

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2018»

Москва, 30 мая — 2 июня 2018 г.

СТРУКТУРА ПОВСЕДНЕВНОГО ДИАЛОГА КАК ПОСЛЕДОВАТЕЛЬНОСТЬ РЕЧЕВЫХ АКТОВ¹

Шерстинова Т. Ю. (sherstinova@spbu.ru)

Филологический факультет СПбГУ;
Национальный исследовательский университет
«Высшая школа экономики»; Санкт-Петербург, Россия

Исследование структуры повседневного диалога проведено на материале 73 микродиалогов повседневной речевой коммуникации из корпуса устной русской речи «Один речевой день» (ОРД корпус). Задачей исследования было выяснение того, какие типы речевых актов чаще всего инициируют и завершают диалог, а также выявление наиболее типичных последовательностей речевых актов в структуре диалога. Была проанализирована речь 30 человек (6 информантов и 24 коммуникантов) в объеме 2230 речевых актов, относящихся как к профессиональным, так и бытовым разговорам. Для подсчета наиболее частотных последовательностей речевых актов использовалась техника n-граммного анализа. Полученные результаты показали, что инициируют диалог чаще всего репрезентативы, т.е. речевые акты, связанные с обменом информацией (38% случаев), «этикетное» начало (приветствия, вокативы) имеет место в 23% диалогов, а в 19% случаев разговор начинается с регулятивной формы. Речевые акты, завершающие диалог, показывают большее разнообразие: это репрезентативы (16% случаев), оценочные суждения (валюативы) (14%), регулятивные формы (14%), по 8% — директивы, комиссивы и этикетные формы и 7% — экспрессивы. Наиболее типичными бинарными последовательностями речевых актов оказались: два репрезентатива подряд (22,35%), регулятивная форма и следующий за ней репрезентатив (6,93%), репрезентатив и регулятивная форма (6,0%), валюатив и следующий за ним репрезентатив (5,21%), репрезентатив и оценочное суждение (4,77%), а также двусторонняя комбинация директива с репрезентативом (по 2,77%).

Ключевые слова: русская разговорная речь, повседневная речевая коммуникация, структура диалога, речевой акт, n-граммы, речевой корпус, устный дискурс

¹ Раздел 4 данной работы подготовлен при поддержке гранта РФ «Система прагматических маркеров русской повседневной речи» (проект № 18-18-00242).

THE STRUCTURE OF EVERYDAY DIALOGUE AS THE SEQUENCE OF SPEECH ACTS

Sherstinova T. Yu. (t.sherstinova@spbu.ru)

Philological Faculty of Saint-Petersburg State University;
National Research University Higher School of Economics;
Saint-Petersburg, Russia

The structure of Russian everyday dialogue was studied on the basis of 73 microdialogues of everyday speech communication from the 'One Day of Speech' corpus (the ORD Corpus). The aim of the research was to find out what types of speech acts commonly initiate and complete everyday dialogues, as well as to reveal the most typical sequences of speech acts in these dialogues. Altogether, 2230 speech acts of 30 people referring to both professional, and household conversations have been analysed. N-gram analysis has been used to calculate the most frequent sequences of speech acts. The obtained results showed that dialogues are usually started by representatives, i. e. speech acts related to the exchange of information (38% of all cases), etiquette beginnings (greetings, vocatives) take place in 23% of the dialogues, and in 19% of cases the conversation begins with a regulative form. Speech acts ending dialogues show a greater variety: representatives contribute 2% of all dialogue ends, evaluative judgments and regulatory forms cover 14% each, further go directives (8%), commissions (8%), etiquette forms (8%) and emotional and expressive form (7%). As for the most typical bigrams of speech acts, they are the following: two consecutive representatives (22.35%), a regulatory form followed by a representative (6.93%), a representative and a regulatory form (6%), a evaluative with a following representative (5.21%), a representative and a evaluative judgment (4.77%), as well as two combinations of a directive with a representative (2.77% each). Besides, the article presents data on the occurrence of the most frequent pairs of speech acts at the subtype level. Here, the most frequent one is the sequence 'question'+ 'answer', which covers 2.45%.

Key words: spoken Russian, everyday speech communication, dialogue structure, speech act, n-gram, speech corpus, oral discourse

1. Введение

Задача данного исследования состоит в выявлении наиболее типичных последовательностей речевых актов (РА) в структуре повседневного диалога. Материалом исследования являются 73 микродиалога повседневного общения из речевого корпуса «Один речевой день» (ОРД). Особенностью звукозаписей корпуса ОРД является их «аутентичность» — они выполняются в естественных условиях коммуникации с применением методики долговременной записи речи информанта в течение всего дня, «от восхода до заката» с помощью диктофона [Asinovsky et al. 2009]; [Bogdanova-Beglarian et al. 2016]. Такая методика сбора речевого материала традиционно используется в японских

лингвистических исследованиях (см., например, [Сибата 1983]; [Campbell 2004]), она также применялась при подготовке материалов для демографического подкорпуса Британского национального корпуса [Burnard 2007].

В настоящее время корпус ОРД содержит более 1250 часов звукозаписей речи, полученной от 130 информантов, мужчин и женщин, представителей разных профессий, разного социального статуса, в возрасте от 17 до 83 лет, и более тысячи их коммуникантов. Богатая коллекция звукозаписей корпуса дает возможность проводить на его материале лингвистические исследования повседневной речи в разных условиях коммуникации для разных языковых уровней (см. например, [Богданова-Бегларян и др. 2017]; [Bogdanova-Beglarian et al. 2016]; [Шерстинова 2016 и др.]).

Кроме того, коллекция ОРД дает уникальную возможность исследовать прагматику и структуру повседневных диалогов. Для прагматического анализа речевой коммуникации в корпусе ОРД была разработана многоуровневая система аннотирования коммуникативных эпизодов на макро- и микроуровнях². Так, вся речевая продукция информанта в течение периода записи членится на *макроэпизоды*, объединенные локусом коммуникации, ее основной задачей и участниками [Sherstinova 2015]. Единицей хранения данных в корпусе являются макроэпизоды речевой коммуникации, записанные в виде отдельных звуковых файлов. Макроэпизоды далее сегментируются на единицы меньшего уровня, именуемые *микроэпизодами*, объединенные по теме разговора и частной коммуникативной задаче [Шерстинова 2015]. Исследование структуры диалога в настоящей работе проведено на материале микроэпизодов (микродиалогов), которые можно считать относительно однородными (по теме и задачам) единицами коммуникации.

2. Система аннотирования речевых актов в корпусе ОРД

Исходное понятие «речевого акта», изначально предложенное Дж. Остином [Austin 1962], затем существенно переработанное Дж. Сёрлем [Searle 1976], а потом и многими другими их последователями, к настоящему времени претерпело значимую трансформацию. Сам термин «речевой акт» является достаточно популярным в современной лингвистике, однако определения РА и конкретные категории РА в трудах разных ученых и в разных научных школах сильно разнятся.

В прагматических исследованиях, проводимых на материале корпуса ОРД, под речевым актом понимается целенаправленное речевое действие, рассматриваемое в контексте прагматической ситуации и обладающее определенной иллюкутивной силой. Предполагается, что любая реплика говорящего состоит из одного или нескольких РА. Более того, мы полагаем, что любое высказывание может быть интерпретировано как речевой акт определенного типа. В этом смысле наш подход отличается от традиционного понимания РА, предложенного

² Разработка системы прагматического аннотирования была поддержана грантом РГНФ № 12-04-12017 «Информационная система коммуникативных сценариев спонтанной русской речи».

Дж. Сёрлем. Более близким для нас является позиция М. М. Бахтина, который считал, что любое высказывание, каким бы кратким (и даже незавершенным) оно ни было, имеет некоторое значение или несет определенную информацию о говорящем [Bakhtin 1986]. Поэтому мы исходим из положения, что любая устная реплика может быть интерпретирована (проаннотирована) с позиции своего прагматического значения (роли) в процессе коммуникации. Формальное аннотирование речевых актов подразумевает их предварительное вычленение (сегментацию) и выполняется вручную в среде ELAN [Sloetjes & Wittenburg 2008].

При разработке классификационной схемы РА для аннотирования корпуса ОРД были проанализированы наиболее известные системы формального представления речевых действий, используемые в разных лингвистических корпусах с прагматической разметкой: например, корпус SPAAC (Speech Act Annotated Corpus, Великобритания) [Weisser 2003]; [Leech & Weisser 2003], система Dialogue Act Markup in Several Layers (DAMSL) [Allen & Core 1997], международный проект Cross-Cultural Study of Speech Act Realization Patterns [Blum-Kulka & Olshtain 1984], VRM-система дискурсивной таксономии (Verbal Response Modes), предложенная В. Стайлесом [Stiles 1992], и некоторые другие [Weisser 2014].

Однако большинство предлагаемых классификаций было разработано для ограниченного перечня коммуникативных сценариев (например, обращение в колл-центр по телефону или покупка железнодорожных билетов) и поэтому не подходит для аннотирования такого сложного жанра, как повседневное речевое общение. Для решения нашей задачи целесообразным показалось использовать классификации речевых поступков, разработанные отечественными лингвистами специально для русской разговорной речи [Борисова 2009].

Основные типы речевых актов, аннотируемых в корпусе ОРД, определяются следующим образом [Шерстинова 2015].

- *Репрезентативы* (ИНФ) — РА, главной целью которых является обмен информацией между участниками диалога.
- *Директивы* (ДИР) — побуждают адресата к действию (или бездействию) или выражают попытку повлиять на его мировоззрение, эмоции и установки.
- *Комиссивы* (КОМ) — связаны с принятием говорящим на себя определенных обязательств.
- *Экспрессивы-эмотивы* (ЭМО) — используются для выражения и передачи чувств и эмоций.
- *Этикетные экспрессивы* (ЭТИ) — стандартизованные формы, регулирующие коммуникацию в этикетных и ритуализированных ситуациях.
- *Валюативы* (ВАЛ) — используются для выражения оценочного мнения или мнения-суждения.
- *Суппозитивы* (СУП) — выражают мнение или предположение говорящего.
- *Коммуникативные регулятивы* (РЕГ) — фатические речевые поступки, связанные с «организационными» аспектами взаимодействия, используемые для структурирования и ведения диалога.

Особенностью реальной речевой коммуникации является тот факт, что часть высказываний остаются незавершенными, оборванными, «брошенными

на полуслове». При этом прозвучавшие фрагменты не всегда дают возможность реконструировать первоначальную интенцию говорящего, т. е. какой тип речевого акта планировался к воспроизведению. Такие сегменты речевого потока в корпусе помечаются кодом (ФРА), что означает *неопределяемый фрагмент* — т. е. незавершенный РА, по которому невозможно определить его иллокутивную силу.

Далее, поскольку записи ОРД выполняются в «полевых» условиях, где естественным является наличие фоновых шумов (шум улицы, посторонние разговоры, звук телевизора и др.), то некоторые речевые фрагменты оказываются *нерасшифруемыми* (НЕР). Так помечаются РА, для которых невозможно получить текстовую расшифровку вследствие особенностей коммуникации или качества звукозаписи.

Наконец, в устной речевой коммуникации не последнюю роль играют *паралингвистические события* (ПАР), многие из которых могут иметь иллокутивную силу (например, смех, вздох, стон и пр.).

В мультивекторных разговорах (когда имеют место несколько параллельных разговоров) аннотируется только основная линия коммуникации, посторонние и параллельные реплики и диалоги не сегментируются на речевые акты и соответствующим образом не аннотируются.

В каждом основном типе речевых актов выделяются подтипы, необходимые для того, чтобы, например, отделить *вопросы* от *ответов* в общей категории *репрезентативов* или *просьбу* от *приказа* среди директивов. Некоторые из наиболее частотных подтипов РА для каждого основного типа приведены в табл. 1.

Таблица 1. Некоторые подтипы РА, выделяемые при аннотировании

№	Основной тип РА	Подтипы РА
1	Репрезентатив	Дескриптив, констатив, сообщение, экспликатив, комментарий, вопрос (рогатив), напоминание, уточнение, репродуктив (передача чужой речи) и др.
2	Регулятив	Тематическая инициатива, речевая поддержка, речевая придержка, запрос на наличие контакта, отказ от темы, смена темы, переспрос, показатель готовности к коммуникации и др.
3	Валюатив	Возражение, отрицание, опровержение, согласие и т. п.
4	Директив	Просьба, требование, приказ, совет, убеждение, разрешение, утешение и др.
5	Этикетная форма	Вокатив, приветствие, прощание, извинение, благодарность, пожелание, поздравление и др.
6	Экспрессив-эмотив	Выражение положительных эмоций, выражение отрицательных эмоций, удивление и т. п.
7	Комиссив	Обещание, согласие выполнить просьбу, заявление о намерениях, отклонение предложения и др.
8	Суппозитив	Предположение, выражение личного мнения и под.
9	Паралингвистическое событие	Смех, вскрик, вздох, стон, свист, цоканье языком и другие невербальные события, значимые для процесса коммуникации.

При аннотировании РА в корпусе ОРД разметка выполняется на следующих четырех уровнях:

1. RAct (речевой акт) — «форма выражения» речевого акта, т. е. орфографическая запись соответствующего фрагмента речи с использованием синтагматической и фразовой разметки.
2. RActSp (говорящий) — код информанта или коммуниканта.
3. RAGenT — общий тип речевого акта.
4. RADetT — детализация речевого акта (подтип).

В табл. 2 приведен пример аннотирования фрагмента диалога с использованием разработанной схемы.

Таблица 2. Пример аннотирования фрагмента диалога

Реплика	Код говорящего	Речевого акт	Тип РА	Подтип РА
P1	Ж1	<i>так //</i>	РЕГ (регулятив)	ИНИ (начало темы)
	Ж1	<i>всё-таки не туда ? <пауза> или туда?</i>	ИНФ (репрезентатив)	ВОПР (вопрос)
P2	M1	<i>мне кажется / что вон туда // до конца //</i>	СУП (суппозитив)	ПРЕДП (предположение)
P3	Ж1	<i>это не она //</i>	ВАЛ (валюатив)	ВОЗР (возражение)
	Ж1	<i>вот туда //</i>	ИНФ (репрезентатив)	КОРР (коррекция)
P4	M1	<i>но там зато фонарь виден хорошо //</i>	ИНФ (репрезентатив)	ЭКСП (экспликатив)
P5	Ж1	<i>угу //</i>	ВАЛ (валюатив)	СОГЛ1 (согласие)

3. Материал и методика исследования

Как уже было отмечено во введении, материалом настоящего исследования стали 73 микроэпизода повседневного общения, включающие как рабочие, так и домашние разговоры и общение друзей. Анализируемые диалоги относятся к 6 макроэпизодам и включают в общей сложности речь 30 человек — 6 информантов и 24 их коммуникантов, мужчин и женщин, представителей разных возрастных и профессиональных групп.

Весь речевой материал был поделен на 2230 речевых актов, которые были вручную проаннотированы. Полученная статистика о распределении основных типов РА в исследуемой выборке приведена в табл. 3.

Таблица 3. Распределение основных типов речевых актов

№	Тип РА	Аббревиатура	Абс. кол-во	%
1	Репрезантатив	ИНФ	884	39,62
2	Регулятив	РЕГ	279	12,51
3	Валюатив	ВАЛ	254	11,39
4	Директив	ДИР	151	6,77
5	Этикетная форма	ЭТИ	93	4,17
6	Паралингвистическое событие	ПАР	83	3,72
7	Экспрессив-эмотив	ЭМО	79	3,54
8	Комиссив	КОМ	62	2,78
9	Суппозитив	СУП	58	2,60
10	Незавершенный фрагмент	ФРА	57	2,55
11	Нерасшифруемый фрагмент	НЕР	18	0,81
12	Смешанные типы	ИНФ/ЭМО, ИНФ/РЕГ, ИНФ/ВАЛ, ДИР/ЭТИ и др.	249	11,16

Таким образом, почти 40% всех высказываний относятся к категории репрезентативов, главная задача которых состоит в обмене информацией, 12,5% речевых актов являются регулятивными формами, задающими и поддерживающими ход диалога, 11,4% высказываний являются валюативами, выражающими оценочные мнения или суждения. Значительно реже встречаются директивы (6,7%), этикетные формы (4,2%) и др. категории РА, а 9,38% всех высказываний выборки представляют собой смешанные типы.

Следует обратить внимание на то, что паралингвистические явления составляют довольно значимый компонент устной коммуникации (3,7%). Подробнее о частоте паралингвистических событий в повседневной речи см. [Sherstinova 2018].

Рассмотрим, какие подкатегории речевых актов оказались самыми частотными для основных типов (см. табл. 4). Процент посчитан по отношению ко всем речевым актам данного типа.

Таким образом, среди репрезентативов самые частые категории — это *вопросы* (28,65%) и *экспликативы* (*разъяснения*, 17,35%), наиболее употребительные регулятивные формы — это разного рода *маркеры границ* (начала, сегмента, конца — *так, вот, ну* и т.п.), *речевая поддержка* (*ага, угу, да* и т.п.), *переспрос*. Среди валюативов на нашей выборке больше всего встретилось выражения *согласия* или *одобрения* (37,84%), в то время как *возражение* имело место почти в 4 раза реже (9,96%). Наиболее частотные директивы — *предложение* (25,83%) и *просьба* (19,87%). Среди этикетных форм выделяются *вокативы* (треть всех реализаций) и *приветствия* (28%), а среди эмотивов — *положительные эмоции* (21%) и *удивление* (14%). Коммисивы и суппозитивы в речевой коммуникации используются нечасто и их подтипы немногочисленны: для коммисивов это *согласие выполнить просьбу* (30%), *заявление о намерениях* (30%) и *обещание* (12%), а для суппозитивов — *предположение* (65%) и *выражение личного мнения* (23%).

Таблица 4. Наиболее частотные подтипы речевых актов в исследуемой выборке

№	Тип РА	Подтип	Абс. кол-во	%
1	Репрезантатив	Вопрос (ВОПР)	251	28,65
		Экспликатив (ЭКСП)	152	17,35
		Сообщение (СООБ)	119	13,58
		Ответ (ОТВ)	105	11,98
2	Регулятив	Маркер начала (ИНИ)	55	19,78
		Речевая поддержка (РП)	41	14,74
		Переспрос (ПЕРЕ)	34	12,23
		Маркер границы (ВОТ)	25	8,99
		Подтверждение понимания	24	8,63
3	Валюатив	Согласие, одобрение (СОГЛ)	95	37,84
		Суждение (СУЖ)	31	12,35
		Возражение (ВОЗ)	25	9,96
4	Директив	Предложение (ПРЕД)	39	25,83
		Просьба (ПРОС)	30	19,87
		Привлечение внимания (ВНИМ)	22	14,57
		Инструкция (ИНС)	9	5,96
5	Этикетная форма	Вокатив (ВОК)	33	35,48
		Приветствие (ПРИВ)	26	27,95
		Прощание (ПРОЩ)	5	5,37
6	Паралингвистическое событие	Смех	57	71,25
		Вздых	10	12,50
7	Экспрессив-эмотив	Положительные эмоции (ЭПОЛ)	16	20,78
		Удивление (УДИВ)	11	14,28
		Реакция на внешний источник (РЕАК)	10	12,99
		Недовольство (НЕД)	9	11,68
		Отрицательные эмоции (ЭОТР)	7	9,09
8	Комиссив	Согласие выполнить просьбу и под. (КСОГЛ)	18	30,51
		Заявление о намерениях (ЗАЯН)	18	30,51
		Обещание (ОБЕ)	7	11,86
9	Суппозитив	Предположение (ПРЕДП)	37	64,91
		Выражение личного мнения (МНЕ)	13	22,81
10	Смешанные типы	Вопрос+выражение эмоций	249	11,16*
		Вопрос+начало темы		
		Ответ+выражение эмоций		
		Просьба+вопрос		
		Просьба+выражение эмоций и др.		

* в данном случае — от общего кол-ва РА

Для подсчета наиболее частотных последовательностей РА использовалась методика n-граммного анализа — довольно популярный статистический метод, позволяющий строить вероятностные лингвистические модели. Следует отметить, что в данном исследовании диалог рассматривается как единая структура, состоящая из последовательности речевых актов; длительность пауз между репликами и смена коммуникативных ходов говорящими при этом не учитываются³.

4. Наиболее типичные бинарные последовательности речевых актов и частотность отдельных типов речевых актов в начале и конце диалога

В табл. 5 приведены данные о частоте встречаемости пар РА по основным типам. В первом столбце приведены начальные элементы биграмм, в первой строке — соответствующие им вторые элементы. Доля встречаемости представлена в процентах. Начало и конец микродиалогов обозначены в таблице как (НД) и (КД), что позволяет определить речевые акты, преобладающие в начале и конце диалога. Наиболее частотные биграммы РА в таблице маркированы.

Таблица 5. Доля встречаемости пар речевых актов в исследуемой выборке (%)

	ЭТИ	ЭМО	СУП	РЕГ	ПАР	КОМ	ИНФ	ДИР	ВАЛ	(КД)	Итого
(НД)	0,33	0,28	0,17	0,55	0,22	0,33	0,67	0,33	0,55	—	3,44
ВАЛ	0,17	0,28	0,61	1,33	1,11	0,39	5,21	1,05	2,55	0,06	12,76
ДИР	0,55	0,33	0,33	1,22	0,11	0,28	2,77	0,78	0,94	0,28	7,60
ИНФ	1,44	1,16	1,44	6,05	1,33	0,89	22,35	2,77	4,77	1,11	43,32
КОМ	0,11	0,11	0,00	1,11	0,00	0,39	0,78	0,55	0,33	0,00	3,38
ПАР	0,11	0,33	0,06	0,06	0,17	0,00	1,66	0,39	0,89	0,06	3,72
РЕГ	0,44	0,44	0,22	2,38	0,33	0,67	6,93	0,94	1,28	0,78	14,42
СУП	0,06	0,00	0,17	0,55	0,11	0,00	1,16	0,22	0,50	0,00	2,77
ЭМО	0,17	0,72	0,06	0,33	0,22	0,00	1,50	0,17	0,44	0,11	3,72
ЭТИ	1,22	0,17	0,00	0,50	0,11	0,22	0,83	0,28	0,61	0,94	4,88

Из таблицы 5 видно, что наиболее типичными бинарными последовательностями РА оказались: два репрезентатива подряд (ИНФ+ИНФ, 22,35%), регулятивная форма и следующий за ней репрезентатив (РЕГ+ИНФ, 6,93%), репрезентатив и регулятивная форма (ИНФ+РЕГ, 6,05%), валюатив и следующий за ним репрезентатив (ВАЛ+ИНФ, 5,21%), репрезентатив и оценочное суждение (ИНФ+ВАЛ, 4,77%), а также двусторонняя комбинация директива с репрезентативом (ДИР+ИНФ и ИНФ+ДИР — по 2,77%).

³ Тем не менее, представляется целесообразным рассмотреть эти факторы при продолжении прагматических исследований корпуса ОРД в будущем.

Из таблицы можно получить и информацию об относительной частоте отдельных типов речевых актов в начале и конце диалога. Однако для удобства интерпретации данных их процент лучше пересчитать по отношению ко всем позициям начала-конца микроэпизода.

В итоге оказалось, что *инициируют диалог* чаще всего репрезентативы, т.е. РА, связанные с обменом информацией (38% случаев от всех начальных позиций микродиалогов), «этикетное» начало (приветствия, вокативы) имеет место в 23% диалогов, а в 19% случаев разговор начинается с регулятивной формы.

Речевые акты, завершающие диалог, показывают большее разнообразие: это репрезентативы (16%), оценочные суждения (валюативы) (14%), регулятивные формы (14%), по 8% — директивы, комиссивы и этикетные формы и 7% — экспрессивы.

Однако данные, приведенные в табл. 5, могут показаться недостаточно информативными, если мы хотим проанализировать структуру диалога на более детальном уровне. Воспользуемся для этой цели частотным списком биграмм, полученных при анализе речевых актов на уровне подтипов (см. табл. 6).

Таблица 6. Наиболее частотные пары речевых актов на уровне подтипов

Ранг	Последовательность	%
1	<i>вопрос — ответ</i>	2,45
2	<i>вопрос — вопрос</i>	1,01
3	<i>экспликатив — экспликатив</i>	0,97
4	<i>ответ — вопрос</i>	0,93
5	<i>экспликатив — вопрос</i>	0,93
6	<i>сообщение — сообщение</i>	0,89
7	<i>вопрос — экспликатив</i>	0,76
8	<i>экспликатив — согласие</i>	0,55
9	<i>согласие — вопрос</i>	0,51
10	<i>сообщение — вопрос</i>	0,46

Видно, что все наиболее частотные пары РА относятся к категории репрезентативов. И, вполне предсказуемо, на первом месте здесь находится пара *вопрос — ответ*, покрывающая 2,45% всех случаев.

Вопрос оказался также весьма типичным началом разговора: с вопроса начинаются 16% всех микродиалогов исследованной выборки. В 17% случаев микродиалог инициируется маркером начала новой темы (например, репликой *так*). Также диалоги регулярно начинаются с приветствия (13%), с сообщения (11%) или с вокатива (10%).

Что касается реплик, завершающих диалог, на уровне подтипов РА почти не оказалось регулярно повторяющихся. Самым частотным финалом диалога оказалось *согласие*, который завершает диалог в 10% случаев, относительно частотны *сообщения* и *регулятивные формы* (напр., *вот*) (по 4% случаев). Таким образом, начало диалога кажется намного проще для формального моделирования структуры диалога, чем его конец.

5. Заключение

Проведенное исследование показало, что инициируют диалог чаще всего репрезентативы, т. е. речевые акты, связанные с обменом информацией (38% случаев), «этикетное» начало (приветствия, вокативы) имеет место в 23% диалогов, а в 19% случаев разговор начинается с регулятивной формы. РА, завершающие диалог, показывают большее разнообразие: это репрезентативы (16%), оценочные суждения (валюативы) (14%), регулятивные формы (14%), по 8% — директивы, комиссивы и этикетные формы и 7% — экспрессивы. Наиболее типичными бинарными последовательностями РА оказались: два репрезентатива (17,5%), репрезентатив и регулятивная форма (5,4%), регулятивная форма и следующий за ней репрезентатив (4,7%), репрезентатив и оценочное суждение (4,1%), валюатив и следующий за ним репрезентатив (3,7%), а также двусторонняя комбинация директива с репрезентативом (по 2,2%).

Разумеется, при интерпретации любых корпусных результатов большое значение имеет используемая в корпусе система аннотирования данных. Схема аннотирования речевых актов, разработанная для корпуса ОРД, как и многие другие формальные классификации, предназначенные для анализа такой сложной и многоаспектной материи, как устная речь, конечно, не является идеальной и лишенной недостатков. Так, не всегда атрибуцию РА по типу и подтипу можно произвести однозначно и непротиворечиво. Однако на сегодняшний день это оптимальная структура, к которой мы пришли в результате анализа реального речевого материала, и которая, на наш взгляд, в большинстве случаев дает вполне корректный и однозначно интерпретируемый результат.

Мне бы хотелось поблагодарить анонимных рецензентов данной работы за их конструктивные замечания и пожелания, большинство из которых обязательно будут приняты во внимание при продолжении исследования. Наиболее интересным мне кажется предложение размечать речевые акты по нескольким независимым параметрам, выделив те или иные аспекты субъективной модальности в независимое поле описания. Развивая эту идею, можно было бы также маркировать и эмоциональную окраску РА, и наличие в нем регулирующих разговор прагматических маркеров — то есть те факторы, которые, как видно из табл. 3, чаще других приводят к появлению «смешанных типов» в используемой эмпирической классификации. При разработке такой системы аннотирования, по-видимому, следует еще раз обратить внимание на VRM-систему дискурсивной таксономии В. Стайлеса [Stiles 1992], в которой подобный подход уже был реализован. Многоаспектное аннотирование речевых актов представляется перспективным, хотя достаточно трудоемким, и приведет к необходимости полного пересмотра используемой классификации (и, следовательно, повторного аннотирования речевого материала).

Надо сказать, что в прагматических исследованиях существует представление о том, что сколько в языке имеется глаголов, связанных с описанием коммуникации, столько разнообразных подкатегорий речевых актов говорящие способны выделять. Действительно, предлагаемые исследователями списки категорий довольно представительны и насчитывают порядка 200 единиц. Но так как многие из них, с одной стороны, близки по значению

или синонимичны, а с другой — не учитывают «реального наполнения» прагматического компонента высказывания (который, как выясняется, до сих пор не имеет адекватного «словарного» описания), построение единой и непротиворечивой классификации РА, которая служила бы основой для их корректной лингвистической и прагматической интерпретации, можно считать одной из «гильбертовых проблем» современной лингвистики.

Литература

1. *Богданова-Бегларян Н. В., Блинова О. В., Мартыненко Г. Я., Шерстинова Т. Ю.* (2017), Некоторые инвариантные характеристики русской разговорной речи: фонетика, морфология, синтаксис // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 31 мая — 3 июня 2017 г.). Вып. 16 (23): В 2 т. Т. 2 — М.: Изд-во РГГУ, 2017, с. 75–86.
2. *Борисова И. Н.* (2009), Русский разговорный диалог. Структура и динамика. М.: Книжный дом «ЛИБРОКОМ», 320 с.
3. *Сибата Т.* (1983), Исследования языкового существования в течение 24 часов // Алпатов В. М., Вардуль И. Ф. (ред.) Языкознание в Японии, М.: Радуга, с. 134–141.
4. *Шерстинова Т. Ю.* (2015), Прагматическое аннотирование коммуникативных единиц в корпусе ОРД: микроэпизоды и речевые акты / Захаров В. П., Хохлова М. В. (ред.) Труды международной конференции «КОРПУСНАЯ ЛИНГВИСТИКА — 2015». СПб.: СПбГУ, с. 436–445.
5. *Шерстинова Т. Ю.* (2016), Наиболее употребительные слова повседневной русской речи (в гендерном аспекте и в зависимости от условий коммуникации) // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 1–4 июня 2016 г.). Вып. 15 (22). М.: Изд-во РГГУ, 2016, с. 616–631.
6. *Allen J. & Core M.* (1997), Draft of DAMSL: Dialog act markup in several layers. Online at: <https://www.cs.rochester.edu/research/speech/damsl/RevisedManual/> (last accessed on Feb. 17, 2018).
7. *Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T.* (2009), The ORD Speech Corpus of Russian Everyday Communication «One Speaker's Day»: Creation Principles and Annotation. In: Matoušek, V., Mautner, P. (eds.) TSD 2009. LNAI, vol. 5729. Springer, Berlin-Heidelberg, pp. 250–257.
8. *Austin J. L.* (1962), How To Do Things With Words, Oxford University Press, Oxford.
9. *Bakhtin M. M.* (1986), Speech Genres and Other Late Essays, Univ. of Texas Press, Austin, TX.
10. *Bogdanova-Beglarian N., Sherstinova T., Blinova O., Ermolova O., Baeva E., Martynenko G., Ryko A.* (2016), Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech / Ronzhin, A. et al. (eds.) SPECOM 2016, Lecture Notes in Artificial Intelligence, LNAI, vol. 9811. Springer, Switzerland, pp. 659–666.

11. *Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G.* (2017), Linguistic Features and Sociolinguistic Variability in Everyday Spoken Russian. In: Karpov A., Potapova R., Mporas I. (eds) *Speech and Computer. SPECOM 2017. Lecture Notes in Computer Science*, vol 10458. Springer, Cham, pp. 503–511.
12. *Blum-Kulka S. & Olshtain E.* (1984), Requests and Apologies: A Cross-Cultural Study of Speech Act Realization Patterns (CCSARP) *Applied Linguistics*, Vol. 5, No. 3, pp. 196–213.
13. *Burnard L.* (Ed.) (2016), Reference guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by Oxford University Computing Services, 2007. Available at: <<http://www.natcorp.ox.ac.uk/docs/URG/>>. Retrieved: February 20, 2018.
14. *Leech G. & Weisser M.* (2003), Generic speech act annotation for task-oriented dialogues, *Proc. of the Corpus Linguistics 2003 Conference*. Lancaster University: UCREL Technical Papers, V. 16.
15. *Searle J. R.* (1976), A classification of illocutionary acts, *Language in Society*, 5(1), pp. 1–23.
16. *Sherstinova T.* (2018), Audible Paralinguistic Phenomena in Everyday Spoken Conversations: Evidence from the ORD Corpus Data, *Language, Music, and Computing (LMAC-2017), Communications in Computer and Information Science*, CCIS, Springer International Publishing Switzerland (in print)
17. *Sloetjes H. & Wittenburg P.* (2008), Annotation by category — ELAN and ISO DCR, *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*.
18. *Stiles W.* (1992), *Describing Talk: A Taxonomy of Verbal Response Modes*, Sage, Newbury Park, CA.
19. *Weisser M.* (2003), SPAACy: A semi-automated tool for annotating dialogue acts. *International Journal of Corpus Linguistics* 8 (1), pp. 63–74
20. *Weisser M.* (2014), Speech act annotation, *Corpus Pragmatics: a Handbook*, CUP, Cambridge, pp. 84–111.

References

1. *Allen, J. and Core, M.* (1997), Draft of DAMSL: Dialog act markup in several layers. Online at: <https://www.cs.rochester.edu/research/speech/damsl/Revised-Manual/> (last accessed on Feb. 17, 2018).
2. *Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T.* (2009), The ORD Speech Corpus of Russian Everyday Communication «One Speaker's Day»: Creation Principles and Annotation, *TSD 2009, LNAI*, vol. 5729, pp. 250–257.
3. *Austin J. L.* (1962), *How To Do Things With Words*, Oxford University Press, Oxford.
4. *Bakhtin M. M.* (1986), *Speech Genres and Other Late Essays*, Univ. of Texas Press, Austin, TX.

5. *Blum-Kulka S. and Olshtain E.* (1984), Requests and Apologies: A Cross-Cultural Study of Speech Act Realization Patterns (CCSARP) Applied Linguistics, Vol. 5, No. 3, pp. 196–213.
6. *Bogdanova-Beglarian N., Sherstinova T., Blinova O., Ermolova O., Baeva E., Martynenko G., Ryko A.* (2016) Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech / Ronzhin, A. et al. (eds.) SPECOM 2016, Lecture Notes in Artificial Intelligence, LNAI, vol. 9811. Springer, Switzerland, 2016, pp. 659–666.
7. *Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G.* (2017), Linguistic Features and Sociolinguistic Variability in Everyday Spoken Russian, SPECOM 2017, LNCS, vol. 10458, pp. 503–511.
8. *Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G.* (2017), Some invariant features of Russian everyday speech: Phonology, morphology, syntax, *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, 2(16), pp. 82–95.
9. *Borisova I. N.* (2009), Russian spoken dialogue. Structure and Dynamics [Russkiy razgovornyy dialog. Structura i dinamika], LIBROKOM [Knizhnyj dom LIBROKOM], Moscow.
10. *Burnard, L.* (Ed.) (2016), Reference guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by Oxford University Computing Services, 2007. Available at: <<http://www.natcorp.ox.ac.uk/docs/URG/>>. Retrieved: February 20, 2018.
11. *Leech G. & Weisser M.* (2003), Generic speech act annotation for task-oriented dialogues, Proc. of the Corpus Linguistics 2003 Conference. Lancaster University: UCREL Technical Papers, V. 16.
12. *Searle J. R.* (1976), A classification of illocutionary acts, *Language in Society*, 5(1), pp. 1–23.
13. *Sherstinova T.* (2015), Approaches to Pragmatic Annotation in the ORD Corpus: Microepisodes and Speech Acts [Pragmaticheskoe annotirovanie kommunikativnykh jedinic v korpuse ORD: mikroepisydy i rechevye akty], Proc. of the Int. Conf. “Corpus linguistics-2015” [Trudy mezhdunarodnoy nauchnoy konferentsii “Korpusnaya linguistica-2013”], pp. 436–446.
14. *Sherstinova T.* (2015), Macro Episodes of Russian Everyday Oral Communication: towards Pragmatic Annotation of the ORD Speech Corpus, SPECOM 2015, LNAI, vol. 9319, pp. 268–276
15. *Sherstinova T.* (2016), The most frequent words in everyday spoken Russian (in the gender dimension and depending on communication settings), *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, pp. 616–631.
16. *Sherstinova T.* (2018), Audible Paralinguistic Phenomena in Everyday Spoken Conversations: Evidence from the ORD Corpus Data, *Language, Music, and Computing (LMAC-2017)*, Communications in Computer and Information Science, CCIS, Springer International Publishing Switzerland (in print).
17. *Sibata, T.* (1983), Issledovanija jazykovogo sushhestvovaniija v techenie 24 časov [Studying the language life with the method of the 24 hour survey]. In: *Alpatov, V. M., Vardul, I. F.* (eds.) *Jazykoznanie v Japonii* [Linguistics in Japan], Raduga, Moscow, pp. 134–141.

18. *Sloetjes H. & Wittenburg P. (2008), Annotation by category — ELAN and ISO DCR, Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008.*
19. *Stiles W. (1992), Describing Talk: A Taxonomy of Verbal Response Modes, Sage, Newbury Park, CA.*
20. *Weisser M. (2003), SPAACy: A semi-automated tool for annotating dialogue acts. International Journal of Corpus Linguistics 8 (1), pp. 63–74.*
21. *Weisser M. (2014), Speech act annotation, Corpus Pragmatics: a Handbook, CUP, Cambridge, pp. 84–111.*
22. *Wittenburg P., Brugman H., Russel A., Klassmann A., Sloetjes H. (2006) ELAN: a Professional Framework for Multimodality Research, Proc. of LREC 2006, Fifth International Conference on Language Resources and Evaluation.*