностранец), доминантный язык, уровень владения русским языком), а также информацию о тексте (письменный/устный, жанровая принадлежность)).

В тексте размечены различные типы так называемых отклонений от стандартного русского языка.

Как отмечают разработчики RLC, «Русский учебный корпус – это ценный ресурс для исследований в области усвоения иностранных языков (Second Language Acquisition), преподавания иностранных языков (Second Language teaching), языковой интерференции, а также других лингвистических исследований.

Данные корпуса и его поисковая система предоставляют основу для сравнительного исследования текстов эритажных носителей и изучающих русский язык как иностранный. Корпус позволяет глубоко изучить такие сложные явления, как нестандартное использование русского вида, падежей, предложных конструкций, а также неверное использование слов или неверное понимание смысла в многословных конструкциях» (Использование корпуса. <http://web-corpora.net/RLC/>).

Наиболее известной базой данных детской речи, куда входят и тексты русскоязычных детей, является международная база данных детской речи CHILDEERS (Children Language Database Exchange System).

Международная база данных детской речи CHILDEERS (<http://chil-des.psy.cmu.edu>) была создана в начале 1980-х гг. американскими учеными К. Сноу и Б. Мак-Винни в Питтсбургском университете для анализа разговорной речи, спонтанной речи детей и их родителей, а также для исследования усвоения второго языка (подробнее см.: [5]). База данных CHILDEERS позволяет любому зарегистрированному пользователю разместить свои материалы по детской речи. Особенности базы данных CHILDEERS связаны с тем, что в ней собраны материалы разговорной, спонтанной речи детей. Указанная база данных является источником материала для многочисленных исследований в области освоения различных языков мира детьми с рождения.

Как видим, в мировой и российской практике наблюдается устойчивый интерес к созданию электронных баз данных, в основу которых положены тексты «нестандартных» носителей русского языка, в речи которых довольно часто встречаются отклонения от нормы. Такого рода продукты используются при решении как исследовательских, так и педагогических задач.

3. Корпус текстов и база данных: сходства и различия

Несмотря на то, что термин «база данных» имеет широкое употребление в самых разных областях, исследователи отмечают, что в настоящее время, в силу аксиоматичности этого понятия, нет единого определения данного термина (см., например, подробный анализ самых разных определений термина «база данных» в статье М.А. Мирошниченко [10]). В то же время, Мирошниченко приходит к выводу, что среди обязательных признаков базы данных следующие: «1) любая БД хранится и обрабатывается в вычислительной системе; 2) БД является сильно структурированной системой; 3) БД включает схему данных» [10, с. 87].

Попытка сравнить близкие термины, имеющие отношение к корпусной лингвистике, такие как «языковой корпус», «лингвистический корпус», «лингвистическая база данных» и другие, представлены в статье М.С. Мальцевой [9]. Любопытно, что автор находит определение языкового корпуса через понятие базы данных в работах зарубежных авторов [18], [20], ср.: «Джоанна Барздал определяет корпус как «языковую (лингвистическую) базу данных», уточняя ее как базу данных, представляющую примеры использования языка, как устно, так и письменно» [9, с. 256].

Как представляется, одно из основных отличий корпуса от базы данных – наличие в нем лингвистической разметки и пользовательского интерфейса. Предполагается, что корпус должен быть удобен для дальнейшей работы с ним. Таким образом, создание электронной базы данных можно рассматривать как предшествующий этап по отношению к разработке языкового корпуса того или иного типа.

4. Проект по созданию электронной базы речевых ошибок детей-инофонов

В данный момент группа лингвистов во главе с одним из авторов данной статьи занимается разработкой электронной базы речевых ошибок детей-инофонов, обучающихся в общеобразовательных школах города Череповца Вологодской области. Несмотря на то, что до сих пор город Череповец и Вологодская область в целом относятся к числу моноэтнических регионов Российской Федерации, число детей-

инофонов в школах г. Череповца постоянно увеличивается, и состав их с точки зрения этнической и языковой принадлежности довольно пестрый [2].

Научная новизна предлагаемого проекта связана с тем, что объектом анализа является письменная речь детей-инофонов, обучающихся в российских школах. Предполагается включить в число параметров базы речевых ошибок детей-инофонов такие категории, как: 1) имя ребенка-инофона, 2) возраст, 3) родной язык, 4) класс обучения, 5) целевое слово/словосочетание, 6) контекст, 7) тип текста, в котором встретилась ошибка, 8) квалификация ошибки.

Одна из самых сложных задач при создании базы – типология ошибок. Основная идея создания указанной базы – возможность поиска ошибок по заданным параметрам, в том числе по заданному типу ошибок.

Включение указанных параметров позволит проводить лингвистические исследования, осуществлять систематизацию речевых ошибок детей-инофонов, обучающихся в российских школах, а также разрабатывать диагностические модели языковой адаптации детей-инофонов разных возрастных и этнических категорий.

Выводы

Проведенное исследование показало интенсивное развитие в мировой, а с некоторым опозданием и в российской практике такого направления в корпусной лингвистике, как создание учебных корпусов (корпусов ошибок). К целевой аудитории корпусов ошибок относятся как основные участники образовательного процесса (преподаватели иностранных языков, студенты и школьники, изучающие иностранные языки), так и лингвисты, изучающие языковые явления, связанные с языковыми/речевыми ошибками.

Несмотря на появление в последние годы русских учебных корпусов, в основу которых положены устные и письменные тексты разных категорий «нестандартных» носителей русского языка (иностранцев, изучающих русский язык как иностранный; эритажных носителей русского языка), за пределами внимания пока оказываются другие категории «нестандартных» носителей, в частности – дети-инофоны, изучающие русский язык в российских школах.

Создание электронной базы речевых ошибок детей-инофонов, обучающихся в общеобразовательных школах города Череповца Вологодской области, призвано заполнить данную лакуну.

Литература

1. Алпатов В.М. История лингвистических учений. М., 1998. 368 с.
2. Бучилова И.А. Языковая и этническая идентичность детей-билингвов, обучающихся в школах г. Череповца // Проблемы порождения и восприятия речи: Материалы XIV выездной школы-семинара / отв. ред. Е.В. Грудева. Череповец, 2017. С. 65–73.
3. Выренкова А.С., Полинская М.С., Рахилина Е.В. Грамматика ошибок и грамматика конструкций: «эритажный» («унаследованный») русский язык // Вопросы языкознания. 2014. №3. С. 3–19.
4. Грудева Е.В. Языковой корпус: апология и критика // Девятая выездная школа-семинар «Порождение и восприятие речи»: Материалы. Череповец, 2010. С. 48–55.

5. Зырянова Е.В. Система CHILDES как метод сбора материалов и изучения детской речи // Известия РГПУ им. А.И. Герцена. 2008. №76-2. URL: <http://cyberleninka.ru/article/n/sistema-childes-kak-metod-sbora-materialov-i-izucheniya-detskoy-rechi>.
6. Копотев М.В. Введение в корпусную лингвистику. Прага, 2014. 230 с.
7. Копотев М.В., Гурин Г.Б. Разметка синтаксической неполноты в корпусе // Труды международной конференции «Диалог-2007». URL: <https://goslide.org/Documents/razmetka-sintaksicheskoy-nepolnoty-v-korpuse-2007>.
8. Лысакова И.П. Русский язык для мигрантов: РКИ или РКН? (к вопросу о содержании терминов) // XXXVIII Международная филологическая конференция. Вып. 21: Русский язык как иностранный и методика его преподавания. СПб., 2009. С. 181–186.
9. Мальцева М.С. К определению термина «лингвистический корпус» и его вариантов // Социально-экономические явления и процессы. 2011. №7. URL: <https://cyberleninka.ru/article/n/k-opredeleniyu-termina-lingvisticheskij-korpus-i-ego-variantov>.
10. Мирошниченко Е.А. К формальному определению понятия «база данных» // Проблемы информатики. 2011. №2. С. 83–87. URL: <http://www.problem-info.sccc.ru/2011-2/12.pdf>.
11. Перцов Н.В. К суждениям о фактах русского языка в свете корпусных данных // Русский язык в научном освещении. 2006. № 1(11). С. 227–245.
12. Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. № 2 (16). С. 7–20.
13. Рахилина Е.В. О новых инструментах описания русской грамматики: корпус ошибок // Симпозиум «Русская грамматика 4.0»: Лингвистика. URL: chrome-extension://ecnphlgnaianjnkcmdbpancdjoidceilk/content/web/viewer.html?source=extension_pdfhandler&file=http%3A%2F%2Fwww.rakhilina.ru%2Ffiles%2Frahilina_foreign_err.pdf.
14. Рахилина Е.В. Степени сравнения в свете русской грамматики ошибок. URL: http://www.rakhilina.ru/files/compar_er.pdf
15. Синклер Джон. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка». URL: http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4-child/texts/sinclair_introduction.pdf
16. Фрей А. Грамматика ошибок. М.: Комкнига, 2006. 304 с.
17. Щерба Л.В. О тройном аспекте языковых явлений и об эксперименте в языкознании // Щерба Л.В. Языковая система и речевая деятельность. Л., 1974. С. 24–39.
18. Barðdal Jóhanna. A Crash Course in Corpus Linguistics. May. 2002. Linguistics Division, UNT.
19. Granger S. The learner corpus: a revolution in applied linguistics. English Today. 1994. 39 (10/3): 25-9.
20. McEnery T. Corpus Linguistics. Edinburgh, 1997.
21. Mustajoki A., Protassova E., Vakhtin N. Instrumentarium of linguistics: sociolinguistic approach to non-standard Russian. Helsinki, 2010.
22. Yukio Tono. Learner corpora: design, development and applications. URL: <http://ucrel.lancs.ac.uk/publications/cl2003/papers/tono.pdf>.
23. Zevakhina N.A., Dzhakupova S.S. Corpus of russian student texts: design and prospects. URL: <http://www.dialog-21.ru/media/1144/zevakhinanadzakupovass.pdf>.

References

1. Alpatov V.M. *Istoriya lingvisticheskikh uchenij* [History of Linguistic Studies]. Moscow, 1998. 368 p.
2. Buchilova I.A. Yazykovaya i etnicheskaya identichnost' detej-bilingvov, obuchayushchihsya v shkolah g. Cherepovca [Language and ethnic identity of children-bilinguals studying in schools of

the city of Cherepovets]. *Problemy porozhdeniya i vospriyatiya rechi* [Problems of generation and perception of speech: Proceedings of the XIV exit school-seminar]. Cherepovets, 2017, pp. 65–73.

3. Vyrenkova A.S., Polinskaya M.S., Rahilina E.V. Grammatika oshibok i grammatika konstrukcij: "ehritazhnyj" ("unasledovannyj") russkij yazyk [Grammar of errors and grammar of constructions: "erytazhny" ("inherited") Russian language]. *Voprosy yazykoznanija* [Questions of linguistics], 2014, no. 3, pp. 3–19.

4. Grudeva E.V. YAzykovoij korpus: apologiya i kritika [Language Corps: Apology and Criticism]. *Porozhdenie i vospriyatie rechi* [Generation and Perception of Speech]. Cherepovets, 2010, pp. 48–55.

5. Zyryanova E.V. Sistema CHILDES kak metod sbora materialov i izucheniya detskoj rechi [The system CHILDES as a method of collecting materials and studying children's speech]. *Izvestiya RGPU im. A.I. Gercena* [Izvestiya RSPU them. A.I. Herzen], 2008, no. 76-2. Available at: <http://cyberleninka.ru/article/n/sistema-childes-kak-metod-sbora-materialov-i-izucheniya-detskoj-rechi>.

6. Kopotev M.V. *Vvedenie v korpusnuyu lingvistiku* [Introduction to corpus linguistics]. Prague, 2014. 230 p.

7. Kopotev M.V., Gurin G.B. Razmetka sintaksicheskoi nepolnoty v korpusе [Marking of syntactic incompleteness in the case]. *Trudy mezhdunarodnoj konferencii "Dialog–2007"* [Proceedings of the international conference "Dialogue-2007"]. Available at: <https://goslide.org/Documents/razmetka-sintaksicheskoi-nepolnoty-v-korpusе-2007>.

8. Lysakova I.P. Russkij yazyk dlya migrantov: RKT ili RKN? (k voprosu o sodержanii terminov) [Russian for migrants: RCT or ILV? (to the question of the content of terms)]. *XXXVIII Mezhdunarodnaya filologicheskaya konferenciya. Vyp. 21: Russkij yazyk kak inostrannyj i metodika ego prepodavaniya* [XXXVIII International Philological Conference. Issue. 21: Russian as a Foreign Language and the Technique of Teaching]. St Petersburg, 2009, pp. 181–186.

9. Mal'ceva M.S. K opredeleniyu termina «lingvisticheskij korpus» i ego variantov [To the definition of the term "linguistic corpus" and its variants]. *Sotsial'no-ehkonomicheskie yavleniya i processy* [Socio-economic phenomena and processes], 2011, no. 7. Available at: <https://cyberleninka.ru/article/n/k-opredeleniyu-termina-lingvisticheskij-korpus-i-ego-variantov>.

10. Miroshnichenko E.A. K formal'nomu opredeleniyu ponyatiya «baza dannyh» [To the formal definition of the concept of "database"]. *Problemy informatiki* [Problems of Informatics], 2011, no. 2, pp. 83–87. Available at: <http://www.problem-info.ssc.ru/2011-2/12.pdf>.

11. Percov N.V. K suzheniyam o faktah russkogo yazyka v svete korpusnyh dannyh [To judgments about the facts of the Russian language in the light of corpus data]. *Russkij yazyk v nauchnom osveshchenii* [The Russian Language in Scientific Lighting], 2006, no. 1(11), pp. 227–245.

12. Plungyan V.A. Korpus kak instrument i kak ideologiya: o nekotoryh urokah sovremennoj korpusnoj lingvistiki [The hull as an instrument and as an ideology: about some lessons of modern case linguistics]. *Russkij yazyk v nauchnom osveshchenii* [Russian language in scientific coverage], 2008, no. 2 (16), pp. 7–20.

13. Rahilina E.V. O novyh instrumentah opisaniya russkoj grammatiki: korpus oshibok [On the new tools for describing Russian grammar: the body of errors]. *Simpozium «Russkaya grammatika 4.0»: Lingvistika* [Symposium "Russian Grammar 4.0": Linguistics]. Available at: http://chrome-extension://ecnpflgnajanjncmbpancdjoidceilk/content/web/viewer.html?source=extension_pdfhandler&file=http%3A%2F%2Fwww.rakhilina.ru%2Ffiles%2Frahilina_foreign_err.pdf.

14. Rahilina E.V. *Stepeni sravneniya v svete russkoj grammatiki oshibok* [Degrees of comparison in the light of the Russian grammar of errors]. Available at: http://www.rakhilina.ru/files/compar_err.pdf.

15. Sinkler Dzhon. *Predislovie k knige "Kak ispol'zovat' korpusa v prepodavanii inostrannogo yazyka"* [Foreword to the book "How to Use Corps in Teaching a Foreign Language"]. Available at:

http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4-child/texts/sinclair_introduction.pdf.

16. Frej A. *Grammatika oshibok* [The grammar of errors]. Komkniga, 2006. 304 p.
17. Shcherba L.V. O troyakom aspekte yazykovykh yavlenij i ob ehksperimente v yazykoznanii [On the Three-Dimensional aspect of Linguistic Phenomena and the Experiment in Linguistics]. *Yazykovaya sistema i rechevaya deyatel'nost'* [Language system and speech activity]. Leningrad, 1974, pp. 24–39.
18. Barðdal Jóhanna. *A Crash Course in Corpus Linguistics*. May. 2002. Linguistics Division, UNT.
19. Granger S. The learner corpus: a revolution in applied linguistics. *English Today*. 1994. 39 (10/3): 25-9.
20. McEnery T. *Corpus Linguistics*. Edinburgh, 1997.
21. Mustajoki A., Protassova E., Vakhtin N. *Instrumentarium of linguistics: sociolinguistic approach to non-standard Russian*. Helsinki, 2010.
22. Yukio Tono. *Learner corpora: design, development and applications*. Available at: <http://ucrel.lancs.ac.uk/publications/cl2003/papers/tono.pdf>.
23. Zevakhina N.A., Dzhakupova S.S. *Corpus of russian student texts: design and prospects*. Available at: <http://www.dialog-21.ru/media/1144/zevakhinanadzhakupovass.pdf>.

Список ресурсов

Корпус русских учебных текстов КРУТ (<https://ling.hse.ru/krut>)
Международная база данных детской речи CHILDEERS (<http://chil-des.psy.cmu.edu>)
Национальный корпус русского языка (www.ruscorpora.ru)
Русский учебный корпус RLC (<http://web-corpora.net/RLC/>)

Для цитирования: Грудева Е.В., Бучилова И.А., Волкова Н.А. Корпусы ошибок: целевая аудитория, возможная архитектура корпуса // Вестник Череповецкого государственного университета. 2018. №5(86). С. 63–72. DOI: 10.23859/1994-0637-2018-5-86-7

For citation: Grudeva E.V., Buchilova I.A., Volkova N.A. Corpora of errors: target audience, a possible architecture of the corpus. *Bulletin of the Cherepovets State University*, 2018, no. 5 (86), pp. 63–72. DOI: 10.23859/1994-0637-2018-5-86-7