

**Будешь задумываться / что говорить / а что не говорить:  
парадокс наблюдателя и речевой контроль в ОРД**

**Аннотация:** Доклад состоит из двух частей. В первой описываются принципы создания корпуса русского языка повседневного общения «Один речевой день» (ОРД). Во второй приводятся некоторые наблюдения, касающиеся эксплицитного проявления т. наз. «парадокса наблюдателя» (observer's paradox) в материалах корпуса.

1. Корпус ОРД фиксирует повседневную устную речь взрослых (старше 18 лет) нативных носителей русского языка, живущих в Санкт-Петербурге. Подбираются информанты-волонтеры, согласившиеся прожить целый день с работающим диктофоном. Запись проводится с использованием цифрового диктофона, способного выполнять многочасовую запись заданного качества, и выносного петличного микрофона. Перед проведением записи диктофон проходит тестирование и проверку настроек.

Информанты перед началом записи проходят инструктаж. Эта работа призвана по возможности обеспечить естественность коммуникативного поведения информантов, а также обеспечить качество звукозаписи, необходимое для дальнейшего анализа полученного материала. Информантам гарантируется анонимность. Проводится технический инструктаж по использованию диктофона, информанты получают специальные памятки.

Выработана методика собирания метаданных и способы юридической защиты собранных полевых записей. Информанты заполняют **Согласие** на дальнейшее использование полученных в ходе эксперимента речевых данных, а также специально разработанные анкеты, в том числе – социологическую (где отмечаются пол, возраст, место рождения, родной язык, место наиболее длительного проживания, образование, профессия информантов и др.). Информантам сообщается о необходимости в течение «речевого дня» предупреждать своих собеседников о ведущейся аудиозаписи. Информанты ведут «дневник речевого дня», где описывают, когда, где с кем они успели пообщаться. Сведения об информантах и их собеседниках (коммуникантах) собираются в базе данных.

Расшифровка звукозаписей выполняется в среде ELAN. Транскрипт записывается в стандартной орфографии. Основные конвенции оформления

транскрипта таковы: начало реплики не выделяется заглавной буквой, вместо точек и занятых используются знаки синтагматического и фразового членения; для обозначения пауз различного типа используется отдельный набор знаков; используются специальные символы для обозначения неразборчивых фрагментов речи, спорных расшифровок слов, обрывов слов. Обозначаются смена говорящего и наложение речи разных говорящих. Выделен набор знаков для обозначения паралингвистических явлений (смех, кашель и др.).

Первичное многоуровневое аннотирование также выполняется в ELAN. Базовые уровни аннотирования таковы: уровень фраз, включающий орфографическую запись с указанием нехизитационных пауз и шумов; уровень говорящих, включающих код говорящего (S1, S2, S39) и его коммуникантов: мужчин, женщин, детей (M1, Ж1, P1); уровень «Events», включающий неречевые аудиособытия; уровень «Voice», описывающий качество голоса (шепот, с эмфазой, скандирующий и др.); уровень фонетического комментария, на котором отмечаются заметные на слух отклонения от стандартного русского языка на фонетическом уровне, ненормативные ударения; уровень фразового комментария, содержащий дополнительную информацию о реализации реплики; уровень «Notes» с общим комментарием (напр., «фрагмент содержит много молодежного сленга»), уровень эпизодов, использующийся для обозначения небольших эпизодов (например, «отвечает на телефонный звонок»).

Сегментация на токены (уровень «Words») производится в ELAN автоматически по орфографической расшифровке фразы (разделитель – пробел). Дальнейшее аннотирование выполняется либо внутри среды ELAN (вручную, иногда с использованием т. наз. «контрольных словарей»), либо разметке подвергаются экспортированные из ELAN списки. Так, например, при морфологической разметке с помощью морфоанализатора MyStem автоматически размечаются выгруженные нумерованные списки словоформ, а омонимия снимается вручную.

2. Уильям Лабов сформулировал следующую известную проблему: «целью лингвистических исследований речевого коллектива является выяснение того, как говорят люди, когда за ними не ведется систематического наблюдения; а получить такие данные можно лишь путем систематических наблюдений» (Лабов У. «Исследование языка в его социальном контексте»).

В идеологии создания ОРД заложено стремление избежать воздействия на речевое поведение носителя языка, просто зафиксировав его речевую продукцию в условиях, максимально приближенных к естественным. Между тем, материалы записей целого ряда информантов свидетельствуют о

наличии речевого контроля. В докладе предлагается первичная классификация феноменов, которые могут свидетельствовать о повышенном речевом самоконтроле вследствие действия «микрофонного эффекта», а также некоторые наблюдения над явлениями, свидетельствующими о стремлении информантов управлять речевым поведением собеседников (коммуникантов).