

УДК 81'25

ПАРАМЕТРЫ ЧАСТОТНОГО СЛОВАРЯ В АСИММЕТРИЧНЫХ И СИММЕТРИЧНЫХ СОЦИАЛЬНЫХ РОЛЯХ ГОВОРЯЩЕГО

Попова Татьяна Ивановна

магистрант кафедры русского языка

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 11. tipopova13@gmail.com

В ходе повседневного общения каждый говорящий выбирает социальную роль из определенного, присущего лично ему, ролевого набора (мать, жена, друг, коллега и т. п.) и в соответствии с этой ролью выстраивает свое поведение, в том числе речевое, в конкретном коммуникативном акте. Применение статистических параметров для анализа частотного словаря информантов корпуса русского языка повседневного общения «Один речевой день» позволило обнаружить показатели, которые маркируют роль в асимметричных отношениях. Можно предположить, что случаи формального общения в рамках повседневной коммуникации могут быть более стандартизированы.

Ключевые слова: русская разговорная речь; корпус; социолингвистика; социальная роль; коммуникативная ситуация; лингвистические корреляты; статистические параметры; частотные списки.

В процессе коммуникации речь варьирует, что зависит от многих факторов, и нередко приобретает не только индивидуальные черты – особенности конкретного идиолекта, но и, как показывают многочисленные исследования, черты, общие для определенных групп (гендерных, возрастных, профессиональных и т. п.) – особенности того или иного социолекта.

На современном этапе развития лингвистики исследования социальных характеристик речи стало возможным (и необходимым) проводить на больших массивах данных, организованных в корпус; важным при этом является наличие обширной дискурсивной разметки в корпусе. Материалом для настоящего исследования стал корпус повседневной русской речи «Один речевой день» (ОРД), наиболее представительный на сегодня ресурс для анализа устного дискурса, который разрабатывается в Санкт-Петербургском государственном университете [Звуковой корпус как материал для анализа русской речи 2013; Русский язык повседневного общения 2016; Asinovskiyy et al. 2009].

Звукозаписи повседневной речевой коммуникации, собранные в корпусе ОРД, выполнены по методике непрерывного многочасового мониторинга. На сегодняшний день общий объем корпуса составляет более 1 млн. словоупотреблений (1250 часов записи от 128 информантов и около 1000 их коммуникантов).

Среди многочисленных параметров метатекстового аннотирования в ОРД выделяется социальная роль говорящего – категория, которая по-

ка является одной из малоизученных. На рисунке представлен фрагмент классификации социальных ролей в ОРД и используемые для их обозначения в базе сокращения (далее в работе все социальные роли обозначаются введенными здесь сокращениями).

Интересным представляется изучение регистров общения как ситуационно обусловленных языковых вариаций в рамках многомерного подхода Д. Байбера [Biber 1993]. Различия в регистрах определяются в неязыковых терминах: различие целей, интерактивность, производственные обстоятельства, отношения между участниками и т. д. (здесь и далее перевод мой. – Т.П.) [там же: 7].

В процессе общения происходит постоянное, обусловленное изменениями в коммуникативной ситуации речевое переключение, в рамках которого, как представляется, «успех общения в значительной мере зависит от того, в достаточной ли степени владеют говорящий и слушающий формами языка, пригодными для данной ситуации» [Крысин 2004: 430]. Так, владение несколькими кодами проявляется и на уровне социальных ролей говорящего, и в его речи можно обнаружить специфические черты, характерные для определенных социальных ролей. Реализация социальной роли будет варьировать вместе с ситуацией, но с условием нахождения внутри рамок, допустимых для той или иной роли: социальная роль порождает «ожидание» определенной манеры поведения человека, в том числе и речевого.

<p>Родственники:</p> <ul style="list-style-type: none"> МУ – муж, ЖЕ – жена ОТ – Отец и пр. <p>Прочие родственники (без указания пола):</p> <ul style="list-style-type: none"> СР – старшие родственники теть, свекровь, теща, дядя и пр. МР – младшие родственники – зять, невестка, и пр. <p>Друзья</p> <ul style="list-style-type: none"> БД – близкий друг (бой-френд) БП – близкая подруга (гёрл-френд) ДР/ПО АД – ассоциируемый друг (друг друзей, родственников) 	<p>Коллеги</p> <ul style="list-style-type: none"> КО – коллега КН – коллега-начальник, командир КД – коллега-друг (подруга) <p>«Парные» отношения</p> <ul style="list-style-type: none"> КЛ – клиент (в том числе – покупатель, в транспорте с водителем-кондуктором, на вахте, арендатор квартиры, пациент, родители пациента) СС – сервис, сервисная служба – все кто оказывают услуги за деньги (продавец, фотограф, квартирная хозяйка, врач) ССВ – врач УЧ – ученик ПР – преподаватель, учитель и пр. 	<p>Группы людей</p> <ul style="list-style-type: none"> СТ – студент(курсант) ПП – «попутчик», а также посетитель того же кафе, магазина, поликлиники и пр. общественных мест <p>Прочее</p> <ul style="list-style-type: none"> ЗН – просто знакомый, не друг, не клиент, не обслуга, не «партнер» ЧУ и пр.
---	--	---

Рис. Классификация социальных ролей в корпусе ОРД

«Ситуативная вариативность проявляется в использовании тех или иных языковых средств – отдельных единиц или целых систем/подсистем в зависимости от социальной ситуации» [Швейцер 2009: 79], которые называют социолингвистическими (ситуативными) переменными. В настоящем исследовании в этом значении используется термин «лингвистические корреляты». Набор выявленных коррелятов не должен носить системного характера: это лишь попытка выбрать из множества лингвистических черт некоторые классифицирующие особенности, которые способны отразить специфичность исследуемого явления.

Исследователи отмечают, что «для социолингвистики релеванты не любые роли, а лишь те из них, которые имеют эксплицитно выраженные языковые или речевые корреляты» [там же: 82]. Какие же роли можно считать «значимыми»? Данные корпуса позволяют выявить ролевой набор участника коммуникации: в ОРД средний показатель количества ролей говорящего оказывается равен 5 [Попова 2018а]. На первых трех позициях по количественным показателям расположились роли, которые можно соединить по принципу парности отношений (см. табл. 1).

Таблица 1
Парное соотношение социальных ролей

СР	Кол-во эпизодов, абс.	%
КО ²	842	26,0
СС – КЛ	389	12,3
МУ – ЖЕ	272	8,5

Видим, что на первые позиции поднимаются роли, которые согласно теоретическим положениям социолингвистики противопоставлены по принципу ассиметричности/симметричности отношений [Беликов, Крысин 2001: 201–202].

Для дальнейшего анализа были отобраны «речевые дни» двух информантов: И38 (муж., 53 года) и И115 (муж., 58 лет), ролевой набор которых равен 8, при этом в речевом материале каждого из информантов есть эпизоды, в которых они выступают как в роли МУ, так и в роли СС. Принцип противопоставленности двух ролей должен способствовать выявлению значимых языковых явлений, т. е. искомым лингвистических коррелятов социальных ролей (данные о материале исследования см. в табл. 2).

Таблица 2
Характеристика материала исследования

Информанты	Объем подкорпуса (токены)	МУ	СС
И38	2 868	1 533	1 335
И115	6 397	2 878	3 519

Поиск лингвистических коррелятов должен, по-видимому, проводиться на всех языковых уровнях. Но какие показатели станут классифицирующими в случае смены ролей, или, иными словами, переключения регистров? Этим вопросом уже задавался американский лингвист Д. Байбер, который, занимаясь изучением жанрового своеобразия, эмпирическим путем отобрал 67 параметров для анализа лингвистических маркеров регистров [Viber 1993]. По мнению Д. Байбера, «лексический выбор обычно не маркирует регистр (так, термин *home run* может встретиться и в трансляции матча, и в новостной статье, и в романтической повести, и в других регистрах)», но существуют «ключевые (core) лексические и грамматические черты как индикаторы регистра <...>, поскольку многие регистры отличаются друг от друга только относительной дистрибуцией ключевых языковых черт» [Байбер 2012].

Значимые, ключевые лексические единицы могут быть обнаружены при составлении частотных списков слов разных социальных ролей и рассмотрении их параметров [Попова 2018б]. Так, анализ верхней зоны частотного списка информанта И38 позволил обнаружить совпадения с подобным списком, составленным на материале аннотированного подкорпуса ОРД для речи мужчин: единицы *ну, да, вот, а, не, это* в обоих списках располагаются в первой десятке (см. табл. 3).

Эти единицы выполняют в устной речи прагматические функции, организуют дискурс. Их позиции (ранги) в разных частотных списках подтверждают их распространенность и универсальность в речи мужчин, значимость в устной коммуникации вне зависимости от социальных ролей.

Предметом внимания настоящего исследования стали параметры частотного словаря речи информантов. Такого рода данные позволяют качественно оценить содержание материала (см. об этом подробнее [Алексеев 2001]). Характеристики словаря информантов в разных социальных ролях представлены в табл. 4.

Таблица 3
Наиболее употребительные слова мужской речи
(по данным ОРД)³

Ранг	Слово	Абс. кол-во	%
1	<i>ну</i>	2062	2,47
2	<i>я</i>	2000	2,40
3	<i>не</i>	1898	2,28
4	<i>вот</i>	1887	2,26
5	<i>в</i>	1669	2,00
6	<i>да</i>	1610	1,93
7	<i>а</i>	1548	1,86
8	<i>и</i>	1530	1,84
9	<i>там</i>	1456	1,75
10	<i>это</i>	1338	1,60
11	<i>что</i>	1335	1,60
12	<i>на</i>	1113	1,33

Таблица 4
Параметры частотного словаря информантов

Параметры	Информанты			
	И38		И115	
	МУ	СС	МУ	СС
Леммы	451	340	606	641
Нарах	375	281	528	530
Кол-во полных слов	389	295	550	577
TTR/TTR (st)	0,3/0,4	0,3/0,3	0,3/0,4	0,2/0,4
Лексическая плотность	25%	21%	20%	14%
Индекс исключительности	0,7	0,8	0,7	0,65

Так, показатель количества уникальных слов (**нарах**), тех, что были произнесены только один раз, характеризует не только лексическое богатство речи информанта, но позволяет также обнаружить специфичные черты языковой личности (особенности идиолекта). Выявление зависимости **нарах** от социальной роли говорящего может рассматриваться как потенциальный лингвистический коррелят данной роли. Этот параметр может быть использован, например, в практике судебной лингвистики, особенно при установлении авторства, атрибуции текста или в случаях предполагаемого плагиата. Поскольку подкорпус речи И115 значительно больше, чем подкорпус речи И38, то сравнивать показатели **нарах** в разных социальных ролях нужно для каждого информанта отдельно. Заметно преобладание уникальных слов у И38 в роли МУ (см. табл. 4).

Лексическая плотность – это доля полных слов в корпусе; **индекс исключительности** – доля однократных употреблений в корпусе. Заметно более низкие, чем у И38, значения лексической плотности и индекса исключительности у И115 в роли СС также свидетельствуют, что эти показатели могут претендовать на статус лингвистических коррелятов социальных ролей.

Однако важно учитывать **индекс разнообразия** TTR – token/type relation (ratio) – отношение количества разных словоформ или лексем к объему текста, которое выводится по формуле L/N , где L – лексем, N – объем текста. При стандартизованном TTR(st) используются выборки одинакового размера. Как видно из табл. 4, показатели равны для обоих информантов в роли МУ, а вот роль СС у И38 и здесь оказалась хоть незначительно, но все же маркирована, что может свидетельствовать о необходимости детального рассмотрения этого аспекта.

Дж. Юр показал, что письменные тексты имеют лексическую плотность более 40%, в то время как плотность устных текстов, как правило, меньше. Разговорные тексты предсказуемы, потому что они обычно тесно связаны с «физической средой» человека и поэтому содержат больше повторов [Ure 1971]. Эту мысль можно развивать на полученных в настоящем исследовании показателях: **лексическая плотность** обоих информантов в обеих ролях сигнализирует о «разговорности», даже в случаях асимметричных отношений.

Поиск лингвистических коррелятов социальных ролей с помощью применения статистических параметров частотного словаря информантов позволил обнаружить показатели, которые маркируют роль СС у обоих информантов. Можно предположить, что случаи формального об-

щения в рамках повседневной коммуникации могут быть более стандартизированы.

Дальнейший анализ материала на разных языковых уровнях сделает возможным получение классификации характерных особенностей, которые отличают речь в рамках одного языкового регистра от речи в рамках другого. Полученные данные могут быть использованы для решения текущих проблем в речевых технологиях, для разработки программ синтеза речи и систем распознавания, а также в судебной фонетике.

Примечания

¹ В корпусной лингвистике принято называть единицы корпуса – минимальные фрагменты анализа – **токенами**, хотя чаще всего они совпадают со словоформами [Baker, Hardie, McEney 2006: 159].

² Социальная роль «коллега» характеризуется ярко выраженной неоднородностью частотного распределения лексических единиц. Следовательно, в структуре «речевого дня» категория «работа/учеба» должна рассматриваться независимо от всех прочих видов деятельности [Русский язык повседневного общения 2016].

³ Таблица взята из монографии «Русский язык повседневного общения: особенности функционирования в разных социальных группах» [2016: 112].

Список литературы

Алексеев П.М. Частотные словари. СПб.: Изд-во СПбГУ, 2001. 153 с.

Байбер Д. Аналитическая рамка для изучения языковых регистров // Социолингвистика и социология языка: хрестоматия / отв. ред. Н.Б. Вахтин. СПб.: Изд-во Европ. ун-та в Санкт-Петербурге, 2012. Т. 1. С. 257–91.

Беликов В.И., Крысин Л.П. Социолингвистика / Рос. гос. гуманит. ун-т. М., 2001. 439 с.

Звуковой корпус как материал для анализа русской речи. Ч. 1: Чтение. Пересказ. Описание / коллект. авт.; отв. ред. Н.В. Богданова-Бегларян. СПб.: Филологический факультет СПбГУ, 2013. 532 с.

Попова Т.И. Говорящий вчера, сегодня, завтра, или ролевой набор участника коммуникации // Матер. Междунар. конф. русистов в Барселонском ун-те – МКР-Барселона 2018. Барселона, 2018а. (В печати.)

Попова Т.И. Частотные списки как инструмент поиска лингвистических коррелятов социальных ролей говорящего // Вестник Бурятского государственного университета. Филология. 2018б. № 2(2). С. 33–41.

Русский язык повседневного общения: особенности функционирования в разных социальных группах / отв. ред. Н.В. Богданова-Бегларян. СПб.: ЛАЙКА, 2016. 244 с.

Швейцер А.Д. Современная социолингвистика: Теория, проблемы, методы. М.: Книжный дом «ЛИБРОКОМ», 2009. 176 с.

Asinovsky A. et al. The ORD Speech Corpus of Russian Everyday Communication “One Speaker’s Day”: Creation Principles and Annotation / Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. // The 12th International Conference on Text, Speech and Dialogue – TSD 2009 / V. Matoušek, P. Mautner (eds.). Vol. 5729. Berlin; Heidelberg: Springer, 2009. Pp. 250–257.

Baker P., Hardie A., McEney T. A Glossary of Corpus Linguistics. Edinburgh: Edinburgh University Press, 2006. P. 123.

Biber D. The Multi-Dimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings // Computers and the Humanities. 1993. Vol. 26. Pp. 331–345.

Ure J. Lexical Density and Register Differentiation // Applications of Linguistics / G. Perren, J. L. M. Trim (eds.). London: Cambridge University Press, 1971. Pp. 443–452.

PARAMETERS OF FREQUENCY LIST IN ASYMMETRIC AND SYMMETRIC SPEAKER’S SOCIAL ROLES

Tatyana I. Popova

**Master student, Russian Language Department
Saint-Petersburg State University**

In the course of daily communication each speaker chooses a social role from a special set of roles (mother, wife, friend, colleague, etc.), and aligns his/her behavior, including speech behavior, in accordance with this role in a particular communicative act. Statistical analysis of the frequency list of informants of the everyday speech Russian corpus “One Day of Speech” shows certain parameters which tag the role in asymmetric relations. It can be assumed that instances of formal communication within everyday communication can be more standardized.

Keywords: colloquial Russian speech; corpus; sociolinguistics; social role; communicative situation; linguistic correlates; statistical parameters, frequency lists.