

ББК 32.965+81.1

А64

Научные редакторы:  
Д. А. Кочаров, П. А. Скрябин

А64 Анализ разговорной русской речи (АР<sup>3</sup>-2021): Труды девятого междисциплинарного семинара – СПб.: Скифия-принт, 2021. —101 с.

ISBN 978-5-98620-528-1

Издание представляет собой сборник докладов, сделанных на заседаниях девятого междисциплинарного семинара «Анализ разговорной русской речи» (АР<sup>3</sup>-2021), проходившего 29–30 июня 2021 года в Санкт-Петербургском государственном университете. Тематика докладов определялась (но не ограничивалась) исследованиями разных аспектов спонтанной речи. Статьи сопровождаются аннотациями на английском языке.

ББК 32.965+81.1

*Издание осуществлено при поддержке  
Special Interest Group on Russian Speech Analysis of  
International Speech Communication Association*

ISBN 978-5-98620-528-1

© Коллектив авторов, 2021

# ПО( )ЧЕМУ? — ПО( )ТОМУ. ЧАСТОТНОСТЬ ОДНОСЛОВНЫХ И НЕОДНОСЛОВНЫХ ИНТЕРПРЕТАЦИЙ ОМОФОНОВ В РУССКОЙ УСТНОЙ РЕЧИ<sup>1</sup>

Владислав Иванович Зубов,  
Елена Игоревна Риехакайнен  
Санкт-Петербургский государственный университет

Омофония — достаточно распространённое в русском языке явление. Как правило, под омофонами имеются в виду отдельные слова, которые совпадают по звучанию, но различаются по морфонологическому составу [Ахманова 2007: 288]. Но в потоке речи омофоничными могут оказаться не только отдельные слова, но и их сочетания. Например, в поэзии известен приём пантограммы, когда строки полностью совпадают по буквенному составу (и, следовательно, ожидается, что будут совпадать и по звучанию), но различаются расстановкой пробелов (*На-ко ней, кучер Нил, / На копейку чернил*) [Бубнов 2002]. Такие примеры можно встретить и в естественной устной речи: см., например, в [Фонетика спонтанной речи 1988: 137] описание того, как словосочетание *Нищий вор*, произнесённое [n'iš'ivó],<sup>2</sup> было ошибочно принято слушающим за слово *ничего*. Обычно интерпретация таких омофоничных отрезков, которые являются следствием случайного, не носящего системный характер совпадения плана выражения слова и комбинации нескольких слов, не вызывает проблем в процессе восприятия естественной речи, поскольку неоднозначность снимается контекстом или благодаря переспросу.

Однако в русском языке есть большая группа омофонов вида «одно слово — комбинация из двух и более орфографических слов», разграничение которых, как кажется, может вызывать затруднения, — это единицы, слитное / раздельное написание которых представляет сложности для носителей языка (*также — так же; итак — и так*). Такие единицы обычно включаются в справочники по русской орфографии. Специалистов же по восприятию устной речи интересует статус таких единиц в ментальном лексиконе слушающего: являются ли неодно-

---

<sup>1</sup> Исследование выполняется при поддержке гранта РФФИ №19-012-00629.

<sup>2</sup> В данном случае использована та же транскрипция, что и в источнике. Во всех остальных примерах в статье транскрипция приводится в символах МФА.

словные сочетания, совпадающие по звучанию с однословными, самостоятельными единицами словаря слушающего наряду с их однословными омофонами и, если да, как происходит их извлечение из лексикона в процессе восприятия речи. Поскольку известно, что в процессе распознавания устной речи активируются не только фонетические репрезентации слов, но и орфографические [Peere, Ziegler 2008], можно предполагать, что однословный и неоднословный вариант интерпретации будут конкурировать в процессе восприятия речи, а тот факт, что у носителей языка возникают сложности с их правильным написанием, по всей видимости, должен свидетельствовать о том, что контекст не всегда позволяет снять неоднозначность. В исследованиях восприятия речи было неоднократно показано, что в случае отсутствия или ограниченности контекста на выбор той или иной интерпретации может влиять частотность единиц [Риехакайнен 2016]. Поэтому для описания того, каким образом происходит выбор интерпретации таких омофонов, необходимо сначала описать, как соотносятся однословные и неоднословные варианты по частотности.

Для решения этой задачи мы, опираясь на справочники по русской орфографии [Лопатин 2009; Розенталь 2011] и на Орфографический академический ресурс «АКАДЕМОС» (<http://orfo.ruslang.ru/>), составили список слов, слитное / раздельное написание которых вызывает сложности у носителей языка. В него вошли преимущественно служебные части речи, а также наречия и сочетания самостоятельных слов с предлогами. В список были включены только слова и словосочетания, для которых ожидается совпадение по звучанию в полном (нередуцированном) произнесении. Следовательно, слова и словосочетания, различающиеся местом ударения (например, *во́время* — *во вре́мя*), в этот список не включались. Все слова из списка были также проверены на наличие в основном подкорпусе Национального корпуса русского языка (<https://ruscorpora.ru/new/>, далее — НКРЯ) начиная с 1956 года, т.е. с года принятия Правил русской орфографии и пунктуации. В итоговый список вошли 204 пары. Список доступен по ссылке: <https://osf.io/bsy2t/>.

Именно для этого списка пар было решено выяснить, как соотносятся по частотности варианты внутри каждой пары. Прежде всего, интерес представляет такое соотношение в устной речи, поскольку при восприятии устной речи слушающему необходимо снимать неоднозначность и выбирать интерпретацию омофона.

Определение частотности единиц в русской устной речи сопряжено с рядом методологических трудностей, потому что на данный момент — пока не стал общедоступным самый большой корпус русской устной речи «Один речевой день» — для этой цели, по-видимому,

можно использовать только устный подкорпус Национального корпуса русского языка (<https://ruscorpora.ru/new/search-spoken.html>), который имеет большой объём, т.е. в нем можно ожидать появления большинства интересующих нас единиц. Предварительная проверка единиц по устному подкорпусу показала, что в нем встречается достаточно много ошибок (например, все примеры сочетания *к\_стати* являются ошибочными написаниями слова *кстати*). Поэтому решено было вычислить частотность по устному подкорпусу со снятой омонимией, выбрав тексты только начиная с 1956 года (205 994 словоупотребления), и сопоставить с результатами, полученными по основному подкорпусу со снятой омонимией за тот же период (4 759 671 словоупотребление). При этом отдельный интерес представляло выявление тех пар омофонов, для которых соотношение по частотности в устном и письменном корпусе окажется разным (т.е. в устном корпусе будет преобладать, например, вариант с раздельным написанием, а в письменном — однословный, и наоборот).

Для 106 омофонов из списка в устном подкорпусе со снятой омонимией не было найдено ни одного примера, для ещё 32 был зафиксирован только один пример. Поэтому далее анализировались только 67 пар, для которых было найдено как минимум два примера (это могли быть примеры как разных употреблений (слитного и раздельного), так и одного). В 52 парах в устном подкорпусе наблюдается преобладание однословного варианта над неоднословным, причём в 34 парах неоднословный вариант вообще не встретился в устном подкорпусе, однословный же может иметь достаточно высокую частотность (например, *наверное* — *на верное, зачем* — *за чем* и др.).

Неоднословный вариант оказался частотнее однословного только в 14 парах. Причём для 6 из них (*также* — *так же, зато* — *за то, при том* — *при том, недаром* — *не даром, заодно* — *за одно, оттого* — *от того*) наблюдается несовпадение результатов, полученных по устному и письменному корпусам: в письменном корпусе однословный вариант частотнее неоднословного. Наибольший интерес представляют те из пар с преобладанием неоднословного варианта, для которых в устном корпусе удалось зафиксировать не единичные примеры, а достаточно большое количество вхождений. Это пары *ввиду* — *в виду, оттого* — *от того, итак* — *и так, зато* — *за то, также* — *так же*. По всей видимости, можно ожидать, что у носителя языка при восприятии соответствующего омофона неоднословный вариант активируется в ментальном лексиконе наряду с однословным или даже быстрее однословного. В целом же подавляющее большинство рассмотренных омофонов при отсутствии снимающего неоднозначность контекста должны интерпретироваться носителем языка в первую очередь как однословные единицы.

При этом необходимо учитывать то, что в естественной устной речи одно и то же слово или словосочетание может произноситься по-разному даже одним и тем же диктором в пределах одного речевого фрагмента [Риехакайнен 2016: 75]. Поэтому, прежде чем переходить к экспериментальному изучению того, как представлены описанные единицы в ментальном лексиконе, необходимо провести корпусное исследование, чтобы выяснить, всегда ли однословная и неоднословная единицы, образующие пару, являются омофонами. Для этого необходимо проанализировать большое количество примеров употребления интересующих нас единиц в русской устной речи. Из существующих общедоступных корпусов устной речи, в которых есть доступ к звуковым файлам, наиболее представительным является Мультимедийный подкорпус Национального корпуса русского языка (<https://ruscorpora.ru/new/search-murco.html>; 5 114 560 словоупотреблений). В нём представлено достаточное количество примеров для всех единиц из пар, в которых, согласно предварительному анализу, неоднословный вариант преобладает над однословным (см. табл. 1; поиск проводился по текстам с 1956 года). Однако в этом подкорпусе нет возможности снять грамматическую омонимию, т.е. не исключены ошибки. Кроме того, в нём отсутствует фонетическая расшифровка, т.е. необходимо провести слуховой и инструментальный анализ всех встретившихся реализаций, что и планируется сделать на следующем этапе нашего исследования. На данный же момент — для получения предварительных результатов — мы обратились к Корпусу русской устной речи (<http://russpeech.spbu.ru/>), который значительно уступает Мультимедийному подкорпусу НКРЯ по объёму (сейчас он включает около 26 тыс. словоупотреблений), но содержит полную фонетическую расшифровку всех текстов. Количество примеров интересующих нас единиц в этом Корпусе представлено в табл. 1.

Таблица 1.

**Количество словоупотреблений единиц в корпусах устной речи**

Однословная единица	Корпус русской устной речи	МУРКО	Неоднословная единица	Корпус русской устной речи	МУРКО
<i>также</i>	18	574	<i>так же</i>	4	719
<i>ввиду</i>	3	32	<i>в виду</i>	2	642
<i>оттого</i>	1	104	<i>от того</i>	2	258
<i>зато</i>	0	395	<i>за то</i>	3	436
<i>итак</i>	4	571	<i>и так</i>	16	2233

Даже то небольшое количество примеров, которые удалось обнаружить в Корпусе русской устной речи, достаточно для того, чтобы отметить некоторые тенденции и сформулировать гипотезы, которые нужно будет проверить на более представительном материале. *Также и так же* практически всегда реализуются одинаково (с точностью до качества конечного безударного гласного, который и в однословном, и в неоднословном варианте может быть представлен одним из следующих звуков — [ɪ], [ɛ] или [ə]). Т.е. в этом случае можно предполагать, что единицы, образующие пару, в большинстве случаев являются омофонами. Для пар же *итак* — *и так* и *ввиду* — *в виду* можно отметить различия в реализациях однословной и неоднословной единиц, которые могут быть связаны с особенностями употребления этих единиц в речи. *И так* во всех 16 встретившихся в корпусе случаях является частью сочетания *и так далее*, чаще всего реализующегося как цельная единица [ɪtag'dalʲe], в то время как во всех четырёх реализациях однословной единицы *итак* на конце произносится глухой согласный [k], и стяжения со следующими словами не происходит. Производный предлог *ввиду* во всех трёх примерах произносится с двумя согласными в начале слова [vvʲ], а в неоднословном варианте, который в обоих встретившихся в корпусе примерах входит в сочетание *иметь в виду*, произносится с выпадением первого [v]. Интересно отметить, что в неоднословном сочетании *в виду* наблюдается тенденция, характерная в целом в русской устной речи для предлога *в*: он подвергается полной количественной редукции (выпадению) чаще всех остальных предлогов (например, в текстах Корпуса русской устной речи он выпал в 110 из 503 примеров). В сочетаниях же согласных [vvʲ] и [vʲvʲ] в начале однословных единиц (например, *ввод*, *ввести* и др.), как правило, не происходит выпадения согласного. Таким образом, получается, что различия в реализациях единиц в паре *ввиду* — *в виду* можно объяснить тем, что в первой проявляется фонетическая тенденция, характерная для отдельных слов, а во второй — для сочетания предлога с последующим словом.

Сделанные наблюдения, хотя и носят на данный момент сугубо предварительный характер, свидетельствуют о том, что тезис об омофоничности единиц, различающихся на письме только наличием / отсутствием пробела, требует проверки на представительном материале. Осуществление такой проверки на материале примеров из Мультимедийного корпуса русского языка станет следующим этапом нашего исследования.

## Список использованной литературы

1. *Ахманова О. С.* Словарь лингвистических терминов. 4-е изд. М.: КомКнига, 2007.
2. *Бубнов А. В.* Палиндромия: от перевертня до пантограммы. Новое литературное обозрение. 2002. Вып. 57. — С. 295–312.
3. *Лопатин В. В.* (ред.) Правила русской орфографии и пунктуации. Полный академический справочник. М., 2009.
4. *Риехакайнен Е. И.* Восприятие русской устной речи: контекст + частотность. СПб.: С-Петербург. гос. ун-т, 2016.
5. *Розенталь Д. Э.* Русский язык. Орфография и пунктуация. М.: Эксмо, 2011.
6. *Фонетика спонтанной речи* / под ред. Н. Д. Светозаровой. Л.: Изд-во Ленинградского ун-та, 1988.
7. *Peere L., Ziegler J. C.* On-line Activation of Orthography in Spoken Word Recognition // Brain Research. 2008. Vol. 1188. — Pp. 132–138.