

# **Автоматическая транскрипция: норма и возможные варианты**

Вера Вячеславовна Евдокимова<sup>1</sup>, Павел Анатольевич Скрелин<sup>2</sup>

<sup>1,2</sup>Санкт-Петербургский государственный университет

<sup>1</sup>v.evdokimova@spbu.ru, <sup>2</sup>p.skrelin@spbu.ru

## **Введение**

Представленная статья связана с интересом Ольги Фёдоровны к автоматической транскрипции текста. Один из авторов, П.А. Скрелин, познакомился с ней в конце 80-х годов прошлого века, когда О.Ф. Кривнова с Н.В. Зиновьевой выступали у нас на кафедре фонетики Ленинградского университета с сообщением на эту тему. Работам над системами автоматической транскрипции текста придавалось настолько большое значение, что в разрабатываемом в то время фонетическом фонде русского языка транскриптор был обязательным четвертым блоком. В Москве в те годы в разработке транскриптора принимала участие О.Ф. Кривнова, у нас — С.Б. Степанова. В качестве основы алгоритмов тогда использовались формализованные правила чтения, поэтому в разговорах и выступлениях Л.В. Бондарко часто называла транскриптор «энциклопедией знаний русского произношения», что не мешало ей планировать «возможность получать сведения о статистических свойствах транскрибируемого материала», а также указывать, что «[с]ущественной исследовательской проблемой является сопоставление знаков автоматического транскриптора с теми реальными звуками, которые появляются в слове или тексте» [Бондарко и др. 1992: 12].

## **Данные и правила, использованные для разработки транскриптора**

«Национальный язык представлен своими разновидностями, имеющими разное значение как для общества, пользующегося этим языком, так и для лингвистов, этот язык изучающих» [Бондарко 1998: 247]. Создание системы автоматической транскрип-

ции является важной исследовательской задачей. Здесь необходимо знать и нормативные варианты произнесения слов, и те, которые не охвачены кодифицированной нормой языка. Таким образом, система автоматической транскрипции, основываясь на правилах кодифицированной нормы литературного языка, должна позволять моделирование и вариантов произношения, которые отражают территориальные и социальные диалекты, просторечие, и разговорные варианты высокочастотной лексики [Бондарко 2001; Вербицкая 2013; Вольская и др. 2005; Зиндер 1979; Кузнецов 1997]. Новые знания о варьировании может дать статистическая обработка больших речевых корпусов, включающих точную сегментацию и транскрипцию речевого материала. Полученная информация о вариативности словоформ, может быть включена в правила с учетом относительных частот каждого отмеченного явления. Использование более одного варианта транскрипции может использоваться в задачах автоматического распознавания и синтеза речи, и в лингвистических исследованиях, предполагающих работу не только с вариантами литературного языка.

В качестве основы последней версии системы автоматической транскрипции использовался набор правил, основанный на результатах обширных фонетических исследований, которые проводились на Кафедре фонетики и методики преподавания СПбГУ с конца прошлого столетия и описаны в работах Л.В. Бондарко [Бондарко 2001; Бондарко и др. 1988, 1993, 1997], Л.А. Вербицкой [Вербицкая 2013], П.А. Скрелина [Скрелин 1999, Skrelin 1997, Skrelin et al. 1998], Е.В. Сомовой [Скрелин, Сомова 2000], Н.В. Богдановой [Бондарко и др. 1988, Богданова 2009] и других. Основной перечень правил был сформулирован в работах С.Б. Степановой [Степанова 1988] и К.Б. Шалоновой [Бондарко и др. 1997, Skrelin et al. 1998, Shalnova 1997, 1999].

К этой основе были добавлены правила и статистические данные, полученные в ходе исследований XXI века.

В рамках проекта СПбГУ «Фонетическая вариативность звуковых единиц в аспекте взаимодействия уровней системы современного русского литературного языка» (2010–2014) проводилось фундаментальное научное исследование с целью развития лингвистической теории в части описания условий и пределов фонетической вариативности звуковых единиц в результате взаимодействия и взаимовлияния разных уровней языковой системы в разных видах устной речи.

Для достижения этой цели решались следующие задачи:

— Создание модели фонетического и фонематического варьирования состава морфем в зависимости от позиции в интонационной единице (по отношению к интонационному центру, логическому ударению, интонационной периферии, выделенности) и в зависимости от принадлежности к разным типам морфем или лексем;

— Создание словаря фонетических и фонематических вариантов морфологических и лексических единиц в разных условиях межуровневых взаимодействий;

— Создание списка релевантных фонетических признаков количественной и качественной редукции и включение их в единую мультипараметрическую систему с иерархической организацией;

— Разработка автоматизированных процедур обработки речевого материала.

Исследование проводилось на материале двух крупных речевых корпусов, созданных на кафедре фонетики СПбГУ: корпуса спонтанной речи и чтения INTAS [Bondarko et al. 2003] и корпуса профессионального чтения CORPRES [Skrelin et al. 2010].

Корпус профессиональной русской речи CORPRES содержит записи чтения текстов различных стилей 8 профессиональными дикторами. Корпус содержит разметку на 6 уровнях: 1) границы периодов основной частоты; 2) указание границ важных переходных процессов; 3) «акустическая» (реальная) транскрипция, которая выполнялась в условиях десемантизированного предъявления звукового материала и с учетом его акустических характеристик; 4) идеальная транскрипция, в соответствии с орфоэпической нормой; 5) орфографическая расшифровка в виде последовательности слов, с указанием логического ударения и случаев смещения синтагматического ударения (т.е. случаи, когда синтагматическое ударение оказывается не на последнем слове в синтагме); 6) интонационная транскрипция (указание границ синтагм/пауз, и, соответственно, типа интонационной модели / типа паузы). Таким образом, используя данные уровней акустической и идеальной транскрипции, а также уровня слов, можно автоматически выявить все случаи несовпадений акустической и идеальной транскрипций для каждого слова.

В результате исследования был создан словарь фонетических и фонематических вариантов морфологических и лексических единиц в разных условиях межуровневых взаимодействий, а также были установлены закономерности их появления в разных типах речи.

Также исследовалась вариативность фонетических характеристик гласных в пределах четырех основных морфемных структур русского языка: приставок, корней, суффиксов и окончаний.

Материал, отобранный из корпуса CORPRES, изначально сегментировался на морфемы автоматически посредством автоматического перенесения границ морфем, указанных в «Словаре морфем русского языка» А.И. Кузнецовой и Т.Ф. Ефремовой [Кузнецова, Ефремова 1986] (использовался оцифрованный вариант словаря), на орфографическую запись художественных текстов (см. подробнее о методике [Садуртинова 2012]).

Файлы аннотации подверглись обработке с помощью специального скрипта, предназначенного для поиска несоответствий аннотации на уровнях акустической (реальной) и идеальной транскрипции (то есть вариативности либо отклонений от нормы в произношении) и для определения ударности/безударности слога, в котором найдено несоответствие. Результаты обработки аннотационных файлов представляются в виде таблиц, что является удобным для дальнейшей статистической обработки данных в программе Microsoft Office Excel. Результаты исследования представляют собой статистические данные, полученные в результате обработки аннотированного материала скриптом и подсчета результатов обработки в программе Microsoft Excel. Пример структуры таких файлов представлен таблице 1.

*Таблица 1.* Пример представления информации о вариативности морфемы: приведены варианты транскрибирования морфемы по орфоэпическим правилам и варианты реального произнесения морфем, найденные в корпусе

(цифры указывают на степень редукции гласного:

1, 2 — предударные позиции, 0 — ударная, 8 — заударная).

Морфема в орфографической записи (*корень, + суффикс, ~приставка/постфикс, = соединительный элемент)	Варианты транскрипции морфемы по орфоэпическим правилам	Варианты реального произнесения морфемы	Количество реализаций в корпусе
*алекс			
	*a1 <sup>l</sup> i1ks <sup>j</sup>		
		*a1 <sup>l</sup> e8ks <sup>j</sup>	6
		*a1 <sup>l</sup> iks <sup>j</sup>	58
	*a1 <sup>l</sup> e0ks <sup>j</sup>	*e8 <sup>l</sup> iks <sup>j</sup>	14
		*a1 <sup>l</sup> iks <sup>j</sup>	4
		*a1 <sup>l</sup> e0ks <sup>j</sup>	8
	*a2 <sup>l</sup> i1ks <sup>j</sup>		

В ходе анализа данных о длительности гласных в зависимости от фразовой позиции и длины слова была обнаружена интересная тенденция: в немаркированных словах (не несущих фразового или логического ударения и не являющихся первыми в синтагме) при увеличении длины слова длительность безударных гласных уменьшается, а длительность ударных гласных возрастает. В маркированных же позициях длительность как заударных, так и ударных гласных, одинаково уменьшается с ростом длины слова. В ходе анализа длинных слов (5 и более слогов), встретившихся в нейтральной фразовой позиции, оказалось, что в подавляющем большинстве случаев ударение падает на корень. Таким образом, вполне возможно, что в длинных словах говорящий стремится выделить часть слова с наибольшей семантической нагрузкой. Это говорит о том, что длинные слова, не несущие синтагматического или логического ударения, все же обладают определенной выделенностью.

Было установлено, что количество замен и выпадений гласных зависит от:

- положения гласной фонемы в слове относительно ударения: в заударных комплексах число выпадений/замен выше;
- количества фонем в составе заударного или предупредительного комплекса;
- маркированности слова в синтагме: количество замен и выпадений гласных в словах в нейтральной фразовой позиции выше, чем в маркированной.

Таким образом, полученная информация о вариативности фонетических единиц в зависимости от разных факторов может быть использована при поиске «незаметных» вариантов нормы и учтена в правилах работы транскриптора [Скрелин и др. 2022].

Л.В. Бондарко отмечала, что в связной речи в предупредительной части слова происходит «своеобразная губная гармония гласных», приводящая к огублению безударного неогубленного гласного, если следующий за ним безударный гласный является огубленным [Бондарко 1998: 266]. Например, часто слово «голубому» произносится как [gulubomu] вместо [gɐlubomu]. Это засвидетельствовано в орфоэпическом словаре, изданном под ред. М.Л. Каленчук, Л.Л. Касаткина и Р.Ф. Касаткиной [Каленчук и др. 2012: 941–942]. Там же отмечено и обратное явление, т.е. возможная потеря огубленности безударными /у/ в некоторых условиях, как, например, при произнесении [ʒɨrnalʲist] вместо [ʒurnalʲist] в слове «журналист».

В исследовании, направленном на изучение артикуляционной базы подобных случаев, было подтверждено, что в основе этого явления лежит регрессивная дистантная ассимиляция гласных [Kocharov, Menshikova 2019; Кочаров, Кочеткова 2020].

Анализ экспериментальных данных подтвердил, что можно ожидать огубленность в случае, если гласный находится во втором предударном или заударном слоге, а в следующем слоге этого слова находится безударный [u], иницирующий огубленность. В этих случаях происходит коартикуляция гласных в соседних слогах, когда гласный, предшествующий гласному [u] уподобляется ему по огубленности. Оценка частотности возникновения огубленности и ее потери на указанных выше корпусах и спонтанной диалоговой речи [Качковская и др. 2017] показала, что частотность этих явлений в чтении 1,5–2 раза ниже, чем в спонтанной речи. Анализ влияния скорости чтения специального экспериментального материала показал, что при быстром чтении эти явления происходят чаще, чем при медленном. Поэтому можно предположить, что на частотность этих явлений влияет не столько стиль речи, сколько скорость произнесения. Изменение гласных по наличию или отсутствию огубленности происходят довольно часто, но не всегда, поэтому в системах автоматической транскрипции или произносительных словарях могут применяться вероятностные модели этого процесса, основанные на представленной выше статистике.

Исследование произнесения консонантных сочетаний проводилось на словах и словосочетаниях, имеющих в своем составе последовательность из трех и более согласных фонем [Скрелин, Сомова 2000, 2001]. Такой отбор производился с учетом идеальной фонетической транскрипции, но, кроме этого, к исследованию привлекались и слова, в которых многокомпонентное сочетание согласных не предусмотрено нормативной реализацией, т.е. такие случаи, когда сочетание согласных образовывалось за счет выпадения гласного (или гласных): *представляешь* [pr<sup>1</sup>stavl<sup>1</sup>aeʃ], *циферки* [tsifr<sup>1</sup>k'e], *запоминаются* [zpm<sup>1</sup>inajtsa]. Большую часть материала составляют не отдельные слова, а их сочетания, т.к. скопления согласных на стыках слов — явление достаточно распространенное для русского языка: *бывает спланировать*, *значит в затылке*, *театр Виктюка*. Таким образом были отобраны 945 слов и словосочетаний, из которых 494 взяты из спонтанных тек-

стов и 451 — из прочитанных. Общее количество согласных, подвергнутых обработке, — 2895, из них ненормативные реализации были зафиксированы у 499 согласных, что составляет 17,2% от общего числа исследуемых звуков. Вероятность появления модифицированных согласных в спонтанной речи (далее СР) и в чтении приблизительно одинаковая — 18% для спонтанных текстов и 16% для прочитанных, что в абсолютных значениях составляет 280 ненормативных реализаций согласных в составе консонантных сочетаний для спонтанных текстов и 219 — для прочитанных.

Общие результаты исследования представлены в таблице 2 (все цифры приведены в процентном отношении к числу зафиксированных модификаций у разных групп согласных).

Таблица 2. Влияние дифференциальных признаков (ДП) согласного на степень его устойчивости в спонтанной речи (СР) и в чтении

Тип согласного	Модификаций в СР (%)	Модификаций в чтении (%)
Смычные	18,6	16,6
Щелевые	20,1	17,1
Дрожащие	8,7	7,9
Среднеязычный /j/	57	66
Переднеязычные	20	16,5
Губные	15,8	15,5
Заднеязычные	10	13,6
Твердые	14	12,5
Мягкие	30	26,7
Глухие	16	14,6
Звонкие	28,5	22
Шумные	19,5	16,6
Сонорные	15	15

В таблице 3 приведены результаты анализа статистических характеристик вероятности появления той или иной модификации для отдельных согласных или для разных групп согласных (все значения указаны в процентах от общего числа зафиксированных ненормативных реализаций в СР и в чтении).

Таблица 3. Частота встречаемости той или иной модификации для разных групп согласных

Виды модификаций	Спонтанные тексты (%)	Чтение (%)
Эллипсис согласных	50,7	44,7
Спирантизация смычных согласных	16,8	11,8
Ненормативное оглушение звонких согласных	9	9
Ненормативное смягчение твердых согласных	6,5	5,8
Ненормативное озвончение глухих согласных	5,4	8,2
Замены	4,3	6,4
Вокализация сонорных согласных и /v/, /v <sup>l</sup> /	1,9	7,3
Ненормативная твердость мягких согласных	0,7	2,7
Другие модификации (одновременная реализация 2 согласных, эмфатическое произнесение и удлинение согласных)	4,7	4,1

Самым универсальным изменением, свойственным согласным всех способов образования, является эллипсис. Традиционно под фонетическим эллипсисом понимается «факультативная потеря в разговорной речи части звуковых сегментов или признаков, содержащихся в кодифицированном фонетическом тексте» [Кодзасов 1973: 109]. Он оказывается самой частотной модификацией как в СР, так и в чтении и составляет 50,7% от всех зафиксированных ненормативных реализаций в спонтанных текстах и 44,7% в прочитанных. Несмотря на то, что полное выпадение звука свойственно всем типам согласных, оно затрагивает согласные разного способа образования с неодинаковой степенью вероятности. Кроме этого, частота встречаемости сочетаний с эллиптированными согласными в СР и в чтении оказывается неодинаковой.

Данные по количеству выпадений в каждой группе согласных в СР и в чтении приведены в таблице 4 (цифры приведены в процентном отношении к общему числу выпадений, зафиксированных в спонтанных текстах и в прочитанных)

Таблица 4. Влияние способа образования согласного на частоту его выпадения

Тип согласного	Отмечено выпадений в СР (%)	Отмечено выпадений в чтении (%)
Смычные	55	58
Щелевые	42	38
Дрожачие	3	4



Как видно из таблицы 4, показатели, зафиксированные для СР и для чтения не сильно отличаются друг от друга. Большая часть всех согласных, подвергшихся эллипсису — смычные, что характерно как для спонтанных, так и для прочитанных текстов. Второе место занимают щелевые согласные, и последнее — дрожащие.

Интересные данные были получены по ассимиляции по глухости/звонкости.

Ненормативное оглушение звонких согласных с равной степенью вероятности может встречаться как в СР, так и в чтении. Более последовательно это изменение затрагивает смычные и щелевые согласные, реже — сонанты; 80% согласных, подвергшихся оглушению, — твердые (данные одинаково справедливы для спонтанных и для прочитанных текстов). Однако причины и характер оглушения согласных могут быть разными.

Первая группа сочетаний слов с оглушенными согласными, в идеальной фонемной транскрипции должна была представлять следующие последовательности: /zdr/, /zdʋ/, /zd/, т.е. содержать консонантный комплекс /zd/ перед сонорным или /v/, или перед гласным. Фраз с такими сочетаниями было всего 22, из которых в 11 случаях наблюдалось частичное или полное оглушение /d/, а иногда и /z/. Причем тенденция к оглушению /d/ чаще наблюдается в позиции перед сонорным и /v/, в то время как, положение перед гласным способствует нормативной реализации. В итоге, в ряде случаев, наблюдались замены комплекса звонких согласных на соответствующие глухие.

Следующую группу составляют случаи оглушения звонких согласных в позиции после глухих, т.е. по принципу прогрессивной ассимиляции. В исследуемом материале было зафиксировано 17 таких реализаций: 7 — в спонтанных текстах и 10 — в прочитанных. Например, реализация словосочетания *передать в ГАИ*, в норме должна соответствовать транскрипционной записи /pʲerʲe-zdavɑtʲvɡɑi/, т.е. в соответствии с правилами регрессивной ассимиляции по звонкости, характерной для русского языка; в нашем случае это сочетание было произнесено как [pʲerʲe-zdavɑtʲfkɑi].

Ассимиляции по глухости не препятствует и изменение слогового состава слова, приводящее к стиранию словесной границы между элементами словосочетания за счет выпадения заударных гласных и /j/. Так произошло в сочетании *сессия закончится*, которое было произнесено диктором как [sesʲ:sakontʲitsɑ].

Вокализация согласных — широко распространенное явление, которое характерно как для кодифицированного литературного языка, так и для разговорной речи и затрагивает сонорные согласные, а также /v, v<sup>j</sup>/. Результаты исследования показали, что согласные в чтении более чем в три раза чаще подвергаются вокализации по сравнению с согласными СР. Из всех вокализованных согласных 38% приходится на долю плавных /l, l<sup>j</sup>/ и дрожащих /r, r<sup>j</sup>/; 33% — на долю /v, v<sup>j</sup>/ и 29% составляют носовые /m, m<sup>j</sup>/ и /n, n<sup>j</sup>/. По результатам настоящего исследования можно предположить, что твердые согласные более склонны к вокализации, чем мягкие, т.к. в анализируемом материале 67% вокализаций составляют именно они, а мягкие, соответственно, 33%.

Результаты исследования показали, что согласные, входящие в состав консонантных комплексов, с большей долей вероятности склонны к появлению мягких вариантов твердых согласных, чем к «отвердению» мягких<sup>1</sup>.

В большинстве типичных для русского языка сочетаний согласных разграничение их компонентов и отождествление их с отдельными фонемами не составляет большого труда, однако существуют сочетания глухих щелевых, в которых первым элементом является /f/, где зачастую наблюдается «смешение спектральных характеристик и невозможность членения» [Вербицкая, Зиндер 1969: 48]. Так, например, в исследуемом материале встретились случаи произнесения консонантных комплексов, при котором происходила одновременная реализация 2 согласных.

Эти случаи касаются в первую очередь сочетаний глухой щелевой + /f/. Известно, что подобные сочетания способны реализовываться таким образом, что оба звука, на каком-то участке своего звучания, могут артикулироваться одновременно, а особенности артикуляции фонем /v, f/ и их мягких пар, могут не препятствовать коартикуляции окружающих их согласных [Скрелин 1999: 37]. С помощью слухового и акустического анализа было установлено, что наличие между некоторыми согласными фонемами /f/ для пар глухих согласных и /v/ — для пар звонких, может и не мешать необходимым чередованиям, т.е. наблюдались случаи

---

<sup>1</sup> Эти данные идут вразрез с полученными ранее, которые свидетельствуют о том, что «полумягкие» согласные встречаются в РР в десять раз чаще, чем «полутвердые» (см., например, [Светозарова (ред.) 1988: 50]).

полной ассимиляции с сохранением фонемы /f/: *случилось в штате*. Артикуляторная природа такого процесса ясна и не вызывает никаких недоумений: губной уклад щелевых фонем /v, v<sup>j</sup>, f, f<sup>j</sup>/ не препятствует артикуляции (и коартикуляции) язычных щелевых, однако интересным представляется то, как подобные сочетания будут восприниматься на слух и как слушающий интерпретирует такое звучание с точки зрения фонемной принадлежности. Кроме этого, в некоторых случаях наблюдается условно называемый «эффект присутствия» исчезнувшей фонемы, который проявляется в том, что чередования не происходит, а след губной фонемы остается: *живут в Чечне*. А часто и следа выпавшей фонемы не остается: *будет в четыре*.

Для исследования влияния на реализацию подобных сложных сочетаний шумных согласных разного темпа речи был разработан специальный текст.

В качестве материала были использованы аудиозаписи голосов четырех дикторов, одного мужчины и трёх женщин. Дикторы читали заранее составленный текст, включавший фонетически сложные сочетания согласных на стыках слов и морфем.

Дикторы читали текст несколько раз и с разной скоростью. Скорость произнесения текста определялась на слух и сравнением длительности периодов чтения в звуковых файлах. Таким образом, для каждого диктора были отобраны по три записи, которые представляли разные темпы речи. Аудиозаписи были маркированы метками, обозначающими скорость («медленно», «нейтрально», «быстро», соответственно). По времени записи длились в диапазоне от 1,49 минут до 3 минут.

Для детального изучения были выделены 17 слов и словосочетаний, 13 из которых воспроизводились в связном тексте, а остальные 4 проговаривались в составе отдельных коротких предложений. Искомые сложные сочетания включали в себя от двух до восьми орфографических знаков, записанных непрерывно, в составе слов, и с пробелами, как части нескольких лексических единиц и целые единицы.

Словосочетания были сгруппированы по схожести состава кластеров согласных и процессов, которые модифицируют их в потоке речи. Основой анализа стала идеальная транскрипция, позволяющая выявить изменения и определить их природу. Рассматривались варианты реализаций данных сочетаний у разных дикторов при разном темпе речи (таблица 5).

Таблица 5. Идеальная транскрипция и зарегистрированные варианты произнесения выделенных сочетаний согласных в разных темпах речи

	Идеальная	Нормальный темп	Быстрый темп	Медленный темп
автобус в Житомир	z v ʒ	z ʒ ʒ ʒ v ʒ s v ʒ	v ʒ z ʒ ʒ ʒ	z v v ʒ z ʒ ʒ ʒ
поджарый	d ʒ	dʒ	dʒ	dʒ
подшёрсток	t ʃ	tʃ ʃʃ tʃʃ	tʃ ʃʃ tʃʃ	tʃ ʃʃ tʃʃ
раз в четыре	s f tʃ <sup>i</sup>	s f tʃ s f <sup>i</sup> s tʃ <sup>i</sup>	s tʃ <sup>i</sup> ʃ <sup>i</sup> : s	s f tʃ <sup>i</sup> s tʃ <sup>i</sup> tʃ <sup>i</sup> tʃ <sup>i</sup>
текст Чехова	kst tʃ <sup>i</sup>	x ʃ <sup>i</sup> ʃ <sup>i</sup> x s tʃ <sup>i</sup>	k s tʃ <sup>i</sup> x tʃ <sup>i</sup> x s tʃ <sup>i</sup>	k s tʃ <sup>i</sup> x ʃ <sup>i</sup>
конец четверти	c tʃ <sup>i</sup>	c tʃ <sup>i</sup> s tʃ <sup>i</sup> tʃ <sup>i</sup> tʃ <sup>i</sup>	c tʃ <sup>i</sup> tʃ <sup>i</sup> tʃ <sup>i</sup>	c tʃ <sup>i</sup> tʃ <sup>i</sup> tʃ <sup>i</sup> ʃ <sup>i</sup> : ʃ <sup>i</sup> :
задержавшись в школе	f ʃ y s <sup>i</sup> f ʃ k	s <sup>i</sup> f ʃ	ʃ ʃ	s <sup>i</sup> ʃ ʃ ʃ
забившись в щель	f ʃ y s <sup>i</sup> f ʃ <sup>i</sup> :	ʃ <sup>i</sup> f ʃ <sup>i</sup> s <sup>i</sup> ʃ <sup>i</sup> s <sup>i</sup> ʃ <sup>i</sup> ʃ <sup>i</sup>	s <sup>i</sup> f ʃ <sup>i</sup> : ʃ <sup>i</sup> ʃ <sup>i</sup> ʃ ʃ	ʃ <sup>i</sup> ʃ <sup>i</sup> : ʃ <sup>i</sup> ʃ <sup>i</sup> s <sup>i</sup> ʃ <sup>i</sup> ʃ <sup>i</sup> : ʃ <sup>i</sup> : ʃ <sup>i</sup> :
весь в шелках	s <sup>i</sup> f ʃ	ʃ <sup>i</sup> ʃ	s <sup>i</sup> ʃ ʃ <sup>i</sup> ʃ	s <sup>i</sup> ʃ <sup>i</sup> ʃ s <sup>i</sup> ʃ

Оказалось, что реализации не всех сложных сочетаний шумных согласных зависят от темпа речи. Некоторые из них обнаруживают ассимиляции, которые свойственны речи определенных дикторов, либо относительно постоянны. Тем не менее, изменения некоторых кластеров имеют закономерности, связанные со скоростью чтения. Кроме того, стыки шумных смычных и щелевых согласных в большинстве случаев реализованы аллофонами фонемы /tʃ<sup>i</sup>/, что позволяет предложить их использование в фонематической транскрипции, для уточнения характера этих звуков и повышения качества транскрипции и её производных.

Фонемы, расположенные в конце стечения, перед гласными, реже подвергаются ассимиляции. Точное определение положения и акустических характеристик данных сегментов позволит лучше определять границы сложных консонантных кластеров, и, следовательно, более детально изучить их природу.

Формирование групп сложных стечений согласных позволяет рассматривать происходящие в них процессы комплексно, поэтому такой подход будет полезен при упорядочивании правил и исключений для алгоритмов автоматического транскриптора.

### **Свойства транскриптора**

В разработанной версии автоматического транскриптора русского текста алгоритмы строятся не только на нормативных предписаниях, но и позволяют использовать зарегистрированные допустимые варианты.

Степень необходимой подробности и вариативности обработки транскриптором текста определяется поставленной задачей. Полученная транскрипция может включать в себя только фонематическую интерпретацию звуковых единиц, использовать основные правила коартикуляции, ассимиляции и фонемных чередований. Транскрипция может быть и фонетической, использующей знания о происходящих в речевой деятельности фонетических процессах и их результатах. Разработанный автоматический транскриптор моделирует не только орфоэпическую норму, но и произносительные варианты, систематически наблюдаемые в русской речи. При моделировании фонемного состава речевой последовательности автоматический транскриптор представляет идеальную фонемную последовательность с обозначением некоторых особенностей аллофонного варьирования в тех случаях, когда это необходимо. Программа обрабатывает связный текст, используя правила коартикуляции на стыках слов и морфем, учитывает варианты произношения, свойственные разным формам речи и наличие сегментов как полного, так и неполного типа произнесения.

Программа включает в себя несколько принципиальных программных блоков. Важной особенностью программы является то, что она работает с классами звуков, используя их дифференциальные и полезные признаки, которые и определяют характер их взаимодействия друг с другом.

На первом этапе программа проводит предварительную обработку полученного текста. Убирает все лишние символы и сохраняет все те, которые необходимо исключить в процессе транскрипции, но вернуть в текст после обработки. Так как транскриптору требуется предварительная информация об ударности/безударности гласных, то такая информация должна быть уже в тексте. Также транскриптор может использовать имеющиеся сведения о просодической разметке корпуса, такие как конец синтагмы, пауза hesitation, вдох, речевая запинка и т.д. В соответствии с этой информацией происходит обработка стыков слов, абсолютного конца фразы и учет межсловной ассимиляции или чередования звуков.

На следующем этапе происходит процесс трансформации букв в обозначения фонем и аллофонов. Учитываются возможные исключения и сочетания звуков внутри слова и на границах морфем, внутрисловная ассимиляция согласных по разным признакам, а также вариативность гласных, в зависимости от приписанной ударности, места в слове и вариативности. На этом же этапе происходит учет возможных границ фонетического слова и обработка фонетических явлений, которые происходят между лексемами и фонетическими словами. В зависимости от просодической разметки может происходить ассимиляция согласных на стыке слов по мягкости-твердости, звонкости-глухости, по фокусу (одно-двухфокусные согласные).

В алгоритм программы были включены правила для учета сложных случаев ассимиляции звуков. К таким можно отнести рассмотренные выше фонетические процессы, происходящие на стыке слов с предлогами. Например, «раз в жизни» обычно транскрибируется как /ra0z v z'i0z'n'i/ (0 — обозначает ударность гласного в транскрипции). Процессы ассимиляции по фокусу артикуляции и месту образования приводят к тому, что такое сочетание в связной речи может звучать и как «ra0z v z'i0z'n'i».

В алгоритме учтены возможности образования фонетического слова (в орфографической разметке обозначены ударные гласные, а также дополнительное ударение). Также происходит учет возможных случаев клитизации, например, сочетания с предлогами. Например, сочетание двух знаменательных слов «снег небольшой» предполагает чередование последнего согласного первого слова с глухим парным, а сочетание с предлогом «через минуту»

такого чередования не допускает, принимаются во внимание случаи типа «через окно» и «сквозь окно».

Также в программу был введен целый ряд исключений из правил произношения в русском языке. Например, слово «интересный» читается по правилам и транскрибируется как /in<sup>t</sup>er<sup>e</sup>0snij/ или /in<sup>t</sup>ir<sup>e</sup>0snij/. При этом слово «интернет» является исключением и предполагает варианты фонематической транскрипции /internet/ и /int<sup>r</sup>net/. Ниже представлен пример транскрипции.

ta0k vo0t k<sup>i</sup>ri<sup>i</sup>0ll / palut<sup>f</sup>a0ica fto p<sup>e</sup>0rvij marfru0d gavar<sup>i</sup>u0  
va0m ja0  
da0 u m<sup>i</sup>in<sup>a</sup>0 z<sup>d</sup>e0s<sup>i</sup> f<sup>s</sup>o0 rasp<sup>i</sup>0sana / dava0jt<sup>i</sup> tagda0 sv<sup>e</sup>0r<sup>i</sup>im  
laka0cii/ э- dava0jt<sup>i</sup> papro0buim e0s\* / sl<sup>e</sup>0va napra0va / э- /  
vo0t / э- / sl<sup>e</sup>0va napra0va p<sup>e</sup>0rvai u m<sup>i</sup>in<sup>a</sup>0 raspalaga0ica e0ta  
mavza<sup>l</sup>e0j

В результате описанных выше исследований в транскриптор была добавлена информация о наиболее вероятных неорфоэпических фонетических явлениях, искажающих фонетическую транскрипцию. Таким образом, удалось включить в программу информацию о некоторых пределах фонетической вариативности звуковых единиц в результате взаимодействия и взаимовлияния разных уровней языковой системы в разных видах устной речи. В ходе работы появилась возможность учитывать некоторые случаи несовпадений реальной и идеальной транскрипций для слова и учитывать данные о частотных заменах разных гласных фонем в ударных и заударных слогах, замен и выпадений согласных в чтении, учесть в транскрипции возможность замены /i/ на /e/, рассматривать возможную замену огубленных неогубленными и наоборот, в том числе в позициях, где по правилам русского нормативного произношения никаких замен быть не должно. Эту информацию автоматический транскриптор может учитывать, порождая несколько вариантов транскрипции.

Для учета информации о частотных выпадениях, модификациях и вставках звуков использовался разработанный нами словарь фонетических вариантов морфологических единиц, который показывает разные зафиксированные варианты транскрипции

конкретной лексемы, а транскриптор может использовать только те варианты, которые полностью согласуются с фонетическими явлениями, возможными в окружающем контексте (ниже представлен пример транскрипции).

V naʃ [11]vʲek / 9 / dva0c:atʲ [10]pʲervij / e0ta [10]vazmo0zna/  
to0 jisʲtʲ 9 a0= [10]prʲidumal / nʲeskaʲlʲka [11]varʲia0ntaf  
(virʲia0ntaf(8), verʲea0ntaf (3)) / adʲi0n is [11]varʲia0ntaf  
(virʲia0ntaf(8), verʲea0ntaf (3)) /e0ta aptʲa0gʲivatʲ 9 ə- ʃin\* /  
[+ ]nu0 / abrʲe0zak ə- ʃi0ni aftamabʲi0ʲnʲaj vakru0k [11]stvala0/

В программу добавлена информация о возможных гласных вставках (из работ О.П. Агапкиной и К.В. Евграфовой, см., например, [Агапкина 1982; Evgrafova 2009]). Обычно гласные вставки появляются в консонантных кластерах. Они не имеют фонологического статуса, однако обладают собственными количественными и качественными характеристиками. Чаще всего гласные вставки появляются в сочетании сонанта с другим согласным. Это позволяет учесть гласную вставку уже как возможную на уровне идеальной транскрипции и уменьшить таким образом ошибки процедуры интерпретации речевого сигнала.

С помощью нашего транскриптора был получен уровень фонетической транскрипции в корпусе русской спонтанной речи CoRuSS [Kachkovskaia et al. 2016], который включает студийные записи спонтанных диалогов, снабженные орфографической расшифровкой и просодической аннотацией, а также корпуса Sibling [Kachkovskaia et al. 2020]. Отдельный уровень аннотации содержит фонетическую транскрипцию отобранного материала. Так же, как и орфографическая расшифровка, данная фонетическая транскрипция может считаться в определенной степени идеальной, так как отталкивается от орфографической расшифровки и не полностью отражает реальное произнесение. Однако полученная транскрипция учитывает просодическую разметку и таким образом моделирует явления ассимиляции по различным дифференциальным признакам или ее отсутствие для звуков на границах интонационных единиц, а также в абсолютном начале и конце фразы. В транскрипции учтены также речевые сбои, вдохи, паузы, отмеченные по звуковому сигналу.



В дальнейшем автоматический транскриптор будет улучшен и сможет предлагать наиболее вероятные варианты произнесения слов, учитывая процессы, происходящие в слитной речи. Это является необходимым требованием, поскольку именно неточности в транскрипции вызывают сбои в обработке спонтанной речи.

На данном этапе программа не учитывает специфическое произношение географических названий, аббревиатур или имен собственных. Присоединение такого словаря может быть следующим этапом работы. Первый словарь такого рода (с вариантами транскрипции) «Топонимы Санкт-Петербурга» уже нами создан.

## Заключение

Представленные в статье данные последних исследований разных типов и стилей русской речи показывают широкий спектр зафиксированных, в том числе и «незаметных» для слушателя, отклонений произношения слов от предписанных орфоэпическими словарями последовательностей фонем. Представленные результаты исследований отвечают идее Л.А. Вербицкой о возможности описания таких вариантов в орфоэпическом словаре, предложены способы моделирования подобных фонемных и звуковых последовательностей в разных ситуациях и условиях речевой деятельности с помощью автоматического транскриптора.

## Литература

- Агапкина О.П. Фонетические характеристики гласной вставки в современном русском литературном языке. Автореф. ... канд. дисс. Л., 1982.
- Богданова Н.В. Произношение и транскрипция: учебно-методическое пособие. СПб.: Факультет филологии и искусств СПбГУ, 2009.
- Бондарко Л.В. Фонетическое описание языка и фонологическое описание речи. Л.: ЛГУ. 1981.
- Бондарко Л.В. Фонетика современного русского языка. СПб.: СПбГУ, 1998.
- Бондарко Л.В. Спонтанная речь и организация системы языка // Бюллетень фонетического фонда русского языка. Фонетические свойства русской спонтанной речи. 2001, № 8. СПб. – Бохум. С. 17–23.
- Бондарко Л.В., Богданова Н.В., Овчаренко Е.Б., Рыжов И.В., Степанова С.Б. Автоматическая транскрипция фонетического фонда русского языка // Бюллетень фонетического фонда русского языка. 1988, № 1. СПб. – Бохум. С. 22–26.

- Бондарко Л.В., Светозарова Н.Д., Скредин П.А. «Фонетический фонд русского языка» как исследовательская программа кафедры фонетики Ленинградского университета. // Бюллетень фонетического фонда русского языка, 1992. №4, СПб. – Бохум. С. 5–16.
- Бондарко Л.В., Кузнецов В.И., Степанова С.Б. Теоретические и практические проблемы транскрипции текста // Николаева Т.М. (отв. ред.). Проблемы фонетики. М.: Прометей, 1993. С. 8–20.
- Бондарко Л.В., Кузнецов В.И., Скредин П.А., Шалонина К.Б. Звуковая система русского языка в свете задач компилятивного синтеза // Бюллетень фонетического фонда русского языка. 1997. №6. СПб. – Бохум. С. 60–84.
- Вербицкая Л.А. Русская орфоэпия. СПб.: СПбГУ, 2013.
- Вербицкая Л.А. Роль фонологии и морфонологии в формировании особенностей произносительной нормы (на материале современного русского языка). // Каленчук М.Л. (ред.). Фонетика сегодня: Материалы докладов и сообщений VIII международной научной конференции, 28–30 октября 2016 г. М.: Нестор-История, 2016. С. 19–24.
- Вербицкая Л.А., Зиндер Л.Р. К вопросу о сочетаниях согласных в русской речи // Филологические науки. М., 1969. С. 43–53.
- Вольская Н., Коваль А., Коваль С., Опарин И., Погарева Е., Скредин П., Смирнова Н., Таланов А. Синтезатор русской речи по тексту нового поколения // Материалы XXI международной конференции Диалог 2005. Москва, 2005. URL: <https://www.dialog-21.ru/media/2431/volskayan.pdf>
- Зиндер Л.Р. Общая фонетика. М.: Высшая школа, 1979.
- Каленчук М.Л., Касаткин Л.Л., Касаткина Р.Ф. Большой орфоэпический словарь русского языка. М.: АСТ-Пресс, 2012.
- Качковская Т.В., Кочаров Д.А., Вольская Н.Б., Тананайко С.О., Васильева Л.А., Евдокимова В.В., Чукаева Т.В., Скредин П.А. Корпус русской спонтанной речи CoRuSS: состав и структура. // Труды седьмого междисциплинарного семинара анализ разговорной русской речи (АРЗ — 2017). СПб.: Политехника-принт, 2017. С. 40–45.
- Кодзасов С.В. Фонетический эллипсис в русской разговорной речи // Теоретические и прикладные исследования в области структурной и прикладной лингвистики. М.: Изд-во Московского ун-та, 1973. С. 109–133.
- Кочаров Д.А., Кочеткова У.Е. Огубленность безударных гласных в русской речи. // Вопросы языкознания, № 6, 2020, С. 31–47.
- Кузнецов В.И. Вокализм связной речи. СПб: СПбГУ, 1997.
- Кузнецова А.И., Ефремова Т.Ф. Словарь морфем русского языка: Ок. 52000 слов. М.: Русский язык, 1986.

- Садуртинова К.Р. Акустические варианты морфологических единиц. Опыт создания словаря. Магистерская диссертация, СПб., 2012.
- Светозарова Н.Д. (ред.) Фонетика спонтанной речи. Л.: Изд-во Ленингр. ун-та, 1988.
- Скрелин П.А. Сегментация и транскрипция. СПб.: СПбГУ, 1999.
- Скрелин П.А., Сомова Е.В. Консонантные сочетания в связной речи // Материалы XXIX межвузовской научно-методической конференции преподавателей и аспирантов. Вып. 4, ч. 2. Секция фонетики. СПб., 2000. С. 16–21.
- Скрелин П.А., Сомова Е.В. Реализация сочетаний согласных в связной речи (наблюдения и слуховой анализ) // Бюллетень фонетического фонда русского языка. Фонетические свойства русской спонтанной речи. 2001, № 8. СПб. – Бохум. С. 25–36.
- Скрелин П.А., Евдокимова В.В., Кочаров Д.А. «Незаметные» варианты русской произносительной нормы // XLIX Международная научная филологическая конференция, посвященная памяти Людмилы Алексеевны Вербицкой (1936–2019). Избранные доклады / под ред. Т.В. Черниговской. СПб.: СПбГУ, 2022. С. 96–117.
- Степанова С.Б. Фонетические свойства русской речи: реализация и транскрипция: Дис. ... канд. филол. наук. Л., 1988.
- Bondarko L.V., Volskaya N.B., Tananaiko S.O., Vasilieva L.A. Phonetic Properties of Russian Spontaneous Speech. *Proceedings of the 15<sup>th</sup> ICPhS*. 2003. P. 2973–2976.
- Evgrafova K. The Phonetic Characteristics of Vowel Epenthesis in Russian Consonant Clusters. *Proceedings of SPECOM 2009*. 2009. P. 419–422.
- Kachkovskaia T., Kocharov D., Skrelin P., Volskaya N. CoRuSS — a new prosodically annotated corpus of Russian spontaneous speech. *Proceedings of LREC* 2016. P. 1949–1954.
- Kachkovskaia T., Chukaeva T., Evdokimova V., Kholiavin P., Kriakina N., Kocharov D., Mamushina A., Menshikova A., Zimina S. SibLing Corpus of Russian Dialogue Speech Designed for Research on Speech Entrainment. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020. Marseille, France. European Language Resources Association. P. 6556–6561.
- Kocharov D., Menshikova A. Unrounding of vowels in Russian speech. *Proceedings of The Fifth Saint Petersburg Winter Workshop on Experimental Studies of Speech and Language (Night Whites 2019)*. 2019. P. 58.
- Shalnova K. Flexible transcriber for Russian continuous speech. *Proceedings of SPECOM 1997*. 1997. P. 171–175.
- Shalnova K. Automatic modelling of regional pronunciation variation for Russian. *Proceedings of TSD 1999*. 1999. P. 329–332.

- Skrelin P. Concatenative Russian Speech Synthesis: Sound Database Formation Principles. *Proceedings of SPECOM 1997*. 1997. P. 157–160.
- Skrelin P., Shalnova K., Shumara S. Acoustic Transcription (Representation) of Russian Vowels in Speech Technologies. *Proceedings of SPECOM 1998*. 1998. P. 239–243.
- Skrelin P., Volskaya N., Kocharov D., Evgrafova K., Glotova O., Evdokimova V. CORPRES — Corpus of Russian Professionally Read Speech. *Proceedings of TSD 2010*. 2010. P. 386–393.