

УДК 81'33

## ЭЛЕКТРОННЫЙ КОРПУС ПИСЬМЕННЫХ ТЕКСТОВ RUSNEUROPSYCH: СОСТАВ, СТРУКТУРА И ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ

Литвинова Т.А., Рыжкова Е.С.

Работа посвящена описанию формируемого авторами корпуса текстов на русском языке RusNeuroPsych, содержащего метаразметку в виде информации об их авторах (пол, возраст, образование, результаты нейропсихологического обследования и психологического тестирования). Подробно рассматриваются процесс сбора и обработки материала для создания корпуса, его состав и структура. Анализируются возможности применения корпуса RusNeuroPsych в различных областях знаний.

**Ключевые слова:** письменный текст, профиль латеральной организации мозга, нейронауки, корпус текстов, психолингвистический эксперимент, математическая статистика, автоматическая обработка языка, естественная письменная речь.

## THE E-TEXT CORPUS RUSNEUROPSYCH: COMPOSITION, STRUCTURE AND APPLICABILITY

Litvinova T.A., Ryzhkova E.S.

The article describes the text corpus RusNeuroPsych developed in the Russian language. It contains the meta-marking as information about the authors (gender, age, education, results of neuropsychological survey and psychological testing). We consider the details of data collection and procession in formation of the corpus, its composition and structure, and also analyse possibilities of application of the RusNeuroPsych corpus in general knowledge.

**Keywords:** written text, profile of lateral brain organization, neurosciences, corpus of texts, psycholinguistic experiment, mathematical statistics, automatic processing of language, natural written speech.

*Статья подготовлена при поддержке гранта РФФИ «Языковые параметры письменного текста и нейропсихологические характеристики его автора: корпусное исследование», № 16-36-00036.*

Речь человека, в том числе письменная, может дать весьма полное представление о нем. При этом наиболее полную информацию о различных индивидуальных особенностях личности можно получить, анализируя целостное, связанное высказывание (текст) [15, с. 262].

На протяжении всей жизни на речевое поведение человека влияет ряд факторов (пол, возраст, социальный статус, характер) в их взаимодействии друг с другом [3, с. 170], которые в свою очередь неизбежно отражаются в продуктах речевой деятельности – текстах.

Достижения современной науки позволяют говорить об отражении в речевых произведениях не только различных индивидуальных психологических характеристик их авторов, но и особенностей мозговой организации, в частности профиля латеральной организации [17, с. 89-97].

Названные выше обстоятельства свидетельствуют о необходимости исследования письменной речевой продукции носителей языка с опорой на данные современных нейронаук – нейролингвистики нормы [1] и нейропсихологии индивидуальных различий, базирующихся на обследовании практически здоровых людей [10].

Очевидно, что для достаточно глубокого изучения взаимосвязи параметров текстов и нейропсихологических характеристик их авторов требуется материал в виде корпуса письменных текстов, содержащего помимо самих речевых произведений соответствующую информацию об их создателях, в частности, данные о типах профиля латеральной организации мозга. Профиль латеральной организации головного мозга является одной из важнейших нейропсихологических

характеристик человека и определяется как сочетание функциональной асимметрии полушарий, моторной и сенсорной асимметрии [14].

Создание и изучение текстовых корпусов является одним из актуальных направлений современных исследований. Важность корпусов текстов трудно переоценить, поскольку они могут использоваться в работах как теоретического, так и прикладного характера в различных областях знаний (например, лингвистами-теоретиками, переводчиками, преподавателями при обучении языкам, а также историками, социологами, литературоведами и др.) [6, с. 273-274].

Как отмечается в литературе, в мировой науке достаточно широко распространен подход к изучению личности по тексту, состоящий в создании специальных корпусов текстов, в состав которых входят не только собственно тексты, но и метаразметка в виде информации о личности их авторов (пол, возраст, баллы по шкалам психологических тестов и пр.), а также разметка корпусов средствами автоматической обработки языка, извлечении численных значений параметров текста, вычислении корреляций между этими значениями и характеристиками личности автора текста и построении на их основе математических моделей для диагностирования тех или иных характеристик автора текста [8, с. 236-255].

Примерами подобных корпусов являются корпусы текстов, созданные зарубежными исследователями либо на материале текстов интернет-коммуникации, либо текстов, специально написанных респондентами по заданию исследователя (корпус myPersonality3, корпус Дж. Пеннебакера, Stylometry Investigation (CSI) Corpus и др.) [подробный обзор см. 7, с. 109-110]. Такие корпусные исследования строятся иностранными учеными преимущественно на материале текстов английского, датского и других языков, при этом вполне очевидна необходимость их проведения на основе русскоязычных текстов.

На этом пути в российской лингвистике делаются лишь первые шаги. Подтверждением тому служит работа по формированию корпуса русских письменных текстов *RusPersonality*, проводимая с 2012 года научными сотрудниками Регионального центра русского языка при Воронежском государственном педагогическом университете под руководством Т.А. Литвиновой. Данный корпус письменных текстов на русском языке содержит помимо самих речевых произведений (более тысяч) метатекстовую разметку в виде информации о личности их авторов (пол, возраст, уровень образования, результаты психологических тестов) [7, с. 110]. На сегодняшний день это самый большой корпус текстов на русском языке с метаразметкой в виде данных об авторах.

Как мы отмечали выше, одной из актуальных задач современной лингвистики является выявление типологических параметров письменных русских текстов, продуцируемых условно здоровыми людьми с разными профилями латеральной организации головного мозга, с применением современных программных средств обработки языка и методов математической статистики, а также учетом достижений современных нейронаук.

Для решения поставленной задачи в течение 2015 года нами был собран корпус русских письменных текстов, специально написанных респондентами по заданию исследователя, размеченный определенным образом и содержащий, помимо собственно речевых произведений, информацию об их авторах (пол, возраст, родной язык, уровень образования, результаты психологического тестирования и нейропсихологического обследования). Данный корпус под названием *RusNeuroPsych* является одним из подкорпусов корпуса *RusPersonality*, о котором говорилось ранее [7]. Подобного корпуса текстов на русском языке, насколько нам известно, не существует.

Материал корпуса *RusNeuroPsych* был собран в процессе проведенного нами психолингвистического эксперимента, в ходе которого информанты выполняли задания анкет и писали тексты в присутствии исследователя.

Трудоемкость работы над созданием корпуса текстов такого типа определяется тем, что полученный в ходе анкетирования респондентов языковой материал должен быть переведен в электронный вид (каждый текст хранится в отдельном файле в формате Word). Кроме того, необходимо провести обработку результатов нейро- и психологического тестирования участников эксперимента, а также занести полученную информацию в специально разработанную базу данных. Такая база данных представляет собой файл Excel определенной структуры, содержащий в первом столбце имя каждого файла, а в остальных – метаинформацию об авторах (пол, год рождения, уровень образования, результаты нейропсихологического обследования и психотестов) и их речевых произведениях (темы и численные значения лексических и морфологических параметров текстов). При наборе текстов информантов в электронном виде нами были исправлены опечатки, но сохранена авторская пунктуация.

### **Характеристики авторов текстов**

Собранный нами корпус *RusNeuroPsych* содержит 644 текста 455 авторов.

*Пол.* Названный корпус включает в себя тексты, написанные 190 представителями мужского пола и 259 – женского, 6 человек не указали свой пол

*Возраст.* В создании корпуса письменных текстов принимали участие лица в возрасте от 12 до 45 лет.

*Родной язык.* Русский язык является для респондентов родным языком.

*Образование.* Корпус *RusNeuroPsych* содержит тексты лиц с неполным средним образованием (246 человек – учащиеся 6-10 классов МБОУ Лицей №3 и МБОУ Гимназия №9 г. Воронежа), средним образованием (2 человека), с неполным высшим образованием (199 человек – студенты Воронежского государственного университета инженерных технологий 1-4 курсов различных специальностей), высшим образованием (8 человек – представители различных профессий – преподаватели, врачи, инженеры и др.).

*Психологические характеристики.* Все информанты проходили нейропсихологическое обследование на определение моторного, сенсорного и когнитивного латерального профиля (455 человек), а также тестирование для выявления психических состояний и свойств личности. Для школьников был использован опросник Г. Айзенка, адаптированный к подростковому возрасту (246 респондентов), для «взрослых» испытуемых, т.е. студентов и работающих лиц, – Госпитальная шкала тревоги и депрессии HADS и Пятифакторный личностный опросник МакКрае-Коста (209).

### **Характеристики текстов**

Средняя длина текстов корпуса *RusNeuroPsych* составляет 124 слова. Максимальная длина текстов насчитывает 731 слово, минимальная – 5 слов. Все собранные тексты корпуса представляют собой образцы естественной письменной речи [5].

В ходе проведения психолингвистического эксперимента респондентам было предложено написать письмо другу (394 текста) и описать картину (250), которая представлена на страницах опросников (одна и та же картина для всех респондентов).

При этом задания варьировались в зависимости от группы информантов: школьники писали один текст по выбору, «взрослые» испытуемые – два текста.

Перед написанием текстов респондентам давалась установка писать первое, что придет в голову, без предварительного обдумывания и планирования в свободной форме в разговорном стиле, не боясь ошибок.

Анализ собранного нами материала позволяет отметить весьма интересную особенность графического оформления некоторыми авторами своих речевых произведений, написанных в жанре письма другу в виде диалогов в интернет-сетях. Очевидно, это связано с растущей популярностью интернет-коммуникации, что особенно актуально в молодежной среде, к которой и принадлежат (главным образом) наши респонденты.

## **Нейропсихологическое обследование и психологическое тестирование участников эксперимента**

В ходе работы по формированию электронного корпуса русских письменных текстов *RusNeuroPsych* нами создана метаразметка, содержащая информацию о результатах нейропсихологического обследования и психологического тестирования их авторов.

В современной научной литературе описываются разнообразные методики определения профиля латеральной организации детей и взрослых [подробнее см. 4; 13; 16 и др.]. Кроме того, существуют также аппаратные пробы [обзор см. 4], дающие, безусловно, большую точность результатов исследования. В нашей работе мы использовали те методики, которые не требуют специального оборудования и могут применяться в «полевых» условиях на большом числе испытуемых.

Так, для *определения моторного профиля* испытуемых нами были использованы следующие пробы:

– *на определение ведущей руки* (переплетение пальцев рук, скрещивание рук на груди, или «поза Наполеона», аплодироватные [4], задание на заполнение двух квадратов 2х2 см вертикальными линиями (сначала правой рукой – правый квадрат, затем левой рукой – левый квадрат) [14], проба на определение руки, которая чаще всего ловит предмет [там же], тест на поднимание лежащего на полу предмета [4], задание нарисовать на одной руке круг другой рукой и выяснить, какая ладонь рисует [14]);

– *на определение ведущей ноги* (закидывание ноги на ногу, шаг вперед, шаг назад, подъем и схождение со стула, подпрыгивание на одной ноге [4]).

Для *определения сенсорного профиля* респондентов мы применяли такие пробы, как:

– *на определение ведущего глаза* (тест «моргание одним глазом», тест «рассматривание в подзорную трубу», тест на выявление особенностей мышц неведущего глаза [4]);

– на определение ведущего уха (испытуемому предлагают определить, у какого уха прозвучал хлопок в ладоши (его производят за спиной испытуемого, на одинаковом расстоянии от обеих ушей) [14], тест «тиканье часов» [4], тест «шепот» [там же], проба «Телефонная трубка» (отмечается, к какому уху испытуемый подносит трубку) [9]).

С целью определения *когнитивного профиля* информантов в рамках исследования нами были использованы:

– тест И.П. Павлова, согласно которому испытуемому предлагают разделить слова «*карась*», «*орел*», «*овца*», «*перья*», «*чешуя*», «*шерсть*», «*летать*», «*плавать*», «*бегать*» на три группы так, чтобы в каждой группе было что-то общее [14];

– задание на распределение слов «*свет*», «*ухо*», «*зрение*», «*слух*», «*нос*», «*обоняние*», «*глаз*», «*звук*», «*запах*» на три группы на основании их общих признаков [2];

– задание разделить прилагательные «*хороший*», «*неплохой*», «*неумный*», «*плохой*», «*умный*», «*глупый*», «*нехороший*», «*неглупый*» на две группы, чтобы в каждой из них было нечто общее [2];

– распределить цифры *I*, *2* и *I*, *II* на две группы по любому основанию [там же];

– задание расквалифицировать 8 предложений на две группы на основании их общих признаков (*Ваня побил Петю, Петя побил Ваню, Петей побит Ваня, Ваня побит Петей, Петю побил Ваня, Ваню побил Петя, Ваней побит Петя, Петя побит Ваней*) [2].

В научной литературе существуют различные интерпретации перечисленных выше проб на определение профиля латеральной организации мозга. Сказанное особенно касается «*позы Наполеона*» и теста «*моргание одним глазом*» [см. 4; 9; 11; 14]. В нашей работе названные пробы объясняются следующим образом: при скрещивании рук на груди ведущей считается та, локоть которой лежит сверху; при моргании испытуемый закрывает ведущий глаз.

При проведении нейропсихологического тестирования нами были выявлены типы заданий, вызвавшие наибольший интерес у опрошенных. К числу таких заданий относятся пробы на определение моторного и сенсорного профилей головного мозга, выполняемые в паре, когда один из участников эксперимента становится «специалистом-психологом», а другой – «испытуемым». Респонденты поочередно выполняли задания и наблюдали друг за другом, фиксируя результаты в бланках анкет.

Заданиями, наиболее трудными для информантов, при выполнении которых у них весьма часто возникали вопросы, затруднения и даже ошибки, оказались пробы на выявление когнитивного профиля головного мозга. Некоторые опрошенные не справились с отдельными видами проб на определение моторного и сенсорного профиля, в том числе по причине полученных ранее травм (эту информацию респонденты указывали в своих анкетах), а также из-за нехватки времени на выполнение заданий. Общее количество «взрослых» респондентов и «детей», которые не справились с теми или иными заданиями на определение ПЛО, составляет, соответственно 102 и 153.

Кроме того, в ряде случаев некоторые информанты не выполнили также задания психологических тестов. Как упоминалось выше, «детям» был предложен опросник Г. Айзенка, «взрослым» респондентам – Госпитальная шкала тревоги и депрессии HADS и Пятифакторный личностный опросник МакКрае-Коста. Общее количество человек, не ответивших на те или иные вопросы психологических тестов, составляет, соответственно, для опросника Г. Айзенка («дети») – 56, для Госпитальной шкалы тревоги и депрессии HADS («взрослые») – 6, для опросника МакКрае-Коста («взрослые») – 19.

В дальнейшем мы планируем продолжать работу по формированию электронного корпуса связных письменных текстов на русском языке *RusNeuroPsych*.

На наш взгляд, создаваемый корпус русских письменных текстов *RusNeuroPsych*, структурированный определенным образом и содержащий также метаданные об их продуцентах, может быть применим в различных сферах.

Корпус *RusNeuroPsych* позволит впервые провести исследования, направленные на выявление типологических особенностей письменной речи лиц с разным типом профиля латеральной организации.

Кроме того, корпус может быть использован для построения методик моделирования личности автора письменных текстов, найдет свое применение в практике преподавания русского языка как родного и иностранного с целью научно обоснованного выбора наиболее эффективных приемов и методик обучения с учетом индивидуальных различий когнитивной сферы учащихся [12, с. 110].

#### **Список литературы:**

1. Ахутина Т.В. Нейролингвистика нормы // I Международная конференция памяти А.Р. Лурия. Сб. докладов. М., 1998. С.289-298.
2. Балонов Л.Я., Деглин В.Л., Черниговская Т.В. Функциональная асимметрия мозга в организации речевой деятельности // Сенсорные системы. Сенсорные процессы и асимметрия полушарий. Л.: Наука, 1985. С. 99-115.
3. Белянин В.П., Шкуратова И.П. Диалоги о человеке говорящем и пишущем. СПб.: Речь, 2011. С. 161-170.
4. Брагина Н.Н., Доброхотова Т.А. Функциональные асимметрии человека. 2-е изд., перераб. и доп. М.: Медицина, 1988. 240 с.
5. Лебедева Н.Б. Естественная письменная русская речь как объект лингвистического исследования [Электронный ресурс] // Вестник Барнаульского государственного педагогического университета. 2001. Вып. 1. URL: <https://goo.gl/IZjoEo> (дата обращения: 04.08.2016).

6. Литвинова Т.А., Литвинова О.А. Идентификация и диагностирование личности автора письменного текста: монография. Воронеж: Воронежский государственный педагогический университет, 2015. С. 273-274.

7. Литвинова Т.А., Диброва Е.В., Литвинова О.А., Рыжкова Е.С. Корпусные исследования письменной речи в решении задач судебного автороведения // Филологические науки. Вопросы теории и практики. 2015. № 8. Ч. 1. С. 107-113.

8. Литвинова Т.А. Языковые корреляты личностных особенностей автора письменного текста: алгоритм исследования // В мире научных открытий. Сер.: Проблемы науки и образования. 2012. № 9.3 (33). С. 236-255.

9. Москвина Н.В., Москвин В.А. Межполушарные асимметрии и индивидуальные различия человека. Москва: Смысл, 2011.

10. Нейропсихология индивидуальных различий: учеб. пособие для студ. учреждений высш. проф. образования. М.: Издательский центр «Академия», 2011.

11. Николаева Е.И., Борисенкова Е.Ю. Сравнение разных способов оценки профиля функциональной сенсомоторной асимметрии у дошкольников // Асимметрия. 2008. № 1. С.32-39.

12. Рыжкова Е.С. Когнитивные стили обработки информации при обучении русскому языку как иностранному // Международное образование и сотрудничество: сборник материалов IV международной научно-практической конференции «Профессионально направленное обучение русскому языку иностранных граждан» в 3 т. М.: МАДИ, 2015. Т. 3. С. 108-110.

13. Семаго Н.Я., Семаго М.М. Теория и практика оценки психического развития ребенка. Дошкольный и младший школьный возраст. СПб.: Речь, 2005. С. 110-111.

14. Сиротюк А.Л. Нейропсихологическое и психофизиологическое сопровождение обучения. М.: ТЦ Сфера, 2003. 288 с.

15. Фомина, Н.А. Комплексное изучение проявлений личности в речевой деятельности // Психологічні перспективи. Волин. нац. ун-т ім. Лесі Українки, Ін-т соц. та політ. психології. Луцьк, 2011. Вип. 18. С. 257-270.

16. Хомская Е.Д., Ефимова И.В., Будыка Е.В., Ениколопова Е.В. Нейропсихология индивидуальных различий. М.: Российское педагогическое агентство, 1997. 282 с.

17. Шубин А.В., Серпионова Е.И. Асимметрия мозга и особенности вербальной креативности // Вопросы психологии. 2007. № 4. С. 89-97.

#### **Сведения об авторах:**

Литвинова Татьяна Александровна – кандидат филологических наук, руководитель Лаборатории диагностирования личности по тексту Воронежского государственного педагогического университета (Воронеж, Россия).

Рыжкова Екатерина Сергеевна – аспирант, научный сотрудник Лаборатории диагностирования личности по тексту Воронежского государственного педагогического университета (Воронеж, Россия).

#### **Data about the authors:**

Litvinova Tatiana Aleksandrovna – Candidate of Philological Sciences, Head of the Authorship Profiling Lab, Voronezh State Pedagogical University (Voronezh, Russia).

Ryzhkova Ekaterina Sergeevna – graduate student, Staff Scientist of the Authorship Profiling Lab, Voronezh State Pedagogical University (Voronezh, Russia).

**E-mail:** [centr\\_rus\\_yaz@mail.ru](mailto:centr_rus_yaz@mail.ru).

**E-mail:** [ryzhkova.katerina@yandex.ru](mailto:ryzhkova.katerina@yandex.ru).