

¹С. О. Савчук, ²А. А. Махова

¹Институт русского языка им. В. В. Виноградова РАН

²Институт проблем передачи информации

им. А. А. Харкевича РАН, НП «НКРЯ»

(Россия, Москва)

¹savsvetlana@mail.ru, ²discourse@yandex.ru

МУЛЬТИМЕДИЙНЫЙ ПОЭТИЧЕСКИЙ КОРПУС: ПРИНЦИПЫ СОЗДАНИЯ И ОСОБЕННОСТИ РАЗМЕТКИ¹

В статье представлен новый тип корпуса в составе мультимедийного модуля НКРЯ, в котором содержатся не только расшифровки звучащих текстов, но и аудио- / видеозаписи их произнесения. В настоящее время мультимедийный модуль включает в себя основной Мультимедийный корпус (МУРКО), предназначенный для изучения устных текстов разных жанров, и два Мультимедийных параллельных корпуса (МультиПАРКа), которые дают возможность сопоставительного изучения одной и той же реплики, произнесенной разными говорящими в одинаковых или сходных обстоятельствах. Русский МультиПАРК — в разных постановках и экранизациях одной и той же пьесы на русском языке, англо-русский МультиПАРК — в постановках пьес на русском и на английском языке, либо в оригинальной и дублированной версии фильмов.

Мультимедийный поэтический корпус предназначен для изучения особенностей звучащего стихотворного текста. Он объединяет в себе особенности мультимедийного, параллельного и поэтического корпусов. В нем собраны образцы поэтических текстов в авторском, профессиональном (актерском) и непрофессиональном исполнении. Обсуждается состав и структура корпуса и перспективы его использования. Корпус позволяет сравнивать письменный текст с его звучанием, сопоставлять разные интерпретации звучащего стиха, что представляет интерес для изучения фонетических и просодических явлений, открывает большие возможности для стиховедческих исследований, исследований поэтического синтаксиса и семантики и др.

Ключевые слова: мультимедийный поэтический корпус, принципы аннотации, звучащая художественная речь, варианты исполнения стихов.

¹ Работа выполнена при поддержке гранта Министерства науки и высшего образования РФ на реализацию крупных научных проектов по приоритетным направлениям научно-технического развития (соглашение № 075-15-2020-793).

1. Введение

Мультимедийный русский корпус (МУРКО), в котором представлены не только расшифровки звучащих текстов, но и аудио-/видеозаписи их произнесения, в настоящее время является одним из самых больших открытых мультимедийных корпусов в мире. Корпус был создан под руководством Е. А. Гришиной (1958–2016) в 2009–2010 гг., первоначально как корпус кино, в дальнейшем осуществлялось пополнение корпуса и включение в него образцов звучащей речи, функционирующей в различных речевых сферах. Цели и задачи создания корпуса, принципы отбора материала, система аннотации описаны в работах Е. А. Гришиной ([Гришина 2009a], [Grishina 2009b], [Grishina 2009c], [Grishina 2010a], [Кудинов, Гришина 2009], [Гришина 2015] и др.). Корпус успешно используется для изучения различных аспектов устной речи ([Гришина 2009d], [Гришина 2010b], [Гришина 2011], [Кобозева и др. 2019], [Бычкова и др. 2019] и др.). На основе МУРКО построен фундаментальный труд Е. А. Гришиной о системе русской жестикуляции [Гришина 2017].

Дальнейшим развитием МУРКО было создание Мультимедийного параллельного корпуса (МультиПАРКа), который предназначен для сопоставительных исследований ([Grishina et al 2012], [Гришина 2015], [Гришина 2016], [Савчук 2018]). В настоящее время оба корпуса составляют две части мультимедийного модуля НКРЯ, общий объем которого превышает 5 млн словоупотреблений.

Основной Мультимедийный корпус (МУРКО) предназначен для изучения устных текстов разных жанров. В его состав входят следующие разделы (подкорпусы).

1. *Речь кино* включает советские и российские кинофильмы 1930–2000-х годов.

2. *Устная публичная речь* представлена текстами, относящимися к устной научной речи (доклады и дискуссии на конференциях, учебные и популярные лекции, теле- и радиопередачи), устной политической речи (интервью, пресс-конференции, выступления на митингах, собраниях и съездах, ток-шоу на радио и ТВ), публицистике (интервью и беседы на разные темы, документальное кино и др.), рекламе (рекламные ролики).

3. *Устная непубличная речь* включает тексты повседневного бытового общения — диалоги и микродиалоги, разговоры в дружеском и семейном кругу, телефонные разговоры и мн. др.

4. *Театральная речь* представлена аудио- и видеозаписями театральных постановок на сцене, радио- и телеспектаклей.

5. *Авторское и художественное чтение* — эти два раздела содержат озвученную письменную речь (written-to-be-spoken), которая интересна в плане изучения фонетических особенностей звучащего текста, орфоэпии и акцентологии, интерпретации текста. В разделе собраны записи прозаических текстов в авторском исполнении и в исполнении мастеров художественного слова.

Назначение Мультимедийного параллельного корпуса — предоставить материал для изучения особенностей одного и того же текста, воспроизведенного в разных коммуникативных условиях, например, на театральной сцене (разные постановки

пьесы), в учебных ситуациях (чтение или пересказ разными учащимися), в ритуальных типах дискурса (формулы этикета, военные команды, гражданские и церковные обряды). На фоне того, что вербальная составляющая в разных версиях произнесения совпадает, наглядно выступают различия между ними на невербальном уровне — фонетические, ритмико-интонационные, мимико-жестикационные — что дает простор для изучения вариантов невербального оформления высказываний и поиска смысловых различий между ними, для определения диапазона интерпретаций письменного текста, различий между письменным и устным текстом. Помимо вариантов текста на одном языке можно сравнивать интерпретации одного и того же текста на разных языках — оригинального текста и его перевода на иностранный язык.

В настоящее время в составе МультиПАРКа функционируют два модуля. Русский МультиПАРК включает постановки пьес в виде аудио- и видеозаписей театральных спектаклей, фильмов, радиоспектаклей, при этом каждая пьеса представлена в нескольких постановках. Пьеса Н. В. Гоголя «Ревизор» представлена в 9 постановках, пьесы А. П. Чехова «Вишневый сад» и «Дядя Ваня» — в 4 и 5 постановках соответственно. Составлен архив записей авторского и актерского исполнения прозы. Русский МультиПАРК дает возможность сопоставительного изучения одной и той же реплики, произнесенной разными говорящими в одинаковых обстоятельствах. В результате таких исследований могут быть установлены пределы варьирования различных аспектов звучащей речи и ее жестового сопровождения в зависимости от факторов, связанных с личностью актера, временем и стилем постановки, замыслом режиссера т. д.

Англо-русский МультиПАРК включает фрагменты сериалов и фильмов на английском языке с закадровым русским переводом или дубляжем, а также разные постановки пьес — на русском языке и в англоязычной интерпретации. Он дает возможность сравнивать и изучать речевое поведение людей, относящихся к разным культурам, говорящим на разных языках, при этом оказавшихся в сходных ситуациях.

Третий модуль в архитектуре МультиПАРКа — корпус звучащей поэзии, пилотный проект которого представлен в настоящей статье. Идея создания Мультимедийного поэтического корпуса принадлежит Е. А. Гришиной [Grishina et al. 2012], которая начала собирать архив записей для этого корпуса и использовала видеоматериал при работе над монографией в главе о моргани [Гришина 2017: 16].

Развитие проекта в этом направлении приведет к тому, что в мультимедийном модуле будут представлены все виды художественной словесности — драма, эпос, лирика (см. табл. 1).

2. Состав корпуса

Мультимедийный поэтический корпус построен на тех же принципах, что и остальные разделы, однако имеет некоторые особенности в разметке. По составу авторов корпус предполагается сделать максимально открытым. Это означает, что

Таблица 1. Формы представления художественного слова в мультимедийном модуле НКРЯ

Виды словесности	МУРКО	Мультимедийный поэтический	МультиПАРК	
			русский	англо-русский
<i>Драма</i>	Кино, театр	—	Варианты театральных постановок и экранизаций	Оригинальные фильмы и их дублированные переводы, варианты постановок на разных языках
<i>Проза</i>	Авторское и художественное чтение	—	Варианты авторского и художественного чтения прозы	—
<i>Лирика</i>	—	Варианты авторского и художественного чтения стихов	—	—

решение о включении или не включении какого-либо автора в состав корпуса будет определяться наличием аудио-/видеозаписей исполнения его стихов и качеством записей. Предпочтение будет отдаваться авторам, произведения которых присутствуют в поэтическом корпусе НКРЯ и, как правило, имеют несколько вариантов исполнения. На первом этапе используются открытые источники, доступные в сети Интернет, в дальнейшем надеемся привлечь архивные материалы из фондов Гостелерадио, музейных коллекций и т. д.

Согласно разработанной концепции, Мультимедийный поэтический корпус включает в себя стихотворные тексты в разных вариантах исполнения. По составу исполнителей эти варианты могут быть отнесены к трем типам: 1) авторское чтение своих стихов, 2) профессиональное чтение в исполнении мастеров художественного слова (актеров, чтецов, дикторов радио и телевидения), 3) непрофессиональное чтение людьми разных возрастов и профессий, не обученными специальными приемам декламации.

В состав пилотного корпуса вошли стихотворения А. А. Ахматовой в исполнении автора, актеров (С. Н. Крючковой, А. С. Демидовой, М. Б. Тереховой, О. Л. Белявской, Е. А. Ниловой, И. В. Кваши и др.), а также непрофессиональных чтецов. Включены стихи С. А. Есенина в исполнении актеров (О. В. Басилашвили, С. В. Безрукова, Б. С. Галкина, И. Ю. Ильина, М. М. Козакова и др.), профессиональных чтецов (В. Н. Аксенова, В. С. Левашева), других поэтов (Б. А. Ахмадулиной, А. А. Вознесенского, Е. М. Винокурова, Е. А. Евтушенко и др.). Стихи А. С. Пушкина исполняют актеры (И. В. Кваша, М. М. Козаков, Э. Е. Марцевич, Е. В. Миронов, И. М. Смоктуновский); стихи М. Ю. Лермонтова — актеры (В. И. Качалов, В. А. Барин, Н. П. Бурляев, Э. Е. Марцевич) и непрофессиональные исполнители. В архиве корпуса — уникальные записи авторского чтения П. Г. Антокольского, И. А. Бунина, А. Е. Кручёных, О. Э. Мандельштама, Б. Л. Пастернака, А. А. Тарковского, М. Н. Цветаевой, а также видеозаписи современных поэтов.

Исследователи звучащей речи 1920–1930-х годов установили ряд особенностей, различающих авторское исполнение стихов от профессионального чтения ([Эйхенбаум 1969], [Бернштейн 1977], [Золотухин 2010]). Авторское исполнение стремится передать прежде всего фонетическую сторону стиха, его мелодику, в то время как исполнительское чтение ориентировано на синтаксическую структуру, передачу смысла стиха, поэтому в интонационном отношении оно более «прозаическое», ближе к практической речи [Бернштейн 1977: 212]. Для сценического исполнения характерна драматизация, как и при исполнении прозы, актер «разыгрывает» стихотворение, сопровождает чтение жестами, мимикой, сменой поз [Бернштейн 1977: 213]. Что касается непрофессионального исполнения, то оно изучено меньше, но можно предположить, что любительское чтение будет тяготеть либо к авторской, либо к сценической манере исполнения.

Создаваемый корпус аккумулирует большое количество записей чтения и разнообразие исполнителей-чтецов — поэтов, профессиональных чтецов, актеров, любителей поэзии разных возрастов. Анализ этих записей, возможно, позволит определить какие-то ключевые показатели, отличающие один тип чтения от другого, установить вариативность внутри каждого типа, переходные, гибридные варианты. В частности, поможет исследовать вопрос о том, к какому типу отнести чтение поэтом стихов другого поэта — авторскому, актерскому или непрофессиональному (например, исполнение стихов С. Есенина поэтами Е. Евтушенко, Б. Ахмадулиной, А. Вознесенским и др.). Результаты, полученные с помощью корпусных инструментов, было бы интересно сравнить с результатами исследований звучащей стихотворной речи 1920–1930-х годов [Звучащая 2018], что поможет проследить изменение манеры чтения в диахронической перспективе и установить факторы, влияющие на изменения.

3. Организация базы данных и поисковый интерфейс

В архитектуре базы данных Мультимедийного поэтического корпуса использован принцип параллельного корпуса, реализованный в «театральной» части МультиПАРКа, при котором сопоставляются разные исполнительские версии одного и того же текста. Одна версия — печатная публикация по авторитетному изданию — служит каноническим инвариантом. Остальные версии — это записи чтения стихотворения разными исполнителями, представленные в двух модусах — в виде аудио-/видеозаписи и в виде текстовой расшифровки.

Вопрос о том, нужно ли каждую исполнительскую версию сопровождать ее текстовой расшифровкой, является дискуссионным. С одной стороны, читается одно и то же стихотворение, поэтому можно было бы ограничиться одним каноническим, печатным текстом и к нему «привязать» видео- / аудиозаписи с разными версиями исполнения. С другой стороны, как известно из описаний других коллекций (см., например, [Валиева 2015: 154–155]), в разных исполнительских версиях могут присутствовать не только просодические, интонационные, жестикуляционные и пр. различия, которые зафиксированы в видео-/аудиозаписи, но могут встречаться трансформации текста и на вербальном уровне (перестановки, пропуски,

замены слов и др.). Подобные случаи вариативного исполнения могут представлять известный интерес для исследователей устного стиха. Как показало предварительное исследование нашего материала, в пилотной коллекции случаи трансформаций текста на вербальном уровне нередки, особенно в записях чтения наизусть [Савчук 2022]. Поэтому было принято решение сопровождать каждую видеозапись чтения точным транскриптом с соответствующей разметкой.

4. Технология подготовки данных

Технология подготовки базы данных мультимедийных корпусов в составе НКРЯ опирается на две процедуры — фрагментирование и выравнивание. Звучащий текст в Мультимедийном корпусе представлен в виде аудио- и видеофайлов, разрезанных на небольшие фрагменты (клипы) длительностью 10–30 сек., каждому из которых поставлен в соответствие фрагмент текстовой расшифровки. Пара «клип + текст» (или *кликст*, по терминологии Е. А. Гришиной), представляет собой, как правило, относительно законченный в смысловом отношении коммуникативный фрагмент. Как показал многолетний опыт подготовки и использования мультимедийного корпуса, способ деления материала на фрагменты оказался оптимальным выбором и для выравнивания фрагментов при формировании корпуса, и для обеспечения скорости и надежности поиска по корпусу, и для удобства сохранения результатов.

В МультиПАРКе пары «клип + текст» каждой исполнительской версии дополнительно сопоставлены и выровнены с «каноном», или инвариантом — текстом стихотворения по печатной публикации, при этом нумерация фрагментов канона и нумерация текстовых фрагментов и клипов каждой исполнительской версии должны совпадать. Таким образом, выдача по запросу пользователя обязательно включает фрагмент «канона» и некоторое количество (в зависимости от количества версий) соответствующих фрагментов из разных исполнительских версий, которые сопровождаются клипами.

5. Аннотация и поисковый интерфейс

В аннотации корпуса сочетаются разные виды разметки:

- 1) разметка, общая для всех корпусов в составе НКРЯ (метатекстовая, морфологическая, семантическая аннотация);
- 2) разметка, используемая в поэтическом корпусе (метрические признаки, рифма), которая содержится в каноническом инварианте;
- 3) разметка, используемая в корпусах устной речи (акцентная, социологическая аннотация);
- 4) специфическая разметка просодического членения [Кривнова и др. 2016], которая наряду с акцентной и социологической аннотацией представлена в исполнительских версиях.

Метатекстовая аннотация в Мультимедийном поэтическом корпусе имеет некоторые особенности: помимо сведений о тексте она содержит информацию о метрических признаках стихотворения (см. п. 2).

Просодическая разметка, а также разметка различных отклонений от канонического инварианта, о которых было сказано выше, различает разные версии исполнения. Аннотированный текст выглядит следующим образом².

- (1) `<p class="verse"><line meter="X5ж">У меня сегодня много <rhyme-zone/> дѣла:
 <line meter="X5"/>Надо память до конца <rhyme-zone/>убить,
 <line meter="X5"/>Надо, чтоб душа <rhyme-zone/> окаменѣла,
 <line meter="X5"/> Надо снова научиться <rhyme-zone/> жить.</p>`
- (2) `<speech actor="А. Ахматова">У меня сегодня много дѣла: ||
 На́до па́мять до конца́ уби́ть, ||
 На́до, чтоб ду́ша́ окаменѣла, ||
 На́до сно́ва научи́ться жи́ть.||
`
- (3) `<speech actor="С. Крючкова"> У меня́ сегодня́ мно́го дѣла: ||
 На́до па́мять | до конца́ уби́ть, ||
 На́до, чтоб ду́ша́ | окаменѣла, ||
 На́до сно́ва | научи́ться | жи́ть.||
`
- (4) `<speech actor="К. Москаленко">У меня́ сегодня́ мно́го дѣла: ||
 На́до | па́мять до конца́ уби́ть, ||
 На́до, чтоб | ду́ша́ окаменѣла, ||
 На́до сно́ва | научи́ться | жи́ть.||
`
- (5) `<speech actor="Е. Погребижская">У меня́ сегодня́ мно́го дѣла: ||
 На́до | па́мять | до конца́ уби́ть, ||
 На́до, чтоб ду́ша́ | окаменѣла, ||
 На́до | сно́ва научи́ться жи́ть.||
`

Как видно уже по транскриптам, все четыре варианта исполнения четверостишия отличаются друг от друга. Просодическое членение первого стиха одинаково во всех четырех вариантах исполнения, второй и третий стих во всех четырех версиях произносится по-разному, а членение четвертого стиха совпадает в чтении К. Москаленко и С. Крючковой и отличается от двух остальных.

Однако размеченный транскрипт — лишь один из источников информации для исследователя. Мультимедийный корпус дает возможность услышать или увидеть реальное исполнение стихотворения. Поисковый интерфейс позволяет по запросу на какое-либо слово, конструкцию или признак получить содержащий соответствующую единицу фрагмент текста в нескольких вариантах исполнения. При этом каждый текстовый фрагмент будет сопровождаться видео- или аудиоклипком,

² (1) — канонический инвариант с размеченными сильными долями (´), зоной рифмовки (<rhyme-zone/>), типом рифмы и размера (в данном случае X5ж — пятистопный хорей, женская рифма). (2) — транскрипт авторского чтения, (3) — транскрипт актерского исполнения, (4) и (5) — транскрипты непрофессионального исполнения, | и || — границы просодических швов.

который можно прослушать, повторить, сохранить с целью дальнейшего анализа с помощью внешних программ.

В корпусе организован поиск по метатекстовым признакам, на основе которых можно отбирать свой собственный подкорпус, наиболее подходящий для решения исследовательских задач. По признаку «автор» (фамилия, пол, год рождения) можно выбрать все тексты какого-либо конкретного поэта, либо разных поэтов определенного пола или возраста во всех вариантах исполнения. По «названию» можно отобрать все исполнительские версии определенного стихотворения и сравнивать разные трактовки между собой. Признак «исполнитель» (фамилия, пол, год рождения) собирает в один подкорпус все тексты в чтении конкретного исполнителя (автора, актера или чтеца) или исполнителей, подобранных по какому-нибудь социологическому признаку (полу, возрасту). Признак «тип исполнения» дает возможность отбора в подкорпус всех стихотворений в авторском чтении, в профессиональном или непрофессиональном исполнении, что облегчает поиск общих особенностей и различий в разных типах чтения. Метрическая разметка позволяет отбирать тексты, написанные определенным стихотворным размером, имеющие определенный тип рифмовки и др. Если использовать сочетание разных признаков, можно предельно сузить и конкретизировать поиск нужного материала.

6. Использование корпуса

Описанная выше организация материала и поисковый интерфейс дают возможность сопоставительного изучения стихотворного текста. Создаваемый корпус предназначен для научных исследований в области лингвистики и лингвистической поэтики, фоностилистики, мультимодальной коммуникации, искусствоведения и т. д., а также для проведения междисциплинарных исследований.

Приведем несколько примеров. Сравним два варианта чтения стихотворения С. Есенина «Спит ковыль»: в исполнении актера Б. Галкина (6) и профессионального чтеца И. Ильина (7)³.

(6) <speech actor="Б. Галкин"> Знать, [у нас] у всех у нас такая участь. ||
И, пожалуй, ~~всякого~~ [каждого] | спреси — ||
Радуясь, свирепствуя и мучась, |
Хорошо живётся на Руси? ||

(7) <speech actor="И. Ильин"> Знать, у всех у нас такая участь. ||
И, пожалуй, | ~~всякого~~ спреси — ||
Радуясь, свирепствуя и мучась, ||
Хорошо | живется на Руси. ||

В исполнении Б. Галкина видим перестановку слов в первом стихе (*у нас у всех* вм. *у всех у нас*) и замену слов во втором стихе (*каждого спреси* вм. *всякого спреси*). Просодическое членение второго и четвертого стихов отличается в двух

³ Видеозаписи фрагментов см. по ссылке <https://disk.yandex.ru/d/3imKXt0hPwztjA>

версиях, значительно расходится и интонационное оформление четвертого стиха. В исполнении И. Ильина четвертый стих произносится с эмфазой на слове *хорошо* (ИК-7). В интерпретации Б. Галкина второй стих произносится с побудительной интонацией (ИК-2), а четвертый стих звучит как вопрос (ИК-3). В результате в целом строфа С. Есенина в двух версиях исполнения интерпретируется по-разному. В одном случае (у И. Ильина) автор уверен в ответе на вопрос о том, хорошо ли живется на Руси: «*всякого спроси*» — [и получишь ответ]: «*Радуясь, свирепствуя и мучась, / Хорошо живет на Руси!*». Во втором случае Б. Галкин как бы обращается к каждому слушателю-зрителю и предлагает задать вопрос «*Хорошо ли живется на Руси?*». Таким образом, письменный текст стихотворения может реализоваться в широком диапазоне интонационного оформления.

Наличие видеозаписей в мультимедийном поэтическом корпусе дает возможность изучения жестового сопровождения стихотворной речи. Имеет ли оно специфические особенности по сравнению с прозаической речью — сценической и произносимой в естественных условиях? Если да, то какими факторами они обусловлены? Рассмотрим транскрипт отрывка стихотворения М. Ю. Лермонтова «Бородино» в исполнении В. Баранова⁴.

- (8) [⊙] Ну ж был денек! || ^{2Н ОЛ↓} {Сквозь дым летучий} ||
[↻] {Французы} | ^{2Н ОЛ ↗} {двинулись} | ^{2Н ОЛ} {как тучи}, ||
^{2Н↑} {И всё на наш редут.} ||
^{ПГ RH▲} Уланы с пестрыми значками, ||
Драгуны | ^{RH↙} {с конскими хвостами}, ||
^{2Н↑} {Все промелькнули перед нами}, ||
^{RH↑↗} {Все} | побывали | тут. ||

В жестикуляции актера преобладают изобразительные и указательные жесты, которые участвуют в передаче тематической информации, сопровождая динамичный рассказ. Жест ^{2Н ОЛ↓} «две руки с открытыми ладонями, ориентированными вниз» связан с идеей плоскости и большого пространства, лежащего ниже точки зрения наблюдателя, которая совмещена с точкой зрения рассказчика. Действительно, рассказчик-наблюдатель находится на возвышении (редуте), следовательно, в поле его зрения попадает большое количество объектов. Идея большого количества объектов передается жестами [↻] «симметричная окружность» (*французы*), ^{2Н↑} {«разведение рук в стороны»} (*как тучи, и всё на наш редут, все промелькнули*), ^{RH↑↗} «радикальное пересечение» (*все побывали*). Жест ^{2Н ОЛ ↗} «движение двух рук вперед» наглядно иллюстрирует смысл глагола *двинулись*, ^{ПГ RH▲} «поворот головы» вправо и движение правой рукой указывают направление, откуда двигались *уланы*, а движение правой рукой над головой по дуге назад ^{RH↙} «выгнутая дуга» изображает головной убор драгун (*с конскими хвостами*).

⁴ В транскрипте использована система обозначений и интерпретация жестов, предложенная в работах Е. А. Гришиной [Гришина 2017]. Видеозапись фрагмента см. по ссылке <https://yadi.sk/i/2ru7AF8HvYko2A>

Помимо изобразительных жестов исполнитель использует прагматические жесты для передачи аксиологической информации: «головные» жесты $\hat{\sim}$ «кивок сверху вниз» подчеркивает эмфазу (*французы*), жест $\textcircled{\text{R}}$ «качать головой» сопровождает восклицание с отрицательной оценкой *Ну ж был денек!* [Гришина 2017: 455].

Как видим, при исполнении эпического по содержанию стихотворного фрагмента жестикуляция исполнителя близка к той, которой сопровождается прозаический рассказ. Можно предположить, что жестикуляционное оформление нарратива не зависит от формы — стихотворной или прозаической, однако для проверки этого предположения нужен большой материал, который и может предоставить корпус.

7. Заключение

В Мультимедийном поэтическом корпусе собран уникальный материал, который не представлен ни в одном из корпусов НКРЯ. Благодаря многофункциональной разметке и оригинальной архитектуре корпус соединяет в себе возможности поэтического, параллельного и мультимедийного корпусов. Новый ресурс, как кажется, может стать полезным инструментом при изучении

- вариантов ритмической и интонационной организации стиха в соотношении с его синтаксической и логической организацией;
- вариантов жестового сопровождения стихотворной речи в зависимости от разной манеры исполнения и в сопоставлении с прозаической речью;
- особенностей непрофессионального исполнения в сопоставлении с авторской и актерской манерой чтения; произносительных тенденций, собственных стихотворной речи вообще;
- особенностей сценической речи, приемов актерского исполнения в зависимости от индивидуальных качеств исполнителя, принадлежности к актерской школе и мн. др.

Особо следует отметить возможность использования нового корпуса в образовательных целях. Корпус содержит большую коллекцию русской классической поэзии в устном исполнении, в которой значительное место занимают образцы высокого исполнительского мастерства, а также оригинального авторского исполнения. Все это делает Мультимедийный поэтический корпус, так же как и русский МультиПАРК, ценным ресурсом при изучении филологических и искусствоведческих дисциплин, актерского мастерства, искусства художественного слова.

Литература

Бернштейн С. И. Голос Блока. / Публикация А. Ивича и Г. Суперфина // Блоковский сборник. II. Труды Второй научной конференции, посвященной жизни и творчеству А. А. Блока. Тарту, 1977. С. 454–525.

Бычкова П. А., Рахилина Е. В., Слепак Е. А. Дискурсивные формулы, полисемия и жестовое маркирование // Труды Института русского языка им. В. В. Виноградова. 2019. Вып. 21. С. 256–283.

Валиева Ю. М. О звуковом архиве «Лица петербургской поэзии: 1950–1990» // Живое слово: логос, голос, движение, жест. М.: НЛО. 2015. С. 151–158.

Гришина Е. А. Мультимедийный русский корпус (МУРКО): проблемы аннотации // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-история, 2009а. С. 175–214.

Гришина Е. К вопросу о соотношении слова и жеста: вокальный жест *О* в устной речи // Компьютерная лингвистика и интеллектуальные технологии (по материалам ежегодной Международной конференции «Диалог 2009»). Вып. 8 (15). М.: РГГУ, 2009d. С. 80–90.

Гришина Е. Вокальный жест *А* в устной речи. Компьютерная лингвистика и интеллектуальные технологии (по материалам ежегодной Международной конференции «Диалог 2010»). Вып. 9 (16). М.: РГГУ, 2010b. С. 102–112.

Гришина Е. Да в русском устном диалоге // Russian Linguistics. 2011. № 35 (2). С. 169–207.

Гришина Е. А. Мультимодальный модуль в составе Национального корпуса русского языка // Труды Института русского языка им. В. В. Виноградова. 2015. Вып. 6. С. 65–88.

Гришина Е. А. Мультимедийный параллельный корпус (МультиПАРК): новый тип корпуса для сопоставительных исследований // Седьмая международная конференция по когнитивной науке: Тезисы докладов. Светлогорск, 20–24 июня 2016 г. / Отв. ред. Ю. И. Александров, К. В. Анохин. М.: Изд-во «Институт психологии РАН», 2016. 720 с.

Гришина Е. А. Русская жестикуляция с лингвистической точки зрения (корпусные исследования). М.: ЯСК, 2017. 744 с.

Звучащая художественная речь: Работы Кабинета изучения художественной речи (1923–1930) / Сост.: В. Золотухин, В. Шмидт. М.: Три квадрата, 2018. 516 с.

Золотухин В. Колодец и колокол. О влиянии авторской декламации на исполнительскую (1905–1929) // Вопросы театра. Prosaenium. Сборник статей. 2010. № 3–4. С. 205–231.

Кобозева И. М., Иванова О. О., Захаров Л. М. К мультимодальному моделированию верификативных дискурсивных маркеров в русском диалоге // Труды Института русского языка им. В. В. Виноградова. 2019. Вып. 21. С. 284–299

Кривнова О. Ф., Князев С. В., Моисеева Е. В. Исследования просодического членения звучащего текста на материале русского языка // Вестник Московского университета. Серия 9: Филология. 2016. № 4. С. 7–33.

Кудинов М. С., Гришина Е. А. Инструменты полуавтоматической разметки для Мультимедийного русского корпуса (МУРКО) // Компьютерная лингвистика и интеллектуальные технологии (по материалам ежегодной Международной конференции «Диалог 2009»). Вып. 8 (15). М., 2009. С. 248–261.

Савчук С. О. Мультимедийный параллельный корпус: перспективы развития // Восьмая международная конференция по когнитивной науке: Тезисы докладов. Светлогорск, 18–21 октября 2018 г. / Отв. ред. А. К. Крылов, В. Д. Соловьев. — М.: Изд-во «Институт психологии РАН», 2018. С. 1320–1322.

Савчук С. О. Мультимедийный поэтический корпус: общая структура и перспективы использования // Труды Института русского языка им. В. В. Виноградова. 2022 (в печати).

Эйхенбаум Б. М. О камерной декламации // О поэзии. М.: Советский писатель, 1969. С. 512–541.

Grishina E. Multimodal Russian Corpus (MURCO): Types of annotation and annotator's workbenches // Proceedings of the Corpus Linguistics Conference CL2009, University of Liverpool, UK, 20–23 July 2009. [Электрон. ресурс]. 2009b. URL: <http://ucrel.lancs.ac.uk/publications/cl2009/#papers>, #165

Grishina E. Multimodal Russian Corpus (MURCO): general structure and user interface. // NLP, Corpus Linguistics, Corpus Based Grammar Research. Fifth International Conference, Smolenice, Slovakia, 25–27 November 2009. Proceedings. Bratislava: Tribun, 2009c. P. 119–131.

Grishina E. Multimodal Russian Corpus (MURCO): First Steps. // Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. & Tapias, D. (Eds), 7th Conference on Language Resources and Evaluation LREC'2010. Valetta, Malta: ELDA, 2010a. P. 2953–2960.

Grishina E., Savchuk S., Sichinava D. Multimodal Parallel Russian Corpus (MultiPARC): Main Tasks and General Structure // LREC 2012 Workshop on Best Practices for Speech Corpora in Linguistic Research. Istanbul, Turkey, May 21, 2012. ELRA. P. 13–16.

¹S. O. Savchuk, ²A. A. Makhova

*¹V. V. Vinogradov Russian Language Institute,
Russian Academy of Sciences*

*²A. A. Kharkevich Institute
for Information Transmission Problems,
Russian Academy of Sciences,
Nonprofit Partnership 'Russian National Corpus'
(Russia, Moscow)*

¹savsvetlana@mail.ru, ²discourse@yandex.ru

MULTIMODAL POETIC CORPUS: PRINCIPLES OF CONSTRUCTION AND MARKUP FEATURES

The article presents a new type of corpus within the Multimodal module of the RNC, which contains not only transcripts of audio texts, but also audio/video recordings of their pronunciation. The multimodal module presently includes the main Multimodal Corpus (MURCO) for the study of oral texts of different genres, and two Multimodal Parallel Corpora (MultiPARC), which allow the comparative study of the same utterance spoken by different speakers in the same or similar circumstances. Russian MultiPARC —

in different productions and adaptations of the same play in Russian, English-Russian MultiPARC — in productions of plays in Russian and in English, or in the original and dubbed versions of films.

The Multimodal Poetry Corpus is designed to explore the features of a sounding poetic text. It combines the features of multimedia, parallel and poetry corpus. It contains the poetic texts in author's, professional (actor's) and non-professional performance. The composition and structure of the corpus and the prospects for its use are discussed. The corpus allows to compare the written text with its sound, to compare different interpretations of the sounding verse, which is of interest for the study of phonetic and prosodic phenomena. It also opens up many possibilities for verse studies, research of poetic syntax and semantics, etc.

Keywords: multimodal poetic corpus, annotation principles, sounding poetic speech, variety of interpretations.

References

Bernshtein S. I. [Blok's Voice / Publication by A. Ivich and G. Superfin]. *Blokovskii sbornik. II. Trudy Vtoroi nauchnoi konferentsii, posvyashchennoi zhizni i tvorchestva A. A. Bloka* [Blok's Collection. II. Proceedings of the 2nd scientific Conference dedicated to the life and work of A. A. Blok]. Tartu, 1977, pp. 454–525 (In Russ.)

Bychkova P. A., Rakhilina E. V., Slepak E. A. [Discourse formulae, polysemy and gesture marking]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Russian Language Institute], 2019, no. 21, pp. 256–283. (In Russ.)

Eikhenbaum B. M. [On Chamber Recitation]. *O poezii* [On Poetry]. Moscow, Sovetskii pisatel' Publ., 1969, pp. 512–541 (In Russ.)

Grishina E. A. [Multimodal Russian Corpus (MURCO): Problems of annotation]. Plungian V. A. (Ed.), *Natsional'nyi korpus russkogo yazyka: 2006–2008. Novye rezul'taty I perspektivy* [The Russian National Corpus: 2006–2008. New results and perspectives]. St. Petersburg, Nestor-Istoriya Publ., 2009a, pp. 175–213. (In Russ.)

Grishina E. Multimodal Russian Corpus (MURCO): Types of annotation and annotator's workbenches. Proceedings of the Corpus Linguistics Conference CL2009, University of Liverpool, UK, 20–23 July 2009. Available at: <http://ucrel.lancs.ac.uk/publications/cl2009/#papers, #165> (accessed 22.09.2021)

Grishina E. Multimodal Russian Corpus (MURCO): general structure and user interface. *NLP, Corpus Linguistics, Corpus Based Grammar Research. Fifth International Conference*, Smolenice, Slovakia, 25–27 November 2009. Proceedings. Bratislava, Tribun, 2009c, pp. 119–131.

Grishina E. Multimodal Russian Corpus (MURCO): First Steps. Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. & Tapias, D. (Eds), *7th Conference on Language Resources and Evaluation LREC'2010*. Valetta, Malta: ELDA, 2010a, pp. 2953–2960.

Grishina E. [The Multimodal module as part of the Russian National Corpus]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the Vinogradov Russian language Institute], 2015, no. 6, pp. 65–88. (In Russ.)

Grishina E. [On gesture–word correlation (vocal gesture *Oh* in spoken Russian)]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii (Mezhdunarodnaya konferentsiya "Dialog 2009")* [Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue 2009"]. Issue 8 (15). Moscow, RSUH Publ., 2009d, pp. 80–90. (In Russ.)

Grishina E. [Vocal Gesture *Ah* in Spoken Russian]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii (Mezhdunarodnaya konferentsiya "Dialog 2010")* [Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2010)]. Issue 9 (16). Moscow, RSUH Publ., 2010b, pp. 102–112. (In Russ.)

Grishina E. [Yes in Russian oral dialogue]. *Russian Linguistics*, 2011, no. 35 (2), pp. 169–207. <https://doi.org/10.1007/s11185-011-9073-z> (In Russ.)

Grishina E. A. *Russkaya zhestikulyatsiya s lingvisticheskoi tochki zreniya (korpusnye issledovaniya)* [Russian gestures from a linguistic perspective: A collection of corpus studies]. Moscow, Yazyki slav'anskoj kul'tury Publ., 2017, 744 p. (In Russ.)

Grishina E. A. [Multimodal parallel corpus (MultiPARC): a new type of corpus for comparative studies]. Yu. I. Aleksandrov, K. V. Anokhin (Eds.). *7 mezhdunarodnaya konferentsiya po kognitivnoi nauke: Tezisy dokladov* [7th International Conference on Cognitive Science: Abstracts of reports]. Svetlogorsk, 20–24 June 2016. Moscow, "Institute of Psychology RAS" Publ., 2016. 720 s. (In Russ.)

Grishina E., Savchuk S., Sichinava D. Multimodal Parallel Russian Corpus (MultiPARC): Main Tasks and General Structure. *LREC 2012 Workshop on Best Practices for Speech Corpora in Linguistic Research*. Istanbul, Turkey, May 21, 2012. ELRA. pp. 13–16.

Kobozeva I., Ivanova O., Zakharov L. [Towards multimodal modelling of verificational discourse markers in Russian dialog]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova*. 2022 [Proceedings of the V. V. Vinogradov Russian Language Institute]. 2019, no. 21, pp. 284–299. (In Russ.)

Krivnova O. F., Knyazev S. V., Moiseeva E. V. [Studies of prosodic division of a sounding text on the material of the Russian language]. *Vestnik Moskovskogo universiteta*. Series 9. Philology, 2016, no. 4. pp. 7–33. (In Russ.)

Kudinov M. S., Grishina E. A. [Tools for semi-automatic markup for Multimodal Russian corpus (MURCO)]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po materialam ezhegodnoi mezhdunarodnoi konferentsii "Dialog"* [Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue 2009"]. Issue 8 (15). Moscow, RSUH Publ., 2009, pp. 248–261. (In Russ.)

Savchuk S. O. [Multimodal parallel corpus: prospects of development]. A. K. Krylov, V. D. Solov'ev (Eds.). *8 mezhdunarodnaya konferentsiya po kognitivnoi nauke: Tezisy dokladov* [8th International Conference on Cognitive Science: Abstracts of reports]. Svetlogorsk, 18–21 October 2018. Moscow, "Institute of Psychology RAS" Publ., 2018, pp. 1320–1322. (In Russ.)

Savchuk S. O. [Multimodal poetic corpus: general structure and prospects of use]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova*. [Proceedings of the V. V. Vinogradov Russian Language Institute]. 2022 (in print) (In Russ.)

Valieva Yu. M. [About the sound archive “Faces of St. Petersburg poetry: 1950–1990”]. *Zhivoe slovo: logos, golos, dvizhenie, zhest* [The living word: logos, voice, movement, gesture.]. Moscow, NLO, 2015, pp. 151–158 (In Russ.)

Zolotukhin V. [The well and the bell. On the influence of the author's recitation on the performance one (1905–1929)]. *Voprosy teatra. Prosaenium. Sbornik statei* [Theatre issues. Prosaenium. Collection of articles], 2010, no. 3–4, pp. 205–231. (In Russ.)

Zvuchashchaya khudozhestvennaya rech': Raboty Kabineta izucheniya khudozhestvennoi rechi (1923–1930) [Sounding artistic speech: The works of the Cabinet for the Study of Artistic Speech (1923–1930)] / Comp. by V. Zolotukhin, V. Shmidt. Moscow, Tri kvadrata Publ., 2018, 516 p. (In Russ.)