

Т. Ю. Шерстинова

Санкт-Петербургский государственный университет
Национальный исследовательский университет «Высшая школа экономики»
(Россия, Санкт-Петербург)
t.sherstinova@spbu.ru; tsherstinova@hse.ru

ДА-ДА-ДА-ДА-ДА! И ДРУГИЕ КОНСТРУКЦИИ ПОВСЕДНЕВНОЙ РУССКОЙ РЕЧИ, СОСТОЯЩИЕ ИЗ КОНТАКТНО ПОВТОРЯЮЩИХСЯ СЛОВ, В КОНТЕКСТЕ ИССЛЕДОВАНИЯ ПРОСОДИЧЕСКИХ МОДЕЛЕЙ ЯЗЫКА

Яркой характерной чертой повседневной русской речи являются реплики или фрагменты реплик, состоящие исключительно из повтора одних и тех же лексических единиц: *да-да, вот-вот, ну-ну, давай-давай, пока-пока, да-да-да, так-так-так, щас-щас-щас, нет-нет-нет, да-да-да-да-да* и др. Структурная организация и просодические характеристики таких конструкций заслуживают особого внимания фонетистов, поскольку представляется, что с их помощью можно выявить базовые просодические модели языка. Материалом для настоящего исследования стали звукозаписи и расшифровки повседневной устной речи из корпуса «Один речевой день» (ОРД) в объеме 900 тыс. словоупотреблений. С помощью программы построения *n-грамм* по текстам расшифровок были получены данные о частоте реплик, представляющих собой двух- и трехэлементные конструкции типа *да-да* и *да-да-да*, а также предварительные данные о частоте встречаемости повторов из четырех и пяти слов. Оказалось, что чаще всего подобные цепочки образуют слова *да, ну, вот, так, щас (сейчас)* и другие высокочастотные элементы устного дискурса, причем абсолютным «лидером» среди них является частица *да*. При исследовании конструкций, представляющих собой цепочки контактно повторяющихся слов, следует учитывать, являются ли они целостно-неделимыми с просодической точки зрения или же представляют собой комбинацию конструкций меньшей размерности. Поскольку просодическая выделенность может характеризовать более чем один элемент таких конструкций (например начальный и конечный), в некоторых случаях возникают сложности с определением их ритмической структуры. Просодическое моделирование для конструкций, состоящих из контактно повторяющихся слов, следует начать с выявления базовых моделей небольших по размеру (например двух- или трехэлементных) изолированных реплик.

Ключевые слова: русский язык, устная речь, повседневная речь, лексический повтор, эпизевксис, геминация, частицы, дискурсивные слова, п-граммы, просодия, просодическая модель, речевой корпус

Яркой характерной чертой повседневной русской речи являются реплики или фрагменты реплик, состоящие исключительно из повтора одних и тех же лексических единиц, например: *да-да, вот-вот, ну-ну, давай-давай, пока-пока, да-да-да, так-так-так, щас-щас-щас, нет-нет-нет, да-да-да-да-да* и т. д. Как правило, элементами таких повторяющихся цепочек слов выступают частицы и близкие к ним речевые единицы (*да, нет, вот, не, ну, так, ладно, давай* и др.), а также междометия (*ой, пока, кьи* и др.). Чаше других используются двукратные и трехкратные повторы, но встречаются и повторы значительно большего количества единиц — вплоть до 9 слов (как это будет показано ниже). Трудно себе представить подобные многократные повторы в письменных текстах или в устной речи других жанров (например в официальном дискурсе), однако для повседневных разговоров они представляют собой типичное и весьма распространенное явление. В данной работе будем использовать аббревиатуру *КПС* для обозначения любых цепочек *контактно повторяющихся слов*, тогда их типовые (наиболее часто используемые) модели можно назвать *КПС-конструкциями*.

Следует отметить, что в научной литературе (главным образом в риторике и поэтике) для обозначения повторения слов в контактной позиции используются следующие термины: *редупликация* (удвоение) [Бойчук 2014: 107], *эпизевксис* — «повторение слова подряд, без разрыва, в контактной позиции» [Бойчук 2014: 107], *геминация* — «контактный повтор слова, словосочетания или предложения» [Копина 2014: 91], *амплификация* (в значении «повтор») [Гришина 2010].

Регулярное использование в речи контактных повторов слов (в том числе и многократных) может объясняться как намеренным желанием участников разговора «усилить» производимый высказыванием прагматический эффект, так, возможно, и неосознанным стремлением говорящих к формированию речи по имплицитно существующим в языке речевым шаблонам — структурным и темпоральным [Шерстинова 2010; Богданова-Бегларян и др. 2013]. Поэтому думается, что высказывания, представляющие собой КПС-конструкции, заслуживают особого внимания фонетистов. Причем в первую очередь интерес представляют те повторы, которые не просто соседствуют в линейной развертке речи, но и образуют интонационно-ритмическое единство.

Так, можно предположить, что трехэлементная конструкция *да-да-да* с ударением на последнем слове (анапест) строится по схожему со многими другими подобными конструкциями (например *нет-нет-нет, так-так-так* и т. п.) интонационно-темпоральному шаблону¹. Тот факт, что повторяющиеся элементы состоят из одних и тех же цепочек аллофонов, должен упростить выявление типовых просодических структур и их вариантов.

¹ При этом надо, разумеется, учитывать прагматическое значение высказываний.

В исследовании Е. А. Гришиной [Гришина 2010] на материале Мультимедийного русского корпуса (МУРКО) рассматриваются случаи повторов «слов-предложений» *да* и *нет* на материале речи кино: приводится статистика частоты встречаемости таких образований и отмечаются их наиболее типичные просодические характеристики — ритмическая структура и движение интонационного контура. Наиболее частотными как для *да*, так и для *нет* оказались двухэлементные (однократные) повторы. Кроме того, было отмечено, что для двух- и трехэлементных конструкций ударение чаще всего падает на последний слог (что соответствует ритмической структуре ямба или анапеста), а движение тона чаще нисходящее, чем восходящее [Гришина 2010].

Материалом для настоящего исследования стали звукозаписи и расшифровки повседневной устной речи из корпуса «Один речевой день» [Asinovsky et al. 2009; Bogdanova-Beglarian et al. 2015; Bogdanova-Beglarian et al. 2016] в объеме 900 тыс. словоупотреблений. Первой из решаемых задач было выявление списка слов, которые регулярно объединяются в двойные, тройные, четверные и т. д. КПС-конструкции.

Предварительный список таких лексических единиц может быть получен автоматическим методом с помощью программы построения *n*-грамм по текстам расшифровок. *N*-граммы (где *n* — некоторое целое число, обычно в интервале от 2 до 5) представляют собой линейные последовательности *n* слов, идущих друг за другом в тексте (например, биграмы — последовательности из двух слов, триграммы — из трех слов и т. д.), которые традиционно используются в компьютерной лингвистике при решении задач распознавания речи, автоматического машинного перевода и др. для построения вероятностных *n*-граммных моделей, т. е. для оценки того, какое слово является наиболее вероятным следующим словом после данного слова (или нескольких слов) [Jurafsky, Martin 2008]. Помимо графических слов, в *n*-граммы обычно включаются *токены*, отражающие членение текста на предложения или реплики. Это позволяет получать информацию о том, какие элементы занимают начальную или конечную позиции во фразах, а также находить все цепочки слов, полностью совпадающие с репликой.

Для получения списка *n*-грамм было использовано программное обеспечение, разработанное в ООО «Центр Речевых Технологий» [ЦРТ 2016]. На вход программы подавался корпус файлов аннотаций корпуса ОРД в формате *.eaf (формат разметки среды мультимедийного лингвистического аннотирования ELAN [Hellwig et al. 2017]), обработка проводилась по уровню реплик, содержащих расшифровки речи говорящих в орфографическом виде. Всего было обработано 94 155 реплик, в результате чего был получен список *n*-грамм размером до 1 до 5 токенов включительно, состоящий из 1 648 989 разных элементов.

Следующей задачей стал отбор из этого списка самых частотных *n*-грамм, все элементы (т. е. графические слова) которых повторяются. Как было отмечено выше, эмпирические наблюдения показывают, что чаще других происходят повторы частиц (которые и сами по себе являются высокочастотными единицами устной речи), поэтому поиск КПС-конструкций было решено провести по частотному

списку словоформ, полученному на материале корпуса ОРД (см. [Шерстинова 2016]). Тем самым были получены предварительные данные о повторных формах слов с рангом от 1 до 500. В табл. 1 представлены данные для 14 высокочастотных слов, которые регулярно образуют реплики, представляющие собой двух- и трех-элементные конструкции типа *да-да* и *да-да-да*. В таблице приведены данные для тех случаев, когда эти конструкции реализованы в виде изолированной реплики (с паузой или сменой говорящего до и после них) — такая реализация КПС-конструкций является оптимальной для исследования их ритмической и интонационной вариативности. Для оценки их частотности в таблице приводится как абсолютная частота употребления соответствующего слова на материале исследованного подкорпуса, так и количественные данные об однократном употреблении этих слов в тех случаях, когда слово равняется реплике («слова-предложения» по Е. А. Гришиной).

Таблица 1

Частота образования КПС-конструкций в устной русской речи для наиболее частотных слов в виде изолированных реплик (по данным n-грамм анализа расшифровок)

Слово (W)	Абсолютная частота в выборке (W)	Однократное употребление (слово = реплике) W //	Частота повторных употреблений	
			двухэлементная реплика W-W //	трехэлементная реплика W-W-W //
<i>да</i>	14 021	2586	121	106
<i>нет</i>	4270	530	15	14
<i>вот</i>	14 630	775	6	5
<i>ну</i>	15 074	323	5	2
<i>не</i>	14 941	34	5	4
<i>что</i>	11 526	208	7	3
<i>так</i>	7116	424	6	8
<i>всё/все</i>	5961	188	7	1
<i>щас (сейчас)</i>	2880	65	1	2
<i>давай</i>	1473	157	13	4
<i>ой</i>	1211	260	9	5
<i>конечно</i>	1171	124	6	1
<i>хорошо</i>	1388	246	12	2
<i>ладно</i>	921	56	1	0

Табл. 1 показывает, что все эти слова (являющиеся преимущественно частицами) регулярно образуют повторные формы, а для таких слов, как *да*, *давай*, *хорошо*, их функционирование в составе КПС-конструкций доходит до 1% от общего количества употреблений. Обращает на себя внимание и тот факт, что количество трехэлементных конструкций для большинства представленных слов незначительно уступает количеству конструкций из двух элементов (а для слов *щас* и *так* даже оказывается несколько выше). Это говорит о том, что трехэлементные контактные лексические повторы в русской устной речи по популярности близки к двухэлементным.

К сожалению, максимальный размер n -грамм (равный 5) не позволяет получить информацию о том, какое количество КПС-конструкций размером больше трех слов присутствует в корпусе в виде изолированных реплик, так как, помимо самих слов, в n -граммы входят маркеры начала и конца реплики. При этом очевидно, что конструкции, состоящие больше чем из трех повторяющихся элементов (*да-да-да-да-да*, *да-да-да-да-да* и др.), тоже встречаются довольно регулярно. Кроме того, в табл. 1 не попали двух- и трехэлементные КПС-конструкции, входящие в состав развернутых реплик (например: *да-да / так-так-так*; *ну-ну / так я и поверил*). Для анализа их частотности необходимо обратиться к спискам 2-, 3-, 4- и 5-грамм, состоящих исключительно из графических слов. Данные, полученные для тех же 14 слов независимо от их позиции в реплике, представлены в табл. 2.

Таблица 2

Предварительная оценка частоты образования КПС-конструкций в речи для наиболее частотных слов (по данным n -грамм анализа расшифровок устной речи)

Слово (W)	Абсолютная частота в выборке (W)	Частота повторных употреблений			
		двух-элементная конструкция W-W	трех-элементная конструкция W-W-W	четырёх-элементная конструкция W-W-W-W	пяти-элементная конструкция W-W-W-W-W
<i>да</i>	14 021	1360	605	209	91
<i>нет</i>	4270	253	81	22	5
<i>вот</i>	14 630	269	31	8	3
<i>ну</i>	15 074	323	44	11	4
<i>не</i>	14 941	418	91	24	11
<i>что</i>	11 526	98	11	4	1
<i>так</i>	7116	169	78	37	19
<i>всё/все</i>	5961	131	26	5	3
<i>щас (сейчас)</i>	2880	65	25	9	2
<i>давай</i>	1473	137	63	34	18
<i>ой</i>	1211	87	43	15	4
<i>конечно</i>	1171	26	1	0	0
<i>хорошо</i>	1388	33	3	0	0
<i>ладно</i>	921	18	2	1	0

Таким образом, абсолютным «лидером» по образованию КПС-конструкций оказалась частица *да*. Высока частота контактных повторов также для *не*, *вот*, *нет*, *ну*, *так* и *давай*.

В целом числовые данные табл. 2 существенно превышают показатели, представленные в табл. 1. Однако следует отметить, что в этом случае метод n -грамм дает завышенную оценку количества исследуемых конструкций. Во-первых, одно и то же слово при повторах, превышающих два, будет посчитано несколько раз (например, одна 5-элементная конструкция *да-да-да-да-да* даст следующий «общий прирост» в итоговом списке n -грамм: 5-грамма — 1, 4-грамма — 2, триграмма — 3, биграмма — 4). Во-вторых, программа составления n -грамм игнорирует всю информацию о синтагматическом членении, сбоях, паузации и прочих особенностях,

свойственных живой речи. Поэтому по результатам автоматической обработки данных можно ожидать появления определенного количества n-грамм, которые не следует рассматривать как цельные с фонетической точки зрения образования. Так, в биграмму *да да* будут, например, преобразованы все 5 следующих фрагментов расшифровок устной речи: 1) *да да //*, 2) *да / да //*, 3) *да / <смех> да //*, 4) *да / <вдох> да //*, 5) *да <пауза> да //* и т. д. Поэтому для получения точной информации о реализации тех или иных КПС-конструкций необходим экспертный анализ текстов расшифровок, а в некоторых случаях — и обращение к исходным звукозаписям.

Ручной анализ и аудирование вариантов, необходимые для коррекции автоматической обработки (табл. 2), удобно проводить с помощью встроенного поиска программы-аннотатора ELAN, в которой выполняются расшифровки корпуса ОРД. Такой выборочный анализ был проведен для пятиэлементных КПС-конструкций с самой частотной частицей *да*. В итоге, если по списку n-грамм количество пятисловных цепочек *да-да-да-да-да* на исследованном материале равнялось 91, то в результате проверки данных по аннотированным расшифровкам только 59 из них (65%) выдержало проверку аудированием на цельность ритмической структуры. Ниже приводятся несколько примеров на пятиэлементное *да-да-да-да-да*, в которых используются следующие обозначения: % — измененное (анонимизированное) имя, *П — пауза, *Н — неразборчивый фрагмент, (...) — пауза хезитации, # — смена говорящего при наложении речи, () — заминка; дефисы между повторяющимися словами в расшифровках ОРД не ставятся.

- (1) *вот Машенька% давай / так отжимай / так тихонечко / да да да да да / вот так / вот знаешь ?* (И5, жен., 27, психолог; эпизод 01 — разговор с подругой);
- (2) *хорошо / *Н в принципе да / тут до / до Приозерска ходит (э) / ходит автобус / ага / да да да да да / кстати / *П ходит* (И40, муж., 40, врач; эпизод 54 — телефонный разговор с приятелем);
- (3) *да да да да да / всё хорошо / спасибо* (И74, муж., 31, дизайнер; эпизод 03 — официальный разговор) (рис. 1);
- (4) *да да да да да // *П ведь и совершенно верно / ведь оценивается не по максимуму знания ...* (И133, жен., 48, домохозяйка; эпизод 01 — бытовой разговор с отцом);
- (5) *а / вот ещё почему / да да да да да / да / так это вот и было в субботу* (И130, жен., 74, пенсионерка; эпизод 08 — телефонный разговор с подругой);
- (6) *вот / и мы всё хотим ей / поставить / перформансы / но (...) не можем яйцо найти // # да да да да да* (И44, муж., 41, врач; эпизод 10 — чаепитие с коллегами);
- (7) *где ? а / да да да да да / вижу на крыше / ага !* (И137, жен., 75, пенсионерка; эпизод 09 — бытовой разговор с мужем)

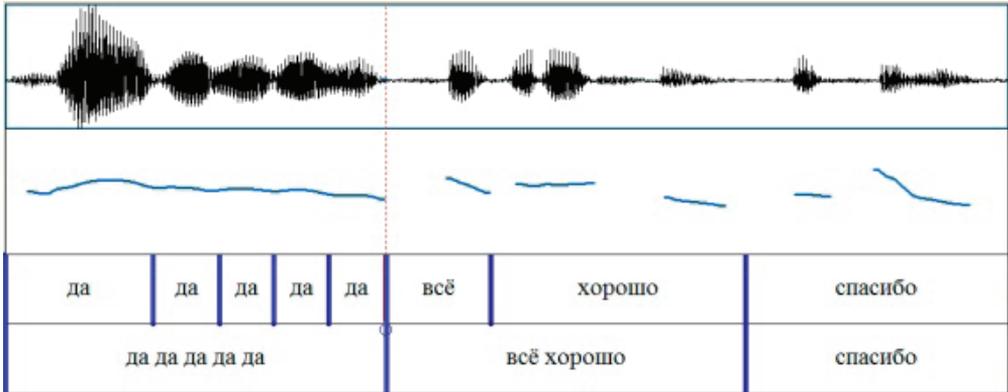


Рис. 1. Осциллограмма и ЧОТ² для фразы **да да да да / всё хорошо / спасибо**
(И74, муж., 31, дизайнер; эпизод 03 — официальный разговор)

- (8) *а она () помнишь которая беленькая такая замуж выходила ? // # да да да да да // # Кристина% ? # да да да* (И77, жен., 39, администратор, эпизод 02 — разговор с коллегами);
- (9) *метро «Балтийская» / да не так уж в принципе и далеко // *П да да да да да* (И92, муж., 46, менеджер; эпизод 09 — телефонный разговор с заказчиком).

Иногда цепочка из 5 повторяющихся слов синтагматически членится на составляющие (например 2+3 — как в (10)), а примеры (11)–(14) показывают, насколько длинными и разнообразными с точки зрения синтагматического членения могут быть контактные лексические повторы слов в повседневной речи:

- (10) *очень обтекаемо и по Украине и по всему по всему // # да да / да да да* (И131, жен., 69, пенсионерка, эпизод 18 — телефонный разговор с подругой);
- (11) *да / да да да да да / да / да / всё / ага / да всё пока // ага* (И24, жен., 63, историк; эпизод 24 — телефонный разговор с коллегой) (рис. 2);
- (12) *да / да / *П да да да да да / да / спасибо // *П угу / да / давай / пока* (И90, жен., 27, IT-специалист, эпизод 14 — телефонный разговор с коллегой);
- (13) *мы с вами вот / долго беседовали / что ... # да // # там такой / терьям-на-пам // терьям-на-пам // # да // да да да / да да да да / да да // # а он такой какой-то странный // # да* (И114-4, жен., 83, преподаватель вуза; эпизод 05 — разговор с педагогом по вокалу);
- (14) *а там же в «Июне» была аптека / её закрыли? # (а-а) да да да да / да да да да / точно / я знаю эту аптеку / да // я забыл* (И78, муж., 52, программист; эпизод 05 — разговор с заказчиком).

² Рисунок получен с помощью программы фонетического анализа Praat [Boersma & Weenink 2017].

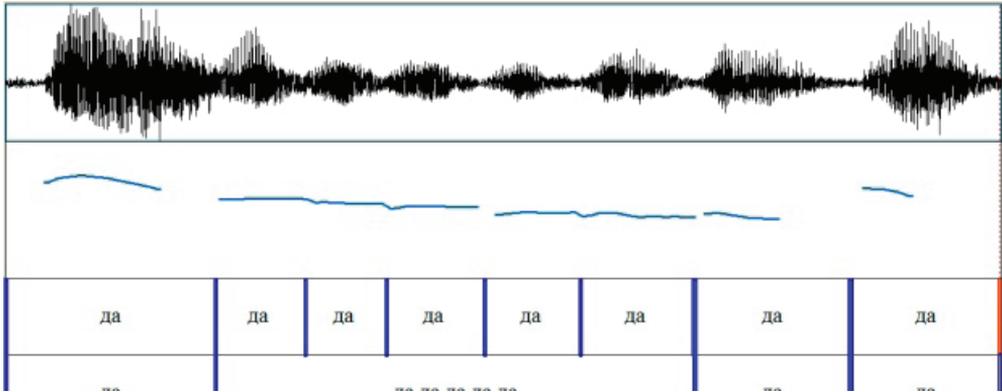


Рис. 2. Осциллограмма и ЧОТ фрагмента с 8-элементным повтором *да* из фразы *да / да да да да да да / да / да / всё / ага / да всё пока // ага* (И24, жен., 63, историк; эпизод 24 — телефонный разговор с коллегой)

При исследовании КПС-конструкций, состоящих из большого количества элементов, возникает вопрос: являются ли они целостно-неделимыми с просодической точки зрения или же представляют собой комбинацию структур меньшей размерности? Например, пятикратный повтор может быть просодически оформлен как цельная конструкция (*да-да-да-да-да*), как комбинация двух- и трехэлементной конструкций *да-да / да-да-да* или *да-да-да / да-да*, состоять из трех синтагм (например, *да / да-да-да / да*), возможны и другие варианты синтагматического членения. При этом чем больше число повторяющихся элементов в анализируемой конструкции, тем сложнее выполнять ее членение на составляющие, и далеко не всегда это можно сделать однозначно. Например, на рис. 3 приводится фраза с семикратным повтором *да-да-да-да-да-да-да*, «внутренние» (все, кроме начального и конечного) слова которой произносятся фактически на одной высоте (без движения мелодии), однако с разной длительностью и интенсивностью, что дает возможность интерпретировать эту последовательность как «просодически составную», не односинтагменную, несмотря на отсутствие пауз. Один из возможных вариантов членения этой фразы на составляющие показан на рис. 3.

Приведенный на рис. 3 речевой пример позволяет обратить внимание еще на один аспект просодического анализа КПС-конструкций. Так, при определении их ритмической структуры сложным может оказаться определение основного ударения (фразового или синтагматического), так как семантически все составляющие ее элементы являются равноправными, а просодическая выделенность может характеризовать более чем один элемент. По-видимому, чаще всего просодически маркируются начальный и конечный элементы КПС-конструкций (см. описание «рамочных» структур ниже).

Следует отметить, что, к сожалению, не для всех речевых фрагментов из корпуса ОРД возможно выполнить качественный просодический анализ. То есть не для всех конструкций, представленных в табл. 1 и 2, можно получить достоверную информацию о частоте основного тона и интенсивности речевого сигнала в том

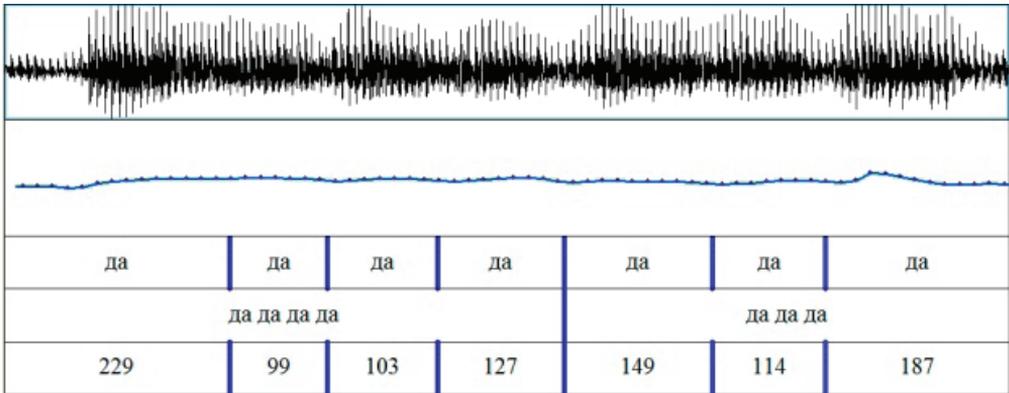


Рис. 3. Осциллограмма, ЧОТ и сегментация на слова фразы **да да да да (/) да да да** (И15, муж., 20, курсант; эпизод 01 — общение с однокурсниками) с указанием возможной границы просодического членения

виде, как это показано на рис. 1–3. Сложность заключается в том, что это записи не студийные, а выполненные на портативный диктофон в «полевых условиях»: 1) очень часто в повседневных звукозаписях на полезный речевой сигнал накладывается шум (это может быть постоянный фоновый сигнал, например работающий телевизор, радиоприемник, фоновая музыка, реклама или гул посторонних разговоров в общественных местах); 2) речь собеседников, записанная на диктофон, бывает довольно разборчива для расшифровки, однако качество сигнала, записанного на значительном (несколько метров) расстоянии от микрофона, во многих случаях не позволяет получать интонограммы и другие акустические характеристики речи; 3) одновременное наложение речи одного или нескольких говорящих присутствует во многих ситуациях речевого общения и приводит к тому, что часть речевого материала, даже записанного в условиях отсутствия внешних шумов, оказывается непригодной для инструментального анализа.

Интерес к строению и просодическим характеристикам КПС-конструкций, как уже было отмечено в начале статьи, объясняется тем, что с их помощью, по нашему мнению, можно выявить базовые просодические структуры языка. Однако для проверки этой гипотезы необходимо проведение статистически представительного исследования с привлечением методов инструментального фонетического анализа.

Приведем пока только один пример. На рис. 4 дана иллюстрация двух схожих интонационно-темпоральных структур для реплики, состоящей из двух КПС-конструкций: цепочки из пяти *щас* (*сейчас*) и следующей за ней последовательности из семи *да*. Начальный элемент каждой из двух цепочек характеризуется подъемом частоты основного тона (более того, углы подъема ЧОТ для обеих частей визуально близки) и по длительности превышает последующие. Далее в первом случае идут три элемента *щас*, во втором — пять элементов *да*, мелодические кривые для обеих цепочек имеют нисходящий характер, длительность составляющих сегментов примерно равна, но последний элемент несколько длиннее предыдущих.

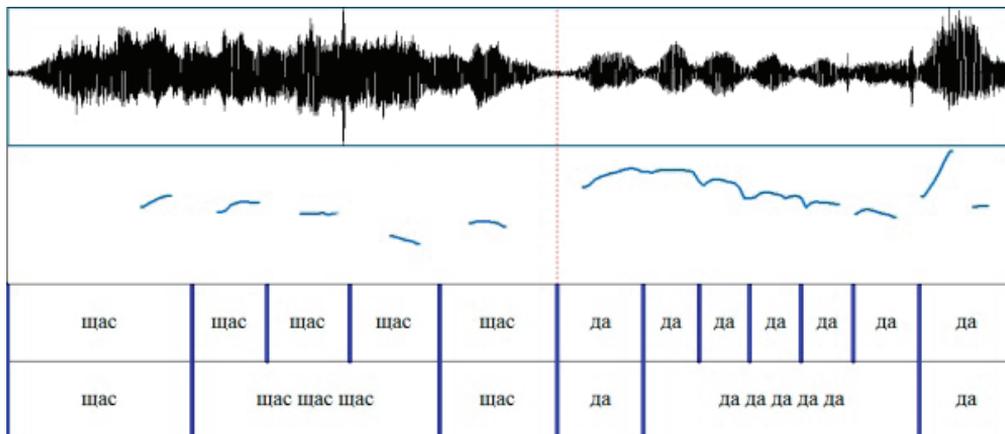


Рис. 4. Осциллограмма, ЧОТ и сегментация на слова для реплики **щас щас щас щас щас / да да да да да да** (И24, жен., 63, историк; эпизод 24 — говорит сама с собой во время работы за компьютером)

Наконец, каждая из конструкций завершается «рамочно» — одним более продолжительным по времени элементом, произносимым на средней частоте основного тона и характеризующимся завершающей интонацией. Из этого следует, что, по-видимому, можно ожидать воспроизведения такой «рамочной» структуры как в других репликах информанта И24, так и в русской повседневной речи вообще, поскольку перцептивно данная реплика воспринимается как вполне нормативная, типичная для спонтанной повседневной речи.

Проведенное исследование позволяет сделать следующие выводы:

1. Изучение расшифровок разговорной речи подтверждает эмпирические наблюдения о том, что многократные повторы частиц и междометий представляют собой достаточно регулярное явление повседневной русской речи.

2. КПС-конструкции используются для усиления прагматического эффекта, реже могут приобретать собственную семантику (как правило, сопровождаемую особой интонацией), и, кроме того, регулярность таких структур может объясняться неосознанным стремлением говорящих к формированию речи по имплицитно существующим в языке речевым шаблонам — ритмическим и темпоральным. Поэтому исследование типичных с точки зрения строения, ритмики и просодической модели конструкций может дать важный материал для понимания просодической организации языка.

3. Для выявления полного списка КПС-конструкций можно использовать метод *n*-грамм-анализа. Однако полученные таким образом количественные данные следует считать предварительными (завышенными по абсолютным показателям) и требующими коррекции, для достоверного результата которой необходимо как минимум обращение эксперта к тексту расшифровки, а в отдельных случаях — проведение перцептивного и акустического анализа соответствующего фрагмента звукозаписи.

4. При исследовании КПС-конструкций, состоящих из большого количества элементов, возникает вопрос о том, являются ли они целостно-неделимыми с просодической точки зрения или же представляют собой комбинацию конструкций меньшей размерности (например *да-да-да-да-да* vs. *да-да / да-да-да, да-да-да / да-да, да / да-да-да / да* и т. п.). Поскольку просодическая выделенность может характеризовать более чем один элемент конструкции (например начальный и конечный), в некоторых случаях возникают сложности с определением их ритмической структуры.

5. Для проверки гипотезы о существовании в языке инвариантных просодических моделей необходимо проведение статистически представительного исследования с привлечением инструментального фонетического анализа. Наиболее достоверные результаты следует, по всей видимости, ожидать при сравнении просодических моделей небольших по размеру (например двух- или трехэлементных) изолированных реплик.

Литература

Богданова-Бегларян Н. В., Шерстинова Т. Ю., Кислощук А. И. О ритмообразующей функции дискурсивных единиц // Вестник Пермского университета. Российская и зарубежная филология. 2013. Вып. 2 (22). С. 7–17.

Бойчук Е. И. Разграничение анадиплозиса, эпаналепсиса и редупликации и особенности их функционирования (на материале французского романа XIX в.) // Вестник Тамбовского университета. Сер. Гуманитарные науки. 2014. Вып. 4 (132). С. 105–110.

Гришина Е. А. Повторы *Да* и *Нет* в устном диалоге // М. Л. Ремнева, А. А. Поликарпов (сост.). IV Международный конгресс исследователей русского языка «Русский язык: исторические судьбы и современность». Труды и материалы. М., 2010. С. 225–226.

Копнина Г. А. Геминияция // Эффективное речевое общение (базовые компетенции). Словарь-справочник. Электронное издание / Сибирский федеральный университет; под ред. А. П. Сквородникова. Красноярск, 2014. С. 91–92.

ЦРТ 2016 — ООО «Центр Речевых Технологий», www.speechpro.ru.

Шерстинова Т. Ю. Об изохронности структурных единиц в спонтанной речи (к постановке проблемы) // Материалы XXXIX Международной филологической конференции. 15–19 марта 2010 г. Санкт-Петербург. Вып. 23: Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб.: Филологический ф-т СПбГУ, 2010. С. 109–118.

Шерстинова Т. Ю. 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.

Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD Speech Corpus of Russian Everyday Communication “One Speaker's Day”: Creation of Principles and Annotation // Lecture Notes in Computer Science. Vol. 5729: Text, Speech and Dialogue / V. Matoušek, P. Mautner (eds.). Berlin–Heidelberg, 2009. P. 250–257.

Boersma P., Weenink D. Praat: doing phonetics by computer: [Computer program]. Version 6.0.36, retrieved 11 November 2017 from <http://www.praat.org/>

Bogdanova-Beglarian N., Martynenko G., Sherstinova T. The “One Day of Speech” Corpus: Phonetic and Syntactic Studies of Everyday Spoken Russian // Ronzhin A. et al. (eds.) SPECOM 2015. Lecture Notes in Artificial Intelligence. Springer, 2015. P. 429–437.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Ermolova O., Baeva E., Martynenko G., Ryko A. Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech // Ronzhin A. et al. (eds.) SPECOM 2016. Lecture Notes in Artificial Intelligence. Vol. 9811. Springer, 2016. P. 659–666.

Hellwig B., Van Uytvanck D., Hulsbosch M. et al. ELAN — Linguistic Annotator. Version 5.0.0-alfa [Электронный ресурс] <http://www.mpi.nl/corpus/html/elan/>, last accessed 2017/03/28.

Jurafsky, D., Martin J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson Prentice Hall, 2008. 988 p.

Tatiana Y. Sherstinova

Saint Petersburg State University

National Research University Higher School of Economics

(Russia, Saint Petersburg)

t.sherstinova@spbu.ru; tsherstinova@hse.ru

**‘YEAH, YEAH, YEAH, YEAH, YEAH!’ AND THE OTHER CHAINS
OF REPEATED WORDS IN EVERYDAY RUSSIAN SPEECH
IN THE LIGHT OF SPEECH PROSODY PATTERNS**

Repetition of words in immediate succession is a common phenomenon of Russian everyday speech. Prosodic characteristics of the chains of repeated words like *da, da, da* (*yeah, yeah, yeah*) or *nu, nu, nu* (*well, well, well*) deserve special attention of phoneticians, since it seems that they may help to reveal the basic speech prosody patterns. The research is based on data obtained from the ORD corpus of everyday Russian speech. N-grams analysis was performed for speech transcripts of 900 000 words, and the frequencies of utterances consisting of two and three identical words in a row were obtained, as well as preliminary data on frequencies for the chains of three and five identical words. It turned out that in Russian everyday speech the following words are used in

such repetitive structures most frequently: *da* (yes/yeah), *nu* (well), *vot* (well, here), *tak* (so, well), *shchas* (right now) and some other high-frequency elements of oral discourse, many of which are commonly referred to as discourse markers. The absolute ‘leader’ among them is the particle *da* (yes/yeah). When studying such repetitive structures, it is important to distinguish whether they are integral (inseparable) from the prosodic point of view or whether they are a combination of smaller word chains. Difficulty may arise in determining their rhythmic structures, since prosodic prominence of different kind may be observed on several elements of such word chains (e. g., on initial and final words). To test the hypothesis of the existence of invariant prosodic models, it is necessary to conduct a study on a representative speech data set, and it seems appropriate to start such research by modelling prosody patterns for small-sized (e. g., consisting of two or three words) isolated utterances.

Key words: Russian, everyday speech, spoken speech, repetition of words, repetitive structures, epizeuxis, gemination, particles, discourse markers, n-grams, prosody, prosody patterns, speech corpus

References

Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD Speech Corpus of Russian Everyday Communication “One Speaker’s Day”: Creation of Principles and Annotation. *Lecture Notes in Computer Science*. Vol. 5729: Text, Speech and Dialogue. V. Matoušek and P. Mautner (eds.). Berlin–Heidelberg, 2009. P. 250–257.

Boersma P., Weenink D. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.36, retrieved 11 November 2017 from <http://www.praat.org/>

Bogdanova-Beglarian N., Martynenko G., Sherstinova T. The “One Day of Speech” Corpus: Phonetic and Syntactic Studies of Everyday Spoken Russian. Ronzhin A. et al. (eds.) *SPECOM 2015. Lecture Notes in Artificial Intelligence*. Springer, 2015. P. 429–437.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Ermolova O., Baeva E., Martynenko G., Ryko A. Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech. Ronzhin A. et al. (eds.) *SPECOM 2016. Lecture Notes in Artificial Intelligence*. Vol. 9811. Springer, 2016. P. 659–666.

Bogdanova-Beglarian N., Sherstinova T., Kysloshchuk A. [On rhythm-forming function of discourse markers]. *Vestnik Permskogo Universiteta. Rossiiskaia i zarubezhnaia filologiya*. 2013. Iss. 2 (22). P. 7–17. (In Russ.)

Boychuk E. I. [Distinction of anadiplosis, epanalepsis and reduplication and features of their functioning (based on French novel of 19th century)]. *Vestnik Tambovskogo universiteta. Ser.: Gumanitarnye nauki*. 2014. Iss. 4 (132). P. 105–110. (In Russ.)

Grishina E. A. [Repetition of Da (Yes) and Net (No) in oral dialogue]. *IV Mezhdunarodnyi kongress issledovatelei russkogo iazyka “Russkii iazyk: istoricheskie sudby i sovremennost”* [Proc. of the 4th Int. Congress ‘The Russian language: historical destinies and the present’]. Moscow, 2010. P. 225–226. (In Russ.)

Hellwig B., Van Uytvanck D., Hulsbosch M. et al. *ELAN — Linguistic Annotator*. Version 5.0.0-alfa [Computer program] <http://www.mpi.nl/corpus/html/elan/> (date of access 2017/03/28).

Jurafsky D., Martin J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson Prentice Hall, 2008. 988 p.

Kopnina G. A. [Gemination]. *Effektivnoe rechevoe obshchenie (bazovye kompetentcii)* [Effective speech communication (Basic competences)]. Slovar-spravochnik. Skovorodnikov A. P. (ed.). Krasnoyarsk, 2014. P. 91–92. [Digital document.] (In Russ.)

Sherstinova T. [The most frequent words in everyday spoken Russian (in the gender dimension and depending on communication settings)]. *Komp'yuternaja Lingvistika i Intellektual'nye Tehnologii*. 2016. No. 2 (16). P. 616–631. (In Russ.)

Sherstinova T. [Some Observations on Isochrony in Russian Spontaneous Speech]. *Materialy XXXIX Mezhdunarodnoi filologicheskoi konferentsii. 15–19 marta 2010 g. Sankt-Peterburg. Vyp. 23. Polevaya lingvistika. Integral'noe modelirovanie zvukovoi formy estestvennykh yazykov* [Materials of the 39th International Philological Conference. March 15–19, 2010. St. Petersburg. Iss. 23. Field linguistics. Integral modeling of the sound form of natural languages]. Bogdanova N. (ed.). Saint Petersburg: Saint Petersburg State University, 2010. P. 109–118. (In Russ.)

Speech Technologies Center (STC Group), www.speechpro.ru (date of access 2017/12/12).