

НКРЯ КАК ИНСТРУМЕНТ СОЦИОЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ. ЭПИЗОД IV. ПОЛ ГОВОРЯЩЕГО И ДЛИНА РЕПЛИКИ

Даниэль М. А. (misha.daniel@gmail.com)
НИУ ВШЭ, МГУ, Москва, Россия

Зеленков Ю. Г. (yuryz@yandex.ru)
Яндекс, Москва, Россия

В статье обсуждается использование Национального корпуса русского языка для социолингвистических исследований. Предлагается использовать имеющуюся в Устном корпусе разметку по полу говорящего в качестве основы для статистического анализа различий между речью мужчин и женщин. Подробно обсуждаются результаты одного направления анализа — сравнительная длина высказываний (реплик) мужчин и женщин, измеренная в количестве словоформ. Обнаружена статистическая значимость различий по этому параметру между мужчинами и женщинами в публичной коммуникации — мужчины говорят больше женщин — при отсутствии статистически значимых различий в частной коммуникации. Также значимым оказывается пол адресата, причем корреляция снова носит существенно разный характер в подкорпусе публичной и частной речи. В частной коммуникации реплики говорящего оказываются несколько длиннее в разговоре с человеком того же пола; в публичной коммуникации реплики говорящего оказываются существенно длиннее при обращении к женщине, чем к мужчине, вне зависимости от пола говорящего. Эти выводы хорошо согласуются с результатами гендерных исследований в западной социолингвистике: различия между мужской и женской речью не абсолютны, но зависят от ситуации общения; публичный дискурс является ареной языкового соревнования, в котором преимущество остается за мужчинами, в то время как в частной коммуникации соревновательность исчезает.

Ключевые слова: гендер, корпусная лингвистика, социолингвистика, русский язык, Национальный корпус русского языка, речевое взаимодействие

RUSSIAN NATIONAL CORPUS AS A PLAYGROUND FOR SOCIOLINGUISTIC RESEARCH. EPISODE IV. GENDER AND THE LENGTH OF THE UTTERANCE

Daniel M. A. (misha.daniel@gmail.com),
NRU HSE / Moscow State University, Moscow, Russia

Zelenkov Yu. G. (yuryz@yandex.ru),
Yandex, Moscow, Russia

The paper discusses sociolinguistic implementations of statistical analysis of the spoken subcorpus of the Russian National Corpus. Given the considerable size of the corpus (about 10 mln tokens), an analysis of co-variation of various linguistic parameters with one of the few sociolinguistic parameters available — the speaker's gender — may give rich and interesting results. One specific example of co-variation is considered in detail: the mean length of the utterance (in tokens). Comparing this parameter in public communication shows statistically significant difference between the speech of men and women (men talk more), while the same difference is absent in private communication. Another important parameter is the gender of the addressee. Again, co-variation is quite different in public and private discourse. In private communication, the utterances are longer when addressing someone of the same sex, the difference between men and women is not statistically significant. In public communication, the utterances are longer when addressing a woman, whether the speaker herself is a man or woman. These conclusions are consistent with the results of sociolinguistic gender studies obtained elsewhere and by other methods. Linguistic difference between men and women are not absolute but depend on the communicative situation (public vs. private). Public discourse is a playground for linguistic competition in which men are the winning party. In private discourse, competition dissolves.

Keywords: gender, corpus linguistics, Russian, Russian National Corpus, interactional sociolinguistics

Как показывают исследования, мужчины считают, что женщины много говорят, не потому, что женщины говорят больше мужчин, а потому, что они вообще открывают рот
Социолнгвист, пожелавший остаться анонимным

Введение

В недавней статье, представленной в рамках конференции «Диалог», один из ведущих российских социолнгвистов В. И. Беликов обосновывает точку зрения, согласно которой НКРЯ не может служить надежным источником данных для социолнгвистического анализа (Беликов 2011). Основным аргументом для В.И. является нерепрезентативность НКРЯ. Мы могли бы добавить многие другие (как нам кажется, более бесспорные) недостатки (устного) Корпуса как источника социолнгвистических данных: например, его ориентированность на лексико-грамматические характеристики, что делает невозможным проведение исследований в одном из самых интересных и популярных на западе направлений социолнгвистики — в сфере социофонетики. Обращает на себя внимание, однако, следующее уточнение: на социолнгвиста НКРЯ «не рассчитан, поскольку ни в коей мере не может считаться сбалансированным с точки зрения решаемых им задач». Очевидно, что это утверждение может быть верно только в зависимости от того, как мы определим круг задач и интересов социолнгвиста. В том разделе цитируемой статьи, которая посвящена социолнгвистике, В. И. Беликов убедительно критикует некоторые неудачные примеры анализа социальной вариативности в сфере лексики. В настоящей статье, однако, мы хотим на примере другой социолнгвистической задачи показать, что по крайней мере Устный корпус, несмотря на существенные недочеты, является социолнгвистическим ресурсом с большим исследовательским потенциалом.

Следует сразу оговориться, что речь пойдет об оффлайновом исследовании текстов НКРЯ. Корпус выступает здесь не в качестве интерактивного инструмента (а именно примеры таких исследований рассматривает В. И. Беликов), но как коллекция аннотированных текстов; а в таком виде он недоступен простому пользователю. Тем не менее, главный тезис остается в силе: с нашей точки зрения, существуют такие социолнгвистические задачи, для которых коллекция текстов НКРЯ, вопреки точки зрения, высказанной в (Беликов 2011), является достаточно репрезентативной и позволяет получать интересные результаты.

По состоянию на момент написания статьи Устный корпус включает около 3,000 документов общим объемом почти 10 млн. словоупотреблений. Все документы Устного корпуса разделены на три группы — подкорпус

публичной речи (около 1,000 документов, около 5,1 млн. словоупотреблений), подкорпус устной непубличной речи (около 1,300 документов, более 1,1 млн. словоупотреблений) и подкорпус речи кино (около 690 документов, около 3,8 млн. словоупотреблений). Подробно мотивацию включения речи кино в Устный корпус (проблема спонтанности) см. в (Гришина, Савчук 2009). Важным свойством Устного корпуса является наличие в разметке простейшей информации о говорящих. Для пользователя портала НКРЯ это значит, что поиск можно ограничивать полом и возрастом говорящего. Другая информация об Устном корпусе для задач, обсуждаемых в настоящей статье, может считаться несущественной.

Включение в разметку базовых сведений о говорящем наталкивает на мысль о возможности офлайн-статистического исследования Устного корпуса НКРЯ с социолингвистической точки зрения: например, на предмет различий в речи мужчин и женщин. В настоящей статье по разным причинам представлен лишь один частный результат такого рода исследований — исследование корреляции длины высказывания с полом говорящего, результаты которого в целом соответствуют теоретическим ожиданиям. В заключительной части статьи мы кратко охарактеризуем другие возможные направления социолингвистического анализа Устного Корпуса НКРЯ.

Средняя длина высказывания как социолингвистическая переменная

Хорошо известно, что предметом социолингвистики являются не только собственно социально значимые языковые феномены, но и рефлексия носителей о взаимосвязи социальных и лингвистических категорий. Одним из самых известных стереотипов о женской речи является представление о том, что женщины много говорят. При помощи Устного корпуса этот стереотип легко верифицировать. Нами было посчитано распределение длин мужских и женских реплик. Репликой считалась последовательность высказываний одного говорящего до момента смены говорящего или конца документа. Длиной реплики считалось количество словоупотреблений в реплике. В следующих таблицах показано распределение длин мужских и женских реплик. Общий объем корпуса в репликах составляет 103 037 и 125 642 для мужчин и женщин, соответственно. На Графике 1 показана зависимость между длиной реплики (в словоупотреблениях, ось X) и долей реплик данной длины в корпусе (процентах от общего числа реплик, ось Y). Длинный штрих отражает эту зависимость для мужских реплик, короткий штрих — для женских, а тонкая непрерывная линия — по устному корпусу в целом.

Для всех трех кривых максимальную долю составляют однословные реплики, двух-, трех- и четырехсловные реплики имеют примерно равную долю, а начиная с пяти словоупотреблений доля реплик соответствующей длины

начинает стремительно падать. Однако видно, что кривые для мужских и женских реплик ведут себя различно, так как для длин реплик от одного до десяти словоупотреблений кривая для женских реплик проходит над той же кривой для мужских реплик. Между 11 и 13 словоупотреблениями кривые практически совпадают, а начиная с 14 кривая для женских реплик навсегда уходит «под» кривую для мужских реплик. Для длинных реплик кривые на Графике 1 располагаются визуально близко, однако вес отличия (который можно посчитать, например, как в Графике 2 — отношение разности долей реплик данной длины у женщин и мужчин к доле реплик этой длины во всем Устном корпусе) начиная с 14 словоупотреблений растет.

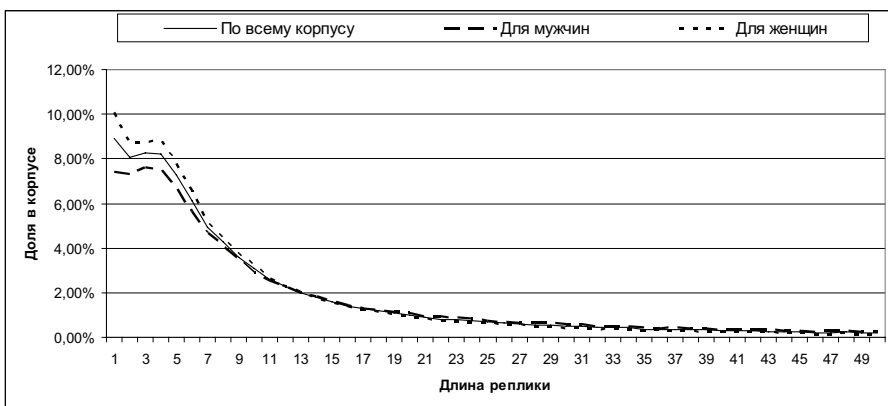


График 1. Зависимость длины реплики и доли в корпусе

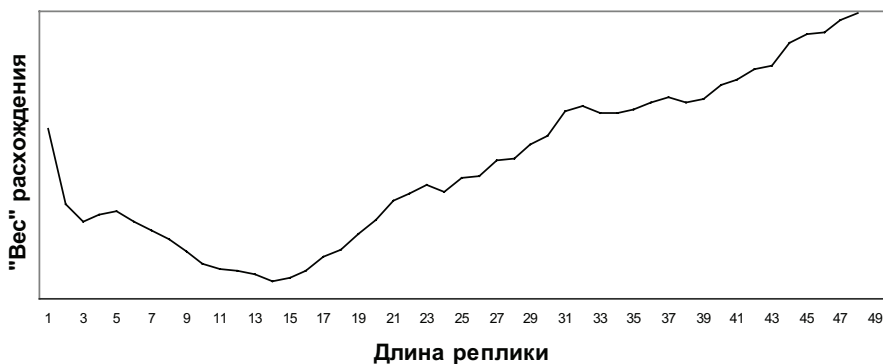


График 2. «Взвешенная» разница долей реплик в мужском и женском корпусе (зависимость от длины реплики)

Далее речь будет идти только о средних длинах (без учета распределений). По всему устному корпусу средняя длина реплики мужчины составляет

26,34 словоформы, а средняя длина реплики женщины — 15,41 словоформы. Результат вполне однозначно указывает на то, что обсуждаемый языковой стереотип неверен. Это соответствует результатам, полученным для других языков и на основании иных методов анализа (например, сравнительная длительность речи мужчин и женщин в рамках публичных семинаров, подробнее см. Вахтин, Головки 2004).

Более интересным и содержательным, но также вполне ожидаемым исходя из накопленных в зарубежной социолингвистике результатов, является следующее уточнение. При разделении исходных данных на публичную и частную речь оказывается, что расхождения в длине реплик у мужчин и женщин выглядят для этих двух подкорпусов существенно разным образом.

Таблица 1. Сравнение средней длины мужских и женских реплик по типам текстов (в словоформах)

| | мужчины | женщины |
|-----------------------|----------------|----------------|
| Устный корпус в целом | 26,34 | 15,41 |
| Публичная речь | 30,36 | 20,65 |
| Частная речь | [10,77] | [10,29] |

Из таблицы видно, что различие между публичной речью мужчин и женщин очень значительно (оно приближается к различию по Устному корпусу в целом, так как именно публичная речь составляет его наиболее значительную долю). При этом в частной речи различие в средней длине реплик оказывается статистически незначимым (здесь и далее статистическая значимость определялась по *t*-критерию Стьюдента на уровне значимости 0,001). Это может быть интерпретировано как подтверждение делаемых в западной социолингвистике обобщений, согласно которым женщины в большей степени ориентированы на частную, чем на публичную речь (Wodak, Benke 1998), а также прекрасно подтверждает один из основных тезисов этого же обзора и общий тренд развития исследований женской речи — различия между речью мужчин и женщин носят не абсолютный, а контекстуальный характер, так что изучение параметров варьирования вообще, в отрыве от коммуникативной ситуации, либо дает вообще неправильные результаты, либо значительно их обедняет.

Иными словам, если в ситуации публичной коммуникации женщины говорят меньше, чем мужчины (или их собеседники не дают им говорить столько же, сколько мужчинам — ср. анализ перебиваний в Tannen 1990, также обсуждаемый в Вахтин, Головки 2004), то в ситуации частного общения они, как кажется, коммуницируют наравне с мужчинами. В этой связи интерес представляет еще один тип калькуляций — относительная длина мужских и женских реплик в зависимости от пола адресата. Ср. следующую таблицу:

Таблица 2. Сравнение средней длины мужских и женских реплик по типам текстов и в зависимости от пола собеседника

| Публичная речь | | | Частная речь | | |
|----------------|--------------|-------|--------------|--------------|-------|
| | пол адресата | | | пол адресата | |
| | муж | жен | | муж | жен |
| муж | 35.75 | 46.58 | муж | 12.94 | 10.84 |
| жен | 27.28 | 39.30 | жен | 9.49 | 12.23 |

Из таблицы видна разнонаправленная корреляция между длиной реплики и соотношением полов говорящего и адресата. В частной речи реплики женщин оказываются статистически значимо (хотя и незначительно) длиннее, если они обращены к женщинам. Разница в длине реплик мужчин в зависимости от пола собеседника оказывается еще менее заметной (хотя, опять же, статистически значимой), но демонстрирует ту же корреляцию. Отметим также, что разница между средней длиной реплики женщины, обращенной к женщине, и средней длиной реплики мужчины, обращенной к мужчине, оказывается статистически незначимой, как и разница между средней длиной реплики мужчины, обращенной к женщине, и средней длиной реплики женщины, обращенной к мужчине. Иными словами, в ситуации частной коммуникации говорящий строит более длинные реплики при общении с человеком того же пола; другие различия нивелированы.

Статистическая картина в публичном подкорпусе принципиально иная. Во-первых, зависимость длины реплик от пола собеседника носит гораздо более ярко выраженный характер. Во-вторых, она устроена совершенно иным образом — если средняя длина реплики женщины, обращенной к женщине, превосходит (и значительно) среднюю длину реплики женщины, обращенной к мужчине, то средняя длина реплики мужчины, обращенной к мужчине, не больше, а меньше, причем значительно меньше — средней длины реплики мужчины, обращенной к женщине.

Таким образом, если для зависимости средней длины реплики от пола адресата в частной речи операциональной является категория гендерной гомогенности / гетерогенности (пол адресата отличается от / совпадает с полом говорящего), то в публичной речи определяющую роль играет категория пола адресата (мужской / женский). Или, выражаясь несколько менее формально: в частном разговоре женщина больше говорит с женщиной, а мужчина — с мужчиной, в то время как в публичной речи женский пол адресата значительно увеличивает продолжительность реплики не только у мужчины, но и у женщины.

Известный тезис о соревновательном характере публичного речевого поведения, в котором мужчины преуспевают заметно более женщин (Tannen 1990), согласно полученным нами данным, оказывается применим только к публичному речевому поведению (см. обсуждение гендерных теорий типа *power/dominance theories* у Wodak, Benke 1998), причем женщины пользуются «гендерной слабостью» адресата так же активно (хотя и менее успешно), как мужчины.

Следует сделать следующую важную оговорку. Как ясно из обсуждения данных, в нашем исследовании контролируруемыми переменными являлись только пол говорящего (собственно экстралингвистическая переменная) и общий тип текста согласно классификации НКРЯ (частный или публичный дискурс; о социолингвистическом статусе таких параметров, задающих функционально-коммуникативный контекст, см. все в том же обзоре Wodak, Benke 1998).

Контроль других классических экстралингвистических переменных (возраст, образование, социальный класс) технически невозможен — эти данные просто отсутствуют в разметке НКРЯ. Впрочем, даже если их можно было бы включить, по нашему мнению, они не оказали бы существенного влияния на полученные результаты. Устный Корпус НКРЯ не является репрезентативным ни (наверняка) с точки зрения возраста, ни (с большой степенью вероятности) социального класса говорящих. Судя по жанровому составу Корпуса, можно было бы высказать предположение, что тексты «центрированы» на испытуемых среднего достатка в возрасте от 20 до 40 лет и слабо отражают более маргинальные (с точки зрения типовых социолингвистических исследований) группы (дети, пожилые люди, люмпены и пр.).

С другой стороны, более внимательный контроль типа текста не только возможен, но и абсолютно необходим для продолжения этого исследования, которое можно таким образом признать лишь пилотным. Во-первых, некоторые жанры являются принципиально неженскими (например, парламентские слушания). Впрочем, доля таких текстов в корпусе пренебрежимо мала (менее двух процентов подкорпуса публичной речи).

Во-вторых, что гораздо более важно, многие устные жанры ведут себя с точки зрения длины реплики совершенно особым образом. Так, в Устный Корпус НКРЯ включены встречи с читателями и слушателями, доклады, интервью, спортивные комментарии, лекции, пресс-конференции, рассказы, репортажи (публичная речь), а также рассказы и пересказы (частная коммуникация). Все эти жанры устроены с точки зрения длины реплик иначе, чем, например, ток-шоу или дискуссии: на одного из говорящих (будем условно называть его «нарратором») приходится большая часть дискурса, в то время как собеседники, если они вообще присутствуют, лишь задают вопросы и вставляют небольшие комментарии. Мы предполагаем, что внутри публичной и частной коммуникации такие «всплески» длительности нивелируются тем, что нарратор в большей части подобных жанров может быть как мужчиной, так и женщиной. Однако сравнение длины реплики в публичном корпусе и в частной коммуникации, на котором мы сделали акцент выше, может от этого сильно страдать (особенно если учесть, что такие особые типы публичных текстов составляют около четверти всего подкорпуса публичной устной речи).

В целом очевидно, что в дальнейшем следует более подробно классифицировать коммуникативные роли говорящих, выделяя по крайней мере роль нарратора. Проще всего просто исключить такого рода данные из анализа. Корреляция между длиной реплики нарратора и его полом составляет предмет для дополнительного исследования, но приходится признать, что «сливать»

этот показатель с длиной реплики в обычной публичной дискуссии, как это сделано в представляемой пилотной версии исследования, в любом случае неправильно — так же неправильно, как сливать длину реплики в частной и публичной коммуникации.

Заключение

Выше было показано, что, несмотря на поверхностность социолингвистической разметки текстов Устного корпуса, оффлайнный статистический анализ текстов позволяет сформулировать нетривиальные интерпретации и обобщения, согласующиеся с результатами, которые были получены иными методами и для других языковых сообществ — например, показать контекстность различий речевого поведения мужчин и женщин, то есть, в частности, перейти от *sex-based* парадигмы к парадигме, опирающейся на гендер как интерактивную социальную категорию. Конечно, это лишь очень частный результат анализа данных, которые, с нашей точки зрения, являются богатейшим полигоном социолингвистических исследований (причем даже этот частный результат требует дополнительной верификации с учетом более детальной жанровой классификации текстов). Попробуем наметить возможные дальнейшие социолингвистические применения устных текстов НКРЯ.

С точки зрения исследования гендерных различий большой интерес представляют статистически значимые различия между частотностью (частотным рангом) конкретных языковых единиц — лексем, грамматических категорий или их сочетаний, а также, возможно, конкретных словоформ. Эти данные относительно легко получить путем статистического анализа Устного корпуса методами, аналогичными методам автоматической рубрикации документов — причем, что важно, без предварительных гипотез о том, что именно мы ищем (ср. Lakoff 1975), так как ранговые списки таких единиц легко составляются автоматически.

Здесь мы позволим себе сделать краткое отступление от основной темы. Параллельно с данным исследованием одним из авторов статьи разрабатывался метод автоматического определения пола автора высказывания по тексту этого высказывания (в графическом представлении). Эта задача в целом аналогична задаче построения автоматического рубрикатора текстов: были выделены «гендерные маркеры», т. е. языковые признаки принадлежности текста к одному из двух классов (мужские vs. женские тексты). Во многих отношениях эта задача отличается от академической задачи поиска расхождений между женским и мужским «вариантами» языка. Например, социолингвисту могут быть интересны статистически значимые расхождения относительной частотности употребления конкретных языковых единиц (скажем, конкретных лексем), а для автоматического определения пола говорящего, который ориентирован в первую очередь на небольшие тексты и отдельные высказывания, сам факт вхождения той или иной лексемы вряд ли окажется достаточным. Мы упоминаем это исследование лишь потому, что алгоритм

опирается на те же данные, что представленный выше анализ длин реплик, и работает с достаточной степенью надежности и уже внедрен в некоторых приложениях поисковой системы Яндекс. Для определения пола говорящих в Устном корпусе НКРЯ он работает в общем с вероятностью более 0,9 (исключением оказывается речь женщин в публичном подкорпусе — факт, который также вполне можно интерпретировать с точки зрения того, что нам известно о различиях в языковом поведении мужчин и женщин; ср. обзор в Wodak, Benke 1998, по-русски также Вахтин, Головкин 2004). Эти результаты указывают на статистическую надежность гендерных интерпретаций данных Устного корпуса и подтверждают целесообразность исследования на том же материале ранговых различий между речью мужчин и женщин. К сожалению, менее лексически ориентированные результаты — например, проверка на материале русского языка тезисов Лакофф об относительной частотности тех или иных синтаксических конструкций — либо вообще не доступны для прямого статистического анализа, либо требуют более тонких косвенных методов оценки.

Важным направлением социолингвистических исследований в области гендерных различий является верификация и интерпретация известного еще с первых работ Лабова тезиса о том, что женская речь по сравнению с мужской в большей степени ориентирована на норму (ср. подробный обзор по этой важной теме в Wodak, Benke 1998). В нашем случае ключевым является вопрос о том, что принять за статистику нормы — с чем сравнивать статистические характеристики устной речи мужчин и женщин. Можно условно считать нормативной статистику корпуса современных газет (сейчас — более 100 млн. словоупотреблений). Несмотря на вполне очевидную условность такого решения, полученные нами предварительные результаты отчасти согласуются с предположениями гендерной теории. Так, количество выделяемых алгоритмом языковых маркеров, отличающие публичную речь женщин от «нормативного» газетного корпуса статистически значительно меньше, чем количество маркеров, отличающих публичную речь мужчин. С другой стороны, аналогичный анализ для частной коммуникации дает обратные результаты, что допускает множество интерпретаций, но в первую очередь требует дополнительной проверки валидности статистического анализа.

Подробное обсуждение этих и других вопросов расширения применения НКРЯ для социолингвистических исследований здесь невозможно в том числе по соображениям объема, поэтому мы оставляем его для отдельной публикации. Единственный момент, который хочется кратко затронуть — это кросслингвистическая применимость предлагаемых методов анализа и сопоставимость их результатов по разным языкам. Единственным условием, необходимым для проведения такого рода исследования на материале других языков, является доступность устных корпусов достаточного объема, обладающих по крайней мере той же разметкой параметров говорящего, что и Устный корпус НКРЯ — указанием пола. Конечно, устные корпуса большого объема — относительная редкость, однако для некоторых языков они уже

существуют. В число таких языков входят не только крупные европейские языки, но и некоторые языки других ареалов — например, Восточноармянский национальный корпус (www.eanc.net). Параллельное применение одного и того же метода статистического анализа к материалу устных корпусов нескольких языков разных структур и ареалов (как лингвистических, так и социокультурных) позволит попытаться ответить на один из самых острых вопросов современной вариационной социолингвистики — в какой степени результаты, полученные ею до сих пор и в подавляющем большинстве случаев опирающиеся на языковые ситуации Западной Европы и других развитых стран (например, ориентация женской речи на норму), носят действительно универсальный характер.

Литература

1. *Lakoff, Robin Tolmach* (1975/2004) *Language and woman's place*. Oxford: Oxford University Press.
2. *Tannen, Deborah* (1990) *You just don't understand. Women and men in conversation*. NY: Ballantine Books.
3. *Wodak, Ruth and Gertraud Benke* (1998) *Gender as a sociolinguistic variable: new perspectives on variation studies*. In Florian Coulmas (ed.) *The handbook of sociolinguistics*. Oxford: Blackwell. Pp. 127–150.
4. *Беликов, В. И.* (2011) Чего не хватает в «оцифрованном мире» лексикографу и социолингвисту. // Труды международной конференции Диалог 2011. Бекасово.
5. *Вахтин, Н. Б. и Е. В. Головки* (2004) *Социолингвистика и социология языка*. СПб.: Гуманитарная академия.
6. *Гришина, Е. А. и С. О. Савчук* (2009) Корпус устных текстов в НКРЯ: состав и структура // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История. Стр. 129–149.

References

1. *Belikov V. I.* (2011) *Chego ne hvataet v "otsifrovannom mire" leksikografu i sotsiolingvistu*. [What a lexicographer and a sociolinguist would want in a digital world] *Trudy Mezhdunarodnoj Konferentsii "Dialog 2011"* [Proceedings of the International Conference "Dialog 2011"]. Bekasovo.
2. *Grishina E. A., Savchuk S. O.* (2009) *Korpus ustnyh tekstov v NKRJa: sostav i struktura* [The Oral Subcorpus of the Russian National Corpus: its structure and composition], in: *Natsionalnyj korpus russkogo jazyka: 2006–2008. Novye rezultaty i perspektivy* [Russian National Corpus: 2006–2008. Current state and future development]. Saint Petersburg, Nestor-Istorija. Pp. 129–149.
3. *Lakoff, Robin Tolmach* (1975/2004) *Language and woman's place*. Oxford: Oxford University Press.

4. *Tannen, Deborah* (1990) *You just don't understand. Women and men in conversation*. NY: Ballantine Books.
5. *Vahin N. B., Golovko E. V.* (2004) *Sotsiolingvistika I sotsiologija jazyka*. [Sociolinguistics and sociology of language]. Saint Petersburg: Gumanitranaja akademija.
6. *Wodak, Ruth and Gertraud Benke* (1998) *Gender as a sociolinguistic variable: new perspectives on variation studies*. In: Florian Coulmas (ed.), *The handbook of sociolinguistics*. Oxford: Blackwell. Pp. 127–150.