

СИСТЕМА CHILDES КАК МЕТОД СБОРА МАТЕРИАЛОВ И ИЗУЧЕНИЯ ДЕТСКОЙ РЕЧИ

*Работа представлена кафедрой речевой коммуникации
Пермского государственного университета.*

Научный руководитель – доктор филологических наук, профессор И. Г. Овчинникова

В статье рассматриваются методы анализа детской речи. Приводится описание международной системы обмена данными по детской речи CHILDES. Делается вывод о том, что записи детской речи, осуществленные в системе CHILDES, позволяют не только исследовать речь, но и составить представление об информанте, установить автора записи. Кроме того, размещение материалов в CHILDES предполагает персональную ответственность за достоверность и качество записи. У системы есть несомненные преимущества: доступность, разноаспектность, единый формат записей, разнообразие материалов и их достоверность.

Methods of infantile speech analysis are observed in the article. The description of CHILDES, an international data exchange system on infantile speech, is given. The author comes to the conclusion that children's speech records made in CHILDES make it possible not only to investigate the speech, but also to make an idea of informants and to determine the author of records. Besides, material arrangement in CHILDES supposes personal responsibility for reliability and quality of recorded information. The system possesses certain advantages: accessibility, variety of aspects, records' common format as well as material diversity and its reliability.

Исследование детской речи имеет довольно богатую историю. Конец XIX – первая половина XX в. стали периодом «дневниковых штудий». Такие известные ученые, как Ипполит Тэн, Чарльз Дарвин, Вильям Прейер, публикуют в журналах фрагменты дневниковых записей речи своих детей. Одним из капитальных трудов, созданных на рубеже веков, явилась книга Клары и Вильгельма Штернов «Die Kindersprache», основанная на дневниковых записях об их детях. И. А. Бодуэн де

Куртене, одним из первых оценив значимость материалов детской речи для теоретической лингвистики, также вел подробные записи речи своих детей.

В нашей стране интерес к речи ребенка возникает в 20-е гг. XX в. в основном благодаря работам Н. А. Рыбникова, которому удалось собрать и издать разрозненные записи родителей, а также провести серьезные исследования в области детской речи. Фундаментальным исследованием являются и дневниковые

записи Н. Гвоздева, в которых он фиксировал речь своего сына, сопровождая их лингвистическими комментариями. На этих наблюдениях в СССР долгие годы базировалась вся лингвистика детской речи.

В конце 20-х гг. прошлого века получили распространение «кросс-секционные штудии», где проводилось сопоставление образцов речи разных детей. В начале 60-х гг. XX в. появились «лонгитюдные срезовые штудии», позволяющие изучать документальные записи фрагментов речи. На смену дневникам пришли документальные записи фрагментов речи ребенка, фиксируемой с заранее назначенными временными интервалами. В этом лонгитюде уже не было фрагментарности и случайных записей, свойственных дневникам.

Кроме дневниковых записей и лонгитюдных наблюдений для анализа детского лексикона используют эксперимент и анкетирование родителей. Одним из вариантов анкетирования является мак-артуровский опросник, русскоязычную версию которого разрабатывает кафедра детской речи Российского государственного педагогического университета (см., напр., [1]).

Очевидно, что для получения достоверных данных необходимо использовать все методы анализа лексикона, дополнять ими друг друга. Так, материал видеозаписей спонтанной речи детей до известной степени объективен, насколько вообще объективна техника: известно, что некий языковой факт может длительное время не попадать в запись, хотя и звучать в речи ребенка, когда запись не ведется. Материал родительских дневников также обычно не вызывает сомнений в достоверности, поскольку в научных целях используются чаще всего дневники людей, причастных к онтолингвистике.

Ценность экспериментальных материалов не одинакова. В каких-то случаях эксперимент в онтолингвистике незаме-

ним: с его помощью можно проанализировать не только продуцирование речи детьми, но и понимание ими того, что говорит взрослый; можно обследовать сразу десятки и даже сотни детей. Вместе с тем известны и недостатки экспериментальных методов. Например, при проведении эксперимента нельзя точно назвать причину отсутствия ответа: то ли ребенок ответа дать не может, то ли не хочет. Тем не менее без эксперимента тоже не обойтись.

Благодаря опросникам можно обнаружить то, что трудно уловить при промежуточных видеозаписях. В то же время почти в половине опросников наблюдается либо родительская переоценка ребенка, либо недооценка, поэтому каждый конкретный опросник – ненадежен, но в целом результатам мак-артуровского опросника можно доверять.

Использование описанных методов анализа детской речи дало возможность уже ко второй половине прошлого века накопить огромный материал. Однако объем расшифрованных стенограмм, сделанных в 1960-е – 1970-е гг., не позволял создавать полные корпуса текстов детской речи. Поэтому исследователям приходилось публиковать работы, основанные на не доступных проверке данных личных наблюдений и примерах из чужих работ. Это породило ситуацию, в которой необработанная эмпирическая база данных для конкретного исследования была сохранена только в частных запасах, не доступных экспертизе широкой публики. Каждый исследователь использовал свою систему транскрипции и кодов. Проблемы методологии, кодирования и межисследовательской надежности становились наиболее очевидными при сравнении рукописных и машинописных расшифровок. В связи с этой ситуацией возникла необходимость создания международной системы обмена данными.

В тот момент, когда стали очевидны проблемы анализа данных, появилась

большая технологическая возможность в форме мощного микрокомпьютера. Системы обработки текстов и система управления базами данных позволила исследователям вводить данные расшифровки стенограммы в компьютерные файлы, которые можно было легко скопировать, отредактировать и обработать с помощью стандартных методов обработки данных. И в 1981 г. был задуман проект CHILDES, который исследователи изначально планировали как большую записную книжку.

Итак, в 1980-х гг. появляется международная система обмена данными по детской речи CHILDES (Children Language Database Exchange System) (<http://chil-des.psy.cmu.edu>), разработанная американскими учеными К. Сноу и Б. Мак-Винни в Питтсбургском университете и применяемая для анализа разговорной речи, спонтанной речи детей и их родителей, а также для исследования усвоения второго языка. На сегодняшний день база данных CHILDES представляет собой значительный корпус, включающий языки различного типологического строя, объемную библиографию по психолингвистике, лингвистике, теории усвоения первого и второго языков, а также правила ввода материала и пакет программ для его анализа.

Каждый исследователь может воспользоваться необходимыми ему данными, а также разместить в CHILDES свои материалы.

Работа с системой начинается с изучения компьютерного формата транскрибирования, необходимого для дальнейшего использования программ. В правилах создания и разметки текста сконцентрированы современные представления о том, на какие особенности речи следует обращать внимание при записи данных. Наиболее часто применяемые программы пакета CHILDES дают возможность определить среднюю длину высказывания каждого собеседника, составить словарь его

речи, найти все случаи употребления последовательностей слов. Система CHILDES имеет широкую сферу применения и может конвертироваться в другие лингвистические системы.

В международном банке данных детской речи есть и записи речи русских детей, представленные благодаря работе таких специалистов по детской речи, как М. А. Воейкова, Н. В. Гагарина, Е. Ю. Протасова (см. обзор: [6]). Однако массив русских записей не так велик: в основном исследователи размещают в CHILDES уже обработанные материалы, поэтому банк данных пополняется очень медленно.

Один из блоков базы данных CHILDES составляют устные рассказы по серии картинок «Frog where are you?» – история приключений мальчика и его собаки в поисках пропавшей лягушки. Подобный блок представлен в CHILDES не случайно: способность ребенка построить развернутый рассказ о событии и основные параметры такого рассказа полагают одним из основных критериев для определения уровня речевого развития.

Банк данных рассказов русских детей по серии картинок «Frog where are you?» включает 170 историй. А 50 историй, полученных нами от пятилетних детей, мы обработали по правилам CHILDES. Каждый из текстов выглядит подобным образом:

```
@Begin
@Name of CHI: Igor' Duleoov
@Age of CHI: 5;0.12
@Birth of CHI: 04-MAI-1998
@Date: 20-JAN-2004
@Filename: Igor'-D_5_ftr.cha
@Tape: DVD 25 (Perm)
@Language: Russian
@Participants: CHI Target_Child,
INV Elena Zyryanova
@Dependent: err, act, exp, com
@Sex of Chi: male
@Situation: Anschauen des Bilderbuchs: Mayer, M. (1969): Frog, where are
```

you? – Printed in Hong Kong by South China Printing Co.

@Transcriber: Elena Zyryanova

@g 01-

*CHI: oni smotrjat na ljagushku.

*INV: ugu, potom chto?

@g 02a

CHI: potom oni # spali # a ljagushka vylazila [].

%err: vylazila = vylezala \$MOR

*INV: tak.

@g 02b

CHI: potom oni prosnulis' # a ljagushki netu [].

%err: netu = net \$PHO

*CHI: tak.

@g 03a

*CHI: a potom oni odevalis' na ulicu # a # oni iskali 0.

%err: 0 = OBJ \$SYN

*INV: ugu.

CHI: i chjo [] dal'she?

%err: chjo = chto \$PHO

%com: CHI obrashchajetsja k INV

*INV: davaj posmotrim.

@g 03b

CHI: i ona vyprygnula pered okoshko [].

%err: vyprygnula pered okoshko = vyprygnula v okoshko \$SYN

*INV: tak.

@g 04a

*CHI: a potom sobachka tozhe vypala iz okna.

@g 04b

CHI: a potom # razbilas' butylka [] # i mal'chik rasserdilsja.

%err: butylka = banka \$LEX

@g 05

CHI: e'to chjo [] # rechka ili chjo e'to?

%err: chjo = chto \$PHO

%act: CHI pokazyvajet na kartinku

*INV: e'to?

%act: INV pokazyvajet na kartinku

*CHI: net # vot.

%act: CHI pokazyvajet na kartinku

*INV: e'to dorozhka.

*CHI: i u+// # i uvideli pchjol.

*INV: a pchjoly gde byli?

*INV: kuda poshli mal'chik s sobachkoj?

CHI: v gnezdo osinoe [].

%err: v gnezdo osinoe = k gnezdu osinomu \$SYN

*INV: tak, a chto dal'she bylo?

@g 06a

CHI: potom # pjosik prygal # k domiku [] # a mal'chik uvidel dyrochku[*].

%err: k domiku = k ul'ju \$LEX

%err: dyrochku = norku \$LEX

*INV: tak.

@g 06b

*CHI: a potom xomjachok ego [%mal'chika] za n+// ukusil za nos.

*INV: ugu.

@g 07

CHI: potom # domik upal # i mal'chik uvidel # druguju norku [].

%err: druguju norku = duplo \$LEX

@g 08

CHI: a potom on svalilsja [] # a tam sova.

*INV: a sobaka?

*CHI: a sobaka ubezhala # ispugalas' # a # pchjol ispugalas'.

@g 09a

*CHI: a potom on # sprjatsja # za kameshkom # a sova # smotrit na nego.

@g 09b

*CHI: a potom mal'chik zalez na kameshek # i zovjot sobachku.

*CHI: mne tak kazhetsja.

%com: CHI obrashchajetsja k INV

*CHI: ili chjo delaet?

%com: CHI obrashchajetsja k INV

*INV: tak.

@g 10a

CHI: a potom ego podxvatil los' [].

%err: los' = olen' \$LEX

@g 10b

*CHI: i los' # potom ubezhal.

@g 11

*CHI: on svoe (?) # vmeste s pjosi-kom +/.

*INV: ugu.
 @g 12a
 CHI: +, i svalilsja v grjaz' [].
 %err: grjaz' = boloto \$LEX
 @g 12b
 *CHI: a potom mal'chik vytashchil sobachku.
 @g 13a
 CHI: a potom on uvidel kakoe-to [] dyrku [*] i on govorit "tshshsh!"
 %err: kakoe-to = kakuju-to \$MOR
 %err: dyrku = brevno \$LEX
 @g 13b
 CHI: i oni stali # zaezzha+// zalezat' [] na e'to derevo.
 @g 14a
 *CHI: a potom # zalezli a tam ljagushki.
 @g 14b
 CHI: a potom oni # stali # slazit' [] a tam staja ljagushek.
 @g 15
 *CHI: a potom oni odnu ljagushku vzjali' # vzja'li # i poshli domoj.
 %err: ostalos' nejasnym, pomnit li CHI tezis o propavshej ljagushke \$VOP
 @End

Для создания подобного файла в программе CLAN необходимо учитывать систему обязательных заголовков. Первой строкой в каждом файле должна быть @Begin – строка заголовка. Второй – @Languages – строка, обозначающая язык. Третьей – @Participants – строка, в которой отражена кодировка участников диалога. В нашем случае это CHI – ребенок и INV – интервьюер. Последней строкой в файле должна быть @End – строка, завершающая файл. В нашем материале, кроме того, присутствуют строки Name of CHI – имя ребенка, @Age of CHI – возраст ребенка на время записи, @Birth of CHI – дата рождения ребенка, @Date –

дата записи материала, @Filename – кодировка имени ребенка и его полных лет, @Type – вид и место записи материала, @Dependent – необходимые комментарии (например, err – ошибка, act – проявление активности, com – вхождение в контакт), @Sex of Chi – пол ребенка, @Situation – ситуация записи или материал, благодаря которому осуществлялась данная запись, @Transcriber – выполняющий расшифровку записи.

Строки, начинающиеся со знака *, содержат в себе то, что было фактически сказано. Их называют «основными строками». В каждой основной строке закодировано только одно произнесение. Строки, начинающиеся с символа %, как правило, включают коды и комментарий того, что было сказано. Их называют «дополнительными» строками. В строке @g указывается номер описываемой картинки (в нашем материале их всего 15). Строка, начинающаяся с %err, как уже было сказано, свидетельствует об ошибке, совершенной ребенком. В связи с тем что нам необходимо знать, какого рода была совершена ошибка, мы сделали дополнительные пометы. Так, \$MOR – ошибка морфологическая, \$PHO – фонетическая, \$SYN – синтаксическая, \$VOP – возможная потеря основного тезиса рассказа.

Таким образом, запись позволяет не только исследовать речь, но и составить представление об информанте, установить автора записи. Размещение материалов в CHILDES предполагает персональную ответственность за достоверность и качество записи. У системы есть несомненные преимущества: доступность, разноаспектность, единство формата записей, разнообразие материалов и их достоверность.

СПИСОК ЛИТЕРАТУРЫ

1. Доброва Г. Р., Королев В. Д. Усвоение детьми раннего возраста слов релятивной семантики: к вопросу о валидности Мак-Артуровского опросника // Проблемы онтолингвистики. СПб., 2007.

ПЕДАГОГИКА, ПСИХОЛОГИЯ, ТЕОРИЯ И МЕТОДИКА ОБУЧЕНИЯ

2. *Протасова Е.Ю.* Эмоциональная регуляция в общении взрослого и ребенка // Эмоции в языке и речи: Сб. ст. / Под ред. И. А. Шаронова. М.: РГГУ, 2005. С. 161–176.

3. *Фотекова Т. А., Ахутина Т. В.* Диагностика речевых нарушений школьников с использованием нейропсихологических методов. М., 2002.

4. *Цейтлин С. Н.* Язык и ребенок. Лингвистика детской речи. М., 2000.

5. *MacWhinney B.* The CHILDES Project: Tools for Analyzing Talk. 3rd Ed. New Uork, 2000.

6. *Polinsky M.* Acquisition of Russian: Uninterrupted and Incomplete Scenarios // The Slavic and Eastern European Language Research Center: <http://seerlc.org/glossos/glossos/seerlc.org>.