

С.О. Савчук
S.O. Savchuk

**УСТНАЯ ПУБЛИЧНАЯ РЕЧЬ
В МУЛЬТИМЕДИЙНОМ МОДУЛЕ НКРЯ¹
ORAL PUBLIC SPEECH IN MULTIMODAL MODULE
OF THE RUSSIAN NATIONAL CORPUS**

Аннотация. В статье представлен опыт создания специализированных мультимедийных корпусов, включающих устную научную и устную политическую речь в разнообразии их жанровых разновидностей и в различных ситуациях произнесения.

Ключевые слова. *Устный научный дискурс, устный политический дискурс, Мультимедийный корпус русского языка.*

Abstract. The article deals with the project compiling specialized multimodal corpora of oral Russian scientific and political discourse in a variety of genres and in different situations. The composition of the corpus, technology of database constructing, types of annotation and the interface are described, as well as the prospects of its use in studying oral public discourse.

Keywords. *Oral scientific discourse, oral political discourse, Multimodal Russian Corpus.*

Введение

Устная научная и политическая речь, а также другие, менее изученные в лингвистике виды коммуникации (деловая, правовая, рекламная, церковная) составляют сферу публичной речи, основными признаками которой являются ориентация на

¹ Исследование выполнено при финансовой поддержке РФФИ, проект № 15-06-04334 и программы ОИФН РАН «Евразийское наследие и его современные смыслы».

публичную аудиторию, социально-значимая тематика, относительная подготовленность (неспонтанность), языковая и стилистическая организованность. Несмотря на разделяемое многими понимание дискурса как речевой деятельности, протекающей в определенном социальном пространстве, объектом изучения в большинстве исследований становятся письменные источники – либо опубликованные тексты, либо стенограммы или транскрипты устных выступлений (см., например, [Лаптева 1985], [Лаптева 2007], [Паршина 2012]). Между тем устное исполнение публичной речи, даже при условии ее предварительной подготовки, нельзя приравнять к печатной публикации или стенограмме, поскольку при письменной фиксации теряются средства актуализации смыслов и эмоционального воздействия на слушателя, которыми обладает устная речь. Если еще недавно лингвисты не имели инструмента, позволяющего фиксировать и анализировать устную русскую речь в единстве с манерой и обстоятельствами ее произнесения, то в настоящее время таким инструментом стал мультимедийный корпус в составе НКРЯ (МУРКО). Инструментарий корпуса позволяет изучать как фонетические и интонационные особенности тех или иных конструкций, так и жестикуляционное сопровождение устной речи. В настоящее время в составе МУРКО активно формируется коллекция текстов, относящихся к устной научной и политической речи. Состав коллекции и особенности ее представления в корпусе описаны в настоящей статье.

Состав корпусов и источники текстов

Формирование текстового состава мультимедийных корпусов устной научной и политической речи осуществляется на основе отбора наиболее частотных жанров и типовых ситуаций, представляющих разновидности научного и политического дискурса. При этом используются только общедоступные источники, размещенные в интернете, а также

собственные записи и материалы, предоставленные партнерами, среди которых Фонотека ИРЯ РАН, видеоархив МГУ им. М.В. Ломоносова, проект «Устная история» и др. Материал представлен в виде аудио- и видеозаписей с текстовыми расшифровками.

Коллекция записей *устной научной речи* (УНР) обширна и разнообразна по составу говорящих (их более 100), областей науки (естественные, гуманитарные, прикладные), временному диапазону, по разнообразию жанрового состава. Ведущими монологическими жанрами, представленными в корпусе, являются *доклад, лекция*; диалогическими – *дискуссия, семинар*. Разновидности научного дискурса – *специальная научная речь, учебная научная речь, научно-популярная речь* – обусловлены различными ситуациями общения. Для каждой разновидности характерны своя иерархия жанров и жанровые варианты, а также специфические жанры, не представленные в других разновидностях. В частности, для учебной научной речи наиболее частотными жанрами являются лекции, семинары и школьные уроки. Для научно-популярной речи – научно-популярные лекции, беседы и интервью, рассказы, комментарии, прежде всего в СМИ.

В настоящее время УНР представлена в составе Мультимедийного корпуса лекциями («Публичные лекции "Полит.ру"» и проект Academia на телеканале «Культура»), а также беседами с ведущими учеными (телепередача «Гордон»). Это обширный интересный материал (объемом более 800 тыс. словопотреблений), на базе которого выполнен ряд исследований (см., в частности, [Гришина 2015]). Текущая работа по формированию корпуса УНР состоит в увеличении жанрового разнообразия текстов и коррекции баланса. С этой целью в состав корпуса включены записи учебных лекций по математике, химии, информатике, экономике, правоведению, психологии, искусствоведению и др., предназначенные для студентов и школьников, записи мастер-классов и семинаров. Для пополнения коллекции

научно-популярной разновидности собраны записи телепередач и интернет-проектов научно-просветительского характера: «Идеи, которые изменили мир», «Истории из будущего», «Вопрос науки», «ЕХперименты», интервью на канале «Эксперт-ТВ», лекторий «Прямая речь», «Постнаука» и др. (см. табл. 1).

Таблица 1. Планируемое пополнение подкорпуса научной речи

Жанры	Специальная научная речь (тыс. с/у)	Научно-популярная речь (тыс. с/у)	Учебно-научная речь (тыс. с/у)	Кол-во словоуп. (тыс. с/у)
Доклад	51	35	10	96
Лекция	24	10	66	100
Семинар	17		14	31
Дискуссия	8			8
Ток-шоу		20		20
Интервью		55		55
Всего	100	120	90	310

Подкорпус *устной политической речи* (УПР) создается впервые. Главная задача состоит в том, чтобы при ограниченном объеме корпуса отразить в нем вариативность политического дискурса по нескольким параметрам. Разнообразие по *составу говорящих* достигается включением в корпус выступлений представителей как власти, так и оппозиции, руководителей федерального и местного уровня, представителей разных ветвей власти, общественных деятелей. Для отражения вариативности по типу *ситуаций общения* и *тематике* в корпус включены обсуждения вопросов внутренней и внешней политики в самых разных ситуациях. Типологически их можно объединить в три группы: 1) выступления профессиональных политиков в профессиональной аудитории, в СМИ, на встречах с избирателями и пр.; 2) выступления политологов, обозревателей и комментаторов в СМИ; 3) выступления на митингах, встречах с

представителями власти, с кандидатами в период предвыборных кампаний и пр. Жанровое разнообразие достигается включением в корпус основных монологических и диалогических жанров политического дискурса: *речь, доклад, комментарий*, с одной стороны, и *интервью, пресс-конференция, парламентские слушания, круглый стол, дискуссия, совещание, встречи с избирателями* – с другой (см. табл. 2).

Таблица 2. Планируемый состав подкорпуса политической речи

Жанры		Кол-во (тыс. с/у)
Монологи- ческие	доклад	20
	лекция	20
	речь	38
	комментарий	20
Диалогиче- ские	интервью	82
	пресс-конференция	50
	парламентские слушания	25
	круглый стол	25
	дискуссия	30
	совещание	10
	ток-шоу	70
	встречи с избирателями, дебаты	60
Всего	450	

Технология организации базы данных и аннотация

Технологическая цепочка опирается на опыт создания Мультимедийного корпуса, подробно описана в работах [Гришина 2007], [Гришина 2015], [Савчук, Махова 2017] и сводится к следующим этапам. Каждая аудио-/видеозапись разрезается на небольшие фрагменты (клипы). На соответствующие фрагменты разрезаются расшифровки этих фрагментов. После

этого клип и расшифровка выравниваются между собой. Клипы и текстовые фрагменты снабжаются одинаковыми именами, при этом нумерация должна строго совпадать, поскольку на этом основан поиск соответствующих фрагментов. На запрос пользователя фрагменты текста, которые снабжены клипами.

Разметка текстовых фрагментов осуществляется по стандартам НКРЯ, а также в соответствии со стандартами мультимедийного корпуса. Каждый текстовый фрагмент снабжен метатекстовой, морфологической, семантической, акцентологической аннотацией; каждой реплике приписана социологическая аннотация (сведения о говорящем). Орфоэпическая разметка и разметка акцентной структуры слова осуществляются в процессе индексации в автоматическом режиме.

Доступ к текстам устной научной и политической речи осуществляется по адресу <http://ruscorpora.ru/search-murco.html>. В дальнейшем планируется внести изменения в интерфейс МУРКО, которые сделают возможным отбор для исследовательских целей пользовательских подкорпусов научной и политической речи, сопоставимых по объему, по однородности состава (диалоги или монологи), жанровой и тематической принадлежности.

Создаваемые специализированные корпуса предоставят возможность исследовать 1) естественную устную русскую речь в мультимедийной форме; 2) особенности диалогической и монологической устной речи; 3) жанровое своеобразие и вариативность устного научного и политического дискурса в зависимости от тематики, типа аудитории (выступление на профессиональной конференции, в учебной аудитории, в СМИ), индивидуальной манеры говорящих; 4) своеобразие научного и политического монолога и диалога в сопоставительном аспекте, сравнивая их по разным параметрам, например жанровому (доклад на научной конференции и на заседании правительства, интервью с политиком и с ученым, научная дискуссия и политические дебаты и т.д.).

Литература

1. *Гришина Е.А.* (2015), Мультиmodalный модуль в составе Национального корпуса русского языка. Труды Института русского языка им. В.В. Виноградова, № 6, с. 65–88
2. *Гришина Е.А.* (2015), О русском жестикуляционном отрицании. Труды Института русского языка им. В.В. Виноградова, № 6, с. 556–604.
3. *Гришина Е.А.* (2009) Мультимедийный русский корпус (МУРКО): проблемы аннотации. Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб, с. 150–174.
4. *Лантева О.А.* (2007), Живая русская речь с телеэкрана. Разговорный пласт телевизионной речи в нормативном аспекте. Изд. 6-е. М.: Издательство ЛКИ.
5. *Лантева О.А.* (ред.). (1985), Современная русская устная научная речь. Общие свойства и фонетические особенности. Т. I. Красноярск; Изд-во Красноярского ун-та.
6. *Паришина О.Н.* (2012), Российская политическая речь: Теория и практика Изд. 3-е. М.: Книжный дом "ЛИБРОКОМ".
7. *Савчук С.О., Махова А.А.* (2017), Мультимедийный модуль в составе Национального корпуса русского языка: направления развития. Анализ разговорной русской речи (АР³-2017): Труды седьмого междисциплинарного семинара. СПб.: Политехника-принт, с. 83–89.

References

1. *Grishina E. A.* (2015), Mul'timodal'nyy modul' v sostave Natsional'nogo korpusa russkogo yazyka [Multimodal module as part of the Russian National Corpus]. Trudy Instituta russkogo yazyka im. V.V. Vinogradova [Proceedings of the V.V. Vinogradov Russian Language Institute], 6, pp. 65–88

2. *Grishina E. A.* (2015), О русском жестикulyatsionном отрицании [On gestural negation in the Russian language] // Trudy Instituta russkogo yazyka im. V.V. Vinogradova [Proceedings of the V.V. Vinogradov Russian Language Institute], 6, pp. 556–604.

3. *Grishina E. A.* (2009) Mul'timediynyy russkiy korpus (MURKO): problemy annotatsii [Multimodal Russian Corpus (MURCO): Problems of Annotation]. Natsional'nyi korpus russkogo yazyka: 2006—2008. Novye rezul'taty i perspektivy [The National Corpus of the Russian Language: 2006-2008. New results and perspectives]. St. Petersburg, Nestor-Istoriya Publ., pp. 150–174.

4. *Lapteva O. A.* (2007), Zhivaya russkaya rech' s teleekrana. Razgovornyi plast televizionnoi rechi v normativnom aspekte. [Russian speech from a television screen. Conversational layer of television speech in the normative aspect]. 6th ed. Moscow, LKI Publ.

5. *Lapteva O. A.* (ed.) (1985), Sovremennaya russkaya ustnaya nauchnaya rech'. Obshchie svoystva i foneticheskie osobennosti [Modern Russian oral scientific speech. General properties and phonetic features]. Vol. 1. Krasnoyarsk, Krasnoyarsk Univ. Press.

6. *Parshina O. N.* (2011), Rossiyskaya politicheskaya rech': Teoriya i praktika [Russian Political Speech: Theory and Practice]. 3d ed. Moscow, LIBROKOM Publ.

7. *Savchuk S.O., Makhova A.A.* (2017), Mul'timediynyi modul' v sostave Natsional'nogo korpusa russkogo yazyka: napravleniya razvitiya [Multimedia module in the Russian National Corpus: directions of development]. Sed'moi mezhdistsiplinarnyi seminar “Analiz razgovornoy russkoy rechi” (AR³–2017) [7th interdisciplinary seminar “Russian Speech Analysis”]. St. Petersburg: Politehnika-print Publ., pp. 83–89.

Савчук Светлана Олеговна

Институт русского языка им. В.В. Виноградова РАН

Svetlana Savchuk
V.V. Vinogradov Russian Language Institute, RAS
E-mail: savsvetlana@mail.ru