

УДК 81'27

## О СПОСОБАХ СОЦИОЛИНГВИСТИЧЕСКОЙ БАЛАНСИРОВКИ УСТНОГО КОРПУСА (на примере «Одного речевого дня») <sup>1</sup>

**Екатерина Михайловна Баева**

к. филол. н., доцент кафедры английского языка

Санкт-Петербургский государственный университет

199034, Санкт-Петербург, Университетская наб., 11. kaverita@yandex.ru

Статья посвящена описанию различных способов социолингвистической балансировки, которые используются при составлении устных корпусов того или иного языка. Дается обзор социолингвистических принципов, лежащих в основе формирования некоторых западных и российских корпусов, а также ставится проблема социолингвистической несбалансированности устного корпусного материала, с которой часто сталкиваются лингвисты, изучающие языковое существование современного города. Основное внимание в статье сосредоточено на создающемся в Санкт-Петербурге звуковом корпусе «Один речевой день». Для достижения максимальной сбалансированности и представительности материала при формировании данного корпуса авторами была разработана собственная методика. На основе данных Федеральной службы городской статистики и общих тенденций лингвистического описания современного города был составлена анкета для информантов, которая в определенной степени стала ориентиром для подбора кандидатов, осуществляемого исследователями в частном порядке. В данной анкете отражены такие социальные факторы, как пол, возраст, место рождения и наиболее длительного проживания информантов, их прошлые и нынешние профессии, характер и уровень образования, родной язык, другие языки, которыми владеют информанты, а также место рождения и социальное происхождение их родителей. Значимость каждого фактора объясняется и обсуждается в статье отдельно. Описываются сбалансированные группы и подгруппы, на которые потенциально делится материал данного устного корпуса.

**Ключевые слова:** социолингвистика; социолингвистическая балансировка; устный корпус; лингвистическое описание; методика корпусного исследования; звуковой корпус «Один речевой день».

Социолингвистические исследования, набравшие силу во второй половине XX в., в настоящее время занимают ведущие позиции в изучении устной речи. На современном этапе невозможно представить себе изучение речи человека в отрыве от его как минимум гендерных и социальных характеристик (см.: [Крысин 1989; Сиротинина 2008; Ерофеева Е. В. 2009; Ерофеева Т. И. 2009, Звуковой корпус... 2013]).

Составление корпуса современной устной речи – очень важная и сложная задача, и балансировка материала – одна из ее составляющих. Когда речь заходит о принципах балансировки такого корпуса, необходимо прежде всего осознать конечные цели его создания и использования. В идеале корпус, как и любой другой научный продукт, должен служить не только теоретическим целям изучающих современный язык лингвистов, но и практическим материалом, кото-

рый был бы полезен, например, студентам, изучающим данный язык как иностранный, исследователям, занимающимся автоматическим распознаванием речи, обычным людям, проявляющим интерес к живому языку<sup>2</sup>. Поэтому составители каждого корпуса стремятся к многообразию и представительности записанной в нем речи, и вопрос классификации материала встает перед исследователями особенно остро.

*Корпусная лингвистика* – одно из самых активно развиваемых направлений современной науки, но пока не существует единого, общепризнанного, принципа подбора материала и информантов, поэтому все корпуса устной речи составлены, как правило, с учетом тех принципов балансировки, которые разделяют их составители.

Например, устная часть *Британского национального корпуса* (BNC), включающая 10 млн

словоупотреблений, сбалансирована с учетом территориального, возрастного и классового признаков информантов. Она состоит из двух частей – демографической и контекстообразованной [Reference Guide... 2007]. В демографической части BNC приняло участие 124 волонтера, набранных с помощью статистического бюро. Это были проживающие в разных районах Великобритании мужчины и женщины разных возрастов из четырех социальных групп, статистически различаемых в британском обществе [Wilmshurst, MacKay 1999]:

1) AB – высший и средний классы (менеджеры и профессионалы высшего и среднего звена, занимающие руководящие должности);

2) C1 – промежуточное звено (младший менеджмент, административные работники, низшие ступени управленческих должностей);

3) C2 – квалифицированные рабочие;

4) DE – полуквалифицированные специалисты и рабочие, не получившие специального образования, а также безработные, пенсионеры и зависимые от государства слои населения.

Составители данного корпуса различают 6 возрастных групп: 0–14 лет, 15–24 года, 25–34 года, 35–44 года, 45–59 лет, 60 лет и старше.

Информантов для демографической части BNC подбирали так, чтобы получилось равное количество мужчин и женщин, примерно равное количество представителей всех возрастных и социальных групп. По возможности, участники записи вели заметки, отмечая пол, возраст, а также социальную и профессиональную принадлежность своих собеседников (коммуникантов).

Контекстообразованная часть корпуса ориентирована на речевые сценарии, а не на информантов и состоит из записей, сделанных в следующих коммуникативных ситуациях:

- образовательные или информативные мероприятия (лекции, занятия, семинары, новостные передачи);

- деловые мероприятия (офисные переговоры, юридические консультации, рекламные акции, встречи профсоюзов, собеседования при приеме на работу);

- публичные мероприятия (церковные службы, политические речи, парламентские процедуры, различные собрания);

- развлекательные мероприятия (спортивные комментарии, послеобеденные речи, встречи в частных клубах, радиопрограммы).

Австралийская национальная база данных устного языка (ANDOSL) подразделяется на три части: записи различных вариантов литературного австралийского английского – «общего»,

«широкого» и «изысканного» [Cox, Palethorpe 2012]; в записях каждой части приняло участие 36 информантов, по 6 человек каждого пола в трех возрастных группах: 18–30 лет, 31–45 лет и более 46 лет.

*Подкорпус устной речи* (часть Национального корпуса русского языка – НКРЯ, – с 2007 г. существующая самостоятельно) составлен из магнитофонных записей публичной и частной устной речи, а также транскриптов кинофильмов. Как и любой национальный корпус, Корпус устной речи (КУП) включает в себя тексты живой речи жителей различных регионов: Москвы, Санкт-Петербурга, Саратова, Екатеринбурга, Новосибирска, Воронежа и др. Исследователи отмечают, что «вопрос о какой бы то ни было социологической сбалансированности устного подкорпуса НКРЯ <...> не ставится по экстралингвистическим причинам <...> однако при конструировании устного подкорпуса мы стараемся ориентироваться, по крайней мере, на жанровую и тематическую сбалансированность» [Гришина 2005: 95]. В КУП присутствует некоторая социологическая разметка: характеристика словоупотребления с точки зрения пола и возраста говорящего, однако эта информация не всегда доступна пользователям [Гришина, Савчук 2009: 137].

Согласно исследованиям Л. П. Крысина, изучающего распределение языковых вариантов в среде носителей русского литературного языка в зависимости от трех факторов: возрастного, профессионального и социально-территориального, наибольшим влиянием на речь человека обладает *территориальный признак*, затем возраст и социальное происхождение [Крысин 1989]. Неудивительно, что в последние десятилетия второе рождение получили исследования так называемого *языка города* [Земская 2004]. Появились целые научные школы, занимающиеся изучением живой городской русской речи: в Перми, Саратове, Омске, Москве и Санкт-Петербурге.

Язык города – это язык его жителей, и при составлении корпуса устной речи представляется важным использование термина *социолект*, введенного Т. И. Ерофеевой и означающего «набор языковых кодов, которым владеют индивиды, объединенные по какой-нибудь страте» [Ерофеева Т. И. 2009: 29], где под *стратой* понимается «конституирующий признак биологических, социальных и психологических свойств, характерных для определенной области (группы)» [там же].

Корпус записей живой речи, создаваемый в Перми, учитывает следующие страты [там же: 34–35]:

- 1) место рождения (областной центр, районный центр, село);
- 2) возраст (группы 20–29 лет, 30–39 лет, 40–49 лет, 50 лет и более);
- 3) образование (высшее и среднее);
- 4) специальность (гуманитарии, негуманитарии);
- 5) род занятий (используется для описания жаргонизмов и аргоизмов);
- 6) пол;
- 7) темперамент.

База данных «Русская устная спонтанная речь Пермского края», создаваемая в Перми, учитывает следующие страты: родной язык, место рождения; возраст; образование; специальность; пол [Боронникова и др. 2014; Ерофеева Т.И., Ерофеева Е.В., Грачева 2000].

М. В. Китайгородская и Н. Н. Розанова, вслед за О. Б. Сиротининой [Сиротинина 2008: 147], утверждают, что «идеально описание языка города должно опираться на информацию социологического характера, однако на современном этапе, в условиях сильной миграции населения и отсутствия надежных социологических данных по социальному, национальному, профессиональному и численному составу жителей того или иного города, такое исследование достаточно сложно осуществить, особенно если речь идет о таком мегаполисе, как Москва» [Китайгородская, Розанова 2010: 25]. Поэтому в цитируемой монографии «языковое существование» Москвы описывается как «единое сложноструктурированное коммуникативное пространство повседневности», и «речь горожан исследуется в структуре городской коммуникации с учетом всего набора ее элементов – говорящий, адресат, средства коммуникации, параметры ситуации» [там же: 32]. Таким образом, основным методом формирования данного корпуса становится жанрово-ситуативный.

При создании *Сбалансированной аннотированной картотеки* (САТ), одного из блоков *Звукового корпуса русского языка* (ЗКРЯ), социолингвистическую балансировку информантов осуществляли путем сознательной выборки, когда отбирают строго определенные группы респондентов, обладающих важными для исследователя признаками, при этом в отдельных исследованиях часть характеристик носителей языка нивелируется для изучения влияния определенного критерия [Звуковой корпус... 2013: 74].

В. В. Куканова, один из создателей САТ, полагает, что «изучение экстралингвистических факторов, влияющих на речь говорящего, удобно проводить путем анализа бинарных оппозиций, т. е. путем противопоставления каждый раз двух групп, различающихся по тем или иным призна-

кам, и выявления их сходства и различия» [Куканова 2007: 45]. В ее статье предлагается следующий набор оппозиций:

- гендерная (мужчины / женщины);
- возрастная (социально активный возраст делится на два периода: *до 35 лет* – молодость / *после 35 лет* – зрелость);
- оппозиция по *уровню речевой компетенции* (УРК – высокий / низкий);
- оппозиция по *отношению говорящего к речи* (профессиональное / непрофессиональное).

Главная задача коллектива исследователей, работающих над проектом «*Один речевого день*» (ОРД), вторым блоком Звукового корпуса русского языка, – научное описание особенностей русского языка повседневного общения больших социальных групп современного российского города (на фонетическом, лексическом, грамматическом и дискурсивном уровнях). При решении подобного рода задач, как представляется, уже невозможно оперировать бинарными оппозициями, особенно в области социальных групп города.

Как уже отмечалось, вопрос выработки методики отбора информантов до сих пор не решен окончательно, поэтому создателям ОРД необходимо было разработать собственную методику, обеспечивающую социальную балансировку корпусных данных.

Для понимания реальной социальной ситуации в Санкт-Петербурге были проанализированы данные Федеральной службы городской статистики, опубликованные по результатам Всероссийской переписи населения в 2010 г. [Социально-демографический портрет... 2012]. На основе этих данных и общих тенденций лингвистического описания современного города была разработана анкета для информантов, которая в определенной степени стала ориентиром для подбора кандидатов, осуществляемого исследователями в частном порядке.

В социологической анкете информантам ОРД предлагается указать следующие собственные характеристики:

- 1) пол,
- 2) возраст,
- 3) место рождения,
- 4) родной язык,
- 5) другие языки, которыми владеет,
- 6) национальность родителей (опционально),
- 7) социальное происхождение родителей,
- 8) образование,
- 9) квалификация/специальность по диплому,
- 10) прошлые профессии,
- 11) кем работает в настоящее время,

12) основные места проживания (регионы, города), где информант жил больше года.

Рассмотрим подробнее все пункты этого списка.

#### **Пол**

На данном этапе все записи для корпуса «Один речевой день» ведутся в Санкт-Петербурге. Согласно переписи 2010 г., здесь проживает 2188,9 тыс. мужчин и 2690,7 тыс. женщин, это 45 и 55 % соответственно [Социально-демографический портрет... 2012: 35], т. е. соотношение полов (гендерных групп) примерно равное. На настоящий момент (сентябрь 2014 г.) в корпусе ОРД записано 100 информантов и соотношение мужчин и женщин в нем такое же.

#### **Возраст**

Информанты указывают в анкете свой точный возраст, что дает возможность сформировать несколько возрастных групп участников записи:

- 18–24 года,
- 25–34 года,
- 35–44 года,
- 45–59 лет,
- 60 лет и старше.

На данный момент формирования корпуса ОРД состав всех перечисленных групп хорошо сбалансирован, за исключением группы старшего поколения, что легко объяснить нежеланием людей пожилого возраста принимать участие в эксперименте с записывающей техникой, боязнию «сломать» диктофон или «сделать что-нибудь не то».

Такая классификация отражает определенные социальные этапы жизни горожанина: до 24 лет – молодость, это время студенчества и первые годы работы, когда человек отрывается от родительского дома и находит свое место в социальной структуре общества; 25–34 и 35–44 – десятилетия, когда человек активно работает, а также заводит семью и из категории «ребенка» окончательно переходит в категорию «родителя»; 45–59 лет – время от социальной зрелости до пенсионного возраста; выше 60 лет – старшее поколение, менее активно принимающее участие в социальной жизни.

#### **Место рождения**

Данный пункт анкеты, равно как и пункт *основные места проживания (регионы, города), где жил больше года*, связан с необходимостью учета влияния на речь говорящего территориального фактора, о важности которого уже шла речь выше. Как правило, информантами становятся люди, длительное время проживающие в Петербурге, однако география мест их рождения варьируется от стран ближнего зарубежья до различ-

ных регионов России, и, по всей вероятности, с учетом происхождения его жителей языковой облик города будет подвержен влиянию регионов. В дальнейшем подобные записи планируется провести и в других городах России, чтобы можно было сделать сравнительный анализ.

#### **Родной язык**

Это одна из бинарных оппозиций, присутствующих в формировании корпуса. Информантам предлагается указать, русский или нерусский язык они считают родным. В настоящий момент ведутся записи только тех информантов, для которых родным языком является русский. Учитывая, что Петербург – это мегаполис, куда приезжают учиться и работать люди из разных регионов России и зарубежья, весьма перспективными кажутся будущие записи информантов, для которых русский будет уже вторым языком.

#### **Другие языки, которыми владеет информант**

Данный пункт был включен в анкету, чтобы проследить в речи информантов возможные проявления межъязыковой интерференции, т. е. тех «случаев отклонения от норм любого из языков, которые происходят в речи двуязычных в результате того, что они знают больше языков, чем один, то есть вследствие языкового контакта» [Вайнрайх 1979: 22].

#### **Национальность родителей**

Этот пункт анкеты является необязательным для заполнения, однако его данные также способствуют выявлению случаев интерференции и возможного территориального влияния; например, если один из родителей является иностранцем, однако сам информант родился и вырос в России. При переписи населения также не обязательно было указывать свою национальность. По последним имеющимся сведениям из всех жителей Санкт-Петербурга и Ленинградской области, указавших свою национальную принадлежность, 92,6 % – русские [Социально-демографический портрет... 2012: 95].

#### **Социальное происхождение родителей**

Социальное происхождение родителей информантов до некоторой степени позволяет определить ту языковую среду, в которой они росли и воспитывались, а также указать на случаи внутриязыковой интерференции, т. е. влияния на повседневную речь человека таких языковых разновидностей, как просторечие, профессиональная речь или арго. Можно также выделить различные способы заполнения этого пункта: так, одни информанты указывают конкретные профессии родителей, другие обобщают и пишут: «крестьянка», «служащие», «рабочий», «ученый».

### Образование

В данном случае предполагается указание не профессии, а уровня образования информанта: среднее, средне-специальное, неполное высшее, высшее, кандидат или доктор наук. Минимальный отмечаемый среди информантов ОРД уровень образования – средний. По данным переписи, на 1000 чел. в возрасте 15 лет и более, указавших свой уровень образования, в Санкт-Петербурге имеют профессиональное образование: высшее – 365 чел., неполное высшее – 74, среднее – 333, начальное – 27; имеют общее образование: среднее – 125 чел., основное – 53, начальное – 21, не имеют начального образования – 2 [Социально-демографический портрет... 2012: 110]. Таким образом, более 40 % опрошенных имеют или получают высшее образование, еще треть – среднее, и это при учете того, что в опросе принимали участие лица от 15 лет, которые, как правило, еще посещают среднюю школу. Степень кандидата наук имеют 45 из 1000 опрошенных в возрасте от 20 лет (4,5 %), доктора наук – 10 (1 %) [там же: 112].

В записях ОРД принимают участие информанты от 18 лет, и доля лиц с высшим образованием среди них больше, чем по данным переписи: около 75 %. Помимо того что в городской среде в целом отмечается рост населения с высшим образованием (что уже является достаточным основанием для такой несбалансированности материала корпуса), стоит отметить, что среди потенциальных информантов с низким уровнем образования многие просто отказывались принимать участие в эксперименте, не понимая его важности.

Пункты 8–10 анкеты указывают на различные контакты информанта в его настоящей и прошлой профессиональной деятельности и должны быть проанализированы в едином контексте:

**квалификация/профессия по диплому;  
прошлые профессии;  
кем работает в настоящее время.**

М. М. Бахтин писал, что «язык в каждый данный момент его становления расслоен не только на лингвистические диалекты в точном смысле слова <...>, но <...> на социально-идеологические языки: социально-групповые, “профессиональные”, “жанровые”, языки поколений и т. п.» [Бахтин 1975: 85]. Б. А. Ларин, говоря о языке города, отмечал, что каждый человек потенциально является носителем городского многоязычия, особых социальных диалектов, которые определяют принадлежность этого человека к той или иной социальной среде [Ларин 1977: 190–191]. Одним из ключевых языковых типов, которыми потенциально владеет городской жи-

тель, безусловно, является профессиональная речь. Профессиональная речь находится в ситуации диглоссии с кодифицированным литературным языком и с разговорной речью. В профессиональной среде в целях непринуждённого или официального общения говорящие пользуются профессиональной речью, а при общении в других ситуациях используют устную литературную речь. Контакт между устной литературной и профессиональной речью отдельных социальных групп – медиков, юристов, «компьютерщиков», философов, преподавателей разной специализации – уже в определенной степени выступал объектом внимания авторов Звукового корпуса русского языка [Звуковой корпус... 2014: 173–194], однако подобные исследования можно и необходимо продолжать, в том числе и с привлечением других профессиональных языков.

Имеющийся на данный момент материал корпуса «Один речевой день» способствует изучению и описанию социолектов следующих социальных групп:

- работники, занятые на *производстве, строительстве, транспорте* (инженер, оператор, кондуктор), а также представители профессий, связанных с физическим трудом (мастер по окнам, столяр, механик и прочие *ремесленники*);
- служащие *силовых структур* (военные, работники правоохранительных органов, курсанты);
- работники *сферы услуг* (продавец, кассир, торговый представитель, курьер, менеджер доставки, официант, мастер по маникюру и проч.);
- специалисты по *экономической деятельности* (экономист, бухгалтер, логист, финансист);
- специалисты по *информационным технологиям* (IT-инженер, программист, сотрудник службы технической поддержки, web-разработчик);
- специалисты по *связям с общественностью и рекламе* (маркетолог, PR-менеджер, пресс-секретарь, менеджер по рекламе);
- специалисты, занятые в *спорте* (спортсмен, фитнес-инструктор);
- представители *творческих профессий* (фотограф, дизайнер, художник, архитектор);
- представители *гуманитарных наук* (лингвист, археолог, историк, философ, психолог);
- представители *естественных наук* (биолог, химик, математик, астроном);
- работники *образования* (преподаватели-практики гуманитарных и негуманитарных специальностей);
- *учащиеся* (студенты гуманитарных и негуманитарных специальностей);
- *неработающие пенсионеры*.

Профессия, указанная в дипломе информанта, помогает исследователю предположить, что как минимум 4 года (срок, за который человек может получить степень бакалавра) человек был связан с определенной профессиональной средой: посещал лекции и занятия, читал учебники, т. е. даже если после окончания учебного заведения он не начинает работать по профессии, то имеет теоретические знания о данном предмете. Прошлые и нынешняя профессии информанта также указывают на его вовлеченность в профессиональные круги.

Всех информантов, независимо от их профессиональной принадлежности, можно также разделить на категории по *должности/служебному положению*:

- руководящие работники высшего звена (руководитель, заместитель руководителя предприятия, учреждения, фирмы);
- руководящие работники среднего звена (руководитель подразделения, департамента, группы, отдела);
- служащие, специалисты, технические исполнители;
- рабочие.

Это распределение информантов на категории не может быть сбалансированным, так как оно осуществляется на основании уже имеющихся сведений. На данный момент доля руководящих работников, принявших участие в записи в качестве информантов, составляет около 5 %, однако потенциально может увеличиться за счет коммуникантов. По условиям эксперимента запись проводится в течение целого дня, и по ее итогам каждый информант составляет таблицу своих коммуникантов, т. е. тех людей, с которыми он общается на протяжении дня, и вносит их персональные данные в специальную анкету, посвященную собеседникам<sup>3</sup>. Эта анкета содержит те же сведения, которые информант предоставляет о себе, однако все ее поля являются опциональными для заполнения, так как информант вполне может не знать о собеседнике почти ничего (например, когда коммуникантом является попутчик в транспорте). Таким образом, каждый информант-служащий в течение своего рабочего дня имеет возможность поговорить с коммуникантом-руководителем, речь которого также записывается и расшифровывается, что позволяет существенно увеличить группу руководящих работников.

На основании данных о специальности и профессии всех информантов также можно разделить на четыре группы по принципу *профессионального или непрофессионального отношения к*

*речи*. Согласно этому подходу в одну группу объединяются говорящие, для которых

1) «язык – только средство коммуникации (все нефилологи, а также люди, профессионально не работающие с речью);

2) язык – средство коммуникации и объект изучения (все филологи – не преподаватели, в том числе студенты);

3) язык – средство коммуникации и орудие труда (нефилологи, чья профессия связана с речью – преподаватели, актеры, дикторы, лекторы и т. п.);

4) язык – средство коммуникации, объект изучения и орудие труда (преподаватели-филологи)» [Богданова 2001: 58].

Исследователи уже второе десятилетие пользуются данной классификацией Н. В. Богдановой, на практическом материале доказывая ее важность и состоятельность (см., например: [Степихов 2003, 2005; Бродт 2007; Сапунова 2009; Хан 2013 и др.]). Стоит отметить, что эта классификация в какой-то степени коррелирует с психолингвистической теорией Ю. Н. Караулова об изучении языковой личности, согласно которой каждый человек имеет некий набор умений, или готовностей к разного рода речемыслительной деятельности. Формирование такого набора готовностей определяется далеко не субъективными характеристиками и психологическими факторами, а в первую очередь социальными условиями и соответствующими ролями языковой личности: «одно дело готовности среднего носителя языка, и другое дело готовности филолога-русиста или преподавателя русского языка» [Караулов 2003: 65].

Очевидно, что отношение к речи формирует речевые способности человека, повышает его готовность решать коммуникативные задачи любой степени сложности. Именно с опорой на эту стратегию, в совокупности с уровнем образования, устанавливается еще один неотъемлемый признак социологической характеристики информанта – *уровень его речевой компетенции*. Уровень речевой компетенции (УРК) – это «степень свободы говорящего в выборе речевых средств, уровень владения этими возможностями, его способность решать те или иные коммуникативные задачи» [Звуковой корпус... 2013: 86].

Можно выделить три уровня речевой компетенции говорящего:

- высокий (высшее образование в совокупности с профессиональным отношением к речи);
- средний (высшее образование и непрофессиональное отношение к речи);

• низкий (отсутствие высшего образования и непрофессиональное отношение к речи).

Среди информантов ОРД можно встретить носителей всех УРК, хотя, например, при составлении картотеки САТ использовалась по преимуществу бинарная оппозиция высокий – низкий УРК на основании того предположения, что средний УРК маркируется, как правило, средними количественными показателями, которые получают при статистической обработке результатов по низкому и высокому уровням [Богданова 1993; Куканова 2007: 48].

Таким образом, можно сказать, что в процессе работы над формированием корпуса ОРД авторам удалось решить многие проблемы балансировки состава информантов, вставшие перед ними в начале пути (см.: [Зобнина 2008]). В частности, сегодня можно говорить о следующих фактах.

1. Достигнута сбалансированность корпуса по гендерному признаку.

2. Достигнуто равновесие в возрастных группах.

3. Информанты со средним образованием по-прежнему составляют лишь 20 % всех участников, однако эта цифра соответствует реальным демографическим данным по Санкт-Петербургу и является отражением облика города.

4. Создана многоступенчатая (по профессии, служебному положению, уровню речевой компетенции) классификация социальных групп информантов и потенциальных социолектов, в которую в перспективе предлагается включить и коммуникантов, что существенно расширит возможности исследования.

5. Обоснована необходимость включения в анкету пункта о родном языке информантов.

Представляется, что описанный способ социолингвистической балансировки устного корпуса удачно сочетает многие существующие подходы к его созданию и может послужить жизнеспособной моделью и ориентиром для дальнейших исследований.

### Примечания

<sup>1</sup> Исследование выполнено при поддержке гранта РНФ № 14-18-02070 «Русский язык повседневного общения: особенности функционирования в разных социальных группах».

<sup>2</sup> Ср. замечание В. П. Селегея в докладе на Международной филологической конференции в СПбГУ (март 2013 г.): «В современной лингвистике корпус – устной или письменной речи – это и объект, и инструмент исследования».

<sup>3</sup> Такая методика применялась в записи устной части Британского национального корпуса (см. выше).

### Список литературы

Бахтин М. М. Слово в романе // М. М. Бахтин Вопросы литературы и эстетики. Исследования разных лет. М.: Худож. лит., 1975. С. 72–233.

Богданова Н. В. Живые фонетические процессы русской речи: пособие по спецкурсу. СПб.: Филол. фак-т С.-Петерб. гос. ун-та, 2001. 189 с.

Богданова Н. В. О синтаксических коррелятах дифференциации уровней речевой культуры // Живое слово в русской речи Прикамья. Пермь: Изд-во Перм. ун-та, 1993. С. 92–100.

Боронникова Н. В. и др. Проект базы данных «Русская устная спонтанная речь Пермского края» / Н. В. Боронникова, Е. В. Ерофеева, Т. И. Ерофеева, Е. С. Худякова // Русско-македонски јазични, културни и литературни врски. Скопје, 2014. Вып. 5. С. 137–145.

Бродт И. С. Спонтанный монолог в социолингвистическом и лингвистическом аспектах: дисс. ... канд. филол. наук. СПб., 2007. 289 с.

Вайнрайх У. Языковые контакты / пер. с англ. Ю. А. Жлуктенко. Киев: Вища школа, 1979. 364 с.

Гришина Е. А. Устная речь в Национальном корпусе русского языка // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005. С. 94–110.

Гришина Е. А., Савчук С. О. Корпус устных текстов в НКРЯ: состав и структура // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 129–149.

Ерофеева Е. В. Языковая ситуация Пермского края: особенности русской спонтанной речи и методы исследования // Вестник Пермского университета. Российская и зарубежная филология. 2013. Вып. 3(23). С. 7–16.

Ерофеева Е. В., Ерофеева Т. И., Грачева И. И. Городские социолекты: Пермская городская речь: звучащая хрестоматия / Пермь; Бохум, 2000. 172 с. CD.

Ерофеева Т. И. Социолект: Стратификационное исследование / Перм. гос. ун-т. Пермь, 2009. 240 с.

Звуковой корпус как материал для анализа русской речи. Ч. 1. Чтение. Пересказ. Описание / отв. ред. Н. В. Богданова-Бегларян. СПб.: Филол. фак-т С.-Петерб. гос. ун-та, 2013. 532 с.

Звуковой корпус как материал для анализа русской речи. Ч. 2. Теоретические и практические аспекты анализа (Т. 1. О некоторых особенностях устной спонтанной речи разного типа. Звуковой корпус как материал для преподавания русского языка в иностранной аудитории) / отв. ред. Н. В. Богданова-Бегларян. СПб.: Филол. фак-т С.-Петерб. гос. ун-та, 2014. 396 с.

Земская Е. А. Разновидности городской устной речи и задачи их изучения // Земская Е. А. Язык как деятельность: Морфема. Слово. Речь. М.: Языки слав. культур, 2004. С. 247–289.

Зобнина Е. А. Социальные характеристики говорящего: объективные данные и экспертная оценка речи (по материалам звукового корпуса «Один речевой день») // Материалы XXXVII междунар. филол. конф. Вып. 21. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 10–15 марта 2008 г. / отв. ред. А. С. Асиновский; науч. ред. Н. В. Богданова. СПб.: Фак-т филологии и искусств С.-Петербург. гос. ун-та, 2008. С. 17–24.

Караулов Ю. Н. Русский язык и языковая личность. М., 2003. 264 с.

Китайгородская М. В., Розанова Н. Н. Языковое существование современного горожанина: На материале языка Москвы. М.: Языки слав. культур, 2010. 496 с.

Крысин Л. П. Социолингвистические аспекты изучения современного русского языка / отв. ред. Ю. Д. Дешериев. М.: Наука, 1989. 186 с.

Куканова В. В. Об одном из способов подбора информантов в ходе полевого исследования // Материалы XXXVI междунар. филол. конф. Вып. 20. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 12–17 марта 2007 г. / отв. ред. А. С. Асиновский, Н. В. Богданова. СПб.: Фак-т филологии и искусств С.-Петербург. гос. ун-та, 2007. С. 45–52.

Ларин Б. А. К лингвистической характеристике города (несколько предпосылок) // Ларин Б. А. История русского языка и общее языкознание. М.: Просвещение, 1977. С. 189–199.

Сапунова Е. М. Неподготовленное чтение как вид речевой деятельности и тип устного спонтанного монолога (на материале русского языка): дисс. ... канд. филол. наук, СПб., 2009. 237 с.

Сиротинина О. Б. Речь города: проблемы, аспекты и трудности изучения // Язык современного города: тез. докл. междунар. конф. Восьмые Шмелевские чтения. М.: Ин-т рус. яз. им. В. В. Виноградова РАН, 2008. С. 146–148.

Социально-демографический портрет России: По итогам Всероссийской переписи населения 2010 года / Федер. служба гос. статистики. М.: ИИЦ «Статистика России», 2012. 183 с.

Степихов А. А. Спонтанный монолог: грамматика жанра // Грамматическая и лексическая семантика. Памяти Л. Л. Буланина / отв. ред. Н. В. Богданова, Б. Н. Осипов. СПб.: Филол. фак-т С.-Петербург. гос. ун-та, 2003. С. 135–154.

Степихов А. А. Соотношение синтаксического и интонационного членения в спонтанном моно-

логе: дисс. ... канд. филол. наук. СПб., 2005. 197 с.

Хан Н. А. Устные спонтанные монологи разного типа в коммуникативно-дискурсивном аспекте (на материале Звукового корпуса русского языка): автореф. дисс. ... канд. филол. наук. СПб., 2013. 24 с.

ANDOSL – Australian National Database of Spoken Language. URL: <http://andosl.rsis.se.anu.edu.au/andosl> (дата обращения: 25.09.2014).

Cox F. M. & Palethorpe, S. Standard Australian English: The Sociostylistic Broadness Continuum // Hickey R. (ed.) Standards of English: Codified Varieties around the World, Studies in Language, Cambridge University Press, 2012. P. 294–317.

Reference Guide for the British National Corpus [XML Edition] // Ed. by L. Burnard, publ. for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services, 2007. URL: <http://www.natcorp.ox.ac.uk/docs/URG.xml> (дата обращения: 25.09.2014).

Wilmshurst J., MacKay, A., The Fundamentals of Advertising, 1999. 386 p.

## References

ANDOSL – Australian National Database of Spoken Language. Available at: <http://andosl.rsis.se.anu.edu.au/andosl> (accessed: 25.09.2014).

Bakhtin M. M. Slovo v romane [A Word in a Novel] Bakhtin M. M. Voprosy literatury i ehstetiki [Issues of literature and aesthetics]. Moscow, 1975. P. 72–233.

Bogdanova N. V. O sintaksicheskikh korreljatakhdifferenciacii urovnej rechevoj kul'tury [On Syntactic Correlates of the Level of Speech Culture]. Zhivoe slovo v russkoj rechi Prikamja [A Live Word in the Russian Speech of the Kama Region]. Perm, 1993. P. 92–100.

Bogdanova N. V. Zhivye foneticheskie processy v russkoj rechi: posobie po speckursu [Live Phonetic Processes in Russian Speech: a Guide to a Special Course]. St. Petersburg, 2001. 189 p.

Boronnikova N. V. et al. Proekt bazy dannyh «Russkaja ustnaja spontannaja rech Permskogo kraja» [A Database Project of Russian Spoken Speech of the Perm Region] / N. V. Boronnikova, E. V. Erofeeva, T. I. Erofeeva, E. S. Hudjakova // Russian-Macedonian Studies of Language, Culture and Literature. Skopje, 2014. Vol. 5. P. 137–145.

Brodt I. S. Spontannyj monolog v sociolingvističeskom i lingvističeskom aspektakh: diss. ... kand. filol. nauk [Spontaneous Monologue in the Sociolinguistic and Linguistic Aspects: PhD philol. sci. diss.]. St. Petersburg, 2007. 289 p.

Cox F. M. & Palethorpe, S. Standard Australian English: The Sociostylistic Broadness Continuum, in



Hickey, R. (ed.) *Standards of English: Codified Varieties around the World*, Studies in Language, Cambridge University Press, 2012. P. 294–317.

Erofeeva E. V., Erofeeva T. I., Gracheva I. I. *Gorodskie sociolekty: Permskaja gorodskaja rech: zvuchashhaja hrestomatija* [Sociolects of a City: The Urban Speech of Perm: a Sounding Anthology]. Perm; Bohum, 2000. 172 p. CD.

Erofeeva E. V. *Jazykovaja situacija sovremennogo goroda i metody ee issledovanija* [Social Component of Lexical Connotation]. Vestnik Permskogo universiteta. Rossijskaja i zarubezhnaja filologija [Perm University Herald. Russian and Foreign Philology]. 2009. Iss. 5. P. 16–23.

Erofeeva T. I. *Sociolekt: Stratifikacionnoe issledovanie* [Sociolect: Stratification Research]. Perm, 2009. 240 p.

Grishina E. A. *Ustnaja rech v Nacional'nom korpuse russkogo jazyka* [Spoken Speech in National Russian Corpus]. Nacional'nyj korpus russkogo jazyka: 2003–2005. [National Russian Corpus 2003–2005]. Moscow, Indrik Publ., 2005. P. 94–110.

Grishina E. A., Savchuk S. O. *Korpus ustnyh tekstov v NKRJa: sostav i struktura* [The Corpus of Spoken Texts in National Russian Corpus: Contents and Structure]. Nacional'nyj korpus russkogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. [National Russian Corpus: 2006–2008. New Results and Prospects]. St. Petersburg: Nestor-Istorija Publ., 2009. P. 129–149.

Karaulov Ju. N. *Russkij jazyk i jazykovaja lichnost'* [The Russian Language and the Linguistic Personality]. Moscow, 2003. 264 p.

Khan N. A. *Ustnye spontannye monologi raznogo tipa v kommunikativno-diskursivnom aspekte (na materiale Zvukovogo korpusa russkogo jazyka): avtoref. diss. ... kand. filol. nauk* [Spoken Spontaneous Monologues of Various Types in Communicative and Discourse Aspects (on the Data of the Spoken Corpus of the Russian Language): Thesis Synopsis of PhD philol. sci. diss.]. St. Petersburg, 2013. 24 p.

Kitajgorodskaja M. V., Rozanova N. N. *Jazykovo sushhestvovanie sovremennogo gorozhanina: Na materiale jazyka Moskvy* [Linguistic Existence of a Modern Citizen; on the Material of the Language of Moscow]. Moscow, Jazyki slavjanskikh kultur Publ. 2010. 496 p.

Krysin L. P. *Sociolingvisticheskie aspekty izuchenija sovremennogo russkogo jazyka* [Sociolinguistic Aspects of Studying Contemporary Russian Language]. Moscow, 1989. 183 p.

Kukanova V. V. *Ob odnom iz sposobov podbora informantov v hode polevogo issledovanija* [On One of the Ways of Choosing Informants for a Field Research] // *Materialy XXXVI mezhdunarodnoj filologicheskoi konferencii*. Vyp. 20. Polevaja ling-

vistika. Integral'noe modelirovanie zvukovoj formy estestvennykh jazykov [Proceedings of XXXVI International philological conference. Iss. 21. Field Linguistics. Integral Modeling of the Spoken Form of Natural Languages]. Ed. by A. S. Asinovskij, N. V. Bogdanova. St. Petersburg, 2007. P. 45–52.

Larin B. A. *K lingvisticheskoj kharakteristike goroda (neskol'ko predposylok)* [On the Linguistic Characteristics of a City (a Few Assumptions)]. B. A. Larin. *Istorija russkogo jazyka i obshchee jazykoznanie* [The History of the Russian Language and General Linguistics]. Moscow, Prosveshhenie Publ., 1977. P. 189–199.

*Reference Guide for the British National Corpus [XML Edition]*. Ed. by L. Burnard, publ. for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services, 2007. Available at: <http://www.natcorp.ox.ac.uk/docs/URG.xml> (accessed: 25.09.2014).

Sapunova E. M. *Nepodgotovlennoe chtenie kak vid rechevoj dejatel'nosti i tip ustnogo spontannogo monologa (na materiale russkogo jazyka): diss. ... kand. filol. nauk* [Unprepared Reading as a Variety of Speech Activity and a Type of a Spoken Spontaneous Monologue (in the Russian Language): PhD philol. sci. diss.]. St. Petersburg, 2009. 237 p.

Sirotinina O. B. *Rech goroda: problemy, aspekty i trudnosti izuchenija* [The City Speech: Problems, Aspects and the Difficulties of Studying]. Jazyk sovremennogo goroda [The language of a Contemporary City]. Moscow: Institute of Russian Language named after V. V. Vinogradov RAS, 2008. P. 146–148.

*Social'no-demograficheskij portret Rossii: Po itogam Vserossijskoj perepisi naselenija 2010 goda*. Federalnaja sluzhba gosudarstvennoj statistiki [Social and demographic portrait of Russia: the result of population census of 2010] [by Federal Agency of Urban Statistics]. Moscow, Statistika Rossii Publ., 2012. 183 p.

Stepikhov A. A. *Sootnoshenie sintaksicheskogo i intonacionnogo chlenenija v spontannom monologe: diss. ... kand. filol. nauk* [The Correlation of Syntactic and Intonation Partitioning in Spontaneous Monologues: PhD philol. sci. diss.]. St. Petersburg, 2005. 197 p.

Stepikhov A. A. *Spontannyj monolog: grammatika zhanra* [Spontaneous Monologue: the Genre Grammar]. Grammaticheskaja i leksicheskaja semantika. Pamjati L. L. Bulanina [Grammatical and Lexical Semantics. In memory of L. L. Bulanin]. Ed. by N. V. Bogdanova, B. N. Osipov. St. Petersburg, 2003. P. 135–154.

Weinreich U. *Jazykovye kontakty* [Linguistic Contacts]. Kiev, Vishcha shkola Publ., 1979.

*Wilmshurst J., MacKay A., The Fundamentals of Advertising, 1999. 386 p.*

*Zemskaja E. A. Raznovidnosti gorodskoj ustnoj rechi i zadachi ikh izuchenija [Varieties of Urban Spoken Speech and the Aims of Their Studying]. Zemskaja E. A. Jazyk kak dejatel'nost': Morfema. Slovo. Rech [Language as an Activity: Morpheme. Word. Speech]. Moscow, 2004. P. 247–289.*

*Zobnina E. A. Social'nye kharakteristiki govorjashhego: objektivnye dannye i ehkspertnaja ocenka rechi (po materialam zvukovogo korpusa «Odin rechevoj den'») [Social Characteristics of the Speaker: Objective Data and Experts' Opinion] // Materialy XXXVII mezhdunarodnoj filologicheskoy konferencii. Vyp. 21. Polevaja lingvistika. Integral'noe modelirovanie zvukovoj formy estestvennykh jazykov [Proceedings of XXXVII International philological conference. Iss. 21. Field Linguistics. Integral Modeling of the Spoken Form of Natural Languages] Ed. by A. S. Asinovskij, N. V. Bogdanova. St. Petersburg, 2008. P. 17–24.*

*Zvukovoj korpus kak material dlja analiza russkoj rechi. Chast' 1. Chtenie. Pereskaz. Opisanie [Sounding Corpus as the Material of Analysis of the Russian Speech. Part 1. Reading. Retelling. Picture Description]. Ed. by N. V. Bogdanova-Beglarjan. St. Petersburg, 2013. 532 p.*

*Zvukovoj korpus kak material dlja analiza russkoj rechi. Chast' 2. Teoreticheskie i prakticheskie aspekty analiza (Tom 1. O nekotoryh osobennostjah ustnoj spontannoj rechi raznogo tipa. Zvukovoj korpus kak material dlja prepodavanija russkogo jazyka v inostrannoj auditorii). Kollektivnaja monografija [Sounding Corpus as the Material of Analysis of the Russian Speech. Collective Monograph. Part 2. Theoretic and Practical Aspects of Analysis. Vol. 1. On Some Particularities of Spoken Spontaneous Speech of Various Kinds. Sounding Corpus as the Material for Teaching Russian as a Foreign Language]. Ed. by N. V. Bogdanova-Beglarjan. St. Petersburg, 2014. 396 p.*

## ON MEANS OF SOCIOLINGUISTIC BALANCING OF A SPOKEN CORPUS (drawn on “One Speaker’s Day”)

**Ekaterina M. Baeva**

Associate Professor in the Department of English Language  
Saint Petersburg State University

The article considers different means of sociolinguistic balancing involved in forming spoken corpora. It gives a review of sociolinguistic principles which underlie some of the foreign and Russian corpora (for instance, the British National Corpus and the Russian National Corpus) and states the problem of unbalanced material, which linguists often encounter. The main object of the study is the sound corpus “One Speaker’s Day”, being recorded in Saint Petersburg, and the methods of its balancing and forming used by its developers. It is the second part of the corpus, the first being an annotated collection of texts which were analyzed by the means of binary oppositions. It has not always proved suitable for the corpus. In order to obtain sociolinguistically balanced material for research, the authors designed a special questionnaire based on the data of Federal Agency of Urban Statistics and general trends of linguistic description of a city. This questionnaire was used to select the candidates to participate in recording for the project. It includes such important sociolinguistic characteristics as gender, age, place of birth and places of living of the informants, their former and current occupations, their profession and education, their native language and other languages they speak, as well as their parents’ place of birth and social origins. The importance of each factor is described and justified in this study. Using certain factors, for example gender and native language, it is possible to divide the informants of a corpus research in two well-balanced groups, but others, such as age and occupation, demand a more precise classification. Therefore, the research aims to categorize the material of “One Speaker’s Day” by possible balanced groups and subgroups and states the necessity for a detailed analysis. It suggests that certain problems of balancing a corpus mentioned in the review can be avoided using this method.

**Key words:** sociolinguistics; sociolinguistic balancing; spoken corpus; linguistic description; sounding corpus; One Speaker’s Day; method of corpus studies.