



# О повторах дискурсивных слов в повседневной речевой коммуникации (на материале русского языка)

Шерстинова Татьяна Юрьевна

Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург,  
Россия

## Multiple repetition of discourse markers in everyday spoken Russian

Sherstinova Tatiana

Saint Petersburg State University (SPSU), St. Petersburg, Russia

### Аннотация

Распространенным явлением разговорной речи является повторение (амплификация) дискурсивных единиц ("ну-ну", "да-да-да" и т. п.). В докладе приводятся статистические данные о реализации таких форм на материале корпуса ОРД, полученные по результатам анализа списка n-грамм. Анализ повторяющихся форм позволяет исследовать имплицитно существующие в языке ритмические паттерны.

### Abstract

The repetition of discourse particles ("well, well", "yes, yes, yes", etc.) is a special feature of everyday spoken language. The paper presents statistical data concerning the frequencies of such forms in the ORD corpus of Russian everyday speech obtained as a result of n-gram analysis. The study of repetitive

forms allows to explore temporal rhythmic patterns implicitly existing in the language.

**Ключевые слова:** повседневный разговорный русский язык, дискурсивные слова, амплификация.

**Keywords:** spoken everyday Russian, discourse markers, repetitive forms

1. Эмпирические исследования, выполненные на представительном объеме речевого материала, показывают, что значительную долю повседневной речевой коммуникации на русском языке составляют шаблонные формулы речи [Sherstinova 2013]. Наиболее частотные из них состоят преимущественно из неизменяемых слов, в первую очередь частиц, которые в последнее время все чаще относят к категории так называемых «дискурсивных» слов

(например, *да, вот, не, ну, так, всё, ладно* и др.).

В повседневной речи эти компактные единицы, выполняющие роль «служебных слов» речевой коммуникации, часто выступают как независимые высказывания ("*Да!*", "*Вот.*", "*Ну?*", "*Так!*" и т. д.), а также объединяются друг с другом для образования частотных речевых штампов ("*Ну да.*", "*Ну вот.*", "*Вот так.*", "*Да нет.*", "*Вот так вот.*", "*Ну всё.*", "*Ну ладно.*", "*Да ладно?*" и т. п.). В списке наиболее употребительных реплик повседневной речи находятся также высказывания, состоящие из двухкратных и трехкратных повторов дискурсивных единиц: "*Да-да.*", "*Угу-угу.*", "*Да-да-да.*", "*Так-так-так.*" и др. [там же]. Для носителей русского языка "на слуху" и другие, пусть и менее частотные, повторяющиеся формы: напр., "*Ну-ну.*", "*Вот-вот-вот.*", "*Давай-давай!*" и многие другие. Подобные образования будем называть «амплифицированными».

2. Исследование конструкций, состоящих из повторяющихся (амплифицированных) дискурсивных слов, было проведено на базе наиболее представительного на сегодняшний день корпуса повседневной устной речи «Один речевой день» (ОРД корпус) [Asinosvsky et al. 2009] и его социолингвистического расширения [Bogdanova-Beglarian et al., 2016]. Материалом исследования стал подкорпус расшифровок звукозаписей устной речи в объеме 900000 словоупотреблений, в котором представлена речь 100 основных информантов корпуса ОРД (мужчин и женщин разного возраста, представителей разных социальных групп) и нескольких сотен их коммуникантов.

3. Предварительную оценку активности слова в построении

амплифицированных структур возможно осуществить по автоматически получаемому списку *n*-граммов текстов расшифровок.

*N*-граммы, которые представляют собой последовательности из *n* подряд идущих слов, относящихся к одной фразе, активно используются в современной прикладной лингвистике для построения вероятностных лингвистических моделей. В настоящем исследовании для получения списка *n*-грамм использовалось программное обеспечение разработки ООО "Центр Речевых Технологий" ([www.speechpro.ru](http://www.speechpro.ru)), при этом анализировались цепочки слов размером до 5 графических единиц.

Поскольку особый интерес вызывают регулярно встречающиеся амплификации, было решено определить количество повторов для всех слов с наивысшим рангом встречаемости (за исключением предлогов и союзов). Перечень этих слов определялся по верхней зоне частотного списка словоформ, построенного на материале ОРД [Шерстинова 2016].

Оказалось, что практически все высокочастотные единицы из первой сотни наиболее частотных словоформ встретились в речи в повторных формах, как минимум в однократном повторе ("*Да-да*"). Как и следовало ожидать, чем из меньшего количества повторных элементов состоит речевая структура, тем чаще она используется. Так, наибольшее число встретившихся в речи структур — двухэлементные ("*Да-да.*"), за ними по частоте следуют трехэлементные амплификации ("*Да-да-да.*"). Реже встречаются четырехкратные и пятикратные повторы, а также более сложные построения, содержащие еще больше



(на исследованном нами материале — до 7) повторяющихся единиц.

Чаще всего в образовании амплифицированных структур использовалась частица *да* (1360 двукратных повторов, 605 — трехкратных, 209 — четырехкратных, 91 — пятикратных) и две отрицательные частицы — *не* (418 двукратных повторов, 91 — трехкратных, 24 — четырехкратных, 11 — пятикратных) и *нет* (253, 81, 22, 5 соответственно). Распространенность повторов "*да*" и "*нет*" в устной речи нашло свое отражение и в речи кино (см. [Гришина 2010]).

Другими регулярно повторяемыми речевыми единицами являются *вот* (269, 31, 8, 3), *ну* (232, 44, 11, 4), *так* (169, 78, 37, 19), *давай* (137, 63, 34, 18), *все/всё* (131, 26, 5, 3), *что* (98, 11, 4, 1), *щас* (*сейчас*) (65, 25, 9, 2), *там* (61, 14, 5, 4), *ой* (32, 18, 15, 4). Встречаются амплификации и других слов: *конечно* (26, 1, 0, 0), *хорошо* (33, 3, 0, 0), *ладно* (18, 2, 1, 0), *надо* (16, 2, 0, 0) и др.

4. При интерпретации списка *n*-граммов следует иметь в виду, что при его подготовке программа автоматической обработки игнорирует всю "несловарную" информацию, представленную в расшифровке, за исключением начала/конца фразы: удаляет всю синтагматическую разметку, информацию о паузах внутри фразы, в том числе о паузах хезитации, символы паралингвистических явлений и другие пометы лингвистической разметки. В результате, часть полученных *n*-граммных цепочек может относиться не только к интересующим нас конструкциям, имеющим цельную ритмическую структуру, но и принадлежать к разным ритмическим группам (синтагмам) или же образовываться в

результате речевых сбоев, явлений хезитации и ряда других факторов. Поэтому для получения более точных результатов необходим ручной анализ и аудирование выявленных вариантов.

Такой выборочный анализ был проведен по пятикратным повторам самой частотной повторяющейся единицы — частицы "*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 (что равняется 65%) выдержало проверку аудированием на цельность ритмической структуры. Поэтому данные, полученные по *n*-граммам следует считать предварительными.

5. Как выяснилось, повторение дискурсивных слов регулярно встречается в речи представителей разных социальных групп говорящих (мужчин и женщин, разных возрастов и профессий), преимущественно в условиях диалога (бытового или профессионального), а также при разговоре информантов самих с собой.

Использование в речи таких конструкций может объясняться как намеренным желанием говорящих "усилить" производимый высказыванием прагматический эффект (ср. "*Да!*" и "*Да-да-да!*"), 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]. Однако для проверки этой гипотезы необходим инструментальный фонетический анализ анализируемых структур и последующая статистическая интерпретация данных.



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

## Литература

- Богданова-Бегларян Н.В., Шерстинова Т.Ю., Кислюк А.И.* О ритмообразующей функции дискурсивных единиц // Вестник Пермского Университета. Российская и зарубежная филология. 2013. Выпуск 2 (22). С. 7–17.
- Гришина Е. А.* Повторы Да и НЕТ в устном диалоге / IV Межд. конгресс исследователей русского языка «Русский язык: исторические судьбы и современность», филологический ф-т МГУ им. М.В.Ломоносова, 20-23 марта 2010 г.  
[https://www.academia.edu/222240/Повторы\\_Да\\_и\\_НЕТ\\_в\\_устном\\_диалоге](https://www.academia.edu/222240/Повторы_Да_и_НЕТ_в_устном_диалоге) (Дата обращения: 25.06.2016)
- Шерстинова Т. Ю.* Об изохронности структурных единиц в спонтанной речи (к постановке проблемы) // Мат-лы XXXIX Межд. филологической конференции. 15-19 марта 2010 г. Санкт-Петербург. Выпуск 23. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб., 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., and Sherstinova, T.* The ORD Speech Corpus of Russian Everyday Communication "One Speaker's Day": Creation of Principles and Annotation. LNCS 5729. Text, Speech and Dialogue. V.Matoušek and P.Mautner, eds. Berlin Heidelberg: Springer-Verlag, 2009. p. 250-257.
- 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, LNAI, vol. 9811. Springer, Switzerland, 2016, Pp. 659–666.
- Sherstinova T.* Russian Everyday Utterances: the Top Lists and Some Statistics. // Thielemann, Nadine and Peter Kosta (eds.), Approaches to Slavic Interaction. 2013. xi, 318 pp. pp. 105–116.

## References