

«КОНДУИТ»: КОРПУС УСТНЫХ ДЕТСКИХ ТЕКСТОВ¹

“KONDUIT”: CORPUS OF CHILD ORAL NARRATIVES

Аннотация. В статье представлен проект КОРПУСА Неподготовленных Детских Устных (Извлеченных) Текстов «Конduit». Корпус состоит из 213 однотипных устных неподготовленных текстов детей в возрасте 2;7–7;6 лет, каждый из которых имеет метатекстовую, структурную, синтаксическую и семантическую разметку. Основной целью создания корпуса является получение базы устных неподготовленных текстов для изучения процессов формирования навыков построения связного текста в онтогенезе.

Ключевые слова. Детская речь, связный текст, устный корпус, связность, цельность, глагольная структура.

Abstract. The project of a new corpus of child oral unprepared elicited narratives “Konduit” is described. The corpus consists of 213 oral narratives based on the same story or a similar sequence of actions, elicited by Russian native children at the age of 2;7–7;6. Each text has semantic, syntactic and discourse annotations. The main purpose of this corpus is to become an effective and useful tool for the studying of narrative acquisition in Russian.

Keywords. Child language, oral corpus, narrative, coherence, cohesion, verb structure.

В последнее время в отечественной лингвистике созданию устных корпусов уделяется значительное внимание — расширяется устный подкорпус Национального корпуса русского языка [Гришина, Савчук 2009], развивается проект «Один речевой день» [Богданова-Бегларян 2016], продолжается работа над полной фонетической расшифровкой Корпуса русских спонтанных текстов [Венцов и др. 2013]. Несмотря на бесспорную значимость для изучения современного русского языка и процессов порождения и восприятия речи в обществе и высокий уровень работы над этими корпусами, каждый из них уделяет большее внимание решению какой-либо конкретной задачи, а все они включают в себя в первую очередь образцы речи взрослых носителей русского языка.

Специфика КОРПУСА Неподготовленных Детских Устных (Извлеченных) Текстов «Конduit» заключается в том, что в нем представлены образцы устной связной речи детей разных возрастных групп от 2;7 до 7;6 лет. Основной целью создания корпуса является получение базы устных неподготовленных текстов для изучения процессов фор-

¹ Публикация подготовлена в рамках поддержанного РГНФ научного проекта №16-04-50114.

мирования навыков построения связного текста в онтогенезе. Всего в корпусе представлено 213 текстов, которые были получены в результате проведения серии экспериментов с детьми, посещающими дошкольные образовательные учреждения г. Санкт-Петербурга. Собранные тексты поделены на 5 возрастных групп: в возрасте от 2,7 до 3,6 (далее — 1 группа) — 37 детей, в возрасте от 3,7 до 4,6 (далее — 2 группа) — 50 детей, в возрасте от 4,7 до 5,6 (далее — 3 группа) — 49 детей, в возрасте от 5,7 до 6,6 (далее — 4 группа) — 42 ребенка и в возрасте от 6,7 до 7,6 (далее — 5 группа) — 35 детей.

При проведении экспериментов с детьми разного возраста был использован разный экспериментальный дизайн: с детьми 1 группы эксперимент проходил в форме игры, в которой дети должны были называть действия, производимые экспериментатором при помощи игрушек-бибабо; дети 2 группы рассматривали с экспериментатором книжку с картинками (сказка В.Г. Сутеева «Три котенка», рис. автора), рассказывая, что на них нарисовано; дети 3, 4 и 5 групп смотрели мультфильм без звука («Как стать большим?», «Союзмультфильм», 1967) и рассказывали происходящее одновременно с просмотром без предварительной подготовки. Таким образом, тексты, полученные при проведении эксперимента с детьми 1 группы, имеют форму диалога, тексты, полученные при проведении эксперимента с детьми 2 группы, также имеют форму диалога, но включают в себя монологические связные фрагменты, тексты, полученные при проведении эксперимента с детьми 3, 4 и 5 групп, имеют форму монологического связного текста. Общий объем корпуса — 25 689 словоформ (сюда входит только речевая продукция детей без высказываний экспериментатора). Каждый текст имеет следующие характеристики: пол и возраст автора (на момент записи), общее число словоформ и высказываний в данном тексте, средняя длина высказывания (MLU) (табл. 1).

Особой отличительной чертой корпуса является то, что все содержащиеся в нем тексты основаны на одинаковой (группы 3, 4 и 5) или схожей (группы 1 и 2) последовательности действий, что сюжетно унифицирует тексты и снимает многие дополнительные факторы, влияющие на семантическую и синтаксическую организацию связного текста. В этом корпус «Конduit», с одной стороны, продолжает традиции изучения особенностей связного устного текста, заложенные в таких крупных исследованиях, как кросслингвистическая экспериментальная серия изучения детского нарратива под руководством Рут Берман и Дэна Слобина [Berman, Slobin 1994] или исследование устно-

Таблица 1. Статистика корпуса по возрасту и полу, характеристика групп по числу словоформ и высказываний, по средней длине высказывания (MLU)

Группа	Средний возраст	М	Ж	Число словоформ (Ме)	Число высказываний (Ме)	MLU (Ме)
1	3 года 1 месяц	21	16	20	13	1,5
2	4 года 2 месяца	30	20	84,5	27	3,12
3	5 лет	25	24	108	23	4,61
4	6 лет	20	22	174,5	34	5,10
5	6 лет 9 месяцев	16	19	190	34	5,39

го спонтанного дискурса «Рассказы о сновидениях» [Кибрик, Подлеская 2009], а с другой, отличается от них, поскольку в последнем случае тексты были объединены одной темой, но были различны по сюжету и эмоциональной вовлеченности рассказчика, а при кросслингвистическом исследовании основной проблематикой является изучение именно сходства и различий в процессе усвоения нарративных принципов говорящими на разных языках, а не построение последовательности этапов усвоения нарративной структуры в онтогенезе.

Применяемая разметка обусловлена целью создания корпуса и включает три основных типа:

- структурная разметка: отражает эпизодическую структуру текстов и полноту описания события, персонажа и обстоятельств каждого из эпизодов. Поскольку тексты либо полностью, либо частично сюжетно идентичны, то такая разметка облегчает задачи исследования особенностей организации цельности текста в рассказах детей разного возраста, а также показывает важность отдельных действий разных типов, деталей и обстоятельств ситуации для детей на разном уровне когнитивного развития;
- синтаксическая разметка: отражает используемые детьми средства обеспечения связности текста на разных этапах усвоения навыков построения связного текста. Также включает в себя коммуникативную разметку, что позволяет выявить взаимосвязь между использованием в связном русскоязычном тексте различных синтаксических структур и организацией коммуникативной (тема-рематической) структуры высказывания;

— семантико-синтаксическая разметка глагольных структур: отражает взаимосвязь семантического класса и синтаксической структуры глагола, а также позволяет проводить сравнительное исследование усвоения семантико-синтаксической структуры глагола.

Взаимодействие всех трех типов разметки отражает роль, которую играют глаголы и их аргументы в организации структуры текста, в отражении событий и действий, в восприятии и понимании окружающей действительности детьми на различных этапах когнитивного и языкового развития, позволяет приблизиться к пониманию взаимодействия семантики и синтаксиса в процессах порождения и восприятия речи.

Данный корпус может использоваться для проведения сравнительных исследований процесса формирования навыков организации связности и цельности текста, формирования разнообразия синтаксических структур, развития лексического запаса, формирования таких когнитивных способностей, как внимание к деталям, эмпатия и эмоциональная оценка поведения другого, развитие фантазийного мышления и многого другого у русскоязычных детей в возрасте от 2,7 до 7,6 лет.

Литература

1. Богданова-Бегларян Н. В. (ред.) (2016), Русский язык повседневного общения: особенности функционирования в разных социальных группах / Н. В. Богданова-Бегларян, Т. Ю. Шерстинова, Е. М. Баева, О. В. Блинова, Г. Я. Мартыненко, О. Б. Ермолова, А. И. Рыко и др. СПб.
2. Венцов А. В., Нигматулина Ю. О., Раева О. В., Риехакайнен Е. И., Слепокурова Н. А. (2013), Корпус русских спонтанных текстов: структура и единицы. Труды международной конференции «Корпусная лингвистика–2013». СПб., с. 223–231.
3. Гришина Е. А., Савчук С. О. (2009), Корпус устных текстов в НКРЯ: состав и структура // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб., с. 129–149.
4. Кибрик А. А., Подлесская В. И. (ред.) (2009), Рассказы о сновидениях: Корпусное исследование устного русского дискурса. М.
5. *Berman R., Slobin D. (eds.) (1994), Relating events in narrative: A crosslinguistic developmental Study. New York.*

References

1. *Bogdanova-Beglaryan N. V.* (ed.) (2016), *Russkij yazyk povsednevnogo obshheniya: osobennosti funkcionirovaniya v raznykh social'nykh gruppakh* [Russian everyday communication: its functioning within different social groups]. Saint Petersburg.
2. *Vencov A. V., Nigmatulina Yu. O., Raeva O. V., Riekhakajnen E. I., Slepokurova N. A.* (2013), *Korpus russkikh spontannykh tekstov: struktura i edinicy* [Corpus of Russian spontaneous texts: its structure and elements]. In: *Trudy mezhdunarodnoj konferencii "Korpusnaya lingvistika — 2013"* [Proceedings of international conference "Corpus linguistics-2013"]. Saint Petersburg, pp. 223–231.
3. *Grishina E. A., Savchuk S. O.* (2009), *Korpus ustnykh tekstov v NKRYA: sostav i struktura* [Corpus of oral texts in National Russian Corpus: elements and structure]. In: *Nacional'nyj korpus russkogo yazyka: 2006–2008. Novye rezultaty i perspektivy* [National Russian Corpus: 2006–2008. New results and perspectives]. Saint Petersburg, pp. 129–149.
4. *Kibrik A. A., Podlesskaya V. I.* (eds.) (2009), *Rasskazy o snovideniyakh: Korpusnoe issledovanie ustnogo russkogo diskursa* [Night Dream Stories: A corpus study of spoken Russian discourse]. Moscow.
5. *Berman R., Slobin D.* (eds.) (1994), *Relating events in narrative: A crosslinguistic developmental Study*. New York.

Эйсмонт Полина Михайловна

Санкт-Петербургский государственный университет
аэрокосмического приборостроения

Eismont Polina

Saint Petersburg State University of Aerospace Instrumentation

E-mail: polina272@hotmail.com