

# СОВРЕМЕННЫЕ НАПРАВЛЕНИЯ ЛИНГВИСТИКИ



УДК 81\23. ББК Ш100.6.  
ГРНТИ 16.21.27. Код ВАК 10.02.19 (5.9.8)

## АКТУАЛЬНАЯ ГРАММАТИКА ДЕТСКОЙ РЕЧИ: КОРПУСНЫЕ ИССЛЕДОВАНИЯ

**Красноперова Е. С.**

Уральский государственный педагогический университет (Екатеринбург, Россия)

ORCID ID: <https://orcid.org/0000-0001-9675-1367>

*А н н о т а ц и я*. В статье рассматривается возможность создания корпуса детской речи и применения корпусных методов исследования в онтолингвистике. Корпус текстов рассматривается автором как инструмент квантитативного лингвистического исследования, что предполагает последовательное решение ряда задач: 1. Отбор материала, наиболее полно и сбалансированно представляющего языковое явление. 2. Определение необходимого объема текстов. 3. Создание разметки или системы разметок, инструмента, отличающего корпус от коллекций текстов, дневниковых записей и словарей. 4. Постобработка и создание корпуса в электронном виде. Расширение сферы корпусных исследований приводит к созданию корпусов детской речи. Показаны сложности составления корпусов детской речи, связанные с ее динамичностью, изменчивостью, индивидуальностью речевых стратегий детей, большим количеством в их речи различных инноваций. Продемонстрирована возможность составления электронного корпуса детской речи с грамматической разметкой на материале издания «Корпус детских высказываний» В. К. Харченко. Представлена авторская разработка такого корпуса с использованием программы AntConc и программы автоматизированной морфологической разметки MyStem. Создание корпуса включало в себя следующие этапы: 1. Перевод материала в электронный вид. 2. Обработка материала. 3. Выполнение автоматизированной грамматической разметки. 4. Постобработка. 5. Представление материала в удобном для дальнейших исследований виде. Представленный проект грамматического корпуса детской речи рассматривается как инструмент моделирования процесса усвоения и использования основных, первичных грамматических категорий. Некоторые грамматические закономерности, выделенные с помощью корпуса, позволят создать инструментарий для исследования грамматического компонента языковой способности ребенка, в частности, выявления актуального грамматикона языковой личности ребенка. Актуальный грамматикон, набор грамматических форм, на основе которых говорящий строит свою речь, может стать базой разработки новых методик обучения грамматике русского языка, в том числе как иностранного (неродного).

*К л ю ч е в ы е с л о в а*: речевой онтогенез; онтолингвистика; языковая личность; грамматикон; корпусные исследования; детская речь; развитие речи.

*Д л я ц и т и р о в а н и я*: Красноперова, Е. С. Актуальная грамматика детской речи: корпусные исследования / Е. С. Красноперова. – Текст : непосредственный // Филологический класс. – 2022. – Т. 27, № 1. – С. 87-100.

## REAL EVERYDAY GRAMMAR OF CHILDREN'S SPEECH: A CORPUS STUDY

**Evgeniya S. Krasnoperova**

Ural State Pedagogical University (Ekaterinburg, Russia)

ORCID ID: <https://orcid.org/0000-0001-9675-1367>

*A b s t r a c t*. The article discusses the possibilities of creating a corpus of children's speech and the use of corpus research in ontolinguistics. The corpus of texts is defined by the author as a tool for quantitative linguistic research, which presupposes a stepwise solution of a number of problems: 1. Selection of the material that most

fully and rationally represents the linguistic phenomenon. 2. Estimation of the required volume of texts. 3. Creation of a system of markup – a tool that distinguishes the corpus from a collection of texts, diaries and dictionaries. 4. Post-processing and creation of the corpus in electronic form. The expansion of corpus research leads to the emergence of children's speech corpora: CHILDES, INFANT.RU, CHILDRU, Emo.Child.Ru, StartWrit. The article demonstrates the difficulties of compiling corpora of children's speech associated with its dynamism, variability, individuality of children's speech strategies, and abundance of neologisms. An electronic corpus of children's speech with grammatical markup was compiled on the basis of the publication "Corpus of Children's Utterances" by V. K. Kharchenko. The corpus was created using the AntConc program and the MyStem morphological markup program. The creation of the corpus included the following stages: 1. Transfer of the material into electronic format. 2. Data processing. 3. Automatic grammatical markup. 4. Post-processing. 5. Presentation of the material in the form convenient for further study. The project of a grammatical corpus of children's speech presented in this study is considered as a tool for modeling the process of acquisition and use of the basic, initial grammatical categories. Some grammatical patterns, singled out with the help of the corpus, can make it possible to create tools for studying the grammatical component of the child's linguistic competencies, and specifically for revealing the real everyday grammar of the child's linguistic personality. The real everyday grammar itself – a set of grammatical forms on the basis of which the speaker constructs their speech – can become a stepping stone for creating new methods of teaching Russian grammar, especially when teaching Russian as a second language.

*Keywords:* speech ontogenesis; ontolinguistics; linguistic personality; real everyday grammar skills; corpus studies; children's speech; speech development.

*For citation:* Krasnoperova, E. S. (2022). Real Everyday Grammar of Children's Speech: A Corpus Study. In *Philological Class*. Vol. 27. No. 1, pp. 87-100.

*Введение: к постановке проблемы.*

Изучение детской речи в нашей стране имеет достаточно долгую историю: первые профессиональные записи детских высказываний и их лингвистическая интерпретация велись еще в 20-х годах XX века. Речь, конечно же, идет о работах А. Н. Гвоздева «От первых слов до первого класса. Дневник научных наблюдений» [Гвоздев 2005], представляющей систематическую запись фактов детской речи, и «Вопросы изучения детской речи», описывающей процесс усвоения языка ребенком [Гвоздев 2007]. Эти работы оказали влияние на следующие поколения онтолингвистов и стали образцом для дальнейших исследований: методы дневниковых записей остаются основными в современной онтолингвистике, например, в работе уже упоминавшейся научной школы онтолингвистики на базе СПбГУ [Речь русского ребенка. Звучащая хрестоматия 1994; От нуля до двух 1997; От двух до трех 1998; Две девочки: Соня и Надя 2001]. Собранные материалы становятся основой для словарей детской речи [Словарь детских словообразовательных инноваций 2001], для работ, посвященных процессу освоения языка, возрастной динамике этого процесса, индивидуальным стратегиям речевого развития в онтогенезе [Гридина 2016].

Современный этап развития онто-

лингвистики характеризуется расширением исследовательского инструментария. Для изучения детской речи и детского языкового сознания используются экспериментальные психолингвистические методы, с помощью которых исследуется процесс становления и развития лексикона, особенности мышления и доминанты детского сознания: «Лексикон младшего школьника (характеристика лексического компонента языковой компетенции)» И. Г. Овчинниковой, Н. И. Бересневой, Л. А. Дубровской, Е. Б. Пенягиной, «Ассоциативный словарь подростка» Е. Н. Гуц, «Возрастная психолингвистика: Толковый словарь русского языка глазами детей» А. Д. Палкина, «Мотивационный словарь детской речи» К. В. Гарганеевой, «Объяснительный словарь детских инноваций» Т. А. Гридиной [Овчинникова, Береснева 2000; Гуц 2004; Палкин 2004; Гарганеева 2007; Гридина 2012].

Расширение инструментария изучения детской речи проявилось и в возможности использования в этой области корпусных методов исследования. Появление корпусов текстов как инструмента количественного лингвистического исследования связано с проникновением в лингвистику математического инструментария и информационных технологий. Первые корпусы появились в США в 60-е гг. XX века.

Распространение информационных технологий и развитие корпусных исследований привели в 90-е и начале 2000-х годов к появлению масштабных проектов – корпусов национальных языков [Захаров 2016: 150; Майорова 2017: 43; Солнышкина, Гатиятулина 2020: 132]. В данный момент существуют американский (American National Corpus, ANC), британский (British National Corpus, BNC), российский (Национальный корпус русского языка, НКРЯ), немецкий (Deutsches Referenzkorpus, DeReKo), чешский (Český národní korpus, ČNK) и другие корпусы национальных языков, отличающиеся определенным составом текстов (репрезентацией различных форм национального и разных стилей литературного языка). Главной ценностью больших национальных корпусов является разметка их текстов, позволяющая осуществлять различные лингвистические исследования на большом объеме материала.

Развитие корпусных методов исследования коснулось и детской речи. Существуют корпусы детской речи на китайском, английском (американский и британский варианты), испанском, французском, немецком, датском и других национальных языках (см. обзор: [Claus, Rosales 2013: 2411; Radoslava 2016: 42-43]).

В международной базе данных CHILDES (Child Language Data Exchange System, <https://childes.talkbank.org>) собраны записи детской речи для более чем 40 языков. Проект создан в 1980-х гг. в Питсбургском университете для анализа разговорной речи, спонтанной речи детей и их родителей, а также для исследования процесса усвоения второго языка. Помимо значительного по объему массива записей, там содержится обширная библиография по психолингвистике, лингвистике, теории усвоения первого и второго языков, а также правила ввода материала и пакет программ для его анализа [Зырянова 2008: 115]. Доступ к материалам свободный, возможно размещение данных собственного исследования. К сожалению, банк данных русскоязычных детей невелик и пополняется медленно из-за необходимости предварительной обработки материала по правилам CHILDES. В данный момент в базе есть транскрибированные записи речи двух девочек. Записи речи первой девоч-

ки – семь 90-минутных записей – были сделаны в середине 1980-х гг. в Москве. Речь девочки фиксировала и расшифровывала ее мать, лингвист Е. Ю. Протасова. Второй ребенок – одноязычная девочка, живущая в русскоязычной семье в США в середине 1990-х гг. Записи сделаны и расшифрованы в рамках проекта Лаборатории детского языка факультета лингвистики Университета Коннектикута. Таким образом, при использовании этих записей в исследовательских целях лингвисты сталкиваются не только с ограниченным объемом материала, но и со сложностью сопоставления речевого опыта и речевого окружения детей.

На русском языке разработан еще ряд корпусов: корпус вокализаций и речи детей первых трех лет жизни INFANT.RU, корпус спонтанной и читаемой речи детей 4–7 лет CHILD.RU, корпус эмоциональной речи детей 4–7 лет Emo.Child.Ru. Все они созданы в рамках работы группы по изучению детской речи Санкт-Петербургского государственного университета [Ляксо, Фролова и др. 2017]. Эти аудиокорпусы, так же как и корпус детской письменной речи StartWrit [Ахапкина, Сосновцева 2017], не находятся в открытом доступе.

Введение в область онтолингвистики корпусных методов исследования позволяет перейти от оптики исключительного (от исследования детской инновации, по определению отклоняющейся от нормы, выделяющейся в потоке речи), от словарного метода работы (согласно которому словарные статьи организованы вокруг отдельных фактов, по той или иной причине привлекающих внимание исследователя), от интерпретации отдельного языкового факта, извлеченного из потока речи, к изучению детской речи «в потоке».

Сказанное выше позволяет утверждать, что корпусные методы исследований могут стать важным инструментом моделирования процесса усвоения и использования грамматики ребенком, динамической, функциональной системности детской речи: «язык не дается ребенку при рождении ..., но зато ему дана уникальная способность самостоятельно выстраивать в собственном сознании все компоненты языковой системы (фонологический, морфологический, синтаксический, лексиче-

ский и ряд иных), а также перестраивать и совершенствовать эту систему в течение всей жизни при взаимодействии с другими людьми» [Цейтлин 2009: 14].

Целостную интерпретацию актуальной грамматики детской речи в настоящий момент выполнить не представляется возможным, над этим активно работают ведущие онтолингвистические школы (в первую очередь – школа С. Н. Цейтлин), но в настоящее время описаны отдельные участки актуального грамматикона ребенка и стадии формирования грамматической способности. С нашей точки зрения, целостная интерпретация актуального грамматикона – дело будущего, что станет возможным после выявления его состава и механизмов формирования. Именно для этого создаются разные методики алгоритмизированного описания грамматикона языковой личности, в числе которых – предлагаемый нами подход «от корпуса» (больших данных) к моделированию грамматикона.

Объект данного исследования – актуальная грамматика детской речи.

Предмет – моделирование корпуса детской речи и разработка его аналога как инструмента изучения детского грамматикона.

Исследование грамматических закономерностей развития детской речи корпусными методами затруднено из-за отсутствия в открытом доступе готовых (размеченных) корпусов. Нами предпринята попытка создания пошагового алгоритма и аналога корпуса детской речи для исследования процесса усвоения грамматики в онтогенезе.

Решение поставленной проблемы предполагает следующие этапы:

1. Отбор материала. В состав корпуса должны попадать тексты, наиболее ярко представляющие языковое явление, изучаемое с помощью корпусных исследований.

2. Определение необходимого объема текстов. Для выявления статистически значимых закономерностей корпусные исследования проводятся на большом объеме текстов. Для каждого изучаемого явления объем определяется отдельно.

3. Создание разметки или системы разметок. Корпусом, а не коллекцией текстов материалы становятся при наличии разметки, которая позволяет использовать

корпусные методы исследования. С помощью разметки осуществляется навигация по тексту, проводится анализ лингвистических явлений, зафиксированных в записях детской речи, обнаруживающий как количественные, так и качественные закономерности. Разметка – это результат работы исследователя над текстом, квалификация текстов и элементов текста, присвоение выбранному элементу определенной информации. Разметка может быть экстралингвистической, то есть включать информацию о дате записи, поле и возрасте автора, обстоятельствах высказывания, например. Она может быть лингвистической: фонетической, морфологической, синтаксической, семантической [Захаров 2016: 153-154]. Один и тот же корпус текстов может содержать разные виды разметки.

4. Постобработка. Современное программное обеспечение позволяет автоматизировать часть операций при работе с корпусом, в том числе процедуру разметки. Однако результаты работы программы автоматизированной разметки требуют проверки и коррекции со стороны исследователя. Чем больше объем корпуса и чем более оригинален (далек от языковой нормы) его состав, тем более трудоемкой является стадия постобработки.

5. Создание корпуса в электронном виде. Объем современных корпусов, необходимость выявления статистических показателей, удобство перемещения по текстам – все это предполагает, что корпус создается и существует как массив данных в электронном виде.

*Предварительные замечания.*

Необходимо сделать несколько предварительных замечаний, касающихся сложности составления корпусов детской речи.

1. Состав корпуса. Основное требование, предъявляемое к корпусу, – это его репрезентативность, то есть способность воссоздать наиболее полную картину того языкового явления, которое он представляет. Для детской речи с ее динамичностью, изменчивостью, индивидуальностью речевых стратегий эта полнота труднодостижима. Важны следующие характеристики материала: массив текстов достаточного большого объема, регулярность записи детских высказываний, временной охват в несколько лет.

2. Создание разметки. Упомянутые выше динамичность, изменчивость, индивидуальность детской речи, большое количество в ней различных типов инноваций затрудняют автоматизацию процесса разметки и увеличивают время, необходимое для проверки результатов и постобработки.

*Цель составления корпуса.*

Изучение процесса формирования грамматического компонента языковой способности ребенка на основе лонгитюдных исследований детской речи позволило нам выявить генетически более ранние формы различных грамматических категорий, сохраняющие свою частотность на дальнейших этапах речевого развития: для существительных это формы именительного и винительного падежа единственного и множественного числа, для глаголов повелительного и изъявительного наклонения – формы 1-го и 3-го лица единственного числа, для прилагательных – формы мужского и женского рода [Красноперова 2016].

Отметим, что речь идет именно об отдельных формах, из постепенного и непрерывного с точки зрения академической грамматики усвоения которых складывается функционирование грамматических категорий. Усвоенные грамматические формы образуют в языковом сознании ребенка сменяющие друг друга функциональные системы, позволяющие ребенку осуществлять речевую деятельность и овладеть языком. Генетически первичные формы, усвоенные ребенком на первых этапах освоения языка, сохраняют свою частотность в речи в дальнейшем, этот факт позволяет нам выдвинуть гипотезу, что «генетически первичные грамматические формы оказываются ядерными в ассоциативно-вербальной сети говорящего» [Коновалова, Красноперова 2020: 11].

Указанные исследования проводились на небольшом объеме данных. Следующий шаг в подтверждении гипотезы – создание грамматического корпуса детской речи. Статистические данные о грамматических закономерностях детской речи, извлеченные из созданного нами корпуса, позволят выявить и уточнить состав генетически более ранних грамматических форм и категорий, выявить частотные грамматические формы и формы, находящиеся на периферии детской грамматической способ-

ности, усваиваемые поздно, остающиеся редкими в грамматиконе ребенка.

*Ход работы.*

Для проведения исследования нами был выбран корпус детских высказываний, созданный В. К. Харченко [Харченко 2012]. Преимущество этого материала в том, что высказывания одних и тех же детей фиксировались на протяжении семи лет наблюдения (2005–2011 гг.). Дети, речь которых фиксировалась: два брата с разницей в возрасте в два года, воспитывающиеся в одной семье. Записи производились регулярно, фиксировались как их спонтанная монологическая речь, так и диалоги между детьми и со взрослыми собеседниками. Таким образом, внешние условия, которые всегда необходимо учитывать при записи детской речи, в этом случае сводятся к минимуму, тогда как внутренние различия, в частности индивидуальные речевые стратегии и коммуникативные тактики, представляющие интерес для исследователя, доступны изучению.

Для представленного материала используется жанровое обозначение «корпус», предполагающее, что «в центре внимания составителя – высказывание (реплика)» [Харченко 2012: 9]. Также для корпуса важен большой объем материала, преимущественно такого, который «отличает детскую речь от речи взрослого носителя русского языка» [Харченко 2012: 9].

В данном корпусе используется авторская разметка нескольких видов:

– дата записи: *20 апреля 2008 г., 22 мая 2008 г.*;

– субъект высказывания: *Лев, Женя, Взрослый*;

– комментарий составителя к ситуации: *Шучу, О духоте, Строит тоннель в песочнице*;

– квалификация лингвистического явления: *Лев: Злата ну что! Каких-то-нибудь пирожков хватается (взрослый оборот)!.. А что? Уговаривай не уговаривай – всё равно не поест! (взрослый оборот)*;

– интерпретация высказывания: *Женя: Бай не! (Не спать!)*;

– ненормативное ударение: *У него нет коОтя... когтя...* [Харченко 2012: 223].

Произведенная В.К. Харченко разметка является выборочной и смешанной (лингвистической и экстралингвистиче-

ской). Этот источник позволяет исследовать детскую речь «как в динамике, по вертикали нарастающих изменений; так и в статике, по горизонтали “одновременно” усваиваемых слов, словоформ, конструкций» [Харченко 2012: 10], однако не является корпусом с точки зрения корпусных исследований. Создание корпуса на основе этого источника требует перевода материалов в электронный вид, создания сплошной лингвистической разметки и постобработки.

Возможная разметка:

Фонетическая.

Речь ребенка фиксировалась и включалась в этот корпус с момента достижения информантом двухлетнего возраста, соответственно, процесс усвоения звукового строя языка не являлся основной целью наблюдения. Хотя этот корпус не стремится фиксировать все фонетические особенности речи конкретного ребенка, однако позволяет отметить некоторые моменты, связанные со становлением ее звуковой стороны, например, ненормативное ударение: *жУкадай! слЕплю! догОню! ЯпЕю! У него нет кОгтя... когТЯ...*

Лексическая.

В рамках этой разметки могут быть маркированы слова «языка нянь»: *кака, нянь, баба*; слова разных тематических групп: термины родства – *баб, папа, мама, дядя*; имена собственные – *Лёва, Вера, Наташа, Денис*; прецедентные единицы – *Позвал дедка Жучку. Жучка за кошку, кошка за собачку... Чем решил стать Бараш, чтоб всех покротИть?* и т. д.

Семантическая.

На материалах корпуса можно проследить изменение значения слова. Расширение значения: *Я свел всю воду! Я скушал всю воду!* Усвоение переносного значения: *Всё, дождик упал! Не идёт дождик!*

Грамматическая.

Этот тип разметки предполагает присвоение каждой единице корпуса информации о частеречной принадлежности, лексико-грамматическом классе, морфологических признаках. Такая разметка чаще всего выполняется автоматизировано, с помощью программ для морфологического анализа текста. В рамках данного исследования будет выполнена именно грамматическая разметка корпуса детских высказываний.

Пошаговый алгоритм создания грамматического корпуса:

1. Перевод материала в электронный вид.
2. Обработка материала.
3. Выполнение автоматизированной грамматической разметки.
4. Постобработка.
5. Представление материала в удобном для дальнейших исследований виде.

*Реализация проекта «Корпус детской речи».*

Реализация проекта продолжается, расширяется материал, попадающий в корпус, уточняются результаты разметки, собираются статистические данные по уже обработанному материалу. К сожалению, количество имеющегося в нашем распоряжении материала не позволяет зафиксировать очень нестабильную, постоянно меняющуюся функциональную систему языка ребенка, например в возрасте от двух до трех лет, то есть нет возможности описать ряд таких систем, сменяющих друг друга за год жизни. Решением этой проблемы может стать расширение источников материала детской речи. Пока доступа к таким данным нет, для целей нашего исследования создается аналог корпуса детской речи, его проект.

1. Перевод материала в электронный вид.

Цель данного этапа – подготовка имеющихся материалов к обработке программными средствами. Для этого необходимо выделить из всего массива зафиксированного материала высказывания конкретного ребенка за определенный возрастной период. Наше исследование сосредоточено на речи детей от двух до пяти лет. Соответственно, составлены три массива записей речи старшего брата (Лев) в возрасте от 2 до 3 лет, от 3 до 4 лет и от 5 до 6 лет и три массива записей речи младшего брата (Женя) в те же возрастные периоды. Записи детской речи за нужный период извлекаются из источника и очищаются от авторской разметки. Продемонстрируем процесс на небольшом фрагменте.

Записи детской речи в разметке  
В. К. Харченко:

МАРТ

4 марта 2006 г.

– *Здесь ходить с тобой будем. Лев подхватывает: Танцы устроивать.*

8 марта 2006 г.

*В эту неделю очень часто говорит: десять тысяч.*

Гуляли, было солнце. Я: Солнце, весна! – Летняя зима!

А 8 марта всех женщин день рождения?  
11 марта 2006 г.

– Где у тебя собачка железная со своим собачонком стеклянным?

– Здесь поместится трёх зверей! (Верх пирамидки) Ой! Ну, толкни вот эту!

– Верблюжье одеяло! Этим одеялом верблюда накрывают!

О подсветке: О! Подсвет! Рыбки на месте!

– А железо не очень хрупкое, вообще не хрупкое! [Харченко 2012: 73].

Записи речи, подготовленные для корпуса и последующей грамматической разметки:

Танцы устраивать.

Десять тысяч.

Летняя зима!

А Восьмого марта у всех женщин день рождения?

Где у тебя собачка железная со своим собачонком стеклянным?

Здесь поместится трёх зверей!

Ой! Ну, толкни вот эту!

Верблюжье одеяло! Этим одеялом верблюда накрывают!

О! Подсвет! Рыбки на месте!

А железо не очень хрупкое, вообще не хрупкое!

2. Обработка материала.

На данном этапе осуществляется обработка материала с помощью программы AntConc, мультиплатформенного инструмента для проведения лингвистических исследований. Возможности программы включают в себя поиск контекстов конкретного слова, поиск кластеров и N-грамм по заданным условиям, подсчет количества словоупотреблений, представление слов в виде упорядоченного по частоте списка.

Важность этого функционала для грамматического исследования объясняется тем, что грамматические формы усваиваются детьми экземплярно, в виде конкретных словоформ. Частотные списки слов позволяют выявить как те лексические единицы, которые могут служить удобной основой для усвоения грамматического значения и механизма образования грамматической формы, так и те единицы, которые являются низкочастотными, но вместе с тем могут выступать основными репрезентантами периферийных грамматических форм.

Текст, обработанный в программе AntConc, выводится в виде частотного списка слов, представленного в таблице 1.

Таблица 1

#Word Types: 47

#Word Tokens: 51

ранг	частотность	слово	ранг	частотность	слово	ранг	частотность	слово
1	2	а	17	1	женщин	33	1	поместится
2	2	не	18	1	зверей	34	1	рождения
3	2	у	19	1	здесь	35	1	рыбки
4	2	хрупкое	20	1	зима	36	1	своим
5	1	верблюда	21	1	летняя	37	1	со
6	1	верблюжье	22	1	марта	38	1	собачка
7	1	вообще	23	1	месте	39	1	собачонком
8	1	восьмого	24	1	на	40	1	стеклянным
9	1	вот	25	1	накрывают	41	1	танцы
10	1	всех	26	1	ну	42	1	тебя
11	1	втолкни	27	1	о	43	1	трёх
12	1	где	28	1	одеяло	44	1	тысяч
13	1	день	29	1	одеялом	45	1	устраивать
14	1	десять	30	1	ой	46	1	этим
15	1	железная	31	1	очень	47	1	эту
16	1	железо	32	1	подсвет			

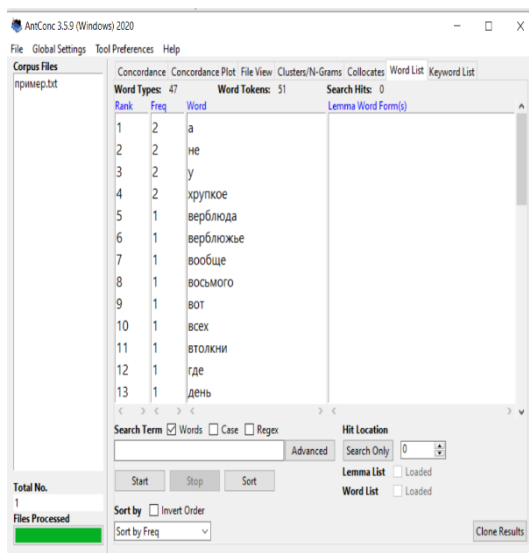
Программа подсчитывает количество слов (Word Types) и количество словоупо-

треблений (Word Tokens), но рассматривает разные словоформы одной лексемы как

разные слова (*одеяло* и *одеялом* посчитаны как два слова), таким образом, в представленном фрагменте 46, а не 47 лексем. Данная особенность программы была учтена

на этапе количественной обработки материала.

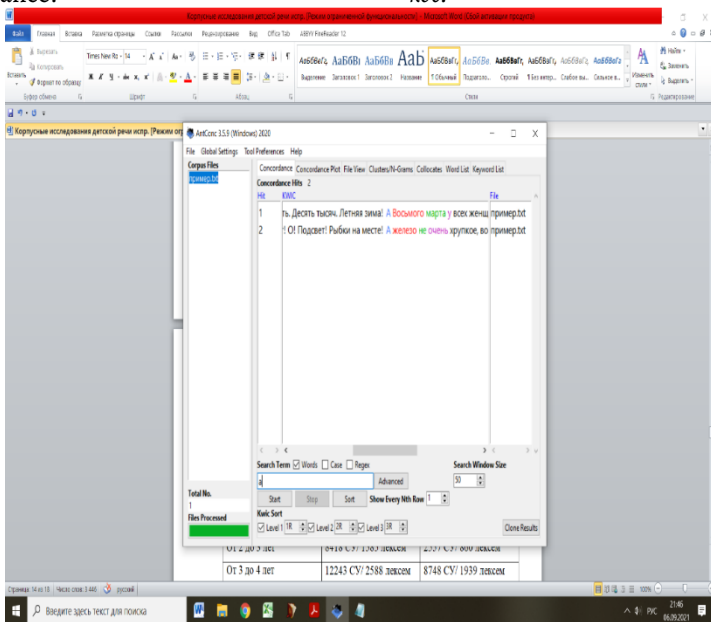
Рабочее окно программы AntConc:



Программа позволяет осуществлять навигацию по тексту и выстраивать контексты употребления конкретного слова на вкладке Concordance:

*А Восьмого марта у всех женщин день рождения?*

*А железо не очень хрупкое, вообще не хрупкое!*



На данном этапе работы размер корпуса составляет более 56 000 словоупотреблений (СУ), около 7 000 различных лемм.

В таблице 2 представлен размер лексикона двух детей в возрасте от двух до пяти лет.



Таблица 2

Возраст	Лев	Женя
От 2 до 3 лет	8 418 СУ / 1 585 лемм	2 557 СУ / 860 лемм
От 3 до 4 лет	12 243 СУ / 2 588 лемм	8 748 СУ / 1 939 лемм
От 4 до 5 лет	13 340 СУ / 2 866 лемм	11 255 СУ / 2 231 лемм

## 3. Грамматическая разметка.

На этапе разметки происходит квалификация грамматических форм. Разметка выполняется автоматизировано с помощью программы MyStem (<https://yandex.ru/dev/mystem/>), которая проводит морфологический разбор всех единиц корпуса.

Фрагмент речи, размеченный программой MyStem.

Танцы{танец=S,муж,неод=(вин,мн|им,мн)} устроить{устроить?=V,несов=инф|устроить?=V,несов=инф}.

Десять{десять=NUM=(вин|им)} тысяча{тысяча=S,жен,неод=род,мн}.

Летняя{летний=A=им,ед,полн,жен} зима{зима=S,жен,неод=им,ед}!

А{a=CONJ=} 8 марта{март=S,муж,неод=род,ед} у{y=PR=} всех{весь=APRO=(пр,мн|род,мн|вин,мн,од)} женщин{женщина=S,жен,од=(вин,мн|род,мн)} день{день=S,муж,неод=(вин,ед|им,ед)} рождения{рождение=S,сред,неод=(вин,мн|род,ед|им,мн)}?

Где{где=ADVPRO=} у{y=PR=} тебя{ты=SPRO,ед,2-л=(вин|род)} собачка{собачка=S,жен,од=им,ед} железная{железный=A=им,ед,полн,жен} со{so=PR=} своим{свой=APRO=(дат,мн|твор,ед,муж|твор,ед,сред)} собачонком{собачонок=S,муж,од=твор,ед} стеклянным{стеклянный=A=(дат,мн,полн|твор,ед,полн,муж|твор,ед,полн,сред)}?

Здесь{здесь=ADVPRO=} поместится{помещаться=V,нп=непрош,ед,изъяв,3-л,сов} трёх{три=NUM=(пр|род|вин,од)} зверей{зверь=S,муж,од=(вин,мн|род,мн)}!

Ой{ой=INTJ=}! Ну{ну=PART=}, втолкни{вталкивать=V,не=ед,пов,2-л,сов} вот{вот=PART=} эту{этом=APRO=вин,ед,жен}!

Верблюжье{верблюжий=A,полн=(вин,ед,сред|им,ед,сред)} одеяло{одеяло=S,сред,неод=(вин,ед|им,ед)}!

Этим{этом=APRO=(дат,мн|твор,ед,муж|твор,ед,сред)} одеялом{одеяло=S,сред,неод=твор,ед} верблюда{верблюду=S,муж,од=(вин,ед|род,ед)} накрывают{накрывать=V=непрош,мн,изъяв,3-л,несов,не}!

О{o=INTJ=}! Подсвет{подсвет=S,муж,неод=(вин,ед|им,ед)}! Рыбки{рыбка=S,жен,од=

(род,ед|им,мн)} на{на=PR=} месте{место=S,сред,неод=пр,ед}!

А{a=CONJ=} железо{железо=S,сред,неод=(вин,ед|им,ед)} не{не=PART=} очень{очень=ADV=} хрупкое{хрупкий=A=(вин,ед,полн,сред|им,ед,полн,сред)}, вообще{вообще=ADV,водн=} не{не=PART=} хрупкое{хрупкий=A=(вин,ед,полн,сред|им,ед,полн,сред)}!

Полученные данные нужны для выделения грамматических закономерностей на определенном массиве текстов. Программа присваивает каждой словоформе грамматический тэг, из которого исследователь извлекает информацию для дальнейшей (на первом этапе – количественной) обработки.

## 4. Постобработка.

Первый этап постобработки – технический, предполагающий снятие грамматической неоднозначности грамматических форм. Для омонимичных грамматических форм программа разметки дает варианты разбора, из которых исследователю необходимо выбрать верный.

Покажем это на примере словоформы «хрупкое», которая может быть квалифицирована как форма именительного или винительного падежа единственного числа прилагательного «хрупкий» (хрупкое {хрупкий=A=(вин,ед,полн,сред|им,ед,полн,сред)}). Только обращение к контексту (А железо не очень хрупкое, вообще не хрупкое!) дает возможность снять омонимию форм: хрупкое{хрупкий=A=им,ед,полн,сред}.

Второй этап постобработки представляет большой интерес для исследователя. Детская речь, как известно, характеризуется значительным количеством инноваций, в первую очередь семантических и мотивационных:

Летняя зима!

А Восьмого марта у всех женщин день рождения?

Верблюжье одеяло! Этим одеялом верблюда накрывают!

Собственно грамматические инновации детской речи могут затруднять работу

морфологического анализатора. Приведем пример: *Хочу всё разобрОсил. Всё разбрОсил! Она вЫгружила, она вЫгружила... Нет, я растЮ!* Такие инновации могут квалифицироваться верно: «разбросил» как глагольная форма совершенного вида (*разбрОсил* {разбрасывать= V=прош,ед,изъяв,муж,сов}), «разбросить» как инфинитив (*разбрОсит*{разбрасывать=V=инф,сов}). Квалификация может быть неверной: «выгрузила» как глагольная форма несовершенного вида, а не совершенного вида, восстановление начальной формы вызывает у программы затруднение (*вЫгружила*{выгрузить?=V,несов,пе=прош,ед,изъяв,жен}). Форма может не распознаваться как глагольная и нуждаться в постпроверке (например, «растЮ» определяется не как глагол, а как потенциальное существительное «растя» (*растЮ*{растя?=S,имя,жен,од=вин,ед}).

Перед исследователем в таком случае

стоит практическая цель: исправление погрешностей работы программы, квалификация грамматических форм вручную. Однако интерес представляет и возможность провести анализ таких форм. Составление списков грамматических форм, отклоняющихся от нормы, т. е. выделение из текста грамматических инноваций автоматизированным способом, само по себе может послужить основанием для отдельного исследования механизмов усвоения грамматических форм, материальных показателей, способов формообразования.

5. Представление материала в удобном для дальнейших исследований виде.

Один из таких вариантов представления данных приведен в таблице 3<sup>1</sup>.

<sup>1</sup> Рубрикация *словоформа, лемма, часть речи* используется в соответствии с традицией, принятой в НКРЯ.

Таблица 3

#Word Types: 46

#Word Tokens: 51

ранг	частотность	словоформа	лемма	часть речи	грамматические признаки
1	2	а	а	PART	
2	2	не	не	PART	
3	2	у	у	PR	
4	2	хрупкое	хрупкий	A	им,ед,полн,сред
5	1	верблюда	верблюд	S	муж,од=вин,ед
6	1	верблюжье	верблюжий	A	полн=им,ед,сред
7	1	вообще	вообще	ADV	вводн
8	1	восьмого	восьмой	ANUM	род,ед,муж
9	1	вот	вот	PART	
10	1	всех	все	SPRO	мн=род
11	1	втолкни	вталкивать	V	пе=ед,пов,2-л,сов
12	1	где	где	ADVPRO	
13	1	день	день	S	муж,неод=им,ед
14	1	десять	десять	NUM	им
15	1	железная	железный	A	им,ед,полн,жен
16	1	железо	железо	S	сред,неод=им,ед
17	1	женщин	женщина	S	жен,од=род,мн
18	1	зверей	зверь	S	муж,од=вин,мн
19	1	здесь	здесь	ADVPRO	
20	1	зима	зима	S	жен,неод=им,ед
21	1	летняя	летний	A	им,ед,полн,жен

Продолжение таблицы 3

ранг	частотность	словоформа	лемма	часть речи	грамматические признаки
22	1	марта	март	S	муж,неод=род,ед
23	1	месте	место	S	сред,неод=пр,ед
24	1	на	на	PR	
25	1	накрывают	накрывать	V	непрош,мн,изъяв,3-л,несов,пе
26	1	ну	ну	PART	
27	1	о	о	INTJ	
28	1	одеяло	одеяло	S	сред,неод=им,ед
29	1	одеялом	одеяло	S	сред,неод=твор,ед
30	1	ой	ой	INTJ	
31	1	очень	очень	ADV	
32	1	подсвет	подсвет	S	муж,неод=им,ед
33	1	поместится	помещаться	V	нп=непрош,ед,изъяв,3-л,сов
34	1	рождения	рождение	S	сред,неод=род,ед
35	1	рыбки	рыбка	S	жен,од=им,мн
36	1	своим	свой	APRO	твор,ед,муж
37	1	со	со	PR	
38	1	собачка	собачка	S	жен,од=им,ед
39	1	собачонком	собачонок	S	муж,од=твор,ед
40	1	стеклянным	стеклянный	A	твор,ед,полн,муж
41	1	танцы	танец	S	муж,неод=вин,мн
42	1	тебя	ты	SPRO	ед,2-л=род
43	1	трёх	три	NUM	вин,од
44	1	тысяч	тысяча	S	жен,неод=род,мн
45	1	устроить	устроить	V	несов=инф
46	1	этим	это	SPRO	ед,сред,неод=твор
47	1	эту	этот	APRO	вин,ед,жен

В результате нашей работы должны быть получены шесть подкорпусов (различные по субъекту и возрасту) в составе единого корпуса с грамматической разметкой, снятой грамматической омонимией, проставленной частотой словоупотреблений, возможностью моделирования грамматического профиля ребенка.

#### *Заключение.*

Грамматикон не сводится к набору грамматических форм [Караулов 2010: 5], более того, грамматикон не может быть описан только как последовательность наборов грамматических форм, сменяющих друг друга. В теории грамматикона, разработанной Ю. Н. Карауловым, центральным элементом являются способ организации грамматической системы, связи, соединяющие ее единицы, закономерности их взаимодействия. Организацию грамматической системы, работу этих взаимосвязей Ю. Н. Караулов демонстрировал на матери-

але масштабного ассоциативного эксперимента [Караулов 1993]. Еще одним подходом к моделированию процесса усвоения и использования грамматики может быть исследование детской речи периода усвоения основных, первичных грамматических категорий. Некоторые грамматические закономерности, выделенные в этот период, позволяют приблизиться к созданию модели грамматикона с учетом некоторых предварительных замечаний:

- выявленные закономерности интерпретируются с учетом данных ассоциативной грамматики Ю. Н. Караулова;
- закономерности выделяются на материале детской речи, с одной стороны, спонтанной, обыденной, с другой – максимально обнажающей механизмы усвоения языка (формирования грамматикона в частности). Эти рабочие механизмы еще не скрыты за поверхностными структурами усвоенного языка;

– эти закономерности статистически значимы благодаря объему однотипного текста (речи, порожденной в определенных, стабильных условиях), с одной стороны, с другой – охватывают некоторые индивидуальные стратегии двух разных детей.

Таким образом, корпус детской речи может стать одним из инструментов моде-

лирования актуального грамматикона языковой личности ребенка в возрасте от двух до пяти лет. В свою очередь актуальный грамматикон, набор грамматических форм, на основе которых говорящий строит свою речь, может послужить базой для воплощения в практике обучения русскому языку идей актуальной грамматики.

### Литература

- Ахапкина, Я. Э. Корпус детской письменной речи: StartWrit / Я. Э. Ахапкина, Е. Г. Сосновцева. – Текст : электронный // Проблемы онтолингвистики. – 2017. – URL: <https://www.elibrary.ru/item.asp?id=30048197> (дата обращения: 09.04.2021).
- Гарганеева, К. В. Мотивационный словарь детской речи / К. В. Гарганеева. – Томск : Изд-во Томского ун-та, 2007. – 120 с.
- Гвоздев, А. Н. Вопросы изучения детской речи / А. Н. Гвоздев. – СПб.: ДетствоПресс, 2007. – 472 с.
- Гвоздев, А. Н. От первых слов до первого класса. Дневник научных наблюдений / А. Н. Гвоздев. – М. : КомКнига, 2005. – 320 с.
- Гридина, Т. А. Актуальная грамматика детской речи: когнитивные модели семантизации и механизмы порождения словоформ / Т. А. Гридина // Языковая личность: аспекты изучения : сборник научных статей памяти члена-корреспондента РАН Юрия Николаевича Караулова. – Москва : МАКС Пресс, 2017. – С. 110-122.
- Гридина, Т. А. Объяснительный словарь детских инноваций : монография и словарь / Т. А. Гридина. – Екатеринбург, 2012. – 202 с.
- Гридина, Т. А. Онтолингвистика. Язык в зеркале детской речи : учеб. пособие / Т. А. Гридина. – Москва : Флинта, 2013. – 152 с.
- Гридина, Т. А. «Своя игра»: ребенок в мире языка : монография / Т. А. Гридина. – Екатеринбург, 2016. – 177 с.
- Гуц, Е. Н. Ассоциативный словарь подростка / Е. Н. Гуц. – Омск : Ом. гос. ун-т, 2004. – 153 с.
- Две девочки: Соня и Надя: дневниковые записи / сост. О. А. Юнтунен, О. Б. Сизова ; отв. ред. С. Н. Цейтлин. – СПб. : Тускарора, 2001. – 191 с.
- Захаров, В. П. Прологомены к корпусной лингвистике / В. П. Захаров. – Текст : электронный // Вопросы психолингвистики. – 2016. – №2 (28). – URL: <https://cyberleninka.ru/article/n/prolegomeny-k-korpusnoy-lingvistike> (дата обращения: 06.04.2021).
- Зырянова, Е. В. Система CHILDES как метод сбора материалов и изучения детской речи / Е. В. Зырянова. – Текст : электронный // Известия РГПУ им. А. И. Герцена. – 2008. – № 76-2. – URL: <https://cyberleninka.ru/article/n/sistema-childes-kak-metod-sbora-materialov-i-izucheniya-detskoj-rechi> (дата обращения: 06.04.2021).
- Караулов, Ю. Н. Ассоциативная грамматика русского языка / Ю. Н. Караулов. – Москва : Русский язык, 1993. – 330 с.
- Караулов, Ю. Н. Русский язык и языковая личность / Ю. Н. Караулов. – Москва : Издательство ЛКИ, 2010. – 264 с.
- Коновалова, Н. И. Актуальный грамматикон языковой личности: проблема моделирования / Н. И. Коновалова, Е. С. Красноперова // Вестник Московского государственного областного университета. Серия: Русская филология. – 2020. – № 3. – С. 6-14.
- Красноперова, Е. С. Усвоение грамматических форм в дошкольном онтогенезе / Е. С. Красноперова // LINGUISTICA JUVENIS. – 2016. – Вып. 18. – С. 69-76.
- Ляко, Е. Е. Корпуса детской речи «infant.Ru», «infant.MAVS», «child.Ru», «EmoChildRu» на материале русского языка и их использование в исследованиях речевого онтогенеза / Е. Е. Ляко, О. В. Фролова, А. С. Григорьев, А. В. Остроухов. – Текст : электронный // Теоретическая и прикладная лингвистика. – 2017. – Т. 3, № 1. – С. 28-58. – URL: <https://www.elibrary.ru/item.asp?id=29161390> (дата обращения: 09.04.2021).
- Майорова, А. Д. Корпусная лингвистика: исторический и лингводидактический аспекты / А. Д. Майорова. – Текст : электронный // МНИЖ. – 2017. – № 5-2 (59). URL: <https://cyberleninka.ru/article/n/korpusnaya-lingvistika-istoricheskij-i-lingvodidakticheskij-aspekty> (дата обращения: 06.04.2021).
- От двух до трех: дневниковые записи / сост. С. Н. Цейтлин, М. Б. Елисеева. – СПб. : Изд-во РГПУ им. А. И. Герцена, 1998. – 159 с.
- От нуля до двух: дневниковые записи / сост. С. Н. Цейтлин. – СПб. : Изд-во РГПУ им. А. И. Герцена, 1997. – 215 с.
- Овчинникова, И. Г. Лексикон младшего школьника (характеристика лексического компонента языковой компетенции) / И. Г. Овчинникова, Н. И. Береснева, Л. А. Дубровская, Е. Б. Пенягина. – Пермь, 2000. – 312 с.
- Палкин, А. Д. Возрастная психолингвистика: Толковый словарь русского языка глазами детей / А. Д. Палкин. – М., 2004. – 360 с.

- Речь русского ребенка: Звучащая хрестоматия / сост.: Т. В. Кузьмина, Э. И. Столярова, С. Н. Цейтлин. – СПб. : Бохум, 1994. – 125 с.
- Ринчинов, О. С. Диахронический корпус бурятского языка как цифровой инструмент исторических исследований: подходы, решения, экспериментальные исследования / О. С. Ринчинов. – Текст : электронный // Историческая информатика. – 2020. – № 2. – URL: <https://cyberleninka.ru/article/n/diachronicheskiy-korpus-buryatskogo-yazyka-kak-tsifrovoy-instrument-istoricheskikh-issledovaniy-podhody-resheniya-eksperimentalnye> (дата обращения: 06.04.2021).
- Рубцова, Ю. В. Построение корпуса текстов для настройки тонового классификатора / Ю. В. Рубцова // Программные продукты и системы. – 2015. – № 1 (109). – С. 72-78.
- Солнышкина, М. И. История развития корпусной лингвистики (на примере англоязычных корпусов) / М. И. Солнышкина, Г. М. Гатиятуллина. – Текст : электронный // Вестник Томского государственного университета. Филология. – 2020. – № 63. – URL: <https://cyberleninka.ru/article/n/istoriya-razvitiya-korpusnoy-lingvistiki-na-primere-angloyazychnyh-korpusov> (дата обращения: 06.04.2021).
- Харченко, В. К. Корпус детских высказываний / В. К. Харченко. – М. : Изд-во Литературного института им. А. М. Горького, 2012. – 520 с.
- Цейтлин, С. Н. Словарь детских словообразовательных инноваций / С. Н. Цейтлин. – Munchen, 2001. – 202 с.
- Цейтлин, С. Н. Очерки по словообразованию и формообразованию в детской речи / С. Н. Цейтлин. – М. : Знак, 2009. – 592 с.
- Claus, F. A survey about databases of children's speech / F. Claus, H. G. Rosales, R. Petrick, H. Hain, R. Hoffmann. – Text : electronic // INTERSPEECH. – 2013. – URL: [https://www.researchgate.net/publication/283719927\\_A\\_Survey\\_about\\_Databases\\_of\\_Children's\\_Speech](https://www.researchgate.net/publication/283719927_A_Survey_about_Databases_of_Children's_Speech) (mode of access: 06.04.2021).
- Radoslava, S. K. Design and development a children's speech database / S. K. Radoslava. – Text : electronic // Computer Science. – 2016. – URL: [https://www.researchgate.net/publication/280917631\\_Design\\_and\\_development\\_a\\_children's\\_speech\\_database](https://www.researchgate.net/publication/280917631_Design_and_development_a_children's_speech_database) (mode of access: 06.04.2021).

## References

- Akhapkina, Ya. E., Sosnovtseva, E. G. (2017). Korpus detskoj pis'mennoj rechi: StartWrit [Corpus of Children's Written Speech: StartWrit]. In *Problemy ontolingvistiki*. URL: <https://www.elibrary.ru/item.asp?id=30048197> (mode of access: 04.09.2021).
- Claus, F., Rosales, H. G., Petrick, R., Hain, H., Hoffmann, R. (2013). A Survey about Databases of Children's Speech. In *INTER\_SPEECH*. URL: [https://www.researchgate.net/publication/283719927\\_A\\_Survey\\_about\\_Databases\\_of\\_Children's\\_Speech](https://www.researchgate.net/publication/283719927_A_Survey_about_Databases_of_Children's_Speech) (mode of access: 06.04.2021).
- Garganeyeva, K. V. (2007). *Motivatsionnyi slovar' detskoj rechi* [Motivational Dictionary of Children's Speech]. Tomsk, Izdatel'stvo Tomskogo universiteta. 120 p.
- Gridina, T. A. (2012). *Ob'yasnitel'nyi slovar' detskikh innovatsii* [Explanatory Dictionary of Children's Innovations]. Ekaterinburg. 202 p.
- Gridina, T. A. (2013). *Ontolingvistika. Yazyk v zerkale detskoj rechi* [Ontolinguistics. Language in the Mirror of Children's Speech]. Moscow. 152 p.
- Gridina, T. A. (2016). «Svoya igra»: *rebenok v mire yazyka* ["Own Game": A Child in the World of Language]. Ekaterinburg. 177 p.
- Gridina, T. A. (2017). Aktual'naya grammatika detskoj rechi: kognitivnye modeli semantizatsii i mekhanizmu porozhdeniya slovoform [Actual Grammar of Children's Speech: Cognitive Models of Semantization and Mechanisms of Generation of Word Forms]. In *Yazykovaya lichnost': aspekty izucheniya: sbornik nauchnykh statej pamyati chlena-korrespondenta RAN Yuriya Nikolaevicha Karaulova*. Moscow, pp. 110-122.
- Guts, E. N. (2004). *Assotsiativnyi slovar' podrostka* [Associative Dictionary of a Teenager]. Omsk, Omskii gosudarstvennyi universitet. 153 p.
- Gvozdev, A. N. (2005). *Ot pervykh slov do pervogo klassa. Dnevnik nauchnykh nablyudenii* [From the First Words to the First Class. Diary of Scientific Observations]. Moscow. 320 p.
- Gvozdev, A. N. (2007). *Voprosy izucheniya detskoj rechi* [Questions of Studying Children's Speech]. Saint Petersburg. 472 p.
- Karaulov, Yu. N. (1993). *Assotsiativnaya grammatika russkogo yazyka* [Associative Grammar of the Russian Language]. Moscow. 330 p.
- Karaulov, Yu. N. (2010). *Russkii yazyk I yazykovaya lichnost'* [Russian Language and Linguistic Personality]. Moscow. 264 p.
- Kharchenko, V. K. (2012). *Korpus detskikh vyskazyvaniy* [Corpus of Children's Utterances]. Moscow. 520 p.
- Kononova, N. I., Krasnoperova, E. S. (2020). Aktual'nyi grammatikon yazykovoj lichnosti: problema modelirovaniya [Actual Grammaticon of the Linguistic Personality: the Problem of Modeling]. In *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Russkaya filologiya*. No. 3, pp. 6-14.
- Krasnoperova, E. S. (2016). Usvoenie grammaticheskikh form v doskol'nom ontogeneze [Assimilation of Grammatical Forms in Preschool Ontogenesis]. In *LINGUISTICA JUVENIS*. Issue 18, pp. 69-76.

Kuz'mina, T. V., Stolyarova, E. I., Tseytlin, S. N. (1994). *Rech' russkogo rebenka: Zvuchashchaya khrestomatiya* [Speech of a Russian Child: Sounding Reader]. Saint Petersburg, Bokhum. 125 p.

Lyakso, E. E., Frolova, O. V., Grigor'ev, A. S., Ostroukhov, A. V. (2017). Korpusa detskoj rechi «infant.Ru», «infant.MAVS», «child.Ru», «EmoChildRu» na material russkogo yazyka i ikh ispol'zovanie v issledovaniyakh rechevogo ontogeneza [Corpus of Children's Speech "infant.Ru", "infant.MAVS", "child.Ru", "EmoChildRu" on the Material of the Russian Language and Their Use in Studies of Speech Ontogenesis]. In *Teoreticheskaya i prikladnaya lingvistika*. Vol. 3. No. 1, pp. 28-58. URL: <https://www.elibrary.ru/item.asp?id=29161390> (mode of access: 04.09.2021).

Maiorova, A. D. (2017). Korpurnaya lingvistika: istoricheskii i lingvodidakticheskii aspekty [Corpus Linguistics: Historical and Linguodidactic Aspects]. In *MNIZh*. No. 5-2 (59). URL: <https://cyberleninka.ru/article/n/korpurnaya-lingvistika-istoricheskii-i-lingvodidakticheskii-aspekty> (mode of access: 04.06.2021).

Ovchinnikova, I. G., Beresneva, N. I., Dubrovskaya, L. A., Penyagina, E. B. (2000). *Leksikon mladshogo shkol'nika (kharakteristika leksicheskogo komponenta yazykovoi kompetentsii)* [Lexicon of a Junior Pupils (Characteristics of the Lexical Component of Language Competence)]. Perm. 312 p.

Palkin, A. D. (2004). *Vozrastnaya psikholingvistika: Tolkovyi slovar' russkogo yazyka glazami detei* [Age-Related Psycholinguistics: Explanatory Dictionary of the Russian Language through the Eyes of Children]. Moscow. 360 p.

Radoslava, S. K. (2016). Design and Development a Children's Speech Database. In *Computer Science*. URL: [https://www.researchgate.net/publication/280917631\\_Design\\_and\\_development\\_a\\_childrens\\_speech\\_database](https://www.researchgate.net/publication/280917631_Design_and_development_a_childrens_speech_database) (mode of access: 06.04.2021).

Rinchinov, O. S. (2020). Diakhronicheskii korpus buryatskogo yazyka kak tsifrovoi instrument istoricheskikh issledovaniy: podkhody, resheniya, eksperimentalnye issledovaniya [Diachronic Corpus of the Buryat Language as a Digital Instrument of Historical Research: Approaches, Solutions, Experimental Research]. In *Istoricheskaya informatika*. No. 2. URL: <https://cyberleninka.ru/article/n/diakhronicheskii-korpus-buryatskogo-yazyka-kak-tsifrovoi-instrument-istoricheskikh-issledovaniy-podkhody-resheniya-eksperimentalnye> (mode of access: 06.04.2021).

Rubtsova, Yu. V. (2015). Postroenie korpusa tekstov dlya nastroyki tonovogo klassifikatora [Building a Text Corpus for Tuning a Tone Classifier]. *Programmnye produkty i sistemy*. No. 1 (109), pp. 72-78.

Solnyshkina, M. I., Gatiyatullina, G. M. (2020). Istoriya razvitiya korpusnoi lingvistiki (na primere angloyazychnykh korpusov) [History of the Development of Corpus Linguistics (on the Example of English-Speaking Corpora)]. In *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya*. No. 63. URL: <https://cyberleninka.ru/article/n/istoriya-razvitiya-korpusnoy-lingvistiki-na-primere-angloyazychnykh-korpusov> (mode of access: 06.04.2021).

Tseytlin, S. N. (1997). *Ot nulya do dvukh: dnevnkovye zapisi* [From Zero to Two: Diary Entries]. Saint Petersburg, Izdatel'stvo RGPU im. A. I. Gertsena. 215 p.

Tseytlin, S. N. (2001). *Slovar' detskikh slovoobrazovatel'nykh innovatsii* [Dictionary of Children's Word-Formation Innovations]. Munchen. 202 p.

Tseytlin, S. N. (2009). *Ocherki po slovoobrazovaniyu i formoobrazovaniyu v detskoj rechi* [Essays on Word-Formation and Form Formation in Children's Speech]. Moscow, Znak. 592 p.

Tseytlin, S. N. (Ed.). (2001). *Dve devochki: Sonya i Nadya: dnevnkovye zapisi* [Two Girls: Sonya and Nadia: Diary Entries]. Saint Petersburg, Tuskarora. 191 p.

Tseytlin, S. N., Eliseeva, M. B. (1998). *Ot dvukh do trekh: dnevnkovye zapisi* [From Two to Three: Diary Entries]. Saint Petersburg, Izdatel'stvo RGPU im. A. I. Gertsena. 159 p.

Zakharov, V. P. (2016). Prolegomeny k korpusnoi lingvistike [Prolegomena to Corpus Linguistics]. In *Voprosy psikholingvistiki*. No. 2 (28). URL: <https://cyberleninka.ru/article/n/prolegomeny-k-korpusnoy-lingvistike> (mode of access: 04.06.2021).

Zyryanova, E. V. (2008). Sistema CHILDES kak metod sbora materialov I izucheniya detskoj rechi [The CHILDES System as a Method of Collecting Materials and Studying Children's Speech]. In *Izvestiya RGPU im. A. I. Gertsena*. No. 76-2. URL: <https://cyberleninka.ru/article/n/sistema-childes-kak-metod-sbora-materialov-i-izucheniya-detskoy-rechi> (mode of access: 04.06.2021).

### Данные об авторе

Красноперова Евгения Сергеевна – аспирант, ассистент кафедры общего языкознания и русского языка, Уральский государственный педагогический университет (Екатеринбург, Россия).

Адрес: 620091, Россия, Екатеринбург, пр. Космонавтов, 26, к. 281.

E-mail: [evgenia113@gmail.com](mailto:evgenia113@gmail.com).

### Author's information

Krasnoperova Evgeniya Sergeevna – Postgraduate Student, Assistant of Department of General Linguistics and Russian Language, Ural State Pedagogical University (Ekaterinburg, Russia).

Дата поступления: 10.09.2021; дата публикации: 30.03.2022

Date of receipt: 10.09.2021; date of publication: 30.03.2022