

УДК 81'23

(В)НАЧАЛЕ БЫЛО СЛОВО: О ЕДИНИЦАХ АНАЛИЗА ДИСКУРСА НА ПРИМЕРЕ КОРПУСА РЕЧИ УЧИТЕЛЕЙ¹

Владислав Иванович Зубов

младший научный сотрудник филологического факультета

Санкт-Петербургский государственный университет

199034, Санкт-Петербург, Университетская наб., 7/9. v.zubov@spbu.ru

Елена Игоревна Риехакайнен

к. филол. н., доцент кафедры общего языкознания им. Л.А. Вербицкой

Санкт-Петербургский государственный университет

199034, Санкт-Петербург, Университетская наб., 7/9. e.riehakajnen@spbu.ru

В статье обсуждаются теоретические и методологические вопросы, возникающие в ходе сегментации устной речи с целью ее дальнейшего многоуровневого анализа. На примере аннотирования записей речи школьных учителей – носителей русского языка рассматривается вопрос о единицах макро- и микроструктуры педагогического дискурса. Предлагается деление урока на макроблоки в зависимости от решаемой учителем педагогической задачи. На уровне микроструктуры размечаются все межпаузальные интервалы и в качестве основной единицы выделяются фрагменты речи, находящиеся между синтаксическими паузами. Минимальными единицами разметки в разрабатываемом нами корпусе речи учителей являются словоупотребления, которым автоматически или вручную приписываются морфологические, лексические, прагматические и прочие лингвистические характеристики.

Ключевые слова: педагогический дискурс; устная речь; корпусный подход; единицы дискурса; русский язык.

1. Введение

Любой дискурс – это сложное многоуровневое явление. Принципы его описания и детальность анализа зависят от того, какие аспекты организации дискурса важны для конкретного исследования. Мы изучаем одну из разновидностей педагогического дискурса – уроки в средней школе. При этом нас интересует то, каким образом речевые средства, используемые учителем, могут способствовать лучшему усвоению материала учениками. Поэтому в данный момент мы сконцентрированы на многоаспектном лингвистическом анализе речи учителей.

В качестве материала мы используем аудиозаписи уроков в 5–8 классах общеобразовательных школ трех городов России: Нижнего Новгорода, Липецка и Костромы. Записи были сделаны в 2022 г. и предоставлены нам нашими партнерами из ООО «СберОбразование».

В статье мы изложим наши представления о том, какие единицы дискурса необходимо выделять и описывать, чтобы получить представление о лингвистических параметрах речи учителей. Важно отметить, что на данный момент нет еди-

ной точки зрения относительно того, что является единицей дискурса (см. обсуждение в [Скребцова 2020: 56–60]). Исходя из понимания дискурса как многоуровневого феномена, мы различаем макроструктуру (суперструктуру, глобальную структуру) и микроструктуру (локальную структуру) дискурса. Первая из них предполагает членение на крупные составляющие (эпизоды, группы реплик, абзацы и т. п.), а вторая – на минимальные единицы анализа дискурса [Кибрик 2003: 30]. В статье мы рассмотрим единицы каждой из этих структур отдельно.

Практическим результатом нашей работы должен стать корпус речи учителей, поэтому в статье мы также уделим внимание техническим аспектам, а именно тому, насколько реалистично и оправданно выделение тех или иных единиц при создании корпуса. Для аннотирования корпуса мы используем программу Praat (<https://www.fon.hum.uva.nl/praat>).

2. Единицы макроструктуры дискурса в корпусе речи учителей

В лингвистических исследованиях дискурса часто предлагается деление дискурса на моноло-

гический и диалогический (полилог в данном случае мы рассматриваем как разновидность диалога), каждый из которых имеет свои структурные особенности. В любом уроке, вне всякого сомнения, можно выделить такие части, но это деление кажется достаточно формальным, поскольку, по всей видимости, нет таких педагогических задач, которые бы были закреплены исключительно за диалогом или монологом. Более оправданным с содержательной точки зрения представляется деление урока на блоки в зависимости от решаемой учителем педагогической задачи. В качестве рабочего перечня мы выбрали следующий:

- 1) организационный этап (постановка цели и задач);
- 2) актуализация знаний с прошлых занятий (повторение и проверка знаний);
- 3) введение учителем нового материала;
- 4) первичное закрепление нового материала, проверка его понимания;
- 5) назначение домашнего задания и инструктаж по его выполнению;
- 6) рефлексия (подведение итогов, обратная связь)

[Бордовская, Реан 2006: 101; Пидкасистый 2011: 194–195].

Дальнейший анализ предлагается проводить внутри каждого из этих блоков. При этом и с педагогической, и с лингвистической точки зрения важным и интересным представляется описание того, как именно строится каждый из этапов: представляет ли он собой диалог или монолог, каков объем речи учителя и учеников во время каждого из этапов урока. Единицами внутри диалога являются реплики учителя и учеников. С технической точки зрения целесообразно их выделять и описывать на разных уровнях, поскольку в процессе естественной коммуникации возможны ситуации наложения реплик коммуникантов. В монологических фрагментах деления на этом уровне не будет.

3. Единицы микроструктуры дискурса в корпусе речи учителей

В научной литературе и в практике создания корпусов устной речи представлены различные варианты деления как реплик диалога, так и монологических фрагментов на составляющие их единицы. Нередко в этом контексте упоминается синтагма, под которой чаще всего понимается фрагмент речи, характеризующийся семантическим, синтаксическим и фонетическим (интонационным) единством (о различных трактовках понятия «синтагма» см. подробнее в [Раева, Риехакайнен 2015]). Однако в естественной устной

речи возможно несовпадение смысловых, грамматических и фонетических границ, поскольку паузы, например, могут возникать между любыми словами в потоке речи. По этой причине ряд исследователей предлагает рассматривать в качестве единиц дискурса прежде всего семантико-грамматические единства. Примером таких единств являются, в частности, элементарные дискурсивные единицы (ЭДЕ) [Кибрик, Подлеская 2009; Carlson, Marcu, Okunowski 2003], которые часто описывают одно событие или состояние и соответствуют одной клаузе. При этом другие признаки (тональный рисунок, темп речи, рисунок громкости, пограничные паузы, акцентный центр) также учитываются при выделении и описании ЭДЕ [Кибрик, Подлеская 2009: 55–61]. В то же время опыт выделения ЭДЕ в русской спонтанной речи, описанный в [Кибрик, Подлеская 2009], показывает, что каноническими, т. е. включающими в себя лексически полный глагольный предикат, является примерно половина ЭДЕ, остальные же случаи – это разного рода неполные клаузы или, напротив, объединение нескольких клауз внутри одного просодического единства. Таким образом, так же как в случае с синтагмами, ведущим фактором при определении границ ЭДЕ оказывается просодический, а не семантико-грамматический.

В нашем корпусе мы предлагаем последовательно разделять семантико-синтаксические и фонетические единицы. Поэтому первоначально мы делим все реплики на межпаузальные интервалы (паузой при этом считается отсутствие полезного речевого сигнала не менее 100 мс, поскольку есть данные о том, что длительность даже самых коротких артикуляторных пауз редко бывает ниже этого порога [Griffiths 1991]). При этом паузы – это не только фрагменты тишины, но и т. н. заполненные паузы (см. описание используемой нами классификации пауз в [Виноградова, Прокаева, Риехакайнен 2023]). Все, что находится внутри межпаузального интервала, мы расшифровываем в орфографии и планируем сопроводить акустико-фонетической транскрипцией, как это сделано в Корпусе русской устной речи (russpeech.spbu.ru). В этом корпусе именно межпаузальные интервалы выдаются в качестве результата при поиске. Однако в ряде случаев в межпаузальный интервал может попасть лишь одно слово или несколько слов, не представляющих собой смысловое целое (о недостатках такого подхода применительно к анализу устного дискурса см. [Бурова 2021]).

При анализе дискурса кажется важным работать именно со смысловыми единствами. На данный момент мы не вводим отдельный уровень для

разметки таких единств (аналогов ЭДЕ), но при разметке характеризуем все паузы как синтаксические (т. е. те, которые возникают между синтаксически не связанными элементами или теми элементами, синтаксические связи между которыми слабые) и несинтаксические (те, которые разрывают сильные синтаксические связи) (см. подробнее [Гаспаров, Скулачёва 2004; Нигмату-

лина, Пальчиков, Риехакайнен 2014; Сонькин 2001]). При наличии такой разметки мы можем при поиске по корпусу в качестве фрагмента дискурса выдавать совокупность межпаузальных интервалов между двумя соседними синтаксическими паузами. На Рисунке такими фрагментами будут отрезки *ну а мы с вами видим* и *что Россия входила в антифранцузскую коалицию*.

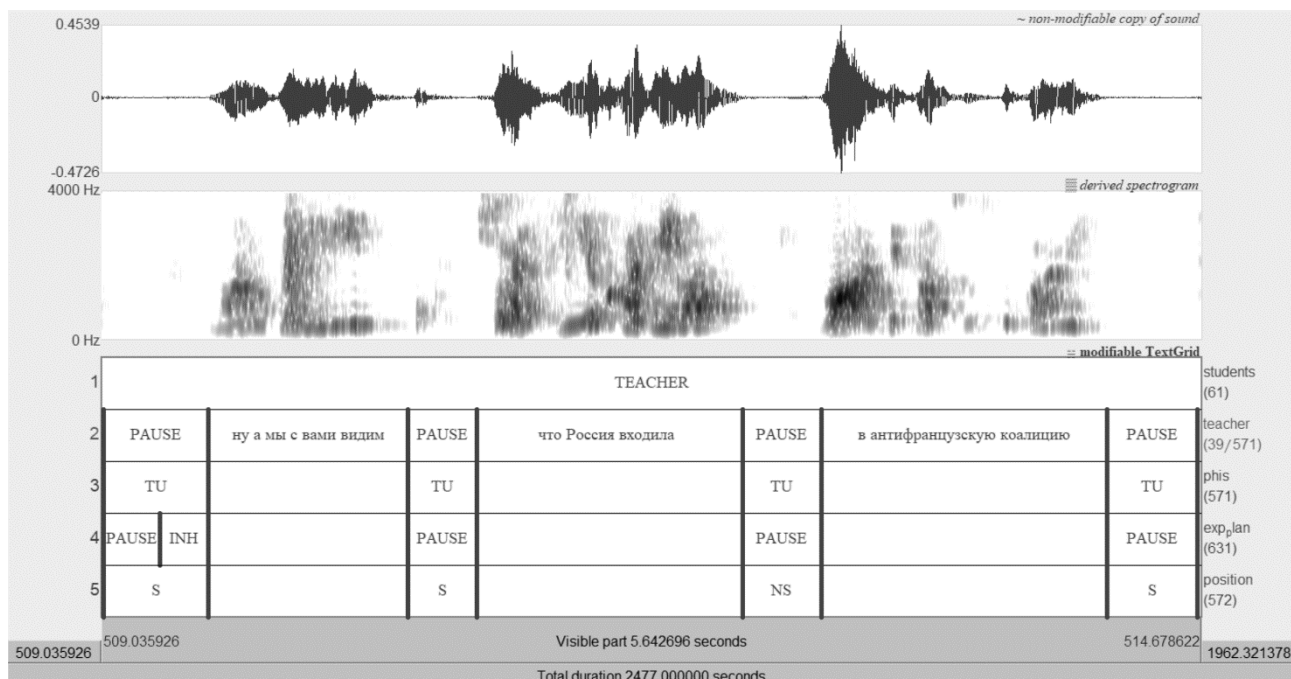


Рисунок. Пример размеченного фрагмента речи учителя
(PAUSE – пауза, TU – темпоральная незаполненная пауза, INH – вдох, S – синтаксическая пауза, NS – несинтаксическая пауза)

Таким образом, мы пытаемся при разметке соединить просодический и синтаксический уровни анализа дискурса, что схоже с тем, как предлагается сегментировать устные тексты, полученные во время интервью, в [Бурова 2021]. При этом мы сознательно уходим от использования термина ЭДЕ, поскольку не проводим на данном этапе детальный анализ всех просодических компонентов, необходимых для последовательного выделения этих единиц.

4. Минимальные единицы анализа в корпусе речи учителей

Что касается сегментации микроструктуры дискурса, то обычно в русскоязычных корпусах устных текстов такой единицей выступает словоупотребление, по аналогии с письменными корпусами, в которых в основном применяется пословный принцип разметки [Ляшевская 2016: 19–40]. Подобный подход реализован в ряде корпусов: Рассказы о сновидениях и другие корпуса звучащей речи (<http://spokencorpora.ru>), Корпус русской устной речи (<http://russpeech.spbu.ru>), Один речевой день (<https://ord.spbu.ru>), Мульти-

медийный корпус русского языка – МУРКО (<https://ruscorpora.ru/new/search-murco.html>). Сегментация речи на словоупотребления позволяет автоматически осуществить лемматизацию и морфологическую разметку, как это сделано, например, в МУРКО. Однако стоит отметить, что сегментация во всех приведенных корпусах осуществляется уже в орфографической расшифровке звучащего текста, а не непосредственно на уровне аудиозаписи. Это обусловлено рядом ограничений, препятствующих сегментации речевого потока на отдельные слова, среди которых не только неприменимость определения «от пробела до пробела» для словоупотребления в устной речи, но и такие особенности речи, как

- 1) наличие стяжений звуков на стыке словоформ (*что они* – [ʃtan'ɪ], *вот дети* – [vod'et'i] и т. п. [Нигматулина 2017]);
- 2) наличие редукции в виде выпадения целых словоформ (например, редукция предлога *в* [Насырова 2022]);
- 3) наличие так называемых неоднословных лексических единиц – составных слов, устой-

чивых сочетаний нескольких печатных слов (*потому что, во что бы то ни стало, Большой театр* и т. п. [Мустайоки, Копотев 2004; Рогожникова 1991]), которые необходимо размечать вручную в том числе из-за наличия омонимии (ср. *с одной стороны* – вводная конструкция и *с одной стороны [дома]* – несколько словоупотреблений).

Даже предварительный анализ записей речи учителей позволяет сделать вывод о значительном количестве перечисленных ограничений в устной речи учителей, несмотря на то что эта речь характеризуется большей степенью подготовленности, чем спонтанная речь. Анализ случайно выбранной минуты из объяснения на уроке истории позволил обнаружить явления редукции (*{в} начале своего правления*), стяжения (*вот эти вот* – [et'et'it], *при вступлении* – [pr'istupl'en'i]), обилие неоднословных единиц (*на самом деле, Павел Первый, Османская империя*).

В связи с этим мы предлагаем осуществлять орфографическую расшифровку только в границах межпаузальных интервалов, без ассоциирования каждого конкретного слова с отрезком звукового файла; а затем автоматически приписывать грамматические и лексические характеристики словоупотреблениям, находящимся между пробелами в орфографической записи, аналогично подходу, реализованному при создании МУРКО. Дальнейший анализ орфографической расшифровки и разметка неоднословных единиц может осуществляться в автоматическом или полуавтоматическом режиме с привлечением словарей таких единиц.

Особыми единицами анализа педагогического дискурса могут стать дискурсивные маркеры, характерные именно для речи учителей, которые могут быть реализованы как в виде отдельного словоупотребления, так и в виде неоднословных единиц. Подобные маркеры подробно описаны на материале русской устной спонтанной речи (ср. корпус «Один речевой день», а также словарь-монографию [Прагматические маркеры повседневной речи 2021]), поэтому мы предлагаем учесть этот опыт при разработке и разметке корпуса устной речи учителей и на отдельном уровне разметки отмечать дискурсивные маркеры. Количественное и функциональное описание таких маркеров в профессиональной речи в сопоставлении с естественной устной коммуникацией может стать отдельным направлением исследований педагогического дискурса.

5. Заключение

В работе мы рассмотрели принципы сегментации и аннотирования педагогического дискурса

на примере записей речи учителей на уроках в средней школе. При разметке педагогического дискурса на уровне макроструктуры предполагается деление урока на блоки в зависимости от решаемой учителем педагогической задачи. При анализе дискурса на уровне микроструктуры в качестве основной единицы анализа предлагается выделять фрагменты речи, находящиеся между синтаксическими паузами: такой подход позволит учесть и просодическую, и смысловую организацию речи [Бурова 2021]. В качестве минимальных единиц анализа речи учителей мы выделяем словоупотребления по аналогии с тем, как это организовано, например, в Мультимедийном корпусе русского языка, с дальнейшей аннотацией словоупотреблений на предмет их морфологических, прагматических и прочих лингвистических характеристик автоматическим или полуавтоматическим способом.

Мы надеемся, что создание корпуса речи учителей и дальнейший лингвистический анализ дискурсивных практик, реализуемых на уроке, позволит нам лучше понять, каким образом происходит взаимодействие учителя и ученика на уроке, какие речевые средства характерны для речи педагога.

Примечание

¹ Исследование выполняется в рамках проекта СПбГУ (ID 101747352).

Список литературы

- Бордовская Н.В., Реан А.А.* Педагогика: учебник для вузов. СПб: Питер, 2006. 299 с.
- Бурова Е.Е.* Сегментация речи при создании корпуса устных текстов // Филологические науки. Вопросы теории и практики. 2021. Т. 14, вып. 9. С. 2657–2663.
- Виноградова Ю.С., Прокаева В.О., Риехакайнен Е.И.* Паузы бывают разные: многомерная классификация пауз для разметки корпусов русской устной речи // Русская речь. 2023. (В печати).
- Гаспаров М.Л., Скулачёва Т.В.* Статьи о лингвистике стиха. М.: Языки славянской культуры, 2004. 288 с.
- Кибрик А.А.* Анализ дискурса в когнитивной перспективе: дис. ... д-ра филол. наук. М., 2003. 90 с.
- Кибрик А.А., Подлесская В.И.* Рассказы о сновидениях: корпусное исследование устного русского дискурса. М.: Языки славянских культур, 2009. 736 с.
- Ляшевская О.Н.* Корпусные инструменты в грамматических исследованиях русского языка. М.: Языки славянских культур: Рукописные памятники Древней Руси, 2016. 520 с.
- Мустайоки А., Копотев М.* К вопросу о статусе эквивалентов слова типа *потому что*, в зави-

симости от, к сожалению // Вопросы языкознания. 2004. № 3. С. 88–107.

Насырова Е.В. Редукция предлога «в» в русской устной речи // Фонетический лицей. СПб.: Скифия-принт, 2022. Вып. 7. С. 78–82.

Нигматулина Ю.О. Стяжение звуков на стыке словоформ в звучащей речи (на материале русского языка) // Вестник СПбГУ. Язык и литература. 2017. Т. 14. Вып. 1. С. 76–88.

Нигматулина Ю.О., Пальчиков Г.Г., Риехакайнен Е.И. Паузация и синтаксические связи в русской спонтанной речи // Проблемы порождения и восприятия речи: Материалы XII выездной школы-семинара (29–30 ноября 2013 г., г. Череповец) / Отв. ред. Е.В. Грудева. Череповец: ЧГУ, 2014. С. 35–40.

Пидкасистый П.И. Педагогика: учебник и практикум для вузов. М.: Юрайт, 2011. 502 с.

Прагматические маркеры русской повседневной речи: словарь-монография / сост., отв. ред. и автор предисловия Н.В. Богданова-Бегларян. СПб: Нестор-История, 2021. 528 с.

Раева О.В., Риехакайнен Е.И. Синтагма и сегментация устной спонтанной речи // Вестник

Пермского университета. Российская и зарубежная филология. 2015. Т. 32, вып. 4. С. 14–20.

Рогожников Р.П. Словарь эквивалентов слова: наречные, служебные, модальные единства. М.: Русский язык, 1991. 254 с.

Скребецова Т.Г. Лингвистика дискурса: структура, семантика, прагматика. Курс лекций. М.: Языки славянских культур, 2020. 312 с.

Сонькин В.В. Белый пятистопный ямб – полигон синтаксической свободы: к вопросу о межстрочных анжамбманах в русской поэзии XX века // Славянский стих: лингвистическая и прикладная поэтика: матер. междунар. конф. / ред. М.Л. Гаспаров, А.В. Прохоров, Т.В. Скулачёва. М., 2001. С. 187–194.

Carlson L., Marcu D., Okurowski M.E. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory // Current and New Directions in Discourse and Dialogue / Eds. J. Kuppevelt, R.W. Smith. Springer, 2003. P. 85–112.

Griffiths R. Pausological Research in an L2 Context: A Rationale, and Review of Selected Studies // Applied Linguistics. 1991. Vol. 12, No. 4. P. 345–64.

(AT) THE BEGINNING THERE WAS A WORD: ABOUT UNITS OF DISCOURSE ANALYSIS IN THE TEACHERS' SPEECH CORPUS

Vladislav I. Zubov

Junior Researcher, Philological Faculty
Saint Petersburg State University

Elena I. Riekhakaynen

Associate Professor, Department of General Linguistics
Saint Petersburg State University

The paper discusses theoretical and methodological issues that arise during the segmentation of speech for the purpose of its further multi-level analysis. We focus on the units of macro- and microstructure of educational discourse that we are going to annotate in the corpus of Russian-speaking school teachers. We propose to divide the lesson into macroblocks depending on the pedagogical task being solved by the teacher. At the microstructure level, all inter-pause intervals are marked and the fragments of speech located between syntactic pauses are regarded as main discourse units. Tokens are the smallest units of the analysis in our corpus. They are automatically or manually assigned morphological, lexical, pragmatic and other linguistic characteristics.

Keywords: educational discourse; speech; corpus-based study; discourse units; the Russian language.