

**РОССИЙСКАЯ АКАДЕМИЯ НАУК**  
**Институт русского языка имени В. В. Виноградова**

---

**RUSSIAN ACADEMY OF SCIENCES**  
**V. V. Vinogradov Russian Language Institute**



**Труды  
Института русского языка  
им. В. В. Виноградова**

**XXI**

**Национальный корпус русского языка:  
исследования и разработки**

МОСКВА  
2019

Издание основано в 2013 г.

Главный редактор А. М. Молдаван

**РЕДАКЦИОННЫЙ СОВЕТ**

А. Е. Аникин, д. ф. н., академик РАН (Новосибирск);  
Ю. Д. Апресян, академик РАН, профессор (Москва, Россия);  
С. Вакарелийская, PhD, профессор (Орегон, США);  
Ж. Ж. Варбот, доктор филол. наук, профессор (Москва, Россия).  
Бьёрн Вимер, доктор филологии, профессор (Майнц, Германия);  
М. Гардзанини, PhD, профессор (Флоренция, Италия);  
А. А. Гиппиус, член-корреспондент РАН, профессор (Москва, Россия);  
О. Йокояма, PhD, профессор (Лос-Анджелес, США);  
Я. Какридис, Dr. habil., профессор (Берн, Швейцария);  
М. Л. Каленчук, доктор филол. наук, профессор (Москва, Россия);  
А. П. Майоров, д. ф. н., профессор (Улан-Удэ, Россия);  
О. Младенова, PhD, профессор (Калгари, Канада);  
Туре Нессет, доктор филологии, профессор (Тромсё, Норвегия);  
В. А. Плунгян, член-корреспондент РАН, профессор (Москва, Россия);  
А. Рабус, Dr. habil., профессор (Фрайбург, Германия);  
Ф. Б. Успенский, д. ф. н., член-корреспондент РАН, профессор (Москва, Россия);  
М. Флайер, PhD, профессор (Кембридж, США);  
Вацлав Чермак, доктор филологии (Прага, Чехия);  
А. Д. Шмелев, доктор филол. наук, профессор (Москва, Россия).

**ОТВЕТСТВЕННЫЙ РЕДАКТОР ВЫПУСКА**

В. А. Плунгян, академик РАН, профессор (Москва, Россия).

**РЕДАКЦИОННАЯ КОЛЛЕГИЯ ВЫПУСКА**

Е. В. Рахилина, д. филол. наук (Москва, Россия)  
С. О. Савчук, к. филол. наук (Москва, Россия).

Выходит 4 раза в год

Адрес редакции:

119019, Москва, ул. Волхонка, д. 18/2

E-mail: [ruslang@ruslang.ru](mailto:ruslang@ruslang.ru)

Издательство зарегистрировано Федеральной службой по надзору  
в сфере связи, информационных технологий и массовых коммуникаций.

Свидетельство о регистрации ПИ № ФС 77-57258

**Proceedings  
of the V. V. Vinogradov  
Russian Language Institute**

**XXI**

**The Russian National Corpus: Research and development**

MOSCOW

2019

The Journal was founded in 2013

Editor-in-Chief Alexander M. Moldovan

EDITORIAL BOARD

- Aleksandr E. Anikin, D.Sc., Full Member of the RAS (Novosibirsk, Russia);  
Yury D. Apresyan, D.Sc., Full Member of the RAS, Professor (Moscow, Russia);  
Václav Čermák, Ph.D., Professor (Prague, Czech Republic);  
Michael S. Flier, Ph.D., Professor (Cambridge, USA);  
Marcello Garzaniti, D. Sc., Professor (Florence, Italy);  
Alexey A. Gippius, D.Sc., Corresponding Member of the RAS, Professor (Moscow, Russia);  
Yannis Kakridis, D. Sc., Professor (Bern, Switzerland);  
Maria L. Kalenchuk, D. Sc., Professor (Moscow, Russia);  
Alexandr P. Mayorov, D.Sc., Professor (Ulan-Ude, Russia);  
Olga Mladenova, Ph.D, Professor em. (Calgary, Canada);  
Alexandr M. Moldovan, D.Sc., Full Member of the RAS, (Moscow, Russia);  
Tore Nettet, D.Sc., Professor, (Tromsø, Norway);  
Vladimir A. Plungian, D.Sc., Full Member of the RAS, Professor (Moscow, Russia);  
Achim Rabus, D. Sc., Professor (Freiburg, Germany);  
Alexey D. Shmelev, D.Sc., Professor (Moscow, Russia);  
Fjodor B. Uspensky, D.Sc., Corresponding Member of the RAS, Professor (Moscow, Russia);  
Cynthia M. Vakareliyska, Ph.D., Professor (Oregon, USA);  
Zhanna Zh. Varbot, D.Sc., Professor (Moscow, Russia);  
Björn Wiemer, D. Sc., Professor, (Mainz, Germany);  
Olga T. Yokoyama, Ph.D., Distinguished Professor (Los Angeles, USA).

CHIEF EDITORS OF THE ISSUE

- A.V. Zanadvorova, PhD (Moscow, Russia);  
R.I. Rozina, D.Sc., (Moscow, Russia).

EDITORIAL BOARD OF THE ISSUE

- A.V. Zanadvorova, PhD (Moscow, Russia);  
R.I. Rozina, D.Sc., (Moscow, Russia).  
N.N. Rozanova, PhD (Moscow, Russia);  
L.P. Krysin, D.Sc., (Moscow, Russia).

Address:

18/2, Volkhonka street, Moscow, 119019  
E-mail: [ruslang@ruslang.ru](mailto:ruslang@ruslang.ru)

The journal is registered by the The Federal service for supervision  
of communications, information technology, and mass-media.  
Registration certificate ПИ № ФС 77–57258.

## СОДЕРЖАНИЕ

Предисловие (В. А. Плунгян) . . . . .	11
---------------------------------------	----

### ***I. Развитие Национального корпуса русского языка и других корпусов***

Е. С. Иншакова, Л. Л. Иомдин, Л. Г. Митюшин, В. Г. Сизов, Т. И. Фролова, Л. Л. Цинман (Москва) СинТагРус сегодня . . . . .	14
Д. В. Сичинава (Москва) Параллельные подкорпуса: новые языки и новые задачи . . . . .	42
Б. В. Орехов, С. О. Савчук (Москва) Акцентологический корпус как инструмент для исследования русского ударения . . . . .	62
А. Е. Поляков (Москва) Орфография Острожской библии в контексте церковнославянского корпуса . . . . .	84
Н. В. Богданова-Бегларян, О. В. Блинова, Г. Я. Мартыненко, Т. Ю. Шерстинова (Санкт-Петербург) Корпус русского языка повседневного общения «Один речевой день»: текущее состояние и перспективы . . . . .	101
Н. В. Богданова-Бегларян, О. В. Блинова, К. Д. Зайдес, Т. Ю. Шерстинова (Санкт-Петербург) Корпус «Сбалансированная аннотированная текстотека» (САТ): изучение специфики русской монологической речи . . . . .	111

### ***II. Национальный корпус русского языка в научных исследованиях***

Г. И. Кустова (Москва) Относительные наречия и синтаксическая иерархия . . . . .	127
С. Ю. Жукова, Б. В. Орехов, Е. В. Рахилина (Москва) Дискурсивные формулы русского языка: диахронический подход . . . . .	145
В. П. Захаров (Санкт-Петербург) Категоризация прилагательных цвета в русских поэтических текстах (корпусное исследование) . . . . .	165
К. М. Корчагин (Москва) Цезура в русском 5-стопном ямбе и ее роль в ритмической эволюции этого размера . . . . .	182
Д. О. Добровольский (Москва) «Показатели безразличия» в русско-немецком корпусе параллельных текстов . . . . .	203

### ***III. Слово и жест: вопросы мультимодальной коммуникации***

Т. Е. Янко (Москва) Просодия и жестикуляция . . . . .	214
О. В. Федорова (Москва) О коммуникативной функции взгляда . . . . .	225

Ю. В. Николаева (Москва)	
Жестовые сбои в диалоге . . . . .	242
П. А. Бычкова, Е. В. Рахилина, Е. А. Слепак (Москва)	
Дискурсивные формулы, полисемия и жестовое маркирование . . . . .	257
И. М. Кобозева, О. О. Иванова, Л. М. Захаров (Москва)	
К мультимодальному моделированию верификативных дискурсивных маркеров в русском диалоге . . . . .	285
А. А. Зинина, Л. Я. Зайдельман, Н. А. Аринкин, А. А. Котов (Москва)	
Подходы к классификации коммуникативных жестов: теория и приложение к роботам-компаньонам . . . . .	300
С. И. Переверзева, Н. А. Ермолаева, А. В. Зуева, Е. А. Слепак (Москва)	
De profundis: проблемы глубокой разметки мультимедийного русского корпуса и пути решения . . . . .	320



## CONTENT

Preface (V. A. Plungian) . . . . .	11
------------------------------------	----

### ***I. The development of the Russian National Corpus and other Corpora***

Ye.S. Inshakova, L. L. Iomdin, L. G. Mityushin, V. G. Sizov, T. I. Frolova, L. L. Tsinman (Moscow)	
The SynTagRus today . . . . .	14
Dmitri V. Sitchinava (Moscow)	
Parallel Sub-Corpora: New Languages and New Tasks. . . . .	42
B. V. Orekhov, S. O. Savchuk (Moscow)	
Accentological Corpus as a tool for studying the Russian accent . . . . .	62
A. E. Polyakov (Moscow)	
Spelling of the Ostrog Bible in context of the Church slavonic corpus . . . . .	84
N. V. Bogdanova-Beglarian, O. V. Blinova, G. Ya. Martynenko, T. Yu. Sherstinova (Saint Petersburg)	
Russian everyday speech corpus “One day of speech”: current state and perspectives. . . . .	101
N. V. Bogdanova-Beglarian, O. V. Blinova, K. D. Zaides, T. Ju. Sherstinova (Saint Petersburg)	
Corpus “Balanced annotated text collection (textotec)” (SAT): Studying the specificity of Russian monological speech . . . . .	111

### ***II. The Russian National Corpus in scientific research***

G. I. Kustova (Moscow)	
Relative adverbs and syntactic hierarchy . . . . .	127
S. Yu. Zhukova, B. V. Orekhov, E. V. Rakhilina (Moscow)	
Discursive Russian language formulas: A diachronic approach . . . . .	143
V. P. Zakharov, A. Ts. Masevitch (Moscow)	
Categorization of adjectives of color in the russian poetical texts (A corpus-based study). . . . .	165
K. M. Korchagin (Moscow)	
Caesura in the Russian iambic pentameter and its impact on the rhythmic evolution of this meter . . . . .	182
D. O. Dobrovol'skij (Moscow)	
“Indicators of indifference” in the Russian-German parallel corpus . . . . .	203

### ***III. Word and Gesture: Multimodal communication issues***

T. E. Yanko (Moscow)	
Prosody and gestures . . . . .	214
O. V. Fedorova (Moscow)	
On the communicative function of the gaze. . . . .	223
Yu. V. Nikolaeva (Moscow)	
Gestural disfluencies in dialogue . . . . .	242
P. A. Bychkova, E. V. Rakhilina, E. A. Slepak (Moscow)	
Discourse formulae, polysemy and gesture marking . . . . .	257

Irina Kobozeva, Olga Ivanova, Leonid Zakharov (Moscow)	
Towards multimodal modelling of verificational discourse markers in Russian dialog . . . . .	285
A. Zinina, L. Zaydelman, N. Arinkin, A. Kotov (Moscow)	
Approaches to the classification of communicative gestures: theory and application to companion robots . . . . .	300
S. I. Pereverzeva, N. A. Ermolaeva, A. V. Zueva, E. A. Slepak (Moscow)	
De profundis: the problems of the deep annotaton of the multimodal russian corpus and the ways of resolution . . . . .	320

## ПРЕДИСЛОВИЕ

Настоящий сборник продолжает серию традиционных публикаций, освещающих состояние Национального корпуса русского языка на текущем этапе и отражающих актуальную проблематику корпусной лингвистики в целом. Внимательный читатель может заметить, что формат и тематика таких сборников менялась со временем: если в первых сборниках основной акцент делался на состав и структуру корпуса, а также на технические и теоретические проблемы, связанные с его созданием, то в последующих сборниках всё большее место начинают занимать исследования, посвященные инструментальным возможностям корпуса. Иначе говоря, происходит постепенный переход от описания корпуса к обсуждению результатов, которые можно получить с помощью корпуса. Разумеется, статьи о структуре корпуса продолжают появляться, но теперь они в первую очередь отражают новые корпусные проекты и новые типы корпусов.

Эта тенденция отчетливо видна и в настоящей публикации. Один из трех разделов сборника посвящен обзору различных корпусных проектов (как в составе НКРЯ, так и вне его), тогда как два других содержат оригинальные исследования, опирающиеся на корпусные инструменты и методики. Охарактеризуем содержание разделов несколько подробнее.

Корпусный раздел открывается статьей большого коллектива сотрудников ИППИ РАН (Е. С. Иншакова, Л. Л. Иомдин и др.), в которой дается характеристика синтаксически размеченного подкорпуса НКРЯ «СинТагРус» на современном этапе. В настоящее время СинТагРус остается единственным синтаксически размеченным подкорпусом (хотя внедрение упрощенной синтаксической разметки на всем объеме основного подкорпуса входит в планы разработчиков НКРЯ); его размеры невелики, но всё время своего существования этот корпус динамично развивается. В статье Д. В. Сичиной описывается современное состояние одного из важнейших модулей НКРЯ — подкорпуса параллельных текстов. Важность параллельных подкорпусов для научных исследований контрастивного и типологического плана давно уже не нужно специально доказывать. Параллельные корпуса в составе НКРЯ — один из самых быстрорастущих разделов корпуса; увеличивается как количество языков (из последних пополнений — латышский, финский, китайский), так и объем отдельных подкорпусов и качество разметки. В статье Б. В. Орехова и С. О. Савчук дается характеристика другого инновационного проекта — акцентологического подкорпуса НКРЯ, объединяющего данные поэтического подкорпуса и подкорпуса устных текстов (именно эти два массива текстов дают достоверные

сведения о реальном ударении в русских словоформах). В статье показано, каким образом данные этого подкорпуса могут быть использованы для исследования русского ударения, в том числе и в плане диахронических изменений.

Ещё один относительно недавний проект в составе НКРЯ — подкорпус церковнославянского языка, включающий современные литургические тексты, используемые в русской православной церкви. О некоторых нетривиальных проблемах (как технических, так и содержательных), связанных с обработкой подобных текстов, рассказывает статья А. Е. Полякова «Орфография Острожской библии в контексте церковнославянского корпуса».

Две оставшиеся статьи данного раздела обобщают корпусный опыт наших петербургских коллег, специализирующихся в изучении устной коммуникации. Совместная статья Н. В. Богдановой-Бегларян, О. В. Блиновой, Г. Я. Мартыненко и др. описывает текущее состояние и перспективы развития хорошо известного корпуса устной речи «Один речевой день»; другая статья группы петербургских авторов (Н. В. Богданова-Бегларян, О. В. Блинова, К. Д. Зайдес и др.) описывает относительно менее известный проект устного корпуса, задуманного как корпус русской монологической речи: проект называется «Сбалансированная аннотированная текстовая» и открывает интересные перспективы.

Второй раздел сборника содержит, главным образом, результаты исследований, полученных с помощью корпусных данных и методик; многие из этих результатов интересны и нетривиальны и хорошо иллюстрируют разнообразные возможности корпусного подхода. Статья Г. И. Кустовой «Относительные наречия и синтаксическая иерархия» иллюстрирует плодотворность привлечения корпусного материала к классическим проблемам русской синтаксической семантики. Совместная статья С. Ю. Жуковой, Б. В. Орехова и Е. В. Рахилиной привлекает внимание к такому своеобразному феномену, как (в авторской терминологии) «дискурсивные формулы», т. е. особого рода выражения, используемые в диалогах в качестве реплик-реакций на слова собеседника. В статье основное внимание уделяется диахроническому аспекту исследования дискурсивных формул (что корпусные данные как раз хорошо позволяют сделать): оказывается, даже в русском языке XIX века стратегии использования дискурсивных формул (и само их значение) в ряде аспектов отличались от современных. Две статьи данного раздела написаны на материале продолжающего интенсивно пополняться поэтического подкорпуса НКРЯ: в статье В. П. Захарова «Категоризация прилагательных цвета в русских поэтических текстах» речь идет о поэтической лексике, тогда как в статье одного из основных разработчиков поэтического корпуса К. М. Корчагина «Цезура в русском 5-стопном ямбе и ее роль в ритмической эволюции» анализируются проблемы русской поэтической метрики. Привлечение корпусного материала и в этом случае оказывается наиболее плодотворным для анализа диахронического аспекта такого сложного и малоизученного явления, как стихотворная цезура. Статья Д. О. Добровольского написана на материале русско-немецкого параллельного корпуса и прекрасно иллюстрирует решение одной частной задачи из классической области

контрастивного лексикологического исследования: автор обсуждает выражение «показателей безразличия» в обоих языках.

Особой тематике, важной для Национального корпуса, посвящен третий раздел сборника: это мультимодальная коммуникация, предполагающая изучение не только устного дискурса, но и сопровождающей речь жестикуляции. Энтузиастом и открывателем этой темы в Национальном корпусе была Елена Александровна Гришина (1958-2016), создавшая русский мультимедийный подкорпус, на основе которого ею была подготовлена монография «Русская жестикуляция с лингвистической точки зрения. Корпусные исследования», опубликованная уже посмертно. Большинство статей настоящего раздела написано по материалам выступлений на конференции «Слово и жест» памяти Е. А. Гришиной (февраль 2018 г.). Статья Т. Е. Янко посвящена взаимодействию просодии и жестикуляции, статья О. В. Федоровой — исследованию коммуникативной функции взгляда, статья Ю. В. Николаевой — особенностям нарушения жестикуляции, сопровождающей спонтанную речь (так называемым жестовым сбоям). Статья П. А. Бычковой, Е. В. Рахилиной и Е. А. Слепак рассматривает уже упоминавшиеся дискурсивные формулы под углом зрения жестикуляции и показывает, что жестикуляция может быть использована для разграничения омонимичных или полисемичных конструкций. Статья И. М. Кобозевой, О. О. Ивановой и Л. М. Захарова посвящена жестовым компонентам так называемых верификативных дискурсивных маркеров (используемых для подтверждения или опровержения информации в диалоге). В интересной статье А. А. Зининой и др. жестовая проблематика анализируется применительно к задачам обучения роботов. Наконец, статья С. И. Переверзевой и др. касается непосредственного продолжения работы над разметкой русского мультимедийного корпуса Е. А. Гришиной, на основе и в развитие идей создателя этого корпуса.

Таким образом, статьи настоящего тома демонстрируют разнообразные направления развития Национального корпуса русского языка на современном этапе и сложность теоретических и описательных задач, в настоящее время решаемых с помощью корпуса.

*В. А. Плунгян*

# І. РАЗВИТИЕ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА И ДРУГИХ КОРПУСОВ

<sup>1</sup>Е.С. Иншакова, <sup>2</sup>Л.Л. Иомдин, <sup>3</sup>Л.Г. Митюшин,

<sup>4</sup>В.Г. Сизов, <sup>5</sup>Т.И. Фролова, <sup>6</sup>Л.Л. Цинман

<sup>123456</sup>Институт проблем передачи информации им. А. А. Харкевича РАН,

<sup>2</sup>Российский государственный гуманитарный университет

(Россия, Москва)

<sup>1</sup>yes@mail.ru, <sup>2</sup>iomdin@gmail.com, <sup>3</sup>mit@iitp.ru,

<sup>4</sup>sizov@iitp.ru, <sup>5</sup>tfrolova@gmail.com, <sup>6</sup>llcinman@gmail.com

## СИНТАГРУС СЕГОДНЯ\*

В статье описывается современное состояние корпуса СинТагРус, содержащего русские тексты с морфосинтаксической разметкой. На различных стадиях работы над корпусом были введены дополнительные типы разметки: лексико-семантическая, лексико-функциональная, анафорическая и микросинтаксическая.

Данные морфосинтаксической разметки предложения включают морфологический разбор каждого слова и синтаксическую структуру предложения в виде дерева зависимостей в соответствии с моделью «Смысл ↔ Текст» И. А. Мельчука и А. К. Жолковского. Лексико-семантическая разметка предполагает, что для каждого слова указана соответствующая ему статья комбинаторного словаря русского языка. Лексико-функциональная разметка представляет собой выделение в текстах словосочетаний, допускающих интерпретацию в терминах лексических функций. Результатом анафорической разметки является маркирование антецедентов местоимений. Микросинтаксическая разметка идентифицирует встретившиеся в текстах синтаксические фраземы и некоторые нестандартные синтаксические конструкции.

Разметка новых текстов корпуса выполняется в несколько стадий. Вначале тексты обрабатываются многофункциональным лингвистическим процессором ЭТАП-3, который в автоматическом режиме вносит в них морфосинтаксическую и лексико-семантическую разметку. Затем результаты работы процессора проверяются и при необходимости корректируются специально подготовленными лингвистами-аннотаторами. После этого ЭТАП-3, используя построенные морфосинтаксические структуры, выполняет лексико-функциональную и анафорическую

---

\* Данная работа поддержана Программой Президиума РАН «Памятники материальной и духовной культуры в современной информационной среде».

разметку. Лингвисты проверяют и корректируют эти типы разметки и вручную вносят в тексты микросинтаксическую разметку.

Корпус СинТагРус используется для целей теоретико-лингвистических исследований, а также для практической лексикографии. Статистика корпуса может учитываться при автоматической обработке текстов для оптимизации принимаемых решений. Весьма перспективно использование данных корпуса в системах, основанных на машинном обучении.

*Ключевые слова:* СинТагРус, синтаксически размеченный корпус, корпус русских текстов, грамматика зависимостей, лексические функции, антецеденты местоимений, микросинтаксис, эллипсис.

## 1. Общие сведения

Глубоко аннотированный корпус русских текстов СинТагРус разрабатывается в Лаборатории компьютерной лингвистики ИППИ РАН им. А. А. Харкевича начиная с 1998 г. При подготовке корпуса используется лингвистический процессор ЭТАП-3 (подробнее о нем см. [Апресян et al. 2003]). Доступ к корпусу имеется на сайте русского национального корпуса (НКРЯ). На сайте доступны морфологическая (со снятой омонимией), синтаксическая (в терминах деревьев зависимостей) и лексико-функциональная разметка. Кроме того, для корпуса разработаны другие типы разметки: лексическая (или лексико-семантическая — разрешение лексической неоднозначности), анафорическая и разметка микросинтаксических конструкций. Эти типы разметки пока не доступны онлайн.

В состав корпуса входит более 650 художественных, публицистических, научно-популярных и новостных текстов. Разговорная речь, поэтические и технические тексты в корпусе не представлены, что позволяет добиться некоторой однородности представленного материала. На данный момент (2018 г.) в корпусе содержится более миллиона словоупотреблений, около 70 тысяч предложений. Корпус постоянно пополняется.

СинТагРус является единственным в мире полностью отредактированным экспертами-лингвистами корпусом текстов на русском языке с аннотацией на морфосинтаксическом уровне.

Этапы развития корпуса отражены в работах [Boguslavsky et al. 2000; Апресян и др. 2005; Богуславский и др. 2008(а,б); Шеманаева, Фролова 2010; Boguslavsky 2014; Дяченко и др. 2015].

Ниже перечислены и кратко рассмотрены виды разметки, уже описанные в предыдущих работах: морфологическая (раздел 2.1), синтаксическая, в том числе разметка эллиптических конструкций (разделы 2.2 и 2.3), лексико-семантическая (раздел 2.4), лексико-функциональная (раздел 2.5). Затем более подробно рассмотрены два новых типа разметки: анафорическая разметка (раздел 3.1) и разметка микросинтаксических конструкций (раздел 3.2). Кроме того, приведены сведения об использовании корпуса СинТагРус для решения исследовательских и прикладных задач (раздел 4).

## 2. Традиционные типы разметки СинТагРус'а

Разметка текстов в СинТагРус'е производится следующим образом: для подготовки к разметке текст обрабатывается программой сегментации текста, которая автоматически разбивает его на отдельные предложения; после этого каждое предложение обрабатывается морфологическим и синтаксическим анализатором многофункционального лингвистического процессора ЭТАП-3. В результате этой обработки формируется морфологическая структура для каждого слововхождения и синтаксическая структура для каждого предложения (см. разделы 2.1 и 2.2). Одновременно выполняется разрешение лексической омонимии (см. раздел 2.4). После автоматического этапа обработки все предложения проверяются лингвистом-аннотатором, который вносит необходимые изменения в структуры слов и предложений. При редактировании в предложения при необходимости вручную вносятся сведения об эллиптических конструкциях (см. раздел 2.3). Такая проверка, хотя и трудоемкая, позволяет достичь высокого уровня точности разметки. При корректировке текста постоянно вносятся уточнения в правила и словари ЭТАП-3. Лексико-функциональная (см. раздел 2.5) и анафорическая (см. раздел 3.1) разметка, а также разметка микросинтаксических конструкций (см. раздел 3.2) выполняются для предложений с построенной и откорректированной синтаксической структурой, при этом лексико-функциональная и анафорическая разметка выполняются в автоматическом режиме с последующей ручной коррекцией, а разметка микросинтаксическими конструкциями производится полностью вручную.

### 2.1. Морфосинтаксическая и лексическая разметка

При морфологической разметке каждое слововхождение снабжается сведениями о его морфологической структуре. Морфологическая структура словоформы представляет собой имя лексемы вместе с информацией о части речи и списком морфологических характеристик. Так, структура словоформы *вытесняя* имеет следующий вид: ВЫТЕСНЯТЬ, V НЕСОВ ДЕЕПР НЕПРОШ. Здесь ВЫТЕСНЯТЬ — имя лексемы, V обозначает глагол, НЕСОВ — несовершенный вид, ДЕЕПР — деепричастие, НЕПРОШ — непрошедшее (настоящее-будущее) время. Полный список частей речи, русских морфологических категорий и характеристик приводится на сайте НКРЯ в разделе «Инструкции к синтаксически размеченному корпусу».

### 2.2. Синтаксическая разметка

Результатом синтаксической разметки является дерево зависимостей, в узлах которого стоят слова предложения, а ветви помечены именами синтаксических отношений. Такое представление о синтаксической структуре предложения восходит к лингвистической модели «Смысл ↔ Текст» И. А. Мельчука



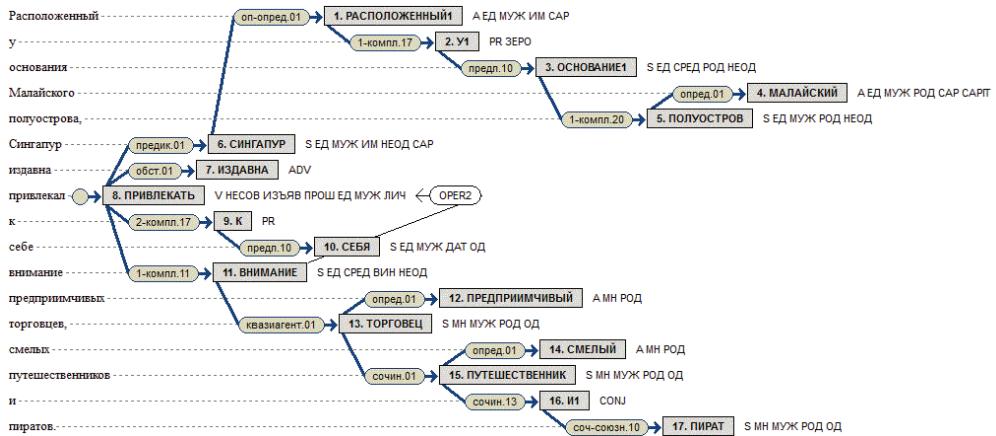


Рис. 1. Синтаксическая структура предложения (1)

и А. К. Жолковского [Жолковский, Мельчук 1967; Мельчук 1974]. В корпусе различается около 70 типов синтаксических отношений. Такая подробная разметка позволяет сделать описание синтаксической структуры более полным и лингвистически содержательным, чем в большинстве существующих корпусов. Перечень синтаксических отношений представлен в онлайн-версии СинТагРус'а на сайте НКРЯ.

Рассмотрим синтаксическую структуру следующего предложения:

(1) *Расположенный у основания Малайского полуострова, Сингапур издавна привлекал к себе внимание предприимчивых торговцев, смелых путешественников и пиратов.*

В узлах этого дерева зависимостей стоят слова предложения, представленные именами лексем (в прямоугольниках) и цепочками грамматических характеристик (справа от прямоугольников), а ветви помечены именами синтаксических отношений (в овалах).

Номера при именах синтаксических отношений указывают на номера правил в блоке синтаксических правил процессора ЭТАП-3, которые установили эти отношения. Как правило эти номера соответствуют разным типам конструкций, в которых такое отношение может быть установлено. Так, сочинительное отношение, соответствующее бессоюзному сочинению однородных членов предложения, и сочинительное отношение, соответствующее сочинению однородных членов с помощью союза И, устанавливаются разными правилами. Первый тип представлен в структуре предложения (1) отношением сочин.01, которое соединяет узлы 13 (*торговцев*) и 15 (*путешественников*). Второй тип можно видеть в том же предложении между узлами 15 (*путешественников*) и 16 (*и*), это отношение обозначено как сочин.13. Эта информация не представлена в онлайн-версии корпуса.

Номера при некоторых именах лексем являются результатом лексико-семантической разметки. О ней см. ниже в разделе 2.4.

Стрелка с овалом, содержащим запись OPER2, соответствует лексико-функциональному отношению, устанавливаемому между узлами 11 и 8. О лексико-функциональной разметке см. ниже в разделе 2.5.

Вершиной предложения (1) является глагол ПРИВЛЕКАТЬ (в форме мужского рода, единственного числа, прошедшего времени, изъявительного наклонения, несовершенного вида — вид трактуется в лингвистическом процессе ЭТАП-3 как словоизменительная характеристика). У вершинного глагола заполнены три синтаксические валентности. Предикативное синтаксическое отношение (на рисунке «предик.01») связывает глагол с подлежащим СИНГАПУР (неодушевленное существительное мужского рода в форме единственного числа, именительного падежа). Первое комплетивное отношение (1-компл.11) связывает вершинный глагол со вторым актантом — существительным ВНИМАНИЕ (неодушевленное существительное среднего рода в форме единственного числа, винительного падежа). Второе комплетивное отношение (2-компл.17) связывает глагол с вершиной предложной группы (предлог К), являющейся третьим актантом. От предлога К идет предложное отношение (предл.10) к существительному СЕБЯ (в ЭТАП-3 для него принята трактовка как одушевленного существительного мужского рода, в данном случае в единственном числе, дательном падеже). Описательно-определятельное отношение связывает подлежащее СИНГАПУР с обособленным определением «расположенный у основания Малайского полуострова», в свою очередь имеющим внутреннюю синтаксическую структуру, где «у основания» является дополнением (1-компл.17) прилагательного РАСПОЛОЖЕННЫЙ, а «полуострова» дополнением (1-компл.20) существительного ОСНОВАНИЕ1. Квазиагентивное отношение (квазиагент.01), предназначенное для указания на субъект предикатного слова, не являющегося глаголом, связывает существительное ВНИМАНИЕ с вершиной сочинительной группы *предприимчивых торговцев, смелых путешественников и пиратов*, внутри которой однородные члены предложения связаны сочинительным отношением.

### 2.3. Особые случаи синтаксической разметки.

#### Разметка предложений с эллипсисом

Чуть более 2%, почти полторы тысячи, предложений в СинТагРус'е описываются в синтаксической структуре как содержащие эллипсис. В этих случаях аннотатором принято решение о том, что для адекватного описания синтаксической структуры предложения в нее необходимо внести узлы, не соответствующие никаким словам в тексте предложения. Такие узлы называются «фантомными».

Чаще всего имена лексем для восстановленных узлов совпадают с именами лексем узлов, уже встречавшихся в предложении. Рассмотрим предложение (2).

(2) *По данным Всемирной организации здравоохранения, впервые в истории современной России средний возраст сильного пола достиг 66 лет, а у представительниц прекрасного пола — 77 лет.*

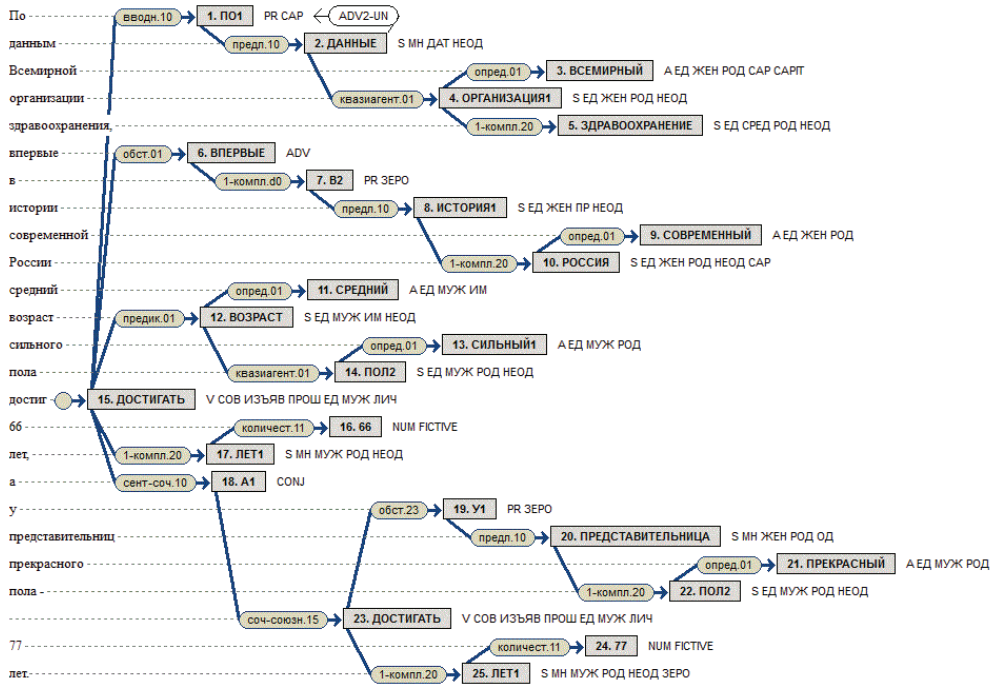


Рис. 2. Синтаксическая структура предложения (2)

В этом сложносочиненном предложении сентенциально-сочинительное отношение (сент-соч.10) связывает вершинный глагол (узел 15) ДОСТИГАТЬ в форме мужского рода, единственного числа, прошедшего времени, изъявительного наклонения, совершенного вида с сочинительным союзом А (узел 18). А сочинительно-союзное отношение (соч-союз.15) связывает этот союз А с фантомным узлом 23, у которого морфологическая структура совпадает с морфологической структурой вершинного узла.

В некоторых случаях восстанавливаются глагольные узлы с размытой семантикой — так называемые неопределенные глаголы. В этих случаях в качестве имени лексемы пишется НЕОПР-ГЛАГОЛ, а в скобках предлагается вариант, представляющий уместной гипотезой. Это интересное, но не очень распространенное явление — таких узлов 169 во всем СинТагРус'е. Рассмотрим предложение (3) и его структуру.

(3) *Остановились на таком маршруте: электричкой до Серпухова (в абсолютно переполненном вагоне), там от вокзала автобусом до пристани, потом на пароходе до Велегожа и оттуда тропкой (4 км) по прекрасному лесу вдоль Оки до заветной поляны.*

В этом предложении восстановлены четыре фантомных глагольных узла с примерным смыслом 'добираться' — 6, 16, 23, 28, все они связаны в сочинительную цепочку.

Более подробно представление эллиптических конструкций рассмотрено в [Дяченко и др. 2015].



винительным падежом (ср. *идти в лес*) и не управляющий местным падежом (ср. *гулять в лесу*). Выбран омоним полнозначного прилагательного СИЛЬНЫЙ, а не часть сложного слова (ср. *200-сильный мотор*). Для слова ПОЛ выбран омоним ‘один из двух генетически противопоставленных разрядов живых существ’, а не ‘нижняя часть помещения’ (ср. *паркетный пол*) и не ‘половина’ (ср. *полгруши*), а также не ‘мужское имя’ (ср. *Пол Хьюитт написал книгу*). Для слова ЛЕТ выбран омоним ‘годов’ (родительный падеж, множественное число), а не отглагольное существительное со значением ‘полет’ (ср. *Снежинки таяли на лету*). Для слова А выбран союз, а не омонимичная частица (ср. *А кто это сказал?*). Для слова У выбор омонима такой же, как в предложении (1).

Информация о разрешении лексической омонимии пока недоступна в онлайн-версии СинТагРус’а.

## 2.5. Лексико-функциональная разметка

При лексико-функциональной разметке обнаруживаются и отмечаются в тексте словосочетания, допускающие интерпретацию в терминах лексических функций.

Под лексическими функциями (ЛФ) в соответствии с теорией ЛФ И. А. Мельчука, А. К. Жолковского и Ю. Д. Апресяна понимаются смыслы, для которых способ выражения определяется не только самим этим смыслом, но и словом-аргументом, при котором выражается этот смысл. Так, значение ЛФ MAGN ‘высокая степень или интенсивность’ выражается в русском языке прилагательным ПРОЛИВНОЙ при существительном ДОЖДЬ и прилагательным ВОЛЧИЙ при существительном ГОЛОД. Выше в предложении (1) значением ЛФ OPER2 при существительном ВНИМАНИЕ является глагол ПРИВЛЕКАТЬ, в то же время для существительного ИНТЕРЕС значением этой же ЛФ является глагол ПРЕДСТАВЛЯТЬ. Аппарат ЛФ удобно использовать в автоматическом переводе, поскольку ЛФ кодируют универсальные смыслы, по-разному представленные в разных языках и при разных аргументах. Корпус, снабженный ЛФ-разметкой, позволяет исследовать контексты, в которых реализуются ЛФ.

В корпусе представлено 129 различных ЛФ. Полный список ЛФ с пояснениями и примерами можно видеть на сайте НКРЯ. На 2018 г. в корпусе отмечено почти 18 тысяч предложений, содержащих около 25 тысяч ЛФ-сочетаний.

ЛФ-разметка, так же как и морфосинтаксическая и лексическая разметка, производится в автоматическом режиме на построенных и отредактированных синтаксических структурах: вначале ЛФ-анализатор лингвистического процессора ЭТАП-3 на основании сведений, записанных в комбинаторном словаре и в правилах распознавания ЛФ-сочетаний, выделяет ЛФ-сочетания в предложении. Затем полученный результат проверяется и при необходимости исправляется и дополняется лингвистом-аннотатором.

Рассмотрим несколько примеров. Выше в предложении (2) узлы 2 (*данным*) и 1 (*по*) связаны ЛФ-отношением AD2-UN (ЛФ, описывающая предлог, сочетание которого с существительным X описывает свойство или состояние второго

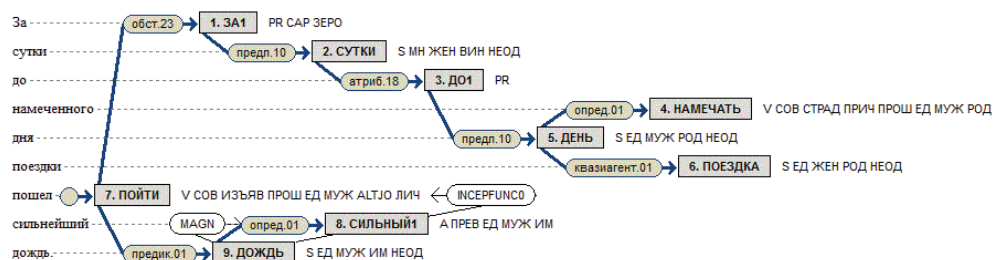


Рис. 4. Синтаксическая структура предложения (4)

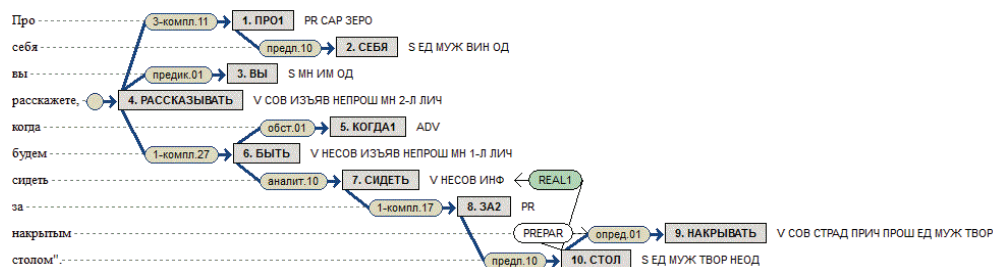


Рис. 5. Синтаксическая структура предложения (5)

участника ситуации X). В предложении (3) узлы 22 (*пароходе*) и 21 (*на*) связаны ЛФ-отношением LOC (ЛФ, описывающая предлог, обозначающий нормальную пространственную или временную локализацию чего-либо по отношению к X-у). В предложениях (4) и (5) ниже имеется по два ЛФ-сочетания с одним аргументом. В предложении (4) *сильный дождь* является сочетанием аргумента ДОЖДЬ со значением ЛФ MAGN ‘высокая степень или интенсивность’, *пошел дождь* является сочетанием аргумента со значением ЛФ INCERFUNC0 ‘начинать иметь место’. В предложении (5) *накрытым столом* представляет собой сочетание аргумента СТОЛ со значением ЛФ PREPAR ‘приготовить X к использованию в соответствии с его назначением’ в форме страдательного причастия, *сидеть за... столом* представляет собой сочетание того же аргумента СТОЛ со значением ЛФ REAL1 ‘использовать X в соответствии с его предназначением’, при этом СИДЕТЬ понимается как значение ЛФ REAL1, а предлог ЗА2 является обязательным контекстом для реализации ЛФ-зависимости. ЛФ-связи, для реализации которых обязателен контекст определенного предлога, отмечаются в структуре записью в овале зеленого цвета, как в предложении (5).

(4) *За сутки до намеченного дня поездки пошел сильнейший дождь.*

(5) *«Про себя вы расскажете, когда будем сидеть за накрытым столом».*

Подробнее о ЛФ-разметке в СинТагРус’е см. в статье [Шеманаева, Фролова 2010].

Поиск по ЛФ доступен онлайн. Однако в настоящее время ЛФ-сочетания в результатах поиска на сайте выделены цветом только в тексте каждого предложения, в структурах же наличие ЛФ-связи и их конкретный вид никак не отражены.



Предложение (4), полученное в результате поиска по запросу ЛФ MAGN, будет иметь в онлайн-варианте вид

(4') За сутки до намеченного дня поездки пошел **сильнейший дождь**.

То же предложение, полученное по запросу ЛФ INCEPFUNC0, будет иметь вид

(4'') За сутки до намеченного дня поездки **пошел** сильнейший **дождь**.

### 3. Новые типы разметки СинТагРус'а

#### 3.1. Анафорическая разметка

Для целей семантического анализа в лингвистический процессор ЭТАП-3 был включен модуль автоматического разрешения анафоры, находящий для каждого местоимения его антецедент — выражение, к которому отсылает это местоимение и которое, как правило, ему предшествует. Например, в предложении

(6) *Директор<sub>1</sub> показался в дверях, натягивая макинтош, его<sub>1</sub> обступили — каждый<sub>2</sub> со своим<sub>2</sub> неотложным делом...*

анафорическое местоимение *он* содержит отсылку к антецеденту *директор* (они кореферентны), а местоимение *своим* — к антецеденту *каждый* (анафора без кореферентности, при которой местоимение отсылает к связанной квантором переменной).

Алгоритм поиска антецедентов представляет собой упорядоченную систему правил, написанных на общем для многоцелевого лингвистического процессора ЭТАП-3 языке FORET. Для него предусмотрено два режима работы: автоматический с возможностью последующего редактирования и ручной. Алгоритм запускается при переводе с русского языка на язык семантических представлений и начинает работать после построения синтаксической структуры парсером ЭТАП-3. Анафорические связи являются ориентированными: они направлены от местоимения к существительному (или, в некоторых случаях, прилагательному), которое является синтаксической вершиной антецедента этого местоимения. Кореферентные связи между неместоименными словами на настоящий момент не устанавливаются.

Технически анафорическая (а также микросинтаксическая) разметка СинТагРус'а (см. ниже раздел 3.2) обеспечивается современной версией программной среды «Редактор структур» (StrEd, см. [Iomdin, Sizov 2009]), разработанной в свое время специально для целей создания и поддержки корпуса.

На рисунке 6 показано, как отражаются анафорические связи в Редакторе структур: в поле COREF под полем с морфосинтаксической разметкой представлены пары словоформ (первая — местоимение, вторая — его антецедент).

В основе наших правил разрешения анафоры лежат принципы, перечисленные в книге [Mitkov 2002], а также наши собственные наблюдения (например, о многих конструкциях с существительными, неспособными быть антецедентами — см. [Иншакова 2016], о некоторых видах нулевых субъектов). Порядок, в котором применяются правила:

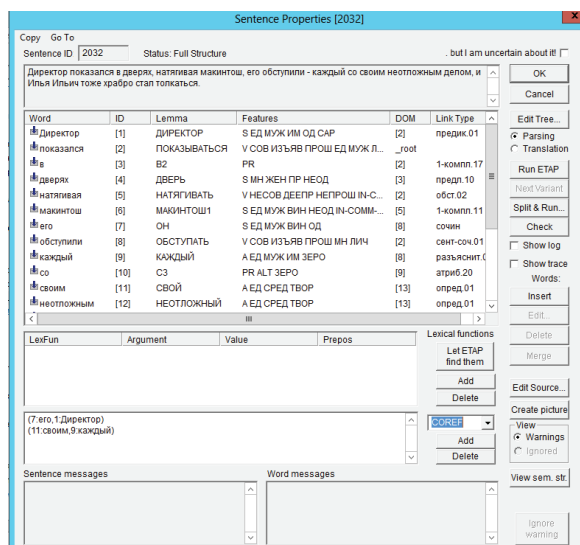


Рис. 6. Представление фразы (6) в Редакторе структур

1. Правило разрешения разных видов нулевой анафоры — нулевых субъектов (PRO) причастных, деепричастных и инфинитивных оборотов, относительных, причинных, временных, условных, уступительных конструкций, конструкций меры и степени (*Акция<sub>i</sub> оказалась такой успешной, что PRO<sub>i</sub> была продлена еще на неделю*), предложений с сентенциальными актантами, а также у событийных существительных в некоторых конструкциях (*Его<sub>i</sub> обвиняли в PRO<sub>i</sub> убийстве*). Это правило проводит «фантомные» синтаксические связи между хо-

зяином нулевого местоимения и его контролером.

2. Правила установления antecedентов возвратных местоимений СЕБЯ и СВОЙ, а также возвратного местоимения друг друга (связь идет от склоняемого элемента ДРУГ).

3. Правила установления antecedентов относительных местоимений КОТОРЫЙ, КТО, ЧТО2 и ЧЕЙ.

4. Правила установления antecedента «местоимения переключения референции» ТОТ1.

5. Правила установления antecedентов местоимений 3-го лица.

В настоящее время поиск antecedента местоимений 3-го лица возможен в пределах трех предложений. Он проводится в три этапа:

а) Поиск и маркировка признаком NON-ANTEC таких существительных, которые вообще не могут быть antecedентами местоимений 3-го лица (сами по себе или в определенных конструкциях), например: *Еще один элемент этой конструкции — рама; признание Мадагаскара независимым государством; в присутствии X-а; сказать в ответ; к сожалению; не в том дело; студия звукозаписи; работать весь день; на 20 лет больше* и мн. др. конструкции (см., например, [Иншакова 2016]).

б) Порождение для каждого местоимения множества гипотетических antecedентов и проведение к ним кореферентных связей. При этом используется морфологическая (согласование), синтаксическая (принципы связывания анафорических местоимений, ограничения на анафору к элементам сочиненной группы), семантическая информация (дескрипторные ограничения на семантику актантов русских лексем в их моделях управления). Данное правило применяется к одному и тому же местоимению до тех пор, пока не будут отобраны все возможные кандидаты на роль его antecedента.



с) Стирание неверных кореферентных связей. За это отвечают три группы правил:

i. правила, использующие семантическую и онтологическую информацию (онтокорреляты русских лексем из онтологии OntoEtap — см. [Богуславский и др. 2012]);

ii. правила, использующие синтаксическую информацию (невозможные синтаксические конфигурации местоимения и его антецедента, конструкции, в которых происходит переключение референции);

iii. правила, использующие дискурсивную информацию (об относительной дискурсивной выделенности разных кандидатов).

Начальная стадия работы над системой разрешения анафоры описана в статье [Иншакова 2016], работа о ее современном состоянии на данный момент не опубликована.

Для оценки качества работы алгоритма разрешения анафоры были созданы и размечены два корпуса в специальном формате Редактора структур .tgt. Первый из них — тренировочный корпус длиной около 3000 предложений, который получен из текстов, входящих в состав СинТагРус'а, в формате .txt. Данные тексты были автоматически проанализированы парсером ЭТАП-3 (без постредактирования), после этого собраны в один tgt-файл и разделены специальными строками — маркерами конца текста, а затем аннотатор вручную разметил корпус эталонными анафорическими связями. Второй — тестовый корпус на основе анафорически размеченного корпуса RuCor, созданного в рамках соревнования систем разрешения анафоры в русских текстах [Toldova et al. 2014; Toldova et al. 2016]<sup>1</sup>. Данный корпус был получен нами путем синтаксического анализа и ручной анафорической разметки текстов в формате .txt, вошедших в RuCor. Причиной того, что мы не использовали синтаксические и анафорические связи, уже имеющиеся в RuCor'е, были разные подходы к разметке у авторов этого корпуса и у нас. Так, в RuCor'е в нумерацию сегментов текста вместе со словоформами включаются и знаки препинания (в ЭТАП-3 — только словоформы), у многих лексем из комбинаторного словаря ЭТАП-3 в разметке имеются номера (ИДТИ1, ИДТИ2 и т.п.), в качестве антецедентов в RuCor'е рассматриваются целые именные группы, тогда как в ЭТАП-3 — словоформы. Корпус на основе RuCor'а был собран по тому же принципу, что и корпус на основе СинТагРус'а; его длина равна 8437 предложениям.

На рисунке 7 представлен фрагмент анафорически размеченного корпуса, где кореферентные связи отображаются в столбце Coref. Координата непосредственно перед словоформой означает ее номер во фразе, а дополнительная координата с минусом присутствует у словоформ, которые относятся к предыдущим фразам. Эта координата содержит также смещение относительно текущей фразы (так, номер -1;10 у словоформы *Василий* означает, что она имеет номер 10 в предшествующем предложении; номер -2;2 у словоформы *дочь* указывает на то, что она идет под номером 2 во втором из предшествующих предложений).

<sup>1</sup> Корпус находится в свободном доступе по адресу [rucoref.maimbava.net](http://rucoref.maimbava.net).

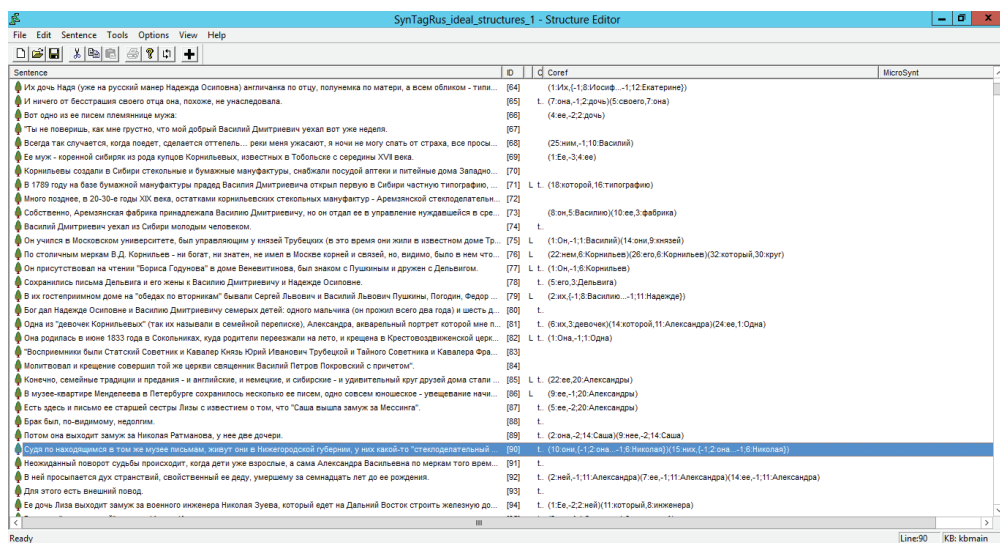


Рис. 7. Фрагмент анафорически аннотированного текста корпуса СинТагРус «А он, мятежный, просит бури...»

Для разметки корпусов был разработан функционал «Рабочее место» (рис. 8). Он представляет собой диалоговое окно, вызываемое из главного меню Редактора структур. Функционал позволяет проводить кореферентные связи в пределах одного или нескольких предложений как вручную, так и автоматически.

Здесь в окне “Sent. number” задается диапазон предложений, в которых могут строиться кореферентные связи. Если диапазон включает предыдущие предложения, они выводятся в окне “Previous sentences”.

В окне, находящемся непосредственно под окном “Previous sentences”, выводится текущее предложение, со словами которого пользователь будет работать. Ниже окна с этим предложением выводится список кореферентных связей, построенных для текущего предложения. В нем указываются координаты в тексте и словоформы для слова-хозяина и слова-слуги.

Блок “Manual addition” позволяет добавлять разметку вручную. Слово, из которого выходит кореферентная связь, выбирается в выпадающем списке Pronoun. Слово-антецедент, в которое связь входит, выбирается в выпадающем списке Begin. После указания слов, входящих

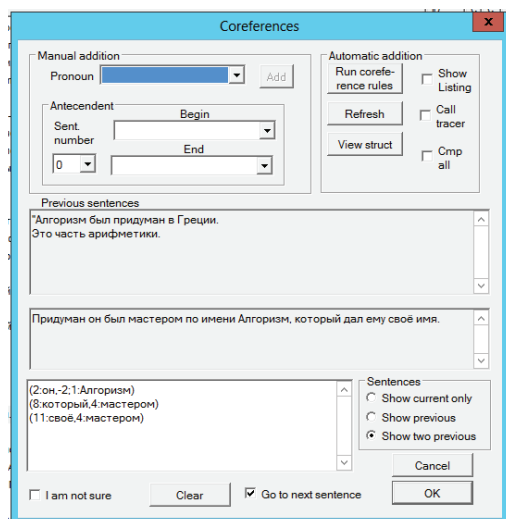


Рис. 8. Рабочее место разметчика

в добавляемую кореферентную связь, данная связь добавляется в список связей и запоминается в разметке нажатием кнопки Add. Связи, проведенные вручную, в дальнейшем используются в качестве эталона.

Непосредственная оценка качества работы алгоритма осуществляется с помощью созданной в лаборатории программы, которая автоматически строит кореферентные связи для фраз из корпуса. Построенные связи сравниваются с эталонными, и при обнаружении расхождений выводится сообщение об этом. Также в ходе работы программа запоминает количество правильных (совпадающих с эталоном) связей и общее количество построенных и эталонных связей. На основании этой информации в конце работы вычисляется точность (отношение числа построенных правильных связей к числу всех построенных связей) и полнота (отношение числа построенных правильных связей к числу эталонных связей), используемые для оценки качества системы.

### ***3.2. Микросинтаксическая разметка***

Разметка СинТагРус'а микросинтаксическими единицами (то есть, по существу, идентификация таких единиц) представляет собой самое последнее принципиальное обновление структуры корпуса. Она осуществляется недавно, с 2016 г. Этот тип разметки прямо связан с развитием работы по микросинтаксису (из последних публикаций упомянем [Иомдин 2013(а,б), 2014, 2017(а,б), 2018; Iomdin 2015]), в ходе которой стало очевидно, что корпусные данные, включающие информацию о наличии в тексте синтаксически чувствительных фразеологических единиц и их встроенности в структуру предложения, исключительно полезны как для теоретической лингвистики и практической одноязычной и двуязычной лексикографии, так и для решения компьютерно-лингвистических задач, в частности глубокого автоматического семантического анализа. Добавим, что микросинтаксическая разметка корпуса ведется параллельно с созданием Микросинтаксического словаря русского языка, в котором микросинтаксические единицы получают полноценное лексикографическое описание (см. [Iomdin 2016; Иомдин 2017]). Характеристику этого словаря мы в данной статье оставляем в стороне.

Насколько известно авторам, в настоящее время СинТагРус — единственный корпус текстов, в котором эксплицитно размечаются синтаксически чувствительные фразеологические единицы. Разумеется, существуют и активно развиваются одноязычные и параллельные корпуса текстов разных языков, в которых размечаются различного рода коллокации, или *multiword expressions* (см., например, [Bejček, Straňák 2010; Hnátková et al. 2017; Osenova, Simov 2018], а также обзор релевантных подходов в [Rosén et al. 2016] и [Savary et al. 2017]). Однако в подобной разметке каждая такая единица фиксируется, как правило, без учета ее фразеологичности, почти как единое целое, без представления о том, как именно она интегрируется в синтаксическую структуру предложения, какие в точности индивидуальное значения отдельных слов в нее входят и какие синтаксические связи

постулируются между составными элементами единицы. Все это присутствует в микросинтаксической разметке СинТагРус'а.

Микросинтаксические единицы бывают двух типов: нестандартные синтаксические конструкции и синтаксические фраземы.

Нестандартных синтаксических конструкций в русском языке немного: их, по нашим оценкам, несколько десятков. Примерами таких конструкций являются конструкции так называемого малого синтаксиса, такие как инфинитивно-модальные конструкции типа

1) *Y-у X-овать* ('Y должен будет X-овать', ср. *Мне сегодня всю ночь работать*);

2) *Y-у не X-овать* ('Отсутствует перспектива, что Y будет X-овать', ср. *Нашей сборной никогда не пробиться в финал*);

конструкции типа

3) *Y-у не до X-а* ('Y занят более важными делами, чем X, и заявляет, что не будет делать X, считая, что X-м можно пренебречь'; ср. *Коле сейчас не до развлечений*).

Значительное место среди нестандартных синтаксических конструкций занимают разнообразные выражения с обязательным повторением лексических единиц: у каждого такого типа выражений всякий раз своеобразная семантика. Надо сказать, что таких конструкций в русском языке достаточно много; вероятно, их существенно больше, чем, например, в английском, хотя и там их немало<sup>2</sup>. Приведем несколько примеров этих выражений:

4)  $X_{инф}$  не  $X_{лич}$ : *Летать не летал, но два раза сидел в салоне Конкорда* [интернет-форум]: (≈ 'Не летал, но имел место факт «сидел в салоне», что является чем-то более слабым, чем факт «летал»');

5)  $X_{пов}$  не  $X_{нов}$ : *Теперь кричи не кричи, зови не зови — никто не услышит* [В. Солоухин] (≈ 'Независимо от того, будешь ли ты кричать или нет, будешь ли ты звать или нет, это не приведет к желательному результату...');

6)  $X_{им}$  есть  $X_{им}$ : *Мальчишки есть мальчишки* (≈ 'Мальчишки ведут себя так, как следует ожидать от мальчишек'); *Жена есть жена* [А.П. Чехов. Три сестры]; *закон есть закон* (≈ 'закон функционирует так, как следует ожидать от закона') и т. д.;

7)  $X_{им}$  как  $X_{им}$ : *Машина как машина, ничего особенного* (≈ 'Ничем не примечательная машина'); *Нижний Прохора не поразил — город как город* [В.Я. Шишков];

8) *X так X* (≈ 'говорящий согласен с предложением X слушающего и готов его выполнить, поскольку говорящему выбор между X и чем-то другим безразличен'): ср. *Хотите еще кофе? — Лучшие чаю. — Чай так чай* [Л. Юзефович]; *Мне вообще казалось неудобным быть выше мамы, и я ей подчеркнуто во всем подчинялся: в лифт — так в лифт, не заходить в квартиру — так не заходить* [А. Алексин].

Подавляющая часть микросинтаксических элементов принадлежит к классу синтаксических фразем. Мы понимаем под синтаксическими фраземами такие фразеологические единицы, которые, помимо лексической избирательности

<sup>2</sup> Представительный список английских тавтологических конструкций приводится, в частности, в работе [Rhodes 2009].

и семантической некомпозициональности, обладают заметной синтаксической спецификой (например, управляющими свойствами, отсутствующими у элементов таких единиц: так, единица *руки чешутся* управляет инфинитивом и предложением на (ср. *Чешутся руки написать стихи о состязании двух людей в благородстве* [Ф. Искандер]; *Уж очень у меня на этого Попугайчикова руки чешутся* [А. В. Сухово-Кобылин]), а единица *не по себе* управляет союзом *что* (ср. *И было как-то не по себе, что едешь не в ту сторону, соблазнившись этим неожиданным отпуском от войны* [К. Симонов]).

Некоторые такие единицы тяготеют к цельным словам; ср. единицы типа *всё равно* (по крайней мере в двух из значений — ‘в любом случае’, как в *Скажите, что я всё равно приеду* [Д. Гранин], и ‘равносильно’, как в афоризме М. Л. Гаспарова «*Цзи Юнь говорил: поститься — это все равно что не брать взятки по вторникам и четвергам*»), *всё же, тот же, между тем, между прочим, тем не менее, тем более, только что, пока что, разве что, то и дело, что ли, мало ли, во что бы то ни стало, в кои веки*, союз и частица *как бы* и десятки других. Многие из этих единиц в русском словаре лингвистического процессора ЭТАП-3 представлены как безусловные обороты (см. ниже), их насчитывается порядка двухсот.

В процессе языковых изменений некоторые такие единицы движутся по направлению к единому слову, подобно тому, как это произошло с союзом/наречием *якобы* (которое, согласно данным НКРЯ, могло писаться раздельно вплоть до начала XX в.), или подобно партикулам Т. М. Николаевой [1985, 2008], таким как *ибо, или, либо* и *неужели*, исторически образовавшимся из нескольких отдельных формантов.

Другие синтаксические фраземы свести к цельным словам нельзя. Примером тут является единица *всё равно* в значении ‘безразлично’: в одних случаях форманты *всё* и *равно* стоят рядом; ср. *Мне всё равно, куда ехать отдыхать*, в других они разделены; ср. *Зависеть от царя, зависеть от народа — Не всё ли нам равно?* [А. С. Пушкин]. К этому же подклассу относятся такие разнообразные единицы, как *как быть* (≈ ‘как следует поступать’) в примерах типа *как врачу быть с пациентами, которые отказываются принимать лекарства*, дискурсивное выражение *то ли дело* (ср. *Не понравилась мне Москва: лохматая, кривобокая, — как старушки, горбатые домишки. То ли дело Петербург-цеголь* [А. С. Серафимович]), конструкции типа *черт <дьявол, бес, бог...> знает что <кто, где, когда, зачем...>* и многие другие.

В СинТагРус’е микросинтаксическая разметка распространяется на оба перечисленных вида единиц. Следует отметить, что внесение в корпус такой разметки — весьма нетривиальная задача.

С одной стороны, готовых списков микросинтаксических единиц, которые можно было бы хотя бы с натяжкой считать полными, не существует. Даже традиционные фразеологические словари здесь мало помогают: большинство присутствующих в них идиом не имеют какой-либо синтаксической специфики, а многие микросинтаксические единицы, в том числе близкие к служебным лексическим

единицам (составные предлоги, союзы, частицы), во фразеологические словари никогда не попадают.

С другой стороны, бывает далеко не просто отличить микросинтаксическую единицу, которая даже не обязательно принадлежит одной синтаксической группе, от произвольной последовательности слов: ср., например, микросинтаксическую союзную единицу *как бы* в предложении *Боюсь, как бы он не заболел* и случайную цепочку, состоящую из наречия *как* и частицы *бы* в предложении *Как бы ты поступил на моем месте?*

Учитывая недостаток исходных данных, авторы СинТагРус'а используют при его микросинтаксической разметке две дополняющие друг друга стратегии: (1) полный просмотр текста, осуществляемый с целью зафиксировать все возможные микросинтаксические единицы, присутствующие в тексте, и (2) направленный поиск вхождений микросинтаксических единиц, уже известных разметчику. Поисковые запросы при второй стратегии формулируются в терминах линейной последовательности элементов, характерной для таких единиц, и в терминах фрагментов синтаксических поддеревьев. Эта формулировка осуществляется с помощью специальных правил, написанных на языке FORET. В последнее время, когда микросинтаксическая разметка вводится в новые тексты одновременно с их морфосинтаксической разметкой, разумеется, используется первая стратегия (она применяется аннотатором к каждому предложению).

Идентификация микросинтаксических единиц в корпусе заключается в фиксации имен этих единиц (в случае синтаксических фразем это обычно цепочка слов, возможно, с цифрой, указывающей на конкретное значение фраземы, если она многозначна), а также указании линейных границ, в пределах которых оказываются элементы фиксируемой микросинтаксической единицы. В случае, если та или иная единица не представляет собой непрерывной последовательности идущих друг за другом слов, а располагается менее плотно, информация о лексическом составе единицы может оказаться недостаточно определенной (ср., например, представление синтаксической фраземы *как быть* в конструкциях типа *Не знаю уж, как и быть мне с тобой...* [М.Е. Салтыков-Щедрин] и даже, казалось бы, совсем уж неделимой единицы *несмотря на* в контекстах типа *В крещенский праздник, несмотря ни на какие морозы, купаются сотни горожан* [О. Белякова]: здесь линейные отрезки содержат посторонние микросинтаксическим единицам элементы и и ни).

В настоящее время микросинтаксическая разметка представляет собой особое поле в XML-файле, соответствующем размечаемому тексту. В этом поле фиксируется имя микросинтаксической единицы и линейный отрезок, в пределах которого она располагается. Графическое представление этого поля, порожаемое Редактором структур, дается на рисунке 9.

Микросинтаксическая разметка корпуса производится в основном вручную. Исключение составляют ситуации, когда в словаре системы используются в качестве лексических единиц неоднословные токены (так называемые «безусловные обороты»), которые в подавляющем большинстве случаев представляют собой



микросинтаксические единицы (конкретнее, синтаксические фраземы). Примерами таких единиц могут служить обороты *а именно, а также, а то и, без устали, бок о бок, в виду* (для представления выражений *иметь в виду и иметься в виду*), *в качестве, в конечном счете, в конце концов, в настоящее время, в противном случае, в связи с, в силу, в складчину, в соответствии с, в состоянии, в течение, в то время как, в частности, во время, во избежание, во что бы то ни стало, во языцех, вплоть до, вряд ли и др.* Их можно автоматически квалифицировать как таковые, конечно,

с последующей экспертной редакцией. Скажем, единица *как только* практически всегда представляет собой единый союз, однако изредка встречаются и ситуации, когда эта цепочка слов должна трактоваться по отдельности; ср. *Как только тебе не стыдно?* Точно так же последовательность *всё же* в подавляющем числе случаев выступает как единая микросинтаксическая единица — частица; ср. *И всё же мы часто ссорились* [С. Довлатов], но изредка представляет два разных слова; ср. *Не всё же разглагольствовать о том, каким должен быть хороший человек, пора и стать им* [С. Смирнов]. В подобных ситуациях микросинтаксическая интерпретация таких последовательностей оказывается ошибочной и удаляется из разметки.

Иногда работа с такими единицами при разметке корпуса приводит к неожиданным и даже курьезным результатам. Так, у той же синтаксической фраземы *всё же* имеется однословный, почти полностью синонимичный аналог (почти полный синоним) *всё-таки*, а у этого последнего имеется трехсловный вариант *всё ж таки* (который является вполне частотным: в основном корпусе НКРЯ имеется 670 его вхождений), и его разумно признать безусловным оборотом, в отличие от спорной, как мы только что видели, единицы *всё же*.

Легко убедиться, что микросинтаксически размеченный корпус текстов представляет собой ценнейший лингвистический материал для исследования русской, а возможно, и сопоставительной фразеологии.

В частности, по соотношению количества истинных и ложно-положительных вхождений той или иной нестандартной синтаксической конструкции или

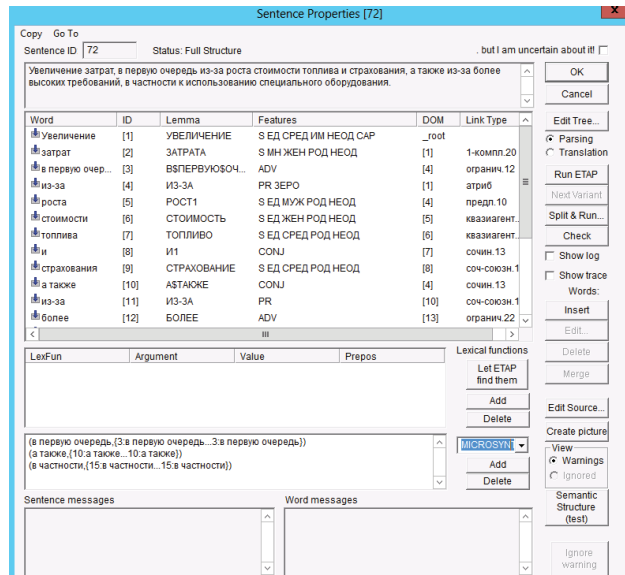


Рис. 9. Представление фразы, содержащей три микросинтаксические единицы — *в первую очередь, а также и в частности*

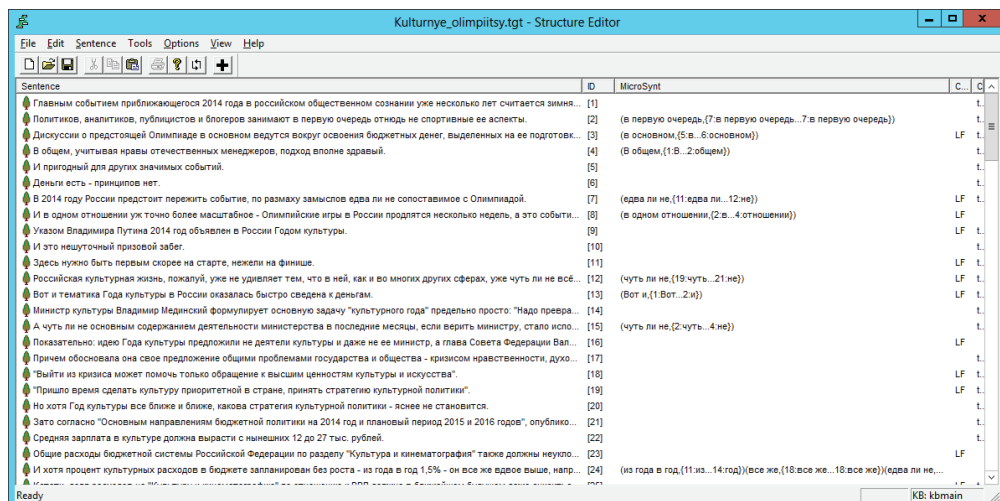


Рис. 10. Фрагмент микросинтаксически аннотированного текста корпуса СинТагРус «Культурные олимпиады». Этот текст содержит 133 предложения, из которых в 34 предложениях (25%) присутствует хотя бы один микросинтаксический элемент

синтаксической фраземы можно судить о ее фразеологической силе: если в подкорпусе предложений, содержащих такие вхождения, их статус достаточно однороден и истинные вхождения микросинтаксической единицы существенно преобладают над ложными, можно считать, что данная единица обладает высокой степенью фразеологичности. Практическим следствием такой оценки может быть утверждение данной единицы безусловным оборотом, даже если теоретически можно допустить, что в реальном тексте могут присутствовать последовательности элементов некоторой единицы, которые ее, однако, не формируют.

Показательным примером является синтаксическая фразема *в первую очередь*. Нетрудно заметить, что в русском тексте последовательность этих словоформ может фигурировать как свободное сочетание (как во фразе *Не становись в первую очередь, она очень медленно движется*) или даже как случайное соположение слов, не образующих смыслового единства (ср. *Работали две кассы, в первую очередь стояла огромная, а во вторую почему-то небольшая*). Тем не менее анализ микросинтаксической разметки СинТагРус'а показывает, что все 125 вхождений последовательности словоформ *в, первую и очередь* представляют собой именно синтаксическую фразему.

Как показывают предварительные статистические оценки, в среднем микросинтаксические единицы присутствуют в четверти предложений типичного текста СинТагРус'а. На рисунке 10 приводится фрагмент одного такого текста.

В сегодняшней версии СинТагРус'а подкорпус, содержащий микросинтаксическую разметку, составляет более 2500 предложений. Общее число встречающихся в нем разных микросинтаксических единиц — около 300.

Более детальное описание микросинтаксической разметки СинТагРус'а можно найти в работах [Маракасова, Иомдин 2016] и [Iomdin 2017].



#### 4. Использование корпуса СинТагРус

Корпус СинТагРус, как и другие глубоко аннотированные корпуса текстов, активно используется для целей теоретико-лингвистических исследований (в первую очередь, разумеется, в области синтаксиса) и для практической лексикографии.

Кроме того, корпус применяется в компьютерно-лингвистических задачах в качестве источника лингвистических данных. Конкретные задачи могут быть самыми разными — от создания автоматических парсеров на основе машинного обучения (см., например, опыт создания Maltparser'a для русского языка в [Nivre et al. 2008]) до использования корпусной статистики при оптимизации правилых парсеров и для регрессионного тестирования синтаксической модели, лежащего в основе СинТагРус'a.

Добавим, что как среди компьютерных лингвистов, так и среди лингвистов-теоретиков большим спросом пользуется оффлайновая версия СинТагРус'a: к настоящему времени образовательным и научным учреждениям и отдельным исследователям выдано свыше 100 лицензий на некоммерческое использование этого ресурса.

Особого упоминания заслуживает сравнительно новая практика использования синтаксически размеченных корпусов текстов для создания унифицированных производных корпусов текстов.

Так, в рамках многоязычного некоммерческого проекта Universal Dependencies (<http://universaldependencies.org/>) СинТагРус был преобразован в формат универсальных зависимостей — гармонизированную систему представления синтаксической структуры в виде деревьев зависимостей, которая в достаточной степени подходит для описания разноструктурных языков. На основании этого формата был создан единый стандарт, с помощью которого аннотируются корпуса текстов на разных языках, если они предусматривают синтаксическую разметку (подробнее о нем см. [Nivre 2015] и [Lyashevskaya et al. 2016]). До настоящего времени не разрабатывались новые корпуса, которые с самого начала размечались бы по данному стандарту, но на сегодняшний момент (релиз от 1 июля 2018 г.) в этот стандарт преобразовано 122 корпуса на 71 языке [Nivre et al. 2018]. В число этих корпусов входит четыре корпуса русского языка — три небольших экспериментальных корпуса, несопоставимых с СинТагРус'ом по объему, и корпус, полученный из СинТагРус'a (без непосредственного участия его разработчиков). Этот корпус фигурирует на портале Universal Dependencies под именем UD\_Russian-SynTagRus и соответствует состоянию СинТагРус'a на середину 2016 г.

#### Литература

Апресян Ю. Д., Богуславский И. М., Иомдин Б. Л., Иомдин Л. Л., Санников А. В., Санников В. З., Сизов В. Г., Цинман Л. Л. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003–2005. М. : Индрик, 2005. С. 193–214.

Богуславский И. М., Диконов В. Г., Тимошенко С. П. Онтология для поддержки задач извлечения смысла из текста на естественном языке // **Информационные технологии и системы 2012 (ИТиС'2012). Труды 35-й междисциплинарной школы-конференции ИППИ РАН. Петрозаводск, 2012. С. 152–161.**

Богуславский И. М., Иомдин Л. Л., Валеев Д. Р., Сизов В. Г. Синтаксический анализатор системы ЭТАП и его оценка с помощью глубоко размеченного корпуса русских текстов // Международная конференция «Корпусная лингвистика — 2008». СПб., 2008а. С. 56–74.

Богуславский И. М., Иомдин Л. Л., Митюшин Л. Г., Сизов В. Г. Длина синтаксических связей в русском аннотированном корпусе // Международная конференция «Корпусная лингвистика — 2008». СПб., 2008б. С. 75–82.

Дяченко П. В., Иомдин Л. Л., Лазурский А. В., Митюшин Л. Г., Подлесская О. Ю., Сизов В. Г., Фролова Т. И., Цинман Л. Л. Современное состояние глубоко аннотированного корпуса текстов русского языка (СинТагРус) // Национальный корпус русского языка: 10 лет проекту. Труды Института русского языка им. В. В. Виноградова. Вып. 6. М., 2015. С. 272–299.

Жолковский А. К., Мельчук И. А. О семантическом синтезе // Проблемы кибернетики. Вып. 19. М.: Наука, 1967. С. 177–238.

Иниакова Е. С. Разрешение синтаксической местоименной анафоры в системе ЭТАП-3 // **Информационные технологии и системы 2016 (ИТиС'2016). Труды 40-й междисциплинарной школы-конференции ИППИ РАН. СПб., 2016. С. 420–429.**

Иомдин Л. Л. Читать не читал, но...: об одной русской конструкции с повторяющимися словесными элементами // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог». Вып. 12 (19). Т. 1. М.: Изд-во РГГУ, 2013а. С. 297–310.

Иомдин Л. Л. Некоторые микросинтаксические конструкции в русском языке с участием слова *что* в качестве составного элемента // Южнославянский филолог. Т. LXIX. Белград: Институт сербского языка Сербской академии наук и искусств, 2013б. С. 137–147.

Иомдин Л. Л. Хорошо меня там не было: синтаксис и семантика одного класса русских разговорных конструкций // Grammaticalization and Lexicalization in the Slavic Languages. По материалам Международного симпозиума «Грамматикализация и лексикализация в славянских языках», 11–14 ноября 2011 г. München ; Berlin ; Washington D. C. : Verlag Otto Sagner, 2014. Bd. 55. S. 423–436.

Иомдин Л. Л. Как нам быть с конструкциями типа *как быть?* // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог». Вып. 16 (23). Т. 1. М.: Изд-во РГГУ, 2017а. С. 161–176.

Иомдин Л. Л. Между синтаксической фраземой и синтаксической конструкцией. Нетривиальные случаи микросинтаксической неоднозначности // SLAVIA, časopis pro slovanskou filologii, ročník 86, sešit 2–3. 2017б. S. 230–243.

Иомдин Л. Л. Еще раз о микроконструкциях, сформированных служебными словами: *то и дело* // Компьютерная лингвистика и интеллектуальные технологии.

По материалам ежегодной Международной конференции «Диалог». Вып. 17 (24). М. : Изд-во РГГУ, 2018. С. 267–283.

**Маракасова А. А., Иомдин Л. Л.** Микросинтаксическая разметка в корпусе русских текстов СинТагРус // Информационные технологии и системы 2016 (ИТиС'2016). Труды 40-й междисциплинарной школы-конференции ИППИ РАН. СПб., 2016. С. 445–449.

Мельчук И. А. Опыт теории лингвистических моделей «Смысл ↔ Текст». М. : Наука, 1974. 314 с. (2-е изд.: 1999. 346 с.)

**Николаева Т. М.** Функции частиц в высказывании (на материале славянских языков). М. : Наука, 1985. 170 с.

**Николаева Т. М.** Непарадигматическая лингвистика (История «блуждающих частиц»). М. : Языки славянской культуры, 2008. 689 с.

Шеманаева О. Ю., Фролова Т. И. Лексико-функциональная разметка текстов в СинТагРус // Информационные технологии и системы 2010 (ИТиС'10). Труды 33-й конференции молодых ученых и специалистов ИППИ РАН. М. : ИППИ, 2010. С. 320–324.

Apresjan Ju., Boguslavsky I., Iomdin L., Lazursky A., Sannikov V., Sizov V., Tsinman L. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning-Text Theory. P., 2003. P. 279–288.

Bejček E., Straňák P. Annotation of Multiword Expressions in the Prague Dependency Treebank // Language Resources and Evaluation. 2010. Vol. 44. No. 1–2. P. 7–21.

Boguslavsky I. SynTagRus — a Deeply Annotated Corpus of Russian // Les émotions dans le discours. Emotions in Discourse / eds. Peter Blumenthal, Iva Novakova, Dirk Siepmann. Peter Lang Edition, 2014. P. 367–381.

Boguslavsky I., Grigorieva S., Grigoriev N., Kreidlin L., Frid N. Dependency Treebank for Russian: Concept, Tools, Types of Information // Proc. of the 18th International Conference on Computational Linguistics (COLING 2000). San Francisco : Kaufmann, 2000. P. 987–991.

**Hnátková M., Petkevič V., Skoumalová H.** Multiword Expressions in Czech: Between Lexicon and Grammar // *Corpus Linguistics* — 2017. St. Petersburg, St. Petersburg State University: 2017. P. 36–42.

Iomdin L. The Challenge of Treating Collocations // International Journal of Lexicography. 2015. Vol. 28. No. 3. P. 376–384.

Iomdin L. Microsyntactic Phenomena as a Computational Linguistics Issue // Grammar and Lexicon: Interactions and Interfaces. Proc. of the Workshop. Osaka, 2016. P. 8–18. Available at: <http://aclweb.org/anthology/W/W16/W16-38.pdf>.

**Iomdin L.** Microsyntactic Annotation of Corpora and its Use in Computational Linguistics Tasks // *Jazykovedný časopis, ročník 86, 2017, číslo 2*. S. 169–178.

**Iomdin L., Sizov V.** Structure Editor: a Powerful Environment for Tagged Corpora // MONDILEX Fifth Open Workshop. Ljubljana, 2009. P. 1–12.

Lyashevskaya O., Droganova K., Zeman D., Alexeeva M., Gavrilova T., Mustafina N., Shakurova E. Universal Dependencies for Russian: A New Syntactic Dependencies Tag-set // NRU HSE. Series WP BRP «Linguistics». No. 44. 2016.

*Mitkov R.* Anaphora resolution. Longman, 2002. 220 p.

*Nivre J.* **Towards a Universal Grammar for Natural Language Processing // Computational Linguistics and Intelligent Text Processing (CICLing 2015). Part 1.** Springer, 2015. P. 3–16.

*Nivre J., Abrams M., Agić Ž. et al.* **Universal Dependencies 2.2 // LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL). Faculty of Mathematics and Physics, Charles University, 2018.** Available at: <http://hdl.handle.net/11234/1-2837>.

*Nivre J., Boguslavsky I., Iomdin L.* **Parsing the SynTagRus Treebank of Russian // Proc. of the 22nd International Conference on Computational Linguistics (COLING'08). Manchester, August 18—22 2008 / eds. D. Scott, H. Uszkoreit. Vol. 1. 2008. P. 641—648.**

*Osenova P., Simov K.* **Modelling multiword expressions in a parallel Bulgarian-English newsmedia corpus // Multiword expressions: Insights from a multi-lingual perspective. Phraseology and Multiword Expressions / eds. M. Sailer, S. Markantonatou. Berlin : Language Science Press, 2018. P. 247—270.**

*Rhodes R.* **Tautological constructions in English ... and beyond // Presented to the Syntax and Semantics Circle, UCB, 2009.**

Available at: [http://linguistics.berkeley.edu/~russellrhodes/pdfs/syntax\\_circle\\_taut\\_qp.pdf](http://linguistics.berkeley.edu/~russellrhodes/pdfs/syntax_circle_taut_qp.pdf).

*Rosén V., De Smedt K., Losnegaard G. S., Bejček E., Savary A., Osenova P.* **MWEs in Treebanks: From Survey to Guidelines // Proc. of the 10th International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia. 2016. P. 2323—2330.**

*Savary A., Sangati F., Candito M. et al.* **The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions // Proc. of the 13th Workshop on Multiword Expressions (MWE 2017), Valencia, Spain, 4 April 2017. P. 31—47.**

*Toldova S., Roytberg A., Ladygina A., Azerkovich I., Vasilyeva M. D.* **Error analysis for anaphora resolution in Russian: new challenging issues for anaphora resolution task in a morphologically rich language // Proc. of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016), co-located with NAACL 2016. San Diego, California, June 16, 2016. Stroudsburg, PA : Association for Computational Linguistics, 2016. P. 74—83.**

*Toldova S., Roytberg A., Nedoluzhko A., Kurzukov M., Ladygina A., Vasilyeva M., Azerkovich I., Grishina Y., Sim G., Ivanova A., Gorshkov D.* **Evaluating Anaphora and Coreference Resolution for Russian // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог». Вып. 13 (20). М. : Изд-во РГГУ, 2014. С. 681—695.**

<sup>1</sup>Ye.S. Inshakova, <sup>2</sup>L.L. Iomdin, <sup>3</sup>L.G. Mityushin,

<sup>4</sup>V.G. Sizov, <sup>5</sup>T.I. Frolova, <sup>6</sup>L.L. Tsinman

<sup>123456</sup>A.A. Kharkevich Institute for Information Transmission Problems,  
Russian Academy of Sciences,

<sup>2</sup>Russian State University for the Humanities  
(Russia, Moscow)

<sup>1</sup>yesi@mail.ru, <sup>2</sup>iomdin@gmail.com, <sup>3</sup>mit@iitp.ru,

<sup>4</sup>sizov@iitp.ru, <sup>5</sup>tfrolova@gmail.com, <sup>6</sup>llcinman@gmail.com

## THE SYNTAGRUS TODAY

The paper describes the current state of the SynTagRus corpus composed of Russian texts tagged with morphosyntactic structures. At certain points of the corpus development, additional kinds of tagging were introduced, namely lexical-semantic, lexical-functional, anaphoric and microsyntactic tagging.

The morphosyntactic tagging of a sentence includes morphological structures of all the words and the syntactic structure of the sentence in the form of a dependency tree, in accordance with I. Mel'čuk and A. Zholkovsky's "Meaning — Text" model. Lexical-semantic tagging implies that each word is assigned a corresponding entry of the Russian combinatorial dictionary. Lexical-functional tagging means finding the phrases in the text which can be interpreted in terms of lexical functions. Anaphoric tagging results in marking the antecedents of pronouns. Microsyntactic tagging identifies syntactic phraseemes and certain nonstandard syntactic constructions occurring in the text.

Tagging of a new text is performed in several stages. First, the text is processed by the multifunctional linguistic processor ETAP-3 which makes morphosyntactic and lexical-semantic tagging in automatic mode. Then the output of the processor is checked and possibly corrected by specially trained annotators. After that, ETAP-3 uses the morphosyntactic structures of the text to make lexical-functional and anaphoric tagging. Finally, the annotators check these types of tagging and manually perform microsyntactic tagging.

The SynTagRus corpus may be used in theoretical linguistic research as well as practical lexicography. The corpus statistics may help optimize decision making in various automatic text processing procedures. Another very promising possibility is to use the SynTagRus data as input for the modern machine learning systems.

*Key words:* SynTagRus, syntactically tagged corpus, corpus of Russian texts, dependency grammar, lexical functions, pronoun antecedents, microsyntax, ellipsis

## References

Apresjan Ju., Boguslavsky I., Iomdin L., Lazursky A., Sannikov V., Sizov V., Tsinman L. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT. *MTT 2003, First International Conference on Meaning-Text Theory*. Paris, 2003, pp. 279—288.

Apresjan Ju.D., Boguslavsky I. M., Iomdin B. L., Iomdin L. L., Sannikov A. V., Sannikov V. Z., Sizov V. G., Cinman L. L. [Syntactically and semantically tagged corpus of Russian: state of the art and prospects]. *Natsional'nyi korpus russkogo yazyka: 2003—2005* [The Russian National Corpus: 2003—005. Results and Prospects]. Moscow, Indrik Publ., 2005, pp. 193—214. (In Russ.)

Bejček E., Straňák P. Annotation of Multiword Expressions in the Prague Dependency Treebank. *Language Resources and Evaluation*, 2010, vol. 44, no. 1—2, pp. 7—21.

Boguslavsky I. SynTagRus — a Deeply Annotated Corpus of Russian. *Les émotions dans le discours. Emotions in Discourse*. Eds. Peter Blumenthal, Iva Novakova, Dirk Siepmann. Peter Lang Edition, 2014, pp. 367—381.

Boguslavsky I. M., Dikonov V. G., Timoshenko S. P. [Ontology for supporting the tasks of meaning extraction from texts in natural languages]. *Informatsionnye tekhnologii i sistemy 2012 (ITiS'2012). Trudy 35-i mezhdistsiplinarnoi shkoly-konferentsii IPPI RAN* [Information Technologies and Systems 2012 (ITiS'2012). Proc. of the 35th Interdisciplinary School-Conference of IITP RAS]. Petrozavodsk, 2012, pp. 152—161. (In Russ.)

Boguslavsky I., Grigorieva S., Grigoriev N., Kreidlin L., Frid N. Dependency Treebank for Russian: Concept, Tools, Types of Information. *Proc. of the 18th International Conference on Computational Linguistics (COLING 2000)*. San Francisco, Kaufmann, 2000, pp. 987—991.

Boguslavsky I. M., Iomdin L. L., Mitjushin L. G., Sizov V. G. [The length of syntactic links in the Russian tagged corpus]. *Mezhdunarodnaya konferentsiya "Korpusnaya lingvistika — 2008"* [Proc. of the International Conference "Corpus Linguistics — 2008"]. St. Petersburg, 2008b, pp. 75—82. (In Russ.)

Boguslavsky I. M., Iomdin L. L., Valeev D. R., Sizov V. G. [A syntactic analyzer of the ETAP system and its evaluation with the help of a deeply annotated corpus of Russian texts]. *Mezhdunarodnaya konferentsiya "Korpusnaya lingvistika — 2008"* [Proc. of the International Conference "Corpus Linguistics — 2008"]. St. Petersburg, 2008a, pp. 56—74. (In Russ.)

Dyachenko P. V., Iomdin L. L., Lazursky A. V., Mityushin L. G., Podlesskaya O. Yu., Sizov V. G., Frolova T. I., Tsinman L. L. [A deeply annotated corpus of Russian texts (SynTagRus): contemporary state of affairs]. *Natsional'nyi korpus russkogo yazyka: 10 let proektu. Trudy Instituta russkogo yazyka im. V. V. Vinogradova. Vyp. 6* [The Russian National Corpus: 10 Years of the Project. Proc. of the V. V. Vinogradov Russian Language Institute. Iss. 6]. Moscow, 2015, pp. 272—299. (In Russ.)

**Hnátková M.**, Petkevič V., Skoumalová H. Multiword Expressions in Czech: Between Lexicon and Grammar. *Proc. of the Conference "Corpus Linguistics — 2017"*. St. Petersburg, St. Petersburg State University, 2017, pp. 36—42.

Inshakova E. S. [Resolution of syntactic pronominal anaphora in the ETAP-3 system]. *Informatsionnye tekhnologii i sistemy 2016 (ITiS'2016). Trudy 40-i mezhdistsiplinarnoi shkoly-konferentsii IPPI RAN* [Information Technologies and Systems 2016 (ITiS'2016). Proc. of the 40th Interdisciplinary School-Conference of IITP RAS]. St. Petersburg, 2016, pp. 420—429. (In Russ.)



Iomdin L. L. [“Chitat’ ne chital, no...”: on one Russian construction with repeating items]. *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”]. Iss. 12 (19), vol. 1. Moscow, RSUH Publ., 2013a, pp. 297—310. (In Russ.)

Iomdin L. L. [Certain microsyntactic constructions in Russian which contain the word “chto” as a constituent element]. *Južnoslovenski filolog*, vol. LXIX. Beograd, 2013b, pp. 137—147. (In Russ.)

Iomdin L. L. [Good thing I wasn’t there: syntax and semantics of a class of Russian colloquial constructions]. *Grammaticalization and lexicalization in the Slavic languages. Papers from the 36th meeting of the commission on the grammatical structure of the Slavic languages of the International committee of Slavists*. München/Berlin/Washington D. C., Verlag Otto Sagner, 2014, band 55, pp. 423—436. (In Russ.)

Iomdin L. The Challenge of Treating Collocations. *International Journal of Lexicography*, 2015, vol. 28, no. 3, pp. 376—384.

Iomdin L. Microsyntactic Phenomena as a Computational Linguistics Issue. *Grammar and Lexicon: Interactions and Interfaces. Proc. of the Workshop*. Osaka, 2016, pp. 8—18. Available at: <http://aclweb.org/anthology/W/W16/W16-38.pdf>.

Iomdin L. L. [What to do about constructions like “what to do”?]. *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”]. Iss. 16 (23), vol. 1. Moscow, RSUH Publ., 2017a, pp. 161—176. (In Russ.)

Iomdin L. L. [Between a syntactic phraseme and a syntactic construction. Nontrivial cases of microsyntactic ambiguity]. *SLAVIA, časopis pro slovanskou filologii*, ročník 86, sešit 2—3. 2017b, pp. 230—243. (In Russ.)

**Iomdin L. Microsyntactic Annotation of Corpora and its Use in Computational Linguistics Tasks. *Jazykovedný časopis, ročník 86, číslo 2, 2017c, pp. 169—178.***

Iomdin L. L. [Once more about microconstructions formed with function words: “to i delo”]. *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”]. Iss. 17 (24). Moscow, RSUH Publ., 2018, pp. 267—283. (In Russ.)

**Iomdin L., Sizov V. Structure Editor: a Powerful Environment for Tagged Corpora. *MONDILEX Fifth Open Workshop. Ljubljana, 2009, pp. 1—12.***

Lyashevskaya O., Drozanova K., Zeman D., Alexeeva M., Gavrilova T., Mustafina N., Shakurova E. Universal Dependencies for Russian: A New Syntactic Dependencies Tagset. *NRU HSE. Series WP BRP “Linguistics”*, 2016, no. 44.

Marakasova A. A., Iomdin L. L. [Microsyntactic tagging in the SynTagRus corpus of Russian texts]. *Informatsionnye tekhnologii i sistemy 2016 (ITiS’2016). Trudy 40-i mezhdistsiplinarnoi shkoly-konferentsii IPPI RAN* [Information Technologies and Systems 2016 (ITiS’2016). Proc. of the 40th Interdisciplinary School-Conference of IITP RAS]. St. Petersburg, 2016, pp. 445—449. (In Russ.)

Mel'čuk I. A. Opyt teorii lingvisticheskikh modelei “Smysl ↔ Tekst” [The theory of linguistic models of the Meaning — Text type]. Moscow, Nauka Publ., 1974. 314 p. (In Russ.)

(2nd ed.: 1999. 346 p.)

Mitkov R. Anaphora resolution. Longman, 2002. 220 p.

Nikolaeva T. M. *Funktsii chastits v vyskazyvanii (na materiale slavyanskikh yazikov)* [The function of particles in an utterance (based on the materials of Slavonic languages)]. Moscow, Nauka Publ., 1985. 170 p. (In Russ.)

Nikolaeva T. M. *Neparadigmaticheskaya lingvistika (Istoriya “bluzhdayushchikh chastits”)* [Non-paradigmatic linguistics (History of the “wandering particles”)]. Moscow, LRC Publ., 2008. 689 p. (In Russ.)

Nivre J. Towards a Universal Grammar for Natural Language Processing. *Proc. of Computational Linguistics and Intelligent Text Processing (CICLing 2015)*. Part 1. Springer, 2015, pp. 3—16.

Nivre J., Abrams M., Agić Ž. et al. Universal Dependencies 2.2. *LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL)*. Faculty of Mathematics and Physics, Charles University, 2018. Available at: <http://hdl.handle.net/11234/1-2837>.

Nivre J., Boguslavsky I., Iomdin L. Parsing the SynTagRus Treebank of Russian. *Proc. of the 22nd International Conference on Computational Linguistics (COLING'08)*. Manchester, August 18—22 2008. Eds. D. Scott, H. Uszkoreit, 2008, vol. 1, pp. 641—648.

Osenova P., Simov K. Modelling multiword expressions in a parallel Bulgarian-English newsmedia corpus. *Multiword expressions: Insights from a multi-lingual perspective. Phraseology and Multiword Expressions*. Eds. M. Sailer, S. Markantonatou. Berlin, Language Science Press, 2018, pp. 247—270.

Rhodes R. Tautological constructions in English ... and beyond. *Presented to the Syntax and Semantics Circle, UCB, 2009*.

Available at: [http://linguistics.berkeley.edu/~russellrhodes/pdfs/syntax\\_circle\\_taut\\_qp.pdf](http://linguistics.berkeley.edu/~russellrhodes/pdfs/syntax_circle_taut_qp.pdf).

Rosén V., De Smedt K., Losnegaard G. S., Bejček E., Savary A., Osenova P. MWES in Treebanks: From Survey to Guidelines. *Proc. of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia, 2016, pp. 2323—2330.

Savary A., Sangati F., Candito M. et al. The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. *Proc. of the 13th Workshop on Multiword Expressions (MWE 2017)*, Valencia, Spain, 4 April 2017, pp. 31—47

Shemanaeva O. Yu., Frolova T. I. [Tagging with lexical functions in SynTagRus]. *Informacionnye tehnologii i sistemy 2010 (ITiS'10)*. *Trudy 33-i Konferencii molodykh uchenykh i spetsialistov IPPI RAN* [Information Technologies and Systems 2010. Proc. of the 33rd Conference of Young Scientists and Specialists of IITP RAS]. Moscow, IITP, 2010, pp. 320—324. (In Russ.)



Toldova S., Roytberg A., Ladygina A., Azerkovich I., Vasilyeva M. D. Error analysis for anaphora resolution in Russian: new challenging issues for anaphora resolution task in a morphologically rich language. *Proc. of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016), co-located with NAACL 2016, San Diego, California, June 16, 2016*. Stroudsburg, PA, Association for Computational Linguistics, 2016, pp. 74—83.

Toldova S., Roytberg A., Nedoluzhko A., Kurzukov M., Ladygina A., Vasilyeva M., Azerkovich I., Grishina Y., Sim G., Ivanova A., Gorshkov D. Evaluating Anaphora and Coreference Resolution for Russian. *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”]. Iss. 13 (20). Moscow, RSUH Publ., 2014, pp. 681—695.

Zholkovsky A. K., Mel'čuk I. A. [On semantic synthesis]. *Problemy kibernetiki. Vyp. 19* [Problems of cybernetics. Iss. 19]. Moscow, Nauka Publ., 1967, pp. 177—238. (In Russ.)

**Д. В. Сичинава**  
Институт русского языка им. В. В. Виноградова РАН;  
НИУ «Высшая школа экономики»  
(Россия, Москва)  
*mitrius@gmail.com*

## **ПАРАЛЛЕЛЬНЫЕ КОРПУСА В СОСТАВЕ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА: НОВЫЕ ЯЗЫКИ И НОВЫЕ ЗАДАЧИ\***

В статье рассказывается об основных направлениях пополнения и содержательного развития параллельных корпусов НКРЯ за 2015—2019 гг. В разделе «Новые языки» речь идёт о новых языковых парах, возникших за этот период, об архитектуре и разметке соответствующих корпусов. По сравнению со списком языков, образующих двуязычные параллельные пары с русским и доступных в 2015 году, в НКРЯ появились следующие новые языки: башкирский, бурятский, китайский, литовский, финский, чешский, шведский. Продолжается традиция создания параллельной части НКРЯ при помощи ряда автономных российских и зарубежных команд, координирующих свои усилия с группой разработчиков Корпуса в Москве. Практически все новые языки ставили перед разработчиками Корпуса те или иные особые задачи, связанные с их морфологической или иной пословной разметкой. Существенно вырос за четыре года объём также некоторых из уже доступных в 2015 г. языковых пар.

В разделе «Новые задачи» раскрываются основные содержательные направления, разрабатываемые в рамках разных языковых корпусов — региональное разнообразие языка, жанровое разнообразие стилей, расширение функциональности и типов разметки и др. На настоящем этапе, как благодаря увеличению типологического разнообразия задействованных языков и письменностей, так и благодаря использованию более сложных морфологических анализаторов, существенно расширился набор дополнительных параметров разметки, поиск по которым может быть релевантен для работающего с параллельным корпусом. Цель включения в корпус образцов полицентричных языков стала одной из важнейших. Жанровое

---

\* Статья написана (и большинство обсуждаемых корпусов разработано) при поддержке проекта РФФИ № 17-29-09154 «Динамика языковой системы: корпусное исследование синхронной вариативности и диахронических изменений в текстах разных типов» (руководитель Г. И. Кустова).

разнообразие корпусов, в 2015 году лишь намечавшееся, в 2019 году является одной из главных целей, причем эта цель учитывается с самого начала создания новых языковых пар.

В статью включен отдельный исследовательский сюжет — изучение плюсквамперфекта по многоязычному корпусу. Анализ многоязычного текста по материалам расширенной коллекции позволяет построить сеть расстояний между данными для граммемы плюсквамперфекта в 24 идиомах Европы.

*Ключевые слова:* параллельные корпуса; двуязычные корпуса; многоязычные корпуса; разметка; репрезентативность; плюсквамперфект

Со времени публикации предыдущей статьи о параллельных (включая многоязычный) (под)корпусах в составе Национального корпуса русского языка [Сичинава 2015] архитектура и функциональность параллельных корпусов в составе НКРЯ не претерпели радикальных изменений. При этом существенно увеличился объём этих корпусов: на май 2019 г. в поиске доступно 98,2 млн словоупотреблений — близко к психологически важной отметке в 100 млн — по сравнению с 70 млн в 2015 г.; таким образом, соответствующий раздел НКРЯ становится одним из крупнейших лингвистических параллельных корпусов, наряду с InterCorp в составе Чешского национального корпуса (<http://ucnk.korpus.cz/intercorp/>). Появились новые языки, представленные в двуязычных парах, появились новые задачи, связанные как с разметкой корпусов, так и с функциональностью поиска.

Существенно вырос за четыре года также объём некоторых из уже доступных в 2015 г. языковых пар:

Язык	Текущий объём, обе языковые пары, словоформы на обоих языках, млн словоупотреблений	Прирост объёма по сравнению с 2015 г.
Английский	28,3	15 %
Белорусский	10,9	63 %
Испанский	2,4	88 %
Итальянский	4,8	21 %
Латышский	3,0	324 %
Немецкий	9,7	27 %
Французский	5,1	120 %
Эстонский	0,6	51 %

В дальнейшем, как и в работе [Сичинава 2015], сокращённое обозначение «L-й корпус», где L-й — название языка, используется как практическая замена для более громоздкого сочетания «параллельные L-русский и русско-L-й корпус».

## 1. Новые языки

По сравнению со списком языков, образующих двуязычные параллельные пары с русским и перечисленных в [Сичинава 2015], в НКРЯ появились следующие новые языки: башкирский, бурятский, китайский, литовский, финский, чешский, шведский. Совокупный объём только этих новых корпусов составляет 12

млн размеченных словоупотреблений, или токенов, причем более половины этого количества приходится на шведский корпус, благодаря его главной разработчице Н. В. Перковой вошедший в пятерку самых крупных параллельных корпусов НКРЯ по объёму. В разработке этих корпусов активно принимали участие специалисты, работающие на территории распространения соответствующих языков. Таким образом, продолжается традиция создания параллельной части НКРЯ при помощи ряда автономных российских и зарубежных команд, координирующих свои усилия с группой разработчиков Корпуса в Москве. Во многих случаях речь идёт о тесном сотрудничестве с командами уже существующих одноязычных корпусов, национальных или иных, об обмене текстами, а также и об интеграции с другими «семействами» параллельных корпусов, например, для чешского или финского. Практически все новые языки ставили перед разработчиками Корпуса те или иные особые задачи, связанные с их морфологической или иной пословной разметкой. Отметим, что почти во всех указанных случаях (кроме чешского) речь идёт о «внешней» разметке по отношению к базовой используемой в Корпусе морфологической разметке «Яндекса», а затем о той или иной ее адаптации под используемый формат Корпуса. Здесь также естественным было заимствование тех или иных решений из существующей практики одноязычных корпусов вместе с унификацией грамматических обозначений, которую диктует общая архитектура параллельных корпусов НКРЯ.

### *1.1. Башкирский корпус*

В состав НКРЯ включены параллельные башкирско-русские и русско-башкирские тексты, подготовленные в 2016 г. под общим руководством Б. В. Орехова командой разработчиков-волонтеров из Башкирии для компании «Яндекс», совокупным объёмом 550 тысяч слов. Для включения в Корпус эта коллекция потребовала некоторой доработки, в частности, создания хотя бы упрощенной метатекстовой разметки (в первоначальной версии она отсутствовала, а также было широко представлено как слияние, так и разделение изначальных текстов). Все файлы снабжены башкирской морфологической разметкой, разработанной Б. В. Ореховым [2014], а также переведены им же специально для НКРЯ в принятый в корпусе формат XML.

Основная масса текстов — новости и статьи из двуязычных газет и электронных изданий 2010-х годов, но широко представлены и другие жанры. В частности, впервые в практике НКРЯ в корпус включаются нетекстовые (при этом изначально именно двуязычные параллельные) жанры — башкирско-русский словарь и разговорник. Это способствует резкому повышению охвата представленных в корпусе лексики, фразеологии и конструкций, хотя, безусловно, подобные источники имеют скорее пограничный статус между текстом как таковым и базой данных. Ср. полезное для диахронического исследования лексики и фразеологии включение в украинский одноязычный корпус ГРАК [Шведова та ін. 2017—2019] полностью «Словаря украинского языка» Б. Д. Гринченко и «Русско-украинского словаря

устойчивых выражений» И. О. Виргана и М. М. Пилинской, которые, между прочим, тоже являются двуязычными — толкования или соответствия украинской лексики и фразеологии даются в них по-русски. Отметим также, что функция «словарь как корпус» позволяет искать внутри словосочетаний и текстовых примеров из статей по самым разным параметрам, а не только по вокабуле статьи или даже отдельным словам из ее текста.

Кроме того, в башкирский корпус входят, полностью или в отрывках, художественные произведения (в том числе изначально написанные на третьем языке, но переведенные на башкирский с русского, например, «Маленький принц» Сент-Экзюпери), официально-деловые тексты (Конституция Башкирии или тексты с сайта Курултая) и развернутые научно-популярные тексты из Википедии, весьма точно переложенные с русского: *Гассман Сальериҙы танылған опера либреттоһы оҫтаһы, һарай шағиры Пьетро Метастазио менән таныштыра, уның йортонда Вена интеллектуалдары һәм артистары йыйылыр була...* (из имеющей статус «избранной статьи» в обеих языковых версиях биографии Антонио Сальери). Кроме Википедии и художественных произведений, направление перевода в этой коллекции текстов не всегда очевидно, что, разумеется, является обычной ситуацией для двуязычных сайтов вообще и СМИ в частности, особенно на постсоветском пространстве. Таким образом, разбиение этого корпуса на башкирско-русскую и русско-башкирскую параллельные части носит в значительной степени условный характер.

## 1.2. Бурятский корпус

Создание бурятского корпуса поддерживалось специальным проектом РФФИ № 15-46-04417 («Бурятско-русский параллельный корпусный модуль») и осуществлялось совместно с Институтом монголоведения, буддологии и тибетологии РАН (Улан-Удэ) под руководством Л. Д. Бадмаевой. В этот корпус объемом в 400 тыс. словоупотреблений вошли только художественные тексты: один переведенный с русского (пушкинская «Капитанская дочка») и несколько произведений довоенной и послевоенной бурятской литературы XX в., включая один из ключевых текстов бурятской культуры «Путь праведный» Б. Дандарона; авторство, текстология и история перевода этого романа представляют собой отдельную исследовательскую проблему.

Важным «вызовом» для создателей этого корпуса являются достаточно вольные переводы, в том числе с пропусками, что, опять-таки, характерно для традиции художественного перевода с «языков народов СССР» и на них; в значительном проценте случаев находимые примеры не имеют точных соответствий на другом языке. Тем не менее подобным образом выровненные тексты имеют большую ценность как для истории перевода, так и для лингвистического анализа. Тексты получают морфологическую разметку, разработанную Т. А. Архангельским и принятую в одноязычном бурятском корпусе (<http://web-corpora.net/BuryatCorpus>); большинство лемм снабжено также русским переводом с опорой на специально оцифрованный бурятско-русский словарь.

### 1.3. Китайский корпус

Создание, поддержка и развитие китайского корпуса, насчитывающего в настоящее время 280 тыс. токенов (число уникальных слов меньше, поскольку в эту сумму входят как разделенные на «слова»-сегменты китайские тексты в оригинальной графике, так и дублирующая их строка транслитерации), безусловно, является одной из самых сложных задач, стоящих перед коллективом НКРЯ. Это диктуется, прежде всего, спецификой китайского иероглифического письма, требующего особой автоматизации процессов словоделения, транскрипции и выравнивания.

В его разработке участвует коллектив сотрудников и студентов из российских (ВШЭ, РГГУ, МГУ, РАНХиГС, Алтайский государственный университет) и китайских вузов; в частности, ключевые роли в проекте играют Л. С. Холкина, К. И. Семенов, О. Р. Валиулин, С. П. Дурнева, М. Н. Якубов. Поиском и предоставлением данных, наряду с выравниванием текстов, занимаются Синь На (и коллектив под руководством Е Цисуна) Института лексикографии Хэйлунцзянского университета, а также Юань Мяосюй (и коллектив под руководством Ван Юн) Института иностранных языков Чжэцзянского университета.

В настоящее время в поиске НКРЯ представлены лишь художественные произведения (три рассказа Лу Синя, русская и советская классика — романы при этом представлены в отрывках), однако выровнены и готовятся к вывеске в Корпусе также тексты разнообразных жанров: это корпус деловой переписки, разработанный К. А. Ульяновой, Евангелие в различных, относящихся к разным историческим периодам и конфессиям, переводах на китайский, выровненное с русским синодальным переводом, учитывавшимся при создании части этих переводов, а также другие художественные тексты: Пушкин, Гоголь, Тургенев, Чехов, Горький, современные китайские авторы — Мо Янь, Лю Чжэньюнь, Юй Хуа.

В настоящее время для выравнивания текстов и словоделения используется коммерческая утилита Skuuper (с некоторым усовершенствованием ее механизма, разработанным М. Н. Якубовым и другими участниками). В текстах, представленных в настоящее время в поиске НКРЯ, использовался «жадный» алгоритм словоделения, при котором сегментирование осуществлялось слева направо автоматически на основании наиболее длинной цепочки иероглифов, доступной как словарный вход. Это давало в принципе приемлемый результат (по данным М. Н. Якубова [2017] и В. А. Морозовой [2018] — 79%), но иногда приводило к неверным анализам, при которых алгоритм при движении по тексту слева направо «попадал» в более длинное и менее вероятное в тексте редкое слово. Например, в рассказе Чехова «Человек в футляре» трижды выделяется имеющееся в словаре слово 的姐 [dǐjiě] ‘таксистка’, поскольку частотный иероглиф 的, выступающий в подавляющем большинстве случаев с чтением [de] как показатель атрибутивной связи, имеет также гораздо более редкое чтение [dǐ] ‘такси’, которое объединяется с первым иероглифом идущего далее в тексте слова 姐姐 [jiějie] ‘старшая сестра’.

Для словарной разметки используется свободно доступный электронный китайско-английский словарь CEdict. В статьях этого словаря выделены разные поля для разметки — транскрипция, толкование и требуемый показатель класса (счетное слово); ряд иероглифов, имеющих грамматические значения, размечен дополнительно как грамматические показатели. На основании словарной транскрипции китайский текст продублирован в виде строки сплошной транскрипции с указанием всех точек, где чтение теоретически может быть неоднозначно. Соответствующий алгоритм разметки разработан Е. А. Кузьменко и усовершенствован М. Н. Якубовым (ср. также [Якубов 2017]).

Принципиальной сложностью такой разметки является неоднозначность словоделения и транскрипций и высокая степень омонимии и полисемии: для большинства слов, состоящих из одного иероглифа, и для значительной части более длинных приводятся все альтернативные значения, представленные в словаре (включая крайне малочастотные). Отдельной проблемой является разметка транскрибируемых иероглифами имен собственных (и тем самым само выделение этих слов в тексте), далеко не все из которых имеются в словаре — хотя набор включенных в CEdict имен собственных, в том числе даже названий знаменитых текстов, велик, фамилии русских литературных персонажей или русские микротопонимы, разумеется, в нем отсутствуют. Ниже эта разметка будет разобрана несколько подробнее с точки зрения представленной в ней информации (раздел 2.1)

#### *1.4. Литовский корпус*

Литовский корпус (как и латышский<sup>1</sup>, шведский, эстонский и финский) представляет собой часть подпроекта по созданию параллельного корпуса с участием литературных языков циркумбалтийского ареала [Sitchinava, Perkova 2019]; об этом ареале с лингвистической точки зрения см. [Dahl, Koptjevskaja-Tamm (ed.) 2001]. Выбор, подготовка и выравнивание текстов осуществлены Н. В. Перковой; морфологическая аннотация корпуса основывается на системе онлайн-разметки, разработанной в университете Витовта Великого в Каунасе [Rimkutė, Daudaravičius, Utkā 2007]. Объем корпуса, доступного в Интернет-поиске — 550 тысяч словоупотреблений, пополнение продолжается. Корпус и его подготовленное пополнение включают литовские художественные тексты XX — начала XXI в. (Вайжгантас, А. Шкема, И. Мерас, К. Сая, Ю. Апутис, Р. Гавялис, М. Ивашквичюс), а также переводы с русского (Чехов, Хармс, Бунин). Параллельный корпус НКРЯ (вместе с некоторыми нехудожественными текстами, например, эссе А. Венцловы) используется в разработанном О. Н. Ляшевской литовском модуле синтаксического корпуса Universal Dependencies ([https://universaldependencies.org/treebanks/Lt\\_hse/index.html](https://universaldependencies.org/treebanks/Lt_hse/index.html)).

<sup>1</sup> Латышский корпус (о котором см. [Perkova, Sitchinava 2016]), как уже указывалось выше, — наиболее значительно пополнившаяся за последние 4 года из существующих двуязычных пар (более чем вчетверо).

### **1.5. Финский корпус**

Финский корпус НКРЯ развивается в тесном сотрудничестве с Тамперским университетом и проектом FinCLARIN, в рамках которого уже разработаны параллельный финско-русский и русско-финский корпуса ПарФин и ПарРус [Михайлов, Хярме 2015]. Стороны обмениваются текстами с целью избежать дублирования работы, кроме того, в проекте НКРЯ используется открытый морфологический анализатор OMorfi (<https://github.com/flammie/omorfi>), используемый и в проектах университета Тампере.

В финский корпус НКРЯ (подбором текстов, сканированием, распознаванием, вычиткой и выравниванием занимается К. О. Мищенко) входит более 1 млн словоупотреблений, доступных в поиске на лето 2019 г., и подготовлено пополнение в 200 тыс. словоупотреблений. В финско-русскую часть включены художественные, научные и публицистические тексты, в том числе 50 статей новостного агентства YLE. Хронологический охват корпуса — с 1880-х годов, включая финскую классическую прозу (А. Киви, Ю. Ахо, М. Лассила, Ф. Э. Силланпяя, М. Валтари, М. Ларни и др.) Русско-финская часть представлена пока только романом Солженицына «В круге первом», а также теми договорами в коллекции межгосударственных договоров с Финляндией, которые «диктовались» российской (советской) стороной и с большой долей вероятности были переведены с русского, хотя, как известно, определение направления перевода является существенной проблемой для нехудожественных текстов в целом, далеко не только юридических.

В дальнейшем планируется пополнить финский корпус текстами, уже входящими в разрабатываемые в университете Тампере корпуса ПарРус и ПарФин.

### **1.6. Чешский корпус**

Чешский язык долгое время был заметной лакуной в параллельном корпусе НКРЯ — и это вызывало несомненные сожаления, учитывая как большое число славянских языков, уже представленных в двуязычных парах, так и существенное развитие корпусной лингвистики в Чехии, в том числе и в области параллельных корпусов. С 2018 года чешский корпус разрабатывается при участии Т. А. Малышевой (инициатора проекта), в том числе на базе текстов, взятых из параллельного корпуса А. Барентсена ASPAC. Как и в случае финского проекта, установлено сотрудничество (взаимовыгодный обмен текстами) с параллельным корпусом InterCorp в рамках Чешского национального корпуса (<http://ucnk.korpus.cz/intercorp/>). Для пословной разметки используется морфологический анализатор компании «Яндекс».

Объём корпуса уже в его пилотной версии, доступной с 2019 г., превосходит миллион словоупотреблений. В него входит несколько современных публицистических текстов, а также классические художественные произведения (четыре части гашековского «Швейка», «Невыносимая лёгкость бытия» М. Кундеры, Пушкин, Чехов, Булгаков, Ильф и Петров, Набоков и др.)



### 1.7. Шведский корпус

Шведский корпус, инициатором и основной разработчицей которого является Н. В. Перкова, как уже сказано, успел войти в пятерку крупнейших параллельных корпусов НКРЯ. Подобно ранее начавшемуся проекту Н. В. Перковой (латышскому корпусу), ставится цель хронологически репрезентативно охватить шведский «культурный канон» от Стриндберга до Бойе и Лагерквиста, представленный в русских переводах, а также наиболее важные произведения современной шведской прозы, довольно активно переводящейся на русский (Д. Ваттин, К.-Й. Вальгрэн, К. Киери и др.). При этом не забывается и противоположное — русско-шведское направление, также представленное как классикой, так и современными романами (М. Шишкин, М. Степнова, Л. Петрушевская и др.). На следующем этапе в корпус введены также современные новостные и аналитические тексты из шведской прессы; соответствующими работами под руководством К. Окерман-Саркисян и О. Янссон занимались студенты Упсальского университета М. Лундгрэн и Э. Маттссон.

Тексты шведского корпуса размечены при помощи анализатора Stagger. Данный анализатор [Östling 2013] в современной его версии отличается высоким качеством морфологической разметки с автоматическим снятием омонимии, при разметке реализуется нейросетевая модель.

## 2. Новые задачи

### 2.1. Лингвоспецифичная разметка

На этапе, описанном в [Сичинава 2015], масштаб лингвоспецифичности разметки в параллельных корпусах НКРЯ был сравнительно невелик: конкретный набор тегов для грамматических категорий разных языков различался, однако схема морфологической разметки была одинаковой и заключалась в наборе частеречных и грамматических тегов (линейная упорядоченность этого набора при поиске роли не играет), аналогично принятому в НКРЯ с неснятой омонимией. В современном состоянии корпуса соблюдается принцип, согласно которому сопоставимые части речи и значения грамматических категорий в разных языках размечаются одинаково. В основном разметка ранее не встречавшихся граммем в новых языках (а инвентарь их вырос значительно: например, различные направительные и местные падежи в прибалтийско-финских, типы конвербов в бурятском и др.) следует «Лейпцигским правилам глоссирования» (<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>), хотя и с отклонениями, необходимыми для сохранения унифицированной разметки «со старыми» корпусами в составе НКРЯ (например, прошедшее время — латинизированное **praet**, а не англизированное **PST**).

Однако на настоящем этапе, как благодаря увеличению типологического разнообразия задействованных языков и письменностей, так и благодаря использованию

более сложных морфологических анализаторов, существенно расширился набор дополнительных параметров разметки, поиск по которым может быть релевантен для работающего с параллельным корпусом. Соответственно, разные языки имеют теперь не только разные наборы тегов (tagsets), но и разный набор атрибутов в пословной XML-разметке.

Наиболее масштабно отличается от среднекорпусного стандарта разметка китайских текстов. Это, разумеется, диктуется как «экзотической» (иероглифической) письменностью, требующей для неспециалиста дублирования в латинской транслитерации и хотя бы упрощенного глоссирования (ранее в корпусе такое решение, как и словарная разметка переводов лексики, уже принималось для армянского языка), так и типологическими отличиями китайского от ранее представленных в корпусе языков, в основном индоевропейских. Простой дихотомии «лексика vs. грамматика» в пословной разметке в целом ряде случаев недостаточно, кроме того, проблемным, в том числе в связи с письменностью, не знающей словоразделов, здесь является само словоделение. Размечаемые сегменты (токены), с точки зрения устройства XML-разметки аналогичные словам в других корпусах, в китайском корпусе делятся на «лексические» (содержащие один или несколько иероглифов) и «грамматические» (один иероглиф). «Грамматические» сегменты задаются списком (разработан Л. С. Холкиной), в большинстве случаев они могут быть омонимичны «лексическому» сегменту. «Грамматические» сегменты получают специфические грамматические пометы: «MOD» (модальная частица 了), «PFV» (перфектив), «PRG» (прогрессив), «PST» (прошедшее время), «EVAL» (оценка действия), «QUEST» (общий вопрос), «CAUS» (каузатив), «PL» (множественное число), «BA» (вынесение объекта в позицию перед глаголом), «ATRN» (определение к имени), «ATRV» (определение к глаголу), «PASS» (пассив), «DIR» (директивные частицы). Кроме того, часть «лексических» сегментов получают по словарю разметку одного грамматического (в данном случае словоклассифицирующего) параметра — классификатора, с которым сочетается данное слово, аналогично роду в европейских языках. Для разметки, как уже указано, используется китайско-английский словарь CEdict, из которого берутся все представленные в словаре толкования для того или иного лексического сегмента. Кроме того, китайским словам и предложениям приписана транслитерация в системе пиньинь, взятая из указанного словаря (с заменой цифрового обозначения тонов на диакритики) и снабженная «европеизированной» пунктуацией. Если один и тот же сегмент имеет несколько транслитераций, в том числе различающихся тонами, то приводятся через знак «/» все варианты (сохраняем автоматически введенное словоделение):

- (1) Tiānqì shì nàiyàng cháoshì **hé/hè/hú/huó/huò** duō wù, hǎobù róngyì cái tiānliàng. **cóng/cōng/zòng** chēxiāng chuāngkǒu wàng qù, tiělù zuǒyòu 10 bù lù **yuǎn/yuàn de/dī/dí/dì dìfāng/dìfāng** jiù hěn nánkàn qīng shénme **dōngxī/dōngxī**.

Было так сыро и туманно, что насилу рассвело; в десяти шагах, вправо и влево от дороги, трудно было разглядеть хоть что-нибудь из окон вагона. [Достоевский, Идиот].

В начальной части «Идиота» на 46 671 китайский сегмент (аналог слова), выделенный программой словоделения, приходится 16 893 знака неоднозначности «/» — то есть в среднем один лишний транскрипционный вариант на три сегмента.

Таким образом, каждой китайско-русской выровненной паре предложений отвечает тройка строк — одна с русским текстом и две с китайским, в иероглифической и транслитерированной форме; кроме того, транслитерация сегментов продублирована в качестве дополнительного поля пословного разбора в иероглифической версии текста.

Разумеется, такая система имеет определенные недостатки, поскольку предусматривает маловероятные варианты разбора и транслитерации (а значительное количество альтернативных переводов и чтений характерно как раз для высокочастотных сегментов) и в то же время «жесткий» выбор единственного варианта словоделения, однако она открыта для дальнейших улучшений и фильтрации маловероятных словарных разборов.

Переводная словарная информация (в поле *sem*) содержится также в бурятском корпусе, куда она введена Т. А. Архангельским на основании отсканированного и распознанного В. Ивановым словаря. Данная разметка покрывает большинство словоупотреблений, но нуждается в усовершенствовании (ручная чистка переводов, добавление отсутствующих лемм). Пример из «Капитанской дочки» Пушкина:

- (2) `<para id="17"><se lang="ru" variant_id="0">Я считался в отпуску до окончания наук.</se>  
<se lang="bu" variant_id=»1»><w><ana lex=»зүгөөр» gr=»A» sem=»но, однако_же, хотя»</ana>Зүгөөр</w> <w><ana lex=»би» gr=»S,nom» sem=»я»</ana>би</w> <w><ana lex=»эрдэм» gr=»S,nom» sem=»наука»</ana>эрдэм</w> <w><ana lex=»шудалха» gr=»V,conv,simult1» sem=»изучать, исследовать»</ana>шудалжа</w> <w>дүүргэтэрээ</w>, <w><ana lex=»табилга» gr=»S,dat» sem=»разрешение»</ana>табилгада</w> <w><ana lex=»гэхэ» gr=»V,conv,simult1» sem=»говорить»</ana>гэжэ</w> <w>тоотой</w> <w><ana lex=»байха» gr=»V,partcp,nonpast,1p,sg» sem=»быть»</ana>байгааб</w>.</se>  
</para>`

Дополнительный единообразный атрибут «словообразование» введен для эстонского (корпус этого языка уже существовал с 2015 г.) и финского корпусов. Как известно, в морфологии прибалтийско-финских языков большую роль играют продуктивные сложные слова, получающие в поле *lex* разбор как единая лексема, а в поле *wordf* — как сочетание нескольких основ (разделенных знаком плюса). Такая лексика получает дополнительную помету **compos**:

- (3) Финский ‘мировая война’:  
`<w><ana lex="maailmansota" wordf="maa+ilman+sota" gr="S,compos,gen,sg"/>maailmansodan</w>`

- (4) Эстонский ‘отечественная война’:  
`<w><ana lex=“isamaasõda“ wordf=“isa+maa+sõda“ gr=“S,compos,gen,sg/>isamaasõja</w>`

И наоборот, на данном этапе в параллельном корпусе уже разрешены орфографически неоднословные леммы и словоформы неизменяемых частей речи, содержащие пробел (в русском корпусе [Плунгян, Ляшевская, Сичинава 2005] они относились к «сложным лексическим единицам», обязательно сопровождаемым пословным разбором):

- (5) Литовский ‘когда-нибудь’  
`<w><ana lex=»kada nors« gr=»ADV=pos«/>kada nors</w>`
- (6) Латышский ‘ведь’, ‘уже’  
`<w><ana lex=“jau arī“ gr=“PART=“/>jau arī</w>`

## 2.2. Региональная вариативность языков

Параллельный корпус лишь в очень ограниченной мере может быть репрезентативным — его представительность ограничивается необходимым «оппортунистическим» критерием, то есть он в принципе может включать только тексты, которые уже были переведены, что искажает пропорции разных метатекстовых параметров по сравнению с генеральной совокупностью оригинальных текстов. Поэтому особую ценность представляют доступные в переводах тексты, представляющие те или иные региональные варианты литературного языка. Ранее в параллельных корпусах НКРЯ уже присутствовали образцы как британского, так и американского английского, а также тексты, созданные в конце XIX и начале XX в. на обоих исторических вариантах литературного украинского языка — западном (И. Франко, О. Кобылянская и др.) и центральном (И. Нечуй-Левицкий, М. Коцюбинский и др.). Сейчас же цель включения в корпус образцов полицентричных языков стала одной из важнейших.

Например, выросший почти вдвое за последние годы трудами К. О. Мищенко испанский корпус, помимо образцов собственно кастильской прозы, по замыслу основной разработчицы, теперь включает в себя художественные и публицистические тексты, созданные авторами из следующих латиноамериканских стран: Аргентина (это не кто иной, как Че Гевара), Колумбия (Г. Гарсиа Маркес), Куба (А. Карпентьер), Мексика (К. Фуэнтес), Парагвай (А. Роа Бастос), Перу (М. Варгас Льоса; причем включенный в корпус его роман посвящен событиям в Доминиканской республике); а помимо них, двуязычный автор из Барселоны Э. Мендоса, пишущий также по-каталански. Тексты СМИ (переведенные на сайте [inosmi.ru](http://inosmi.ru)), пополнявшие испанский корпус, происходят с испанских, аргентинских, венесуэльских, а также международных (имеющих локальную версию в нескольких испаноязычных странах) сайтов.

Финский корпус, также развиваемый К. О. Мищенко, включает в себя тексты Ю. Конкка, ингерманландца, писавшего на стандартном финском, но испытавшем определенное влияние родного диалекта. В свою очередь, в шведский корпус, архитектурой художественной части которого занимается Н. В. Перкова, входят (или

запланированы к включению) произведения представителей «финляндского шведского»: Г. Парланда, Т. Янссон, М. Фагерхольм, Ч. Вестё, переводчика Я. Даля.

В перечисленных случаях региональная вариативность литературного языка значима в разной степени, а в некоторых из них, возможно, само ее наличие является исследовательской задачей, требующей особого доказательства, тем не менее разнообразие текстов по географическому признаку — важный параметр, который потенциально следует учитывать в лингвистическом исследовании.

Вот примеры латиноамериканизмов в текстах из испанского параллельного корпуса:

- (7) Seguro que si estabas *vos* lo sacabas y Ana María creo que también, ya que no tienen esos complejos nochísticos que me dan a mí. [Ernesto Che Guevara. Notas de viaje (1952–1953)].  
Уверен, что, будь *ты* на моем месте, ты бы его вытащила, и Ана Мария, думаю, тоже, поскольку у вас нет связанных с темнотой комплексов. [Эрнесто Че Гевара. Дневник мотоциклиста (А. Ведюшкин, 2014)].
- (8) «No podía dormir por la rabia de estar pensando en él, pero lo que más rabia me daba era que *mientras más* rabia sentía, más pensaba». [Gabriel García Márquez. Vivir para contarla (2002)].  
Я не могла спать от бешенства, что он не идет у меня из головы, но *чем больше* я бесилась, тем больше о нем думала. [Габриэль Гарсиа Маркес. Жить, чтобы рассказывать о жизни (С. Марков, Е. Маркова, А. Малоземова, В. Федотова, 2012)].
- (9) Pero, para hacer frente a sus frugales necesidades, en sus horas libres fue vendedora en un supermercado, *mesera* en una pizzería de Boston... [Mario Vargas Llosa. La Fiesta del Chivo (2000)].  
Однако на жизнь, хотя потребности у нее были более чем умеренные, ей пришлось зарабатывать в свободные часы продавщицей в супермаркете, *официанткой* в бостонской пиццерии... [Марио Варгас Льюса. Нечестивец, или Праздник Козла (Людмила Синянская, 2004)].
- (10) Los comedores donde se hartaban, las aguas infectadas, los *excusados* públicos y apestosos y las recámaras donde ellos cogían y roncaban... [Carlos Fuentes. Gringo viejo (1985)].  
Я сжег грязные столовые, зараженные водоемы, зловонные *отхожие места*; конуры, где бесились и рычали псы... [Карлос Фуэнтес. Старый гринго (Маргарита Былинкина, 2010)].

### 2.3. Жанровое разнообразие корпусов

Жанровое разнообразие корпусов, в 2015 году лишь намечавшееся, в 2019 году является одной из главных целей, причем эта цель учитывается с самого начала создания новых языковых пар. Так, новый башкирский корпус — единственный, где

нехудожественные тексты преобладают над художественными, а также, как указано выше, включены тексты из Википедии и словарей.

Активно используется такой источник нехудожественных текстов, как русские переводы иностранной прессы на сайте ИноСМИ.ру. Несмотря на наличие только одного направления перевода, трудности в автоматизации скачивания, не всегда доступные тексты оригинальных СМИ (часто требующие отдельной подписки на платные версии их сайтов) и нередко вольный перевод, эту коллекцию переводной публицистики на десятках языков из сотен изданий трудно переоценить. В частности, тексты оттуда дополнили испанский, итальянский, шведский, английский, чешский, финский корпуса; готовится значительное пополнение и армянского корпуса.

Научные и философские тексты включаются во французский корпус — в частности, туда вошли работы Ж. Женетта, Ж. Кокто, М.-П. Рей, М. Бахтина, Н. Бердяева. Заметный прирост таких текстов произошёл и в итальянском корпусе (историческая работа Дж. Боффы «От СССР к России», физический трактат Т. Редже «Этюды о вселенной», «Кризис западной философии» В. Соловьева, а также «Проблемы поэтики Достоевского» Бахтина, причем перевод сделан с несколько иной редакции оригинала, чем во французском корпусе).

В испанский корпус введена подборка двуязычных путеводителей по Испании (подготовленная и выровненная Т. Горожанкиной) — данный тип дискурса в параллельном корпусе ранее был не представлен (да даже и в одноязычном НКРЯ он учитывается лишь в очень ограниченной степени). В китайском корпусе представлены деловая переписка и переводы Евангелия, делавшиеся с учетом славянских текстов (ранее библейские тексты включались в белорусский корпус НКРЯ). В итальянский корпус включены пресс-коммюнике и материалы российско-итальянской торговой палаты.

Жанровое разнообразие способствует разнообразию лексики и конструкций в корпусе: например, русское сочетание (грамматикализирующееся как предлог) *в рамках* в итальянском корпусе встретилось 24 раза, причем только 3 раза в художественных текстах; а в армянском корпусе, где нехудожественных текстов пока нет — только 1 раз (как канцеляризм в прямой речи у Толстого).

#### **2.4. Развитие поливариантных двуязычных корпусов**

Напомним, что в рамках особого проекта, начавшегося в 2012 г., французский корпус был в порядке эксперимента пополнен поливариантными переводами («Шинель», «Нос», «Обломов»). Основной целью этого проекта (продолжением его стал целый ряд проектов РНФ, РФФИ, частного фонда «Династия» и Швейцарского научного фонда, выполняемых в ИПИ РАН под руководством Анны А. Зализняк и О. Ю. Иньковой) стало создание надкорпусных баз данных по видо-временным формам, коннекторам и дискусивным словам. Позже в эти базы была включена и соответствующим образом размечена «Шинель» Гоголя, представленная в 15 итальянских переводах, и другие итальянские тексты с одним вариантом

перевода. Надкорпусные базы данных и соответствующие публикации доступны на сайтах <http://a179.frccsc.ru/PublicLingvoProjects/main.aspx> и <http://a179.frccsc.ru/RFH41002/main.aspx>.

Поливариантные корпуса ценны для выявления внутриязыкового пространства возможностей и точек варьирования. Нередко переводчики, будучи знакомы с предыдущими переводами, сознательно от них отталкиваются. Эта тенденция продолжается, как в «старых», так и в «новых» языковых разделах. Например, в латышском корпусе «Четыре поездки» В. Лациса представлены в переводах Г. Цейтлина и В. Ругайса, рассказы Чехова — в различных латышских переводах П. Калвы, Р. Эзеры, А. Гревини, О. Калнциемса, А. Курцийса; в шведском корпусе «Шляпа волшебника» Т. Янссон — в переводах В. Смирнова и Л. Брауде, а «Пеппи/Пиппи Длинныйчулок» — в переводах Л. Лунгиной и Л. Брауде, в немецком корпусе «Капитанская дочка» Пушкина — в переводах Ф. Отто и Ф. Фриш (?). Во французском корпусе, развиваемом в ходе упомянутых выше проектов ИПИ РАН, дважды представлено сочинение «Difficulté d'être» Ж. Кокто — в версиях под названиями «Тяжесть бытия» (Л. М. Цывьян) и «Трудность бытия» (М. Л. Аннинская)<sup>2</sup>. Поливариантными являются и файлы с различными китайскими переводами Евангелия (см. выше). Пока не все эти альтернативные версии перевода сведены в единые файлы, однако могут искаться в Корпусе одновременно и сопоставляться.

## 2.5. Развитие функциональности поиска

На текущей платформе Яндекс.Поиска ряд размеченных параметров пока не реализован (например, типология неточности перевода, о которой см. Сичинава 2015, или словообразование сложных слов). Частично подобный поиск может быть осуществлен в разрабатываемых ИПИ РАН надкорпусных базах данных по французскому и итальянскому корпусам (см. выше, 2.4).

Поставлена задача внедрить эти параметры в поиск после переезда корпуса на новые технологические рельсы. В ожидании же этого момента некоторые функции реализованы на движке «Цакорпус» (доступен по адресу [https://bitbucket.org/tsakorpus/tsakonian\\_corpus\\_platform/src/default/](https://bitbucket.org/tsakorpus/tsakonian_corpus_platform/src/default/); корпуса публикуются на сайте Linghub.ru) Т. А. Архангельским. В частности, возможен одновременный поиск в двух языках (например, можно выяснить, когда шведское *fika* переводится на русский как *[выпить] кофе*, а когда русское *простор* — как шведское *vidd*), получение частотных списков слов, статистика сочетаемости, статистика по дате создания текста.

<sup>2</sup> А в чешском корпусе одновременно (по случайному совпадению) появилась и «Nesnesitelná lehkost bytí» («Невыносимая лёгкость бытия») М. Кундеры; правда, существует как будто бы только один перевод этого романа.



The interface shows two query panels side-by-side. The left panel is for 'Word #1' in Russian, with fields for Word, Lemma (filled with 'простор'), Grammar, and Language/tier (set to 'Russian'). The right panel is for 'Word #2' in Swedish, with fields for Word, Lemma (filled with 'vidd'), Grammar, and Language/tier (set to 'Swedish'). Below these panels is a 'Full-text search' field and a 'Precise match' checkbox. At the bottom, there are buttons for 'Search sentences', 'Search words / lemmata', 'Select subcorpus', and navigation icons.

Search result: 9 occurrences, 9 sentence(s) found in approximately 6 document(s).

Рис. 1. Образец поискового интерфейса на linghub (движок «Цакорпус») для одновременного поиска в обоих языках

## 2.6. Расширение многоязычных корпусов

В рамках проекта Йенского университета (руководители Р. фон Вальденфельс и И. С. Левин) ведется работа по созданию многоязычного параллельного корпуса ParaPooh, который включает в себя переводы «Винни-Пуха» на более чем 50 языков из коллекции автора настоящей статьи (некоторые переводы приобретены специально для корпуса). Сканированием и распознаванием текстов занимается О. Д. Шиншинова (МГУ). По окончании работ тексты пополняют как многоязычный

корпус ParaSOL [von Waldenfels 2006] так и, по крайней мере частично, НКРЯ (в котором в настоящее время «Винни-Пух» представлен на 16 языках).

Анализ многоязычного текста «Винни-Пуха» по материалам расширенной коллекции позволяет построить следующую сеть NeighbourNet (о построении сетей расстояний между данными и их использовании в кластерном анализе см. Сичинава 2015) для грамлеммы плюсквамперфекта в 24 идиомах Европы.

Всего в тексте представлено 258 предложений, в которых плюсквамперфект отмечен хотя бы в одном языке. В 89 из этих

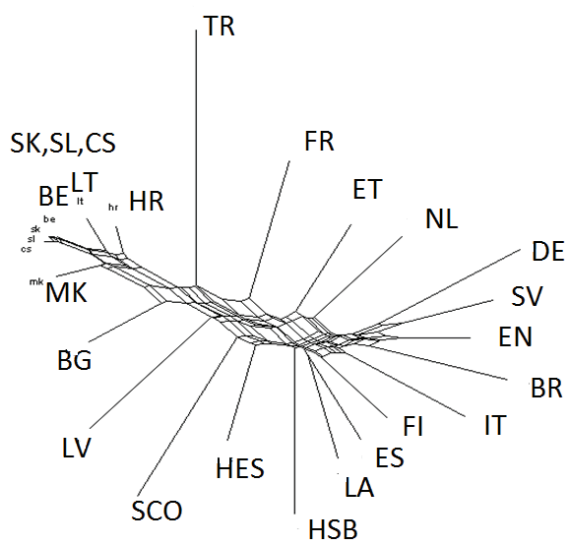


Рис. 2. Сеть NeighbourNet для плюсквамперфекта в переводах «Винни-Пуха»



случаев плюсквамперфекта нет в английском оригинале; часто это связано с использованием нефинитных средств выражения (перфектного инфинитива, номинализации, герундия). Однако выделяется, в частности, класс случаев «ошибочное решение / исправление»; для его кодирования в английском используется претерит, а некоторые языки выбирают плюсквамперфект, маркирующий отмененный результат или смежное значение:

- (11) англ. *I forgot*.  
швед. *Det hade jag glömt*.  
латыш. *Es biju aizmirsis*.  
верхнелуж. *Běch to zabył*.  
‘Я позабыл [что сам съел мёд]’.

Интересен случай *ирреального* действия, предшествовавшего реальному; здесь, например, в близкородственном нидерландском языке также выбирается плюсквамперфект, в то время как в оригинале выступает простое прошедшее:

- (12) англ. *You gave him don't you remember — a little — a little — I gave him a box of paints to paint things with. — That was it. — Why didn't I give it to him in the morning?*  
нид. *Jij hebt hem een... Weet je het niet meer?... een kleine... een kleine... — O, ja...! Een verfdooisje om mee te verven! — Precies. — Maar waarom had ik het hem niet meteen ,s morgens gegeven?*

‘Ты подарил ему, разве ты не помнишь, ну эту... маленькую... маленькую... — Я подарил ему коробочку с красками, чтобы он мог ими рисовать. — Именно так. — А почему я не *подарил* ее ему тем утром?’

Приведенный выше граф NeighbourNet показывает, что языки Европы по употребительности этой формы в различных контекстах распределены между двумя полюсами:

**максимально частое употребление, квазиобязательное использование** в таксисных контекстах: германские языки — английский [EN, язык оригинала], нидерландский [NL], немецкий [DE], шведский [SV] (их кластеризация на графе, несмотря на указанные выше частные различия, демонстрирует схождение генетических и типологических характеристик); бретонский [BR, кельтские]; латинский [LA, италийские] и два романских языка — испанский [ES], итальянский [IT]; прибалтийско-финские языки — финский [FI] и эстонский [ET], которые, однако, не образуют кластера друг с другом);

**редкое использование;** в контрфактических или антирезультативных контекстах (словацкий [SK], словенский [SL], чешский [CS], белорусский [BE]); несколько более частое применение для подчеркивания последовательности временных планов (македонский [MK], болгарский [BG], хорватский [HR], литовский [LT]<sup>3</sup>).

<sup>3</sup> В целом плюсквамперфект используется в литовском языке чаще, чем в славянских. Возможно, это индивидуальная особенность данного перевода. Кроме того, надо учесть, что английский

Об употреблении плюсквамперфекта в славянских языках (с примерами близких к «инвариантным» контекстов) подробнее см. [Сичинава 2019].

Промежуточную позицию между этими двумя полюсами занимают турецкий [TR] (употребления соответствующих форм в этом языке вообще стоят особняком; они маркируют дигрессии, состояния во временном плане прошлого, недостигнутый результат), латышский [LV], французский [FR], верхнелужицкий [HSB]. Отметим, что литературные микроязыки на диалектной базе (скотс [SCO] и гессенский [HES]) демонстрируют заметно более редкое употребление плюсквамперфекта, чем стандартные английский и немецкий соответственно.

### 3. Промежуточные итоги

Дальнейший план развития параллельных корпусов, после стабилизации новой версии Яндекс.Поиска для НКРЯ, включает в себя:

а) дальнейшее добавление новых языковых пар с русским (предварительно речь идёт о словенском, сербском, словацком, венгерском, японском, португальском, а, возможно, и ряде других языков);

б) развитие существующих языковых пар (например, добавление болгарско-русских текстов — сейчас в корпусе представлены только русско-болгарские; добавление нехудожественных текстов в корпуса, где они сейчас отсутствуют, например, в армянский или литовский);

в) полная реализация поиска по всем размеченным параметрам, в том числе по словообразованию и неточностям перевода;

г) подключение средств визуализации, например, построения графиков, аналогично существующим в основном и газетном корпусах НКРЯ.

### Литература

Михайлов М. Н., Хярме Ю. Параллельные корпуса художественных текстов в Тамперском университете // Русский язык за рубежом. Финская русистика (специальный выпуск). М., 2015. С. 16–19.

Морозова В. А. Построение русско-китайского параллельного корпуса текстов: реализация нейросетевой модели для реализации автоматического словоделения иероглифических текстов. Курсовая работа. ВШЭ, 2018.

Орехов Б. В. Проблемы морфологической разметки башкирских текстов // Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2014. Казань: Изд-во «Фэн» Академии наук РТ, 2014. С. 135–140

---

оригинал мог повлиять на выбор времени весьма ограниченно, поскольку перевод В. Чепайтиса первоначально делался с польского текста И. Тувим и лишь потом правился по оригиналу. Как видим, у этого в принципе полезного «нивелирующего» свойства есть и обратная сторона (демонстрирующая, между прочим, и ограниченность возможностей корпусного исследования для граммов, активно конкурирующих с синонимичными).

Ляшевская О. Н., Плунгян В. А., Сичинава Д. В. О морфологическом стандарте Национального корпуса русского языка // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М., 2005. С. 111–135.

Сичинава Д. В. Параллельные тексты в составе Национального корпуса русского языка: новые направления развития и результаты // Труды Института русского языка РАН, 2015. № 6. С. 194–235.

Сичинава Д. В. Славянский плюсквамперфект: пространства возможностей // Вопросы языкознания. 2019, № 1. С. 30–57.

Шведова М., фон Вальденфельс Р., Яригин С., Крук М., Рисин А., Возняк М. Генеральный регионально анотированный корпус украинської мови (ГРАК). Київ, Осло, Сна, 2017–2019. [Электронный ресурс]. URL: uacorpus.org.

Якубов М. Н. Построение китайско-русского корпуса параллельных текстов: проблемы выравнивания, словоделения и лексико-семантической разметки. Выпускная квалификационная работа. ВШЭ, 2017.

Dahl, Ö., Koptjevskaja-Tamm, M. (eds.) Circum-Baltic languages. Typology and contact. Vol. 1-2, Amsterdam—Philadelphia: Benjamins, 2001.

Östling, R. Stagger: an Open-Source Part of Speech Tagger for Swedish. // Northern European Journal of Language Technology, 2013, Vol. 3, Article 1, pp 1–18.

Perkova, N., Sitchinava, D. On the Development of a Latvian-Russian Parallel Corpus // Skadiņa, I., Rozis, R. (eds.). Human Language Technologies — The Baltic Perspective: Proceedings of the Seventh International Conference Baltic HLT 2016, pp. 130–135. IOS Press, Amsterdam, 2016.

Rimkutė, E., Daudaravičius, V., Utkā, A. Morphological Annotation of the Lithuanian Corpus. 45th Annual Meeting of the Association for Computational Linguistics. // Workshop Balto-Slavonic Natural Language Processing, 2007, pp. 94–99.

Sitchinava D., Perkova, N. Bilingual Parallel Corpora Featuring the Circum-Baltic Languages within the Russian National Corpus. Digital Humanities in the Nordic Countries Proceedings of the Digital Humanities in the Nordic Countries 4th Conference Copenhagen, Denmark, March 5-8, 2019, pp. 495–502.

von Waldenfels R. Compiling a parallel corpus of Slavic languages. Text strategies, tools and the question of lemmatization in alignment // Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV) 9. München, 2006, S. 123–138.

**Dmitri V. Sitchinava**

*Vinogradov Russian Language Institute of the Russian Academy of Sciences*

*National Research University Higher School of Economics*

*(Russia, Moscow)*

*mitrius@gmail.com*

## **ON PARALLEL CORPORA WITHIN THE RUSSIAN NATIONAL CORPUS : NEW LANGUAGES AND NEW CHALLENGES**

The paper discusses the main trends in the update and conceptual development of the parallel corpora within the RNC since 2015. The New languages section deals with new language pairs that emerged during this period, their architecture and tagging.

Compared with the list of languages that form bilingual parallel pairs with Russian available in 2015, the following new languages have appeared in the RNC: Bashkir, Buryat, Chinese, Czech, Finnish, Lithuanian, Swedish. Creating the parallel subcorpora comes as combined efforts of autonomous Russian and foreign teams coordinated by the team of developers of the RNC in Moscow. Virtually all the new languages offer specific challenges for the Corpus developers with regard to their annotation. The size of some of the language pairs already available in 2015 has also significantly increased in four years.

The New challenges section goes further to explore some general trends of parallel corpora across different language pairs such as regional/national language varieties, representativeness with regard to text genres, new annotation and search types etc. At the present stage, both due to an increase in the typological diversity of the languages and scripts involved, and through the use of more complex morphological analyzers, the set of additional markup parameters has been significantly expanded. The purpose of incorporating polycentric language samples into the corpus has become one of the most important. The genre representativeness of corpora, which was only being planned in 2015, is one of the main goals in 2019, and this goal is taken into account from the very beginning of the creation of new language pairs.

A case study is dedicated to studying pluperfect in 24 languages of Europe in an expanded multilingual corpus. The analysis of a multilingual text based on the material of an extended collection allows us to construct a network of distances between the data for the Pluperfect category in 24 European languages/lects.

*Key words:* parallel corpora; bilingual corpora; multilingual corpora; annotation; representativeness; pluperfect

## **References**

Dahl, Ö., Koptjevskaja-Tamm, M. (eds.) *Circum-Baltic languages. Typology and contact*. Vol. 1-2, Amsterdam—Philadelphia: Benjamins, 2001.

Lyashevskaya O.N., Plungian V.A., Sitchinava D.V. [On morphological standard of the Russian National Corpus] // *Natsional'nyi korpus russkogo yazyka: 2003-2005. Rezul'taty i perspektivy* [The Russian National Corpus: 2003-2005. Results and prospects]. Moscow, 2005, pp. 111–135. (In Russ.)

Mikhailov M.N., Härme J. [Parallel Corpora of Fiction at the University of Tampere] // *Russkii yazyk za rubezhom. Finskaya rusistika* (spetsial'nyi vypusk) [Russian language abroad. Finnish Russian studies (special issue)]. Moscow, 2015, pp. 16–19. (In Russ.)

Morozova V.A. *Postroenie russko-kitaiskogo parallel'nogo korpusa tekстов: realizatsiya neirosetevoi modeli dlya realizatsii avtomaticheskogo slovodeleniya ieroglificheskikh tekстов*. Kursovaya rabota. [Building Russian-Chinese parallel corpus : applying of neural network model to word segmentation of hieroglyphic texts] NRU HSE, 2018. (In Russ.)

Östling, R. Stagger: an Open-Source Part of Speech Tagger for Swedish // *Northern European Journal of Language Technology*, 2013, Vol. 3, Article 1, pp 1–18.

Orekhov B. V. [Problems of morphological marking of Bashkir texts]. *Trudy Kazanskoi shkoly po komp'yuternoi i kognitivnoi lingvistike TEL-2014* [Proceedings of the Kazan summer school in computer and cognitive linguistics TEL-2014]. Kazan', "Fen" Publ., 2014, pp. 135–140. (In Russ.)

Perkova N., Sitchinava D. On the Development of a Latvian-Russian Parallel Corpus // Skadiņa, I., Rozis, R. (eds.). *Human Language Technologies — The Baltic Perspective: Proceedings of the Seventh International Conference Baltic HLT 2016*, pp. 130–135. IOS Press, Amsterdam, 2016.

Rimkutė E., Daudaravičius V., Utkā A. Morphological Annotation of the Lithuanian Corpus. 45th Annual Meeting of the Association for Computational Linguistics. *Workshop Balto-Slavonic Natural Language Processing*, 2007, pp. 94–99.

Sitchinava D., Perkova N. Bilingual Parallel Corpora Featuring the Circum-Baltic Languages within the Russian National Corpus. *Digital Humanities in the Nordic Countries*. Proceedings of the Digital Humanities in the Nordic Countries 4th Conference Copenhagen, Denmark, March 5-8, 2019, pp. 495–502.

Sichinava D. V. [Parallel texts within the Russian National Corpus: new trends of developments and new results]. *Trudy Instituta russkogo yazyka RAN* [Proceedings of the V. V. Vinogradov Russian Language Institute], 2015, no. 6, pp. 194–235. (In Russ.)

Sichinava D. V. [Slavic Pluperfect: foci of variation] *Voprosy jazykoznanija* [Topics in the study of language]. 2019, no. 1, pp. 30–57. (In Russ.)

Shvedova M., von Waldenfels R., Jaryhin S., Kruk M., Rysin A., Woźniak M. *Henerval'nii rehional'no anotovanyi korpus ukrajins'koji movy (HRAK)* [General Regionally Annotated Corpus of Ukrainian (GRAC)]. Kyiv, Oslo, Jena, 2017–2019. Available at: [uacorp.us.org](http://uacorp.us.org).

von Waldenfels R. Compiling a parallel corpus of Slavic languages. Text strategies, tools and the question of lemmatization in alignment. *Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV)* 9. München, 2006, pp. 123–138.

Yakubov M. N. *Postroenie kitaisko-russkogo korpusa parallel'nykh tekстов: problema vyravnivaniya, slovodeleniya i leksiko-semanticheskoi razmetki*. Vypusknaya kvalifikatsionnaya rabota. [Building Chinese-Russian parallel corpus : alignment, word segmentation, lexical and semantic annotation]. NRU HSE, 2017. (In Russ.)

<sup>1</sup>Б.В. Орехов, <sup>2</sup>С.О. Савчук

<sup>1</sup>НИУ «Высшая школа экономики»,

<sup>2</sup>Институт русского языка им. В.В. Виноградова РАН

Россия, Москва

<sup>1</sup>nevmenandr@gmail.com, <sup>2</sup>savsvetlana@mail.ru

## АКЦЕНТОЛОГИЧЕСКИЙ КОРПУС КАК ИНСТРУМЕНТ ДЛЯ ИССЛЕДОВАНИЯ РУССКОГО УДАРЕНИЯ\*

В настоящей статье рассмотрено несколько вопросов, связанных с разработкой и использованием акцентологического корпуса в качестве инструмента для исследования ударения: состав и структура корпуса, текущее состояние, перспективы развития, пополнение новым материалом. Особое внимание уделено подкорпусу наивной поэзии в составе акцентологического корпуса как источнику акцентологических данных. Возможности этого ресурса, его эффективное использование проверены на нескольких участках акцентологической системы.

Было проведено корпусное исследование акцентных вариантов форм единственного и множественного числа кратких прилагательных, а также падежных форм имен существительных. В ходе изучения падежных форм существительных было обследовано несколько зон активной конкуренции акцентных вариантов: существительных женского рода на *-а* (*стена, доска*), существительных женского рода с основой на мягкий согласный (*кисть, челюсть*) и существительных мужского рода с твердой основой (*ветер, шарф*).

Результаты корпусного исследования подтвердили предположение о том, что материал акцентологических корпусов может быть использован как достоверный источник получения акцентологических данных. Увеличение объема корпуса делает эти данные статистически достоверными, а также расширяет круг исследуемых форм и способствует обнаружению новых точек вариативности.

*Ключевые слова.* Акцентологический корпус, Национальный корпус русского языка, «наивная поэзия», русское словесное ударение, акцентные варианты.

Русское ударение, его сложная система и состояние неустойчивого равновесия привлекают внимание как отечественных, так и зарубежных лингвистов

---

\* Работа выполнена при поддержке РФФИ (грант № 17-29-09154).

к изучению его истории и тех процессов, которые наблюдаются в области словоизменения и словообразования. Обычно изучение основывается на использовании источников материала, среди которых наиболее известными являются:

— данные словарей и справочников, содержащие сведения об ударении в разные исторические периоды и на разных территориях [Грамм.; Еськова 1967; Зализняк 1985].

— акцентуированные тексты — памятники письменности [Зализняк 1985], а также учебные тексты;

— поэтические тексты [Воронцова 1979];

— транскрипты звучащей речи с проставленными ударениями;

— экспериментальный материал разного рода, полученный в ходе направленных интервью, анкетирования, опросов и др. [Куракина 2011; Marklund Sharapova 2000; Ukiah 2002; Каленчук и др. 2017];

— корпуса текстов [Савчук 2010; Гришина 2012; Пожарицкая и др. 2017].

Можно предположить, что объединение разных источников в один корпус текстов многократно повышает их информативность и производительность. Идея создания корпуса для изучения русского ударения в составе НКРЯ, принадлежащая Е. А. Гришиной, была реализована в 2008 году, когда Акцентологический корпус был открыт для всеобщего доступа. За десять лет существования корпус многократно увеличился в объеме и продолжает пополняться. В настоящей статье будет рассмотрено несколько вопросов, связанных с разработкой и использованием акцентологического корпуса в качестве инструмента для исследования ударения: текущее состояние, состав и структура корпуса; перспективы развития, пополнение новым материалом. Особое внимание будет уделено подкорпусу наивной поэзии в составе акцентологического корпуса как источнику акцентологических данных. Возможности этого ресурса, его эффективное использование будут проверены на нескольких участках акцентологической системы.

## **1. Акцентологический корпус: состав и структура, перспективы роста**

Акцентологический корпус первоначально разрабатывался как инструмент для изучения истории русского ударения [Гришина 2009]. Его основу составили поэтические тексты с размеченными долями (иктами), которые в подавляющем большинстве случаев позволяли вычислить место ударения в слове<sup>1</sup>.

---

<sup>1</sup> Правила вычисления места ударения по стихотворной разметке подробно описаны в [Гришина 2009, 153-155], там же рассмотрены случаи, когда однозначно определить место ударения не помогает ни метрическая организованность, ни рифма [там же, 155-157]. Как показывает практика применения акцентологического корпуса, таких контекстов в целом немного, не более 10%. Стиховеды утверждают, что «в размерах русской силлаботоники нет таких слабых позиций, которые вообще не могли бы быть заняты ударными слогами, и нет таких сильных позиций, которые не могли бы быть заняты безударными слогами» [Илюшин 2004, 71], но разработанная Е. А. Гришиной методика акцентологической работы позволяет получать нужные результаты.



Поэтическая часть акцентологического корпуса позволяет изучать изменения, которые произошли в системе ударения на протяжении XVIII–XXI вв. Однако при использовании корпуса следует иметь в виду, что информация об этих изменениях извлекается из текстов, относящихся к особому роду словесности — поэзии, которая обладает специфическими свойствами как в отношении лексического состава, так и в отношении к языковым нормам, в частности, акцентологическим. В научной литературе неоднократно отмечалось, что пользуясь поэтическим материалом, необходимо учитывать его «особость» — «большую вариантность, возможность использования ударения в стилистических целях, намеренную в некоторых случаях его архаизацию и др.» [Воронцова 1979, 6]. Поэтому дальнейшее развитие акцентологического корпуса требовало расширения его текстового состава за счет включения в него устных текстов, функционирующих в различных речевых сферах, что и стало задачей создания прозаической части подкорпуса [Savchuk 2009, Гришина и др. 2009].

Прозаический подкорпус содержит транскрипты устной речи, в которых представлены ударения в соответствии с реальным произношением и которые представляют устную русскую речь в ее функциональных разновидностях. В подкорпусе собраны образцы спонтанной бытовой речи, публичной устной речи разной степени спонтанности (устной научной речи, устной политической речи, радио- и телепублицистики, современной деловой речи, религиозных наставлений и др.). Значительную долю материала составляют транскрипты кинофильмов, театральных постановок, образцы художественного и авторского чтения. Прозаическая часть корпуса предназначена для изучения современного состояния акцентологических норм и их варьирования в зависимости от сфер устной речи.

В настоящее время объем текстов акцентологического корпуса, доступных на сайте [ruscorgpora.ru](http://ruscorgpora.ru), составляет более 17 млн словоупотреблений (без учета пилотного корпуса наивной поэзии, о котором речь пойдет дальше). Распределение текстов по двум зонам представлено в таблице 1.

Примерное распределение по периодам представлено в таблице 2.

Ценность акцентологического корпуса состоит в предоставляемой исследователю возможности получения ответов на как можно большее число вопросов.

Таблица 1

**Распределение текстов  
акцентологического корпуса  
по функциональным разновидностям**

Зона		Объем, млн	%
Поэзия		11	63 %
Проза		6,4	37 %
В т. ч.	Речь кино и театра	3,5	20 %
	Публичная речь	2,4	14 %
	Непубличная речь	0,4	2 %
	Художественное и авторское чтение	0,1	0,5 %

Таблица 2

**Распределение текстов  
акцентологического корпуса по периодам**

Период	Поэзия		Проза	
	Объем, млн	%	Объем, млн	%
1700–1800	0,92	8,1 %	—	—
1801–1910	4,47	39,4 %	0,02	0,3 %
1911–1950	4,43	39,2 %	0,4	6,2 %
1951–1980	1,26	11,1 %	2,14	33,1 %
1981–2000	0,24	2,1 %	1	15,4 %
2001–2017	0,02	0,1 %	2,9	45 %



Причем ответов не единичных, а статистически значимых, позволяющих строить гипотезы о тенденциях в развитии акцентологических норм. Поэтому ответ на главный вопрос о том, в каком направлении следует развивать акцентологический корпус, напрашивается сам собой: нужно увеличивать его объем, включая новые тексты. Но тогда возникает другой вопрос: какими именно текстами следует пополнять корпус? Дело в том, что специфика корпуса предъявляет определенные требования к отбору текстов. Во-первых, акцентологические изменения затрагивают далеко не каждое слово русского языка, большинство слов (по оценкам А. А. Зализняка, около 80 % лексического состава) не меняют свой просодический профиль на протяжении столетий, сохраняя стабильное поведение при словоизменении и словообразовании: *август, дорога, жизнь, книга, кошка, нога, подруга, поле, привет, строитель, хвост, яблоко* и др. Поэтому было бы желательно, если бы корпус расширялся за счет ускоренного увеличения доли интересного в акцентологическом отношении материала, таких нестабильных участков, как глаголы с подвижным ударением, существительные 2-го склонения, заимствованные слова и др.

Во-вторых, технология создания акцентологического корпуса, особенно подготовка прозаических текстов, отличается большой трудоемкостью. Поэтому увеличение темпов и объемов пополнения корпуса ставит задачу снижения трудозатрат на подготовку текстов, причем не в ущерб качеству подготовки, прежде всего точности акцентологической разметки. Значительно сократить трудозатраты помогают специальные программы, позволяющие автоматизировать некоторые процессы подготовки, в частности, используемое при обработке транскриптов рабочее место лингвиста *Scripter* (автор Л. Д. Алексеевский). Еще больший эффект имело бы применение надежных программ автоматической акцентуации текстов, которые способны за короткое время разметить большие текстовые массивы. В последнее время на основе разных подходов активно ведутся разработки таких программ-акцентуаторов (см. [Зеленков и др. 2014]; [Ponomareva и др. 2017]; [Reynolds, Tyers 2015]; [Хомицевич и др. 2008]). Программа Ю. Зеленкова предназначена для разметки стихотворного текста, ее алгоритм позволяет обрабатывать русские словоформы с точностью 0.95–0.97, что вполне приемлемо для использования автоматически размеченного стихотворного материала в исследованиях ударения, поскольку в случае ошибочного предсказания программой места ударения есть возможность определить его исходя из метрической разметки. Подробнее о принципах работы программы см. [Гришина и др. 2015, 260–263].

## 2. Подкорпус «наивной поэзии» как источник акцентологических данных

Исходя из этих соображений было принято решение использовать в качестве источника для пополнения акцентологического корпуса коллекцию текстов «наивной поэзии», размещенной на сайте *stihi.ru*. Пилотный корпус «наивной поэзии» объемом около 13 млн словоупотреблений был сформирован и прошел апробацию, результаты которой изложены в работе [Гришина и др. 2015, 257–268].

Понятие наивной поэзии поначалу появилось в среде фольклористов [«Наивная литература» 2001]. В самом деле, культурный статус этих текстов сильно отличается от того, который закреплён за поэтической классикой. Если последняя имеет репутацию золотого фонда языка, непререкаемых образцов, тематически разнообразных и формально совершенных текстов, то тексты, которые обычно публикуются в Интернете непрофессиональными писателями, не выдерживают эстетической критики, построены формульно, однообразно и ограничены небольшим тематическим набором. Такое описание в гораздо большей степени подходит под шаблонно выстроенный фольклорный текст (с определенной оговоркой относительно эстетических достоинств фольклорных произведений, которые в любом случае второстепенны для системы устного народного творчества).

Типичность этих текстов позволяет оценить их количественно. В рамках концепции *distant reading* [Moretti 2013] перед нами открываются перспективы компьютерного анализа больших корпусов художественных текстов и литературоведческих обобщений. Такие оценки были сделаны и для наивной поэзии, размещенной авторами на сайте *stihi.ru* [Бонч-Осмоловская, Орехов 2014, 20–36]. Стихи «наивных» авторов в основном посвящены внутренним переживаниям, моментальным эмоциям и психологическому пейзажу.

Объемы текстов на ресурсе *stihi.ru* поистине циклопические: каждые 7 секунд на нем появляется новое стихотворение, поэтому размер корпуса *stihi.ru* превышает не только 100 млн словоупотреблений, которыми мы собираемся расширить акцентологический корпус, но и — в несколько раз — объем основного корпуса НКРЯ. Большинство авторов при этом формально стараются повторить известные им поэтические образцы и поэтому следуют силлабо-тоническому шаблону, что, в свою очередь, облегчает работу с акцентологическим материалом.

Тексты собраны с сайта *stihi.ru* в автоматическом режиме. С помощью программы Ю. Зеленкова для каждой строки был установлен метр и расставлены икты. В дальнейшем тексты, в которых для более чем 15 % строк не удалось определить силлабо-тонический метр, отсеивались. Согласно правилам разметки акцентологического корпуса, в текстах была выделена зона рифмовки, которая дает дополнительную информацию о позиции ударения. Для достижения порога в 100 млн словоупотреблений потребовалось около 1 млн стихотворений «наивных» поэтов.

В настоящее время подготовлен и акцентологически размечен подкорпус «наивной» поэзии в объеме 100 млн словоупотреблений. Предпринимая предварительный анализ этого корпуса, мы ставили задачу получить ответы на следующие вопросы. Достаточен ли объем корпуса для принципиального повышения статистических показателей в акцентологически «проблемных» точках? Достаточно ли лексическое разнообразие текстов для того, чтобы охватить наибольшее количество «проблемных» точек, а не только увеличить его объем? Отражает ли корпус наивной поэзии живое русское произношение и позволяет ли проследить динамику изменений акцентологических предпочтений? Коррелируют ли данные, полученные в корпусных исследованиях, с результатами, полученными экспериментальным путем?

### 3. Анализ корпуса наивной поэзии

С целью проверки возможностей 100-миллионного корпуса было проведено исследование акцентных вариантов в области словоизменения существительных и прилагательных. Для исследования были отобраны лексемы существительных и прилагательных, которые относятся к проблемным точкам акцентологической системы, то есть при словоизменении имеют вариативное ударение во всех или в отдельных формах. Для каждой лексики был установлен состав вариантов и их частотные показатели в каждом из трех описанных ниже корпусов.

Первый корпус составили тексты основного акцентологического корпуса за период до 2000 года. Его объем — 14 468 529 словоупотреблений. В дальнейшем будем обозначать его **Акц\_2000**.

Второй корпус включает тексты после 2000 года, в основном это записи устной речи и пилотный корпус наивной поэзии. Объем корпуса — 17 262 123 словоупотреблений, условное наименование — **Акц\_2000+**.

Третий корпус составлен из текстов сайта *stihi.ru*, его объем около 100 млн словоупотреблений, условное наименование **Стихи.ру-100**.

По результатам анализа составлены базы данных, фрагменты которых показаны в таблицах 3 и 4<sup>2</sup>.

Таблица 3

Варианты форм кратких прилагательных

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
верный	ве'рный'	54	59	21	135	182	843
глупый	глу'пы'	41	5	37	98	499	302
строгий	стро'ги'	127	5	51	14	516	81
честны	че'стны'	12	8	9	32	103	250
быстрый	бы'стры'	81	11	18	10	164	125
жалкий	жа'лки'	61	1	19	1	147	26
властный	вла'стны'	83	8	78	2	536	4
вкусный	вку'сны'	14	4	8	2	90	59
вредный	вре'дны'	25	10	2	7	20	71
жадный	жа'дны'	23	2	3	0	51	13
ловкий	ло'вки'	8	11	2	4	76	36
тусклый	ту'склы'	21	4	5	0	88	11
храбрый	хра'бры'	22	5	2	5	16	24
звучный	звуч'ны'	21	4	5	1	15	2
душный	ду'шны'	6	0	2	0	14	0

<sup>2</sup> Во второй графе таблиц представлены акцентные варианты словоформ, в следующих графах вариант с ударением на основе обозначен как вариант 1, с ударением на окончании — как вариант 2. В некоторых трех- и четырехсложных формах возможны три варианта ударения (*бо'ро'зда'м*, *по'ло'са'х* и пр.), при этом обе формы с ударением на основе приводятся в графе варианта 1 через вертикальную черту.

Продолжение табл. 3

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
пресный	пре'сны'	1	0	4	0	47	11
спелый	спе'лы'	8	1	0	0	23	1
знатный	зна'тны'	7	0	1	0	12	0
смуглый	сму'глы'	8	7	0	0	0	4

Таблица 4

## Акцентологические варианты падежных форм существительных

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
Стена	сте'ну'	182	27	113	5	3074	153
	сте'на'м	69	159	104	29	1142	245
	сте'на'ми	32	124	57	35	505	288
	сте'на'х	97	339	225	139	2048	1000
Волна	во'лна'м	52	229	83	184	687	1474
	во'лна'ми	52	410	107	172	1065	1322
	во'лна'х	76	385	107	210	1020	1468
Доска	до'ску'	59	14	66	10	417	122
	до'со'к	30	50	23	13	198	122
	до'ска'м	26	9	4	2	62	10
	до'ска'ми	10	4	6	3	72	23
	до'ска'х	26	16	9	7	50	59
Щека	щё'ку'	67	27	90	25	415	184
	щё'ка'м	6	86	14	469	35	2359
	щё'ка'ми	0	23	0	9	3	110
	щё'ка'х	5	207	7	286	23	1710
Полоса	по'лосу'	26	29	18	34	32	294
	по'ло'с	4	28	0	0	102	560
	по'лоса'м	1	1	1	3	3 2	32
	по'лоса'ми	3	41	2	12	1	0
	по'лоса'х	3	5	2	9	3 2	0
Борозда	бо'розду'	3	16	12	2	102	0
	бо'ро'зд	7	11	1	11	3	38
	бо'ро'зда'м	1	7	1	0	0 1	10
	бо'розда'ми	0	3	2	0	0	19
	бо'ро'зда'х	1 3	6	1	0	1 2	12
Сковорода	ско'вороду'	2	12	0	12	3	57
	ско'воро'д	1	4	0	0	0	4
	ско'ворода'м	0	0	0	0	0	0
	ско'ворода'ми	0	0	0	0	0	0
	ско'ворода'х	0	0	0	0	0	5
Изба	и'збу'	75	81	20	20	168	287
	и'зба'м	23	5	3	1	32	4

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
Изба	и'зба'ми	7	0	4	1	28	1
	и'зба'х	31	4	9	0	87	6
Челюсть	че'люсте'й	8	5	2	3	49	6
	че'люстя'м	0	1	0	0	9	0
	че'люстя'ми	3	7	0	8		
	че'люстя'х	6	3	0	1	12	1
Кисть	ки'сте'й	2	32	4	16	21	174
	ки'стя'м	0	2	1	1	4	15
	ки'стя'ми	3	42	4	10	40	206
	ки'стя'х	2	11	4	3	30	39
Ветер	ве'тры/ветра'	513	109	310	480	3793	5324
	ве'тро'в	114	194	57	272	581	3641
	ве'тра'м	31	78	24	150	231	1797
	ве'тра'ми	29	115	33	163	468	2125
	ве'тра'х	4	36	2	55	36	828
Шарф	ша'рфа'	13	0	7	1	61	17
	ша'рфу'	1	0	2	0	6	3
	ша'рфо'м	17	1	21	16	192	89
	ша'рфе'	6	0	9	6	60	30
	ша'рфы'	14	3	8	7	68	60
	ша'рфо'в	2	0	1	0	10	19
	ша'рфа'м	0	0	0	0	2	1
	ша'рфа'ми	2	1	0	3	15	18
	ша'рфа'х	3	0	0	1	12	11

Первый вопрос, поставленный в исследовании, касался количественных параметров корпуса и состоял выяснении того, насколько 100-миллионный корпус наивной поэзии обеспечивает пополнение акцентологического корпуса в «проблемных» точках. Подобный анализ проводился ранее на пилотном корпусе [Гришина и др. 2015, 263–268], в настоящей работе мы прежде всего сосредоточились на лексемах, которые не получили заметного прироста в пилотном корпусе.

Как видно из таблицы 3, пополнение акцентологического корпуса текстами наивной поэзии обеспечивает существенный прирост материала, значимого для изучения акцентной вариативности. Так, например, в современном корпусе для вариантов *ло'вки* — *ловки'*, *че'стны* — *честны'*, *вку'сны* — *вкусны'*, *вре'дны* — *вредны'* не было достоверных данных, на основании которых можно было бы сделать вывод о соотношении вариантов, а 100-миллионный корпус предоставил десятки примеров, позволяющих определять это соотношение и строить гипотезы о конкуренции вариантов и о тенденциях развития вариативности.

Однако в отдельных случаях корпус Стихи.ру-100 не дал ожидаемого прироста. Подобный отрицательный результат для отдельных лексем был получен ранее и на материале пилотного корпуса: в частности, для таких форм, как *знатны*, *смуглы*,

спелы, тучны, серъгам не зафиксировано ни одного вхождения в исследуемом корпусе, формы *храбры, жадны, дуины, звучны, тусклы* получили 1–2 новых вхождения [Гришина и др. 2015, 265–268]. Но тогда высказывалось предположение, что благодаря 100-млн корпусу десятикратное увеличение объема поможет преодолеть бо2льшую часть проблем с полнотой представления акцентологически значимых форм. Как выяснилось, это не совсем так. То же можно сказать и о вариантах форм существительных, представленных в таблице 4. Для многосложных форм существительных *сковорода, борозда, ведомость* и под. обращение к 100-млн корпусу не дает существенного увеличения данных, и количественные показатели находятся на уровне основного акцентологического корпуса. Возможно, это объясняется «непозтичностью» подобных слов, которые нетипичны для стихотворного текста как в плане содержания, так и в плане выражения: в силу своей многосложности они с трудом вписываются в стихотворную строку. Видимо, исследовать редкие варианты придется другими методами, в частности, на основе экспериментов.

Содержательный анализ данных ставил своей целью выяснение возможностей корпуса для исследования вариативного ударения. Для этого в исследуемых корпусах был собран материал: в каждом подкорпусе для каждой лексемы определена акцентная парадигма (а.п.) и для каждой словоформы установлено количественное соотношение акцентологических вариантов. Результаты представлены в таблице 5<sup>3</sup>.

Таблица 5

**Количественное соотношение акцентологических вариантов  
падежных форм существительных**

Слово	а.п.	Варианты	Соотношение вариантов		
			Акц_2000	Акц_2000+	Стихи.ру-100
Стена	d'/f'	стену	87:13	96:4	95:5
		стенам	30:70	78:22	82:18
		стенами	20:80	62:38	64:36
		стенах	22:78	62:38	67:33
Волна	f//d	волну	2:98	0,8:99,2	—
		волнам	19:81	31:69	32:68
		волнами	11:89	38:62	45:55
		волнах	17:83	34:66	41:59
Доска	f//d	доску	80:20	87:13	77:23
		досок	38:62	64:36	62:38
		доскам	74:36	100:0	86:14
		досками	71:39	67:33	76:24
		досках	62:38	56:44	46:54

<sup>3</sup> Ударные гласные в вариантах словоформ в графе «Варианты» обозначены полужирным шрифтом. Соотношение вариантов складывается из отношения количества каждого из акцентных вариантов словоформы к общему количеству акцентуированных вхождений словоформы. В случае вариативного наосновного ударения соотношение с флексийными ударными формами рассчитано отдельно, и оба варианта приводятся в соответствующей графе через /. Соотношения вариантов, рассчитанные на основе единичных данных (см. табл. 4), обозначены !! Прочерк обозначает отсутствие данных.

Слово	а.п.	Варианты	Соотношение вариантов		
			Акц_2000	Акц_2000+.	Стихи.ру-100
Щека	f//f	щёку	71:29	78:22	69:31
		щёкам	7:93	3:97	1:99
		щёками	0:100	0:100	3:97
		щёках	2:98	2:98	1:99
Полоса	f//f	полосу	47:53	29:71	10:90
		полос	13:87	0	15:85
		полосам	50:50!	25:75!!	9:91/6:94
		полосами	7:93	8:92	100:0!!
		полосах	38:62!!	18:82!	100:0
Борозда	f//f	борозду	16:84	86:14	100:0
		борозд	39:61	8:92	7:93
		бороздам	12:88	100:0!	0:100/10:90
		бороздами	0:100	100:0!	0:100
		бороздах	14:43:43	100:0!	8:92/14:86
Сковорода	f//f	сковороду	14:86	0:100	5:95
		сковород	20:80	—	0:100
		сковородам	—	—	—
		сковородами	—	—	—
		сковородах	—	—	0:100
Изба	d//d'	избу	48:52	50:50	37:63
		избам	82:28	75:25	89:11
		избами	100:0	80:20	97:3
		избах	89:11	100:0	94:6
Ведомость	e//a	ведомостей	—	—	—
		ведомостям	—	—	—
		ведомостями	—	—	—
		ведомостях	—	67:33	—
Челюсть	a//e	челюстей	61:39	40:60	89:11
		челюстям	0:100	—	100:0
		челюстями	30:70	0:100	—
		челюстях	67:33	0:100!!	92:8
Кисть	e	кистей	6:94	20:80	11:89
		кистям	0:100!!	50:50!!	21:79
		кистями	7:93	29:71	16:84
		кистях	15:85	57:43!!	43:57
Ветер	e//a//c	ветры/ветра	82:18	39:61	42:58
		ветров	37:63	17:83	14:86
		ветрам	28:72	14:86	11:89
		ветрами	20:80	17:83	18:82
		ветрах	10:90	4:96	4:96
Шарф	a//b	шарфа	100:0	88:12	78:22
		шарфу	100:0	100:0	67:33
		шарфом	94:6	57:13	68:32
		шарфе	100:0	60:40	67:33
		шарфы	82:18	53:47	53:47
		шарфов	100:0	100:0	34:66
		шарфам	—	—	67:33
		шарфами	67:33!!	—	45:55
		шарфах	100:0	0:100	52:48

Задачей сравнительного анализа данных, представляющих соотношение вариантов в трех корпусах, было проверить, отражают ли корпусные данные реальное

соотношение вариантов в живой речи или в какой-то ее разновидности. Отражают ли полученные соотношения те процессы, которые происходят в акцентологической системе и описаны в лингвистической литературе?

Было обследовано несколько зон активной конкуренции акцентных вариантов падежных форм имен существительных: существительных женского рода на *-а*, существительных женского рода с основой на мягкий согласный и существительных мужского рода с твердой основой.

Напомним, что ударение в падежных формах имен существительных может быть описано с помощью 6 основных и 4 второстепенных схем (Грам. 2003, 31–32)<sup>4</sup>.

Таблица 6

## Основные схемы ударения в субстантивном склонении

Формы	а	б	с	д	е	ф
Ед.ч.	■○	□●	■○	□●	■○	□●
И. мн.	■○	□●	□●	■○	■○	■○
Р.Д.Т.П. мн.	■○	□●	□●	■○	□●	□●
	<i>карта, спор</i>	<i>стол, очко</i>	<i>сад, море</i>	<i>лист, вино</i>	<i>зуб, вещь</i>	<i>губа, конь</i>

Таблица 7

## Второстепенные схемы ударения в субстантивном склонении

Формы		б'		д'	ф'	ф''
И.Р.Д.П. ед.		□●		□●	□●	□●
В. ед.		□●		■○	■○	□●
Т. ед.		■○		□●	□●	■○
И. мн.		□●		■○	■○	■○
Р.Д.Т.П. мн.		□●		■○	□●	□●
		<i>вошь</i>		<i>спина</i>	<i>рука</i>	<i>грудь (ф''//е)</i>

Для имен существительных активность акцентных вариантов определяется направлением эволюции к формированию акцентной оппозиции единственного и множественного числа: ед. *па'рус* — мн. *паруса'*, ед. *высота'* — мн. *высо'ты*. Исторически процесс складывается из двух компонентов: 1) акцентная поляризация И. ед. и И. мн.; 2) выравнивание каждой из двух подпарадигм (Зализняк 1985, 373).

У существительных женского рода на *-а* акцентная вариантность наблюдается в формах В. ед. (*сте'ну', ре'ку'*) и в формах Д., Т., П. мн. (*сте'на'м, сте'на'ми, сте'на'х*)<sup>5</sup>.

Сравнение акцентных вариантов форм Д., Т., П. мн. по корпусу Акц\_2000 (напомним, что тексты представляют поэзию XVIII–XX вв., кино, театр и художественное

<sup>4</sup> В схемах ударная основа обозначена черным квадратом, безударная — белым, соответственно ударное окончание — черным кружком, безударное — белым.

<sup>5</sup> В Грам. это колебание в выборе схемы ударения при словоизменении обозначается как принадлежность лексемы к одной из двух акцентных парадигм *f//d*, *f''//d'*, *f''//f*, *d''//d*.



слово) и корпусу Акц\_2000+ (в основном это записи устной речи и наивная поэзия) показало, что соотношение вариантов изменилось более или менее существенно. Зависит ли степень изменений от лингвистических характеристик слов или экстралингвистических факторов, станет понятнее при рассмотрении конкретных лексем.

**Стена** (d'//f'). У существительного *стена* преобладание форм с ударением на основе (*сте'нам, сте'нами, сте'нах*, соответствующих схеме d') в корпусе Акц\_2000+ отражает рекомендацию Грам., в котором эта модель рассматривается как современная, а модель с ударением на окончаниях (*стена'м, стена'ми, стена'х*, соответствующая схеме f') — как устаревшая. Предпочтение этой устаревшей схемы в текстах корпуса Акц\_2000 объясняется высокой долей в нем поэзии XIX в. и высокой частотностью этого слова в поэтических текстах, отличающихся, как уже говорилось, бо'льшим консерватизмом, чем проза. Закрепление ударения на основе в формах Д., Т., П. мн. приводит к выравниванию внутри парадигмы множественного числа и к акцентному противопоставлению парадигм единственного и множественного числа. Корпус Стихи.ру-100, на порядок увеличивая количество вхождений, демонстрирует то же соотношение вариантов, что и современный корпус Акц\_2000+.

**Волна** (f//d). Соотношение в пользу вариантов Д., Т., П. мн. с флексийным ударением (*волна'м, волна'ми, волна'х*) сохранилось в корпусе современных текстов, однако доля вариантов с наосновным ударением (*во'лнам, во'лнами, во'лнах*) выросла в 2–3 раза, в чем можно усмотреть действие тенденции к выравниванию внутри парадигмы множественного числа. Интересно, что в 100-млн корпусе на фоне многократного увеличения абсолютной частоты вхождений словоформ Д., Т. и П. мн. увеличивается доля вариантов Т. и П. мн. с ударением на основе, но сохраняется неизменным (на уровне Акц\_2000+) соотношение вариантов Д. мн., — возможно, здесь проявляется сдерживающее влияние частотной формулы «по моря'м, по волна'м».

**Доска** (f//d). Все три корпуса показывают преобладание варианта с ударением на основе (*до'скам, до'сками, до'сках*) над вариантом с конечным ударением (*доска'м, доска'ми, доска'х*), причем в количественном отношении это соотношение также стабильно. Таким образом, вопреки нормативным рекомендациям, для лексемы *доска* отдающим предпочтение схеме ударения f, в современном узусе, представленном в акцентологическом корпусе, приоритетной является схема ударения d'. Кроме того, следует отметить, что в корпусных данных зафиксировано и зеркальное изменение соотношения вариантов *до'сок* — *досо'к* в формах Р. мн.: 38% : 62% в Акц\_2000 и 64% : 36% в Акц\_2000+, 62% и 38% в Стихи.ру-100.

**Щека** (f'//f). В формах Д., Т., П. мн. во всех корпусах преобладание варианта с флексийным ударением *щека'м, щека'ми, щека'х* близко к 100%, что свидетельствует о закреплении в этих формах традиционного ударения<sup>6</sup>.

<sup>6</sup> Слово *щека* входит в семантическую группу названий частей тела (наряду с *голова, борода, нога, рука, спина, коса*), устойчиво сохраняющую схему ударения d' или f' (см. об этом [Зализняк

**Полоса, борозда, сковорода** (f/f'). Конкуренция вариантов в формах Д., Т., П. мн., видимо, еще продолжается и по-разному протекает в разных падежных формах. С одной стороны, в Д. и П. мн. выросла доля форм с конечным ударением (*полоса'м, полоса'х*), с другой стороны, увеличилась доля форм Т. мн. с наосновным ударением в соотношении *по'лосами* и *полоса'ми* — от 7% до 14%. То же можно сказать о формах *бо'роздах, бо'роздами, бо'роздах*, которые представлены в корпусе Акц\_2000+. В корпусе Стихи.ру-100 преобладают варианты с конечным ударением: *полоса'м, борозда'м, борозда'ми, борозда'х, сковорода'х*. Впрочем, эти лексемы имеют невысокую частоту как в современном акцентологическом корпусе, так и в корпусе наивной поэзии, так что для выводов по всем интересующим формам требуется прибегнуть к другим акцентологическим источникам.

В целом можно сказать, что поведение акцентных вариантов в рассматриваемых лексемах постепенно ведет к выравниванию ударения внутри парадигмы множественного числа и способствует вовлечению данных лексем в группу существительных со сформированной акцентной оппозицией чисел.

Более сложную картину дает исследование соотношения форм Вин. ед. рассматриваемых существительных женского рода.

С одной стороны, у существительных **волна, грядя, сковорода, полоса, изба** предпочтительной оказывается форма с конечным ударением: *волну', грядю', сковороду', полосу', избу'*. При этом в первых трех словах конечное ударение в формах Вин. ед. доминирует; форма *полосу'* показывает в современном корпусе Акц\_2000+ значительное перераспределение в свою пользу по отношению к варианту *по'лосу*, а в корпусе Стихи.ру-100 оно доминирует. Сохраняющееся колебание вариантов *и'збу* — *избу'* (примерно 50% : 50%, хотя в 100-млн корпусе все-таки преобладает вариант *избу'*) может объясняться тем, что ударение *и'збу* поддерживается употреблением в поэтических текстах (они составляют 55 из 75 всех вхождений формы в корпусе Акц\_2000), в том числе хрестоматийных: *Прибежали в и'збу дети, В торопях зовут отца; Дурачина ты, прямой простофия! Выпросил, простофия, и'збу!* (А. С. Пушкин); *Коня на скаку остановит, В горящую и'збу войдет!* (Н. А. Некрасов); *И'збу освещает Огонек светца; Зимний вечер длится, Длится без конца...* (И. З. Суриков); *Метель мысли путает, Метель в и'збу ломится.* (М. И. Цветаева). В современной живой речи слово практически вышло из употребления и усваивается, включая и фонетический облик, по известным литературным цитатам. Достаточно сказать, что четверть вхождений формы *и'збу* в корпусе наивной поэзии встречается в перифразах цитаты из поэмы Н. А. Некрасова «в горящую избу войдет».

Преобладание формы с флексийным ударением в форме Вин. ед. отражает тенденцию к выравниванию подпарадигмы единственного числа и акцентному противопоставлению единственного и множественного числа (Зализняк 1985, 374).

Наряду с этим у существительных **стена, доска, борозда, щека, борода** в конкуренции акцентных вариантов формы Вин. ед. побеждает вариант с наосновным

1985, 29], то есть в сохранении стабильности соотношения вариантов можно предположить влияние семантического фактора.

ударением: *сте'ну, до'ску, щёку, бо'роду, бо'розду*. При этом в корпусе Акц\_2000+ доля победившего варианта выше.<sup>7</sup>

В преобладании форм Вин. ед. с наосновным ударением проявляется локальная тенденция к акцентной оппозиции Им. ед. и Вин. ед. (*трава' — тра'ву*), которую А. А. Зализняк относит к более поздним инновациям [Зализняк 1985, 374]. Рост доли этих форм в корпусах Акц\_2000+ и Стихи.ру-100, отражающих по преимуществу речь молодого поколения, согласуется с этим положением.

Вместе с тем, оба разнонаправленных процесса, обуславливающих тяготение вариантов Вин. ед. существительных ж. р. на *-а* к одной из акцентологических схем (с ударением на основе или на окончании), является проявлением глобальной тенденции к грамматикализации ударения, что выражается в закреплении его за определенными грамматическими формами.

Еще одной зоной вариативности, возникающей под воздействием грамматического фактора, являются существительные ж. р. с основой на мягкий согласный. Тенденция к акцентному противопоставлению форм ед. и мн. числа (*ветвь' — ве'тви, ве'тви — ветве'й*) проявляется в существовании вариантов с конечным ударением в формах Р., Д., Т., П. мн., которые конкурируют с вариантами, имеющими наосновное ударение, то есть схемы *е//а*: **ве'домость** — *ве'домосте'й, ве'домостя'м, ве'домостя'ми, ве'домостя'х*; **че'люсть** — *че'люсте'й, че'люстя'м, че'люстя'ми, че'люстя'х*; **кре'пость** — *кре'посте'й, кре'постя'м, кре'постя'ми, кре'постя'х*; **кисть** — *ки'сте'й, ки'стя'м, ки'стя'ми, ки'стя'х* и др.

Акцентная оппозиция чисел ограничена в этой группе морфонологическими правилами, по которым форма Им. мн. имеет ударение только на основе. Вследствие этого вступает в действие противоположная тенденция — акцентное выравнивание парадигмы множественного числа по форме Им. п., имеющей наосновное ударение [Зализняк 1985, 375]. Перечисленные лексемы являются малочастотными в обследованных корпусах, хотя для некоторых форм можно отметить наличие вариантов с флексийным ударением (*челюсте'й, челюстя'ми, челюстя'х*) как проявление тенденции к числовому противопоставлению, а для других — рост вариантов с наосновным ударением (*ки'стей, ки'стями, ки'стях*) как проявление противоположной тенденции.

В группе существительных мужского рода акцентная перестройка коснулась слов с наосновным ударением в единственном числе, в результате которой возникли варианты Им.-Вин. мн. (*по'рты — порты', ве'кторы — вектора'*) и варианты косвенных падежей мн. ч. (*ве'кторо'в, ве'ктора'м, ве'ктора'ми, ве'ктора'х*).

**Ветер** (*е//а//с*). В косвенных падежах мн. ч. во всех корпусах преобладают варианты с ударением на флексии, при этом в формах Р. мн и Д. мн. доля флексийных вариантов выросла в современных корпусах. В соотношении вариантов Им. мн.

<sup>7</sup> Исключение составляет форма *бо'розду'*, для которой разные корпуса фиксируют зеркальное соотношение вариантов: в Акц\_2000 — 16:84% в пользу *борозду'*, в Акц\_2000+ — 86:14% в пользу *бо'розду*. Корпус Стихи.ру-100 дает безальтернативный ответ в пользу формы *бо'розду* (102 вхождения). Возможно, это связано с низкой частотностью этого слова в корпусе и в речи вообще, что требует проверки на альтернативном материале.

*ве'тры* — *ветра'* произошла смена доминанты: если в корпусе Акц\_2000 преобладал вариант с наосновным ударением, то в корпусах Акц\_2000+ и Стихи.ру-100 его место занял конкурирующий вариант *ветра'*. Все это отражает тенденцию к акцентному противопоставлению форм единственного и множественного числа в парадигме склонения.

Модель Им.-Вин. мн. с ударным окончанием *-а* является очень продуктивной, в нее вовлекаются новые существительные, что особенно наглядно на примере заимствованных слов.

**Шарф** (а//b). В корпусе Акц\_2000 преобладают формы с неподвижным ударением на основе (что соответствует а. п. а, приписанной этой лексеме в Грам.), варианты отдельных форм с флексийным ударением единичны (*шарфы'*, *шарфо'м*, *шарфа'ми*). Однако по данным корпуса Акц\_2000+ и особенно Стихи.ру-100, доля вариантов с ударением на окончании значительно выросла, хотя и неравномерно в разных падежных формах. Особенно заметен рост вариантов с флексийным ударением в формах единственного числа и в формах *шарфы'* (до 47%), *шарфо'в* (до 66%), *шарфа'ми* (до 55%).

Большую роль в появлении вариантов с конечным ударением играет прагматический фактор: у освоенных слов растет доля вариантов с подвижным ударением, в то время как неосвоенная лексика (книжные, редкие, новые слова) сохраняет тривиальные схемы словоизменения [Зализняк 1985, 376].

#### 4. Заключение и выводы

Подводя итоги анализа акцентных вариантов в области именного словоизменения, можно сказать, что результаты корпусного исследования подтверждают выводы, сформулированные в лингвистической литературе на основе изучения словарных материалов, акцентуированных источников, поэтических текстов. Следовательно, материал акцентологических корпусов может быть использован как достоверный источник получения акцентологических данных. Увеличение объема корпуса делает эти данные статистически достоверными, а также расширяет круг исследуемых форм и способствует обнаружению новых точек вариативности.

Надежность результатов, полученных с помощью корпусов, подтверждается их количественным совпадением с результатами экспериментальных исследований акцентных вариантов. В частности, описанные выше соотношения вариантов *до'ску* — *доску'*, *по'лосу* — *полосу'*, *ско'вороду* — *сковороду'*, *ще'ку* — *щеку'*, *сте'нам* — *стена'м*, *ша'рфы* — *шарфы'* совпадают с данными, приведенными в диссертационном исследовании Е.Б. Куракиной, посвященном изучению места ударения в вариативных акцентных формах в речи молодых москвичей и построенном на экспериментальном материале [Куракина 2011].

При этом корпус наивной поэзии, как показало проведенное исследование, отражает тенденции в современном состоянии русской акцентной системы: соотношение словоизменительных вариантов в текстах этого корпуса в целом совпадает с результатами, полученными на материале современных текстов основного

акцентологического корпуса. Объем корпуса в 100 млн, видимо, следует признать достаточным для обеспечения статистической надежности результатов: количество большинства исследуемых вариантов выросло многократно, те же малочастотные лексемы, которые не обнаружили роста в 100-миллионном корпусе, вряд ли увеличат частоту при увеличении общего объема корпуса.

Однако идея корпуса наивной поэзии не всеми принимается безоговорочно. Возражения, в частности, вызывает сомнительный в содержательном плане материал. Далее, обращают внимание на специфичность материала: стиховая форма может накладывать ограничения на выбор лексики и акцентных вариантов. Можно предположить, что в корпусе будут слабо представлены или вообще отсутствовать «непоэтические» слова (в частности неологизмы, термины и под.) из интересующего акцентолога списка лексем и словоформ, что подтвердилось в процессе нашего исследования, когда ряд «непоэтических» слов не обнаружил никакого роста на фоне десятикратного увеличения объема корпуса. Вызывает замечания и анонимность материала: отсутствие социологических данных об авторах текстов, в отличие основного акцентологического корпуса или экспериментального материала (за исключением допущения о возрастной принадлежности авторов к молодому поколению).

Отдельно следует остановиться на ошибках в акцентологической разметке. На фоне экспериментальных данных или размеченного вручную основного акцентологического корпуса, в котором количество ошибок стремится к нулю, погрешности в акцентологической разметке корпуса наивной поэзии кажутся значительными. Ввиду ее автоматического характера ошибки неизбежны, и задача разработчиков корпуса — добиться их приемлемого уровня. Анализ пилотного корпуса показал, что в целом погрешности невелики и на большом объеме материала не будут иметь статистического эффекта. Однако в отдельных случаях доля ошибок в разметке могла достигать 30%. Пожалуй, это наиболее серьезный аргумент, который требует более подробного пояснения.

Причины ошибок могут быть различными. Программа автоматической разметки, как об этом говорится в статье [Гришина и др. 2015], расставляет ударения в несколько этапов. Сначала акцентуация каждого слова определяется в соответствии с наиболее вероятным вариантом, выводимым на основе модели машинного обучения, построенной на окончаниях русских словоформ. Затем, исходя из этой информации, фиксируется наиболее вероятный метр строки, который верифицируется данными о соседних строках. Чаще всего дальнейшие этапы анализа позволяют исправить ошибки первого шага, когда ударения расставляются в соответствии со статистической языковой моделью, а не опираются на реальное положение в строке. Так, ударение слова «вьюга» с точки зрения машинной модели могло бы приходиться на последний слог по аналогии с «юга́» или «города́». Но в разметке достаточное число случаев правильных вхождений формы «вью́га», подсказанных метрической схемой: *«С о́паской смóтрим дру́г на ду́га <так в тексте! — Б.О., С.С.>, / Но ме́жду нами снóва вью́га»; «Когда́ унима́ется вью́га / стано́вится снёг безмяте́жным...»* В то же время наивные поэты часто

не справляются с силлабо-тонической версификацией, пропускают или добавляют слоги, осуществляют наращення и усеечения, подобные цезурным, но не в позиции цезуры, фактически превращая силлабо-тонический стих в дольник. Нарушения регулярного стиха приводят к ошибкам в автоматической расстановке ударений. Так, следующее стихотворение наивного автора начинается как правильный трех-стопный анапест: *«Он дарил ей цветы и духи, / а она поцелуи дарила / сочинял для неё он стихи, / а его она боготворила / они летом любили бродить / в сосняке, и в лесу где берёзы, / с родника воду чистую пить / и вдыхать аромат белой розы»*. Однако в следующем стихе метр нарушается и в третьей стопе недостает одного безударного слога: *«а когда начинала петь / за окошком седая вьюга»*. Так как в данном случае это нарушение регулярное (то есть наблюдается в нескольких строках подряд), его можно отнести на счет сознательно примененного приема: переход на дольник совпадает с поворотным моментом в лирическом сюжете. Однако программа продолжает считать стих силлабо-тоническим и пытается расставить ударения в нем так, чтобы они соответствовали метрической схеме. В результате мы получаем неверное решение: «вьюга».

Случай другого типа, приводящий к ошибкам в разметке, — разница между письменным и произносимым вариантом. Формально из четырех строк *«Возможно, за какие-то заслуги, / И более желанный ждёт обмен, / Я буду, слушать завывание вьюги, / И не бояться больше перемен»* три строки (первая, вторая и четвертая) написаны пятистопным ямбом, а третья — снова нарушает версификацию, наращивая лишний безударный слог перед пятой стопой. Эта формальная ошибка снова приводит к неправильной разметке ударения в слове «вьюги». На деле же третья строка точно так же представляет собой правильный пятистопный ямб, но слово, произнесенное как «завыванье» написано иначе: «завывание». Возможно, это результат коррекции орфографии, встроенной в офисный текстовый процессор. Добавление лишней графемы, обозначающей гласный, на месте отсутствующего слога, негативно влияет на точность работы программы разметки. Такого рода «графическая» ошибка не единичный случай, вот аналогичное ему нарушение метра (здесь — четырехстопный ямб) в письменном представлении при очевидно правильном произносительном варианте: *«Ужё не быть в твоёй судьбё / Любимой девушки, подруги! / А колыбельная тебе — / Ночное завывание вьюги...»*.

Как представляется, однако, сильным аргументом в пользу включения материала наивной поэзии в состав акцентологического корпуса, который подтвердился в ходе предпринятого исследования, являются достоинства корпуса. К их числу относится, во-первых, большой объем, который, как уже говорилось, позволяет на порядок увеличить данные об акцентных вариантах. Во-вторых, репрезентативность материала: в текстах корпуса представлено достаточно большое количество словоформ, значимых для акцентологических исследований. В-третьих, относительная возрастная однородность авторов, которые в большинстве своем относятся к молодому поколению носителей языка, поэтому в текстах в основном представлена «младшая норма», что позволяет более уверенно делать выводы



о направлении развития акцентологических норм. Дальнейшие исследования с использованием корпуса наивной поэзии, проведенные на других участках колебания акцентологической системы, возможно, подтвердят или уточнят полученные результаты и характеристики материала.

## Литература

Бонч-Осмоловская А. А., Орехов Б. В. Корпусно-статистические подходы к наивной поэзии // Корпусный анализ русского стиха: Сборник научных статей. Вып. 2. / Отв ред. В. А. Плунгян, Л. Л. Шестакова. М. : Издательский центр «Азбуковник», 2014. С. 20–36.

Воронцова В. Л. Русское литературное ударение XVIII–XX вв. Формы словоизменения. М. : Наука, 1979.

Грам. — Зализняк А. А. Грамматический словарь русского языка. М. : Русские словари, 2003.

Гришина Е. А. Корпус «История русского ударения» // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб : Нестор-История, 2009. С. 150–174.

Гришина Е. А. Микроизменения в акцентологической системе русских прилагательных по материалам Национального корпуса русского языка // Вопросы культуры речи. Вып. 11. М., 2012

Гришина Е. А., Корчагин К. М., Плунгян В. А., Сичинава Д. В. Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб. : Нестор-История, 2009. С. 71–113.

Гришина Е. А., Зеленков Ю. Г., Орехов Б. В. Наивная поэзия в акцентологическом корпусе // Труды Института русского языка им. В. В. Виноградова. 2015, № 6. С. 257–271.

Гришина Е. А., Савчук С. О. Устный корпус в Национальном корпусе русского языка: состав и структура // Национальный корпус русского языка 2006–2008: Новые результаты и перспективы, СПб. : Нестор-История, 2009, 71–113.

Еськова Н. А. Картотека по современному русскому ударению // Лингвистические источники. Фонды Института русского языка. М.: Наука, 1967.

Зализняк А. А. От праславянской акцентуации к русской. М. : Наука, 1985.

Зеленков Ю. Г., Зобнин А. И., Маслов М. Ю., Титов В. А. Илья Сегалович и развитие идей компьютерной лингвистики в Яндексе // Труды международного семинара «Диалог14» по компьютерной лингвистике и ее приложениям (электронный документ). URL: <http://www.dialog-21.ru/digests/dialog2014/materials/pdf/ZelenkovYuG.pdf>. 2014.

Илюшин А. А. Русское стихосложение. М. : Высшая школа, 2004. 240 с.

Каленчук М. Л., Савинов Д. М., Скачедубова Е. С. Активные процессы в просодической системе русского языка: акцентуация прилагательных. // Русский язык в научном освещении. № 2 (34). 2017. С. 9–28.

Куракина Е. Б. Тенденции развития акцентологических норм в речи современных молодых москвичей. АКД Дисс. канд. филол. н. М., 2011.

«Наивная литература»: исследования и тексты / Сост. С. Ю. Неклюдов. М., 2001.

Пожарицкая С. К., Добрушина Е. Р. Орфоэпический взгляд на некоторые варианты явления русского литературного языка в эпоху корпусной лингвистики // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2017». М., 2017. URL: <http://www.dialog-21.ru/media/3940/pozharickajaskdobrushinaer.pdf>

Савчук С. О. Активные процессы в системе русского словоизменения: опыт корпусного исследования акцентологических норм // Труды II Международной конференции «Русский язык и литература в международном образовательном пространстве: современное состояние и перспективы». Т. 2. Гранада, 2010. С. 1549–1554.

Хомичевич О. Г., Рыбин С. В., Таланов А. О., Опарин И. В. Автоматическое определение места ударения в незнакомых словах в системе синтеза речи // Материалы XXXVI Международной филологической конференции. СПб., 2008.

Bonch-Osmolovskaya A., Orekhov B. Distant reading of naive poetry: corpora comparison as research methodology // Digital humanities Lausanne — Switzerland '14. URL: <http://dharchive.org/paper/DH2014/Paper-777.xml>

Grishina, E., Polyakov, A., Savchuk, S. Design and data collection for the Accentological corpus of Russian // LREC 2010 Proceedings, p. 618–622. URL: <http://www.lrec-conf.org/proceedings/lrec2010/>

Lagerberg, Robert. Variation and Frequency in Russian Word Stress. München : Verlag Otto Sagner, 2011. 184 p. URL: [https://www.researchgate.net/publication/45172817\\_Variation\\_and\\_frequency\\_in\\_Russian\\_word\\_stress](https://www.researchgate.net/publication/45172817_Variation_and_frequency_in_Russian_word_stress).

Marklund Sharapova, E. Implicit and Explicit Norm in Contemporary Russian Verbal Stress (Acta Universitatis Upsaliensis. Studia Slavica Upsaliensia 40). Uppsala: Uppsala University, 2000. URL: <https://uu.diva-portal.org/smash/get/diva2:160762/FULLTEXT01.pdf>

Moretti F. Distant reading. London, New-York: Verso, 2013.

Ponomareva, M., Milintsevich, K., Chernyak, E., Starostin, A. Automated Word Stress Detection in Russian. In Proceedings of the First Workshop on Subword and Character Level Models in NLP, pp. 31–35, Copenhagen, Denmark, September 7, 2017. Available at: <http://aclweb.org/anthology/W17-4104>.

Reynolds, R., Tyers, R. (2015) Automatic word stress annotation of russian unrestricted text. In Proceedings of the 20<sup>th</sup> Nordic Conference of Computational Linguistics (NODALIDA 2015). Linköping University Electronic Press, Sweden, Vilnius, Lithuania, pp. 173–180. URL: <http://www.aclweb.org/anthology/W15-1822>.

Savchuk, S. Spoken texts representation in the Russian National Corpus: Spoken and Accentologic sub-corpora // “NLP, Corpus Linguistics, Corpus Based Grammar Research”, 5th International conference SLOVKO Smolenice, Slovakia, 25–27 November 2009 Brno, Tribun EU, 2009, pp. 310–320.

Ukiah, N.: 2002, ‘The stress of Russian nouns in -a and -я of Zaliznjak’s pattern f (rybá type)’, Australian Slavonic and East European Studies, 16/1–2, 1–39.



<sup>1</sup>**B.V. Orekhov**, <sup>2</sup>**S.O. Savchuk**

<sup>1</sup>*National Research University Higher School of Economics*

<sup>2</sup>*Vinogradov Institute of Russian Language, Russian Academy of Sciences  
Russia, Moscow*

<sup>1</sup>*nevmenandr@gmail.com*, <sup>2</sup>*savsvetlana@mail.ru*

## ACCENTOLOGICAL CORPUS AS A TOOL FOR STUDYING RUSSIAN WORD STRESS

The article addresses several issues related to the development and use of the Accentological corpus as a tool for studying word stress. The current state of the corpus, its composition and structure, development prospects, replenishment with new material are described in this article. Particular attention is paid to the subcorpus of “naive” poetry within the Accentological corpus as a source of accentological data. The capabilities of this resource, its effective use are demonstrated in the study of variants of nominal inflection.

A corpus-based study of inflectional stress variation of singular and plural forms of short adjectives and case forms of nouns has been conducted. While studying case forms of nouns several areas of active competition of accent variants have been examined: feminine nouns with –a ending (*doska* ‘desk’, *stena* ‘wall’), feminine nouns with a soft stem (*kist* ‘brush’, *chel’ust* ‘jaw’) and masculine nouns with a hard stem (*veter* ‘wind’, *sharf* ‘scarf’).

The results of the corpus based study have shown that the material of accentological corpora can be used as a reliable source of accentological data. Increasing the volume of the corpus makes these data statistically reliable, as well as expanding the range of the studied forms and contributes to the detection of new points of variability.

*Key words:* Accentological corpus, the Russian National Corpus, “naive poetry”, Russian word stress, accent variants.

## References

Bonch-Osmolovskaya A. A., Orekhov B. V. [Corpus-statistical approaches to naive poetry] // *Korpusnyi analiz russkogo stikha: Sbornik nauchnykh statei* [Corpus analysis of Russian verse: Collection of articles]. Iss. 2. / V. A. Plungyan, L. L. Shestakova (eds.). Moscow, Azbukovnik Publ., 2014, pp. 20–36. (In Russ.)

Bonch-Osmolovskaya A., Orekhov B. Distant reading of naive poetry: corpora comparison as research methodology // *Digital humanities Lausanne—Switzerland’14*. Available at: <http://dharchive.org/paper/DH2014/Paper-777.xml>

Es’kova N. A. [Card index on modern Russian accent] // *Lingvisticheskie istochniki. Fondy Instituta russkogo yazyka* [Linguistic sources. The funds of the Institute of the Russian language]. Moscow, Nauka Publ., 1967. (In Russ.)

Grishina E. A. [Corpus “The History of Russian word stress”] // *Natsional’nyi korpus russkogo yazyka: 2006–2008. Novye rezul’taty i perspektivy* [The Russian National

Corpus: 2006–2008. New results and prospects]. Saint-Petersburg, Nestor-Istoriya Publ., 2009, pp. 150–174. (In Russ.)

Grishina E. A. [Micro-changes in the accentological system of Russian adjectives based on the materials of the Russian National Corpus] // *Voprosy kul'tury rechi* [Issues on culture of speech]. Iss. 11. Moscow, 2012. (In Russ.)

Grishina E. A., Korchagin K. M., Plungyan V. A., Sichinava D. V. [The poetic corpus within the RNC: General structure and future use] // *Natsional'nyi korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [The Russian National Corpus: 2006–2008. New results and prospects]. Saint-Petersburg, Nestor-Istoriya Publ., 2009, pp. 71–113.

Grishina E. A., Zelenkov Yu. G., Orekhov B. V. [Naive poetry within the Russian accentological corpus] // *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Russian Language Institute]. 2015, no. 6, pp. 257–271. (In Russ.)

Grishina E. A., Savchuk S. O. [Oral Speech corpus within the Russian National Corpus: composition and structure] // *Natsional'nyi korpus russkogo yazyka 2006–2008: Novye rezul'taty i perspektivy* [The Russian National Corpus: 2006–2008. New results and prospects]. Saint-Petersburg, Nestor-Istoriya Publ., 2009, pp. 71–113. (In Russ.)

Grishina E., Polyakov A., Savchuk S. Design and data collection for the Accentological corpus of Russian // *LREC 2010 Proceedings*, p. 618–622. Available at: <http://www.lrec-conf.org/proceedings/lrec2010/> (accessed 30.10.2018).

Ilyushin A. A. *Russkoe stikhoslozhenie* [Russian poetry]. Moscow, Vysshaya shkola Publ., 2004. 240 p. (In Russ.)

Kalenchuk M. L., Savinov D. M., Skachedubova E. S. [Active processes in the prosodic system of the Russian language: accentuation of adjectives] // *Russkii yazyk v nauchnom osveshchenii* [Russian Language and Linguistic Theory]. 2017, no. 2 (34), pp. 9–28. (In Russ.)

Kurakina E. B. *Tendentsii razvitiya aktsentologicheskikh norm v rechi sovremennykh molodykh moskvichei* [Trends in the development of accentological norms in speech of contemporary young moscovites]. Abstract of Ph. D. thesis. Moscow, 2011. (In Russ.)

Khomitsevich O. G., Rybin S. V., Talanov A. O., Oparin I. V. [Automatic determination of the place of stress in unknown words in the speech synthesis system] // *Materialy XXXVI Mezhdunarodnoi filologicheskoi konferentsii* [Proceedings of XXXVI<sup>th</sup> International philological conference]. St. Petersburg, 2008. (In Russ.)

Lagerberg R. *Variation and Frequency in Russian Word Stress*. München: Verlag Otto Sagner, 2011. 184 p. Available at: [https://www.researchgate.net/publication/45172817\\_Variation\\_and\\_frequency\\_in\\_Russian\\_word\\_stress](https://www.researchgate.net/publication/45172817_Variation_and_frequency_in_Russian_word_stress).

Marklund Sharapova, E. *Implicit and Explicit Norm in Contemporary Russian Verbal Stress*, Uppsala: Uppsala University (Acta Universitatis Upsaliensis. Studia Slavica Upsaliensia). 2000. Available at: <https://uu.diva-portal.org/smash/get/diva2:160762/FULLTEXT01.pdf>

Moretti F. *Distant reading*. London, New-York: Verso, 2013.

“*Naivnaya literatura*”: issledovaniya i teksty [“Naive literature”: studies and texts] / S. Yu. Neklyudov (comp.) Moscow, 2001. Available at: <http://www.ruthenia.ru/folklore/luriem43.pdf> (accessed on 14.09.2018). (In Russ.)

Ponomareva, M., Milintsevich, K., Chernyak, E., Starostin, A. Automated Word Stress Detection in Russian. In *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, pp. 31–35, Copenhagen, Denmark, September 7, 2017. Available at: <http://aclweb.org/anthology/W17-4104>. (In Russ.)

Pozharitskaya S.K., Dobrushina E.R. [Orthoepy in an era of corpus linguistics: some case studies in variation in standard Russian] // *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” (2017). Moscow. Available at: <http://www.dialog-21.ru/media/3940/pozharickajaskdobrushinaer.pdf> (In Russ.)

Reynolds, R., Tyers, R. Automatic word stress annotation of russian unrestricted text. In *Proceedings of the 20<sup>th</sup> Nordic Conference of Computational Linguistics (NODALIDA 2015)*. Linköping University Electronic Press, Sweden, Vilnius, Lithuania, pp. 173–180. Available at: <http://www.aclweb.org/anthology/W15-1822>.

Savchuk, S. Spoken texts representation in the Russian National Corpus: Spoken and Accentologic sub-corpora // “*NLP, Corpus Linguistics, Corpus Based Grammar Research*”, 5th International conference SLOVKO Smolenice, Slovakia, 25–27 November 2009 Brno, Tribun EU, 2009, pp. 310–320.

Savchuk S.O. [Active processes in the system of Russian inflection: A corpus-based research of accentological norms] // *Trudy II Mezhdunarodnoi konferentsii “Russkii yazyk i literatura v mezhdunarodnom obrazovatel'nom prostranstve: sovremennoe sostoyanie i perspektivy”* [Proceedings of the II<sup>nd</sup> International conference “Russian language and literature in the international educational space: current state and prospects”]. V. 2. Granada, 2010, pp. 1549–1554. (In Russ.)

Ukiah, N. ‘The stress of Russian nouns in -a and -я of Zaliznyak’s pattern f (рыба type)’, *Australian Slavonic and East European Studies*, 2002, 16/1–2, 1–39.

Vorontsova V.L. *Russkoe literaturnoe udarenie XVIII-XX v. Formy slovoizmeneniya* [Russian literary word stress of the XVIII-XX centuries, Forms of inflection]. Moscow, Nauka Publ., 1979. (In Russ.)

Zaliznyak A. A. *Ot praslavyanskoi aktsentuatsii k russkoi* [From proto-Slavic accentuation to Russian]. Moscow, Nauka Publ., 1985. (In Russ.)

Zaliznyak A. A. *Grammaticheskii slovar' russkogo yazyka* [Grammatical dictionary of the Russian language]. Moscow, Russkie slovari Publ., 2003. (In Russ.)

Zelenkov Yu.G., Zobnin A.I., Maslov M.Yu., Titov V.A. [Ilya Segalovich and development of ideas of computational linguistics to Yandex]// *Trudy mezhdunarodnogo seminara «Dialog14» po komp'yuternoi lingvistike i ee prilozheniyam* [Proceedings of the international seminar “Dialog14” on computational linguistics and its applications (electronic document)]. Available at: <http://www.dialog-21.ru/digests/dialog2014/materials/pdf/ZelenkovYuG.pdf>. (In Russ.)

*А. Е. Поляков*  
(Россия, Москва)  
*pollex@mail.ru*

## **ОРФОГРАФИЯ ОСТРОЖСКОЙ БИБЛИИ В КОНТЕКСТЕ ЦЕРКОВНОСЛАВЯНСКОГО КОРПУСА\***

Современная церковнославянская орфография достаточно стандартизирована и четко фиксирует правила употребления букв и диакритических знаков, использование сокращений (титл и буквотитл), правила словоделения, правила написания конкретных словоформ и различения омонимов. Наличие такой нормы позволило нам создать грамматический словарь и морфологический анализатор для церковнославянского языка (<http://dic.feb-web.ru/slavonic/dicgram/index.htm>), который используется в корпусе церковнославянских текстов (<http://ruscorpora.ru/search-orthlib.html>). Острожская Библия 1581 года, как первое полное издание Библии на церковнославянском языке, также должна быть включена в корпус, но при этом возникла масса проблем. Орфография этого издания сильно отклоняется от современной нормы и отличается крайней нерегулярностью и нестабильностью. В данной статье анализируются орфографические особенности Острожской Библии и возможность ее автоматического морфологического анализа. Специально рассматриваются правила употребления дублетных букв и других графических средств, которые порождают массу вариантов написания для конкретного слова. В результате делается вывод о том, что морфологический анализатор должен быть существенно переделан, чтобы он мог обрабатывать тексты в старой орфографии, а также определяются направления дальнейшей работы.

*Ключевые слова:* церковнославянский язык, орфография, корпусная лингвистика, автоматический морфологический анализ

---

\* Работа выполнена при поддержке РФФИ, проект № 17-04-12064-ОГН «Разработка модулей НКРЯ для автоматической разметки и словарной поддержки старорусских и церковнославянских текстов».



## 2. Источники текста

Острожская Библия была подготовлена в текстовом виде благодаря усилиям автора и отдана в общее пользование в качестве открытого интернет-ресурса, который доступен по адресу: <http://dic.feb-web.ru/slavonic/corpus/0/bible1581/index.htm>.

Электронный текст подготовлен на основе следующих источников:

1) Факсимильная (отсканированная) копия Острожской Библии [ОБ факсимиле].

2) Острожская Библия: современный набор с параллельным переводом на украинский язык. Подготовил Рафаил (Роман) Торконяк [ОБ 2006].

Факсимильная копия сама по себе не может заменить текстовый корпус, поэтому в качестве основного был выбран текстовый набор Торконяка, несмотря на то, что он не является точным воспроизведением оригинала. Так, в оригинале клитики пишутся слитно с основным словом, а в тексте они отделены пробелами согласно современной норме (вѣнѣмъ порѣдѣ на землѣ → вѣ нѣмъ по рѣдѣ на землѣ). Кроме того, там есть отдельные опечатки и отклонения от оригинала, которые мы постарались исправить при возможности. Однако мы сохранили раздельное написание клитик, поскольку это необходимо для работы анализатора и поиска в корпусе.

Кодировка текста была приведена в соответствие с современным стандартом «Church Slavonic typography in Unicode» [Unicode TN41], насколько это возможно. Дело в том, что типографика Острожской Библии устроена весьма сложно, а ее интерпретация неоднозначна и допускает различные толкования. Например, в оригинале камора визуально практически неотличима от придыхания, поэтому нам пришлось восстанавливать это различие на основании теоретических правил, а не реального текста (см. п. 6).

Для улучшения читаемости текста были сделаны некоторые упрощения при передаче символов, которые отсутствуют в стандартном наборе, чтобы пользователь мог видеть текст без установки специальных шрифтов. В частности, в данной статье буквотитла заменены на буквы с верхним индексом (ап<sup>л</sup>лѣ, во<sup>м</sup>ме<sup>т</sup>, вѣ<sup>л</sup>л<sup>н</sup>нѣ, ѿ<sup>л</sup>лѣ, зѣ<sup>н</sup>нѣ), а придыхания воспроизводятся только в разделе 6.2, специально посвященном этому вопросу. Далее мы часто будем давать примеры в несколько упрощенном виде, что, впрочем, никак не влияет на результаты анализа.

## 3. Общая характеристика графико-орфографической системы

Основной особенностью средневековой кириллической графики является наличие большого числа дублетных (омофоничных) букв, для которых нужно было придумать какие-то правила употребления [Карский 1979; Каверина 2010; Кусмалъ 2014]. Любопытно сравнить эту ситуацию с латиницей, где, наоборот, букв было мало и для изображения недостающих фонем приходилось использовать диграфы (ch sh th gn nh ll lh ae oe ue) и надстрочные знаки (ç š ž ñ ä ö ü).

В древнейшей кириллице были дублетные буквы И=І, О=Ѡ. В собственно славянских словах употреблялись в основном И О, буква І использовалась

в сокращениях и как компонент диграфов **ІѦ** **ІѸ** **Ю Ѧ** **ІѦ**, а буквы **Ѹ** **Ѩ** **Ѧ** **Ѧ** **Ѧ** **Ѧ** должны были использоваться только для греческих слов. Буква **Ѹ** первоначально произносилась как греческое [y], но в результате фонетических изменений в самом греческом совпала с **И**=**І**=**Ѹ**. Буква **Ѧ** произносилась как [ф] или [т] в разных традициях, но в русском варианте совпала с **Ф**=**Ѧ**.

В русской кириллице за счет фонетических изменений появились новые пары дублетов **ІѦ**=**ІѦ**, **Оу**=**Ѩ**, **Ю**=**ІѦ**, **З**=**С**, затем вторые варианты фактически перестали употребляться на письме. Позже некоторые из этих букв были частично восстановлены как подражание южнославянским текстам, вместе с другими орфографическими особенностями.

В рукописной традиции появились графические варианты букв, не имеющие аналогов в древнейшей кириллице, например, **О** широкое—**О** узкое—**О** очное, **Є** широкое—**Е** узкое, **Оу**—**Ѩ**, а также сложились некоторые правила их употребления. Например, **О** **Є** **Оу** обычно пишутся в начале слова или слога, тогда как **О** **Е** **Ѩ** в середине слова или слога (после согласной), но это правило нестрогое и часто нарушается (см. п. 4).

В древнейшей кириллице было четкое различие йотированных и простых гласных. Буквы **ІѦ** **ІѸ** **Ю Ѧ** **ІѦ** обозначали сочетания [й+гласная] или палатальность согласного (**земля**, **земле**, **земль**, **любити**), тогда как нейотированные буквы **Е**, **Ѧ** обозначали просто (полу)мягкость согласного. В ходе эволюции русского письма это различие стерлось из-за фонетических изменений, в результате буквы **ІѦ**=**Ѧ**, **ІѸ**=**Е** фактически стали дублетами — каждая из них могла обозначать сочетание [й+гласная] или палатализацию согласного. Прежнее различие частично сохранилось как контекстное правило — обычно **ІѦ** **ІѸ**/**Є** пишутся в начале слова или слога, **Ѧ** **Е** в середине слова или слога, но это правило нестрогое и не всегда соблюдается (см. п. 4).

Ниже мы рассмотрим правила употребления отдельных букв и графических знаков в тексте Острожской Библии по сравнению с современной церковнославянской нормой. Основное внимание будет уделено графическим дублетам, которые создают массу вариантов написания для конкретного слова.

## 4. Правила употребления отдельных букв

### 4.1. **И**=**І**=**Ѹ**

В современном церковнославянском буквы **И**—**І** являются позиционными вариантами в славянских словах: **І** пишется перед гласным, **И** перед согласным или на конце слова. Противопоставление **И**—**І** используется для различения нескольких омонимов: **міръ** vs. **міръ**, **кни́** vs. **кни́**. В греческих словах довольно четко соблюдается этимологическое различие **И**—**І**=**Ѹ**: **іа́аѸъ** vs. **іа́аѸъ**, **ікні́Ѹа** vs. **ікү́менѸ**, **іі́мѸнѸ** vs. **іү́меіѸнѸ**.

В Острожской Библии **И** достаточно регулярно заменяется на **І** перед гласной, в том числе на границе слов, которые в оригинале пишутся без пробела: **і** **ізнѸіѸ**,



І ОБЛАДАНТЕ, І МАНСА, І ОУПЕ, І ИЗЫНДОША. Впрочем, перед гласной может также писаться И: ВЪЗНАНѢТЪ-ВЪЗНАІЕТЪ, ВЪПНѢШИ-ВЪПНІЕШИ, ВЪСПРНАТИ-ВЪСПРНІАТИ, ЖРѢНИ-ЖРѢНІИ, ЗМНАІ-ЗМНІА, ЗМНІНЪ-ЗМНІНЪ, ИЗНАНІЕМЪ-ИЗНАНІЕМЪ, НЕПРНАЗНЬ-НЕПРНІАЗНЬ.

В греческих словах этимологическое различие І—И почти не соблюдается, причем И часто пишется перед гласной: *авна, д. аръ-авіа, д. аръ, архіерее-архирее, асавна-асавіа-асавіа, веннамінъ-веніамінъ, годоннілъ-годонілъ, елиезеръ-еліезеръ, іаковъ-іаковъ, ідолъ-ідолъ, незавель-іезавель, незекіа-іезекіа, нерел-іерел, исаакъ-ісаакъ, іавъ-іавъ, іаіа-іаіа-іаіа, иона-іона-іона, иоавъ-іоавъ, иоакимъ-іоакимъ, иона, д. анъ-іона, д. анъ, ионнъ-іоннъ-іуаннъ, іосифъ-іосифъ, манасіина-манасіина.*

Буква *У* регулярно пишется в следующих словах и их производных, иногда даже вместо этимологического правильного *И* (такие случаи помечены \*): *асу́рїа*, *асу́рїнскїй*, *баву́лонъ*, *баву́лонскїй*, *егу́петъ*, *егу́петскїй*, *ку́мвалы*, *ку́минъ*, *ку́парисъ-кппарисъ*. *ку́пръ-кн́пръ*, *лев́итинъ-лев́у́тинъ\**, *моу́сеи*, *моу́сеѡвъ*, *моу́ей*, *моу́ейѡвъ*, *нав́инъ-нав́унъ\**, *су́рїа*, *су́рїнскїй*, *су́ринъ*, *ску́менъ-скн́менъ*, *скн́їа-скун́їа\**, *скн́петръ-ску́петръ\**, *ту́мпаы-тн́мпаы*, *ту́ръ*, *ту́рскїй*.

## 4.2. $O=\mathbb{C}=\mathbb{G}$

В современном церковнославянском **О** пишется только в начале слова, а также в начале корня в сложных словах (**о**тѣцъ, прѣ**о**тѣцъ).

В Острожской Библии О регулярно заменяется на Ѡ в начале слова и слога: бѣгосѡбразенъ, косѡрѣженіе, добросѡбразна, изсѡбіліе-изсѡбіліе, изсѡбілне-изсѡбілне, обещалъ, обнѣтя, ѡблакъ, обнѣчаетъ, обѡе, ѡбразъ, обрѣщаете. Особенно характерно слово изсѡбіліе-изсѡбіліе, где Ѡ пишется после немого Ъ, но меняется на О при его устранении.

Это правило соблюдается абсолютно четко и распространяется на греческие слова: **антио́хх-антио́хх**, **антиохію-антио́хію**, **ахио́рх-ахио́рх**, **воо́зз**, **гавло́нх**, **гедео́нх**, **елесискаа**, **еліо́на**, **едіо́піа**, **ноавъз-іоавъз**, **ноакімъз-іоакімъз**, **ноанадд́анх-іоанадд́анх**, **ноа́ннх-іоа́ннх-іоа́ннх**, **іо́на-іо́на-іо́на**, **норд́анх-іорд́анх**, **нолаф́атъх-іолаф́атъх**, **но́иғх-іо́иғх**, **искарио́тъхкїи-искарио́тъхкїи**, **кнотъх-кїотъх**, **клеоп́атра**; а также на суффиксы/флексии: **андре́овъх**, **андре́ови**, **архі́ереовъх**, **архі́ереомъх**, **евре́омъх**, **нере́омъх-іере́омъх**;

Буква **Ѓ** по первоначальному замыслу должна была использоваться для греческих слов, но в реальности она стала применяться в русских словах, обычно в составе лигатуры **Ѓ**=ОТ.

В современном церковнославянском буква **Ґ** используется в следующих случаях:

- 1) в греческих словах в соответствии с этимологией (иωάννнз, ίώкз, είμωнз);
- 2) в приставке ω(ε)-, особенно в глаголах (ωενтáти vs. ѿвѣтѣль, ωблáчáти vs. ѿблáкз, ѿбразовáти vs. ѿбразз/ѿбразз, ѿбщíй);
- 3) для различения грамматических форм (рáбѡмз vs. рáбѡмз, мíллѡгтн vs. мíллѡгтн, нѡбáго vs. нѡбáгѡ, гíльно vs. гíльнѡ).









БЛА<sup>2</sup>КА, ВЪСКРѢСЕНІЕ-ВЪСКРѢНІЕ-ВЪСКРЕСЕНІЕ, ГЛАТН, ГЛАНІЕ, ГЪ, ГА, ГНЪ, ГНѦ, ГННМЪ, ДБАЖ-  
ДБАЖ-ДВ<sup>2</sup>Ъ, ДБДОВЪ-ДБДВЪ, ДБА, ДБАЦА, ДНЪ, ДНѢ, ДХЪ-ДХЪ, ДТѢ-ДТѢ, ДША, ЕУ<sup>2</sup>ЛІЕ,  
ІЕР<sup>2</sup>ЛНМЪ-ІЕР<sup>2</sup>ЛНМЪ, ІНЛЪ-ІНЛЪ, ІНЛЕВЪ-ІНЛЕВЪ, ІНЛЪТАНННЪ, ІѢ, ІЛА, ІДОВЪ, КНЪ-КНЪ,  
КНАЪ-КНАЪ, КР<sup>2</sup>ТЪ, КР<sup>2</sup>ТНТН, КРЦЕНІЕ, М<sup>2</sup>РДІЕ-МНОСЕРДІЕ<sup>2</sup>, М<sup>2</sup>ТНВЪ-МНОСТНВЪ,  
М<sup>2</sup>ТЪ-МНОСТЪ, МЛТВА-МОЛНТВА, МЦЪ, МРТВЫН, МРТВЕЦЪ, МТН, МТѢРН, НБД, НБЕ,  
НБННН, ННѢ, ОЦЪ-ОЦЪ, ОЧЕ-ОЧЕ, ОЧЕСТВА-ОЧЕСТВА, ПРП<sup>2</sup>БННХЪ, ПРР<sup>2</sup>КЪ-ПРДРКЪ,  
СНЦЕ, СМРТЬ, СМРТННН, СНЪ, СПЪ, СПНТЕЛЪ, СПЕНІЕ, СПЕТЪ, СТ<sup>2</sup>ТЪ, СТЪН, СТНТН,  
СТЫН, СЦЕНННН, ОУЧНН-ОУЧННН, ОУЧНТЕЛЪ-ОУЧТЪ<sup>2</sup>, ХЕ, ХА, ХНЪ, ХБЪ, ЦРЪ, ЦРНЦА,  
ЦРКН-ЦРСКН, ЦРТВІЕ-ЦРСТВІЕ, ЦРТВОКАТН-ЦРСТВОВАТН, ЦРКВЪ, ЦРКОВННН, ЧЛКЪ-  
ЧЛКЪ, ЧЛЧКН-ЧЛЧКН-ЧЛЧКН.

Список слов и корней под титлом в основном совпадает с современной нормой, однако здесь не проводится различие между сакральным и обычным употреблением. Например, слова ѿѿѿ=ѿѿѿ и ѿѿѿ пишутся под титлом применительно к обычным людям, тогда как в современном церковнославянском в этом случае они пишутся без титла.

## 5.2. Выносные буквы

Выносные буквы (буквотитла) использовались как способ сокращения слов при нехватке места в строке. Поскольку выносная буква ставится над предыдущей буквой, а не после нее, слово становится короче по горизонтали, но немного выше по вертикали. Например, написание **во<sup>3</sup>мет** vs. **во<sup>3</sup>метъ** занимает 4 позиции vs. 7, **вса<sup>3</sup>ни<sup>к</sup>** vs. **всадникъ** — 5 vs. 8, **лю<sup>3</sup>н** vs. **люди** — 2 vs. 4.

Выносные буквы создают массу проблем и являются источником ошибок. То, что легко делается в рукописном тексте, создает большие трудности при печати, поскольку поставить выносную букву на нужное место не всегда возможно. Во-первых, есть высокие буквы Ъ ъ, над которыми физически невозможно поставить надстрочный знак, поэтому его ставят над предыдущей буквой, то есть еще левее от нормальной позиции. Во-вторых, выносные буквы часто сдвигаются со своего места по непонятным причинам или в результате неаккуратности наборщика. В результате появляются слова с неправильным порядком букв, которые можно прочитать, только если передвинуть букву на правильное место: БЕЗ<sup>м</sup>ыныи→БЕЗ<sup>д</sup>мыни, Бѣгор<sup>о</sup>ныхъх→Бѣгоро<sup>д</sup>ныхъх, бл<sup>а</sup>ѡ→бл<sup>а</sup><sup>д</sup>, б<sup>н</sup>ѣ→б<sup>н</sup>ѣ<sup>п</sup>, брац<sup>к</sup>ѣ→брац<sup>к</sup>ѣ<sup>х</sup>, бр<sup>ш</sup>ана→бр<sup>ш</sup>на, б<sup>с</sup>ѣ→б<sup>с</sup>ѣ<sup>с</sup>, бѣд<sup>а</sup>ѡ→бѣд<sup>а</sup><sup>т</sup>, б<sup>х</sup>ѣ→б<sup>х</sup>ѣ<sup>х</sup>, б<sup>д</sup>онаѡ→бо<sup>д</sup>наѡ, велич<sup>и</sup>ѣ→велич<sup>и</sup>ѣ<sup>м</sup>, бе<sup>л</sup>аю→бе<sup>л</sup>аю<sup>д</sup>, вѣд<sup>а</sup>ѣ→вѣд<sup>а</sup><sup>т</sup>ѣ<sup>в</sup>, вѣд<sup>а</sup>ѣ→вѣд<sup>а</sup><sup>т</sup>ѣ<sup>п</sup>, виногр<sup>д</sup>а→виногра<sup>д</sup>, в<sup>к</sup>ѣ→в<sup>к</sup>ѣ<sup>к</sup>, возд<sup>а</sup>ѡ→возд<sup>а</sup><sup>х</sup>, вр<sup>а</sup>тъѣ→вр<sup>а</sup>тъѣ<sup>х</sup>, в<sup>р</sup>ѣ→в<sup>р</sup>ѣ<sup>д</sup>, вс<sup>ѣ</sup>ѣ→вс<sup>ѣ</sup>ѣ<sup>м</sup>, вс<sup>ѣ</sup>ѣ→вс<sup>ѣ</sup>ѣ<sup>х</sup>, вш<sup>ѣ</sup>→вш<sup>ѣ</sup><sup>д</sup>, възв<sup>д</sup>нгоша→въз<sup>д</sup>нгоша, в<sup>ѣ</sup>ѣнаго→в<sup>ѣ</sup>ѣ<sup>н</sup>аго, възз<sup>р</sup>ѣ→възз<sup>р</sup>ѣ<sup>в</sup>, гла→гла<sup>д</sup>, гн<sup>ѣ</sup>→гн<sup>ѣ</sup><sup>в</sup>, гра<sup>д</sup>ѣ→гра<sup>д</sup><sup>ѣ</sup><sup>х</sup>, гр<sup>д</sup>а→гр<sup>д</sup><sup>а</sup>, гр<sup>ѣ</sup>ѣ→гр<sup>ѣ</sup><sup>ѣ</sup><sup>х</sup>, д<sup>ѡ</sup>ѡ→д<sup>ѡ</sup><sup>ѣ</sup>, др<sup>ѡ</sup>ѡ→др<sup>ѡ</sup><sup>ѣ</sup>, еб<sup>н</sup>ѡ→еб<sup>н</sup><sup>ѡ</sup><sup>х</sup>, мж<sup>ѡ</sup>→мж<sup>ѡ</sup><sup>ж</sup>, нар<sup>ѡ</sup>→нар<sup>ѡ</sup><sup>д</sup>, насл<sup>ѣ</sup>ѣтѣѣ→насл<sup>ѣ</sup>ѣ<sup>д</sup>ѣтѣѣ, пл<sup>ѡ</sup>ѡ→пл<sup>ѡ</sup><sup>д</sup>, тр<sup>ѡ</sup>ѡ→тр<sup>ѡ</sup><sup>д</sup>, тр<sup>ѡ</sup>ѡна→тр<sup>ѡ</sup><sup>д</sup>на.

Человек может легко прочитать слово с неправильным порядком букв, но программа этого сделать не сможет, если ей не указать специально. Поскольку изначально неизвестно, что буквы переставлены и где именно, программа будет

вынуждена перебирать весь словарь, чтобы найти подходящий вариант, и не факт, что этот вариант будет правильным.

Выносные буквы открывают возможность для орфографического разнообразия: любая согласная перед согласной или на конце слова может быть записана как выносная. Дополнительную путаницу вносит немой Ъ, который часто пишется не только в конце слова, но и в середине, а также может заменяться на надстрочный знак (паерок). В результате такой орфографической свободы в тексте Острожской Библии некоторые слова могут иметь более десятка (!) вариантов написания (если учесть дублетные буквы и необязательное ударение, см. п. 6): **авѡаѡм-авѡаѡмѡ-авѡаѡмѡ-авѡаѡмѡ, аммонѡ-аммонѡ-а<sup>м</sup>мо<sup>н</sup>-аммо<sup>н</sup>, бѣз<sup>на</sup>-бѣз<sup>на</sup>-бѣз<sup>на</sup>-бѣз<sup>на</sup>, бѣ<sup>ч</sup>ертѣ-бѣ<sup>ч</sup>ертѣ-бѣ<sup>ч</sup>ертѣ-бѣ<sup>ч</sup>ертѣ-бѣ<sup>ч</sup>ертѣ-бѣ<sup>ч</sup>ертѣ, ѡавѡлонскѣн-ѡавѡлонскѣн-ѡавѡлонскѣн-ѡавѡлонскѣн, ѡа<sup>д</sup>ртѡ-ѡа<sup>д</sup>ртѡ-ѡа<sup>д</sup>ртѡ-ѡа<sup>д</sup>ртѡ-ѡа<sup>д</sup>ртѡ, ѡз<sup>н</sup>бѣжѣ-ѡз<sup>н</sup>бѣжѣ-ѡз<sup>н</sup>бѣжѣ-ѡз<sup>н</sup>бѣжѣ, зѣмны<sup>х</sup>-зѣмны<sup>х</sup>-зѣмны<sup>х</sup>-зѣмны<sup>х</sup>-зѣмны<sup>х</sup>.**

Выносные буквы иногда употребляются и в других случаях. Например, выносное <sup>л</sup> довольно часто употребляется в составе диграфа **ж<sup>л</sup>**, который теоретически может обозначать специфическое сочетание [*жс'дж*'], являющееся звонким коррелятом *ц*=[*ш'ч*'], хотя на основании графики определить его фонетическое значение невозможно. Примеры (знак ~ означает фонетически близкие варианты): **блж<sup>л</sup>**ение ~ блж<sup>л</sup>ение, **блгооугож<sup>л</sup>**у ~ блгооугож<sup>л</sup>у, **вельбж<sup>л</sup>**ь ~ вельбж<sup>л</sup>ь, **вжж<sup>л</sup>**а<sup>л</sup>ете, **вжж<sup>л</sup>**ел<sup>л</sup>ение, **вжж<sup>л</sup>**ел<sup>л</sup>ти, **вжзбж<sup>л</sup>**аю ~ вжзбж<sup>л</sup>аю, **вжзграж<sup>л</sup>**, **вжздаж<sup>л</sup>**ь, **бж<sup>л</sup>**а, **браж<sup>л</sup>**а, **вжхж<sup>л</sup>**ах<sup>л</sup>, **вжхж<sup>л</sup>**ение ~ вжхж<sup>л</sup>ение, **вжж<sup>л</sup>**ах<sup>л</sup> ~ вжж<sup>л</sup>ах<sup>л</sup>, **граж<sup>л</sup>**ано<sup>л</sup> ~ **граж<sup>л</sup>**ано<sup>л</sup>и, **граж<sup>л</sup>**ение, **дож<sup>л</sup>**амн, **дожж<sup>л</sup>**ает<sup>л</sup>ь ~ **дожж<sup>л</sup>**ает<sup>л</sup>ь, **жж<sup>л</sup>**а, **жж<sup>л</sup>**ает<sup>л</sup>ь, **кж<sup>л</sup>**о, **мжж<sup>л</sup>**, **многаж<sup>л</sup>**ы, **нажж<sup>л</sup>**а ~ **нажж<sup>л</sup>**а, **нжж<sup>л</sup>**а ~ **нжж<sup>л</sup>**а, **оажж<sup>л</sup>**а ~ **оажж<sup>л</sup>**а.

Выносное <sup>ж</sup> часто употребляется вместо финального же: *а́ще<sup>ж</sup>, вьзави<sup>ж</sup>, да<sup>ж</sup>, до́нде<sup>ж</sup>, егда<sup>ж</sup>, еа<sup>ж</sup>, зане<sup>ж</sup>, нде<sup>ж</sup>, н<sup>ж</sup>, него<sup>ж</sup>, ни<sup>ж</sup>, ника́ко<sup>ж</sup>, никто<sup>ж</sup>, ничто<sup>ж</sup>, поне<sup>ж</sup>, та́ко<sup>ж</sup>, та́ко<sup>ж</sup>.*

Выносное <sup>г</sup> иногда употребляется вместо финального го (адъективная флексия): бѣз<sup>г</sup>ъмнѧ, влѣжнѧ, вѣше<sup>г</sup>, велѣикѧ, вѣдаѹщѧ, все<sup>г</sup>, вѣсѣка<sup>г</sup>, вѣшна<sup>г</sup>, вѣчна<sup>г</sup>, е<sup>г</sup>, едина<sup>г</sup>, земнѧ, мое<sup>г</sup>, нѣше<sup>г</sup>. В других случаях оно заменяет обычное гж: блѧ<sup>г</sup>, вре<sup>г</sup>, вѣздѣнѧ, вѣрѧ<sup>г</sup>, залогѧ, кннѧ<sup>г</sup>, ковче<sup>г</sup>, корча<sup>г</sup>, многѧ<sup>г</sup>, сѣтѣро<sup>г</sup>, полѣнѧ<sup>г</sup>.

В единичных случаях употребляется выносное <sup>а</sup>: бѣш<sup>а</sup>, бѣж<sup>а</sup>, кѣш<sup>а</sup>, кѣшим<sup>а</sup>, взынѣш<sup>а</sup>, взыш<sup>а</sup>, коевѣд<sup>а</sup>, възпрошаше, вѣк<sup>а</sup>, вѣк<sup>а</sup>, гдѣнѣ, гнѣв<sup>а</sup>, гдѣ<sup>а</sup>, десницѣ, добрѣ, дрѹг<sup>а</sup>, дѣлѣ, егдѣ, едѣнѣ, ефрѣдѣ<sup>а</sup>; а также выносное <sup>и</sup> (обычно в составе финального дн): вѣкѣ<sup>дн</sup>, влѹдн, кѣш<sup>и</sup>, вѣльвѣлѹдн, гдѣ<sup>дн</sup>, гдѣ<sup>дн</sup>, зѣ<sup>дн</sup>, лѹдн, мѣ<sup>дн</sup>, нарѣ<sup>дн</sup>, плѣ<sup>дн</sup>, погрѣ<sup>дн</sup>, рѣ<sup>дн</sup>, рѣ<sup>дн</sup>, сѣзѣ<sup>дн</sup>, сѣ<sup>дн</sup>.

В современном церковнославянском буквититла используются только в составе сакральных слов вместо титла: **лпꙗз**, **вцꙗ**, **блꙗтъ**, **влꙗка**, **кокꙗрꙗнїе**, **гдꙗ**, **ѡꙗꙗїе**, **крꙗтъ**, **млꙗдїе**, **млꙗтъ**, **мꙗꙗнїй**, **мцꙗ**, **нбꙗнїй**, **прꙗꙗнїй**, **прꙗтꙗча**, **прꙗтꙗз**, **прꙗꙗꙗз**, **рꙗꙗтꙗко**, **тꙗꙗца**, **хꙗꙗтꙗꙗз**, **цꙗꙗтꙗко**, **чꙗꙗтъ**. По типографским соображениям, выносная буква обычно не ставится над начальной буквой слова, а сдвигается вправо, поэтому чтобы прочесть слово, нужно сдвинуть выносную букву влево: **вцꙗ**→**вꙗцꙗ**, **гдꙗ**→**гꙗдꙗ**, **мꙗꙗнїй**→**мꙗꙗꙗнїй**, **мцꙗ**→**мꙗꙗꙗз**. В современной орфографии список таких слов конечен и эту проблему можно легко решить словарно, а не алгоритмически.

## 6. Ударение и придыхание

Знаки ударения и придыхания не употреблялись в древнейшей кириллице и были введены позже как подражание греческому письму [Карский 1979]. В греческом было три знака ударения (´ ` ~) и два знака придыхания, хотя фонетическое различие между ними было утрачено задолго до создания славянской письменности: все три ударения звучали одинаково, а придыхание не произносилось. В славянских языках ударение имело смысловоразличительный характер, а придыхание оказалось чисто графическим знаком, не несущем никакого смысла. В греческом придыхание ставилось в начале слова и тем самым обозначало словесную границу в слитном письме, а в славянском оно нередко ставилось внутри слова и обозначало слоговую границу, которая и так очевидна из контекста.

### 6.1. Ударение

В современном церковнославянском основным знаком ударения является оксия (´), которая регулярно заменяется на варию (˘) в конце фонетического слова, согласно правилам греческого языка. Камора (˘) употребляется для различения форм единственного vs. множественного/двойственного числа (благѣмъ vs. благѣмъ, дѣла́нѣ vs. дѣла́нѣ, раба́ vs. раба́, раба́ vs. раба́), хотя в этой функции с ней конкурирует различие Е—Є и О—О (см. п. 4).

В Острожской Библии ударение употребляется крайне непоследовательно. Одно и то же слово может писаться с ударением или без него, ударение может стоять в разных местах, ударений может быть несколько: а́внмелѣхъ-авнмелѣхъ-авнмелѣхъ, авраа́мла-авраа́мла-авраа́мла, а́рхїерей-архїерей-архїерей-архїерей, безако́новаша-безако́новаша-безако́новаша, боа́щенѣ-боа́щенѣ-боа́щенѣ. Наконец, ударение может быть сдвинуто со своего места из-за высоких букв (Ѣ Ѥ) или типографской небрежности, в результате чего ударение фактически оказывается над согласной (!): безѣ́мнын-безѣ́мнын-безѣ́мнын, бра́тѣ-бра́тѣ-бра́тѣ, бра́шна-бра́шна-бра́шна, бѣ́детѣ-бѣ́детѣ-бѣ́детѣ, бѣ́дешн-бѣ́дешн-бѣ́дешн-бѣ́дешн, бѣ́ла-бѣ́ла-бѣ́ла, бѣ́гы-бѣ́гы-бѣ́гы, велелѣ́пота-велелѣ́пота-велелѣ́пота-велелѣ́пота, вѣ́тѣхъ-вѣ́тѣхъ, вѣ́звѣщѣ-вѣ́звѣщѣ-вѣ́звѣщѣ, вѣ́зложитѣ-вѣ́зложитѣ-вѣ́зложитѣ, вѣ́зметѣ-вѣ́зметѣ-вѣ́зметѣ-вѣ́зметѣ, вѣ́змѣ-вѣ́змѣ-вѣ́змѣ, вѣ́зней-вѣ́зней-вѣ́зней, вѣ́зпн-вѣ́зпн-вѣ́зпн-вѣ́зпн-вѣ́зпн, вѣ́зрн-вѣ́зрн-вѣ́зрн-вѣ́зрн, вѣ́дѣша-вѣ́дѣша-вѣ́дѣша, вѣ́дѣрѣхъ-вѣ́дѣрѣхъ-вѣ́дѣрѣхъ.

Возможно, некоторые случаи сдвига ударения имеют под собой фонетическое основание, но при такой орфографической свободе их невозможно отличить от ошибок набора.

Основным знаком ударения также является оксия (´), которая регулярно заменяется на варию (˘) в конце фонетического слова. В некоторых односложных/служебных словах вместо оксии также употребляется камора (˘), которая обычно ставится над согласной: бо́-бо́-бо́, бы́-бы́-бы́-бы́, ѣ́ѣ-ѣ́ѣ-ѣ́ѣ, вѣ́а-вѣ́а-вѣ́а, вѣ́ѣ-вѣ́ѣ-вѣ́ѣ, вѣ́е-вѣ́е-вѣ́е, вѣ́езѣжѣнїе-вѣ́езѣжѣнїе-вѣ́езѣжѣнїе-вѣ́езѣжѣнїе-вѣ́езѣжѣнїе,

БѢА-БѢА́-БѢА́-БѢА, БѢЮ-БѢЮ́-БѢЮ, БѢѢ-БѢѢ́-БѢѢ, БѢАКХ-БѢА́КХ-БѢА́КХ-БѢАКХ,  
БѢы-БѢы́-БѢы́-БѢы, ГДѢ-ГДѢ́-ГДѢ́-ГДѢ, ДА-ДА́-ДА, ДБА-ДБА́-ДБА́-ДБА, ДВѢ-ДВѢ́-ДВѢ,  
ДБАДЕСАТЬ-ДБА́ДЕСАТЬ-ДБА́ДЕСАТЬ-ДБАДЕСАТЬ, ДБАНА́ДЕСАТЬ-ДБАНА́ДЕСАТЬ-ДБАНА́ДЕСАТЬ  
ДБАНА́ДЕСАТЬ, ДВОЮ́-ДВОЮ́-ДВОЮ́-ДВОЮ, ДВѢСѢ́-ДВѢСѢ́-ДВѢСѢ́, ДНН-ДНН́-ДНН, ДЦН-  
ДЦН́-ДЦН́-ДЦН́-ДЦН, ЕА-ЕА́-ЕА́-ЕА, ЗДѢ-ЗДѢ́-ЗДѢ́, ЗЛО́-ЗЛО́-ЗЛО́-ЗЛО́-ЗЛО́-ЗЛО́-ЗЛО́-  
ЗЛО, ЗЛО́ТРАДАНІЕ, ЗЛО́ТВОРЕНІЮ, НЛН-НЛН́-НЛН́-НЛН, КТО́-КТО́-КТО́, ЛН-ЛН́-ЛН́-ЛН,  
МН-МН́-МН, МНѢ́-МНѢ́-МНѢ́-МНѢ́-МНѢ́-МНѢ́, МЫ́-МЫ́-МЫ́-МЫ, МА́-МА́-МА́,  
НЕЖЕЛН́-НЕЖЕЛН́-НЕЖЕЛН́-НЕЖЕЛН́-НЕЖЕЛН́-НЕЖЕЛН́, НН́-НН́-НН́-НН, ННКТО́ЖЕ-ННКТО́ЖЕ-  
ННКТО́ЖЕ-ННКТО́ЖЕ-ННКТО́ЖЕ-ННКТО́ЖЕ-ННКТО́ЖЕ, ННЧТО́ЖЕ-ННЧТО́ЖЕ-ННЧТО́ЖЕ-НН-  
ЧТО́ЖЕ, НО́-НО́-НО́-НО́, НЫ́-НЫ́-НЫ́, НЮ́-НЮ́-НЮ́, НА́-НА́-НА́, НѢКТО́-НѢКТО́-НѢКТО́,  
ПОЧТО́-ПОЧТО́-ПОЧТО́-ПОЧТО́, РЦН́-РЦН́́-РЦН́-РЦЫ́-РЦЫ́-РЦЫ́, СЕ́-СЕ́-СЕ́-СЕ, СН́-СН́-СН́-СН,  
СТО́-СТО́-СТО́, ТА́-ТА́-ТА, ТН́-ТН́-ТН́, ТМА́-ТМА́-ТМА́-ТМА́-ТМА, ТО́-ТО́-ТО́-ТО, ТРН́-  
ТРН́́-ТРН́́, ТЫ́-ТЫ́-ТЫ́-ТЫ, ТА́-ТА́-ТА́, ТѢ́-ТѢ́, ЧТО́-ЧТО́-ЧТО́-ЧТО. Подробнее  
об употреблении каморы в данных случаях см. [Кусмауль 2017].

Поскольку ударение в Острожской Библии употребляется крайне нерегулярно, то с точки зрения морфологического анализатора придется его просто игнорировать. При этом мы теряем возможность различить некоторые омонимы (вѣды vs. кѣды), но это небольшая потеря на общем фоне.

## 6.2. Придыхание

В современном церковнославянском придыхание автоматически ставится над начальной гласной слова и не имеет смыслоразличительной функции.

В Острожской Библии придыхание регулярно ставится над начальной гласной слова, если только оно не пропущено по типографской небрежности. Кроме того, придыхание иногда ставится над начальной гласной слога внутри слова, особенно в следующих случаях:

1) довольно часто над гласной в начале слога в иностранных словах: аа́ро́нз, а́виа́а́.а́рх-а́ви́а.а́рх, а́виоу́дз, а́ермо́нз, а́нтіо́хїа, а́рхне́рен-а́рхте́рен, бе́ннамі́нз-бе́нтіа́мінз, б́и.а́.ле́емз-б́и.а́.ле́омз, бо́о́зз, га́лаа́дз, ге́део́нз, е́лаа́зрз, еоу́сен, заоу́лонь, на́фе́тз, не́забе́ль-і́е́забе́ль, не́зе́кіа-і́е́зе́кіа, не́ре́н-і́е́ре́н, не́ремі́а-і́е́ремі́а, не́рихо́нз-і́е́рихо́нз, не́ровоа́мз-і́е́ровоа́мз, не́ро́лмз-і́е́ро́лмз, і́есен, і́нль, і́нльта́нінз, на́оа́до́рз, но́а́вз-і́о́а́вз, но́а́кнімз-і́о́а́кнімз, но́а́на.а́.а́нз-но́а́на.а́.а́нз-і́о́а́на.а́.а́нз, но́а́ннз-і́о́а́ннз, но́на.а́.а́нз-і́о́на.а́.а́нз, но́ра́днз-і́о́ра́днз, но́са́фа́тз-і́о́са́фа́тз, но́сі́фз-і́о́сі́фз, ноу́да-ноу́да-іоу́да-іоу́да, ноу́деа-іоу́деа, ноу́де́нска-іоу́де́нска, ноу́де́нска-іоу́де́нска, ка́пернаоу́мз, ка́рна.а́.іа́рнімз-ка́ріа.а́.фіа́рнімз-ка́ріа.а́.іа́рнімз, кио́тз-кио́тз-кіо́тз-кіо́тз, ле́гео́нз, ма́діа́мз, ма́насі́ебз, ма́насі́нінз-ма́насі́ннз, мо́а́вз, моу́сі́о́вз, на́лсе́о́нз, не́ема́нз, не́ема́а, но́е́лмі́ннз, са́мо́наз, са́мџ́наз, саоу́лз-саоу́лз, сі́мео́нз-сі́мео́нз-сџ́мео́нз, фа́рао́нз-фа́рао́нз, ха́наа́нз-ха́наа́нз;

2) довольно часто над гласной в начале корня в сложных словах: бѣгоуобразенъ, бѣгоуучіиѣ, вѣдѣтроуѣдѣ, вѣдоуѣдѣ, войсѣтнинѣ, въобразитѣлѣ, въоружитѣлѣ, двѣйзыченѣ, доидѣоша, законооучитель, заоутра, звѣроудинны, свѣрообразно, извѣбнѣе, извѣстрѣнѣ, извоумѣніѣ, наѣстѣилѣ, наоучитѣлѣ, извѣбитѣ-извѣбитѣ,



изидѣніе, мимоидеть, найдѡша, неизреченныхъ, немѣніе, неистовствѣа,  
нейцѣлаю, необрѣзанныхъ, неопісанъ, неудѡбъ, неукрашена, неѣсати, обоудъ-  
обоудъ, обѡймо, обзоученіе, обзоушіе, обзѣмлютъ, обзѣкнѣтъ, обзѣденіе,  
обзѣти, обзидѡша, обзѣмлютъ, ѡбнѡдъ, ѡѣмла, ѡѣти-ѡѣти, ѡнде, ѡнметъ,  
полѣмлетъ, полѣремникъ-полѣремникъ, поѣдиномъ, поѣмлютъ, поѣмше, поимнѣ,  
пойстиннѣ, поитнѣ, поицѣши, поустѣти, поучѣнѣа, поидѡша-поидѡша,  
прѣидеть, прѣидѡша, прѣобразѣа, прѣпоѣсана, прѣпоѣсаніе, прѣукрашена, прѣмше-  
прѣмше, прїидѡша, прїимати, прїобрѣа, прїѣти-прїѣти-прїѣти-прїѣти,  
прондѣте, пронзѡблѣтъ, прозѣблѣ, сзѡбѣщнѣ, оутроудѣ, четверооугólnыхъ;

[illegible]

4) иногда в определенных глагольных флексиях и суффиксах причастий: даёте, даётъ, даёцин, движущаѣся, дѣлаётъ, дѣлаёшин, дѣлайте, дѣлають, дѣлающе, избывающе, исповѣдаѣся, исповѣдающеся, исповѣдаѣи, кланяѣтеса, кланяющеся, негнюще, ѿкрѣютъ, работаѣмъ, работаётъ, работайте, работаютъ, работающе, разумѣмъ, разумѣете, разумѣетъ, разумѣёшин, разумѣють, разумѣюще, стоите, стоятъ, стоиши, стоѣла, стоѣше, стоѣца, требѣмъ, требѣте, требуютъ, требующа, требующе;

5) в отдельных словах: бо́ѡзнь-боѡзнь-боѡзнь, боѡтиса, боѡи҃са, боѡхѣса,  
боѡшесѧ, боѡщесѧ, боѡщисѧ, коѡбанїи, коѡбѣти, коѡбѣша, коѡвода, коѡнхъ, коѡнна,  
коѡини, коѡинникъ, коѡинство, коѡинствїи, плѡхїина.

Придыхание внутри слова почти всегда необязательно и может отсутствовать. Достаточно регулярно оно ставится только над буквами **Oy**, **O** (реже над **И**), таким образом, здесь диакритика становится как бы частью буквы. Это создает дополнительное визуальное противопоставление между «автономными» **Oy**, **O**, которые могут образовывать самостоятельный слог, и «связанными» **V**, **O**, которые всегда стоят после согласной.

## 7. Отдельные особенности.

Перечислим кратко другие орфографические особенности Острожской Библии:

### 7.1. Отсутствие буквы Й

Буква Й исторически представляет собой сочетание И с надстрочным знаком «краткая». В Острожской Библии этот знак употребляется крайне редко и непоследовательно в конце слова или флексии: *кѡй—кѡн, вѣй—вѣн, ѿй—ѿн, ѿйже—ѿйже, смѣй—смѣн, мой—мой—мон, коѡй—коѡн, гѣй—гѣн, гѣй—гѣн, тѡй—тѡн*. Обычно краткость И никак не обозначается, что порождает множество грамматических омонимов (*мой=мон, край=кран*). Это создает огромную проблему для морфологического анализа: приходится переделывать всю систему флексий и часть словаря.



В некоторых корнях достаточно часто пишутся архаичные/южнославянские сочетания Р/Л+Ъ/Ь или Ъ/Ь+Р/Л вместо более обычных русских ЕР, ОР, ОЛ: влѣкъ, влѣчѣцъ, влѣна, влѣнѣнїе—влѣнѣнїе, влѣхъ—влѣхъ, влѣхвобанїе—влѣхвобанїе, влѣхвобачи—влѣхвобачи, врѣхъ—врѣхъ, врѣтѣпъ—врѣтѣпъ, врѣтоградъ—врѣтоградъ, врьвъ, врьжетъ, врьзи, вьврѣже—вьврѣже, вьздрѣжанїе, вьстрѣгачи, дѣгъ, дѣготъ, дѣготерпѣнїе—дѣготрѣпѣнїе, дѣготрѣпѣчи—дѣготрѣпѣчи, дѣготрѣпѣливъ—дѣготрѣпѣливъ, дѣженъ, дѣжннѣхъ, дрѣжачица, дрѣзачи, дрѣзновѣнїе, дрѣзотъ, исплѣннѣчи, истрѣгнетъ, наплѣннѣчи, оплѣчѣнїе, оплѣчачица, плѣкъ, плѣнъ, плѣтъ, прѣхъ—прѣхъ, прѣвенѣцъ—прѣвенѣцъ, прѣвороднын—прѣвороднын, прѣтрѣпѣчи, прѣси, прѣскѣи, прѣсчѣнъ, прѣсчѣ, прѣсчѣ, рѣстрѣгнетъ, рѣстрѣзачи, рѣстрѣзанїе.

В приставках **вз**, **взз/взи**, **из** часто пишется Ъ независимо от произношения, которое может быть [о] или нуль, причем точно это установить невозможно: **въвлече́**, **въврѣго́ша**, **възбра́нѣти**, **възбѣа́ти**, **възвесели́ти**, **възвра́тити**, **възвышѣ́ние**, **възвѣстѣ́ти**, **възда́ние**, **въздави́ати**, **възража́ние**, **възложѣ́ти**, **възлюбѣ́ти**, **възнестѣ́ти**, **възопѣ́ти**, **възрадова́тиса**, **възрѣ́ти**, **въпросѣ́ти**, **въспрѣа́ти**, **въсѣта́ти**, **въсѣто́къ**, **въходо́ни**, **въѿрѣ́тиса**, **събла́зна**, **съблюсти́**, **събо́ръ**, **събра́ти**, **сѣтворѣ́ти**, **сверши́ти**, **свлеци́**, **свѣща́ти**, **схргѣши́ти**, **сзба́ти**, **сзнда́ти**, **скрѣбнице**, **скры́ти**, **сѣсѣа́вити**, **сѣтворѣ́ти**, **сѣчета́ти**.

### 7.2.3. А вместо А после гласной

Иногда по болгарскому образцу вместо **А** пишется **А** после гласных **А, О**, реже после **Ъ, Оу**. Обычно это происходит во флексиях, реже в суффиксах и основах. Примеры:

1) А+А: бѣга, боѣщаѣ, бываѣн, бѣга, бѣлаа, ваѣнѣ, велѣкаа, велѣчаа, вѣзвѣщаѣн, вѣздаѣнѣ, вѣздаѣти, вѣздѣнѣаѣн, вѣзлагѣа, вѣзлаѣбленаа, вѣзскаѣа, блѣаѣ, внимаа, вселѣннаа, всаѣческаа, вѣтораа, вѣчнаа, даѣнѣ, даѣти, дѣвнаа, дѣбраа, жѣвѣщаа, запрѣщаа, земнаа, злаа, нѣваѣно, нѣпытаѣн, нѣтѣннаа, кѣа, крѣа, крѣснаа, мѣншаа, морѣскаа, наѣпаа (!), наѣпаѣти, наѣуѣчаѣн, нѣчаѣнный, нѣѣкаа, ѣваѣнниѣн, ѣвѣѣтаа, ѣвѣѣтаѣн, ѣвѣѣчаа, ѣѣлаѣннаа, паѣдаа, поѣвѣаа, поѣкаѣнѣ, поѣкаѣтиѣа, приѣкаѣаѣа, прѣчаа, прѣѣваа, расѣкаѣнѣ, расѣкаѣтиѣа, сѣѣѣѣаа, сѣѣѣаѣн, сѣѣѣѣннаа, сѣѣѣаа, сѣѣѣѣнаа, сѣѣѣѣаа, ѣѣмѣнѣраа, ѣѣдаѣѣтаа, ѣѣѣѣѣѣкаа, ѣѣаа, ѣѣаѣнѣ;

97

Ѧноа, поаѡх, поаѡша, покоа, преѡстоаѡнїе, преѡстоаѡти, препоаѡсанїе, препоаѡсанѡ, своа, твоа, тоа, тоаѡ, тоаѡн, тоаѡло, тоаѡнїе, тоаѡти, оубоаѡтиѡѡ;

3) Ъ+А: бѣгодѣаніе, възвѣаша, вѣано, грѣа, дѣаніе, дѣати, имѣа, имѣан, любодѣаніе, надѣаніе, надѣатиса, одѣанх, одѣаніе, одѣаша, поимѣаніе, поѣаша, прелюбодѣаніе, сѣаніе, сѣати, разсѣаніе, чародѣаніе;

4) У+А: бегѣдѣа, враждѣа, вѣрѣа, вѣрѣанѣ, испытѣа, недѣгѣа, неправѣа, обвинѣа, обѣаетѣ, обѣаа, повинѣанѣ, повинѣаа, показѣа, пррѣчетѣа, свѣдѣтельствѣа, радѣаа, требѣа.

Замена **А** на **А** также встречается после **Е, І** в славянских словах, но это явление наблюдается и современном церковнославянском, хотя значительно реже.

## 9. Заключение

Анализ орфографических особенностей Острожской Библии приводит нас к следующему выводу. Морфологический анализатор для церковнославянского языка требует существенной переделки, чтобы он мог обрабатывать тексты в старой орфографии, особенно с учетом царящей там орфографической свободы. Можно выделить следующие направления работы:

1) Применять более «агрессивные» методы унификации исходных написаний, чтобы свести вместе все возможные орфографические варианты, даже если это приведет к некоторой потере информации.

2) Научиться работать с неполной и неточной информацией, сделать нечеткий поиск в словаре и грамматических таблицах.

3) Ввести в словарь дополнительные варианты написания, которые невозможно вычислить алгоритмическими методами.

## Литература

Алипий (Гаманович). Грамматика церковно-славянского языка. Jordanville (N.Y.), 1964.

Плетнева А. А., Кравецкий А. Г. Церковнославянский язык. 4-е доп. изд. М., 2006.

Поляков А.Е. Корпус церковнославянских текстов: проблемы орфографии и грамматики // *Przegląd wschodnioeuropejski*. V. 1, 2014. S. 245–254. URL: <http://www.ruslang.ru/doc/church-slav/conf4/05-polyakov.pdf>.

Поляков А. Е. Грамматический словарь церковнославянского языка (по материалам корпуса). // Труды международной конференции «Корпусная лингвистика — 2017». СПб., 2017. С. 295–298.

Добрушина Е. Р., Кравецкий А. Г., Поляков А. Е. Корпус и частотный грамматический корпусный словарь церковнославянского языка в составе Национального корпуса русского языка // Национальный корпус русского языка: 10 лет проекту. Труды Института русского языка им. В. В. Виноградова. Вып. 6. М., 2015. С. 116–141.

Острожская Библия (факсимиле). URL: <http://www.vechnoe.info/bible/pdf>, <http://samstar-biblio.ucoz.ru/load/46-1-0-84>.

Острожская Библия: современный набор с параллельным переводом на украинский язык. Подготовил Рафаил (Роман) Торконяк. Львов, 2006. URL: <http://www.vechnoe.info/bible/ostrog-bible-ukranian/view>.

Church Slavonic typography in Unicode. URL: <http://www.unicode.org/notes/tn41/>.

Кириллица в Юникоде. URL: [https://ru.wikipedia.org/wiki/Кириллица\\_в\\_Юникоде](https://ru.wikipedia.org/wiki/Кириллица_в_Юникоде).

Карский Е. Ф. Славянская кирилловская палеография. М.: Наука, 1979. XIX, 494 с.

Каверина В. В. Становление русской орфографии в XVII–XIX вв. : правописный узус и кодификация / дисс. ... докт. фил. наук. М., 2010. 436 с. URL: <http://www.ruslang.ru/doc/autoref/kaverina.pdf>

Кусмауль С. М. Книжная справа 40-х годов XVII века // Slověne. 2014. № 1. С. 72–101.

Кусмауль С. М. Эволюция функций знака каморы в богослужебных изданиях кон. XVI — перв. пол. XVII в. // Вестник ПСТГУ. Серия III: Филология. 2017. Вып. 51. С. 21–47. URL: [http://pstgu.ru/download/1498120889.2\\_Kusmaul\\_21-47.pdf](http://pstgu.ru/download/1498120889.2_Kusmaul_21-47.pdf).

*A. E. Polyakov*  
(Russia, Moscow)  
*pollex@mail.ru*

## SPELLING OF THE OSTROG BIBLE IN CONTEXT OF THE CHURCH SLAVONIC CORPUS

Modern Church Slavonic spelling is fairly standardized and defines the rules for letters and diacritics, abbreviations (titlo and letter-titlos), word separation, spelling of word forms and distinction of homonyms. This standard helped us to create a grammar dictionary and morphological analyzer for Church Slavonic (<http://dic.feb-web.ru/slavonic/dic-gram/index.htm>), which is used in the corpus of Church Slavonic texts (<http://ruscorpora.ru/search-orthlib.html>). The Ostrog Bible (1581) is the first complete edition of the Bible in Church Slavonic and should be included into the corpus, but a number of problems have occurred. The spelling of this edition deviates from the modern standard and is extremely irregular and unstable. This article analyzes the spelling peculiarities of the Ostrog Bible and the feasibility of its automatic morphological analysis. Special attention is paid to duplicate letters and other graphic symbols which generate a lot of spelling variations for a particular word. We conclude that the morphological analyzer should be significantly reworked in order to process the old spelling, and determine the directions of further work.

*Keywords:* Church Slavonic, spelling, corpus linguistics, automatic morphological analysis

## References

Alipii (Gamanovich). *Grammatika tserkovno-slavyanskago yazyka*. [Grammar of Church Slavonic] Jordanville (N. Y.), 1964. (In Russ.)

Pletneva A. A., Kravetskii A. G. *Tserkovnoslavyanskii yazyk*. [Church Slavonic] 4th aug. ed. Moscow, 2006. (In Russ.)

Polyakov A. E. [Corpus of Church Slavonic texts: problems of spelling and grammar]. *Przegląd wschodnioeuropejski*. V. 1, 2014. pp. 245–254. Available at: <http://www.ruslang.ru/doc/church-slav/conf4/05-polyakov.pdf>. (In Russ.)

Polyakov A. E. [Grammatical dictionary of Church Slavonic (corpus-based)]. *Trudy mezhdunarodnoi konferentsii «Korpusnaya lingvistika—2017»* [Proceedings of the International Conference “Corpus Linguistics—2017”]. St. Petersburg, 2017. pp. 295–298. (In Russ.)

Dobrushina E. R., Kravetskii A. G., Polyakov A. E. [Corpus and frequency grammatical dictionary of Church Slavonic within the National corpus of Russian]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Russian Language Institute]. 2015, no. 6, pp. 116–141.

Ostrog Bible (facsimile). Available at: <http://www.vechnoe.info/bible/pdf>, <http://samstar-biblio.ucoz.ru/load/46-1-0-84>.

*Ostrozhskaia Bibliia: sovremennyyi nabor s parallel'nym perevodom na ukrainskii yazyk* [Ostrog Bible: modern typesetting with parallel Ukrainian translation] Prepared by Rafail (Roman) Torkonyak. L'vov, 2006. Available at: <http://www.vechnoe.info/bible/ostrog-bible-ukranian/view>.

Church Slavonic typography in Unicode. Available at <http://www.unicode.org/notes/tn41/>.

Cyrillic script in Unicode. Available at: [https://en.wikipedia.org/wiki/Cyrillic\\_script\\_in\\_Unicode](https://en.wikipedia.org/wiki/Cyrillic_script_in_Unicode).

Karskii E. F. *Slavyanskaya kirillovskaya paleografiya* [Slavic Cyrillic paleography]. Moscow, Nauka Publ., 1979. XIX, 494 p. (In Russ.)

Kaverina V. V. *Stanovlenie russkoi orfografii v XVII–XIX vv.: pravopisnyi uzus i kodifikatsiya*. Diss. dokt. fil. nauk. [Formation of Russian spelling in 17th–19th centuries: spelling usage and codification. Dr. phil. sci. diss.] Moscow, 2010. 436 p. Available at: <http://www.ruslang.ru/doc/autoref/kaverina.pdf> (In Russ.)

Kusmaul' S. M. [Book Correction in the 40s of the 17th Century]. *Slověne*. 2014. № 1. pp. 72–101. (In Russ.)

Kusmaul' S. M. [Evolution of the functions of the kamora sign in liturgical books from the late 16th to the first half of the 17th centuries]. *Vestnik PSTGU. Ser. III: Filologiya*. Moscow, 2017. Vyp. 51. pp. 21–47. Available at: [http://pstgu.ru/download/1498120889.2\\_Kusmaul\\_21-47.pdf](http://pstgu.ru/download/1498120889.2_Kusmaul_21-47.pdf).

**Н. В. Богданова-Бегларян, О. В. Блинова, Г. Я. Мартыненко, Т. Ю. Шерстинова**  
*Санкт-Петербургский государственный университет*  
*(Россия, Санкт-Петербург)*  
*n.bogdanova@spbu.ru, o.blinova@spbu.ru,*  
*g.martynenko@spbu.ru, t.sherstinova@spbu.ru*

## **КОРПУС РУССКОГО ЯЗЫКА ПОВСЕДНЕВНОГО ОБЩЕНИЯ «ОДИН РЕЧЕВОЙ ДЕНЬ» (ОРД): ТЕКУЩЕЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ\***

В статье описываются этапы создания и расширения корпуса повседневной русской речи «Один речевой день» (ОРД), а также требования к публикации его онлайн-версии. Методологическая основа ОРД — осуществление звукозаписей в максимально естественных условиях. Для участия отбирались информанты-добровольцы, готовые прожить день «с диктофоном на шее» и заполнить ряд анкет. Перед аудиозаписью информанты проходили инструктаж, на котором учились пользоваться звукозаписывающей техникой и получали «Памятку информанта». Мы просили вмешиваться в процесс звукозаписи как можно меньше и вести себя «как обычно». Все записи осуществлялись анонимно, информанты и их собеседники обозначались кодами, а частная информация в транскрипте специальным образом маркировалась. В результате материалы корпуса содержат преимущественно разговоры из частной жизни информантов. Что же касается требований к публикации материалов онлайн, важным является сохранение анонимности авторства речевого материала. Таким образом, транскрипты звукозаписей могут быть опубликованы только при условии анонимизации личных имен, фамилий, прозвищ, а также исключения из текстов, представленных на сайте, любой другой информации, которая может повлечь раскрытие личности говорящего. В статье предлагается способ анонимизации личных данных и ставится проблема цензурной редакции транскриптов.<sup>1</sup>

*Ключевые слова:* русский язык, устная повседневная речь, речевой корпус, многоуровневое аннотирование, личная информация, подготовка корпуса к публикации.

---

\* При поддержке гранта РНФ (проект № 18-18-00242 «Система прагматических маркеров русской повседневной речи»).

## 1. Введение

Корпус повседневной русской речи «Один речевой день» (ОРД) создается, пополняется и активно анализируется на филологическом факультете СПбГУ с 2007 г.<sup>1</sup> Целью создания корпуса является изучение устной речи, бытовой и профессиональной коммуникации. Методологическая основа ОРД — осуществление звукозаписей в условиях, максимально приближенных к естественным, для чего используется методика многочасового мониторинга, непрерывной 24-часовой записи всей речевой продукции информантов (см. [Asinovsky et al. 2009; Богданова-Бегларян и др. 2017]). Подобная методика сбора речевого материала используется в японских лингвистических исследованиях (см. об этом, например: [Сибата 1983; Campbell 2004]); кроме того, она применялась при создании демографического подкорпуса Британского национального корпуса [Burnard 2007].

Сегодня корпус ОРД содержит около 1250 часов звучания, более 2800 коммуникативных макроэпизодов, включает материалы речи 128 информантов (из них 68 мужчин и 60 женщин в возрасте от 18 до 83 лет), а также более 1000 их основных коммуникантов. Объем текстовых расшифровок корпуса превышает 1 млн словоупотреблений (см. о нем подробнее: [Русский язык... 2016]). В настоящей статье описываются этапы создания и расширения корпуса ОРД и его текущее состояние, а также возможные условия публикации его онлайн-версии.

## 2. Технология создания корпуса

### 2.1. Отбор и инструктаж информантов

Для участия в записи отбирались информанты-добровольцы, готовые прожить день «с диктофоном на шее» и заполнить ряд анкет. При отборе учитывался возрастной фактор (принять участие в записи могли лишь лица, достигшие 18 лет) и фактор родного языка (русский язык для информанта должен быть родным и основным). К участию в эксперименте приглашались исключительно информанты, живущие в городской среде.

Перед аудиозаписью добровольцы проходили инструктаж, в ходе которого учились пользоваться звукозаписывающей техникой, и получали «Памятку информанта», где отражены основные требования к организационным и техническим особенностям проведения исследования. В «Памятке» указывалось, в частности, что диктофон следует отключать только на время смены батареек (приблизительно раз в 6 часов). Мы также просили информантов по возможности выключать источники фоновых шумов (телевизор, радио и др.) и обязательно предупреждать собеседников о проводимой звукозаписи. Перед началом записи информанты подписывали «Согласие», подготовленное юридической службой СПбГУ.

---

<sup>1</sup> Работа по созданию корпуса была начата при поддержке гранта РГНФ (проект №07-04-94515е/Я).

## **2.2. Методика звукозаписи**

Первые фонограммы корпуса ОРД собраны с помощью цифрового диктофона WS-320M. В ходе расширения корпуса, выполненного в рамках проекта РНФ 2014–2016 гг.<sup>2</sup>, запись производилась с помощью профессиональных диктофонов Roland R09-HR с внешними конденсаторными микрофонами SONY ECM-T140 в формате PCM (WAV) 44 100Гц, 16 бит, стерео.

Мы стремились, чтобы в корпус ОРД вошли «обычные дни» наших информантов (без редких событий или из ряда вон выходящих происшествий). Мы просили также вмешиваться в процесс звукозаписи как можно меньше и вести себя как обычно. В упомянутой выше «Памятке информанта» оговорены основные режимы звукозаписи — «стационарный» и «мобильный». В «стационарном» режиме — при длительном нахождении в одном помещении — мы рекомендовали снимать диктофон и размещать его вблизи себя на какой-то поверхности (например, на столе), а выносной микрофон отключать. В «мобильном» режиме, предполагающем активные перемещения информанта, мы советовали вести запись с помощью петличного микрофона.

## **2.3. Анкетирование информантов, социологическая информация в базе данных**

В ходе записи все информанты вели «Дневник речевого дня». Это шаблон, в котором фиксируются основные события, произошедшие в течение дня, и описываются основные коммуниканты. «Дневник» содержит четыре поля: «время» (когда происходил разговор), «место» (где происходил разговор), «собеседники» (с кем происходил разговор), «вид деятельности» (что сопровождало разговор).

Кроме того, в ходе записи проводилось анкетирование: информанты корпуса заполняли одну социологическую анкету и проходили три психологических теста.

Социологическая анкета, в которую записывались данные информантов и его основных собеседников (коммуникантов), разработана на основе данных Федеральной службы городской статистики и традиционных для социолингвистики параметров описания говорящих [Bogdanova-Beglarian et al. 2016]. В анкете отражены пол, возраст, место рождения и наиболее длительного проживания информантов и др. Информация о каждом коммуниканте дополнена его социальной ролью по отношению к информанту.

Все данные поступают в информационную базу данных формата MS Access. Таблицы с данными об информантах и коммуникантах, взятыми из социологических анкет, содержат следующие основные поля: (1) код информанта, (2) пол информанта, (3) возраст информанта на момент записи, (4) место рождения, (5) родной язык, (6) другие языки, которыми владеет информант, (7) национальность

---

<sup>2</sup> Грант РНФ № 14-18-02070 «Русский язык повседневного общения: особенности функционирования в разных социальных группах».

родителей (заполняется по желанию), (8) социальное происхождение, (9) уровень образования (среднее специальное, высшее и т. п.), (10) квалификация (специальность) по диплому, (11) прошлые профессии или опыт работы, (12) профессия или род деятельности на момент записи, (13) места наиболее длительного проживания, (14) комментарии [Русский язык... 2016].

Записи велись анонимно, все информанты и их собеседники обозначены кодами. Код в формате S01, S10, S100 присваивается в зависимости от порядкового номера информанта. Коммуникантам (мужчинам, женщинам и детям) присваиваются коды типа M1, M2, Ж1, Ж2, P1, P2 — в зависимости от момента их появления в эпизоде коммуникации (звуковом файле и файле расшифровки со стандартным названием типа ordS01-nn, где 01 — код информанта, nn — номер макроэпизода «речевого дня»). Возможны коды НМ или НЖ — в случае, когда идентифицировать коммуниканта по записи не удается.

Три стандартных психологических теста (анкеты), которые выполняли участники эксперимента, — это тест Г. Айзенка, тест FPI и тест Р. Кеттела, с бланками ответов, подготовленными для заполнения. Психологическая информация представлена для 85 информантов корпуса (это звукозаписи, выполненные в 2014–2016 гг.).

### **3. Обработка речевого материала**

#### ***3.1. Форматирование файлов звукозаписи. Членение на «макроэпизоды»***

Форматирование файлов ОРД подразумевает преобразование первичного формата записи в формат PCM (22 050 Гц, 16 бит, моно), удаление пауз продолжительностью более 2–5 минут и членение исходных файлов звукозаписи на файлы, содержащие отдельные коммуникативные макроэпизоды.

Коммуникативные макроэпизоды — это фрагменты «речевого дня», объединенные местом, условиями и участниками коммуникации. Эпизоды описываются по определенной схеме с учетом параметров «тип коммуникации», «условия коммуникации», «социальные роли говорящих», «место коммуникации» (см. [Шерстинова 2013]).

#### ***3.2. Техника расшифровки, система аннотирования***

Расшифровка звукозаписей выполняется в программе ELAN [ELAN] и предполагает заполнение восьми базовых уровней: Frase (реплики говорящих), Speaker (код говорящего), Events (невербальные аудиособытия), Voice (качество голоса говорящего), FonetCom (фонетический комментарий), FraseComment (фразовый комментарий), Notes (общий комментарий), Episode (мини-эпизод речевой коммуникации) (подробнее см. [Sherstinova 2015; Русский язык... 2016]).

Уровень «Frase» — основной. Он содержит транскрипт звукозаписи, на нем производится членение звукозаписи на сегменты (боксы), не содержащие речевого



сигнала (паузы, обозначаемые символом <\*П>), и реплики говорящих. Транскрипт выполняется в стандартной орфографии.

Некоторые отступления от правил стандартной орфографии допускаются при расстановке пробелов, оформлении дефисных написаний. Эти отступления введены для того, чтобы избежать ряда проблем, влияющих на качество дальнейшей обработки текстов (лемматизации, морфологического анализа и др.) (см. [Блинова 2015]). Рекомендации для расшифровщиков требуют, чтобы цепочки символов, которые считаются отдельными токенами, были отделены пробелами с обеих сторон. Например, таким образом оформляются: словоформы, символы, обозначающие паралингвистические явления, паузы, наложения речи, знаки фразового и синтагматического членения </>, </> и др. Для неоднословных выражений введено одно правило: названия из двух и более слов объединяются с помощью нижнего подчеркивания (например, *От\_заката\_до\_рассвета*).

Инструкция для расшифровщика рекомендует оформлять написания, в графическом представлении которых задействован дефис, следующим образом: <\*то, \*таки> даются как отдельные слова, за исключением неопределенных местоимений и местоименных наречий (*где-то, куда-то* и т. д.), а также *всё-таки*; сложные слова с дефисом (за исключением прилагательных, обозначающих цвет, имен собственных и некоторых наречий, имеющих в своей основе дуплеты), а также лексические повторы принимаются за два отдельных слова. В более подробной инструкции оговаривается, что раздельно (через пробел, не по правилам орфографии) пишутся, во-первых, различные дуплеты: прилагательные (*быстрый быстрый*), наречия (*долго долго*), глаголы (*сидел сидел*), частицы (*да да*), междометия (*ха ха*) и др.; во-вторых, частицы *-то, -ка, -таки, -де* (*я то знаю, покажи ка, сказал таки*); в-третьих, существительные с приложением (*девушка красавица*); в-четвертых, числительные со значением приблизительного количества (*два три*).

Конвенции дискурсивной транскрипции ОРД подробно описаны в [Шерстинова и др. 2009]. В последнее время список стандартных помет расширен за счет обозначения паралингвистических явлений (<\*З> — зевок, <\*Ц> — цыканье, <\*S> — шмыганье носом, <\*G> — гортанные неречевые звуки, <\*Г> — причмокивание и др.). Введены новые обозначения для маркирования незавершенных реплик, продолжающихся после паузы. Для обозначения сегментов транскрипта, подлежащих анонимизации (прежде всего личных имен), введена дополнительная помета <%>.

### 3.3. Подготовка транскриптов к лингвистической разметке

Перед лингвистическим аннотированием (лемматизацией, морфологической разметкой, синтаксической разметкой и др.) файлы расшифровки \*.eaf подвергаются ручному экспертному редактированию и автоматической обработке. Во-первых, все файлы обрабатываются с помощью утилиты «Corrector» (собственная разработка научного коллектива), позволяющей автоматически исправить возможные технические огрехи (например, убрать из транскрипта лишние пробелы), а также выявить несоответствия между уровнями «Fraser» и «Speaker» в случаях наложения

речи (периодах одновременного говорения) нескольких собеседников. Наложения в повседневной речи наблюдаются достаточно часто и существенно затрудняют ее анализ. В случаях выявления несоответствий аннотаций на указанных уровнях проводится ручная коррекция соответствующих фрагментов расшифровок и их повторный анализ.

Во-вторых, файлы расшифровки \*.eaf проходят обработку программой «Eafer» (также собственная разработка коллектива). С ее помощью осуществляется преобразование одноуровневого представления речевого материала, принятого за основу транскрибирования корпуса ОРД с момента его основания в 2007 г., в многоуровневое представление, где каждый речевой уровень относится только к одному говорящему. Этот этап обработки позволяет разделить речевой материал разных говорящих.

В-третьих, выполняется ручная коррекция границ боксов аннотаций для реплик с наложением речи. Коррекция границ аннотаций осуществляется непосредственно в среде ELAN для каждого из уровней, содержащих реплики говорящих. После этого файлы аннотаций формата \*.eaf считаются готовыми к автоматической обработке.

#### **4. Подготовка ОРД к публикации: анонимизация данных и цензурная редактура**

Звукозаписи корпуса ОРД содержат преимущественно разговоры из частной жизни информантов. Анонимность участников звукозаписи (при открытости их социологических характеристик) — один из ключевых моментов методики сбора данных, который позволял участникам проживать свой «речевой день» свободно и «естественно». Поэтому наиболее важным условием публикации материалов корпуса онлайн является сохранение анонимности авторства речевого материала.

Для исключения атрибуции говорящего по акустическим свойствам голоса и манере речи сами звукозаписи повседневного общения, по-видимому, не могут быть опубликованы в свободном доступе. Что касается транскриптов звукозаписей, то они могут быть опубликованы только при условии полной анонимизации личных имен, фамилий, прозвищ, а также исключения из текстов, представленных на сайте, любой другой информации, которая может повлечь раскрытие личности говорящего (номера телефона, паспортных данных, конкретных мест работы и прочей озвученной информации). Кроме того, не все эпизоды речевой коммуникации могут быть опубликованы по этическим соображениям.

Сегодня актуальными задачами, стоящими перед авторским коллективом, работающим с корпусом ОРД, являются следующие:

1) определение типов эпизодов, которые не могут быть представлены на сайте ни в каком виде, к решению этой задачи будут привлечены квалифицированные юристы;

2) отбор коммуникативных эпизодов, которые могут быть опубликованы после их анонимизации;

3) анонимизация транскриптов: замена всей личной информации — в первую очередь имен и фамилий — на иные, имеющие ту же акцентно-ритмическую структуру;

4) «цензурная» редакция.

Повседневная устная речь содержит заметное количество непечатной лексики. Необходимо принять решение, каким образом такие выражения будут представлены на сайте (как известно, публикация нецензурных слов в сети запрещена). Однако для лингвистических исследований желательно, чтобы по тексту расшифровки можно было однозначно восстановить произнесенный текст, следовательно, предстоит получить полный список цензурируемых единиц и выработать способы их передачи.

## **5. Корпус ОРД онлайн**

В рамках текущего проекта РНФ коллективом запланирована публикация в сети Интернет со свободным доступом материалов корпуса с возможностью текстового поиска и фильтрации речевого материала по социологическим характеристикам говорящих и условиям коммуникации. Объем подкорпуса, планируемого к публикации, — 300 тыс. словоупотреблений.

Сайт корпуса ОРД будет опубликован в свободном доступе на сервере СПбГУ по адресу <http://www.ord-corpus.spbu.ru>. Пользовательский интерфейс онлайн-версии корпуса будет обеспечивать текстовый поиск по заданному слову (подстроке). Каждая реплика, полученная по запросу, будет сопровождаться социологической информацией о говорящем (пол, возраст, профессия и др.), а также о типе коммуникативной ситуации (бытовой разговор, профессиональный разговор, учебный разговор и др.). Будет предусмотрена возможность развертывания выдачи до нескольких реплик.

## **6. Заключение**

Корпус ОРД задуман как инструмент мониторинга современной русской речи, направлен на фиксацию материала естественной коммуникации и тем самым — создание условий для научного описания современного повседневного русского языка и анализа особенностей его функционирования в разных социальных группах, в разных ситуациях, в бытовом и профессиональном общении, при паритетных и непаритетных отношениях между говорящими — носителями различных социальных ролей. Корпус достиг значительного для речевого корпуса объема и продолжает развиваться. В настоящее время на его основе разрабатывается система прагматических маркеров повседневной русской речи. Сегодня готовится его публикация на сайте. Мы надеемся, что ОРД станет значимым общедоступным ресурсом для лингвистических исследований самых разных направлений.

## Литература

Блинова О. В. «Немашинное обучение»: экспериментальная проверка применимости правил ручной токенизации // Труды международной конференции «Корпусная лингвистика-2015». СПб. : Изд-во СПбГУ, 2015. С. 111–120.

Богданова-Бегларян Н. В., Шерстинова Т. Ю., Блинова О. В., Мартыненко Г. Я. Корпус «Один речевой день» в исследованиях социолингвистической вариативности русской разговорной речи // Анализ разговорной русской речи (АР<sup>3</sup>-2017): Труды седьмого междисциплинарного семинара / науч. ред. Д. А. Кочаров, П. А. Скрелин. СПб. : Политехника-принт, 2017. С. 14–20.

Русский язык повседневного общения: особенности функционирования в разных социальных группах: коллективная монография / отв. ред. Н. В. Богданова-Бегларян. СПб. : ЛАЙКА, 2016. 244 с.

Сибата Т. Исследования языкового существования в течение 24 часов // Языкознание в Японии / ред. В. М. Алпатов, И. Ф. Вардуль. М. : Радуга, 1983. С. 134–141.

Шерстинова Т. Ю., Степанова С. Б., Рыко А. И. Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Материалы XXXVIII Международной филологической конференции. Секция: «Формальные методы анализа русской речи». СПб. : Изд-во СПбГУ, 2009. С. 66–75.

Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD Speech Corpus of Russian Everyday Communication «One Speaker's Day»: Creation Principles and Annotation // Matoušek, V., Mautner, P. (eds.) TSD 2009. LNAI. Vol. 57292009. Berlin-Heidelberg : Springer, 2009. P. 250–257.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Baeva E., Martynenko G., Ryko A. Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech // SPECOM 2016. Lecture Notes in Artificial Intelligence, LNAI. Vol. 9811. Switzerland : Springer, 2016. P. 659–666.

Burnard L. (ed.) Reference Guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by Oxford University Computing Services, 2007 [Электронный ресурс]. URL: <http://www.natcorp.ox.ac.uk/docs/URG/>.

Campbell N. Speech & Expression; the Value of a Longitudinal Corpus // Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26–28. Lisbon, Portugal. 2004. P. 183–186.

ELAN [Программное обеспечение]. Nijmegen: Max Planck Institute for Psycholinguistics. URL: <https://tla.mpi.nl/tools/tla-tools/elan/>.

Sherstinova T. Macro Episodes of Russian Everyday Oral Communication: towards Pragmatic Annotation of the ORD Speech Corpus / Ronzhin A. et al. (eds.) SPECOM 2015. Lecture Notes in Artificial Intelligence, LNAI. Vol. 9319. Switzerland : Springer, 2015. P. 268–276.

Sherstinova T. The Structure of the ORD Speech Corpus of Russian Everyday Communication // Text, Speech and Dialogue. Proceedings of the 12th International Conference, TSD-2009. LNAI 5729. Switzerland : Springer, 2009. P. 258–265.

**N. V. Bogdanova-Beglarian, O. V. Blinova, G. Ya. Martynenko, T. Yu. Sherstinova**

*Saint Petersburg State University*

*(Russia, Saint Petersburg)*

*{n.bogdanova, o.blinova, g.martynenko, t.sherstinova}@spbu.ru*

## **RUSSIAN EVERYDAY SPEECH CORPUS “ONE DAY OF SPEECH”: CURRENT STATE AND PERSPECTIVES**

The article describes some stages of creation and extension of the Russian everyday speech corpus “One Day of Speech” (the ORD corpus), as well as peculiar conditions for publishing its online version. Speech material of the ORD corpus was obtained in natural communicative situations. Volunteer-respondents were people who expressed their willingness to live a day “with a dictaphone dangling around their necks” and fill out several questionnaires. Before recording, the respondents were instructed; they learned to use sound recording equipment and received the “Informant’s memo”. We asked them to intervene in recording process as little as possible, not to turn off the dictaphone and behave “as usual”. All records are anonymous, respondents and their interlocutors are named using codes, personal information in transcripts is marked in a special way. The ORD corpus contains mainly private conversations. An important requirement for the publication of its online version is ensuring anonymity of speakers. Thus, transcripts of sound recordings can be published only after the anonymization of personal names, surnames, and nicknames, and only after any information that may lead to the disclosure of the speaker’s identity is excluded from the texts.. The article describes the method proposed for anonymization of personal data and poses the problem of censorship of text transcripts.

*Key words:* Russian language, everyday spoken speech, speech corpus, personal information, pre-publishing processing.

### **References**

Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD Speech Corpus of Russian Everyday Communication “One Speaker’s Day”: Creation Principles and Annotation. TSD 2009. Ed. by V. Matoušek, P. Mautner. *LNAI*, vol. 5729/2009. Berlin-Heidelberg, Springer, 2009, pp. 250–257.

Blinova O. V. [“Non-Machine Learning”: Experimental Verification of Manual Tokenization Rules]. *Trudy mezhdunarodnoi konferentsii “Korpusnaya lingvistika-2015”* [Proceedings of International Conference CORPORA 2015]. St. Petersburg, St. Petersburg St. Univ. Publ., 2015, pp. 111–120. (In Russ.)

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Baeva E., Martynenko G., Ryko A. Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech. *SPECOM 2016, Lecture Notes in Artificial Intelligence, LNAI*, vol. 9811. Switzerland, Springer, 2016, pp. 659–666.

Bogdanova-Beglarian N. V. (ed.) *Russkii yazyk povsednevnogo obshcheniya: osobennosti funktsionirovaniya v raznykh sotsial’nykh gruppakh. Kollektivnaya monografiya*

[Everyday Russian Language in Different Social Groups. Collective. Monograph]. St. Petersburg, 2016. 244 p. (In Russ.).

Bogdanova-Beglarian N. V., Sherstinova T. Ju., Blinova O. V., Martynenko G. Ja. [Corpus “One Speaker’s Day” in Studies of Sociolinguistic Variability of Russian Colloquial Speech]. *Analiz razgovornoj russkoi rechi (AR<sup>3</sup>-2017): Trudy sed'mogo mezhdistiplinarnogo seminar* [Analysis of Spoken Russian Speech (AR<sup>3</sup>-2017): Proceedings of the 7<sup>th</sup> Interdisciplinary Seminar]. St. Petersburg, 2017, pp. 14–20. (In Russ.)

Burnard L. (ed.) Reference Guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by Oxford University Computing Services, 2007. Available at: <http://www.natcorp.ox.ac.uk/docs/URG/> (accessed 23.06.2018).

Campbell N. Speech & Expression; the Value of a Longitudinal Corpus. *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC-2004, May 26–28*. Lisbon, Portugal, 2004, pp. 183–186.

ELAN [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics. Available at: <https://tla.mpi.nl/tools/tla-tools/elan/> (accessed 23.06.2018).

Sherstinova T. Macro Episodes of Russian Everyday Oral Communication: towards Pragmatic Annotation of the ORD Speech Corpus. Eds. A. Ronzhin et al. *SPECOM 2015, Lecture Notes in Artificial Intelligence, LNAI*, vol. 9319. Springer, Switzerland, 2015, pp. 268–276.

Sherstinova T. The Structure of the ORD Speech Corpus of Russian Everyday Communication. Text, Speech and Dialogue. *Proceedings of the 12th International Conference, TSD-2009. LNAI 5729*. Switzerland, Springer, 2009, pp. 258–265.

Sherstinova T. Ju., Stepanova S. B., Ryko A. I. [Annotation System in the Russian Sound Corpus “One Day of Speech”]. *Materialy XXXVIII Mezhdunarodnoi filologicheskoi konferentsii. Sektsiya: “Formal’nye metody analiza russkoi rechi”* [Proceedings of the XXXVIII International Philological Conference. The Conference Section “Formal Approaches to Analysis of Russian Speech”]. St. Petersburg, St. Petersburg State University Publ., 2009, pp. 66–75. (In Russ.)

Sibata T. [A Twenty-Four-Hour Survey of the Language Life]. *Yazykoznanie v Yaponii* [Linguistics in Japan]. Ed. by V. M. Alpatov, I. F. Vardul’. Moscow, Raduga Publ., 1983, pp. 134–141. (In Russ.)

**Н. В. Богданова-Бегларян, О. В. Блинова, К. Д. Зайдес, Т. Ю. Шерстинова**  
*Санкт-Петербургский государственный университет*  
*(Россия, Санкт-Петербург)*  
*n.bogdanova@spbu.ru, o.blinova@spbu.ru, kristina.zaides@student.spbu.ru,*  
*t.sherstinova@spbu.ru\**

**КОРПУС «СБАЛАНСИРОВАННАЯ АННОТИРОВАННАЯ ТЕКСТОТЕКА»  
(САТ): ИЗУЧЕНИЕ СПЕЦИФИКИ РУССКОЙ МОНОЛОГИЧЕСКОЙ  
РЕЧИ\***

Статья представляет один из корпусов русской устной речи: коллекцию спонтанных монологических текстов, известную как «Сбалансированная аннотированная текстотека» (САТ). Данный корпус собирается в Санкт-Петербургском государственном университете в течение уже более чем 20 лет с использованием авторской (Н. В. Богдановой-Бегларян) методики сбора данных, предполагающей достаточно строгий набор экспериментальных процедур. САТ предназначен для изучения спонтанных монологов разного типа (чтение (сюжетного и несюжетного исходных текстов), пересказ прочитанных текстов, описание изображения (также сюжетного и несюжетного), рассказ на заданную тему) и содержит тексты, записанные от пяти профессионально-ориентированных групп носителей языка (медики; юристы; «компьютерщики»; филологи, преподаватели русского языка как иностранного; преподаватели-философы), несколько блоков речи студентов (филологов и нефилологов), а также четыре блока интерферирующей русской речи носителей других языков: американского, английского, китайского, французского и нидерландского. Всего в составе САТ сегодня около 700 текстов и около 50 часов звучания. В статье на фоне других русскоязычных и иноязычных устных корпусов дано описание данного лингвистического ресурса, отмечены основные темы, разрабатываемые на его материале, а также намечены перспективы продолжения работы.

*Ключевые слова:* современный русский язык, устная монологическая речь, звуковой корпус, обработка естественного языка, база данных, лингвистический

---

<sup>1</sup> Исследование выполнено при поддержке гранта РФФИ № 17-29-09175 «Диагностические признаки социолингвистической вариативности повседневной русской речи (на материале звукового корпуса)».



эксперимент, чтение, описание изображения, пересказ текста (репродуктив), спонтанный монолог, социолингвистика, психолингвистика.

Исследованием разных аспектов русской спонтанной монологической речи в разные годы занимались многие лингвисты, что лишний раз свидетельствует о важности этой темы для традиционной русистики. В этой связи следует отметить работы фонетистов — Л. В. Бондарко, Н. Д. Светозаровой, Н. Б. Вольской, Н. И. Гейльман, О. Ф. Кривновой, С. В. Кодзасова, Л. Л. Касаткина, психолингвистов — Л. С. Выготского, Н. И. Жинкина, А. А. Леонтьева, А. Р. Лурии, И. Н. Горелова, К. Ф. Седова, социолингвистов — Б. Н. Головина, Л. П. Крысина, А. П. Мартынюка, Т. И. и Е. В. Ерофеевых, А. В. Кириловой, М. Краузе, специалистов в области лингвистической экспертизы — к примеру, Е. И. Галяшиной, диалектологов — Е. А. Оглезневой, И. А. Букринской, О. Е. Кармаковой. В контексте исследования устного монолога необходимо упомянуть также имена Н. Ю. Шведовой, О. А. Лаптевой, О. Б. Сиротининой, Е. В. Красильниковой и ряд других. Однако эта ниша в коллоквиалистике все еще нуждается в детальном и многоаспектном исследовании. Именно этой цели служат материалы корпуса «Сбалансированная аннотированная текстотека» (САТ), которому и посвящена настоящая статья.

САТ представляет собой коллекцию монологических текстов, которая собирается в Санкт-Петербургском государственном университете в течение уже более чем 20 лет (см., например: [Богданова и др. 2008; Богданова 2010; Богданова-Бегларян и др. 2017a]) с использованием авторской (Н. В. Богдановой-Бегларян) методики сбора данных, предполагающей достаточно строгий набор экспериментальных процедур. САТ — это во многом уникальный лингвистический ресурс, предназначенный для изучения спонтанных монологов разного типа. К настоящему времени корпус содержит тексты, записанные от пяти профессионально-ориентированных групп носителей языка (медики; юристы; «компьютерщики»; филологи, преподаватели русского языка как иностранного и преподаватели-философы), несколько блоков речи студентов (филологов и нефилологов), а также четыре блока интерферированной русской речи носителей других языков: американцев, китайцев, франкофонов и голландцев (см. подробнее об исследованиях на материале русской речи иностранцев: [Метлова 2013; Казак 2015; Зайдес 2016, 2017; Замковец 2018; Чэн Чэнь 2018]). Всего в составе САТ сегодня около 700 текстов и около 50 часов звучания.

С учетом состава информантов монологи в составе САТ можно разделить на следующие группы:

- 150 монологов медиков;
- 201 монолог юристов;
- 172 монолога студентов (филологи и нефилологи);
- 32 монолога преподавателей РКИ;
- 28 монологов «компьютерщиков»;
- 12 монологов преподавателей-философов;
- 12 монологов информантов разных профессий;



• 62 монолога изучающих русский язык инофонов (8 текстов, записанных от носителей американского варианта английского языка, 8 — от франкофонов, 30 — от носителей китайского языка, 16 — от носителей нидерландского языка).

Данный материал по типам спонтанных монологов распределяется следующим образом:

- чтение предтекста — 72 монолога (36 монологов-чтений сюжетного и 36 — несюжетного предтекстов);
- пересказ первичного текста — 204 монолога (101 пересказ сюжетного предтекста и 103 — несюжетного);
- описание изображения — 254 монолога (136 описаний сюжетного изображения и 118 — несюжетного);
- рассказ — 139 монологов.

Общее количество информантов, речь которых вошла в САТ, — 212 человек. Состав информантов с учетом их социальных и психологических характеристик:

- 1) по гендерному признаку: 89 мужчин, 123 женщины;
- 2) по возрасту:
  - 1 группа, поздняя юность (18–24 года) — 85 информантов;
  - 2 группа, ранняя зрелость (25–34 года) — 41 информант;
  - 3 группа, средняя зрелость (35–44 года) — 39 информантов;
  - 4 группа, поздняя зрелость (45–54 года) — 13 информантов;
- 3) по уровню речевой компетенции:
  - 43 информанта с высоким УРК;
  - 54 информанта со средним УРК;
  - 73 информанта с низким УРК;
- 4) по профессиональной принадлежности:
  - 95 студентов (носителей русского языка и иностранцев);
  - 48 юристов;
  - 30 медиков;
  - 7 преподавателей РКИ;
  - 14 «компьютерщиков»;
  - 6 преподавателей-философов;
  - 12 информантов разных профессий;
- 5) по психологическим характеристикам (представлены не для всех групп информантов; общее количество говорящих, психологические характеристики которых известны, — 97 человек):
  - 38 экстравертов;
  - 31 амбиверт;
  - 28 интровертов.

Тексты для корпуса записывались в разное время, разными собирателями и часто в разных целях (как правило, в рамках той или иной исследовательской работы, от студенческой курсовой до кандидатской диссертации), первые записи были сделаны еще до рождения самой идеи корпуса [Степихов 2005; Бродт 2007], поэтому материал САТ не вполне однороден и далеко не все его блоки отвечают тем

требованиям, которые позже были сформулированы. Требования эти (принципы создания корпуса) в общем и целом таковы.

Коллекция САТ предполагает *балансировку* и материала, и состава информантов:

- *собственно лингвистическая балансировка*: все тексты записывались по специально разработанной программе, в рамках разных коммуникативных (речевых) сценариев: чтение и пересказ сюжетного и несюжетного текстов, описание сюжетного и несюжетного изображения, рассказ на заданную тему; для некоторых конкретных исследований записывались тексты только одного типа, которые и подвергались углубленному анализу: *чтение* [Сапунова 2009], *пересказ* [Куканова 2009; Малиновская 2012], *описание* [Филиппова 2010], *рассказ* [Иванова 2011; Казак 2015; Зайдес 2016];

- *социолингвистическая балансировка*: информанты подбираются с учетом их социальных характеристик, что фиксируется в специальных анкетах затем попадает в базу данных корпуса;

- *психолингвистическая балансировка*: информанты обладают разными психологическими характеристиками (психотип и в ряде исследований — уровень невротизма/нейротизма), что устанавливалось в ходе специального психологического тестирования (тест Г. Айзенка).

Лингвистическая (подбор текстов-стимулов) и социолингвистическая (подбор информантов) балансировка проводились до записи, психологическое тестирование участников эксперимента — после записи. В результате полной психолингвистической балансировки не получилось, но в целом материал САТ дает некоторую возможность его анализа с учетом психологических факторов (см., например: [Куканова 2009; Хан 2013; Зайдес 2016]).

За годы существования САТ был проанализирован и продолжает исследоваться в самых разных направлениях. Можно сказать, что на его основе проводится фундаментальное многоуровневое (фонетика, грамматика, лексика, структура дискурса) описание звучащей монологической речи, которая является основной формой человеческого общения. Фундаментальность исследований, проводимых на материале САТ, определяется до некоторой степени новым подходом к анализу монологической речи.

Прежде всего в научный оборот вводится ряд новых понятий: *профессиональное/непрофессиональное отношение говорящего к языку/речи*, *степень естественности спонтанности* устной речи, *степень лингвистической мотивированности* устного монологического текста, *уровень речевой компетенции* (УРК) говорящего. В целом ряде работ авторского коллектива предлагается исследование этих понятий в широком междисциплинарном контексте (с привлечением методов психолингвистики, социолингвистики, дискурсивного анализа, акустического инструментального анализа). Пилотный анализ показал, что эти параметры действительно являются дистинктивно значимыми для анализа спонтанных монологов разных типов (см., например, трехтомную коллективную монографию: [Звуковой корпус... 2013, 2014, 2015]). При аннотировании корпуса САТ каждый звучащий текст был атрибутирован с точки зрения степени его спонтанности и лингвистической

мотивированности. База данных САТ содержит информацию обо всех информантах, участвовавших в записи.

Другая важная научная задача, решению которой служит корпус САТ, состоит в исследовании степени вариативности порождаемых монологических текстов в зависимости от характера стимула: визуального или текстового. В частности, предварительные наблюдения показывают, что сюжетность/несюжетность предтекста или описываемого изображения, а также степень знакомства говорящего с темой свободного монолога, заданного вопросом экспериментатора (исходным стимулом), оказывают влияние на выбор говорящим тех или иных речевых средств и в целом на лингвистическую природу вторичного текста (см. многочисленные наблюдения в: [Звуковой корпус... 2013, 2014, 2015]). Этот аспект порождения устных монологических текстов еще мало изучен.

Социолингвистическая и психолингвистическая вариативность речевой продукции в жанре монолога — еще один важный аспект исследования корпуса САТ. Предлагаемая разработчиками САТ методика позволяет анализировать речевую продукцию говорящих с разными психологическими и социологическими характеристиками, полученную фактически в одинаковых коммуникативных ситуациях, что дает основания для ее научного анализа в сопоставительном аспекте.

Спонтанная устная речь представляет собой весьма сложный для научного анализа объект. По форме выражения — по крайней мере, внешне — она значительно менее упорядочена и структурирована, чем письменная, выглядит более хаотичной и непредсказуемой, что тем не менее не мешает ей выполнять свои основные функции. Поэтому в настоящее время лингвисты все чаще говорят о принципиальном отличии *грамматики* спонтанной устной *речи* от существующих академических грамматик, описывающих стандартный, кодифицированный, литературный язык в его письменной форме. Правила порождения спонтанной речи и ее грамматика до сих пор системно не описаны, приблизиться к решению этой задачи может помочь анализ материала САТ.

Более того, наблюдения показывают, что не только грамматика речи обладает существенным своеобразием по сравнению с письменными формами языка, но и *грамматика* устного *монолога* в определенных аспектах принципиально отличается от *грамматики* устного *диалога* на фонетическом, морфологическом, синтаксическом, лексическом и дискурсивном уровнях (см., например: [Богданова 2012; Завадская 2018]). В первую очередь эти различия касаются принципов текстообразования монологической спонтанной речи и описательного дискурса, которые непосредственно влияют на синтаксическую организацию текста. Такой аспект анализа становится возможен при сравнении материала САТ (монологические тексты) и корпуса повседневной русской речи «Один речевой день» (ОРД; в основном диалоги и полилоги) (см. о нем, например: [Asinovsky et al. 2009; Русский язык повседневного общения... 2016; Bogdanova-Beglarian et al. 2016a, b, 2017; Богданова-Бегларян и др. 2017]).

Материалы САТ становятся доступными широкому кругу читателей и пользователей двумя путями.

Первый — через серию публикаций текстов устной спонтанной речи, которая начала выходить с 2008 г.: в первых трех выпусках этой серии содержатся спонтанные тексты, записанные от информантов-юристов, — свободные рассказы [Русская спонтанная речь 2008], монологи-репродуктивы [Русская спонтанная речь 2010] и монологи-описания [Русская спонтанная речь 2011]. Сборник текстов, записанных от информантов-медиков (монологи всех типов), готовится к печати в Германии [Русская спонтанная речь 2018]. Отличительной чертой этих публикаций является наличие, помимо, собственно транскриптов всех текстов и метаданных информантов, лексических материалов: частотных списков, как общих, так и упорядоченных по типам текстов.

Второй путь, которым материалы САТ могут попасть в руки широкого круга читателей и пользователей, — их постепенная передача в Национальный корпус русского языка (его устный подкорпус). Эта работа проводится в настоящее время. Однако предпринимаемых усилий по «сближению» материалов САТ с их потенциальным пользователем пока явно недостаточно. Поэтому в перспективах работы авторского коллектива — разработка системы онлайн-доступа к звуковому корпусу русской монологической речи.

Возвращаясь к существующим исследованиям спонтанной речи, еще раз повторим, что центральным объектом внимания лингвистов по сей день остается русская разговорная речь в форме бытового диалога: в работах исследователей оказывается затронутой специфика спонтанной речи, которая раскрывается прежде всего на материале записанных и расшифрованных диалогов. Бытовые спонтанные монологи как исследовательский материал представлены явно недостаточно.

Традиционно для сбора речевых данных используется ряд методик. Наиболее распространенными из них являются устное интервью, разнообразные эксперименты, включенное наблюдение.

Из отечественных работ следует отметить, например, записи, сделанные М. В. Китайгородской и Н. Н. Розановой и нашедшие отражение в книгах «Речь москвичей: коммуникативно-культурологический аспект» (1999) и «Языковое существование современного горожанина: На материале языка Москвы» (2010). В сборнике «Жанр интервью: Особенности русской устной речи в Финляндии и Санкт-Петербурге» (2004) анализировался корпус текстов полуструктурированных интервью, записанных российско-финским коллективом социологов (И. И. Травин, Е. М. Порецкина, Т. Пиирайнен, Ю. Симпура и др.) от жителей Санкт-Петербурга в 1999–2003 гг., а также тексты, записанные от русскоговорящих жителей Финляндии. Необходимо упомянуть также о записях русской речи коми-пермяков и татар, собранных под руководством Т. И. Ерофеевой (2007, 2010, 2012). Материалы русской речевой коммуникации (монологи, диалоги, полилоги) представлены также в изданиях «Живая речь уральского города: Тексты» (1995), «Город в зеркале своего языка» (1996), «Воспоминания работницы М. Н. Колтаковой “Как я прожила жизнь”» (1997), «Разговорная речь носителей массовой городской культуры (на материале г. Омска)» (2007) и многих других.

Разнообразные материалы устной монологической речи представлены в проекте А. А. Кибрика, В. И. Подлесской и др. «Рассказы о свидениях и другие корпуса звучащей речи».

В целом подавляющее большинство коллекций русских устных спонтанных монологов собрано в рамках единого коммуникативного сценария, чаще всего — «нарративы на заданную тему».

Если обратиться к западному опыту исследования устных спонтанных монологов, то он также весьма разнообразен и опирается на многочисленные корпуса и лингвистические базы данных. В основном это материалы интервью, в том числе записанные в лабораторных условиях, например BACKBONE (записи интервью с нативными носителями английского, французского, немецкого, польского, испанского, турецкого и ненативными носителями английского языка), SLX Corpus of Classic Sociolinguistic Interviews (на материале английского языка), TAUS (фиксирует речь жителей Осло, собранную в 1970-х гг.), ELISA (включает интервью носителей английского языка как второго) и др. По разнообразию методик получения речевого материала можно выделить Lincoln Laboratory Speech Enhancement Corpus (LLSEC), собранный в рамках целого ряда сценариев, моделирующих широкую вариативность условий речи. Крупные корпуса устной речи зачастую включают материалы как устных диалогов, так и монологов: это и Lancaster/IBM Spoken English Corpus (SEC), и The Wellington Corpus of Spoken New Zealand English, и ANDOSL (Australian National Database of Spoken Language), и ELFA (English as a Lingua Franca in Academic Settings), и MICASE (Michigan Corpus of Academic Spoken English).

В том числе можно особо выделить направление исследования устных академических монологов — в основном на материале английского языка (см. работы S. E. Thompson о структуре текста, интонации академических лекций, работы J. Rendle-Short о паузах в академическом монологе, работы M. Cribb о дискурсивных маркерах, интонации и других явлениях в текстах ненативных носителей английского языка и др.), направление исследования детских монологов (работы L. J. Harriet).

Важность любого научного ресурса для исследователей определяется не в последней степени его доступностью для пользователей. К сожалению, онлайн-ресурсов, представляющих материалы русской устной монологической речи, в настоящее время несоизмеримо меньше, чем посвященных этой теме научных работ.

Параллельно со сбором речевого материала коллекции САТ проводится анализ монологической русской речи. В результате был получен целый ряд наблюдений, связанных со спецификой устной спонтанной, в первую очередь монологической, русской речи.

В *общетеоретическом аспекте* обсуждалось, в частности, само понятие спонтанности речи и было предложено два подхода к его трактовке. Прежде всего, спонтанность может пониматься фактически как синоним неподготовленности, и спонтанной в этом случае признается речь неподготовленная, непринужденная и осуществляемая в неофициальной обстановке. Большая часть

проведенных исследований основана именно на таком понимании спонтанности и спонтанной речи. В другой трактовке спонтанность отражает не только и не столько предварительную неподготовленность речи, сколько плохую ее согласованность с конкретными условиями речевой коммуникации, которые порождаются и определяются не в последнюю очередь личностью говорящего. Спонтанность в таком понимании является маркером не первичности, а несогласованности мысли и речи с условиями коммуникации. В материалах собранной коллекции монологических текстов можно найти примеры проявления спонтанности и такого рода.

Столь же общетеоретическими можно считать и поиск единиц описания устной речи, анализ способов сокращения и приращения устного текста, установление градаций (степеней) естественности спонтанной речи и некоторые другие вопросы, включая принципы формирования речевой коллекции и разработку способов балансировки материала как в лингвистическом, так и в психо- и социолингвистическом отношении. Поворот от исследования по преимуществу лабораторной устной речи к более естественным ее типам, порождаемым в условиях повседневной коммуникации, с учетом всех возможных корреляций лингвистических параметров материала с характеристиками говорящего и типом спонтанного монолога можно считать отличительной чертой предлагаемого подхода к анализу устной речи.

На *фонетическом уровне* анализировались паузы хезитации разного типа, являющиеся неотъемлемым свойством спонтанного речепорождения и возникающие во всех видах речи и у всех говорящих; фонетические ошибки, в первую очередь в чтении; а также темп речи — разных говорящих и в разных типах речи.

*Лексический аспект* исследования представлен, в частности, анализом лексических трансформаций исходных текстов при пересказе и детальной проработкой всех типов эзоединиц, противопоставленных эндоединицам репродуктивов; а также анализом лексических ошибок при чтении. Кроме того, лексический аспект исследования представлен анализом новых, специфичных явлений, свойственных именно повседневной спонтанной речи: новых слов, новых значений или новых коннотаций старых слов, особенностей сочетаемости тех или иных лексических единиц или даже новых идиом, отличающих нашу спонтанную речь; наконец, анализом проявлений внутриязыковой интерференции — в первую очередь между профессиональной и бытовой речью говорящего индивида. Такие проявления, как показали исследования, обнаруживаются прежде всего именно на лексическом уровне.

*Морфологический уровень* представлен анализом грамматических ошибок в чтении, номинативной лексики в монологах-описаниях, функционирования в разных видах текстов и в речи разных групп информантов различных глагольных форм (инфинитива, причастия и деепричастия), а также некоторыми другими частными наблюдениями морфологического характера.

На *синтаксическом уровне* были описаны синтаксические трансформации исходных текстов при пересказе, структура предикативных единиц, длина текста

в «предложениях» и длина «предложения» в словах (также по преимуществу в репродуктиве, в сравнении с исходными текстами-стимулами) (о способе вычленения в устном тексте единиц, соответствующих предложению, см.: [Bogdanova-Beglarian 2017]). Кроме того, были описаны вставные конструкции и способы передачи чужой речи в разных типах текстов, дана общая синтаксическая характеристика речи говорящих с разным уровнем речевой компетенции, описано функционирование в спонтанных монологах изолированного номинатива.

На *дискурсивном уровне* проанализированы повторы, перебивы, случаи самокоррекции, коммуникативные установки и стратегии говорящего в разных типах текстов, элементы метакоммуникации, отличающие именно спонтанную речь и зависящие как от типа текста, так и от характеристик говорящего. Особенно подробно описаны сценарный и композиционный уровни построения устного монолога-описания и специальные конструкции описательного дискурса.

Лингвистический характер полученных наблюдений (см. [Звуковой корпус... 2013, 2014, 2015]) сопряжен с психо- и социолингвистическим аспектом описания: почти во всех случаях сделана попытка установления корреляции между лингвистическими и экстралингвистическими параметрами материала. В паралингвистическом аспекте были рассмотрены функции смеха и вздохов в спонтанной монологической речи. Прагматический аспект исследования представлен серией методических разработок с использованием материалов Звукового корпуса и предназначенных для преподавания русского языка как иностранного; а также попыткой описания идиолекта — речи конкретной языковой личности, записанной многократно во всех типах коммуникативных ситуаций.

Уже из этого перечня видно, что научным коллективом исследователей — разработчиков САТ получены результаты действительно многоаспектного описания русской спонтанной речи, в том числе по четырем основным коммуникативным сценариям нашей повседневной жизни — чтению, пересказу, описанию изображения и свободному нарративу. Однако большинство результатов анализа спонтанных монологов было получено лишь на материале изолированных компонентов коллекции САТ, системный анализ и описание монологической речи на всем корпусе данных никогда ранее не проводился — препятствием этому была техническая разнородность элементов коллекции. Поэтому перспективной задачей коллектива ставится не только унификация форматов и стандартов коллекции САТ, но и формирование на ее базе уникальной информационно-исследовательской системы анализа монологической речи, на базе которой можно будет не только проверить на представительном речевом материале выдвинутые ранее научные гипотезы, но и проводить ряд новых исследований, в том числе квантитативных.

В заключение скажем, что результаты таких исследований важны как для собственно лингвистики, так и для ряда смежных научных дисциплин: акустики речи, прагматики, семантики, психолингвистики, социолингвистики, когнитивной лингвистики, исследований дискурса, антропологии, культурологии, этнолингвистики.



### *Литература*

Богданова Н. В. О корпусе текстов живой речи: новые поступления и первые результаты исследования // Компьютерная лингвистика и интеллектуальные технологии. Вып. 9 (16). По материалам международной конференции «Диалог» (2010) / гл. ред. А. Е. Кибрик. М.: РГГУ, 2010. С. 35–40.

Богданова Н. В. Метакоммуникация в устной спонтанной речи (диалог vs. монолог) // Коммуникация в социально-гуманитарном знании, экономике, образовании. Материалы III Международной научно-практической конференции. 29–31 марта 2012 г., Минск. Научное электронное издание. Минск, 2012. С. 330–331.

Богданова Н. В., Бродт И. С., Куканова В. В., Павлова О. В., Сапунова Е. М., Филиппова Н. С. О «корпусе» текстов живой речи: принципы формирования и возможности описания // Компьютерная лингвистика и интеллектуальные технологии. Вып. 7 (14). По материалам ежегодной международной конференции «Диалог» (2008) / гл. ред. А. Е. Кибрик. М.: РГГУ, 2008. С. 57–61.

Богданова-Бегларян Н. В., Шерстинова Т. Ю., Зайдес К. Д. Корпус «Сбалансированная Аннотированная Текстотека»: методика многоуровневого анализа русской монологической речи // Анализ разговорной русской речи (АР<sup>3</sup>-2017): Труды седьмого междисциплинарного семинара / науч. ред. Д. А. Кочаров, П. А. Скрелин. СПб.: Политехника-принт, 2017а. С. 8–13.

Богданова-Бегларян Н. В., Шерстинова Т. Ю., Блинова О. В., Мартыненко Г. Я. Корпус «Один речевой день» в исследованиях социолингвистической вариативности русской разговорной речи // Анализ разговорной русской речи (АР<sup>3</sup>-2017): Труды седьмого междисциплинарного семинара / науч. ред. Д. А. Кочаров, П. А. Скрелин. СПб.: Политехника-принт, 2017б. С. 14–20.

Бродт И. С. Спонтанный монолог в лингвистическом и социолингвистическом аспектах (на материале текстов разного типа): дис. ... канд. филол. наук. СПб., 2007. 289 с. (машинопись).

Завадская Ю. О. Оговорки в русской устной спонтанной речи: монолог vs. диалог. Курсовая работа. СПб., 2018. 78 с.

Зайдес К. Д. Метакоммуникативные вставки в русской устной спонтанной речи на родном и неродном языке // Коммуникативные исследования. 2016. №3 (9). С. 19–35.

Зайдес К. Д. Типология метакоммуникативных единиц русской спонтанной монологической речи // Компьютерная лингвистика и вычислительные онтологии. Вып. 1 (Труды XX Международной объединенной научной конференции «Интернет и современное общество», IMS-2017, Санкт-Петербург, 21–23 июня 2017 г. Сб. научных статей). СПб.: Университет ИТМО, 2017. С. 46–56.

Замковец К. С. Поисковые хезитативы в русской спонтанной речи носителей нидерландского языка // Материалы международного молодежного научного форума «Ломоносов-2018» / отв. ред. И. А. Алешковский, А. В. Андриянов, Е. А. Антипов. Москва, МГУ им. М. В. Ломоносова, 9–13 апреля 2018 г. М.: МГУ, 2018 [Электронный ресурс].



*Звуковой корпус* как материал для анализа русской речи: коллективная монография. Ч. 1: Чтение. Пересказ. Описание / отв. ред. Н.В. Богданова-Бегларян. СПб.: Филологический ф-т СПбГУ, 2013. 532 с.

*Звуковой корпус* как материал для анализа русской речи: коллективная монография. Ч. 2: Теоретические и практические аспекты анализа. Т. 1: О некоторых особенностях устной спонтанной речи разного типа. Звуковой корпус как материал для преподавания русского языка иностранной аудитории / отв. ред. Н.В. Богданова-Бегларян. СПб.: Филологический ф-т СПбГУ, 2014. 396 с.

*Звуковой корпус* как материал для анализа русской речи: коллективная монография. Ч. 2: Теоретические и практические аспекты анализа. Т. 2: Звуковой корпус как материал для новых лексикографических проектов / отв. ред. Н.В. Богданова-Бегларян. СПб.: Филологический ф-т СПбГУ, 2015. 364 с.

Иванова О. А. Специфика бытовой речи различных профессионально ориентированных групп: дис. ... маг. лингв. СПб., 2011. 112 с. (машинопись).

Казак М. В. Паузы хезитации в спонтанной речи на родном и неродном языках (на материале речи франкофонов). Saarbrücken: Lambert Academic Publishing, 2015. 74 с.

Куканова В. В. Лингвистический анализ репродуцированных текстов (на материале звукового корпуса русской речи юристов): дис. ... канд. филол. наук. СПб., 2009. 328 с. (машинопись).

Малиновская А. И. Репродуктив как объект многоаспектного анализа (на материале Звукового корпуса русского языка): дис. ... маг. лингв. СПб., 2012. 103 с. (машинопись).

Метлова В. А. Темп речи и паузы хезитации в речи на родном и неродном языках: монография. Saarbrücken: Palmarium Academic Publishing, 2013. 79 с.

*Русская спонтанная речь*. Свободные монологи-рассказы на заданную тему. Тексты. Лексические материалы / сост. В.В. Куканова; отв. ред. и авт. предисл. Н.В. Богданова. СПб.: Филологический ф-т СПбГУ, 2008. 208 с.

*Русская спонтанная речь*. Монологи-репродуктивы. Тексты. Лексические материалы / сост. В.В. Куканова; отв. ред. и авт. предисл. Н.В. Богданова. СПб.: Филологический ф-т СПбГУ, 2010. 132 с.

*Русская спонтанная речь*. Монологи-описания. Тексты. Лексические материалы / сост. В.В. Куканова; отв. ред. и авт. предисл. Н.В. Богданова. СПб.: Филологический ф-т СПбГУ, 2011. 140 с.

*Русская спонтанная речь*. Спонтанные монологи разных типов. Тексты. Лексические материалы (CD) / сост. Н.В. Богданова-Бегларян, И.С. Бродт; отв. ред. М. Краузе // Бюллетень Фонетического Фонда. Бохум (Германия), 2018. 100 с. (в печати).

*Русский язык повседневного общения*: особенности функционирования в разных социальных группах: коллективная монография / отв. ред. Н.В. Богданова-Бегларян. СПб.: ЛАЙКА, 2016. 244 с.

Сапунова Е. М. Неподготовленное чтение как вид речевой деятельности и тип устного спонтанного монолога (на материале русского языка): дис. ... канд. филол. наук. СПб., 2009. 237 с. (машинопись).

Степихов А. А. Соотношение синтаксического и интонационного членения в спонтанном монологе: дис. ... канд. филол. наук. СПб., 2005. 197 с. (машинопись).

Филиппова Н. С. Принципы построения устного описательного дискурса (на материале русской спонтанной речи): дис. ... канд. филол. наук. СПб., 2010. 220 с. (машинопись).

Хан Н. А. Устные спонтанные монологи разного типа в коммуникативно-дискурсивном аспекте (на материале Звукового корпуса русского языка): дис. ... канд. филол. наук. СПб., 2013. 277 с. (машинопись).

Чэн Чэнь. Хезитации в русской устной речи носителей китайского языка: дис. ... канд. филол. наук. СПб., 2018. 205 с. (машинопись).

Asinovsky, A., Bogdanova, N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD Speech Corpus of Russian Everyday Communication “One Speaker’s Day”: Creation Principles and Annotation // Text, Speech and Dialogue. 12th International Conference, TSD 2009. Proceedings / eds. V. Matoušek, P. Mautner. Pilsen, Czech Republic, September 2009. P. 250–257.

Bogdanova-Beglarian N. V. In Search of Phrase Boundaries in Spontaneous Speech // *SPECOM 2017. Lecture Notes in Artificial Intelligence*, LNAI. Vol. 10458. Springer, Switzerland, 2017. P. 456–463.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Baeva E., Martynenko G., Ryko A. Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech // Speech and Computer. 18<sup>th</sup> International Conference, SPECOM 2016. Budapest, Hungary, August, 23–27, 2016a. Proceedings / eds. A. Ronzhin, R. Potapova, G. Németh. P. 659–666.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G. An Exploratory Study on Sociolinguistic Variation of Spoken Russian // Speech and Computer. 18<sup>th</sup> International Conference, SPECOM 2016. Budapest, Hungary, August, 23–27, 2016b. Proceedings / eds. A. Ronzhin, R. Potapova, G. Németh. P. 100–107.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G. Linguistic Features and Sociolinguistic Variability in Everyday Spoken Russian // *SPECOM 2017. Lecture Notes in Artificial Intelligence*, LNAI. Vol. 10458. Springer, Switzerland, 2017. P. 503–511.

**N. V. Bogdanova-Beglarian, O. V. Blinova, K. D. Zaides, T. Ju. Sherstinova**  
*Saint-Petersburg State University (Russia, Saint-Petersburg)*  
*n.bogdanova@spbu.ru, o.blinova@spbu.ru, kristina.zaides@student.spbu.ru,*  
*t.sherstinova@spbu.ru*

## **CORPUS “BALANCED ANNOTATED TEXT COLLECTION (TEXTOTEC)” (SAT): STUDYING THE SPECIFICITY OF RUSSIAN MONOLOGICAL SPEECH**

The article represents one of the Russian speech corpora: a collection of monologic texts, known as the “Balanced Annotated Text Collection (Textotec)” (SAT). This corpus

was being assembled in St. Petersburg State University for more than 20 years, using the author's (N. V. Bogdanova-Beglarian's) methodology of data collection, which involves a fairly strict set of experimental procedures. SAT is designed to study various types of spontaneous monologues (reading, retelling, image description, story on the topic) and it contains texts recorded from five professionally-oriented groups of native speakers (medical doctors, lawyers, computer specialists, philologists, teachers of Russian as a foreign language, and teachers-philosophers), several blocks of students' speech (philologists and non-philologists), as well as four blocks of the interfered Russian speech of native speakers of other languages: Americans, Chinese, Francophone and Dutch. In total, there are about 700 texts in the SAT and about 50 hours of sound recording. In the article, against the background of other Russian-speaking and foreign speaking corpora, a description of this linguistic resource is given, the main topics developed on its material are marked, and prospects for continuing work are outlined.

**Keywords:** modern Russian language, oral monologic speech, speech corpus, natural language processing, database, linguistic experiment, reading, description of the image, retelling of the text (reproductive), spontaneous monologue, sociolinguistics, psycholinguistics.

### References

Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD Speech Corpus of Russian Everyday Communication "One Speaker's Day": Creation Principles and Annotation. *Text, Speech and Dialogue. 12th International Conference, TSD 2009. Proceedings*. Eds. V. Matoušek, P. Mautner. Pilsen, Czech Republic, September 2009, pp. 250–257.

Bogdanova N. V. [Metacommunication in Oral Spontaneous Speech (Dialogue vs. Monologue)]. *Kommunikatsiya v sotsial'no-gumanitarnom znanii, ekonomike, obrazovanii. Materialy III Mezhdunarodnoi nauchno-prakticheskoi konferentsii*. 29–31 marta 2012 g. [Communication in Socio-Humanitarian Knowledge, Economics, Education. Materials of the III International Scientific and Practical Conference. 29–31 March, 2012]. Minsk, Scientific electronic publication, 2012, pp. 330–331. (In Russ.)

Bogdanova N. V. [On the Corpus of Texts of Living Speech: New Acquisitions and First Results of the Research]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Po materialam mezhdunarodnoi konferentsii "Dialog" (2010)* [Computer Linguistics and Intellectual Technologies. Based on the Materials of the International Conference "Dialogue" (2010)]. Iss. 9 (16). Moscow, 2010, pp. 35–40. (In Russ.)

Bogdanova N. V., Brodt I. S., Kukanova V. V., Pavlova O. V., Sapunova E. M., Filipova N. S. [On the Corpus of Texts of Living Speech: the Principles of Formation and the Possibility of Description]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Po materialam ezhegodnoi mezhdunarodnoi konferentsii "Dialog" (2008)* [Computer Linguistics and Intellectual Technologies. Based on the Materials of the Annual International Conference "Dialogue" (2008)]. Iss. 7 (14). Moscow, 2008, pp. 57–61. (In Russ.)

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Baeva E., Martynenko G., Ryko A. [Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech]. *Speech and Computer. 18<sup>th</sup> International Conference, SPECOM 2016*. Budapest, Hungary, August, 23–27, 2016a. Proceedings. Eds. A. Ronzhin, R. Potapova, G. Németh, pp. 659–666.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G. [An Exploratory Study on Sociolinguistic Variation of Spoken Russian]. *Speech and Computer. 18<sup>th</sup> International Conference, SPECOM 2016*. Budapest, Hungary, August, 23–27, 2016b. Proceedings. Eds. A. Ronzhin, R. Potapova, G. Németh, pp. 100–107.

Bogdanova-Beglarian N., Sherstinova T., Blinova O., Martynenko G. [Linguistic Features and Sociolinguistic Variability in Everyday Spoken Russian]. *SPECOM 2017. Lecture Notes in Artificial Intelligence, LNAI*, vol. 10458. Springer, Switzerland, 2017, pp. 503–511.

Bogdanova-Beglarian N. V. (ed.) [*Everyday Russian Language in Different Social Groups*. Collective. Monograph]. *Russkii yazyk povsednevnogo obshcheniya: osobennosti funktsionirovaniya v raznykh sotsial'nykh gruppakh. Kollektivnaya monografiya*. St. Petersburg, 2016. 244 p. (In Russ.).

Bogdanova-Beglarian N. V. (ed.) *Zvukovoi korpus kak material dlya analiza russkoi rechi. Kollektivnaya monografiya. Ch. 1. Chtenie. Pereskaz. Opisanie* [Speech Corpus as a Base for Analysis of Russian Speech. Collective Monograph. Part 1. Reading. Retelling. Description]. St. Petersburg, 2013. 532 p. (In Russ.).

Bogdanova-Beglarian N. V. (ed.) *Zvukovoi korpus kak material dlya analiza russkoi rechi. Kollektivnaya monografiya. Ch. 2. Teoreticheskie i prakticheskie aspekty analiza. T. 1. O nekotorykh osobennostyakh ustnoi spontannoi rechi raznogo tipa. Zvukovoi korpus kak material dlya prepodavaniya russkogo yazyka v inostrannoi auditoria* [Speech Corpus as a Base for Analysis of Russian Speech. Collective Monograph. Part 2. Theoretical and Practical Aspects of Analysis. Vol. 1. On Some Features of Oral Spontaneous Speech of Different Types. Speech Corpus as a Material for Teaching Russian in a Foreign Audience]. St. Petersburg, 2014. 396 p. (In Russ.).

Bogdanova-Beglarian N. V. (ed.) *Zvukovoi korpus kak material dlya analiza russkoi rechi. Kollektivnaya monografiya. Ch. 2. Teoreticheskie i prakticheskie aspekty analiza. T. 2. Zvukovoi korpus kak material dlya novykh leksikograficheskikh proektov* [Speech Corpus as a Base for Analysis of Russian Speech. Collective Monograph. Part 2. Theoretical and Practical Aspects of Analysis. Vol. 2. Speech Corpus as a Material for New Lexicographic Projects]. St. Petersburg, 2015. 364 p. (In Russ.).

Bogdanova-Beglarian N. V. [In Search of Phrase Boundaries in Spontaneous Speech]. *SPECOM 2017. Lecture Notes in Artificial Intelligence, LNAI*, vol. 10458. Springer, Switzerland, 2017, pp. 456–463.

Bogdanova-Beglarian N. V., Brodt I. S. (orig.), Krause M. (ed.) [Russian Spontaneous Speech. Spontaneous Monologues of Different Types. Texts. Lexical Materials (CD)]. *Byulleten' Foneticheskogo Fonda* [Bulletin of the Phonetic Fund]. Bochum (Germany), 2018. 100 p. (In Russ.) (In Print.)

Bogdanova-Beglarian N. V., Sherstinova T. Ju., Blinova O. V., Martynenko G. Ja. [Corpus “One Speaker’s Day” in Studies of Sociolinguistic Variability of Russian Colloquial Speech]. *Analiz razgovornoj russkoirechi (AR<sup>3</sup>-2017): Trudy sed'mogo mezhdistsiplinarnogo seminar* [Analysis of Spoken Russian Speech (AR<sup>3</sup>-2017): Proceedings of the 7<sup>th</sup> Interdisciplinary Seminar]. St. Petersburg, 2017b, pp. 14–20. (In Russ.)

Bogdanova-Beglarian N. V., Sherstinova T. Ju., Zaides K. D. [Corpus “Balanced Annotated Text Library”: Methodology Multi-level Analysis of the Russian Monologue Speech]. *Analiz razgovornoj russkoi rechi (AR<sup>3</sup>-2017): Trudy sed'mogo mezhdistsiplinarnogo seminar* [Analysis of Spoken Russian Speech (AR<sup>3</sup>-2017): Proceedings of the 7<sup>th</sup> Interdisciplinary Seminar]. St. Petersburg, 2017a, pp. 8–13. (In Russ.)

Brodt I. S. *Spontannyi monolog v lingvisticheskom i sotsiolingvisticheskom aspektakh (na material tekstov raznogo tipa)*. Dis. kand. filol. nauk [Spontaneous Monologue in Linguistic and Sociolinguistic Aspects (Based on the Material of Texts of Different Types)]. Dis. cand. philol. sci.]. St. Petersburg, 2007. 289 p. (In Russ.)

Cheng Chen. *Khezitatsii v russkoi ustnoi rechi nositelei kitaiskogo yazyka*. Dis. kand. filol. nauk [Hesitations in the Russian spoken language of Chinese speakers. Dis. cand. philol. sci.]. St. Petersburg, 2018. 205p. (In Russ.)

Filippova N. S. *Printsipy postroeniya ustnogo opisatel'nogo diskursa (na materiale russkoi spontannoi rechi)*. Dis. kand. filol. nauk [Principles of Constructing of Oral Descriptive Discourse (Based on Russian Spontaneous Speech)]. Dis. cand. philol. sci.]. St. Petersburg, 2010. 220 p. (In Russ.)

Ivanova O. A. *Spetsifika bytvoi rechi razlichnykh professional'no orientirovannykh grupp*. Dis. mag. lingv. [Specificity of Everyday Speech of Various Professionally Oriented Groups. Dis. mast. ling.]. St. Petersburg, 2011. 112 p. (In Russ.)

Kazak M. V. *Pauzy khezitatsii v spontannoi rechi na rodnom i nerodnom yazykakh (na material rechi frankofonov)* [Pauses of Hesitations in Spontaneous Speech in Native and Non-Native Languages (on the Material of the Speech of Francophone)]. Saarbrücken, Lambert Academic Publishing, 2015. 74 p. (In Russ.)

Khan N. A. *Ustnye spontannye monologi raznogo tipa v kommunikativno-diskursivnom aspekte (na material Zvukovogo korpusa russkogo yazyka)*. Dis. kand. filol. Nauk [Spoken Spontaneous Monologues of Various Types in Communicative and Discourse Aspects (Based on the Spoken Corpus of Russian Language)]. Dis. cand. philol. sci.]. St. Petersburg, 2013. 277p.

Kukanova V. V. (orig.), Bogdanova N. V. (ed.) *Russkaya spontannaya rech'. Monologi-opisaniya. Teksty. Leksicheskie materialy* [Russian Spontaneous Speech. Description Monologues. Texts. Lexical Materials]. St. Petersburg, 2011. 140 p. (In Russ.)

Kukanova V. V. (orig.), Bogdanova N. V. (ed.) *Russkaya spontannaya rech'. Svoobodnye monologi-rasskazy na zadannuyu temu. Teksty. Leksicheskie materialy* [Russian Spontaneous Speech. Free Monologues-Stories on a Topic. Texts. Lexical Materials]. St. Petersburg, 2008. 208 p. (In Russ.)

Kukanova V. V. (orig.), Bogdanova N. V. (ed.) *Russkaya spontannaya rech'. Monologi-reproduktivny. Teksty. Leksicheskie materialy* [Russian Spontaneous Speech. Monologues-Reproductives. Texts. Lexical Materials]. St. Petersburg, 2010. 132 p. (In Russ.)

Kukanova V. V. *Lingvisticheskii analiz reprodutsirovannykh tekstov (na material zvukovogo korpusa russkoi rechi yuristov)*. Dis. kand. filol. Nauk [Linguistic Analysis of Reproduced Texts (Based on the Material of the Russian speech corpus of Lawyers). Dis. cand. philol. sci.]. St. Petersburg, 2009. 328 p. (In Russ.)

Malinovskaya A. I. *Reproduktiv kak ob'ekt mnogoaspektного analiza (na materiale Zvukovogo korpusa russkogo yazyka)*. Dis. mag. lingv. [Reproductive as an Object of Multi-Aspect Analysis (Based on the Material of the Speech Corpus of Russian). Dis. mast. ling.]. St. Petersburg, 2012. 103 p. (In Russ.)

Metlova V. A. *Temp rechi i pauzy khezitatsii v rechi na rodnom I nerodnom yazykakh. Monografiya* [The Rate of Speech and Pause of the Hezation in Speech in Native and Non-Native Languages. Monograph]. Saarbrücken, Palmarium Academic Publishing, 2013. 79 p. (In Russ.)

Sapunova E. M. *Nepodgotovlennoe chtenie kak vid rechevoi deyatel'nosti i tip ustno-go spontannogo monologa (na materiale russkogo yazyka)*. Dis. kand. filol. nauk [Unprepared Reading as a Type of Speech Activity and Type of Oral Spontaneous Monologue (Based on the Material of Russian Language). Dis. cand. philol. sci.]. St. Petersburg, 2009. 237 p. (In Russ.)

Stepikhov A. A. *Sootnoshenie sintaksicheskogo i intonatsionного chleneniya v spontannom monologe*. Dis. kand. filol. Nauk [The Correlation of Syntactic and Intonation Partitioning in Spontaneous Monologues. Dis. cand. philol. sci.]. St. Petersburg, 2005. 197 p. (In Russ.)

Zaides K. D. [Metacommunication Inserts in Russian Oral Spontaneous Speech in Native and Non-Native Language]. *Kommunikativnye issledovaniya* [Communicative Research], 2016, no. 3 (9), pp. 19–35. (In Russ.)

Zaides K. D. [Typology of Metacommunicative Units of Russian Spontaneous Monologic Speech]. *Komp'yuternaya lingvistika i vychislitel'nye ontologii. Vyp. 1 (Trudy XX Mezhdunarodnoi ob'edinennoi nauchnoi konferentsii "Internet i sovremennoe obshchestvo", IMS-2017, Sankt-Peterburg, 21–23 iyunya 2017 g. Sb. nauchnykhstatei)* [Computer Linguistics and Computing Ontologies. Iss. 1 (Proceedings of the XX International Joint Scientific Conference "Internet and Contemporary Society", IMS-2017, St. Petersburg, June 21–23, 2017, Collection of Scientific Articles)]. St. Petersburg, 2018, pp. 46–56. (In Russ.)

Zamkovec K. S. [Search Hesitatives in the Russian Spontaneous Speech of Native Speakers of the Netherlands]. *Materialy mezhdunarodного molodezhного nauchного foruma "Lomonosov-2018". Moskva, MGU im. M. V. Lomonosova, 9–13 aprelya 2018 g.* [Proceedings of the International Youth Scientific Forum "Lomonosov-2018". Moscow, Moscow State University M. V. Lomonosov. April 9–13, 2018]. Moscow, 2018 [Electronic resource]. (In Russ.)

Zavadskaya Yu. O. *Ogovorki v russkoi ustnoi spontannoi rechi: monolog vs. dialog. Kursovaya rabota* [Bloopers in Russian Oral Spontaneous Speech: Monologue vs. Dialogue. Student course work]. St. Petersburg, 2018. 78 p. (In Russ.)



## II. НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА В НАУЧНЫХ ИССЛЕДОВАНИЯХ

**Г. И. Кустова**

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

*(Россия, Москва)*

*galinak03@gmail.com*

### ОТНОСИТЕЛЬНЫЕ НАРЕЧИЯ В КОНТЕКСТЕ ГЛАГОЛОВ И ПРИЛАГАТЕЛЬНЫХ (ПО ДАННЫМ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА)\*

Работа посвящена наречиям с субстантивными корнями, образованным от относительных отсубстантивных прилагательных: *аварийно, административно, амбулаторно, анатомически, денежно, коммерчески, композиционно, лекарственно, медицински, музыкально, научно, планоно, промышленно, профессионально, социальнo, сюжетно, территориально, технически, туристически, физически, финансово, химически, экономически, юридически* и т.п. Таких наречий, как свидетельствуют данные Национального корпуса русского языка (НКРЯ), немало, хотя не все они отражаются даже в электронных словарях, не говоря уже о бумажных. Они употребительны, в основном, в книжной и устной литературной речи. Относительные наречия рассматриваются с точки зрения их способности выражать валентности предикатных слов — глаголов и прилагательных. Относительные прилагательные могут выражать валентности субъекта, объекта, инструмента, а также употребляться в значении функционального пространства и аспекта. Относительные наречия при глаголах практически не обозначают актанты, исключение — средство (*обозначить **буквенно***) и аспект (***административно** и **георгафически** делиться на две части*). Относительные наречия в контексте прилагательных обозначают главным образом аспект (*экономически **привлекательный**, коммерчески **успешный***). Аспект (сфера) — это специфическая валентность интерпретационно-оценочных слов.

*Ключевые слова:* относительное прилагательное, относительное наречие, предикат, валентность

---

\* Исследование выполнено в рамках проекта РФФИ № 17-29-09154 офи\_м «Динамика языковой системы: корпусное исследование синхронной вариативности и диахронических изменений в текстах разных типов».



## 1. Материал и постановка задачи

Материалом настоящей работы являются наречия с субстантивными корнями, образованные от относительных отсубстантивных прилагательных: *аварийно, административно, амбулаторно, анатомически, буквенно, денежно, духовно, коммерчески, композиционно, лекарственно, медицински, музыкально, научно, планоно, практически, промышленно, профессионально, социально, стратегически, сюжетно, тактически, теоретически, территориально, технически, туристически, фабрично, физически, финансово, химически, экономически, энергетически, юридически* и т. п., которые мы также будем называть относительными. Относительные значения противопоставлены качественным. Важнейшей особенностью качественных значений является то, что они соотносятся с градуируемыми признаками (а также состояниями и процессами). Градуируемость проявляется, в частности, в сочетаемости с показателями степени [Кустова 2018а], ср. *довольно широко, не совсем честно, вполне дружески, совершенно абстрактно*. Соответственно, слова с относительным значением (т. е. обозначающие неградуируемый признак) — относительные прилагательные и наречия — такой сочетаемости не имеют, ср.: *\*очень территориальный, \*довольно медицински, \*вполне туристически*.

Сразу оговоримся, что относительные прилагательные и наречия употребляются или могут быть употреблены в качественных значениях, ср.: *музыкальное сопровождение* (отн.) vs. *очень музыкальный ребенок* (кач.); *туристический сезон* (отн.) vs. *Город настолько туристический, что местных жителей почти не видно* (кач.); *научно обоснованные выводы* (отн.) vs. *Ты как-то слишком научно рассуждаешь* (кач.); *профессионально ориентированная подготовка* (отн.) vs. *Довольно профессионально выполнил прыжок* (кач.). Однако в данной работе нас интересуют именно относительные значения наречий на *-о, -и* с субстантивными корнями, т. е. значения в контекстах типа *финансово самостоятельный* (*\*очень финансово*); *композиционно делится на две части* (*\*довольно композиционно*). В этих значениях, как будет показано ниже, относительные наречия наследуют основные семантические типы употребления (интерпретации) относительных прилагательных.

ЗАМЕЧАНИЕ. Относительные (т. е. обозначающие неградуируемые признаки) прилагательные бывают не только отсубстантивные, но и отглагольные, ср.: *стиральный, строительный, ткацкий, подъемный, выдвигной*. Однако от глагольных прилагательных наречия не употребляются — и этот запрет носит системный характер, — хотя технически образовать наречия типа *стирально* можно.

Относительные наречия являются гибридной группой слов и заключают в себе некоторое противоречие — в них соединяются субстантивная семантика и адвербиальные синтаксические свойства. В силу своей гибридной природы они остаются за пределами классификаций.

Если прилагательные в грамматиках традиционно делятся на качественные и относительные, то наречия — не на качественные и относительные, а на характеризующие и обстоятельственные. Характеризующие, в свою очередь, делятся

на наречия образа действия (качественно-характеризующие, ср. *весело, мелодично, по-медвежьи, ежиком*, включая местоименные *как, так*, ср. [РГ-1980, т. I]) и степени (количественные), но в широком смысле количественные тоже можно считать качественными (при этом некоторые наречия образа действия, ср. *леишком, вприсядку*, качественными в оговоренном выше смысле не являются, т. к. не сочетаются с показателями степени, ср.: *\*очень вприсядку*).

Наречия типа *планово, административно, юридически* по словообразовательным и морфосинтаксическим свойствам похожи на качественные наречия — они имеют суффиксы *-о / -и* (как у качественных прилагательных типа *внезапно, дружески*) и отвечают на вопрос «как?». Поэтому теоретически они должны были бы попасть в разряд качественно-характеризующих, т. е. образа действия. Но у них субстантивный корень и относительное значение, что плохо совместимо с семантикой образа действия. И действительно, как будет показано ниже, они крайне редко употребляются в значении образа действия. В результате с классификацией этой группы наречий возникают проблемы. В разделах грамматик, учебных пособий и специальных исследований (ср. [Панков 2008]; [РГ 1980]; [Филипенко 2003]; [Greenbaum 1969]; [Thomason, Stalnaker 1973]), посвященных наречиям на *-о, -и*, о наречиях типа *планово, юридически* ничего не говорится. Они упоминаются лишь в отдельных работах (ср., например, [Битехтина 1979]) и иногда называются ограниченительными.

Языковое поведение относительных наречий определяется двумя основными факторами — контекстом (сочетаемостью) и собственной семантикой. Относительные наречия сочетаются с глаголами (*коммерчески освоить сеть мелких месторождений Таймыра*) и прилагательными (*коммерчески эффективная транспортная ограничена определенным расстоянием*). И такая сочетаемость для наречий является предсказуемым и закономерным, т. е. вполне тривиальным, свойством. Нетривиальным является их субстантивный корень и субстантивная семантика. Задача настоящей работы — показать, что относительные наречия сохраняют связь с относительными прилагательными, а через них — с существительными, и именно это определяет закономерности интерпретации их сочетаний с глаголами и прилагательными. В частности, существенно то, является ли семантика субстантивного корня предметной (ср. *карандашно*) или абстрактной (ср. *административно*).

Поскольку интерпретация наречий в сочетаниях с глаголами и прилагательными во многом — хотя и не полностью — различается, мы будем рассматривать эти сочетания отдельно.

## 2. Относительные наречия в контексте глаголов

Итак, если относительные наречия рассматривать как третью ступень (звено) цепочки «существительное → относительное прилагательное → относительное наречие», то они семантически и словообразовательно непосредственно связаны с однокоренными относительными прилагательными

и опосредованно — с существительными. Поскольку производные элементы в этой цепочке являются результатом синтаксической транспозиции, то относительные прилагательные и относительные наречия можно считать более низким рангом оформления субстантивного корня. При этом связь с субстантивом у относительных прилагательных и относительных наречий легко обнаружить на уровне синтаксического преобразования, ср.: *автомобильное колесо* — *колесо автомобиля*; *обойный клей* — *клей [предназначенный] для обоев*; *помогать денежно* — *помогать деньгами*; *территориально провинция делится на два района* — *с точки зрения территории*.

Разница между относительными прилагательными и относительными наречиями (помимо морфосинтаксических различий) состоит в том, что относительные прилагательные сочетаются с существительными самой разной семантики — предметами (*книжная обложка*, *детские ботинки*), пространствами (*вертолетная площадка*), в том числе — учреждениями (*обувная мастерская*), лицами (*финансовый инспектор*), ситуациями (*административная поддержка*), а относительные наречия сочетаются с собственно предикатными словами (глаголами и прилагательными), т.е. ситуациями. Поэтому при сопоставлении относительных наречий с относительными прилагательными последние будут рассматриваться в контексте существительных с предикатной семантикой. Речь идет прежде всего об отглагольных существительных, т.е. номинализациях, ср. *финансовая проверка*. Однако не все такие существительные производны от глаголов, ср. *финансовый кризис*. Поэтому мы будем называть существительные с предикатной семантикой ситуативами (при этом ситуации могут быть как динамическими, ср. *экономическое сотрудничество*, так и статическими, ср. *лекарственная зависимость*).

Существуют своего рода пары соотносительных сочетаний «прилагательное + существительное» (далее — адъективно-субстантивные сочетания, или AS) и «наречие + глагол» (далее — наречно-глагольные сочетания), ср.: *После аварийного приземления аэропорт закрыли на два с лишним часа* [Комсомольская правда, 2002.02.06] — *В июле из-за перегрева правого двигателя «боинг» аварийно приземлился в Ростове* [Комсомольская правда, 2011.05.03]. Наречно-глагольное сочетание нельзя представить как результат транспозиции адъективно-субстантивного, т.к. у из элементов разные направления транспозиции (*аварийный* → *аварийно*, но: *приземлиться* → *приземление*). Это, конечно, не означает, что для каждого наречно-глагольного сочетания имеется адъективно-субстантивная пара (некоторые существительные, ср. *крах / аудит*, как уже говорилось, не образованы от глаголов). Тем не менее между ними существует очевидная корреляция, и это позволяет сопоставлять, так сказать, «однокоренные сочетания». Данные НКРЯ показывают, что если употребляются оба сочетания, наречно-глагольных существенно меньше, чем адъективно-субстантивных, ср.: *технически перевооружаться* — 7 вхождений vs. *техническое перевооружение* — ок. 900; *технически превосходить* — 3 vs. *техническое превосходство* — ок. 40 (здесь и далее приводятся суммарные данные по основному и газетному корпусам). И вообще, по данным НКРЯ, употреблений относительных наречий существенно меньше, чем однокоренных относительных

прилагательных, ср. *технически* — ок. 3700 вхождений, *технический* — более 60000; *туристически* — 10 вхождений, *туристический* — больше 10000.

Поскольку у относительных наречий и относительных прилагательных один и тот же корень и мы рассматриваем их в контексте предикатов (не важно, глаголов или отпредикатных существительных), логично предположить, что интерпретация адъективно-субстантивных и наречно-глагольных сочетаний будет сходной.

Начнем с обзора значений прилагательных в контексте ситуативов.

Относительные прилагательные могут выражать (см. [Кустова 2015]):

– **актант**, в том числе:

Субъект: *прокурорская проверка* — ‘прокурор проверил’;

Объект: *грузовые перевозки* — ‘перевозить грузы’, *рыбная ловля* — ‘ловить рыбу’;

Инструмент: *автомобильные перевозки* — ‘перевозить автомобилями, на автомобилях’;

**функциональное пространство** (собственно пространство, помещение, учреждение, где происходит деятельность), ср.: *садовые работы* — ‘работы в саду как функциональном пространстве’ (например, обрезание засохших веток; работа на компьютере в саду или вязание в саду не являются садовыми работами); *заводские испытания* — ‘испытания на заводе’;

**функциональную сферу**: *медицинское обследование*; *юридическое обслуживание*.

Теперь обратимся к данным НКРЯ, чтобы выяснить, какую интерпретацию имеют относительные наречия в контексте глаголов и могут ли они, в частности, выражать актант, функциональное пространство и функциональную сферу.

### **Актант**

В типологических исследованиях языковые явления часто рассматриваются в рамках различных иерархий. Пребывание на первых позициях в таких иерархиях означает более высокий ранг языковой единицы в тексте, что выражается, в частности, в наличии большего количества возможностей. Так, для именных групп выделяются иерархии синтаксических позиций (подлежащее > прямое дополнение > не прямое дополнение > косвенное дополнение), иерархия семантических ролей (Агенс > Пациенс); иерархически организованы и другие характеристики участников ситуации (одушевленный > неодушевленный; референтный > нереферентный). Теоретически можно представить себе и иерархию способов кодирования (морфологического выражения) актантов. Основным способом для флексивных языков типа русского являются падежные формы. При этом актанты могут «понижаться» и до непадежных способов выражения.

Можно ли рассматривать субстантивные прилагательные и наречия как дальнюю периферию в иерархии способов выражения актантов; иначе говоря, можем ли мы считать относительные прилагательные и относительные наречия иерархически низким, но возможным способом кодирования актантов?

Что касается прилагательных, то относительные прилагательные действительно традиционно рассматриваются как один из способов выражения актантов при номинализациях. Так, в [ТКС 1984] под ред. И. А. Мельчука и А. К. Жолковского наряду с падежными формами (ср. *просьба Пети*) в модель управления абстрактных (отглагольных) существительных включаются прилагательные и притяжательные прилагательные — речь идет о случаях типа: *его просьба, Петина просьба, суворовские победы*.

Однако здесь необходимо сделать важное уточнение. В первую очередь речь должна идти не вообще о прилагательных, а именно о притяжательных прилагательных — и, конечно, притяжательных местоимениях; ключевое слово здесь притяжательность, т. е. референтность. *Суворовские победы* — это притяжательное значение. Здесь на иерархию способов выражения накладывается другая иерархия — референциальная. Чем менее референтный, так сказать, актант, тем ниже его статус. Актанты, выраженные «обычными», не притяжательными, относительными прилагательными тоже имеют трансформационные свойства актантов: сочетания AS трансформируются в конструкцию с падежом: *прокурорская / милицейская / экспертная проверка* — ‘прокуроры / милиционеры / эксперты проверяют’. Некоторые сочетания AS имеют такую же неоднозначность, как падежные конструкции, ср., с одной стороны, *приглашение писателя* — ‘писатель пригласил’ (субъект) или ‘писателя пригласили’ (объект), а с другой стороны, *юридическая проверка* — ‘юристы (субъект) проверяют кого-л. / что-л.’ или ‘кто-то проверяет юридический аспект деятельности’ (объект). Однако актанты, выраженные не-притяжательными прилагательными, нереперентны, а следовательно, неполноценны. Актант понижается в ранге не только синтаксически, но и референциально. Тем самым, оставаясь актантом, он превращается в характеристику, — что и является естественной функцией прилагательного.

**ЗАМЕЧАНИЕ.** Относительные прилагательные могут относиться и к референтному актанту, но тогда они употребляются в притяжательном значении, ср.: *Еще в коридоре услышал прокурорский голос* — конкретного прокурора.

Свойством относительного прилагательного, неоднократно отмечавшимся в литературе (см. [Павлов 1960; 1996; 2005]; [Панов 1999]), является выделение подкласса из класса объектов, обозначаемых существительными, т. е. обозначение разновидности данного типа объектов, например, *рыбный нож* и *хлебный нож* — разновидности ножей, *книжный шкаф* и *посудный шкаф* — разновидности шкафов. Такие сочетания имеют характер устойчивой номинации, причем в той или иной степени терминологической номинации — своего рода номенклатурной единицы. Сочетания относительных прилагательных с ситуативами имеют сходную природу: нереперентный актант «понижается» до характеристики и служит для обозначения типа, разновидности ситуации: *прокурорская проверка* — специальный вид проверки, *грузовые перевозки* — особая разновидность перевозок (наряду с пассажирскими), *читательская конференция* — вид конференции. Т. е. это тоже своего рода устойчивые номинации, номенклатурные единицы.

Наречно-глагольное сочетание по обоим характеристикам — обозначению типа ситуации и выражению актанта — проигрывает адъективно-субстантивному.

Во-первых, наречно-глагольное сочетание не может служить номенклатурной номинативной единицей. Во-вторых, наречие при глаголе — еще менее подходящий способ выражения актанта, чем прилагательное при ситуативе. Наречное выражение актанта при глаголе заведомо проигрывает конкуренцию падежному выражению актанта. Во-первых, наречие не может различать «глубинные падежи» (семантические роли) ни за счет морфологических показателей, ни за счет позиции при глаголе, т. е. является семиотически дефектным способом выражения актанта, тогда как разные падежные формы коррелируют с разным синтаксическими позициями и семантическими ролями (субъект vs. объект vs. инструмент). Во-вторых, наречие, в отличие от именных групп, не может различать референциальные статусы. Поэтому вполне естественно, что говорящие выбирают падежную форму выражения актанта при глаголе, чтобы иметь возможность различать референтность и роль.

Кроме того, запрет выражения наречиями основных участников ситуации имеет прагматическое объяснение: наречие — это дополнительная, необязательная характеристика ситуации, так что странно превращать основных участников в такие дополнительные характеристики (а именно это происходит при их выражении наречиями). Однако этот запрет не абсолютный. Возможность для наречия употребляться в качестве актанта глагола зависит от семантики субстантивного корня и семантической роли актанта в ситуации.

Наречия от существительных с предметными корнями в функции актантов практически не употребляются — независимо от семантической роли: *\*Нас прокурорски проверили* (субъект); *\*автомобильно перевозить грузы* (на автомобилях / автомобилями — инструмент); *рисовать карандашно* (карандашом); *\*добираться на работу велосипедно* (на велосипеде — инструмент); *\*Он лечится не лекарственно, а травами* (средство).

Если относительное прилагательное в контексте ситуатива может выражать объект-пациенс (*гаражное строительство* — ‘строить гаражи’; *пассажирские перевозки* — ‘перевозить пассажиров’), то однокоренное наречие при глаголе не интерпретируется как объект, а имеет статус дополнительной характеристики ситуации, например, сочетание *\*пассажирски перевозить* (будь оно употребительно) не означало бы ‘Х перевозит пассажиров’, т. к. позиция объекта при глаголе сохраняется и требует заполнения, ср.: *\*РЖД за прошлый год пассажирски перевезло 10 миллионов человек* — ‘в качестве пассажиров’.

Необходимо подчеркнуть, что системный запрет касается именно относительного, так сказать, субстантивного (и тем самым актантного) значения типа *\*прокурорски проверять*. Те же наречия с предметными корнями свободно употребляются в качественных значениях (за которыми стоит сравнение): — *Ты почему не позвонил?* — *прокурорски спросила жена* [Эльдар Рязанов, Эмиль Брагинский. Тихие омуты (1998)] — ‘как прокурор’; *Рисунок сделан углем, но выглядит он как-то карандашно* — ‘как будто нарисован карандашом’; *Из-за пробок ехали на мотоцикле практически велосипедно* — ‘как будто на велосипеде’.



Более или менее очевидно, почему наречия с предметными корнями не могут использоваться в качестве выражения актантов высокого ранга (субъект, объект), связанного с прямыми падежами, — они, как уже говорилось, не выдерживают конкуренции с падежными формами. Однако актанты более низкого ранга — инструмент и средство (стандартно выражаемые творительным падежом), — вообще говоря, входят в семантический блок «способ действия» и соотносятся с обстоятельствами образа действия (и сам творительный часто переходит в наречия образа действия: *боком, бегом*). Тем не менее и в этой роли наречия почти не употребляются, и ограничения здесь касаются характера семантики самого наречия. Наречия с «предметной» семантикой в роли инструмента или средства (*\*ехал автомобильно, \*рубил топорно*) практически не встречаются. В НКРЯ встретился случай, когда наречие можно перифразировать с помощью падежной формы и интерпретировать как валентность средства: *Специальные классы звезд обозначаются буквенно в таком порядке: О, В, А, F, G, К, М — от очень горячих голубых звезд с поверхностной температурой 100 000° до красных с температурой в 3000 °* [И. А. Ефремов. Туманность Андромеды (1956)] — ‘обозначаются буквами’. Тут важно также, что само это средство — в данном случае буквы — является нерелевантным, и кроме того, буква — особый, семиотический, «предмет».

Еще один характерный пример — наречия *денежно* и *финансово*: *деньги* и *финансы* тоже употребляются, в основном, нерелевантно, а кроме того, это тоже особые «предметы»: хотя они и существуют в виде купюр, но обычно имеется в виду их знаковая ипостась (средство платежа).

В текстах 19 и 20 вв. наречие *денежно* употреблялось на месте разных падежных форм. В примерах: *захотят помочь им денежно, Ему денежно помогали многие из его добрых друзей; Наиболее соответствующие приказу постановки — денежно поощрять; журнал денежно не обеспеченный; Во всяком случае, решено было за день 6 июня наградить его денежно* [Е. В. Тарле. Павел Степанович Нахимов (1802-1855) (1943)] — наречие *денежно* соответствует актанту, выражаемому творительным падежом: *помогать / наградить / обеспечить деньгами* (при этом надо отметить, что эти глаголы не относятся к семантическому классу физических действий, и актант, выражаемый творительным, не является средством в собственном смысле); в примерах: *Я не нуждался денежно, а чтение лекций для меня в профессорском реноме играло всегда вторую роль* [В. И. Вернадский. Дневники (1926-1934)]; *был стеснен денежно* — *денежно* соответствует предложному падежу, ср. *не нуждался в деньгах; стеснен в деньгах*. При этом иногда наречие входит в сочинительный ряд с падежной формой, что подчеркивает его актантный характер: *За покорность, за демонстрацию лояльности отменно ублажали, и не только денежно, но и разными наградами и поощрениями* [Эльдар Рязанов. Подведенные итоги (2000)]; *Подполковник начинал с Аргентины, послужил там мало, был стеснен и денежно, и в передвижениях по стране* [Анатолий Азольский. Монахи, 2000]

Для современного языка *денежно* звучит довольно архаично. В XX в. в таких контекстах больше используются наречия *финансово* и *материально* (иногда



также — *экономически*), — но обозначают они все равно деньги: *Грант Артурович вполне мог финансово обеспечить неработающую жену* [А. Маринина] — **финансово обеспечить** является синонимом *денежно обеспечить* и обозначает именно деньги (ср. *лекарственно обеспечить*; *продовольственно обеспечить*).

Таким образом, в некоторых случаях актант наречием все-таки выражается, но это неререферентный актант и притом низкого ранга.

### Функциональное пространство → образ действия

Наречия от прилагательных со значением функциональных пространств (*садовые работы*) не употребляются для обозначения пространства (ср. *\*работать фабрично* — т.е. ‘на фабрике’). При этом некоторые наречия от функциональных пространств широко употребляются в значении образа действия, однако это не любые функциональные пространства, а, так сказать, агентивные пространства, где человек что-то производит. Такое значение образа действия имеют, например, наречия *фабрично* и *промышленно*.

Фабрика — это с одной стороны функциональное пространство, а с другой — производство (промышленность — это совокупность производств), а это имплицитно соответствует способу производства: фабричное и промышленное производство противопоставлены ручному, кустарному, ремесленному. *Фабричный* и *промышленный* значит не просто ‘на фабрике’, ‘в промышленности’, а ‘фабричным / промышленным способом’ (*фабричное / промышленное производство*). Это значение есть уже в прилагательном, поэтому оно как бы автоматически передается наречию: *фабрично закупоренные бутылки*; *Производитель гарантирует качество и соблюдение условий хранения только для фабрично упакованного корма!* [Комсомольская правда, 2007.08.27]; *Во-первых, такая вещь может появиться у каждого, во-вторых, все, что сделано фабрично, обладает слабой энергетикой и быстро надоедает* [Комсомольская правда, 2002.04.02] — ‘фабричным / машинным способом, не вручную’. Аналогично: *промышленно произведенные товары* = ‘произведенные промышленным способом’.

Таким образом, относительные наречия сближаются с качественными в том смысле, что обозначают образ действия — однако не становятся качественными (ср. *\*очень промышленно*; *\*довольно фабрично*).

В значении функционального пространства выступает наречие *амбулаторно*, которое при этом содержит и компонент ‘способ’: *А едут к нам со всей России. Лечим мы амбулаторно и в стационаре* [«Сельская новь», 2003.10.07].

### Сфера / аспект

Наконец, рассмотрим наречия от прилагательных со значением сферы деятельности человека, сферы общественной жизни — *административно*, *дипломатически*, *научно*, *медицински* и т. п. (рассмотренное выше наречие *финансово*, разумеется, тоже может иметь значение сферы).

Во-первых, они тоже могут обозначать образ действия. Этому способствует то, что у них, как и у производящих прилагательных, непредметная (хотя и субстантивная) семантика. Сфера — это абстрактное пространство. Но управление, дипломатия, наука, медицина — не просто сфера жизни общества, это люди и их деятельность. И вот эту деятельность можно превратить в образ действия. Наречия *административно, дипломатически, научно, медицински* и под., употребляясь в значении образа действия, эксплуатируют, акцентируют семантику действия, методов, способов, свойственных данному типу деятельности: *Я имел возможность административно поддерживать или запрещать какие-то творческие направления* [Никита Хрущев. Воспоминания (1971)] — ‘административными методами’; *Сталин дипломатически, а Тельманы агитаторски требовали от социал-демократического правительства не впускать меня в Германию — надо думать, во имя интересов пролетарской революции* [Л. Д. Троцкий. Моя жизнь (1929-1933)] — ‘дипломатическими / агитационными методами’; *Страдание не всегда можно снять, но человеку помочь (медицински или душевно) его вынести — можно* [митрополит Антоний (Блум). Вопросы медицинской этики (1994)]; ср. также: *медицински подтвержденный диагноз; научно доказанный факт; юридически оспорить документ* — ‘медицинскими / научными / юридическими методами’ (*подтвержденный* и *доказанный* здесь выступают как причастия, т. е. формы глаголов).

Но, конечно, для относительных наречий с субстантивными корнями от абстрактных существительных семантика образа действия не является основной, и их употребления в такой функции, по данным НКРЯ, немногочисленны. Более типичным для таких наречий является употребление в значении сферы или аспекта, ср. примеры (а) и (б): (а) *Малый и средний бизнес — основа рыночной экономики. И все развитые государства законодательно, административно и экономически его поддерживают* [«Аргументы и факты», 2003.06.04] — в (а) административно значит ‘административными методами’, а не просто с административной точки зрения, но: (б) *Да, Эстремадура — сплошные границы. Административно она делится на две провинции. Географически (Толедским хребтом) — на Верхнюю и Нижнюю Эстремадуру* [«Вокруг света», 1984] — в (б) административно — это не способ, а именно аспект, противопоставленный географическому: с административной и географической точек зрения территория делится по-разному.

Значение аспекта / сферы реализуется в контексте глаголов с абстрактной семантикой и нефизическими актантами, которые требуют соответствующего уточнения, ср.: *Когда за сольный концерт мне предложат половину от гонорара «Квартета И» за корпоратив, буду считать, что коммерчески и музыкально я состоялся* [Известия, 2013.02.06] — ‘с точки зрения коммерции / музыки’, ‘в плане коммерции / музыки’; *Отдуваться в таком случае придется малому и среднему бизнесу, который, скорее всего, «разгрузят» административно и «нагрузят» финансово* [«Вслух о...», 2003.10.24]; *Далеко не каждое предприятие, желающее участвовать в выставке, может финансово и организационно осилить такую работу* [«Биржа плюс свой дом» 2002.09.16]; *Можно будет взять с собой и ребёнка — и при этом не тревожиться, что он со своими «хочу это!» вас*

**финансово опустошит**: цены тут вполне умеренные [«Домовой», 2002.04.04], ср. духовно / морально / психологически опустошит; Полное крушение гегемонии композиторской музыки связано с деятельностью американских минималистов, ибо практически все они так или иначе стали **профессионально осваивать** некомпозиторские музыкальные системы [Владимир Мартынов. Конец времени композиторов (2002)], ср.: практически / теоретически / технически осваивать.

Интересно, что относительные наречия, подобно падежным формам, могут выступать в позиции так называемого детерминанта (термин Н.Ю. Шведовой, см. [Шведова 1964; 1968]), т.е. распространителя всего предложения, относящегося к сказуемому (выраженному не только глаголом, но и, например, прилагательным): **Анатомически ухо представляет собой совокупность трех частей: наружного, среднего и внутреннего уха** [«Наука и жизнь», 2006]; **Необычна приспособляемость американских актеров. Когда они понимают, что в «Женитьбе» не нужно играть грубо, как в рекламе, то молниеносно перестраиваются и играют очень нежно. Профессионально они готовы и к тому, и к другому** [Анатолий Эфрос. Профессия: режиссер (1975-1987)]; **Композиционно фотографии идентичны: людей снимают в одинаковых позах** [РБК Дейли, 2013.08.30]; **И рядом с этими двумя контрастирующими тенденциями все напряженнее к концу повести звучит третья, от «неистой» поэтики идущая. Сюжетно она отражается в трагической развязке** [В.В. Виноградов. Гоголь и натуральная школа (1925)].

### 3. Относительные наречия в контексте прилагательных

Наиболее типичным контекстом для относительных наречий в значении аспекта / сферы является контекст прилагательных.

Аспект (или сфера) — это специфическая валентность интерпретационно-оценочных слов. В частности, эта валентность характерна для интерпретационно-оценочных прилагательных, см. [Кустова 2018б] (а также глаголов с абстрактной семантикой, см. предыдущий раздел). Такие прилагательные, в силу абстрактности своей семантики, нередко нуждаются в уточнении, в проекции в какую-то конкретную сферу. Эту функцию выполняет валентность аспекта.

Валентность аспекта у прилагательных может выражаться предложной группой по Дат.: картина **интересна по композиции**, а может выражаться относительным наречием, ср., например, контексты прилагательного **грамотный**: **идеологически грамотная** околесица; **исторически грамотные** люди, **коммерчески грамотные** собственники; **культурно грамотная** буржуазия; **методически грамотная** работа; **музыкально грамотные** любители; *Работает прекрасно — это раз. Политически грамотная — это два* [Л.К. Чуковская. Софья Петровна (1939-1940)]; **профессионально грамотное** решение задач психологической службы; **психологически грамотный** подход к воспитанию; **радиотехнически грамотная** сила; **готовить статистически грамотных** специалистов в области экономики; **стратегически грамотное** решение иркутских властей; **тактически грамотное**

общение со средствами массовой информации; **технически грамотные** американцы; У нас немало **художественно грамотных**, но бесталанных людей [«Искусство», 1957]; Правительство за все время своей работы не приняло ни одного **экономически грамотного** решения [Борис Немцов. Провинциал в Москве (1999)]; **юридически грамотные** контракты. Эту валентность можно перифразировать с помощью оборотов «с точки зрения», «в плане», «в аспекте»: **юридически грамотные** контракты — ‘с юридической точки зрения / в юридическом аспекте’; **методически грамотная** работа — ‘с точки зрения / в плане / в аспекте методики’.

Заметим, что даже при присоединении показателей степени наречия сохраняют относительное значение и «пропускают через себя» интенсификатор, который семантически относится к качественному прилагательному, т. е. относительное наречие не входит в сферу действия интенсификатора (о сфере действия см. [Богуславский 1985; 1996]): Любой **мало-мальски экономически грамотный** строитель вкупе с инвестором никогда не поддержит технологию, повышающую инвестиционную стоимость проекта [«Наука и жизнь», 2008] — **мало-мальски грамотный**, а не \*мало-мальски экономически; А наши жители **достаточно юридически грамотны**, чтобы понять, что за это из квартиры их не выселят [«Встреча» (Дубна), 2003.05.07] — **достаточно грамотны**, а не \*достаточно юридически.

Как показывают данные НКРЯ, если валентность аспекта выражима и предложным оборотом, и наречием, то чаще она выражается наречием: сочетания **беспомощный по** Дат. (ср. *Вся эта отрасль литературы представлялась Лему реакционной по форме и беспомощной по содержанию* [Александр Генис. Три «Соляриса» (2002) // «Звезда», 2003]) в НКРЯ встретились 4 раза, а сочетания с наречием (типа **юридически беспомощный**) — 29 раз; **выгодный по** Дат. (ср. *Финская компания известна как производитель надежных, выгодных по цене сухих отделочных смесей* [«Строительство», 2003.09.29]) — около 10 вхождений в НКРЯ; относительное наречие на -ски + **выгодный** (коммерчески / стратегически / тактически / экономически / энергетически **выгодный**) — около 200 вхождений.

Как и в контексте глаголов, в контексте прилагательных относительные наречия могут перифразироваться не только с помощью оборотов ‘в аспекте X-а’, ‘с точки зрения X-а’, но и с помощью других падежных форм и предложных оборотов, ср.: В частности, Москва, являясь наиболее **транспортно неблагоприятным** городом, отказалась от проведения Дня без автомобилей [Новый регион 2, 2005.09.22] — ‘неблагополучным с точки зрения транспорта’; **Лесных ресурсов, на которые ориентирована промышленность, почти не осталось, по крайней мере, в транспортно доступных районах** [РБК Daily, 2004.12.02] — ‘доступных для транспорта’; Укрупнение — **нормальный закономерный процесс для рынка, через который прошли все туристически развитые страны** [Новый регион 2, 2007.04.27] = ‘развитые в туристическом отношении, в сфере туризма’ vs. В России, помимо Питера, есть и другие **туристически привлекательные города** [Труд-7, 2010] — ‘привлекательные для туристов (и с точки зрения туризма)’; ср. еще: **Появились и лекарственно устойчивые формы туберкулеза** [Комсомольская правда,

2008.03.26] — ‘устойчивые к лекарствам’; *Я лекарственно зависим и знаю, что медикаменты следует принимать таким образом, чтобы лекарства не смешивались* [Сергей Есин. Марбург (2005)] — ‘зависим от лекарств’. Таким образом, помимо валентности аспекта относительные наречия при прилагательных могут выражать и другие валентности.

#### 4. Заключение

Итак, проблема синтаксической иерархии и выбора кодирования актанта применительно к относительным наречиям выглядит следующим образом. У глаголов основной способ кодирования актантов — падежи, хотя используются — очень редко — и наречия, ср.: *хорошо охарактеризовал* (но это качественные наречия!). Что касается относительных наречий при глаголах, то, как показывают рассмотренные выше примеры из НКРЯ, они могут кодировать некоторые актанты (средство, аспект), но в целом для глаголов это не характерно. У прилагательных ситуация другая. У прилагательных, особенно семантически соотносимых с глаголами, есть глагольные валентности, которые выражаются падежными формами, нередко даже совпадающими с глагольным управлением, ср. *готовиться к чему — готовый к чему*. Но это, так сказать, глагольный способ выражения валентностей, и прилагательное в таком случае смещается в сторону глагольного предиката. А «настоящее», прототипическое прилагательное — характеристика существительного, и оно не должно иметь при себе зависимых существительных (оно само зависит от существительного), *интересная книга* — это семантически полноценное сочетание, оно не требует выражения субъекта, хотя и допускает его (ср.: *интересная для детей книга*). Напротив, наречие — это типичный распространитель прилагательного.

Но у прилагательных с интерпретационно-оценочной семантикой есть валентность аспекта (сферы), специфическая именно для таких слов. И эту валентность, как свидетельствуют данные НКРЯ, для прилагательного более типично выражать наречием. Так, если взять одно и то же наречие, то его сочетаний с прилагательными окажется на порядок больше, чем сочетаний с глаголами. Так, сочетаний «*коммерчески* + прилагательное» (*коммерчески выгодный, коммерчески грамотный, коммерчески успешный, коммерчески эффективный* и т. п.) в основном и газетном корпусе более 400 (при этом причастия употребляются в таких контекстах в значении прилагательных, ср. *коммерчески оправданный*), а сочетаний «*коммерчески* + глагол» (ср. *коммерчески окупился*) — меньше 20.

Таким образом, выражение валентности (актанта) наречием в русском языке не запрещено, просто этот способ входит в другую иерархию — в иерархию распространителей прилагательных, — и в этой иерархии он даже «более главный», чем падежи. Т. е. относительные наречия могут системно кодировать валентность, что не характерно для качественных наречий (там это бывает случайно). Но это не любая валентность, а специфическая валентность абстрактных интерпретационно-оценочных слов — валентность аспекта.

Итак, относительные наречия — и по семантике, и по функциям — ближе к относительным прилагательным, чем к другим наречиям. Одна из главных функций относительных субстантивных прилагательных, как уже было сказано выше, — уточнение, конкретизация. Для относительных наречий, во всяком случае в контексте прилагательных, функцию конкретизации аспекта рассмотрения ситуации или признака также следует признать основной.

### Литература

*Битехтина Г. А.* Семантико-синтаксические разряды определительных наречий в современном русском языке и условия их функционирования. Автореф. ... канд. филол. наук. М., 1979.

*Богуславский И. М.* Исследования по синтаксической семантике. М.: Наука, 1985. 175 с.

*Богуславский И. М.* Сфера действия лексических единиц. М.: ЯСК, 1996. 460 с.

*Кустова Г. И.* Скрытая грамматика русских атрибутивных конструкций // Изв. РАН. Сер. лит. и яз. 2015. № 5. С. 5–21.

*Кустова Г. И.* Прилагательные // Материалы к корпусной грамматике русского языка. Вып. 3. Части речи и лексико-грамматические классы. СПб.: Нестор-История, 2018а. С. 40–107.

*Кустова Г. И.* Семантические и конструктивные валентности прилагательных // *Prace filologiczne*. Warszawa, 2018b. № 73. Р. 223–238.

*Павлов В. М.* О разрядах имен прилагательных в русском языке // Вопросы языкознания. 1960. № 2. С. 65–70.

*Павлов В. М.* Качественность и субстанциальная семантика // Теория функциональной грамматики. Т. 5. Качественность. Количественность. СПб., 1996. С. 8–53.

*Павлов В. М.* Предметно-относительный атрибут как типологическая константа // Проблемы функциональной грамматики. Полевые структуры. СПб.: Наука, 2005. С. 169–192.

*Панков Ф. И.* Опыт функционально-коммуникативного анализа русского наречия. М.: МАКС Пресс, 2008. 448 с.

*Панов М. В.* Позиционная морфология русского языка. М.: Наука–ЯРК, 1999. 275 с.

РГ-1980, т. I — Русская грамматика. В 2-х тт. / Под ред. Н. Ю. Шведовой. М.: Наука, 1980. Т. I.

ТКС 1984 — *Мельчук И. А., Жолковский А. К.* Толково-комбинаторный словарь современного русского языка. Wien. 1984. 992 с.

*Филипенко М. В.* Семантика наречий и адвербиальных выражений. М.: Азбуковник, 2003. 304 с.

*Шведова Н. Ю.* Детерминирующий объект и детерминирующее обстоятельство как самостоятельные распространители предложения // Вопросы языкознания, 1964. № 6. С. 77–93.

*Шведова Н. Ю.* Существуют ли все-таки детерминанты как самостоятельные распространители предложения // Вопросы языкознания, 1968. № 2. С. 39–50.



Greenbaum S. Studies in English adverbial usage. London, 1969.

Thomason R., Stalnaker R. A semantic theory of adverbs // Linguistic Inquiry. Cambridge (Mass.), 1973, v. 4, № 2. P. 195–220.

**G. I. Kustova**

*Vinogradov Russian Language Institute of the Russian Academy of Sciences  
(Russia, Moscow)  
galinak03@gmail.com*

## **RELATIVE ADVERBS IN THE CONTEXT OF VERBS AND ADJECTIVES (ACCORDING TO THE NATIONAL CORPUS OF THE RUSSIAN LANGUAGE' DATA)**

The paper is devoted to adverbs with noun roots, derived from the relative substantive adjectives: *avarijno* ('emergency'), *administrativno* ('administratively'), *anatomicheskii* ('anatomically'), *kommercheskii* ('commercially'), *meditsinskii* ('medically'), *muzykal'no* ('musically'), *nauchno* ('scientifically'), *promyshlennno* ('industrially'), *professional'no* ('professionally'), *social'no* ('socially'), *territorial'no* ('geographically'), *tekhnicheskii* ('technically'), *fizicheskii* ('physically'), *finansovo* ('financially'), *himicheskii* ('chemical-ly'), *ekonomicheskii* (economically), *yuridicheskii* ('legally'), etc. Such adverbs are numerous, as evidenced by data from the National corpus of the Russian language (RNC), though not all of them are included even in electronic dictionaries, not to mention paper dictionaries. They are used mainly in written and oral literary speech. Relative adverbs are considered in terms of their ability to express the valence of predicate words — verbs and adjectives. Relative adjectives can express valences of the subject, object, instrument, and also be used in the meaning of functional space and aspect. Relative adverbs with verbs practically refer to the actants, with the exception of means (*oboznachit' bukvenno* 'to denote by letters') and aspects (*Administrativno i geograficheskii oblast' delitsya na dve chasti* 'administratively and geographically the region is divided into two parts'). Relative adverbs in the context of adjectives refer mainly to the aspect (*ekonomicheskii privlekatel'nyj* 'economically attractive', *kommercheskii uspekhnyj* 'commercially successful'). Aspect (sphere) is a specific valence of interpretative words.

*Key words:* relative adjective, relative adverb, predicate, valence

## **References**

Bitekhtina G. A. *Semantiko-sintaksicheskie razryady opredelitel'nykh narechij v sovremennoy russkoy yazyke i usloviya ikh funkcionirovaniya* [Semantic and syntactic discharges of definitive adverbs in the modern Russian and conditions of their functioning]. Avtoref. ... kand. filol. nauk. M., 1979.



Boguslavskii I. M. *Issledovaniya po sintaksicheskoi semantike* [Research on syntactic semantics]. Moscow: Nauka, 1985. 175 p.

Boguslavskii I. M. *Sfera deistviya leksicheskikh edinits* [The scope of lexical units]. Moscow, YSK, 1996. 460 p.

Filipenko M. V. *Semantika narechij i adverbial'nyh vyrazhenij* [Semantics of adverbs and adverbial expressions]. Moscow, Azbukovnik, 2003. 304 p.

Greenbaum S. *Studies in English adverbial usage*. London, 1969.

Kustova G. I. [Hidden grammar of Russian attributive constructions]. *Izvestiya RAN. Seriya literatury i yazyka*. 2015, № 5, pp. 5–21.

Kustova G. I. [Adjectives]. *Materialy k korpusnoj grammatike russkogo yazyka. Vyp. 3. Chasti rechi i leksiko-grammaticheskie klassy* [Materials to corpus grammar of Russian. Issue 3. Parts of speech and lexical and grammatical classes]. St. Petersburg, Nestor-Istoriya, 2018a, pp. 40–107.

Kustova G. I. [Semantic and constructive valences of adjectives]. *Prace filologiczne*. Warszawa, 2018b, № 73, pp. 223–238.

Pavlov V. M. [On the discharges of adjectives in Russian]. *Voprosy yazykoznaniya*. 1960, № 2, pp. 65–70.

Pavlov V. M. [Quality and Substantial Semantics]. *Teoriya funkcional'noj grammatiki*. T. 5. Kachestvennost'. Kolichestvennost' [Theory of functional grammar. V. 5. Quality. Quantitative]. St. Petersburg, 1996, pp. 8–53.

Pavlov V. M. [Subject-relative attribute as a typological constant]. *Problemy funkcional'noj grammatiki. Polevye struktury* [Problems of functional grammar. Field structures]. St. Petersburg, Nauka, 2005, pp. 169–192.

Pankov F. I. *Opyt funkcional'no-kommunikativnogo analiza russkogo narechiya* [The experience of the functional-communicative analysis of the Russian adverb]. Moscow, MAKSPress, 2008. 448 p.

Panov M. V. *Pozitsionnaya morfologiya russkogo yazyka* [Positional morphology of Russian]. Moscow, Nauka–YRK, 1999. 275 p.

Russkaya grammatika [The Russian Grammar]. V. II. Ed. by N. Yu. Shvedova. Moscow, Nauka, 1980.

Shvedova N. Yu. [Determining object and determining circumstance as independent spreader of a sentence]. *Voprosy yazykoznaniya*. 1964, № 2, pp. 77–93.

Shvedova N. Yu. [Are there any determinants as independent spreader of a sentence?]. *Voprosy yazykoznaniya*. 1968, № 2, pp. 39–50.

TKS 1984 — Mel'chuk I. A., Zholkovskij A. K. *Tolkovo-kombinatornyj slovar' sovremennogo russkogo yazyka* [Defining combinatorial dictionary of the modern Russian]. Wien, 1984, 992 p.

Thomason R., Stalnaker R. A semantic theory of adverbs. *Linguistic Inquiry*. Cambridge (Mass.), 1973, v. 4, № 2, pp. 195–220.

<sup>1</sup>С. Ю., Жукова, <sup>1</sup>Б. В. Орехов, <sup>1,2</sup>Е. В. Рахилина

<sup>1</sup>Национальный исследовательский университет «Высшая школа экономики»

<sup>2</sup>Институт русского языка им. В. В. Виноградова РАН

(Россия, Москва)

syupuzhaeva@gmail.com, nevmenandr@gmail.com, rakhilina@gmail.com

## ДИСКУРСИВНЫЕ ФОРМУЛЫ РУССКОГО ЯЗЫКА: ДИАХРОНИЧЕСКИЙ ПОДХОД\*

Статья посвящена проблеме описания дискурсивных формул русского языка с позиции диахронии. Под дискурсивными формулами понимаются устойчивые, легко воспроизводимые изолированные конструкции, ср.: *Еще бы! Не то слово! То-то же* и под. Они служат ответными репликами в диалоге и в отличие от традиционных конструкций не содержат переменных внутри себя: свободным слотом для них становится предшествующая реплика другого говорящего. Важным аспектом описания дискурсивных формул является динамика их изменений во времени. В силу своей частотности, семантической опустошенности и прагматической нагруженности дискурсивные формулы так быстро появляются, исчезают и сменяют друг друга, что эти изменения видны даже на выбранном нами временном отрезке в 200 лет, который охватывает XIX–XX век. На базе статистического исследования, которое упорядочивает составленный нами список русских дискурсивных формул на временной оси, в статье последовательно анализируются примеры уходящих, новых и стабильных дискурсивных формул. На этих примерах исследуются механизмы возникновения новых формул, связанные с прагматикализацией (то есть превращением в идиоматичные выражения с прагматическим значением прежде композиционных сочетаний), а также причины старения формул и природа неустойчивости их формы и семантики.

**Ключевые слова:** дискурсивные формулы, диахронические исследования, корпусная лингвистика, грамматика конструкций, Русский конструктикон, дискурс пьес.

---

<sup>1</sup> Работа выполнена при поддержке РФФИ (грант № 17-29-09154).

## 1. Введение: постановка задачи

Одной из наиболее значимых лингвистических теорий XX века является теория Грамматики конструкций [Fillmore 1988; Fillmore, Kay 1992], согласно которой основной лингвистической единицей признается конструкция. При этом под конструкцией понимается «языковое выражение, у которого есть аспект плана выражения или плана содержания, не выводимый из значения или формы составных частей» [Рахилина, Кузнецова, 2010: 19], см. подробнее [Hoffmann, Trousdale 2013]. Разные конструкции различаются по своему синтаксису и семантике, так что можно говорить о типах конструкций и классификации конструкций, которая для русского языка предпринята в рамках проекта Русский конструктикон (подробнее см. Janda et al. 2018).

Согласно этой классификации, одним из важных и нетривиальных классов конструкций являются так называемые *дискурсивные формулы* (термин восходит к [Fillmore 1984]): законченные неоднословные реплики в диалоге в ответ на вербальный стимул собеседника, не обусловленный частной социально значимой ситуацией, ср. *Не тут-то было!*, *Как же это?* *Вон оно как!* и под. Структурно дискурсивные формулы (ДФ) включают в себя преимущественно частицы, союзы, местоимения, но также и исходно полнозначные, но десемантизированные слова.

Подробно особенности дискурсивных формул как конструкций рассмотрены в статье настоящего выпуска Трудов [Бычкова и др. 2018] на примере полисемичной формулы *да ну*. Там приводятся сходные классы языковых явлений и близкие термины<sup>1</sup>, а также дается синхронный взгляд на семантико-синтаксические свойства формул, которые маркируются прежде всего интонацией и жестикulyацией. Пример *да ну, ты что* и других формул показывает, что все особенности формул мотивированы дискурсивным характером их значения, возникшего в результате выветривания семантики и потери композициональности исходной для формулы сложной языковой единицы (словосочетания).

Именно этот, диахронический аспект функционирования в языке дискурсивных формул (ДФ) будет нас интересовать. Он тоже имеет свои особенности для дискурсивных формул по сравнению с обычными конструкциями: в их случае конструкционализация заканчивается образованием не единицы со значением из универсального грамматического набора, а единицы с особым, прагматическим значением — эмфатического согласия, одобрения, отрицания и проч. Эмфатичность и оценочность этих единиц становится причиной довольно быстрых, наблюдаемых даже на протяжении короткого промежутка времени, диахронических изменений в этих конструкциях, и это свойство нужно учитывать при их описании: в частности, оно способствует вариативности ДФ — большей, чем обычных конструкций, и в целом значительной нестабильности этих единиц. Нашей задачей было «измерить» эту нестабильность, понять ее природу и проиллюстрировать

---

<sup>1</sup> См., например, *formulae* [Coulmas 1981], *conversational routines* [Aijmer 1996], *pragmemes* [Mey 2001], *formulaic sequence* [Wray 2002], коммуникативы [Шаронов 1996] и др.

соответствующие лингвистические процессы примерами изменений, которым подверглись конкретные формулы русского языка.

Таким образом, наша работа укладывается в широкий круг исследований, которые посвящены изучению микродиахронии русского языка и начинались с классических трудов [Булаховский 1948; Виноградов, Шведова 1964]. По большей части тщательному изучению подвергались грамматические категории (см, в частности, RusConstr 2017), недавнюю статью [Падучева 2018] на материале романа Л. Н. Толстого «Анна Каренина». Между тем значительная часть работ, посвященных языку периода А. С. Пушкина, показывает, что не только грамматические, но и «лексические и прагматические особенности речи <...> очень сильно изменились за последние 160-170 лет» [Добровольский 2001], что может выражаться в порой заметных для читателя семантических сдвигах [Пеньковский 2005], [Зализняк Анна 2012] и др.

В последнее время задача исследования микродиахронических изменений решается с привлечением корпусного материала, ср. [Даниель, Добрушина 2016], предпринимаются попытки выявления семантических сдвигов на основе больших массивов данных [Kuzmenko, Kutuzov 2017], создан и специальный корпус языка XIX века [Рахилина и др. 2016, Рахилина 2017]. Все эти публикации служат методологической основой для нашей микро-диахронической работы с ДФ как лексико-грамматическими единицами.

В диахронических исследованиях в рамках грамматики конструкций в зоне внимания в конечном итоге всегда оказывается история отдельной конструкции [Hilpert 2013]. Однако ДФ представляют более гомогенный класс, важный для изучения таких системных процессов, как грамматикализация и прагматикализация [Norper, Traugott 2003]. Как кажется, для него можно выделить определенные закономерности, которые могут быть видны даже на достаточно коротком временном промежутке.

С этой целью мы полуавтоматически составили список ДФ русского языка и затем провели статистический анализ этого списка в диахронии, определив динамику частотности каждой формулы по Национальному корпусу русского языка (НКРЯ) в течение примерно последних двух веков (то есть начиная с XIX века). Эти подготовительные этапы нашей работы будут коротко описаны в разделе 2. Статистика показала, что несмотря на то, что мы взяли такой короткий временной интервал, внутри него в нашем списке есть «новые», то есть недавно возникшие формулы и «старые», уходящие или даже уже ушедшие из языка. В разделах 3 и 4 мы рассмотрим примеры таких формул и проанализируем ход изменений, которые они претерпевают, на материале Национального корпуса русского языка. Раздел 5 посвящен формулам, которые стабильно присутствуют в русском языке на протяжении двух веков. Раздел 6 Заключение подводит итоги нашего исследования.

## 2. Дискурсивные формулы: динамика изменений в ДФ по данным НКРЯ

Процедура создания списка ДФ подробно описана в [Пужаева и др. 2018]. Вначале была проведена ручная разметка более 60 пьес, которая послужила основой для дальнейшего машинного обучения. Пьесы были выбраны в качестве исходного материала, поскольку они в большей степени имитируют разговорную речь. После нескольких этапов улучшения автоматического модуля по извлечению ДФ из текстов мы применили готовую программу к 420 пьесам XIX–XXI вв., взятым из двух корпусов драматических текстов [Lubimovka 2018, Russian Drama Corpus 2018].

Первоначально в таблице, созданной после автоматической обработки материала, было около 3000 структурно устойчивых единиц, которые в дальнейшем были классифицированы с точки зрения принадлежности их к синтаксическим конструкциям, этикетным выражениям и иным коллокациям. Собственно дискурсивных формул<sup>2</sup>, у которых на первый план выходит прагматическая функция, оказалось не так много: учет вариативности и объединение их по структурному и семантическому сходству позволил сократить список, в котором сейчас осталось 925 единиц. Понятно, что этот список неоднородный — и в формальном и в семантическом отношении, и требует дальнейшей работы: систематизации, классификации и подробного анализа. Однако и теперь в нем явно просматривается противопоставление устаревших и новых, недавно появившихся ДФ.

Отправной точкой нашего исследования был гипотеза о том, что даже на таком коротком этапе, как 200 лет, мы увидим динамику изменения этих единиц, потому что они шаблонны как реплики и обладают высокой частотностью, и это должно ускорять процесс их контрукционализации — а точнее прагматикализации.

Материалом для проверки этого предположения послужила текстовая коллекция Национального корпуса русского языка (НКРЯ). Благодаря богатой метатекстовой разметке мы знаем дату создания каждого текста и можем расположить найденные в этих текстах ДФ на временной шкале.

Поиск по текстам производился автоматически. Мы использовали список ДФ (полученный на предыдущем этапе), и искали их вхождения в текстах таким образом, чтобы они оказались между знаками препинания, традиционно отграничивающими предложение или вставную конструкцию (таким образом, сочетания внутри запятых как знаков пунктуации не учитывались). Такое дополнительное

---

<sup>2</sup> Заметим, что наши формулы часто очень многозначны, и в разных значениях «пересекают границы» собственно дискурсивных, представляя в других своих значениях единицы другого плана. В этом отношении ручная «чистка» итогового списка при любом, даже самом высоком качестве программы, будет неизбежной. Хорошей иллюстрацией может служить формульное сочетание *скажите пожалуйста*, которое прежде всего является известной застывшей формулой вежливости в вопросе, однако ввиду своего дальнейшего развития приобрело круг дискурсивных употреблений: — *Скажите лучше, что за дама была с Евлампием. — А это Евлампий? — Он. — Скажите пожалуйста! Важный, со спины и не подумаешь, что лакей.* [Леонид Юзефович. Дом свиданий (2001)].

ограничение на поиск накладывалось, чтобы избежать смещений, представленных в парах типа (1) и (2):

- (1) — *Ну что ж, — сказал он хрипло, — давай бог. Давай, давай. Может быть, тут тебя постигнет удача. Не вышло с романом, **кто знает**, может быть, с пьесой выйдет.* [М. А. Булгаков. Записки покойника (Театральный роман) (1936-1937)].

Ср. Обычно ей нравились читающие люди. Но в случае Жени-сана чтение имело непредсказуемые последствия. — Разве есть такие книжки? — **Кто знает?** И вообще — все это неважно. Важно, чтобы с Марсиком не случилось ничего плохого. Костик был полностью согласен с мамой. [М. С. Аромштам. Мохнатый ребенок (2010)].

- (2) *Её работа (Елена Андреевна) в «Дяде Ване» Льва Додина, **без сомнения**, стала одним из важных событий прошедшего фестиваля. Они с Соней ровесницы, почти сёстры.* [Легкое дыхание (2004) // «Экран и сцена», 2004.05.06].

Ср. — Ты уверен, что это она? — проговорил «священник». Его лицо по-прежнему было скрыто капюшоном. — **Без сомнения!** Она рождена под знаком числа света. Четырежды семь, я это проверил. Она рождена в царском доме, в ее жилах течет священная кровь... [Наталья Александрова. Последний ученик да Винчи (2010)].

Для каждой ДФ фиксировалось только одно вхождение в тексте: если ДФ попала в одном документе большее число раз, остальные вхождения игнорировались: такой подход был избран для того, чтобы исключить влияние индивидуального стиля автора. При этом абсолютное число найденных в текстах за определенный год ДФ делилось на совокупное число слов в документах за этот год. Этот способ нормализации выбран потому, что вероятность появления ДФ зависит именно от объема корпуса, а не от числа документов. Иными словами, в большом корпусе, состоящем из ограниченного числа длинных текстов ДФ встретить проще, чем в большом корпусе, включающем много коротких произведений. Разумеется, это утверждение справедливо не всегда, но как выяснилось, оно хорошо соотносится со спецификой жанрового состава НКРЯ.

Так как некоторые ДФ довольно редки, а деление их вхождений производилось на число на много порядков превышающее количество этих вхождений, в итоге получалась дробь с длинным «хвостом» цифр после запятой, из которых первые несколько были нулями. Для удобства мы умножили получившийся результат на 1000000.

Полученные данные подтвердили наше предположение о высокой динамичности этих единиц: значительную часть ДФ из списка можно отнести к категории устаревших и устаревающих (*да что такое, как ты смеешь, сделай милость*), в то же время мы видим новые единицы, появившиеся в русском языке только конце XX — начале XXI века (*без проблем, все в порядке, так не бывает*),

и наконец, — небольшую группу ДФ, которые употребляются с той или иной мерой постоянства на протяжении практически всего анализируемого периода (*что такое, что это, как же*). Рассмотрим примеры каждого из типов.

### 3. Устаревшие и устаревающие формулы

Группу, занимающую по объему около 40% от всех формул, составляют устаревшие и устаревающие формулы — те, которые стабильно не употребляются на протяжении хотя бы последних четырех десятилетий или имеют тенденцию к уменьшению количества употреблений, хотя продолжают фигурировать в текстах вплоть до XXI века, такие как *тьфу, пропасть; сделай милость; почему я знаю* и проч. Они служат яркой лингвистической характеристикой классических текстов середины XIX века и приемом стилизации для писателей следующих поколений, которые обращаются к минувшей эпохе, ср. (3):

- (3) *Фома попробовал подтянуть силным волчьим тенорком, сбился и сбил Маманю. — Тьфу ты, пропасть! — раздраженно плюнула она. — Такую песню испортил!* [Сергей Осипов. Страсти по Фоме. Книга вторая. Примус интер парэс (1998)].

Примерно в 20% случаев, как показывает материал, устаревание конструкций происходит за счет устаревания или существенного семантического изменения исходно полнозначных слов в их составе. Хорошим примером такого рода может служить вопросительное слово *почем*, которое до сих пор существует в разговорной речи в значении ‘сколько стоит / какая цена’, но устарело и исчезло в современном языке в значении ‘почему’.

Действительно, если в конце XVIII — начале XIX в. мы видим разнообразные варианты формул с элементом *почем*: *почем знать, почем знаешь, я почем знаю, почем угадать, почем ты знаешь* и т. д., и они оказываются более употребительными, чем слово *почем* в значении ‘сколько стоит’, то к середине XX в. ситуация

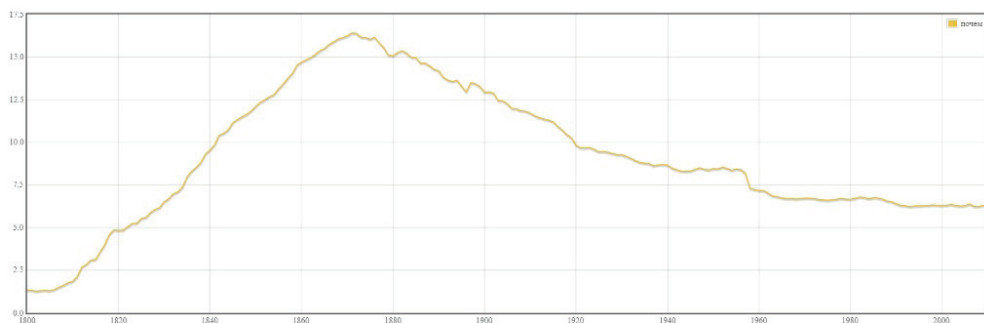


Рис. 1. Частотность употребления слова *почем*

Здесь и далее поиск производится с учетом введенного ранее ограничения: ДФ с двух сторон ограничена знаками конца предложения. К графикам применена функция сглаживания



в корне меняется: более редким становится формульное употребление и на первый план выходит использование этого слова в значении ‘сколько стоит’. (Одновременно появляется и быстро набирает частотность также некомпозициональное фразеологизованное сочетание *почем зря*).

Между тем сама формула в этом случае сохранилась: место устаревшего значения вопросительного *почем* заняло другое вопросительное слово *откуда* в значении ‘почему’, так что на смену формуле *почем я знаю* (рис. 2) пришла формула *откуда я знаю*:

- (4) Ты скажи, куда восемьсот рублей делись? — **Почем я знаю?** Разве я знаю, куда вы тратите? [И. А. Гончаров. Обломов (1859)].
- (5) Отец, когда умер Достоевский? — **Откуда я знаю?** — сказал мистер Браун, улыбаясь. [Б. Б. Вахтин. Портрет незнакомца (1966)].

Заметим, что для вопросительного *откуда* семантика причины тоже не единственная — исходным для него является пространственное значение исходной точки (ср.: *откуда X пришел?*). Вообще говоря, сдвиг в причинную зону из пространственной в контексте *знать* вполне естествен, потому что для *знать* источник знания и является причиной-основанием. Однако он сопровождается повышенной эмфатичностью: классические вопросительные контексты в этом случае выражают вопрос-недоверие, когда сообщенная собеседником информация входит в конфликт с презумпциями говорящего. Соответствующая дискурсивная конструкция (по форме — риторический вопрос говорящего к себе) с семантикой ‘не знаю’

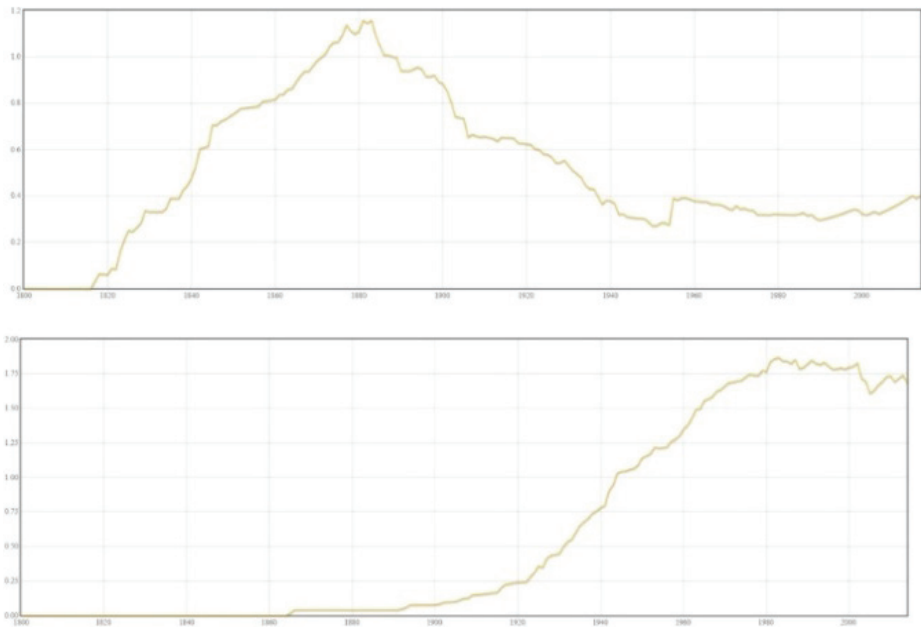


Рис. 2. Употребление формул *почем я знаю* и *откуда я знаю*

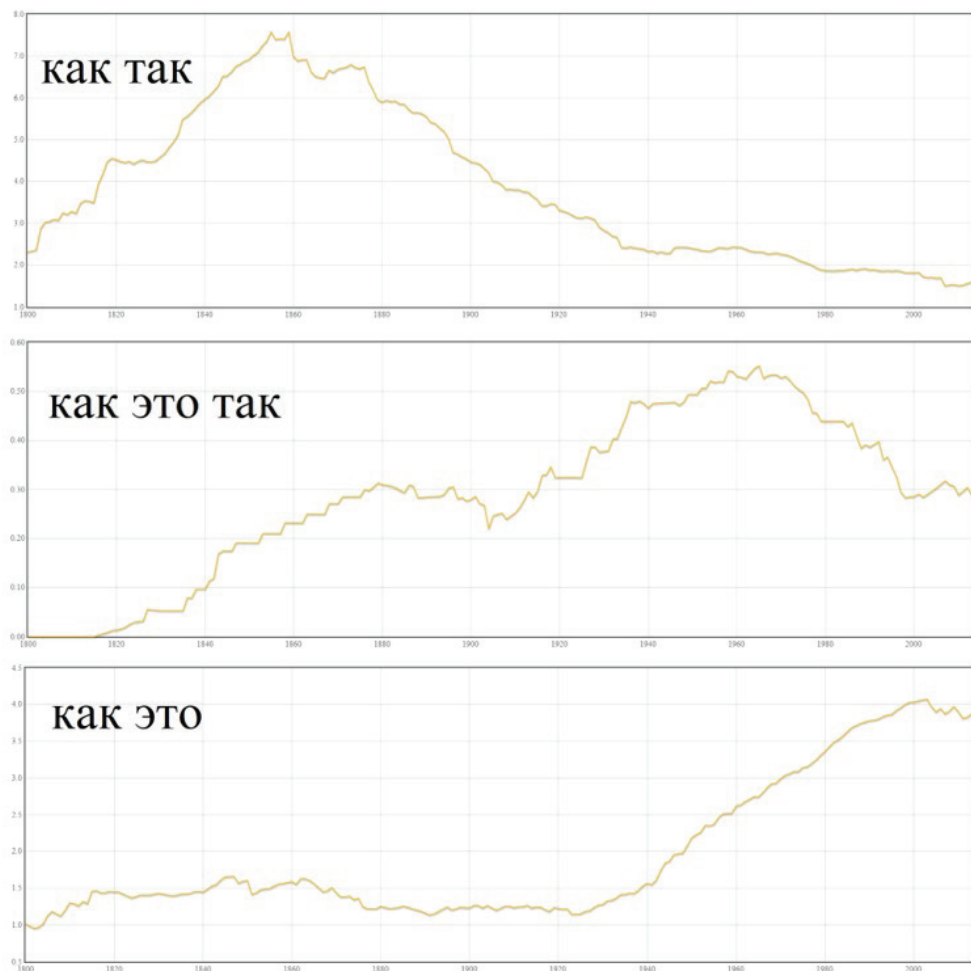


Рис. 3. Функционирование ДФ *как так*, *как это так*, *как это*

и прагматикой отказа от коммуникации помимо эмфатичности, как и *почем я знаю*, характеризуется еще и сниженностью. В этом смысле замена устаревшей лексики на альтернативную действительно сохранила формулу в той функции, которая ей была свойственна прежде — что, по всей видимости, свидетельствует о значимости этой функции как таковой.

Между тем устаревание конкретного элемента конструкции — не единственный стимул к изменению ДФ. Примером другого рода служит история формулы *как так*, в которой нет однозначных устаревших компонентов — тем не менее, статистика НКРЯ показывает, что она постепенно уходит из употребления, конкурируя с *как это* и *как это так* с «равноценным» набором местоименных составляющих.

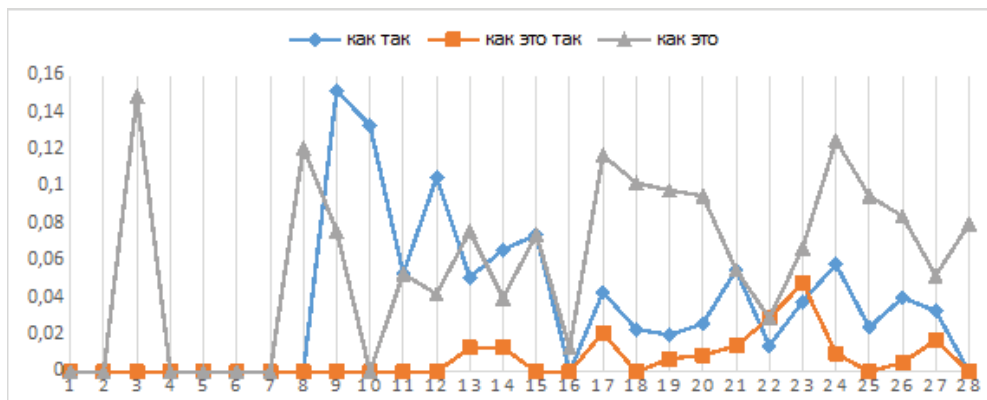


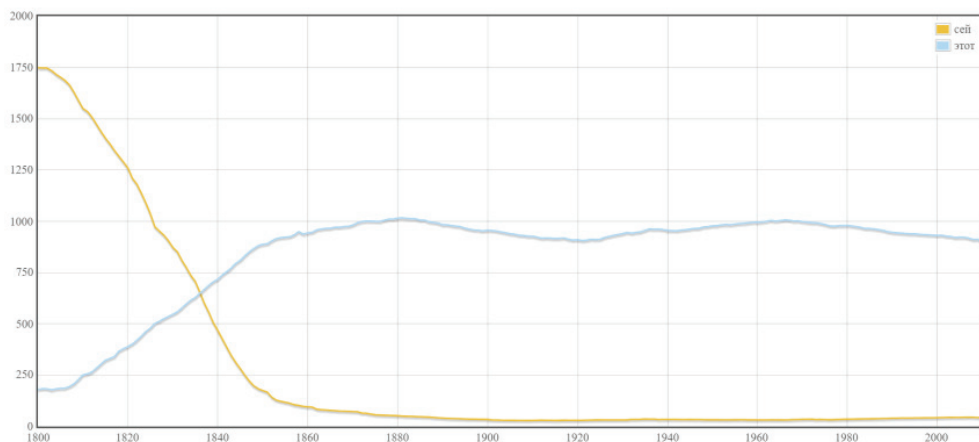
Рис. 4. Соотношение употребления ДФ как так, как это так, как это

При этом все три формулы к середине XX в. выражают одно и то же прагматическое значение удивления, недоумения из-за несоответствия сообщаемого тому, что знает собеседник или что он предполагал:

- (6) — *Поскольку моя практика была недостаточно длительной, я родился снова. Но уже не вампиром. Гера засмеялась, и ее смех был холодным и злым. — Как так? — спросила она. — Вы нам только что целый час объясняли, что никакого «я» на самом деле нет. А потом говорите, что родились снова.* [Виктор Пелевин. Бэтман Аполло (2013)].
- (7) *Тебе собирались позвонить! Не торопись, не делай ошибки! Пойми, я просто не успел объяснить. — Как это так? Чай выпить успел, опиум схватить успел, а главного не сказал? Убайдулла сунул острие лома в костер и начал его нагревать: — Ччас уши дыра делать.* [Михаил Гиголашвили. Чертово колесо (2007)].
- (8) *Почему бы на это не купиться?.. — ... Игорь Рюрикович, не жалеете, что ушли с завода? — Я не ушел с завода. — Как это? Мы вроде с вами сейчас не на заводе, — я оглядываю кабинет в резиденции. Ни станков, ни кранов — стол, кресла, дорогая подставка для ручек. — Мне президент поручил работать.* [Дарья Данилова, Дмитрий Карцев. Вышел в дамки // «Русский репортер», 2013].

Все эти формулы прошли примерно одинаковый путь от композиционного вопроса к удивлению, при этом они являются структурными вариантами одной единицы.

Из Рис. 4 видно, что в определенный момент все эти формулы обладают почти одинаковой частотой употребления, однако в дальнейшем частотность «как так» и «как это так» снижается, в то время как частотность формулы «как это» растет. Со временем в функции *как так* начинает употребляться формула *как это*, оказавшаяся более жизнеспособной, чем гибридная единица *как это так*.

Рис. 5. Соотношение употребления местоимений *сей* и *этот*

Ясно, что динамика ДФ объясняется тем, что внутри вариантов одной формулы происходит перестройка с заменой компонентов, способных сохранить общее прагматическое значение формулы, при этом на смену менее частотному приходит более частотный элемент. С точки зрения статистики, это значит, что распределение частот диахронических вариантов ДФ должно напоминать распределение форм *сей* и *этот* (см. рис. 5), которые можно считать вариантами указательного местоимения и которые представлены в языке начала XIX века с явным преобладанием одного над другим. Затем, как видно из графика, тенденция меняется на противоположную: *этот* начинает успешно конкурировать с *сей* и в итоге получает преимущество.

Дизайн исследования заключался в следующем. Мы сгруппировали варианты одной ДФ таким образом, чтобы в разметке заключалось указание на то, что, например, *а как иначе*, *а как же иначе*; *так что ж*, *ну так что* и т. д. относятся к одной и той же группе ДФ. Затем для каждого из вариантов были получены значения вхождений в корпусе за отдельные десятилетия, а после подсчитан коэффициент корреляции для каждой пары вариантов. Мы ожидали, что конкурирующие формы получат высокое отрицательное значение этого коэффициента. Это бы означало, что в тот момент, когда частотность одного варианта ДФ растет, частотность противопоставленного ему варианта должна падать и наоборот. Само по себе отрицательное значение коэффициента должно было подтверждать системные отношения вариантов ДФ во времени.

Однако на имеющихся у нас небольших данных наша гипотеза пока не подтвердилась статистически. Ожидаемый отрицательный коэффициент мы получили только в тех случаях, когда формулы довольно редки, что позволяет им «не пересекаться» во времени, однако при этом частота их употребления настолько мала, что мы не можем говорить, что полученная статистика свидетельствует о сменяемости структурных элементов.



Рис. 6. Употребление формул *тут уж ничего не поделаешь*, *ничего тут не поделаешь*

Как показывает рис. 6, в течение интересующего нас периода варианты ДФ склонны существовать одновременно, по всей видимости, дополняя друг друга.

Тем не менее, кажется, что пример с формулами *как так* и *как это* не единичный и отражает некоторый глобальный процесс, связанный со слиянием значения формулы с формулой иной структурной организации, которая впоследствии становится более частотной и заменяет первоначальный вариант. Другое дело, что скорость этого процесса в разных случаях может быть разной — и для *как так* и *как это* она выше средней. Это значит, что наша гипотеза не подтвердилась из-за того, что для большинства формул временной промежуток, который был рассмотрен, оказался недостаточен для формальной статистической процедуры<sup>3</sup>.

В связи с этим интересными кажутся рис. 7 и рис. 8, отражающие соотношение вариантов в формулах *да быть не может* / *не может такого быть* и *вот то-то же* / *то-то вот*. Правда, во втором случае оба варианта в итоге оказываются

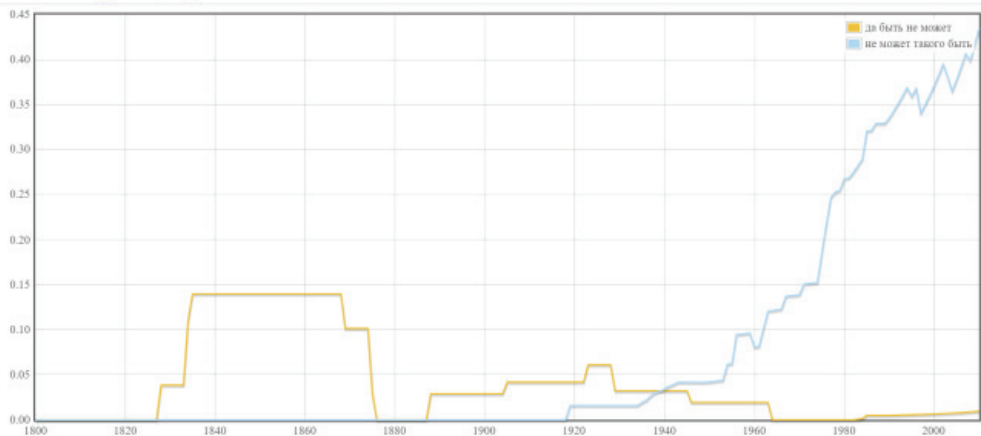
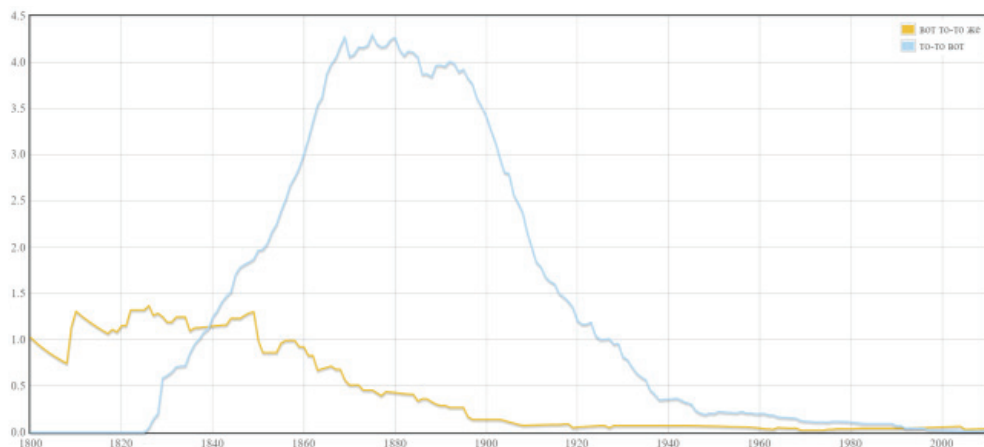


Рис. 7. Частотность употребления ДФ *да быть не может*, *не может такого быть*

<sup>3</sup> Ср. также большое количество «мнимых» нулей, получившихся в нашей таблице из-за округления значения вхождений формул в тех случаях, когда частота ДФ в целом по данным НКРЯ оказывается невысокой.

Рис. 8. Частотность употребления ДФ *от то-то же, то-то от*

устаревающими, однако на определенном этапе их сосуществования заметно доминирование одной единицы над другой.

В целом можно сказать, что высокая скорость прагматикализации ДФ не вызывают сомнений — хотя для их статистической визуализации при том наборе сегодняшних инструментов и при том объеме данных, которые собраны, не всегда хватает мощностей. Материал свидетельствует, что когнитивно значимые формулы не уходят бесследно — они замещаются своими вариантами. Главная задача — это полный анализ имеющегося материала, который позволит описать механизмы такого замещения: установить закономерности, которые лежат в основе процессов устаревания формулы, принципы выбора нового варианта и замены одного варианта на другой.

#### 4. Новые ДФ

Новыми мы называем те ДФ, которые появляются преимущественно во второй половине XX в. или в более ранних текстах, после чего число их вхождений (по данным НКРЯ) продолжает расти, такие как: *не факт*, *без проблем*, *что за дела*, *без вопросов* и проч.

В качестве иллюстрации на рис. 9 представлена динамика употреблений ДФ *не факт* по данным НКРЯ. Из примеров видно, что в качестве формулы *не факт* появляется в 1970-е годы, и наш график показывает, что после этого частота его употребления заметно растёт:

- (9) *А ножны с финкой на правом бедре... Левиша?... Не факт...* [Владимир Богомолов. Момент истины (В августе сорок четвертого...) (1973)].

Само значение слова *факт* включает в себя не только идею соответствия Р действительности (10), как это написано в МАС [Евгеньева 1999], но и компонент оценки говорящего, утверждающий несомненность этого соответствия. Этот компонент объясняет семантику его сдвига в дискурсивную зону: МАС квалифицирует

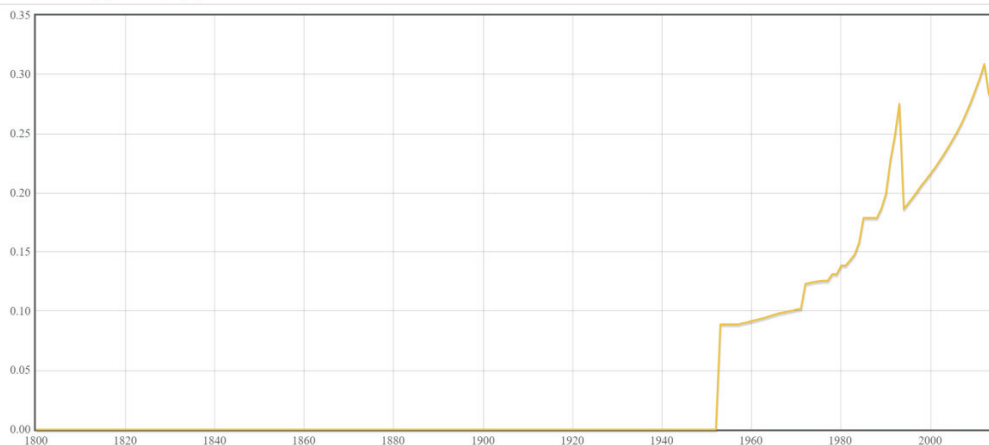


Рис. 9. Частотность употребления ДФ не факт

результатирующую единицу как утвердительную частицу, синонимичную *верно, несомненно* и под. По данным НКРЯ, этот сдвиг произошел в 60-е годы XIX века (0,67 ipm), соответствующие ему изолированные употребления достигли пика в 1930 г. (3,06 ipm), а к концу XX в. значительно сократились (в период с 1980 по 2000 максимум 1 ipm).

- (10) *По крайней мере он хорошо знает этот важный исторический факт.* [А. В. Никитенко. Дневник (1827)].
- (11) — *Ну, не все же пропадает, — возразила Лиза. — Факт. — Я, впрочем, не читала Прудона.* [Н. С. Лесков. Некуда (1864)].
- (12) — *И ты с нами, окурок? — Факт! — улыбнулся Роман и поправил сползающий с плеча карабин.* [Г. Г. Белых. Дом веселых нищих (1930)].
- (13) *И передовик так же говорит... Гнать, говорит, анкаголиков. Они показатели снижают. — Факт.* [Михаил Анчаров. Как Птица Гаруда (1989)].

Между тем, как видно из рис. 9, ДФ не факт ведет себя совершенно иначе.

Первые его употребления (полнозначные, в единственном и множественном числе) — это контексты противопоставления, в которых отрицание относится не к истинности как таковой (и не дает значения ‘ложь’): в сферу действия отрицания попадает оценочный компонент несомненности истины:

- (14) *<...> долгое время были в виду именно все одни только идеи, а не факты, за которые можно было бы ухватиться.* [М. А. Корф. Записки (1838-1852)].

Следующий шаг семантического развития — оценочный предикат с валентностью объекта: говорящий оценивает как недостоверную некоторую ситуацию Р. Р вводится придаточным с союзом *что*; такие контексты появляются в 1920-х годах и резко растут к концу XX в. (с 3 ipm до 11 ipm в начале XXI в.). Морфологический маркер этого этапа — запрет на формы множественного числа (15).



- (15) — *Совсем не факт (\*не факты), что пойдет за другого, — сказала Лена.* [В. Ф. Панова. Спутники (1945)].

Сомнение в утверждении собеседника — это как раз очень характерное значение из принадлежащих к так называемому «универсальному прагматическому набору» [Рахилина и др. 2019] — именно семантика сомнения запускает процесс превращения *не факт* в ДФ с этим значением. В конечном счете приобретение прагматического значения несогласия с предположением собеседника маркируется изменением синтаксиса: формула изолируется, объектную валентность можно не постулировать — по крайней мере она не выражается поверхностно, поскольку предположение становится частью дискурса другого участника диалога (16), (17).

- (16) — *Если б знал — поставил бы на него, что ли? — Не факт, — сказала Агата.* [Андрей Лазарчук, Михаил Успенский. Посмотри в глаза чудовищ (1996)].
- (17) *Вот у кого их больше, тот и возглавит совет кредиторов. А значит, и протащит в арбитражном суде свою кандидатуру. — Тогда это будет банк «Орбита»! — Не факт! — не согласился Коломнин. — У банка всего пять миллионов. — Хорошенькое «все».* [Семен Данилюк. Бизнес-класс (2003)].

Как видим, слово *факт* дало жизнь двум (антонимичным) дискурсивным единицам: *Факт!* и *Не факт!* Первая появилась в середине XIX в. и уже выходит из употребления, тогда как вторая сформировалась на век позже: только к 70-м годам XX в. она стала использоваться как изолированная ответная реплика со значением несогласия. Однако разница не только в том, когда начался процесс их семантического изменения, но и в ходе самого процесса. Именно для отрицательной формы *не факт* характерна стадия превращения в предикат и приобретения валентности, которая выражается изъяснительным придаточным, и ее потеря как маркер превращения в ДФ. Похожий путь, по-видимому, проделало *не суть*, ср.:

- (18) *Лучшие, какие могут существовать приятели, Александры, не суть лучшие друзья, именно потому **не суть, что** они, как круглые, катятся ко всем, ни в кого не втыкаясь острым ребром, но и ни за кого не зацепляясь.* [П. А. Флоренский. Имена (1926)].
- (19) *Как это работает? **Не суть.** Современные технологии — штука сложная.* [Свое кино (2004)],

а также многие другие новые формулы: их источником становятся полноценные конструкции оценки ситуации, у которых затем объект оценки из придаточного «переезжает» в предшествующую реплику, как бы обрезая исходную конструкцию справа. ср., например (20–23):

- (20) *С чего вы взяли, что получают отрицательный результат?* [коллективный. Форум: Как врачи перелили годовалому ребенку ВИЧ-инфицированную кровь (2012)] → *(Они) получают отрицательный результат. — С чего вы взяли?*

- (21) *Ну и что с того, что параллельному классу досталась Лидия Тимофеевна?* [Андрей Геласимов. Фокс Малдер похож на свинью (2001)] → *Параллельному классу досталась Лидия Тимофеевна. — И что с того?*
- (22) *Трудно сказать, по кому история с «Тремя столицами» и «Трестом» ударила сильнее.* [Иван Толстой. Барин из Парижа // «Русская жизнь», 2012] → *По кому история с «Тремя столицами» и «Трестом» ударила сильнее? — Трудно сказать.*
- (23) *Что поделать: картина пострадала в процессе бытования.* [Дина Рубина. Белая голубка Кордовы (2008-2009)]. → *Картина пострадала в процессе бытования — Что поделать!*

Таким образом, по-видимому, внутри языка существуют какие-то стандартные модели, по которым образуются новые формулы. Однако есть и примеры внешне-го заимствования, прямого калькирования ДФ непосредственно из языка-донора, как, например, это произошло с формулой *нет проблем*:

- (24) <...>*as her friend, I'm going to have to ask you to leave now* <...> “*Dude, no problem whatsoever.* [Lauren Weisberger. The Devil Wears Prada (2003)].
- (25) — *Короче, как ее друг, я прошу тебя сейчас уйти.* <...>. — *Чувак, нет проблем.* [Лорен Вайсбергер. Дьявол носит Прада (перевод М. Маяков, Т. Шабаяева, 2006)].

ДФ *Нет проблем* появляется в текстах НКРЯ в середине 1960-х годов, однако в дальнейшем, по всей видимости, она служит источником для формул *Какие проблемы?* (конец 1970-х) и *Без проблем* (1980-е), которые строятся по уже существующим в языке моделям конструкций и формул, ср., например: *О чем ты говоришь! Какие деньги!*, или: *Без шансов / Без вопросов* и под.

- (26) *Угу. Милорд, вы ведь им покажете? Покажете, да? — Какие проблемы! Сейчас отдохнем немного, перекусим, а потом как начнем мордовать всех подряд. Полетят клочки по закоулочкам! — Это будет только справедливо, — серьезно кивнула Лия, подтягивая к нам еду.* [Андрей Белянин. Свирепый ландграф (1999)].
- (27) — *Ромыч, подбрось меня до церкви, а?* — *попросил Кирилл. — Без проблем. — Ты что, уходишь караулить?* — *удивился Валерий. — А что, Годовалов сейчас прийти не может?* [Алексей Иванов (Алексей Маврин). Псоглавцы (2011)].

Как видим, здесь срабатывает обычный для заимствованных конструкций механизм «синтаксической имитации», который перестраивает их под известные говорящему шаблоны.

## 5. «Стабильные» ДФ

К «стабильным» относятся дискурсивные формулы, которые продолжают использоваться на протяжении всего рассматриваемого периода, при этом сюда попадают и те ДФ, частота употребления которых может в значительной степени колебаться. В данном случае нам важно, что они не демонстрируют выраженной тенденции к спаду или увеличению частотности: в начале и в конце рассматриваемого периода они оказываются в практически равной степени частотными.

Например, в эту группу попадают слова, употребление которых оказывается неравномерным (*представь себе, что это со мной* и проч.), поэтому в определенный период количество текстов, включающих эти ДФ снижается. Так, в начале XIX в. частота единицы *представь себе* составляет 0,15 ipm, далее в 1920-е гг. мы наблюдаем спад до 0,04 ipm, но к началу XXI в. частота этого сочетания составляет уже 0,19 ipm.

Первоначально, как и ожидалось, данное словосочетание употребляется в прямом значении:

- (28) *Сколько новых ощущений, какой разнообразный мир забав, радостей, удовольствий ожидают тебя в этом роскошном, обольстительном Париже! **Представь себе**... Вдруг барон замолчал, он поглядел робко вокруг себя и, схватив меня за руку, проговорил торопливо:* [М. Н. Загоскин. Искуситель (1838)].

В 1920-е гг., когда мы наблюдаем спад употребления этой единицы, мы видим, что она уже оформилась как ДФ и используется как ответная реплика, служащая для подтверждения своих слов в предыдущем сегменте диалога:

- (29) — *Какой Орлов? — Да наш чекист знаменитый! — Что ты говоришь? — **Представь себе!** Поймали вчера на вокзале. Понесу-ка газетку Маргоше. Пусть отвлечется немножко.* [Б. А. Лавренев. Рассказ о простой вещи (1924)].

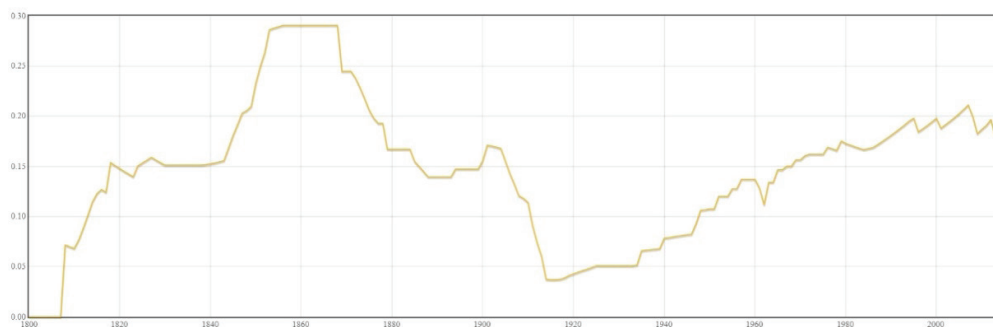


Рис. 10. Динамика употребления формулы *представь себе*

Далее с 1970-х гг. ДФ *представь себе* начинает выражать просто подтверждение говорящим того, что удивляет его собеседника, заставляет его недоумевать:

(30) — *Что, подружились с ним за эту поездку?* — *спросил про Василия Ивановича Лева Степанов.* — **Представь себе.** — *Вы с ним или он с вами?* — **Представь себе,** и я с ним, и он со мной. — *По-моему, вы первый — во всяком случае, у нас в редакции.* [Константин Симонов. Так называемая личная жизнь/ Мы не увидимся с тобой... (1978)].

(31) *Галя и так повсюду смотрит, был ты или нет? Думаешь, она не знала, что ты придешь?* — *А ты знала?* — **Представь себе.** Да, знала, и я это чувствовал, я шел на зов. [Г. Я. Бакланов. Мой генерал // «Знамя», 1999].

Количество стабильных единиц, по результатам работы программы, составляет менее 10% от общего числа ДФ, однако нужно учесть, что эта группа выделена исключительно по формальному признаку неизменности частотности. Пример с *представь себе* свидетельствует, что она нуждается в дополнительном лингвистическом исследовании, потому что такой «стабильный» характер существования может оказаться иллюзией.

## 6. Заключение и перспективы исследования

В этой статье рассмотрен особый класс конструкций русского языка — класс дискурсивных формул — в целом, с диахронической точки зрения. ДФ носят преимущественно разговорный характер и достаточно эмфатичны — этим обусловлены два их важнейших свойства: высокая вариативность и необыкновенно высокая скорость изменения. Каждое из этих свойств имеет интересные в лингвистическом отношении аспекты, а кроме того, они тесно связаны друг с другом.

С одной стороны, по определению, ДФ имеют жесткую структуру и не содержат слотов, заполнение которых могло бы способствовать ее изменению. Одновременно, как и любые другие типы фразем [Баранов, Добровольский 2008], они допускают дополнительные эмфатические частицы и замену одних эмфатических компонентов на другие. Для ДФ как прагматических единиц такая вариативность особенно значима. С учетом этих возможностей, ДФ функционируют «семействами» (и для них это более характерно, чем для обычных конструкций), причем внутри семейства, ввиду того, что эти единицы гораздо более изолированы, чем обычные конструкции, сосуществуют полностью тождественные в отношении семантики и прагматики единицы. Это подтверждает наша статистика: часто у этих вариантов, по крайней мере некоторое время, нет очевидных статистических преимуществ друг перед другом.

С другой стороны, ДФ очень динамичны. Среди них почти нет стабильных, даже на временном отрезке в чуть более 200 лет можно увидеть оформление композиционного словосочетания в фиксированную в отношении формальной структуры единицу, и сопровождающее этот процесс вымывание первоначальной семантики и ее прагматикализацию. На том же выбранном нами небольшом

временном отрезке можно наблюдать и устаревание формулы, и исчезновение из языка, а иногда и полный ее жизненный цикл, который, как это и должно быть у эмфатических единиц, недолог.

Общие контуры динамических процессов, которые были здесь проиллюстрированы, можно описать так. Формула может устаревать, во-первых, за счет устаревания определенного лексического компонента, а во-вторых, за счет «стирания» данной комбинации служебных элементов. Тогда ей на смену приходит структурно близкий вариант или более высокочастотный «член семейства». В последнем случае новая формула, которая ее замещает, состоит практически из тех же элементов. Что касается структурно новых ДФ, то для них есть разные источники, и здесь мы обсудили два: внутриязыковой и внеязыковой. Внутренним источником ДФ служат полноценные конструкции данного языка, в частности конструкции оценки некоторого события, его достоверности, вероятности, последствий и проч. Такого рода конструкции легко разделяются на презумптивную часть, которая становится предшествующей репликой собеседника, и ассертивную, которая прагматизируется и превращается в формулу. Внешним источником служит заимствование и калькирование, в ходе которого, однако, как мы показали, все равно включаются внутренние механизмы русского языка.

Ближайшая задача нашего исследования — уточнить детали исторических изменений, которые претерпевают ДФ, представляющие разные структурные и семантические классы.

## Литература

Баранов А. Н., Добровольский Д. О. Аспекты теории фразеологии. М. : Знак, 2008.

Булаховский Л. А. Русский литературный язык первой половины XIX века. Киев : Ряданьска школа, 1948.

Бычкова П. А., Рахилина Е. В., Слепак Е. А. Дискурсивные формулы, полисемия и жестовое маркирование // Труды Института русского языка им. В. В. Виноградова (в печати)

Виноградов В. В., Шведова Н. Ю. Очерки по исторической грамматике русского литературного языка XIX века. М. : Наука, 1964.

*Два века в двадцати словах* / Отв. ред.: М. А. Даниэль, Н. Р. Добрушина. М.: Изд. дом НИУ ВШЭ, 2016.

Добровольский Д. О. К динамике узуса (язык Пушкина и современное словоупотребление) // Русский язык в научном освещении. 2001. № 1. С. 161–178.

Зализняк Анна А. Об эффекте ближней семантической эволюции // PHILOLOGICA. Двухязычный журнал по русской и теоретической филологии. 2012. Т. 9. № 21/23. С. 11–22.

Падучева Е. В. Из наблюдений над языком Л. Толстого (к вопросу о малых диахронических сдвигах) // Вопросы языкознания. 2018. № 5. С. 49–63. DOI: 10.31857/S0373658X0001396-4.

Пеньковский А. Б. Загадки пушкинского текста и словаря. Опыт филологической герменевтики. М.: Языки славянских культур, 2005.

Рахилина Е. В. Говорю я, Карл // Компьютерная лингвистика и интеллектуальные технологии. По материалам международной конференции «Диалог'2017». Вып. 16 (23). М.: РГГУ, 2017. С. 361–369.

Рахилина Е. В., Кузнецова Ю. Л. Грамматика конструкций: теория, сторонники, близкие идеи // Лингвистика конструкций / Под ред. Е. В. Рахилиной. М.: Азбуковник, 2010. С. 18–79.

Рахилина Е. В., Резникова Т. И., Бородин М. А. «Тамань сегодня»: корпусное исследование русского языка XIX века. Труды Института русского языка им. В. В. Виноградова, Вып. 11, 2016.

Русский язык: конструкционные и лексико-семантические подходы 2017 (RusConstr). Тезисы конференции 2017 г. URL: <https://iling.spb.ru/confs/rusconstr2017/papers.html>

Словарь русского языка. В 4-х т. / Под ред. А. П. Евгеньевой. 4-е изд., стер. М., 1999.

Шаронов И. А. Коммуникативы как функциональный класс и как объект лексикографического описания // Русистика сегодня, №2, 1996, с. 89-112.

Aijmer, K. Conversational Routines. Convention and Creativity. Longman, London. 1996.

Coulmas, F. Conversational Routine: Explorations in Standardized Communicative Situations and Prepatterned Speech. The Hague, Mouton. 1981.

Fillmore, Ch. J. The mechanisms of “construction grammar” // Annual Meeting of the Berkeley Linguistics Society. Berkeley, 1988. Vol. 14. P. 35–55.

Fillmore, Ch. J., Kay, P. Construction Grammar Course Book. Berkeley: University of California, 1992.

Fillmore, Ch. J. Remarks on contrastive pragmatics // Contrastive linguistics: Prospects and problems. 1984. P. 119–141.

Gerasimenko E., Puzhaeva S., Zakharova E., Rakhilina E. Defining discourse formulae: computational approach (in press)

Hilpert, M. Constructional change in English: Developments in allomorphy, word formation, and syntax. — Cambridge University Press, 2013.

Hoffmann, T., Trousdale, G. (Eds.). The Oxford handbook of construction grammar. Oxford: Oxford University Press, 2013.

Hopper, P. J., Traugott, E. C. Grammaticalization. Cambridge: Cambridge University Press, 2003 [1993].

Janda, L. A., Lyashevskaya, O., Nessel, T., Rakhilina, E., Tyers, F. M. A Constructicon for Russian: Filling in the Gaps. Constructicography: Constructicon development across languages. Amsterdam: John Benjamins. 2018. P. 165–182.

Kuzmenko E., Kutuzov A. Two centuries in two thousand words: neural embedding models in detecting diachronic lexical changes // Quantitative Approaches to the Russian Language. Routledge, 2017. P. 105–122.

Lubimovka (2018). The Lubimovka Young Russian Playwrights Festival. Available at <http://lubimovka.ru/piesy> (accessed January 2019).

Mey, J. Pragmatics. Blackwell, Oxford. 2001.

Russian Drama Corpus (2018). Available at <https://github.com/lehkost/RusDraCor/tree/master/tei> (accessed January 2019).

Wray, A. Formulaic Language and the Lexicon. Cambridge: Cambridge University Press. 2002.

<sup>1</sup>S. Zhukova, <sup>1</sup>B. Orekhov, <sup>1,2</sup>E. Rakhilina

<sup>1</sup> National Research University Higher School of Economics

<sup>2</sup> V. V. Vinogradov Russian Language Institute of the Russian Academy of Sciences  
(Russia, Moscow)

[syupuzhaeva@gmail.com](mailto:syupuzhaeva@gmail.com), [nevmenandr@gmail.com](mailto:nevmenandr@gmail.com), [rakhilina@gmail.com](mailto:rakhilina@gmail.com)

## DISCOURSE FORMULAE IN RUSSIAN: DIACHRONIC APPROACH

This article deals with the description of Russian discourse formulae in diachronic perspective. Discourse formulae are defined as easily reproduced phraseological constructions mostly used as isolated responds in the dialogs. Unlike traditional constructions, they do not have variables in their structure. The preceding remark of an interlocutor is what serves as a construction slot. Being emphatic and semantically bleached, discourse formulae are extremely dynamic diachronically, the changes of their form and semantics are observable even within a short span of time, like two centuries. To get this diachronic information the list of Russian discourse formulae relevant for 19-20 centuries was arranged statistically by frequencies for every 10-years. Examples of old, obsolete formulae, new formulae which appeared recently and those which seem to be stable in frequency for the last two centuries, have been thoroughly analyzed. Through these examples one could see the mechanisms of pragmatization of a compositional word combination (i.d. turning it into fixed idiomatic expressions with pragmatic meanings), and also the reasons for formulae becoming obsolete and the origin of their structural and semantic instability.

*Key words:* Discourse formulae, diachronic studies, corpus linguistics, Construction grammar, Russian Constructicon, discourse of drama

## References

Aijmer, K. *Conversational Routines. Convention and Creativity*. Longman, London. 1996.

Baranov A. N., Dobrovol'skij D. O. *Aspekty teorii frazeologii* [Aspects of the theory of phraseology]. Moscow, "Znak" Publ., 2008. (In Russ.)

Bulakhovskii L. A. *Russkii literaturnyi yazyk pervoi poloviny XIX veka* [The Russian literary language of the first half of the 19 century]. Kyev, "Ryadan'ska shkola" Publ., 1948. (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 V. V. Vinogradov Russian language Institute] (in press)

Coulmas, F. *Conversational Routine: Explorations in Standardized Communicative Situations and Prepatterned Speech*. The Hague, Mouton. 1981.

Dobrovol'skij D. O. [On the dynamics of usage (Pushkin's language and modern usage)]. *Russkii yazyk v nauchnom osveshchenii*. 2001, no. 1, pp. 161–178. (In Russ.)

*Dva veka v dvadtsati slovakh* [Two centuries in twenty words] / M. A. Daniehl', N. R. Dobrushina (eds.). Moscow, NRU Higher School of Economics Publ. House, 2016. (In Russ.)

Fillmore, Ch. J. The mechanisms of “construction grammar” // *Annual Meeting of the Berkeley Linguistics Society*. Berkeley, 1988. Vol. 14, pp. 35–55.

Fillmore, Ch. J., Kay, P. *Construction Grammar Course Book*. Berkeley: University of California, 1992.

Fillmore, Ch. J. Remarks on contrastive pragmatics // *Contrastive linguistics: Prospects and problems*. 1984, pp. 119–141.

Gerasimenko E., Puzhaeva S., Zakharova E., Rakhilina E. *Defining discourse formulae: computational approach* (in press).

Hilpert M. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge University Press, 2013.

Hoffmann T., Trousdale G. (Eds.). *The Oxford handbook of construction grammar*. Oxford: Oxford University Press, 2013.

Hopper P. J., Traugott E. C. *Grammaticalization*. Cambridge: Cambridge University Press, 2003 [1993].

Janda, L. A., Lyashevskaya, O., Nessel, T., Rakhilina, E., Tyers, F. M. A Constructicon for Russian: Filling in the Gaps. *Constructicography: Constructicon development across languages*. Amsterdam: John Benjamins. 2018, pp. 165–182.

Kuzmenko E., Kutuzov A. Two centuries in two thousand words: neural embedding models in detecting diachronic lexical changes // *Quantitative Approaches to the Russian Language*. Routledge, 2017, pp. 105–122.

*The Lubimovka Young Russian Playwrights Festival*. Available at <http://lubimovka.ru/piesy> (accessed January 2019).

Mey, J. *Pragmatics*. Blackwell, Oxford. 2001.

Paducheva E. V. [Some remarks on the language of Leo Tolstoy and small-scale diachronic shifts]. *Voprosy Yazykoznanija*. 2018, no. 5, pp. 49–63. DOI: 10.31857/S0373658X0001396-4. (In Russ.)

Pen'kovskii A. B. *Zagadki pushkinskogo teksta i slovarya. Opyt filologicheskoi ger-menevtiki* [Mysteries of Pushkin's text and vocabulary. An attempt of philological hermeneutics]. Moscow, Yazyki Slavyanskikh Kul'tur Publ., 2005. (In Russ.)

Rakhilina E. V. [I say, Carl]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Po materialam mezhdunarodnoi konferentsii “Dialog’2017”*. [Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”]. No. 16 (23). Moscow, Russian State Univ. for the Humanities, 2017, pp. 361–369. (In Russ.)

Rakhilina E. V., Kuznetsova Yu. L. [Introduction. Construction Grammar: Theories, Adherents, Similar Approaches]. E. V. Rakhilina (Ed.). *Lingvistika konstruktiv* [Construction linguistics]. Moscow: Azbukovnik Publ., 2010, pp. 18–79. (In Russ.)

Rakhilina E. V., Reznikova T. I., Borodina M. A. [“Taman today”: corpus case study of the Russian language of the XIX<sup>th</sup> century]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of V. V. Vinogradov Russian language Institute]. Vol. 11. 2016, pp. 240–253. (In Russ.)

*Russian Drama Corpus*. Available at <https://github.com/lehkost/RusDraCor/tree/master/tei> (accessed January 2019).

*Russian language: constructional and lexical-semantic approaches* (RusConstr) 2017. Abstracts. Available at <https://iling.spb.ru/confs/rusconstr2017/papers.html>

Sharonov I. A. [Kommunikatives as a functional class and the objekt of lexicographical description]// *Rusistika segodnia* [Russian studies today]. 1996, no. 2, pp. 89–112. (In Russ.)

*Slovar' russkogo iazyka* [The Russian language dictionary]. In 4 v. / A. P. Evgen'eva (ed.). Moscow, 1999. Available at <http://feb-web.ru/feb/mas/> (In Russ.)

Vinogradov V. V., Shvedova N. Yu. *Ocherki po istoricheskoi grammatike russkogo literaturnogo yazyka XIX veka*. [Essays on the historical grammar of the Russian literary language of the 19 century]. Moscow, Nauka Publ., 1964. (In Russ.)

Wray, A. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. 2002.

Zaliznyak Anna A. [On the effect of small-scale semantic evolution]. *PHILOLOGICA. Dvuyazychnyi zhurnal po russkoi i teoreticheskoi filologii* [PHILOLOGICA. Bilingual journal on Russian and theoretical Philology]. 2012. Vol. 9, no. 21/23, pp. 11–22. (In Russ.)

**<sup>1</sup>В.П. Захаров, <sup>2</sup>А.Ц. Масевич**

<sup>1</sup>*Санкт-Петербургский государственный университет*

<sup>2</sup>*Санкт-Петербургский государственный институт культуры*

*v.zakharov@spbu.ru, andmasev@mail.ru*

*(Россия, Санкт-Петербург)*

## **КАТЕГОРИЗАЦИЯ ПРИЛАГАТЕЛЬНЫХ ЦВЕТА В РУССКИХ ПОЭТИЧЕСКИХ ТЕКСТАХ (КОРПУСНОЕ ИССЛЕДОВАНИЕ)**

Статья описывает некоторые количественные данные использования прилагательных, обозначающих цвета в русских поэтических текстах. Предложена методика выявления прилагательных цвета из текстов поэтического корпуса НКРЯ посредством использования тегов семантической разметки. Произведено сравнение частотности прилагательных цвета в четырех корпусах НКРЯ, при этом выявлено, что частота прилагательных цвета в текстах поэтического корпуса значительно выше, чем в текстах остальных трех корпусов. На основе анализа 170 лемм прилагательных цвета, извлеченных из текстов поэтического корпуса, произведена категоризация полученного списка прилагательных. Приведены количественные данные о представленности (репрезентации) выделенных типов прилагательных в русских поэтических текстах. Выявлена заметная корреляция частоты встречаемости 11 прилагательных, означающих базовые цветовые категории, с эволюционной теорией базовых цветовых терминов Берлина — Кэя.

*Ключевые слова:* прилагательные цвета, теория базовых цветовых терминов Берлина — Кэя, Национальный корпус русского языка, поэтический корпус, семантическая разметка.

### **Введение**

Обозначение цвета в языке — исключительно популярная тема в лингвистических исследованиях. В данной работе мы пытаемся ввести в рассмотрение несколько принципиально новых аспектов лингвистической статистики с применением лингвистических корпусов.

Цели исследования:

1) выявление возможной корреляции между эволюционной теорией базовых цветовых терминов Брента Берлина и Пола Кэя (Brent Berlin & Paul Kay) [Berlin, Kay 1969] и статистикой частоты встречаемости прилагательных цвета;

2) изучение частотного поведения единиц данной лексической группы на материале поэтических текстов в сравнении с другими типами текстов;

3) систематизация прилагательных цвета.

Исследование выполнялось по следующему плану:

1. Выявление и анализ частотности прилагательных цвета в четырех корпусах Национального корпуса русского языка (НКРЯ).

2. Создание авторских подкорпусов для отдельных авторов.

3. Выявление количества прилагательных цвета, встречающихся в текстах отдельных авторов.

4. Систематизация прилагательных цвета на основе полученных данных.

В качестве материала исследования использовались:

1) поэтический корпус НКРЯ и авторские корпуса, созданные на его основе;

2) корпуса поэтических текстов, созданные специально для целей исследования в системе SketchEngine на базе различных источников;

3) основной, газетный и устный корпуса НКРЯ.

В качестве инструментов исследования использовались:

1) поисковый аппарат (корпусный менеджер) НКРЯ, в том числе поиск по семантическому признаку;

2) корпусный менеджер SketchEngine;

3) электронная таблица MS Excel 2010.

### **Теория базовых цветовых терминов Берлина и Кэя**

Теория базовых цветовых терминов (категорий) Б. Берлина и П. Кэя состоит в том, что языки мира содержат полностью или частично общий набор слов цветообозначений, и эти слова в процессе исторического развития появляются в языках в определенном порядке. В английском языке авторы выделяют 11 базовых прилагательных цвета, в славянских, в частности в русском, таких прилагательных 12.

Исследовав десятки европейских языков, с одной стороны, и десятки языков целого ряда примитивных культур — с другой, Берлин и Кэй открыли «всеобщий» эволюционный закон. По их мнению, существует семь ступеней (стадий) развития цветовой терминологии, отражающих последовательность появления в лексике языка каждого такого слова [Berlin, Kay 1969; Kay, McDaniel 1978; Kay, Regier 2006].

Критерии принадлежности к базовым категориям следующие:

1) слово должно быть непроеизводным и не относиться к сложным словам;

2) его значение не должно быть уже значения другого имени цвета, указывающего на какой-либо близкий оттенок;

3) слово должно обладать широкой сочетаемостью;

4) для носителей данного языка слово должно быть психологически выделенным, значимым (salient).

На «самой низшей» стадии, стадии I, в языке есть только два основных цветообозначения, и это во всех случаях оказываются слова, указывающие на белый и черный цвета. На стадии II к двум цветам добавляется третий, и это всегда слово,

обозначающее красный цвет. На стадиях III—V добавляется каждый раз по одному слову из трех — «синий», «зеленый», «желтый». Седьмым словом (стадия VI) всегда бывает «коричневый», а высшая стадия VII характеризуется появлением четырех цветов — «розовый», «оранжевый», «фиолетовый» и «серый» [Василевич, Кузнецова, Мищенко 2005].

## Анализ прилагательных цвета по четырем корпусам НКРЯ

Одной из замечательных особенностей НКРЯ, отличающей его от многих других корпусов, является наличие семантической разметки [Кустова и др. 2005]. Ее назначение — выделение (поддержка поиска) групп лексических единиц, объединенных общими семантическими признаками. Семантические признаки приписываются всем словоформам знаменательных слов корпуса. При поиске семантические признаки задаются с помощью вкладок — меню, которые раскрываются по ссылке на входной странице поиска «Семантические признаки: выбрать». Существует восемь меню с перечнем семантических признаков для пяти частей речи. Существительным соответствуют три меню — для имен существительных предметных, непредметных и собственных. Семантический признак «цвет» может быть задан для трех частей речи — прилагательных, непредметных существительных и наречий (рис. 1).

Семантические признаки - Mozilla Firefox

www.ruscorpora.ru/reqsem.html

Имена: [предметные](#) | [непредметные](#) | [собственные](#) | **[Прилагательные](#)** | [Числительные](#) | [Местоимения](#) | [Плаголы](#) | [Наречия](#)

**Прилагательные**

**Разряд**

☐ качественные

☐ относительные

☐ приткательные

☐ неизменяемые

**Таксономия**

☐ размер в том числе:

☐ большой ☐ малый

☐ абсолютный

☐ расстояние в том числе:

☐ большое ☐ малое

☐ количество в том числе:

☐ большое ☐ малое

☐ абсолютное

☐ место

☐ направление

☐ время в том числе:

☐ длительность в том числе:

☐ большая ☐ малая

☐ абсолютная

☐ возраст в том числе:

☐ большой ☐ малый

☐ абсолютный

☐ скорость в том числе:

☐ большая ☐ малая

☐ физические свойства в том числе:

☐ форма

☐ цвет

☐ вкус

☐ запах

☐ температура

☐ качества человека

☐ Оценка в том числе:

☐ положительная

☐ отрицательная

**Словообразование**

☐ диминутивы

☐ аугментативы

☐ аттенуативы

☐ хабитивы

☐ каритивы

☐ потенциальные в том числе:

☐ лососибилитивы

☐ импосибилитивы

☐ отыменные прилагательные в том числе образованные от имен со значениями:

☐ лица

☐ животные

☐ растения

☐ вещества и материалы

☐ пространство и место

☐ здания и сооружения

☐ инструменты и приспособления

☐ еда и напитки

☐ тексты

☐ части тела и органы

☐ отлагательные прилагательные

☐ отадавербальные прилагательные

Рис. 1. Меню выбора семантических признаков для прилагательных

Таблица 1

**Результат поиска по семантическому признаку «цвет» в четырех корпусах**

Корпус	Число слово-употреблений (вхождений)	ipm
Поэтический корпус	85286	7776,48
Основной корпус	771638	2722,48
Газетный корпус	280630	1228,02
Устный корпус	12682	1046,93

На первом этапе мы отбирали все контексты с прилагательными цвета в четырех корпусах: поэтическом, основном, газетном и устном, — опираясь именно на семантический тег «Физические свойства: цвет». Результаты представлены в таблице 1.

Анализ данных таблицы показывает, что значение ipm прилагательных,

означающих цвет, в поэтическом корпусе в 2,9 раза выше, чем в основном корпусе, в 6,3 раза выше, чем в газетном корпусе, и в 7,4 раза выше, чем в устном корпусе.

Для сопоставления частотности цветообозначающих прилагательных отобраны 11 прилагательных, которые в известных экспериментах Берлина и Кэя назывались базовыми цветовыми категориями.

Результаты поиска по этим прилагательным по четырем корпусам приведены в таблице 2. Кроме того, в таблице приводится частотность отобранных прилагательных по Частотному словарю современного русского языка (ЧССРЯ) [Ляшевская, Шаров 2009].

Таблица 2

**Частотность 11 базовых цветовых категорий в четырех корпусах и в Частотном словаре**

Лемма	Поэтический корпус		Основной корпус		Газетный корпус		Устный корпус		ЧССРЯ
	слово-употр.	ipm	слово-употр.	ipm	слово-употр.	ipm	слово-употр.	ipm	ipm
белый	10730	978,37	111109	392,01	54079	236,65	2115	174,60	339,6
черный	9556	871,33	106635	376,23	42210	184,71	2286	188,72	337,5
красный	5582	508,97	86762	306,11	47301	206,99	1958	161,64	240,5
синий	4951	451,44	29371	103,63	8722	38,17	450	37,15	82,8
зеленый	4871	444,14	37891	133,69	17006	74,42	854	70,50	124,4
серый	2692	245,46	38297	135,12	7753	33,93	453	37,40	96,1
желтый	2423	220,93	26833	94,67	8445	36,95	388	32,03	78,1
розовый	1883	171,69	16436	57,99	3938	17,23	216	17,83	49,3
оранжевый	260	23,71	3944	13,92	6825	29,87	67	5,53	15,60
коричневый	181	16,50	6960	24,56	1491	6,52	110	9,08	25,5
фиолетовый	125	11,40	3325	11,73	727	3,18	57	4,71	12,4
среднее		358,54		149,97		78,96		67,20	127,44

Мы видим, что частота встречаемости прилагательных цвета в поэтическом корпусе заметно выше, чем их частота в других корпусах. При этом разница в ipm прилагательных цвета между поэтическим и основным корпусом меньше, чем между поэтическим и газетным и устным корпусами. Это легко объясняется тем, что

основной корпус сбалансирован по жанровому составу. Из таблицы 2 видно, что значение *ipm* базовых категорий в основном корпусе близко к значениям ЧССРЯ.

Для 11 прилагательных цвета статистическая картина встречаемости их в четырех корпусах НКРЯ в целом повторяет данные, полученные при поиске по семантическому тегу «цвет» (см. табл. 1), что хорошо видно как по отдельным цветам, так по средним значениям *ipm*. Лишь в двух случаях (прилагательные «коричневый» и «фиолетовый») *ipm* в основном корпусе незначительно превышает *ipm* в поэтическом корпусе, и в одном случае в газетном корпусе *ipm* прилагательного «оранжевый» превышает частотность этого же прилагательного в поэтическом корпусе.

Обратим внимание на некоторые различия в порядке снижения частотности прилагательных. Для этого расположим слова в порядке убывания частоты для каждого корпуса (табл. 3). Под таблицей расположим эволюционную схему Берлина — Кэй.

Таблица 3

Порядок убывания частоты прилагательных цвета в четырех корпусах НКРЯ в сопоставлении с эволюционной схемой Берлина — Кэй

Корпус Номер	1	2	3	4	5	6	7	8	9	10	11
Поэтический	белый	черный	красный	синий	зеленый	серый	желтый	розовый	оранжевый	коричневый	фиолетовый
Основной	белый	черный	красный	серый	зеленый	синий	желтый	розовый	коричневый	оранжевый	фиолетовый
Газетный	белый	красный	черный	зеленый	синий	желтый	серый	оранжевый	розовый	коричневый	фиолетовый
Устный	черный	белый	красный	зеленый	серый	синий	желтый	розовый	коричневый	оранжевый	фиолетовый

Берлин — Кэй	Стадия I		Стадия II	Стадия III— IV		Стадия V	Стадия VI	Стадия VII			
Берлин — Кэй	белый	черный	красный	зеленый	желтый	синий	коричневый	серый	розовый	оранжевый	фиолетовый

Как видно из таблицы 3, из 11 отобранных прилагательных в трех корпусах самым частотным оказывается «белый», в одном случае (устный корпус) — «черный». Во всех четырех корпусах три прилагательных «белый», «черный» и «красный» составляют тройку самых частотных. Следующие по частоте — четыре прилагательных «синий», «серый», «зеленый» и «желтый». Наименее частотными оказываются прилагательные «розовый», «оранжевый», «коричневый» и «фиолетовый». Порядок следования этих прилагательных по убыванию частоты в целом коррелирует со стадиями появления слов-цветообозначений в различных языках



по Берлину — Кэю, хотя прямо не совпадает с ними. Например, прилагательное «желтый» при полной корреляции должно было бы быть более частотным, чем прилагательное «синий», а по данным НКРЯ его частотность ниже «синего». Также прилагательное «серый», относящееся к седьмой, последней, исторической стадии, по нашим данным имеет сравнительно высокую частотность и соответствует стадиям IV и V.

Еще два наблюдения. Во-первых, порядок убывания частотности во всех корпусах различается не очень существенно. Можно, как описано выше, разбить массив из 11 прилагательных на три группы — в первой три прилагательных, в двух других по четыре. Обращает на себя внимание, что в пределах одной группы порядок убывания частотности у прилагательных может различаться, но выявлен всего один случай «выхода» за пределы группы. Во-вторых, частотность прилагательных «белый», «черный» и «красный» почти полностью коррелирует с первыми двумя стадиями эволюционной схемы.

### Создание авторских корпусов

Корпус поэтических текстов включает стихотворные произведения от XVIII в. до современности. Помимо обычной семантической и морфологической разметки, имеется специальная стиховедческая разметка [Гришина и др. 2009]. Для создания авторских корпусов были отобраны наиболее известные поэты, представляющие различные хронологические периоды. При этом выяснилось, что тексты некоторых показательных авторов, в основном второй половины XX в., в НКРЯ отсутствуют. В связи с этим авторские корпуса этих поэтов были созданы с помощью системы SketchEngine на основе текстов, представленных на авторитетных веб-сайтах.

Таблица 4

#### Список авторов с указанием источников текстов для создания авторских корпусов

Автор	Источник
1. Аронзон	НКРЯ
2. Асадов	Интернет
3. Ахмадулина	Интернет
4. Ахматова	НКРЯ
5. Бальмонт	НКРЯ
6. Баратынский	НКРЯ
7. Блок	НКРЯ
8. Бродский	Интернет
9. Вертинский	НКРЯ
10. Вознесенский Андрей	Интернет

Автор	Источник
11. Высоцкий	Интернет
12. Державин	НКРЯ
13. Евтушенко	Интернет
14. Есенин	НКРЯ
15. Жданов	НКРЯ
16. Заболоцкий	НКРЯ
17. Кибиров	Интернет
18. Кривулин	НКРЯ
19. Лермонтов	НКРЯ
20. Мандельштам	НКРЯ

Автор	Источник
21. Матвеева Новелла	Интернет
22. Маяковский	НКРЯ
23. Набоков	НКРЯ
24. Некрасов	НКРЯ
25. Пастернак	НКРЯ
26. Пушкин	НКРЯ
27. Рубцов	НКРЯ
28. Светлов	НКРЯ

Автор	Источник
29. Симонов	НКРЯ
30. Слуцкий	НКРЯ
31. Тарковский	НКРЯ
32. Твардовский	НКРЯ
33. Толстой А. К.	НКРЯ
34. Тютчев	НКРЯ
35. Хлебников	НКРЯ
36. Цветаева	НКРЯ

## Методика исследования

### *Создание таблиц частотности прилагательных цвета для отдельных авторов средствами НКРЯ*

По созданному авторскому корпусу выполняется запрос на прилагательные с семантическим признаком «цвет». В НКРЯ это осуществляется посредством вкладки, описанной выше (см. рис.1). Следующий шаг — статистическая обработка полученных результатов. Идеальный вариант — получить сразу всю статистику по данному запросу — как правило, не проходит из-за ограничений на объем выдачи (рис. 2). Опытным путем мы установили, что максимальное число документов, отражаемых на одной странице, не должно превышать 100. Число примеров в одном документе мы ограничивали числом 25 (см. рис. 2).

Рис. 2. Параметры представления результатов при поиске в авторском корпусе НКРЯ

В нижней части каждой страницы с результатами поиска находится раздел, который называется «Частоты найденного для этой страницы» (рис. 3).

С каждой страницы поиска производится перенос всех таблиц «Леммы» в MS Excel, где формируется сводная таблица на весь результат поиска по данному авторскому подкорпусу. Средствами MS Excel осуществляется подсчет числа употреблений каждой леммы, ее доли в процентах среди других лемм цветообозначения, а также относительной частоты (ipm) по отношению к числу словоупотреблений во всем авторском подкорпусе.

Формула подсчета доли леммы данного прилагательного цвета среди лемм всех прилагательных цвета в данном подкорпусе следующая:

**Частоты найденного для этой страницы**

Словоформы			Леммы		
1	яркие	3	1	яркий	4
2	яркими	1	2	красный	4
3	холодном	1	3	синий	2
4	теплый	1	4	светлый	2
5	темный	1	5	лимонный	2
6	синим	1	6	золотой	2
7	синее	1	7	холодный	1
8	серые	1	8	теплый	1
9	светлому	1	9	темный	1
10	светлое	1	10	серый	1
11	розовыми	1	11	розовый	1
12	лимонный	1	12	зеленый	1
13	лимонного	1			
14	красным	1			
15	красный	1			
16	красное	1			
17	класное	1			
18	золотым	1			
19	золотое	1			
20	зеленый	1			

Рис. 3. Статистика запроса — частоты лемм и словоформ

пусе, А — общее число словоупотреблений в исследуемом авторском корпусе.

В результате описанных операций создается таблица частотности прилагательных цвета в корпусе данного автора (рис. 4).

**Создание таблиц частотности прилагательных цвета для отдельных авторов средствами SketchEngine**

В системе SketchEngine нет семантической разметки, поэтому методика обработки данных отличалась от описанной выше для корпусов НКРЯ. Запрос задавался в явном виде, а именно перечислением всех прилагательных цвета, полученных на основе данных НКРЯ. В связи с ограничением на длину запроса проводилось два сеанса поиска: по простым (однокоренным) (рис. 5) и сложным (двусоставным) (рис. 6) прилагательным.

1	Лемма	автор	Число с/у	IPM	%
2	белый	Блок	225	1982,5	13,69
3	темный	Блок	193	1700,55	11,74
4	черный	Блок	182	1603,62	11,07
5	светлый	Блок	159	1400,97	9,67
6	красный	Блок	131	1154,26	7,97
7	синий	Блок	118	1039,71	7,18
8	бледный	Блок	103	907,54	6,27
9	голубой	Блок	97	854,68	5,9
10	зеленый	Блок	73	643,21	4,44
11	серый	Блок	52	458,18	3,16
12	алый	Блок	41	361,26	2,49
13	розовый	Блок	41	361,26	2,49
14	лазурный	Блок	38	334,82	2,31
15	желтый	Блок	33	290,77	2,01
16	сумрачный	Блок	26	229,09	1,58
17	пестрый	Блок	17	149,79	1,03
18	золотистый	Блок	16	140,98	0,97
19	мутный	Блок	11	96,92	0,67
20	златой	Блок	8	70,49	0,49
21	рыжий	Блок	8	70,49	0,49
22	серебристый	Блок	8	70,49	0,49

Рис. 4. Фрагмент таблицы MS Excel частотности прилагательных, обозначающих цвет в авторском подкорпусе текстов А. А. Блока

$$N/B * 100,$$

где N — число словоупотреблений данной леммы в исследуемом авторском корпусе, В — число словоупотреблений прилагательных, обозначающих цвета в результатах поиска по авторскому корпусу.

Формула подсчета ipm следующая:

$$N/A * 1000000,$$

где N — число словоупотреблений данной леммы в исследуемом авторском корпусе.

агатовый | аленький | алый | аметистовый | багровый | багряный | бежевый | белейший | беленький | белесоватый | белесый | беловатый | белоснежный | белый | бесцветный | бирюзовый | бледный | бордо | бордовый | буланный | бурый | вороной | гнедой | голубенький | голубоватый | голубой | двухцветный | дымчатый | желтенький | желтый | зелененький | зеленоватый | зеленый | златой | золотистый | изумрудный | караковый | карий | кауры | клетчатый | коричневый | красненький | красноватый | красный | лазоревый | лазурный | лиловатый | лиловый | линиялый | мутный | одноцветный | опаловый | оранжевый | палевый | пегий | пестрый | поблеклый | пожелтелый | поседелый | посинелый | потускнелый | почернелый | пунцовый | пурпурный | пурпуровый | пятнистый | пятнистый | разноцветный | розоватый | розовый | румяный | русый | рыжебородый | рыжеватый | рыжеволосый | рыженький | рыжий | рябенький | рябоватый | саврасый | самоцветный | светлый | седенький | серебристый | серенький | сероватый | серый | сивый | сизокрылый | сизый | синеватый | синенький | синий | соловый | сумрачный | сумрачный | темнехонький | темный | трехцветный | фиолетовый | цветастый | цветной | чалый | червленький | черненький | чернехонький | черненький | черноватый | черноволосый | черномазый | черный

Рис. 5. Простые (однокоренные) прилагательные цвета

**.\*-красный|. \*-оранжевый|. \*-желтый|. \*-зеленый|. \*-голубой|. \*-синий|. \*-фиолетовый|. \*-коричневый|. \*-белый|. \*-черный|. \*-розовый|. \*-лиловый|светло-.\*|темно-.\*|ярко-.\*|черно-.\*|сине-.\*|нежно-.\*|серо-.\***

Рис. 6. Двусоставные прилагательные цвета (с переменной первой или второй частью)

По результатам автоматически, с использованием функции Frequency, формировался частотный список лемм, затем сводный список по двум сеансам поиска. Из сводного списка удалялись все слова, не имеющие значения «цвет» (например, ярко-тугой, ярко-положительный, нежно-хмельной). Все нелемматизированные формы (например, ярко-золотистым, ярко-желтым, ярко-розовою, черно-золотая) приводились к лемме. Обработанный таким образом список переносился в таблицу Excel в соответствующие столбцы («Лемма» и «Число с/у»), где далее формировалась сводная таблица на весь результат поиска по данному авторскому подкорпусу, как это описано выше.

### Исследование соотношений частоты употреблений прилагательных цвета по текстам 36 авторов

Полученные таблицы по каждому из 36 авторов были слиты в единую таблицу Excel, составившую 1571 строку. Материал в таблице отсортирован по столбцу «Лемма». Всего в таблице оказалось 170 лемм, и средствами Excel было подсчитано количество авторов, у которых каждая из них встретилась. В таблице 5 показаны первые (наиболее частотные) 25 лемм из 170.

Таблица 5

**Наиболее часто встречающиеся леммы прилагательных, обозначающих цвета**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. белый	36	10. бледный	35	19. белоснежный	28
2. голубой	36	11. серый	34	20. золотистый	26
3. желтый	36	12. розовый	33	21. лиловый	24
4. зеленый	36	13. рыжий	31	22. сизый	24
5. красный	36	14. цветной	31	23. сумрачный	24
6. светлый	36	15. багровый	30	24. серебристый	23
7. синий	36	16. пестрый	30	25. багряный	21
8. темный	36	17. алый	29		
9. черный	36	18. мутный	29		

**Категоризация прилагательных цвета**

Сделана попытка разбить 170 прилагательных на классы, иначе говоря, разработать основы их категоризации. Естественно, состав этих классов открыт, а возможно, появятся (следует ввести) и какие-то другие.

**Первая группа** — собственно цвета, таких слов мы обнаружили 30 из 170 (17,6 %) (табл. 6).

Таблица 6

**Прилагательные, означающие собственно цвета**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. белый	36	12. лиловый	24	23. пурпуровый	10
2. голубой	36	13. сизый	24	24. сивый	10
3. желтый	36	14. багряный	21	25. палевый	9
4. зеленый	36	15. лазурный	20	26. бордовый	5
5. красный	36	16. бурый	18	27. червлёный	5
6. синий	36	17. оранжевый	18	28. бежевый	3
7. черный	36	18. пурпурный	16	29. бордо	3
8. серый	34	19. коричневый	13	30. опаловый	3
9. розовый	33	20. лазоревый	13	Среднее	20,5
10. багровый	30	21. фиолетовый	12		
11. алый	29	22. пунцовый	10		

В эту группу вошли все 11 основных цветов (12 для русского языка). В пределах одного частотного ранга слова отсортированы по алфавиту. Отметим, что и здесь частотность (число авторов) у прилагательных «серый», «розовый», «оранжевый»,

«коричневый» и «фиолетовый», соответствующих VI и VII эволюционным ступеням, ниже частотности прилагательных, относящихся к более ранним ступеням, что еще раз говорит о корреляции между частотностью слов в поэтических текстах и стадиями эволюции по Берлину и Кэю.

Среди этих прилагательных встречаются синонимы. Вряд ли можно сделать уверенное различие между значениями прилагательных «красный», «алый», «пунцовый», «червлёный» или «фиолетовый», «лиловый», «пурпуровый», а также, например, «бурый» и «коричневый».

Кроме того, пожалуй, нет ни одного слова из тех, что относятся к базовым категориям цветообозначений, которое бы не употреблялось в переносном смысле — метафорически либо символически — в составе устойчивых словосочетаний. Достаточно вспомнить выражения «белый свет», «черный список», «красная нить», «оранжевая революция», «желтая пресса», «тоска зеленая» «голубой» (о мужчине нетрадиционной сексуальной ориентации), «синий чулок», «коричневая чума».

**Вторую группу** слов мы условно назвали «квазицвета» (19 лемм, 11,2%) (табл. 7). В число прилагательных, обозначающих цвета (то есть имеющих тег «цвет» в НКРЯ), попадают слова «темный», «светлый», «бледный», «пестрый», «сумрачный», «цветной», которые, собственно, означают не цвет, а его интенсивность, оттенок (темный, светлый, бледный), просто наличие цвета (цветной) — как оппозиция черному и белому, множество цветов (пестрый) или даже настроение (сумрачный), которое может вызвать цвет или комбинация цветов.

Таблица 7

**Прилагательные «квазицвета»**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. светлый	36	8. разноцветный	21	15. трехцветный	3
2. темный	36	9. бесцветный	20	16. двухцветный	1
3. бледный	35	10. клетчатый	10	17. крапчатый	1
4. цветной	31	11. самоцветный	10	18. одноцветный	1
5. пестрый	30	12. пятнистый	9	19. потускнелый	1
6. мутный	29	13. цветастый	8	среднее	16,42
7. сумрачный	24	14. линиялый	6		

С одной стороны, вызывает сомнение приписывание этим прилагательным тега «цвет». Они описывают визуальные характеристики объекта, но не цвет. С другой стороны, многие из этих слов хотя и не обозначают, но характеризуют именно цвет. Некоторые из них образуют сложные «цветные» прилагательные типа «бледно-зеленый», «светло-серый», «темно-красный» и т. д. (см. ниже табл. 11).

Как видно из таблицы 7, употребляемость некоторых из этих прилагательных достаточно высока. Особенно это касается прилагательных «светлый» и «темный».

Отдельные авторы утверждают, что именно смена светлого и темного времени суток обусловила появление цветообозначений «белый» и «черный» на самой ранней стадии эволюции цветообозначений в языках.

**Третью группу** составляют прилагательные, которые «привязываются» к определенному, чаще одушевленному, предмету. Это может быть цвет волос или глаз, масть или окрас животных (28 лемм, 16,5 %) (табл. 8).

Таблица 8

**Прилагательные, привязанные к определенному объекту**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. рыжий	31	11. сизокрылый	7	21. седоватый	2
2. румяный	21	12. буланный	6	22. соловый	2
3. русый	21	13. порыжелый	4	23. каурый	1
4. гнедой	15	14. рыжеволосый	4	24. муругий	1
5. карий	14	15. рыженький	4	25. мухортый	1
6. пегий	10	16. рябоватый	3	26. рыжебородый	1
7. вороной	9	17. саврасый	3	27. рябенький	1
8. поседелый	9	18. чалый	3	28. чубарый	1
9. рыжеватый	7	19. черноволосый	3	среднее	6,93
10. седенький	7	20. черномазый	3		

Некоторая часть этих слов также не означает собственно цвета. В некоторых случаях они означают цветовую неоднородность, наличие пятен («пегий», «рябенький», «чубарый», «чалый»), в других — указывается на результат некоторого процесса («поседелый», «порыжелый»). В некоторых случаях образуется прилагательное, содержащее одновременно объект и цветовую характеристику («рыжебородый», «черноволосый»). Значительное место в этой группе занимают масти лошадей и окрасы собак. Известно, что масть или окрас это не столько цвет, сколько сочетание цветов, генетически детерминированный тип распределения пигментов в шерсти животного.

**Четвертая группа** (85 лемм, 50 %) объединяет прилагательные-дериваты. Она может быть разделена на три подгруппы, отражающие словообразовательные модели образования прилагательных.

*В первую подгруппу* четвертой группы вошли прилагательные-диминытивы (13 лемм, 7,6 %) (табл. 9).

В их число вошли не только собственно прилагательные цвета — мы видим среди них и прилагательные «темнохонький» и «бледненький», исходные формы которых «темный» и «бледный» мы отнесли к «квазицветам». Некоторые из них мы включили в таблицу 7: «рыженький», «рябенький» и «седенький». Здесь мы сталкиваемся с типичной проблемой любой классификации — два значимых классификационных признака у одного объекта.



Таблица 9

**Прилагательные-диминытивы**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. беленький	13	6. черненький	5	11. темнехонький	1
2. серенький	13	7. зелененький	3	12. чернехонький	1
3. синенький	9	8. желтенький	2	13. чернешенький	1
4. аленький	6	9. красненький	2	среднее	4,77
5. голубенький	5	10. бледненький	1		

*Вторая подгруппа* — это прилагательные, которые мы условно назвали прилагательными подобия (10 лемм, 5,9%) (табл. 10).

Таблица 10

**Прилагательные подобия**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. розоватый	13	5. голубоватый	7	9. белесоватый	2
2. синеватый	11	6. лиловатый	6	10. беловатый	1
3. зеленоватый	9	7. желтоватый	4	среднее	6,5
4. красноватый	8	8. сероватый	4		

В этой форме мы видим некоторые прилагательные из таблицы 8: «рябоватый», «рыжеватый» и «седоватый». Обращает внимание, что их исходные слова те же, от которых были образованы диминытивы.

*Третья подгруппа* — это сложные прилагательные, имеющие обычно два (редко три) корня (62 леммы, 36,5%) (табл. 11). Они разными способами описывают интенсивность или оттенок цвета.

Таблица 11

**Сложные прилагательные**

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. темно-синий	14	8. ярко-желтый	4	15. ярко-красный	3
2. темно-голубой	9	9. молочно-белый	3	16. бледно-зеленый	2
3. черно-белый	9	10. нежно-белый	3	17. бледно-лиловый	2
4. темно-зеленый	8	11. светло-серый	3	18. медно-красный	2
5. бледно-голубой	7	12. светло-синий	3	19. нежно-голубой	2
6. черно-синий	6	13. серо-зеленый	3	20. светло-голубой	2
7. светло-зеленый	5	14. серо-синий	3	21. сине-зеленый	2

Продолжение табл. 11

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
22. синий-синий	2	36. нежно-зеленый	1	50. черно-бархатный	1
23. темно-желтый	2	37. нежно-золотой	1	51. черно-бурый	1
24. темно-красный	2	38. нежно-розовый	1	52. черно-золотой	1
25. темно-серый	2	39. потерто-белый	1	53. черно-красной	1
26. бледно-розовый	1	40. пречерный-черный	1	54. черно-серебряный	1
27. буйно-красный	1	41. прозрачно-синий	1	55. электродово-синий	1
28. грязно-белый	1	42. розовато-белый	1	56. ярко-белые	1
29. грязно-розовый	1	43. светло-пегий	1	57. ярко-голубой	1
30. желто-зеленый	1	44. светло-сиреневый	1	58. ярко-голубой	1
31. зеленовато-голубой	1	45. серо-голубой	1	59. ярко-зеленый	1
32. красно-синий	1	46. сине-красный	1	60. ярко-золотистый	1
33. кроваво-красный	1	47. темно-бирюзовый	1	61. ярко-оранжевый	1
34. медно-зеленый	1	48. темно-вишневый	1	62. ярко-розовый	1
35. небесно-голубой	1	49. темно-лиловый	1	среднее	2,24

Число лемм сложных прилагательных значительно превышает число членов других групп и подгрупп. В том числе элемент «бледно» образует 4 слова, «нежно» — 5 слов, «светло» — 6 слов, «темно» — 9 слов, «ярко» — 9 слов.

Наконец, **пятая группа**, которую мы назвали «аналоговые прилагательные» (табл. 12). Они передают цвета и оттенки через цвета и оттенки хорошо известных предметов. Г. И. Герасимов называет их относительными прилагательными [Герасимов 1969]. Семантическая разметка НКРЯ плохо учитывает слова такого типа, причем многие из них не имеют тега «цвет», поэтому при поиске по семантическому признаку в корпусе найдено всего 8 (4,7%) таких прилагательных. На самом деле их гораздо больше.

Таблица 12

## Аналоговые прилагательные

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. золотистый	26	4. изумрудный	18	7. агатовый	3
2. серебристый	23	5. бирюзовый	10	8. аметистовый	2
3. златой	21	6. дымчатый	8		

Проблема, видимо, в полисемии этих прилагательных. Рассмотрим одно из наиболее характерных прилагательных этого типа — «малиновый». Его описание в корпусе содержит следующие теги семантической разметки:

«Семантика основная der:s, dt:food, dt:fruit, r:rel  
Семантика дополнительная; ev:posit, r:rel, t:sound».

Эти теги указывают на происхождение слова от существительного «малина», то, что оно относится к еде (тег «food»), тег «fruit» подразумевает понятие «ягода», его оценка позитивная «ev:posit», тег «t:sound» указывает на отношение к звукам (по-видимому, «малиновый звон»), но ни в основной, ни в дополнительной семантике цвета нет.

Ниже приведены 10 самых частотных биграмм с прилагательным «малиновый», полученных с помощью системы Google Books Ngram Viewer.

1996	малиновый перезвон_NOUN
малиновый цвет_NOUN	малиновый кафтан_NOUN
малиновый звон_NOUN	малиновый закат_NOUN
малиновый бархат_NOUN	малиновый Евсюков_NOUN
малиновый оттенок_NOUN	малиновый чай_NOUN
малиновый свет_NOUN	

В этом списке видно, что преобладающим значением прилагательного является именно цвет. Даже комиссар Евсюков из рассказа Б. Лавренева «Сорок первый» назван малиновым потому, что носил куртку малинового цвета.

Вывод, казалось бы, ясен. При ближайшем редактировании семантической разметки НКРЯ аналоговым прилагательным следует приписать тег «цвет». Однако все не так просто. Не у каждого прилагательного этого ряда значение цвета является доминирующим. Например, у таких прилагательных, как «кровавый», «стальной», цвет будет, скорее, периферийным значением, в результате чего применение этого тега могло бы вызвать значительный информационный шум.

## Заключение

Прилагательные, которые рассмотрены нами в данной статье, имеют один общий признак — наличие тега «цвет» в семантической разметке НКРЯ. По приведенным данным частота встречаемости прилагательных цвета в поэтическом корпусе выше, чем их частота в основном корпусе, и заметно выше, чем в устных речевых произведениях или газетных статьях.

Прилагательные цвета могут быть разделены на пять групп. Причем четвертая группа (формы прилагательного) может быть, в свою очередь, разбита на три подгруппы. В ходе исследования выявлены некоторые недостатки семантической разметки, которые не позволили в нашем случае учесть истинное число лемм прилагательных, которые мы назвали аналоговыми.

Что касается частотного поведения прилагательных цвета, то здесь можно обозначить два аспекта. С одной стороны, имеется корреляция частоты прилагательных с эволюционной теорией Берлина и Кэя, то есть наиболее частотны базовые категории прилагательных, а среди них наиболее частотны прилагательные, относящиеся к более ранней стадии эволюции. С другой стороны, эта корреляция не является абсолютной, но зависит от контекста. При этом отметим, что схема

Берлина и Кэя, как, впрочем, и любая логическая схема, в применении к языку не может быть соблюдена полностью, тем более, что она представляет собой некоторый усредненный результат анализа обширного лингвистического материала.

Употребляемость авторами поэтического корпуса прилагательных собственно цвета выше частотности прилагательных других обозначенных нами групп.

Обращает на себя внимание, что наряду с небольшим числом наиболее частотных прилагательных существует очень большое число сложных цветовых терминов единичного употребления. Подобная структура лексического массива напоминает частотное распределение по закону Ципфа.

### Литература

Василевич А. П., Кузнецова С. Н., Мищенко С. С. Цвет и названия цвета в русском языке / под общ. ред. А. П. Василевича. М.: Ком Книга, 2005. 216 с.

Герасимов Г. И. Прилагательные, обозначающие цвет // Русский язык за рубежом. 1969. №3. С. 26—35.

Гришина Е. А., Корчагин К. М., Плунгян В. А., Сичинава Д. В. Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 71–113.

Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. М., 2005. С. 155–174.

Berlin B., Kay P. Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press, 1969.

Kay P., Regier T. Language, thought and color: recent developments // Trends in cognitive sciences. 2006.No.10(2). P. 51–54.

Kay P., McDaniel Ch.K. The Linguistic Significance of Meanings of Basic Color Terms //Language. 1978. No. 54 (3). P. 610–646.

<sup>1</sup>*V.P. Zakharov, <sup>2</sup>A.Ts. Masevitch*

<sup>1</sup> *St. Petersburg State St. University*  
*v.zakharov@spbu.ru*

<sup>2</sup> *St. Petersburg State institute of culture*  
*andmasev@mail.ru*

<sup>1, 2</sup> *(Russia, St. Petersburg)*

### CATEGORIZATION OF ADJECTIVES OF COLOR IN THE RUSSIAN POETICAL TEXTS (A CORPUS-BASED STUDY)

The paper describes certain quantitative data of usage of color adjectives in Russian poetic texts. We suggest methodology of revealing adjectives of color in the poetic

corpus of the Russian National Corpus with the help of semantic tags. We have also compared frequency of adjectives of color in four corpora of the National corpus of Russian Language and found that frequency of adjectives of color in the poetic corpus significantly higher than in other three corpora. Having analyzed 170 lemmas of adjectives of color retrieved from the texts of the poetic corpus, we have systematized these adjectives and suggested criteria of their typification. The paper presents quantitative data on the representation of suggested types of adjectives in Russian poetic texts. We have revealed a noticeable correlation between the frequency of eleven adjectives of color and evolutionary theory of basic color terms by Berlin and Kay.

**Keywords:** adjectives of color, theory of basic color terms by Berlin & Kay, National corpus of Russian language, poetic corpus, semantic tagging.

### References

Berlin B., Kay P. *Basic Color Terms: Their Universality and Evolution*. Berkeley, University of California Press, 1969.

Gerasimov G.I. [Adjectives denoting colour]. *Russkii yazyk za rubezhom* [Russian abroad]. 1969, no. 3, pp. 26–35. (In Russ.)

Grishina E.A., Korchagin K.M., Plungyan V.A., Sichinava D.V. [Poetic corpus within the framework of the RNC: general structure and prospects of use]. *Natsional'nyi korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [The National Corpus of the Russian language: 2006–2008. New results and prospects]. St. Petersburg, Nestor-Istoriya Publ., 2009, pp. 71–113. (In Russ.)

Kay P., McDaniel Ch.K. The Linguistic Significance of Meanings of Basic Color Terms. *Language*, 1978, no. 54 (3), pp. 610–646.

Kay P., Regier T. Language, thought and color: recent developments. *Trends in cognitive sciences*, 2006, no. 10(2), pp. 51–54.

Kustova G.I., Lyashevskaya O.N., Paducheva E.V., Rahilina E.V. [Semantic annotation of the lexicon in the Russian National Corpus: principles, problems, prospects]. *Natsional'nyi korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy* [The Russian National Corpus: 2003–2005. Results and Prospects]. Moscow, 2005, pp. 155–174. (In Russ.)

Vasilevich A.P., Kuznecova S.N., Mishchenko S.S. *Tsvet i nazvaniya tsveta v russkom yazyke* [Color and colour names in Russian]. Ed. A.P. Vasilevich. Moscow, Kom Kniga Publ., 2005. 216 p. (In Russ.)

**К. М. Корчагин**  
*Институт русского языка им. В. В. Виноградова РАН*  
*(Россия, Москва)*  
*stivendedal@gmail.com*

## **ЦЕЗУРА В РУССКОМ 5-СТОПНОМ ЯМБЕ И ЕЕ РОЛЬ В РИТМИЧЕСКОЙ ЭВОЛЮЦИИ ЭТОГО РАЗМЕРА\***

Русский 5-стопный ямб — один из самых распространенных и один из самых хорошо исследованных размеров. Исчерпывающе подробное описание его ритмики на материале XVIII—XIX вв. было сделано К. Тарановским еще в середине XX в. (Тарановский 2010), а затем дополнено М. Л. Гаспаровым (Гаспаров 1974; 2002). Тем не менее в истории этого размера остаются «слепые зоны», связанные как с тем, что за время, прошедшее с упомянутых классических работ, были опубликованы важные тексты, проясняющие некоторые детали его эволюции, так и с тем, что подход классиков русского стиховедения предполагал детальное исследование профилей ударности, но почти игнорировал другие ритмические факты стиха, прежде всего цезуру. Отчасти это было связано с тем, что русское стиховедение 1960—1970-х гг. стремилось создать теорию метрики с чистого листа, подвергнуть ревизии все то, что составляло предмет старых руководств по поэтике и стихосложению. В этом оно было близко лингвистическому структурализму, хотя последний воспринимался скептически М. Л. Гаспаровым и его последователями.

Для того чтобы показать влияние этих несвоевременно отброшенных ритмических фактов на эволюцию 5-стопного ямба, я произведу частичную ревизию истории этого размера в том виде, в каком она изложена у К. Тарановского и М. Л. Гаспарова. Это позволит показать, что граница между классическими размерами (такими как 5-стопный ямб) и неклассическими (такими как дольник и тактовик) не должна восприниматься как непреодолимая: последние являются прямыми следствиями эволюции ритма первых. Однако это невозможно показать, если руководствоваться принятыми в русском стиховедении способами анализа ритмики стиха, стремящимися изолировать различные типы размеров, избегая вопроса

---

\* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-29-09154 «Динамика языковой системы: корпусное исследование синхронной вариативности и диахронических изменений в текстах разных типов».

о том, как они могут быть связаны друг с другом. Для того чтобы сделать эту связь ясной, я вернусь к понятию цезуры как к ключевому для анализа стиха и попытаюсь показать, каким образом это понятие позволяет уточнить эволюцию русского 5-стопного ямба.

*Ключевые слова:* метрика, стиховедение, 5-стопный ямб, цезура, эволюция русского стиха, поэтический корпус, НКРЯ

У русского 5-стопного ямба было, как минимум, три зарубежных источника: французский *vers commun* (10-сложник 4 + 6), немецкий 5-стопный ямб и затем английский 5-стопный ямб<sup>1</sup>. При этом наиболее влиятельными эти образцы (особенно первый из них) были на ранних стадиях развития размера — начиная с определенного момента русский 5-стопный ямб начинает развиваться независимо, что непосредственно сказывается на интерпретации как цезуры, так и других элементов стиха. Во многом это связано с тем, что к XIX в. 10-сложник во французской поэзии практически выходит из употребления (Гаспаров 2003: 110), а немецкий и английский виды этого размера уже мало отличаются от русского в трактовке цезуры.

Для 5-стопного ямба важно различать два вида цезур, впервые в таком виде выделенных В. М. Жирмунским [Жирмунский 1975], а затем подробно проанализированных Дж. Бейли [Бейли 2004: 220—251]. Это *слабая* и *сильная* цезуры. В случае *слабой* цезуры речь идет только о регулярном словоразделе, который не сопровождается дополнительной метрической (ударной) константой. Напротив, *сильная* цезура — это регулярный словораздел, который таковой сопровождается. В истории русского стиха эти два типа цезур сменяют друг друга: если классическая силлабо-тоника XVIII—XIX вв. предпочитает слабую цезуру, то в XX в. в тех же размерах начинает использоваться сильная. По всей видимости, это было связано с тем, что стих с цезурой воспринимается как экзотика, и поэтам требуется дополнительное средство, чтобы привлечь внимание читателя к тому, что они не просто соблюдают классический размер, но используют его в наиболее строгом варианте — с цезурой.

В истории 5-стопного ямба цезура закрепляется отнюдь не сразу и держится очень недолго, хотя и оставляет после себя достаточно долгую память: до сих пор в практических руководствах по стихосложению 5-стопный ямб упоминается как

---

<sup>1</sup> Немецкий вариант существовал как в цезурном, так и в бесцезурном виде: первый воспринимался как аналог французского 10-сложника, второй — итальянского 11-сложника (Wagenknecht 2007: 86—87). Английский вариант, как мы отмечали в первой главе, в лучшем случае обладает лишь колометрической цезурой. Русский вариант 5-стопного ямба почти всегда бесцезурен. Однако тот факт, что бесцезурный 5-стопный ямб генетически восходит к 5-стопному ямбу с цезурой после второй стопы и часто сохраняет словораздел в этой позиции, заставляет некоторых исследователей говорить о *свободной* цезуре, если стихотворение содержит 80—95% строк со словоразделом в обычной цезурной позиции (Тарановский 2010: 153; Бейли 2004: 182). Используемая в этой статье классификация цезур обходится без этого понятия, и стих такого типа в общем случае будет считаться либо бесцезурным, либо цезурным с отдельными отклонениями.



цезурный размер, хотя количество его бесцезурных образцов в сотни и тысячи раз больше. На материале поэзии XVIII в. видно, что 5-стопный ямб не сразу закрепляется в метрическом репертуаре русской поэзии. В поэтическом подкорпусе НКРЯ обнаруживается всего девять образцов этого размера: они принадлежат В. Тредиаковскому (один образец, если не считать иллюстраций к стиховедческим работам), М. Ломоносову (переводы Горация и Марциала), А. Сумарокову, Я. Княжнину, П. Словцову и М. Муравьеву (последнему — целых шесть текстов, больше, чем у любого другого поэта эпохи). Все эти образцы в известном смысле маргинальны — они не формируют заметной традиции даже в рамках индивидуальных поэтик; только Муравьев писал этим размером более или менее систематически.

Тредиаковский определял 5-стопный ямб следующим образом: «Стих пентаметр ямбический имеет два полстишия, то есть он пресекается. В первом полстишии мужеского стиха содержит он две стопы, из сих последняя есть пресечение, а во втором три стопы равно» (Тредиаковский 2009: 80). За этим следует утверждение, что полстишие «никогда не долженствует пресекаемо быть пиррихием, но всегда и непременно ямбом» (Там же: 81). Впрочем, далее он говорит о допустимости 5-стопного ямба без цезуры («в самых малых шутках: сему я не советую в важном Сочинении подражать»). Другими словами, Тредиаковский использовал в 5-стопном ямбе сильную цезуру, хотя не настаивал на ее обязательности:

Приятный брѣг! || Любезная страна!  
Где свой Невá || поток стремится к пучине.  
О! прежде дѣбрь, || се коль населена!  
Мы град в тебѣ || престольный видим ныне...

*В. Тредиаковский, «Похвала Ижерской земле...», 1752*

Следующий опыт сильной цезуры в этом размере обнаруживается у Сумарокова в VI идиллии, что может показаться неожиданным на фоне того, как по-разному понимали Тредиаковский и Сумароков цезуру в 6-стопном ямбе: первый настаивал на том, что она должна быть всегда сильной, второй же — что всегда слабой. Но в трактовке 5-стопного ямба эти непримиримые соперники совпали. Причина этого, по-видимому, в том, что 5-стопный ямб в то время воспринимался как размер периферийный, не требующий особо тщательной «русификации» и сохранявший некоторые особенности своего французского прототипа, 10-сложника 4 + 6, где цезура могла быть только сильной.

Действительно, 5-стопный ямб XVIII в. в основном ведет себя непоследовательно с точки зрения цезуры, и о каких-либо общих тенденциях трудно говорить и в силу бедности материала. Так, в переводе небольшой эпитаграммы Марциала, выполненном Ломоносовым для «Риторики» (1759), цезура не соблюдается:

О имя, купно с рѡ||<sup>2</sup>замирожденно!  
Тобой зовется лѣ||<sup>2</sup>та часть прекрасна.

Ты сладко, как цветы|| и мед Иблейский  
И как пресладкий нѣк||<sup>п</sup>тар на Олимпе.  
Тобою бы желал|| назваться отрок,  
Зевесов виночерпец.

Еще более интересный случай — перевод Ломоносова знаменитой оды Горация (III, 30) «Ехegi monumentum aere perennius...» <1747>:

Я знак бессмѣртія || себе воздвигнул  
Превыше пирамид || и крепче меди,  
Что бурный аквилон || сотреть не может,  
Ни множество веков, || ни едка древность...

Здесь слабая цезура занимает позицию не после второй, а после *третьей* стопы (Тарановский 2010: 153). Эксперимент, однако, остался единичным в этой эпохе: разработка неклассических цезурных вариаций 5-стопного ямба возобновится лишь в XX в.

В 1770—1780-е гг. 5-стопный ямб периодически использует М. Муравьев (напомним, что образцов этого размера в его корпусе больше, чем у какого-либо другого поэта XVIII в.). У Муравьева употребляется цезура после второй стопы, но, в отличие от Тредиаковского и Сумарокова, она всегда слабая, что можно воспринимать как признак «русификации» размера, сближения его с русским александрийским стихом:

Товарищи,    наставники, друзья,	д.ц.
О книги! к вам    украдываюсь я	м.ц.
Мгновенье скрыть    оставшейся прохлады,	м.ц.
Вкушая в вас    полезные отрады.	м.ц.
И время течь    скорые обяжу...	м.ц.
«Товарищи, наставники, друзья...», 1779	

Среди 5-стопных ямбов Муравьева обнаруживается лишь одна строка (из 141-й, написанных этим размером), где позиция цезуры нарушается: *Вошедша гост||<sup>п</sup>тя зрят. Они мне святы* («Итак, опять убежище готово...», 1780). В этих стихах Муравьева употребляется 26 % дактилических полустиший и 74 % мужских. Речевая модель цезурного варианта этого размера, предложенная М. Л. Гаспаровым для поэзии XIX в., дает 35,9 % дактилических полустиший и 64,1 % мужских (Гаспаров 1974: 103), что заметно отличается от данных по Муравьеву. По всей видимости, для этого поэта было важно подчеркивать цезуру мужским окончанием именно в силу того, что размер еще не казался ему привычным и естественным.

За опытами Муравьева следует 5-стопный ямб Княжнина («Послание трем грациям», <1790>), где также используется слабая цезура:

О не́жные    сестрицы неразлучны!	д.ц.
Сопу́тницы    прекрасного всего!	д.ц.
Сама красá    не стоит ничего,	м.ц.
Черты её    бездушны, вялы, скучны,	м.ц.
Коль, ва́шего    знакомства лишена,	д.ц.
Жема́нится    без прелестей она...	д.ц.

В этом послании дактилический цезурный словораздел встречается почти так же часто, как мужской, — 43,5% дактилических полустипий и 56,5% мужских (всего в стихотворении 139 строк, причем одна из них — 4-стопная).

Первые бесцезурные 5-стопники появляются в то же десятилетие — у П. Словцова в двух довольно длинных натурфилософских одах, опубликованных в 1796 г.: «К Сибири» и «Материя», почти на два десятилетия раньше, чем переводы Жуковского из Гебеля (1816), упоминаемые К. Тарановским в качестве первых образцов этого размера. Словцов был из семьи священника, он родился в горно-заводской части Сибири (ныне Свердловская область), учился в Тобольской духовной семинарии, а затем за успехи в обучении был отправлен в Петербург в Александро-Невскую семинарию, где его соучеником был М. М. Сперанский. Еще в ранние годы он был увлечен идеей «научной» поэзии и вдохновлен естественнонаучными опытами Ломоносова и его же «Письмом о пользе стекла». Авторское примечание к оде «Материя», сочиненной еще в семинарии, гласит: «Сочинитель сей пиесы хотел только испытать, можно ли физические истины предлагать в стихах» [Поэты 1790—1810-х гг. 1971: 212]. Два упомянутых стихотворения стилистически близки к торжественной оде XVIII в., но тематика их далека от тематики таких од. Возможно, именно это расхождение подтолкнуло поэта использовать необычный для оды и редкий в то время размер — бесцезурный 5-стопный ямб:

Дщерь Азии, богато наделенна!  
 По статным и дородным раменам  
 Бобровую порфирой облеченна,  
 С соболями хвостами по грудям,  
 Царевна! серебряный венец носяща  
 И пестрой насыпью камней блестяща!  
 Славян наперсница, орд грозных мать,  
 Сибирь — тебя мне любо вспоминать...  
*П. Словцов, «К Сибири», 1796*

Словораздел после второй стопы встречается здесь всего в 36,8% строк (из 136), то есть отсутствует даже минимальная тенденция к тому, чтобы сосредоточивать словоразделы после второй стопы<sup>2</sup>.

<sup>2</sup> Сходные показатели дает другой ранний пример бесцезурного 5-стопного ямба — драма В. Т. Нарезного «Кровавая ночь, или Конечное падение дому Кадмова» (1799). Строго говоря, эта

Таким образом, можно считать, что в этот период 5-стопный ямб уже разделился на два варианта — цезурный и бесцезурный. Второй вариант быстро стал широко употребительным размером и сохранял эти позиции до самого последнего времени, превратившись, по сути, в наиболее популярный классический размер [Гаспаров 2002: 306—314]. История же цезурного 5-стопного ямба более прерывистая — в некоторые эпохи, особенно когда возрождается интерес к поэтической классике, он мог ненадолго входить в моду, в другие — почти совсем пропадать.

В первые десятилетия XIX в. цезурный 5-стопный ямб переживает кратковременный расцвет. В основном в поэзии этого периода употреблялся 5-стопный ямб со слабой цезурой (как у Муравьева и Княжнина); отступления от этой закономерности были очень редкими и нуждались в специальной мотивировке. Согласно данным К. Тарановского<sup>3</sup>, стремление к мужскому окончанию первого полустишия в 5-стопном ямбе у поэтов до 1840 г. выражено достаточно четко: больше всего мужских полустиший у Полежаева в 1834 г. (95,5% — это почти ударная константа), меньше всего — у Батюшкова в 1817 г. (58,9%; впрочем, сразу за ним, хотя и с большим разрывом, следует Баратынский с 67,3%, близкий к речевой модели). Таким образом, количество мужских полустиший не опускается ниже речевой модели: дактилическое окончание ни у кого, кроме Батюшкова, сознательно не выводится на первый план. Если усреднить все показатели, предварительно изъяв из списка Батюшкова с его резко индивидуальной ритмикой, то для всего периода получится 22,4% дактилических полустиший и 78,6% мужских. Другими словами, тенденции, обнаруживаемые в стихах Муравьева, становятся в первой половине XIX в. еще более заметными: предцезурное полустишие гораздо чаще несет мужское, а не дактилическое окончание — в случае, когда используется цезурный вариант размера, требуется дополнительно подчеркнуть это ритмическими средствами.

Сильную цезуру в 5-стопном ямбе использует, например, Жуковский в переводе песни Миньоны из романа Гёте «Годы учения Вильгельма Мейстера» (под названием «Мина», 1817) и в «Верности до гроба» (переложение Теодора Кернера, 1818) [Тарановский 2010: 170—174]:

Я знаю кра́й! || там негой дышит лес,  
Златой лимон|| горит во мгле древес,  
И ветерок|| жар неба холодит,  
И тихо мирт|| и гордо лавр стоит...  
Там счастье, друг! туда! туда  
Мечта зовёт! || Там сердцем я всегда!  
*«Мина», 1817*

драма написана вольным ямбом, но 5-стопные строки в ней преобладают (88,4%). О стихе этой драмы см. подробнее в: [Бейли 2004: 104—108].

<sup>3</sup> В этих расчетах были учтены: И. Крылов, В. Жуковский, П. Вяземский, К. Батюшков, А. Пушкин, А. Дельвиг, П. Плетнев, И. Козлов, Н. Языков, Е. Баратынский, С. Шевырев, А. Подолинский, А. Полежаев, В. Одоевский. Данные по ударности 5-стопного ямба для всех этих поэтов почти не отличаются с точки зрения предцезурной позиции.

Младой Рогёр || свой острый меч берет,  
За веру, чéсть || и родину сразиться.  
Готов он в бóй... || но к милой он идет,  
В последний рáз || с прекрасною проститься...  
*«Верность до гроба», 1818*

В «Мине» пятая строка каждой строфы 4-стопная и сохраняется словораздел после второй стопы. В немецком оригинале песни Миньоны также используется 5-стопный ямб с сильной цезурой, но с единичным смещением словораздела на слог вправо. Стихотворение же Кернера устроено несколько иначе: в нем катрены 5-стопного ямба чередуются с катренами 4-стопного ямба, причем позиция цезуры иногда нарушается:

Kennst du das Háus? || Auf Säulen ruht sein Dach.  
Es glänzt der Sáal, || es schimmert das Gemach,  
Und Marmorbíll||<sup>?</sup>der stehn und sehn mich an:  
Was hat man dir, du armes Kind, getan?—  
Kennst du es wohl?  
Dahin, dahin  
Möcht ich mit dír, || o mein Beschützer, ziehn!  
*J. W. Goethe, «Mignon», 1795*

Der Ritter múß || zum blut‘gen Kampf hinaus,  
Für Freiheit, Rúhm || und Vaterland zu streiten;  
Da zieht er nóch || vor seines Liebchens Haus,  
Nicht ohne Ab||<sup>?</sup>schied will er von ihr scheiden.  
*T. Kerner, «Treuer Tod», <1813>*

Важно, что цезурный вариант 5-стопного ямба в немецком стихе почти не употреблялся, так что приведенные стихотворения Гёте и Кернера в известном смысле исключения, демонстрирующие встречное влияние французского десятисложника, всегда содержащего цезуру [Тарановский 2010: 174]. Связанный с этим фактом эпизод в истории размера — полемический обмен колкостями между Жуковским и Пушкиным:

Послушай, дедушка, мне каждый раз,  
Когда взгляну на этот замок Ретлер,  
Приходит в мысль, что, если это проза,  
Да и дурная?.....  
*<1818>*

Эти строки Пушкина пародируют начало длинного драматического стихотворения Жуковского «Тленность» (1816), перевода стихотворения Гебеля «Die

Vergänglichkeit». Обычно считается, что Пушкин в этой эпиграмме метит в использованный старшим товарищем белый стих, однако такое прочтение может показаться странным на фоне того, что сам Пушкин впоследствии неоднократно писал белым стихом. Можно предположить, что предмет полемики иной — отсутствие цезуры в размере, избранном Жуковским. Действительно, все 5-стопные ямбы молодого Пушкина содержат цезуру, в том числе и белые ямбы драмы «Борис Годунов» (1825), что отнюдь не тривиально для драматического белого стиха, который в дальнейшей истории русской поэзии избегает цезуры так же, как и его английский прототип. Так, драматическая трилогия А. К. Толстого (1866—1870), непосредственно ориентировавшегося на «Бориса Годунова», уже бесцезурна<sup>4</sup>. В черновике предисловия к драме Пушкин писал: «Я сохранил цезурку французского пентаметра на второй стопе — и, кажется, в том ошибся, лишив добровольно свой стих свойственного ему разнообразия» (цит. по: [Тарановский 2010: 216]).

5-стопный ямб с цезурой утрачивает популярность в 1840—1850-е гг., когда его окончательно вытесняет бесцезурный стих. Это также связано с Пушкиным, который начинает активно пользоваться бесцезурным ямбом в 1830-е гг. — когда его практика уже становится знаковой для большинства поэтов. И первое же вступление на поле бесцезурного ямба, поэма «Домик в Коломне» (1830), может быть названа образцом для всех дальнейших опытов в этом роде — как и у самого Пушкина, который далее пишет этим размером и лирику, и «Маленькие трагедии», так и у его современников. Пример оказался настолько заразителен, что среди заметных поэтов послепушкинского времени цезурный вариант используют только немногочисленные одиночки — А. Майков, К. Павлова и И. Никитин: их средние показатели — 23,9% дактилических полустихий и 76,1% мужских. Таким образом, уже в это время 5-стопный ямб с цезурой становится размером, который любят одни поэты и игнорируют другие, причем первых, как правило, немного.

В цезурном 5-стопном ямбе единичные нарушения цезурного словораздела встречаются чаще, чем в александрийском стихе: такие строки присутствуют у В. Жуковского, А. Пушкина, И. Козлова, А. Полежаева, А. Подолинского, В. Одоевского, А. Майкова и К. Павловой, хотя их никогда не бывает больше 3%. Отклонения, так же как и в 6-стопном ямбе, бывают двух типов: смещение цезурного словораздела на один слог вправо (строки типа *Устало солнце, || жегшее спокойно*)<sup>5</sup> и исчезновение словораздела в окрестности цезуры (*Нет муки сла||?дострастней и больней*). У Жуковского, Пушкина, Подолинского и Козлова встречается только первый тип, в то время как у остальных поэтов — оба.

Начиная со второй половины XIX в. цезурный 5-стопный ямб выходит из употребления; по данным Дж. Бейли у поэтов 1880—1922 гг. словораздел в позиции цезуры встречается лишь в 60—80% строк [Бейли 2004: 211—219], по данным М. Л. Гаспарова у советских поэтов — не чаще чем в 50% строк [Гаспаров 1974:

<sup>4</sup> Автор благодарит Александра Мурашова за сообщение об этом сюжете и последующее его обсуждение.

<sup>5</sup> Дж. Бейли называл такое явление передвижной цезурой.





карьеры, — почти всегда у него наблюдается либо цезура «бальмонтовского» типа, с тяготением к мужской, либо слабая цезура в духе XIX в. Исключительно слабой цезурой пользуется Брюсов, у которого, впрочем, гораздо больше бесцезурного стиха.

В отдельных стихотворениях Бунин, чье скептическое отношение к модернистской поэзии хорошо известно, идет даже дальше модернистов и использует сильную мужскую цезуру, причем последовательное ее соблюдение воспринимается, по всей видимости, как способ придать привычному размеру привкус экзотики:

В шелках песко́в || лишь сизые полыни  
Растит Алла́х || для кочевых отар,  
И небеса́ || здесь несказанно сини,  
И солнце в ни́х — || как адский огонь, Сакар.  
*«Ковсесь», <1905>*

Несколько дальше идет И. Анненский, воспринимающий мужскую цезуру как характерную черту обновленного Бальмонтом пятистопника. У Анненского систематическое использование сильной мужской цезуры дополнительно подчеркивается синтаксисом — так же, как в 6-стопном ямбе (таких стихотворений у него порядка 25, то есть это отнюдь не случайность):

Среди мирóв, || в мерцании светил  
Одной Звезды|| я повторяю имя...  
Не потому́, || чтоб я Её любил,  
А потому́, || что я томлюсь с другими.  
*1909*

Я — слабый сын|| больного поколенья  
И не пойду́|| искать альпийских роз,  
Ни ропот во́лн, || ни рокот ранних гроз  
Мне не дадут|| отрадного волненья.  
*«Ego», 1909*

Строго говоря, почти все 5-стопные ямбы Анненского написаны таким образом, в том числе и пародия на Бальмонта «Моя душа эбеновый гобой...». Вдохновившее эту пародию стихотворение «Моя душа оазис голубой...» было построено несколько иначе: начало стихотворения выдерживает мужскую цезуру, которая затем ослабляется, а в последних строках вовсе исчезает. Анненский усиливает тенденцию оригинала: цезурный словораздел выдерживается везде, и почти везде он мужской<sup>7</sup>:

---

<sup>7</sup> Эхом этого стихотворения Бальмонта можно считать текст раннего Н. Тихонова «Моя душа в бездумности озер...» (1913—1919?), где происходит возвращение к цезурной стратегии старшего

Моя душа || оазис голубой,  
Средь бледных душ || других людей, бессильных.  
Роскошный сон || ниспослан мне судьбой,  
Среди пустынь, || томительных и пыльных.

Везде пески. || Свистя, бежит самум.  
Лазурь небес || укрылася в туманы.  
Но слышу я || желанный звон и шум,  
Ко мне сквозь мглу || подходят караваны.

**Веселые, || раскинулись на миг,**  
Пришли, ушли, || до нового свиданья,  
В своей душе || лелеют мой двойник,  
Моей мечты || воздушной очертанья.

И вновь один, || я вновь живу собой,  
**Мне снится ра||?дость вечно молодая.**  
Моя душа || оазис голубой,  
Мои мечты || цветут, не отцветая.  
Моя душа || эбеновый гобой,  
И пусть я ниц || упал перед кумиром,  
С тобой, дитя, || как с медною трубой,  
Мы всё ж, пойми, || разъяты целым миром.

О будем же || скорей одним вампиром,  
Ты мною будь, || я сделаюсь тобой,  
Чтоб демонов || у Яра тешить пиром,  
Будь ложкой мне, || а я тебе губой...

Пусть демоны || измаялись в холере,  
Твоя коза || с тобою, мой Валерий,  
А Пантеон || открыл над нами зонт,

Душистый зонт || из шапок волькамерий.  
Постой... Но ложь — || гобой, и призрак — горизонт.  
Нет ничего || нигде — один Бальмонт.

1899

В тех случаях, когда у Анненского сильная мужская цезура все-таки не соблюдается, имеет место слабая цезура, причем в некоторых строках она дополнительно

---

поэта: мужская цезура вначале, затем ее ослабление и частичное устранение в конце. Тем самым, можно говорить о том, что эксперименты с цезурой у Бальмонта не остались незамеченными.

ослабляется за счет более плотных синтаксических связей между второй и третьей стопой или, напротив, усиливается, чтобы создать впечатление ритмического курсива. Особенно показательно в этом отношении стихотворение «Кошмары», построенное на тонкой игре с различными типами цезуры:

«Вы ждете? || Вы в волненьи? Это бред.  
Вы отворять || ему идете? Нет!  
Поймите: к вам || стучится сумасшедший,  
Бог знает где || и с кем всю ночь проведенный,  
Оборванный, || и речь его дика,  
И камешков || полна его рука;  
<...>  
Послушайте!.. || Я только вас пугал:  
Тот далеко, || он умер... Я солгал.  
И жалобы, || и шепоты, и стуки, —  
Все это «ше||<sup>2</sup>лест крови», голос муки...  
<1910>

В этом отрывке мужское предцезурное окончание используется в случаях, когда нужно передать аффектацию, овладевающую персонажем: *Поймите: к вам..., Бог знает где...* и т. п. При этом эти строки могут быть прочитаны и без подчеркнутой цезуры: синтаксическая связь между полустипиями в них довольно сильная, заставляющая читать их слитно. И все же то, что словораздел удерживается на протяжении всего стихотворения, заставляет внимательнее к ним приглядеться. Строку, где цезурное членение нарушается (*Все это «шелест крови»...*), стоит читать как ритмический курсив — она содержит скрытую отсылку к повести «После смерти (Клара Милич)» Тургенева.

Такой же тип стиха предпочитают и М. Кузмин и М. Волошин, представители младшего по сравнению с Анненским и Бальмонтом поколения. Они довольно часто прибегают к цезурному пятистопнику (у М. Кузмина 14 из 104 всех 5-стопных ямбов, у Волошина — 25 из 54), хотя у них использование этого размера ограничено рубежом 1900—1910-х гг.

У Кузмина граница между размерами с сильной цезурой и слабой выражена более отчетливо: если у Бальмонта и его современников можно говорить о тенденции к сильной мужской цезуре, то в его случае нужно говорить о двух группах текстов — с сильной мужской цезурой и с традиционной слабой:

О тихий край, || опять стремлюсь мечтою      сильная м.ц.  
К твоим лугам|| и дремлющим лесам,  
Где я бродил, || ласкаемый тоскою,  
Внимал лесным|| и смутным голосам...

М. Кузмин, «О тихий край, опять стремлюсь мечтою...», 1908—1909

И, ду́маю,    туманится их взгляд:	д.ц.
«Ах, ю́ноши,    когда б пришел к вам опыт.	д.ц.
Ты, мо́лодость,    вернешься ль к нам назад?»	д.ц.

Наве́ки уж    утрачен резвый топот	д.ц.
Веселых но́г,    не заблестает глаз,	м.ц.
Любовью́ полн,    и тщетен грустный ропот...	м.ц.

М. Кузмин, «Стрелы», <1909>

В 85 строках последнего текста 28% дактилических словоразделов и 72% мужских, что довольно близко к показателям XIX в. Однако на этом фоне можно заметить разного рода вольности — отсечение проклитики цезурным словоразделом (*Которой, без || любви, бежал я сам?*) и смещение позиции словораздела на слог вправо (*Где снят признать||?я, девичий испуг!*), которые напоминают ритмический курсив.

У Волошина в целом наблюдается либо тяготение к сильной мужской цезуре с единичными отклонениями (более уверенная тенденция, чем у Бальмонта), либо избегание цезурного словораздела после второй стопы. Причем почти всегда сильная цезура наблюдается в сонетах, которые, в свою очередь, объединены в венки: цезура, по всей видимости, позволяет поэту делать эту предельно твердую форму еще более твердой. В цикле сонетов «Киммерийские сумерки» (1907—1909) и более позднем венке «Lunaria» (1913) мужская цезура присутствует лишь как тенденция, только в некоторых стихотворениях, зато в венке «Corona Astralis» (1909) она употребляется в каждом:

Кому земл́я —|| священный край изгнанья,  
Того простóр|| полей не веселит,  
Но каждый ша́г, || но каждый миг таит  
Иных мирóв|| в себе напоминанья.

В этом же русле движется И. Северянин — великий новатор цезурного стиха, который в 5-стопном ямбе выступает, напротив, почти консерватором. Размеры с подчеркнутой мужской цезурой приобретают у него оттенок стилизации (прежде всего любимого им И. Мятлева) и допускают отдельные отступления в согласии с развитием темы. Наиболее известное стихотворение в этом духе может быть названо манифестом поэтического стилизаторства<sup>8</sup>:

В те временá, || когда роились грезы  
В сердцах люде́й, || прозрачны и ясны,  
Как хороши́, || как свежи были розы  
Моей любви́, || и славы, и весны!..

<sup>8</sup> Интересно, что молодой Г. Иванов, находящийся под сильным впечатлением от творчества Северянина, вслед за ним также пользуется цезурным 5-стопником, всегда исключительно со слабой цезурой.

<...>

Прошли лета́, || и всюду льются слёзы...  
 Нет ни страны́, || ни тех, кто жил в стране...  
 Как хороши́, || как свежи были розы  
 Воспомина́||'ний о минувшем дне! \*

*И. Северянин, «Классические розы», 1925*

Нужно заметить, что цезурный вариант размера в XX в. часто употреблялся в стилизациях (Гаспаров 2002: 239), причем и размеры с сильной цезурой, и размеры со слабой одинаково отсылают к поэтической классике XIX в. В силу этого поэты могут «маскировать» цезурный словораздел — так, что опознать его может только подлинный ценитель классического стиха:

Ты белый сти́х   в обыча́й ввел отныне	м.ц.
Для дру́жеских   посланий. В добрый час,	д.ц.
Далекий дру́г,    но смутно близкий часто	м.ц.
Моей душо́,    как Иппокрены звон	м.ц.
Сквозь голосо́   толпы любимцу Муз	м.ц.
Изда́лека   ежеминутно внятн...	м.ц.

*Вяч. Иванов, «Послание на Кавказ», 1912*

Не радуйся    возвышенному дару,	д.ц.
Богатая,    звучащая душа;	д.ц.
Не верь словам,    что просятся, спеша,	м.ц.
Преодолеть    немых прозрений мару.	м.ц.

*Ю. Верховский, <1917>*

В стихотворении Иванова цезура не сопровождается ослаблением синтаксических связей в окрестности словораздела, а распределение полустиший (35 % дактилических, 65 % мужских на 140 строк) практически не отличается от речевой модели: таким образом, позиция цезурного словораздела никак не подчеркивается дополнительно и даже скрывается тем, что слова с левой и с правой ее окрестности синтаксически плотно примыкают друг к другу (строки вроде *Сквозь голосо́ || толпы любимцу Муз*).

В младшем поэтическом поколении интерес к цезуре в 5-стопном ямбе в целом ослабевает: у акмеистов изредка используется цезура «бальмонтовского» типа — иногда в более, иногда в менее регулярном виде. Явный интерес к цезурному ямбу испытывает только М. Цветаева (18 текстов) и в какой-то мере О. Мандельштам (7 текстов). У них в отдельную группу выделяются стихотворения, где мужская цезура доминирует, но допускает разного рода отступления — такие же, как у И. Северянина, с которым они в этом сходятся:

Андрей Шеньё|| взошел на эшафот,  
А я живу́—|| и это страшный грех.  
Есть времена́—|| железные — для всех.  
И не певец, || кто в порохе — поет.

И не отец, || кто с сына у ворот  
Дрожа срыва||'ет воинский доспех.  
Есть времена́, || где солнце — смертный грех.  
Не человек—|| кто в наши дни живет.

*М. Цветаева, «Андрей Шеньё», 1918*

Огромный па́рк. || Вокзала шар стеклянный.  
Железный ми́р|| опять заморожен.  
На звучный пи́р|| в элизиум туманный  
Торжественно|| уносится вагон:  
Павлиний кри́к|| и рокот фортепьянный.  
Я опозда́л. || Мне страшно. Это — сон.

*О. Мандельштам, «Концерт на вокзале», 1921*

В поэзии эмиграции интерес к цезуре, напротив, повышается — видимо, в последний раз в русском стихосложении. Причиной этого можно считать ту же ориентацию на классические образцы (пусть даже с привкусом мифологизации), о которой шла речь выше. При этом интерес к цезуре замечен в большей мере у поэтов «второго» ряда: среди «европейских» эмигрантов к ней регулярно прибегает Б. Божнев и В. Набоков, особой популярностью она пользуется у «китайской» части русской эмиграции — у А. Несмелова и В. Перелешина, перед которыми, видимо, более остро стояла задача сохранения классической метрики. У всех этих поэтов преобладает слабая цезура, часто с единичными нарушениями словораздела, а некоторые стихотворения демонстрируют «бальмонтовскую» тенденцию к мужской:

К глубокому || столбу привязан пруд...  
Свисает цепь, || своих длиннее звеньев...  
И две ступе||'ни черные ведут,  
Нет, лишь одна — || в чистилище гниенья.

*Б. Божнев, <1939>*

А горизонт — || он выгоренно-дымчат —  
Как ветхих ряс || тепло шуршащий шелк...  
Мой взгляд грустнел, || в нем растворялось «нынче»,  
К далекому || парящий дух ушел.

*А. Несмелов, <1922>*

Во второй половине XX в. цезурный 5-стопный ямб фактически исчезает. Им немного пользуется молодой Бродский («Воротишься на родину. Ну что ж...», 1961), тяготеющий к мужской цезуре, и Аронзон («Троллейбусы уходят в темноту...», <1962>), чаще всего предпочитающий слабую цезуру. Но даже у них количество таких стихотворений не доходит до десяти, у других же поэтов (Б. Ахмадуллиной, О. Чухонцева, А. Еременко, М. Айзенберга) редко превышает два-три образца с заметным количеством отступлений.

Совсем другая история у экспериментов с «неканонической» цезурой, которые продолжились спустя почти два столетия после ломоносовского перевода из Горация. Известно одно стихотворение Черубины де Габриак, в котором вместо привычного цезурного членения 2 + 3 использовано членение 3 + 2 [Гаспаров 2001: 82—83]:

Он долго говорѣл,    и вдруг умолк.	м.ц.
Мерцали нам со стѣн    сияньем бледным	м.ц.
Инфант Велѣскеса    тяжелый шелк	д.ц.
И русый Тициѣн   с отливом медным...	м.ц.

*«Он улыбается», 1910*

Заметим, что цезура в этом стихотворении — слабая, что вкупе с самым выбранным членением, по словам М. Л. Гаспарова, может восприниматься как «знак установки на избранный круг мастеров и знатоков стиха, которые только и смогут расслышать и оценить непривычную новацию» (Там же: 83). Среди более поздней поэзии такие примеры хотя и очень редко, но также встречаются, например в «Стрельнинской элегии» (1960) Иосифа Бродского, где неканоническая позиция цезуры подчеркнута регулярным мужским окончанием в согласии со вкусом эпохи:

Как будто бы зимой|| в деревне царской  
является мне тѣнь|| любви напрасной,  
и жизнь опять бежит|| во мгле январской  
замерзшею волной|| на берег прекрасный.

Существует несколько стихотворений, в которых цезура после третьей стопы стремится быть дактилической — такой размер фактически совпадает с 4-стопным ямбом с дактилическими наращенными на цезуре. Иногда тяготение к этому размеру присутствует как тенденция, например в некоторых стихотворениях Зинаиды Гиппиус возникает стремление удержать дактилическую цезуру, но на этом фоне случаются отдельные отклонения. Лишь в позднем стихотворении «Дни» (1918) эта тенденция выражена настолько последовательно, что 5-стопный ямб превращается в размер с цезурными наращенными:

К стыду и гордости — || равнопрезрение...  
Всему покорственный|| привет без битвы...



Тяжеле всех грехо́в — || Богоубьение,  
Жизнь без прокля́тия — || и без молитвы.  
*«Что есть грех?..», 1902*

Она шерша́вая, || она колючая,  
Она холо́дная, || она змея.  
Меня изра́нила|| противно-жгучая  
Ее колéнчата||<sup>?</sup>я чешуя.  
*«Она», 1905*

Все дни изло́маны, || как преступлением,  
Седого Вре́мени|| заржавел ход.  
И тело ско́вано|| оцепенением,  
И сердце сда́влено, || и кровь — как лед.  
*«Дни», 1918*

Подобные же «переходные» формы встречаются и у современников Гиппиус, например у Валерия Брюсова, который может более резко нарушать возникшую было дактилическую тенденцию:

В наемной ко́мнате|| все ранит сердце:  
И рама зе́ркала, || и стульев стиль,  
Зачем-то со стены|| глядящий Герцен,  
И не смети́нная|| с комода пыль.  
*«В наемной комнате», <1912>*

По всей видимости, первые два из этих стихотворений были источниками для того полномасштабного пересмотра репертуара цезурных размеров, который проделал Игорь Северянин. В то же время позднее стихотворение Гиппиус «Дни» отражает уже следующую стадию развития цезуры — когда цезурные эффекты становятся законной частью новой русской метрики.

Наконец, последнее, но едва ли не самое важное — размеры на основе 5-стопного ямба, которые допускают амбивалентную интерпретацию. Такие размеры, по всей видимости, были одним из основных источников для развития неклассической метрики — дольников, некоторые ритмические формы которых совпадают с силлабо-тоническими размерами. Характерный пример — стихотворение Осипа Мандельштама «Дано мне тело...» (1909), которое формально укладывается в 5-стопный ямб. В нем нет канонической цезуры после второй стопы и даже намека на нее — словораздел в этой позиции намеренно избегается: он всегда смещен либо на один, либо на несколько слогов вправо. В то же время словоразделу после третьей стопы часто предшествуют дактилические окончания (как в примерах из Гиппиус), благодаря которым возникает ощущение, что строка должна делиться на две части; они соблюдаются нерегулярно, но достаточно часто, чтобы

остальные строки прочитывались в этом же русле. В итоге 5-стопный ямб начинает звучать как 4-иктный дольник с вариациями в окрестности цезуры:

Дано́ мне те́ло —    что мне де́лать с ни́м,	ж.ц.
Таки́м еди́ным   и таки́м мои́м?	ж.ц.

За ра́дость ти́хую   дыша́ть и жи́ть	д.ц.
Кого́, скажи́те,    мне благода́ри́ть?	ж.ц.

Я и садо́вник,    я же и цвето́к,	ж.ц.
В темни́це ми́ра   я не одино́к.	ж.ц.

На сте́кла ве́чности   уже́ легло́	д.ц.
Моё́ дыха́ние,    моё́ тепло́.	д.ц.

Запечатле́ется   на нём узо́р,	д.ц.
Неузнава́емый   с неда́вних по́р.	д.ц.

Пулка́й мгнове́ния   стека́ет му́ть —	д.ц.
Узо́ра ми́лого   не зачеркну́ть.	д.ц.

Таким образом, размер этого стихотворения можно считать одновременно и 5-стопным ямбом, и 4-иктным дольником с цезурными эффектами<sup>9</sup>. Подобные «амбивалентные» размеры представляют особый интерес в качестве источников для неклассической метрики, начинающей активно развиваться уже в следующую эпоху — в 1910-е гг.

## Литература

Бейли Дж. Избранные статьи по русскому литературному стиху. М.: Языки славянских культур, 2004.

Гаспаров М. Л. Современный русский стих. М.: Наука, 1974.

Гаспаров М. Л. Русский стих начала XX в. в комментариях. 2-е изд., доп. М.: Фортуна Лимитед, 2001.

Гаспаров М. Л. Очерк истории русского стиха. Метрика. Ритмика. Рифма. Строфика. 2-е изд., доп. М.: Фортуна Лимитед, 2002.

Гаспаров М. Л. Очерк истории европейского стиха. 2-е изд., доп. М.: Фортуна Лимитед, 2003.

---

<sup>9</sup> В качестве параллели к этому размеру можно привести стихи Бальмонта, написанные, как обозначал это сам поэт, «прерывистыми строками», — тактовики 1900-х гг., также полученные на основе 5-стопного ямба, но путем прицельного расшатывания исходного ямбического ритма. Подробнее см.: [Полилова 2017].

Жирмунский В. М. Теория стиха. Л.: Советский писатель, 1975.

Полилова В. О неклассическом стихе Бальмонта: ритмическая структура «прерывистых строк» // Сборник Матице српске за славистику. 2017. Књ. 92. С. 731—746.

Поэты 1790—1810-х гг. / вступ. ст. и сост. Ю. М. Лотмана; под ред. М. Г. Альтшуллера. Л.: Советский писатель, 1971.

Тарановский К. Русские двусложные размеры // Тарановский К. Русские двусложные размеры. Статьи о стихе / пер. с серб. В. В. Сонькина. М.: Языки славянских культур, 2010. С. 13—363.

Тредиаковский В. К. Способ к сложению российских стихов, против выданного в 1735 году исправленный и дополненный // Тредиаковский В. К. Сочинения и переводы как стихами, так и прозою. Литературные памятники. СПб.: Наука, 2009. С. 69—99.

Шанур М. И. Metrum et rhythmus sub specie semioticae // Шанур М. И. Universum versus: Язык — стих — смысл в русской поэзии XVIII—XX веков. М.: Языки славянских культур, 2000. С. 91—130.

Wagenknecht Ch. Deutsche Metrik. Eine historische Einführung. 5. Aufl. München: Verlag C. H. Beck, 2007.

**K. M. Korchagin**

*Vinogradov Russian Language Institute  
of the Russian Academy of Sciences  
(Russia, Moscow)  
stivendedal@gmail.com*

## **CAESURA IN THE RUSSIAN IAMBIC PENTAMETER AND ITS IMPACT ON THE RHYTHMIC EVOLUTION OF THIS METER**

The present paper regards iambic pentameter which is one the most frequent meters in Russian poetry and one of the best examined. Its rhythmic patterns had been studied by Kirill Taranovsky with Mikhail Gasparov in their notably works on Russian versification. Nevertheless, a number of rhythmic relevant facts had been neglected in these studies, especially an issue of *caesura*. In popular handbooks, iambic pentameter is still regarded as *caesured meter* although most of the examples of this meter definitely demonstrate lack of caesura. Despite that, the history of caesura in Russian iambic pentameter can be a subject of certain investigation because of its connection to the history of non-traditional modernist metrics. In the first iambic pentameters, the caesura after the second foot preceded by fixed accent was applied by Vasilij Trediakovsky and Aleksander Sumarokov. This interpretation of caesura was quite different from such in the iambic hexameter where these poets always implemented caesura without a fixed accent. Quite soon, the iambic pentameter was vanishing from the repertoire of the epoch and returned at the end of the century. In the late 1810s, there was a dispute on caesura in the iambic pentameter

between Vasilii Zhukovskiy and Aleksandr Pushkin: the elder poet in his translation of Johann Jakob Hebel's *Die Vergänglichkeit* showed implementation of un-caesured verse in narrative poetry, and then Pushkin wrote a quite malign epigram on this Zhukovsky's oeuvre in order to return to caesured verse in *Boris Godunov* several years later and, afterwards, to decline this verse in *Domik v Kolomne*. After several decades, at the early *fin-de-siècle* time, Konstatin Balmont partially returned to caesured verse in his attempts to renovate traditional Russian metrics. In his pentameters, there was no constant caesura as in Pushkin's *Boris Godunov*; instead of this, there was a bias to caesura keeping a strong resemblance to the old caesured pentameter. This bias was repeated by many poets from the early 1900s to the late 1910s where attempts to renovate the old metrics were replaced by attempts to invent the new ones. This kind of iambic pentameter was implemented by Vyacheslav Ivanov, Ivan Bunin, Zinaida Gippius, Vladislav Khodasevich, Nikolay Tikhonov and other poets. A significant poem demonstrating how these two different kinds of metrics replace each other is Osip Mandelshtam's "Danomnetelo..." (1909). The verse of this poem can be interpreted both as iambic pentameter and as four-accented *dolnik* (strict stress meter in terms of English versification). This ambiguity is an effect of lacking routine caesura after the second foot followed by the frequent paroxytonal and proparoxytonal accents after the third. All of these make an impression of a symmetric two-segmented and two-accented verse pattern. One can say this poem was a summit in the history of caesured pentameter: in the following 20th century poetry, caesured iambic pentameter was used only by the poets who wanted to stylize 19<sup>th</sup> century poetry. Unlike the un-caesured iambic pentameter, the caesured iambic pentameter was one of the rarest meters in the 20th century Russian poetry.

*Key words:* metrics, theory of verse, iambic pentameter, caesura, the evolution of Russian verse, poetic corpus, Russian National Corpus

## References

Bailey J. *Izbrannye stat'i po russkomu literaturnomu stikhu* [Collected Works on the Russian Literary Verse]. Moscow, Yazyki slavyanskikh kul'tur Publ., 2004. (In Russ.)

Gasparov M. L. *Ocherk istorii evropeiskogo stikha* [A History of European Versification]. Moscow, Fortuna Limited Publ., 2003 (In Russ.)

Gasparov M. L. *Ocherk istorii russkogo stikha. Metrika. Ritmika. Rifma. Strofika* [History of the Russian Verse Outline]. Moscow, Fortuna Limited Publ., 2002. (In Russ.)

Gasparov M. L. *Russkii stikh nachala XX v. v kommentariyakh* [Early 20th Century Russian Verse with Notices]. Moscow, Fortuna Limited Publ., 2001. (In Russ.)

Gasparov M. L. *Sovremennyyi russkii stikh* [Modern Russian Verse]. Moscow, Nauka Publ., 1974. (In Russ.)

*Poety 1790—1810-kh gg.* [1790s—1810s poets]. Leningrad, Sovetskii pisatel' Publ., 1971. (In Russ.)

Polilova V. *O neklassicheskom stikhe Bal'monta: ritmicheskaya struktura «pre-ryvistykh strok»* [On the Balmont's Non-Classic Verse in "pre-ryvistyestroki"]. *Zbornik Matitse srpske za slavistiku*, 2017, iss. 92, pp. 731—746. (In Russ.)

Shapir M. I. *Metrum et rhythmus sub specie semioticae*. In: Shapir M. I. *Universum versus: Yazyk — stikh — smysl v russkoipoezii XVIII—XX vekov* [Universum versus: Language — Verse — Meaning in 18<sup>th</sup>—20<sup>th</sup> century Russian Poetry]. Moscow, Yazyki slavyanskikh kul'tur Publ., 2000, pp. 91—130. (In Russ.)

Taranovskii K. [Russian Binary Meters]. Taranovskii K. *Russkie dvuslozhnye razmery. Stat'i o stikhe* [Russian Binary Meters. Works on Verse]. Moscow, Yazyki slavyanskikh kul'tur Publ., 2010, pp. 13—363. (In Russ.)

Trediakovskii V. K. [A New and Brief Way for Composing of Russian Verses]. Trediakovskii V. K. *Sochineniya I perevody kak stikhami, tak I prozoyu* [Oeuvres and Translations in Verse and Prose]. Saint Petersburg, Nauka Publ., 2009, pp. 69—99. (In Russ.)

Wagenknecht Ch. *Deutsche Metrik. Eine historische Einführung*. München, Verlag C. H. Beck, 2007.

Zhirmunskii V. M. *Teoriya stikha* [Verse Theory]. Leningrad, Sovetskii pisatel' Publ., 1975. (In Russ.)

## **«ПОКАЗАТЕЛИ БЕЗРАЗЛИЧИЯ» В РУССКО-НЕМЕЦКОМ КОРПУСЕ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ\***

Цель данного исследования — выявить на основе анализа корпусных данных, насколько интуитивные представления об эквивалентности структурно схожих немецких и русских конструкций соответствуют их реальному употреблению. В качестве материала используются данные русско-немецкого и немецко-русского параллельных корпусов НКРЯ. Анализируются контексты с такими конструкциями, как *родственники родственниками, но..., мнения мнениями, а...,* то есть фразами, структурно основанными на паттерне [X X-ом, но/а Р] и обладающими специфической семантикой. Несколько огрубленно такие конструкции можно охарактеризовать как «показатели безразличия», поскольку, употребляя эту структуру, говорящий как бы выводит сущность X из сферы релевантных объектов обсуждения. При этом на первый план может выходить идея снижения значимости аргумента собеседника или несогласие с его предложением, чисто уступительное значение или равнодушие говорящего по отношению к объективным обстоятельствам. На первый взгляд паттерн [X X-ом, но/а Р] имеет в немецком языке хороший, то есть близкий и по форме, и по значению, эквивалент — фразеологизм-конструкцию [X *hin*, X *her*]. Обращение к параллельному корпусу позволило сделать вывод об отсутствии стабильной эквивалентности между этими конструкциями. Их параллельное употребление в аутентичных контекстах крайне нетипично. Причины отсутствия функциональной эквивалентности анализируются в статье.

*Ключевые слова:* параллельный корпус, лексическая семантика, конструкции, межкультурная эквивалентность, русский язык, немецкий язык.

Современное состояние параллельных корпусов НКРЯ позволяет постепенно начинать исследовать явления, относительно редко встречающиеся в текстах. Тема настоящей работы — сопоставление некоторых идиоматичных способов выражения безразличия в русском и немецком языках. Изучению выражений русского языка с семантикой ‘безразлично А или не-А, потому что В’ и их английских эквивалентов посвящена работа [Голубцов 1999]. В ней описываются, с одной стороны, такие грамматические фразеологизмы, как *мало ли, все равно*, а с другой — восклицательные и вопросительные высказывания типа *Что за дело/беда! Ну и что?, А мне-то что?* и идиомы разной структуры *до лампочки, до фонаря, один черт*,

---

\* Работа выполнена в рамках Программы фундаментальных исследований Президиума РАН «Памятники материальной и духовной культуры в современной информационной среде» по проекту «Пополнение и развитие корпуса параллельных текстов в составе НКРЯ».

хозяин — барин, в том числе речевые формулы *вольному воля, где наша не пропадала, семь бед — один ответ*. В качестве английских коррелятов рассматриваются такие выражения, как *it is all the same (to me), it's all one to me; it is the same thing; it is the same damned thing; it comes to the same damn thing, it doesn't make any difference, What's the difference?, there's no point in it; that is nothing to me; it is no business/concern of mine; I am not concerned; it is not my problem/affair; it doesn't refer to me*.

В данной статье нас интересуют несколько другие выражения. Так, формы [X X-ом, но/а Р], [хоть X, хоть не-X, (а) Р] или [что X, что не-X, (а) Р] хорошо вписываются в группу конструкций с базовой семантикой безразличия. Ср.: *Дождь дождем, а я все равно пойду гулять* или *Хоть отпуск, хоть не отпуск, а работу все равно надо будет сдать в срок* или *Что воля, что неволя...* Формы [X X-ом], [хоть X, хоть не-X] и [что X, что не-X] синонимичны выражениям типа *Что за дело/беда, что X, Ну и что, что X?* и соответствуют формуле ‘безразлично А или не-А, потому что В’, предложенной в [Голубцов 1999].

Наряду с семантикой безразличия интересующие нас конструкции обладают уступительно-противительным значением; ср. [Булыгина, Шмелев 1997; Шмелев 2002; Апресян 2015]. Валентина Апресян [Апресян 2015: 60] указывает на полемический потенциал подобных выражений, предполагающих скрытую полемику с собеседником. Их значение можно приблизительно описать так: ‘хотя ты и говоришь, что X имеет место и, возможно, является для тебя значимым фактором, для меня X несущественен и не является препятствием для Р’. Естественно, эта схема значения весьма приближительна и может модифицироваться в конкретных контекстных условиях. На первый план может выходить идея снижения значимости аргумента собеседника или несогласие с его предложением, чисто уступительное значение или равнодушие говорящего по отношению к объективным обстоятельствам. Таким образом, отнесение этих выражений к «показателям безразличия» до некоторой степени условно.

С одной стороны, достаточно очевидно, что эти выражения характеризуются определенной лингвоспецифичностью. С другой — не менее очевидно, что выражаемые ими смыслы в силу их коммуникативной значимости должны выражаться и в других языках. На первый взгляд в немецком языке у русских конструкций [X X-ом, но/а Р], [хоть X, хоть не-X, (а) Р] и [что X, что не-X, (а) Р] есть хорошие системные эквиваленты. Паттерн [X X-ом] коррелирует с немецким фразеологизмом-конструкцией<sup>1</sup> [X *hin*, X *her*]; ср.: *Krise hin, Krise her. Doch eine genauere Analyse zeichnet ein anderes Bild*. — *Кризис кризисом, а более точный анализ дает другую картину*. Конструкциям с повторяющимися *хоть* и *что* соответствуют немецкие конструкции с союзом *ob*. Ср.: *Ob Krise, ob nicht, ob mit Kindern oder ohne*. — *Хоть кризис, хоть не кризис, хоть с детьми, хоть без*. ≈ *Что кризис, что не кризис, что с детьми, что без*.

<sup>1</sup> О классе фразеологизмов-конструкций (англ. constructional phrasemes, нем. Phrasem-Konstruktionen) см. [Dobrovolskij 2011; 2015; Баранов, Добровольский 2013; Добровольский 2016].



Соответствующие русские и немецкие конструкции схожи не только семантически, но и по предпочтениям в заполнении свободного слота. Так, русское выражение, построенное по схеме [X X-ом], однозначно требует заполнения позиции X существительным, причем и первый (существительное в именительном падеже), и второй слот (существительное в творительном падеже) должны обязательно заполняться одним и тем же существительным. На аналогичные ограничения в заполнении слота X немецкой конструкции [X *hin*, X *her*] неоднократно указывалось в специальной литературе [Černyševa 1986: 213; Fleischer 1997: 132]<sup>2</sup>. Напротив, русские конструкции с *хоть* или *что* и их немецкие аналоги с *ob* допускают в позиции X разные слова без ограничений на их частеречную принадлежность: *Ob dieses, ob jenes, kommt auf eins hinaus. — Что то, что это <хоть то, хоть это> — все равно кончится одним и тем же*. Ср. также пример (1) из русско-немецкого параллельного корпуса НКРЯ:

- (1) И! Пахомовна, у нас *что свисти, что нет*: а денег все нет как нет. [А. С. Пушкин. Дубровский (1832—1833)]  
Ih! Pachomowna, *ob bei uns gepfiffen wird oder nicht*: Geld ist niemals da. [Alexander Puschkin. Dubrowskij (Arthur Luther, Michael Pfeiffer, 1984)]

Цель данного исследования — выявить на основе анализа корпусных данных, насколько интуитивные представления об эквивалентности структурно схожих немецких и русских конструкций соответствуют их реальному употреблению. В этой статье мы сконцентрируемся на функционировании конструкции [X X-ом, но/а Р] и ее немецких соответствий.

Обратимся к русско-немецкому параллельному корпусу. При поиске были заданы следующие параметры:

Слово1 Грамм. признаки: S,nom

Слово2 Грамм. признаки: S,ins

Слово2 Доп. признаки и языки: lexred

В результате поиска были собраны контексты, в которых форма существительного в творительном падеже непосредственно следует за формой того же существительного в именительном падеже. После исключения из материала примеров с омонимичными конструкциями типа *дурак дураком* и взаимного местоимения

<sup>2</sup> Анализ материала корпуса DeReKo показал, что данные ограничения не существенны для современного употребления конструкции [X *hin*, X *her*]. В немецкоязычных корпусах встречаются случаи, когда позицию X занимают слова других частей речи; ср. *einverstanden hin, einverstanden her*, что можно перевести на русский язык как *согласиться-то можно, но...* или *согласен-то я согласен, но...* Что касается требования заполнения слота X идентичным словом, оно также не соблюдается. По данным корпуса DeReKo, больше половины употреблений этой конструкции обнаруживают разные слова в позиции перед *hin* и *her*. Ср.: *Schauer hin, Graupel her, das Spiel findet statt. ≈ Хоть ливень, хоть град — матч всё равно состоится*. В дальнейшем нас будут интересовать только случаи заполнения слота X идентичным существительным, так как только в этих случаях ожидается потенциальная эквивалентность форм [X *hin*, X *her*] и [X X-ом].

друг другом остались единичные контексты (2)—(5). Несмотря на их немногочисленность, эти примеры позволяют обнаружить определенную тенденцию: ни один из контекстов с конструкцией [X X-ом, но/а Р] не переводится на немецкий язык с помощью конструкции с формой [X *hin*, X *her*], хотя эта форма, казалось бы, полностью эквивалентна русскому паттерну [X X-ом].

- (2) — Надеюсь увидеть вас еще, но надеюсь тоже, что вы перемените свои мнения, мой милый мсье Пьер, — сказала она. Когда она сказала ему это, он ничего не ответил, только наклонился и показал всем еще раз свою улыбку, которая ничего не говорила, разве только вот что: «Мнения мнениями, а вы видите, какой я добрый и славный малый». И все, и Анна Павловна невольно почувствовали это. [Л. Н. Толстой. Война и мир, том 1 (1865—1869)].  
«Ich hoffe Sie nun öfter zu sehen, aber ich hoffe auch, daß Sie Ihre Meinungen ändern werden, mein lieber Monsieur Pierre.» Pierre antwortete nichts darauf, verneigte sich nur und zeigte noch einmal allen sein Lächeln, das nichts anderes besagte als höchstens: *Es gibt nun einmal verschiedene Meinungen*, aber Sie sehen ja, was für ein guter und prächtiger Junge ich bin. Und alle, auch Anna Pawlowna, fühlten das. [Lew Tolstoi. Krieg und Frieden (1. Band) (Hermann Röhl, 1922)].
- (3) — Да что ж пенька? Помилуйте, я вас прошу совсем о другом, а вы мне пеньку суете! *Пенька пенькою*, в другой раз приеду, заберу и пеньку. [Н. В. Гоголь. Мертвые души (1835—1852)].  
„Was soll ich mit dem Hanf? Du lieber Gott, ich bitte Sie um etwas ganz anderes, und Sie bieten mir Hanf an! *Lassen wir den Hanf Hanf sein*, wenn ich ein andermal vorbeikomme, nehme ich auch den Hanf.“ [Nicolaj Gogol. Die toten Seelen (Michael Pfeiffer, 1978)]
- (4) Генерал Бетрищев, близкий приятель и, можно сказать, благотворитель, просил навестить родственников... Конечно, *родственники родственниками*, но отчасти, так сказать, и для самого себя: ибо видеть свет, коловращенье людей — кто что ни говори, есть как бы живая книга, вторая наука. [Н. В. Гоголь. Мертвые души (1835—1852)].  
Der General Betriscschew, ein naher Freund und, ich darf wohl sagen, mein Wohltäter, bat mich, seine Verwandten aufzusuchen? Natürlich, *die Verwandten sind wichtig*, doch ich reise zum Teil sozusagen auch zum eigenen Vergnügen: Denn die Bekanntschaft mit der Welt und dem Treiben der Menschen — da sage einer, was er will, das ist wie ein lebendes Buch, wie eine zweite Wirtschaft. [Nicolaj Gogol. Die toten Seelen (Michael Pfeiffer, 1978)]
- (5) А если видишь еще, что всё это с какой целью творится, как вокруг тебя всё множится да множится, принося плод да доход. Да и я рассказать не могу, что тогда в тебе делается. И не потому, что растут деньги. *Деньги деньгами*. Но потому, что всё это дело рук твоих; потому, что видишь, как ты всему причина, ты творец всего, и от тебя, как от какого-нибудь мага, сыплется изобилие и добро на всё. [Н. В. Гоголь. Мертвые души (1835—1852)].

Und wenn man dann noch sieht, zu welchem Ziel dies alles geschaffen wird, wie sich ringsum alles mehrt und mehrt, Früchte trägt und Gewinn bringt... Ich kann gar nicht erzählen, wie einem da zumute ist. Und nicht deshalb, weil das Geld zunimmt. *Das Geld ist nicht das Wichtigste*. Sondern deshalb, weil all dies das Werk deiner Hände ist, weil du siehst, daß alles von dir ausgeht, daß du der Schöpfer von allem bist und wie ein Zauberer Überfluß und Wohlstand verteilst. [Nicolaj Gogol. Die toten Seelen (Michael Pfeiffer, 1978)].

Вопрос, почему при наличии столь близкого системного эквивалента переводчики предпочитают другие способы передачи этого смысла, достаточно сложен. Проще всего было бы списать это на ограниченность корпусных данных и признать отсутствие переводческой эквивалентности<sup>3</sup> артефактом анализа. Тем более, что три из четырех проанализированных примеров взяты из одного произведения, переведенного одним переводчиком — М. Пфайфером. В такой ситуации естественно предположить, что выбор эквивалента в значительной степени продиктован индивидуальными вкусами переводчика. Это предположение сталкивается, однако, с возражениями двух типов. Во-первых, как будет показано ниже, отсутствие переводческой эквивалентности между [X X-ом] и [X *hin*, X *her*] наблюдается и при обращении к немецко-русскому параллельному корпусу. Следовательно, речь идет о некоей тенденции, не ограничивающейся направлением перевода, то есть о феномене, который явно выходит за рамки индивидуальных предпочтений переводчика. Во-вторых, семантика.

Обращение к одноязычному немецкому корпусу DeReKo (das Deutsche Referenzkorpus des Instituts für Deutsche Sprache in Mannheim — Институт немецкого языка, Мангейм) показало, что выражения, образованные на основе паттерна [X *hin*, X *her*], в отличие от русского [X X-ом], редко употребляются в противительной конструкции. Если русская форма [X X-ом] — это часть конструкции [X X-ом, но/а P], немецкое выражение [X *hin*, X *her*] само часто вводится усилительной частицей *doch*, а после него крайне редко следует клауза, вводимая противительными союзными словами типа *aber*, *doch*<sup>4</sup> или *jedoch*. Соответствующая пропозиция P следует за [X *hin*, X *her*], как правило, без всякого коннектора. Из 6409 встретившихся в DeReKo употреблений выражения [X *hin*, X *her*] только около 30 включены в контекст эксплицитного противопоставления с *aber*, *doch* или *jedoch*. Таким образом, русской конструкции [X X-ом, но/а P] соответствует немецкая [X *hin*, X *her*, P], а не [X *hin*, X *her*, *aber/doch/jedoch* P], как можно было бы предположить на основе структурного сходства.

Заметим, что противопоставление, часто вводимое союзами *но* и *а*, может отсутствовать и в употреблении русской конструкции [X X-ом]. Ср. примеры (3) и (5). В целом, однако, контексты противопоставления, даже и не вводимые

<sup>3</sup> О соотношении понятий системной и переводческой эквивалентности см. [Добровольский 2011].

<sup>4</sup> В случаях, когда *doch* вводит противопоставление, эту лексему принято описывать как союзное слово — грамматический омоним усилительной частицы *doch*.

противительным союзом, характерны для русской конструкции [X X-ом]. Так, в примере (3) фраза *в другой раз приеду, заберу и пеньку* содержательно противопоставлена предшествующей фразе *пенька пенькою*, которая выражает нежелание Чичикова обсуждать с Коробочкой условия покупки пеньки, то есть является показателем безразличия, как бы выводя этот объект обсуждения из области значимых тем. В примере (5) противительный союз *но* формально присутствует, однако он вводит противопоставление, относящееся не к непосредственно предшествующей фразе *деньги деньгами*, а к более раннему контексту, ср.: *И не потому, что растут деньги. <...> Но потому, что всё это дело рук твоих*. Именно по этой причине пример (5) целесообразно рассматривать в одной группе с (3). В этом случае семантика противопоставления как бы разлита в контексте, но формально паттерн [X X-ом, но/а Р] реализуется не полностью. Наличие или отсутствие эксплицитного противопоставления *но/а Р* является значимым моментом для выбора немецкого соответствия. В переводе контекстов эксплицитного противопоставления, вводимого союзами *но* и *а*, более естественными оказываются конструкции типа [*es gibt nun einmal X, aber Р*] или [*X ist wichtig, doch Р*]. Ср. способы перевода русской конструкции [X X-ом] в примерах (2) и (4): *Es gibt nun einmal verschiedene Meinungen, aber Sie sehen ja...* и *die Verwandten sind wichtig, doch ich reise...* Говорящий признает значимость обсуждаемой сущности в первой части высказывания, для того чтобы тут же указать на ее незначимость: ‘X важен, но Р’. Говорящий как бы соглашается с мнением собеседника, чтобы сразу указать на более важные моменты ситуации. Если же эксплицитное противопоставление, вводимое противительным союзом, отсутствует в контексте, конструкция [X X-ом] переводится фразой со значением ‘X неважен’. Ср.: *Das Geld ist nicht das Wichtigste*. В этом случае уступительность, присущая этой русской конструкции, в переводе теряется.

Отсутствие параллелизма между выражениями, образованными по схеме [X X-ом], и выражениями на основе паттерна [X *hin*, X *her*] подтверждается и при обращении к немецко-русскому параллельному корпусу. Поскольку этот подкорпус не содержит ни одного примера употребления немецкой конструкции [X *hin*, X *her*], поиск осуществлялся по русской части на основе тех же параметров и принципов отбора материала, что и в русско-немецком корпусе. Поиск дал следующие результаты:

- (6) „Warum sollte er dir Blumen schicken?“ „Wir kennen uns schon lange,“ sagte sie, „und vielleicht verehrt er mich.“ — „Gut,“ sagte ich, „soll er dich verehren, aber soviel kostbare Blumen, das ist aufdringlich.“ [Heinrich Böll. *Ansichten eines Clowns* (1963)]
- «<...> Почему, собственно, он послал тебе цветы? — Мы старые друзья, — сказала она, — может быть, он мой поклонник. — Очень мило, — сказал я, — *поклонник поклонником*, но дарить такой большой букет дорогих цветов — значит навязываться. [Генрих Бёльль. Глазами клоуна (Р. Райт-Ковалева, 1964)]

- (7) Wenn er sie erst hat, gibt es nur noch Königsberger Klopse mit deutscher Tunke. Und ich habe keine Sorge, daß Gerda das nicht auch weiß. Trotzdem beschließe ich, mit ihr nach dem Essen zusammen wegzugehen. *Vertrauen ist zwar Vertrauen*, aber Eduard hat zuviel verschiedene Liköre in der Bar. [Erich Maria Remarque. Der schwarze Obelisk (1956)]  
Когда он ее получит, ей опять будут подавать только битки по-кенигсбергски с немецким соусом. И я уверен, что Герде это тоже известно. И все-таки я решаю после ужина уйти вместе с нею. *Доверие — доверием*, но у Эдуарда в погребе слишком много крепких напитков. [Эрих Мария Ремарк. Чёрный обелиск (В. Станевич, 1961)]
- (8) „Hast du beachtet, wie sie grüßte? Beinahe gar nicht. Dabei war meiner unmaßgeblichen Ansicht nach ihr Hut ganz unmäßig geschmacklos...“ „Na, was den Hut betrifft... Und mit dem Grüßen warst du wohl auch nicht viel entgegenkommender, meine Liebe. Übrigens ärgere dich nicht; das macht Falten.“ [Thomas Mann. Buddenbrooks (1896—1900)]  
Ты заметил, как она поклонилась? Едва-едва кивнула. И потом, с моей точки зрения, впрочем ни для кого не обязательной, ее шляпа верх безвкусицы. — Ну, *шляпа шляпой*... а что касается поклона, то и ты была не слишком любезна, дорогая моя! Не сердись, Тони, это старит... [Томас Манн. Будденброки (Н. Ман, 1953)]

Примеры (6)—(8), с одной стороны, подтверждают наше наблюдение, что в контексте эксплицитно выраженной противительности немецкий язык предпочитает не конструкцию [X *hin*, X *her*], а другие выражения; ср. *soll er dich verehren, aber...* — *поклонник поклонником, но...*; *Vertrauen ist zwar Vertrauen, aber...* — *доверие — доверием, но...* С другой стороны, эти примеры позволяют предположить, что русская форма [X X-ом] более употребительна, чем немецкая [X *hin*, X *her*]. Русское выражение [X X-ом] появляется в тексте перевода и в случаях, когда оригинал не содержит стимула, который в силу формальных особенностей мог бы спровоцировать употребление данного выражения. Так, вполне можно себе представить в качестве адекватного перевода фразы *soll er dich verehren, aber...* (6) формально более близкую к синтаксису оригинала фразу *пусть он даже и твой поклонник, но...* Выбор Райт-Ковалевой мотивирован, по-видимому, интуитивным ощущением большей укорененности паттерна [X X-ом] в узусе и, соответственно, его большей идиоматичности. Немецкая конструкция, употребленная в оригинальном тексте примера (6), также основана на паттерне с высокой степенью идиоматичности: [soll SUBJ OBJ V] в значении допущения, которое в следующей клаузе, вводимой союзным словом *aber*, признается нерелевантным. Таким образом, перевод в примере (6), далекий от оригинала по структуре, оказывается хорошим функциональным эквивалентом, поскольку и та и другая формы, вводя в дискурс допущение истинности пропозиции, семантически вынуждают последующее противопоставление. К тому же обе эти формы характеризуются высокой частотностью в узусе и выражают идею безразличия говорящего по отношению

к обсуждаемому положению вещей; ср. близкое по семантике высказывание *Мне-то какая разница, поклонник он твой или нет*. Ср. также пример (7), в котором обе формы *Vertrauen ist zwar Vertrauen, aber...* и *доверие — доверием, но...* близки и семантически, и функционально<sup>5</sup>. Вопрос, почему в немецком оригинале употребляется конструкция *[X ist zwar X]*, а не теоретически возможная конструкция *[X hin, X her]*, объясняется не только авторским выбором, но и узуальными ограничениями на заполнение слота X в конструкции *[X hin, X her]*. В корпусе DeReKo не обнаруживается ни одного примера выражения *Vertrauen hin, Vertrauen her* и есть лишь один пример на употребление лексически близкого выражения; ср. *Gottvertrauen hin, Gottvertrauen her* (источник: Z80/SEP.00291 Die Zeit, 19.09.1980, S. 69; Zeitlese). Сходным образом, теоретически возможное в примере (6) *Verehrer hin, Verehrer her* оказывается не закрепленным в узусе: в DeReKo на такое словосочетание не нашлось ни одного примера. Пример (8) интересен тем, что в немецком оригинале полностью отсутствует значение уступительности, характерное для русской конструкции *[X X-ом]*. Ср. *Na, was den Hut betrifft...* vs. *Ну, шляпа шляпой...* Потенциально возможное в этом контексте выражение *Hut hin, Hut her*, с одной стороны, совершенно не характерно для стиля Томаса Манна, а с другой — недостаточно укоренено в узусе. В DeReKo встретился всего один контекст, и тот — как элемент языковой игры: [...] *daß der Kanzler nicht seinen fischen kanadischen Cowboyhut trug. Auf diesen scherzhaften Ton ging der Kanzler schlagfertig ein und sagte: Hut hin, Hut her — Hauptsache, die Opposition zwingt ihn nicht eines Tages, seinen Hut zu nehmen* (источник: Z77/JUL.00351 Die Zeit, 22.07.1977, S. 5; Schmidt bei Carter: Nützlich und offen).

\*\*\*

Наличие в том или ином языке лексических единиц и конструкций, близких по значению и прагматическим характеристикам, всегда ведет к существенной вариативности в формулировании заданного смысла. Такая ситуация имеет место как в русском, так и в немецком языке при выражении смыслов, связанных со снижением релевантности аргумента собеседника, с незначимостью для говорящего тех или иных аспектов обсуждаемого положения вещей и т. п. Несколько условно единицы, выражающие эти смыслы, можно назвать «показателями безразличия». Тот факт, что структурно схожие показатели безразличия русского и немецкого языков редко используются в переводе, объясняется целым рядом факторов. Во-первых, выбор способа перевода заданного смысла может зависеть от индивидуальных предпочтений переводчика. Во-вторых, значимыми факторами оказываются стилистические ограничения и конвенции узуса, в частности максимально близкие по форме и значению конструкции сопоставляемых языков могут отличаться друг от друга степенью употребительности. В-третьих, межъязыковые различия могут

<sup>5</sup> Возможный альтернативный вариант перевода фразы *Vertrauen ist zwar Vertrauen, aber...* с помощью *доверие — это, конечно, хорошо, но...* был бы более близок к оригиналу по форме, но вряд ли обнаруживает серьезные функциональные преимущества.



объясняться особенностями сочетаемости. В работах по грамматике конструкций неоднократно подчеркивалось, что любая конструкция может быть частью некоторой другой, большей конструкции, что влияет на семантику и прагматику этой встраиваемой конструкции. Так, русская конструкция [X X-ом] в стандартном случае встраивается в контекст эксплицитного противопоставления, становясь частью уступительно-противительной конструкции [X X-ом, но/а Р]. Для немецкой конструкции [X *hin*, X *her*] контексты противопоставления, вводимые такими союзными словами, как *aber*, *doch* или *jedoch*, нехарактерны. И наконец, межъязыковые контрасты могут зависеть от предпочтений в заполнении свободного слота X. Хотя на первый взгляд позиция X может заполняться практически любым существительным, анализ больших корпусов типа DeReKo показывает, что здесь существуют явные ограничения.

### Литература

Апресян В.Ю. Уступительность: механизмы образования и взаимодействия сложных значений в языке. М. : Языки славянской культуры, 2015. 288 с.

Баранов А.Н., Добровольский Д.О. Основы фразеологии (краткий курс). М. : Флинта : Наука, 2013. 312 с.

Булыгина Т.В., Шмелев А.Д. Языковая концептуализация мира (на материале русской грамматики). М. : Школа «Языки русской культуры», 1997. 576 с.

Голубцов С.А. Семантика и прагматика показателей безразличия: сопоставительный аспект: дис. ... канд. филол. наук / Кубанский гос. технологический ун-т. Краснодар, 1999. 210 с.

Добровольский Д.О. Сопоставительная фразеология: межъязыковая эквивалентность и проблемы перевода идиом // Русский язык в научном освещении / 2011. № 2 (22). С. 219—246.

Добровольский Д.О. Грамматика конструкций и фразеология // Вопросы языкознания. 2016. № 3. С. 7—21.

Шмелев Д.Н. О «связанных» синтаксических конструкциях в русском языке // Шмелев Д.Н. Избранные работы по русскому языку. М. : Языки славянской культуры, 2002. С. 413—438.

Černyševa I.I. Phraseologie // Stepanova M.D., Černyševa I.I. Lexikologie der deutschen Gegenwartssprache. 2. Aufl. Moskau : Vysšaja škola, 1986. S. 175—230.

Dobrovol'skij D. Phraseologie und Konstruktionsgrammatik // Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze / Alexander Lasch, Alexander Ziem (Hrsg.). Tübingen : Stauffenburg, 2011. S. 111—130.

Dobrovol'skij D. On the systematic variation of German idioms: Converse pairs as a constructional phenomenon // Journal of Social Sciences. 2015. Vol. 11. No. 3. P. 248—257.

Fleischer W. Phraseologie der deutschen Gegenwartssprache. 2. durchges. u. erg. Aufl. Tübingen : Niemeyer, 1997.



### **Электронные ресурсы:**

НКРЯ — Национальный корпус русского языка. URL: <http://www.ruscorpora.ru>

DeReKo — Das Deutsche Referenzkorpus des IDS Mannheim im Portal COSMAS II (Corpus Search, Management and Analysis System). URL: <https://cosmas2.ids-mannheim.de/cosmas2-web>

**D. O. Dobrovol'skij**

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

*(Russia, Moscow)*

*dobrovolskij@gmail.com*

### **“INDICATORS OF INDIFFERENCE” IN THE RUSSIAN-GERMAN PARALLEL CORPUS**

The purpose of this study is to reveal to what extent the intuitive ideas of the equivalence of structurally similar German and Russian constructions correspond to their actual use. The empirical data are based on the Russian-German and German-Russian parallel corpora of the RNC (Russian National Corpus). We analyze such constructions as *rodstvenniki rodstvennikami, no...* ‘relatives are relatives but...’, *mneniya mneniyami, a...* ‘opinions are opinions but...’, i.e. phrases structurally based on the pattern [X X-om, no/a P] ‘X is X but P’. Somewhat simplifying, such constructions can be labelled as “indicators of indifference”, because using this structure, the speaker puts the entity X outside the scope of relevant issues of discussion. Specific semantic features may come to the fore, such as downshifting of the relevance of the argument of the interlocutor or disagreement with his or her proposal, the purely concessive meaning or indifference of the speaker with regard to objective circumstances. At first glance, the pattern [X X-om, no/a P] ‘X is X but P’ has an equivalent in German, that is very close in both form and meaning, namely the constructional phraseme [X *hin*, X *her*]. The analysis allowed us to conclude that there is no functional equivalence between these structures. Their parallel use in authentic contexts from the RNC is extremely atypical. The present article discusses the reasons for the lack of functional equivalence between the analyzed constructions.

*Key words:* parallel corpus, lexical semantics, constructions, cross-linguistic equivalence, Russian, German.

### **References**

Apresyan V. Yu. *Ustupitel'nost': mekhanizmy obrazovaniya i vzaimodeystviya slozhnykh znachenii v yazyke* [Concessiveness: mechanisms of formation and interaction of complex meanings in the language]. Moscow, Yazyki slavyanskoi kul'tury Publ., 2015. 288 p. (In Russ.)

Baranov A. N., Dobrovol'skij D. O. *Osnovy frazeologii (kratkii kurs)* [Foundations of phraseology (a brief course)]. Moscow, Flinta, Nauka Publ., 2013. 312 p. (In Russ.)

Bulygina T. V., Shmelev A. D. *Yazykovaya kontseptualizatsiya mira (na materiale russkoi grammatiki)* [Language conceptualization of the world (on the basis of the Russian grammar)]. Moscow, Shkola "Yazyki russkoj kul'tury" Publ., 1997. 576 p. (In Russ.)

Černyševa I. I. Phraseologie. *Lexikologie der deutschen Gegenwartssprache* / Stepanova M. D., Černyševa I. I. 2. Aufl. Moskva, "Vysshaja škola" Publ., 1986, pp. 175—230.

Dobrovol'skij D. Phraseologie und Konstruktionsgrammatik. *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze* / Alexander Lasch, Alexander Ziem (Hrsg.). Tübingen, Stauffenburg, 2011, pp. 111—130.

Dobrovol'skij D. **On the systematic variation of German idioms: Converse pairs as a constructional phenomenon.** *Journal of Social Sciences*, 2015, vol. 11, no. 3, pp. 248—257.

Dobrovol'skij D. O. [Contrastive phraseology: cross-linguistic equivalence and translation of idioms]. *Russkii yazyk v nauchnom osveshchenii*, 2011, no. 2 (22), pp. 219—246. (In Russ.)

Dobrovol'skij D. O. [Constraction Grammar and phraseology]. *Voprosy yazykoznaniya*, 2016, no. 3, pp. 7—21. (In Russ.)

Fleischer W. *Phraseologie der deutschen Gegenwartssprache*. 2. durchges. u. erg. Aufl. Tübingen, Niemeyer, 1997.

Golubtsov S. A. *Semantika i pragmatika pokazatelei bezrazlichiya: sopostavitel'nyi aspekt*. Diss. kand. filol. nauk [Semantics and pragmatics of indifference indicators: a contrastive aspect. Cand. of philology diss.] / Kubanskii gos. tekhnologicheskii universitet. Krasnodar, 1999. 210 p. (In Russ.)

Shmelev D. N. [About "bound" syntactic constructions in Russian]. *Izbrannye raboty po russkomu yazyku* [Selected works on the Russian language]. Moscow, "Yazyki slavyanskoi kul'tury" Publ., 2002, pp. 413—438. (In Russ.)

### III. СЛОВО И ЖЕСТ: ВОПРОСЫ МУЛЬТИМОДАЛЬНОЙ КОММУНИКАЦИИ

**Т. Е. Янко**

*Институт языкознания РАН*

*(Москва, Россия)*

*tanya\_yanko@list.ru*

#### **Просодия и жестикуляция\***

Рассматриваются два типа русских просодий и соответствующая им жестикуляция. Первый просодический тип характеризуется падением частоты основного тона в существенном диапазоне частот. Это ИК-2 по классификации Е. А. Брызгуновой. В предложениях с определенным выбором словоформы-акцентоносителя такая просодия недифференцированно выражает целый ряд иллокутивных значений: 'объясняю', 'советую', 'напоминаю': *Вам ВРАЧА надо вызвать; Тише. БАБУШКА спит; Одеденья потеплее: ВЕТЕР пронзительный.* При произнесении они сопровождаются выдвижением подбородка вперед по направлению к слушающему, широким открытием глаз и резким подъемом бровей вверх. Второй просодический тип отражает процесс погружения говорящего в его внутренний мир: говорящий мечтает о будущем, вспоминает о прошлом, предается размышлениям. Соответствующая просодия по классификации Е. А. Брызгуновой характеризуется как интонационная конструкция ИК-6. Артикуляция конструкции в данном случае сопровождается существенной растяжкой; время произнесения ударного слога словоформы-акцентоносителя увеличивается по сравнению со средними значениями в полтора-два раза: *Я даже помню, там ДОРО-ОЖКА была.* При артикуляции таких предложений говорящий откидывает голову назад и вбок, приподнимает брови. Артикуляционные и жестикуляционные особенности рассматриваемых иллокутивных типов предложений иллюстрируются фрагментами из фильмов, размещенных в Мультимедийном подкорпусе НКРЯ. Для анализа звучащих текстов используется компьютерная программа Praat. Аудио- и видеозаписи примеров доступны по ссылке: [http://iling-ran.ru/yanko/prosodii\\_zhesty/](http://iling-ran.ru/yanko/prosodii_zhesty/).

*Ключевые слова:* звучащая речь, интонационная конструкция, жестикуляция, Praat, речевые действия, иллокуции, НКРЯ.

Просодические и жестикуляционные параметры коммуникации традиционно исследуются независимо. В новейшей работе по русской жестикуляции [Гришина 2017] рассмотрено множество вопросов, связанных с выражением смыслов

---

\* Работа выполнена при финансовой поддержке РФФИ (грант № 16-06-00226).

лингвистическими и жестикуляционными средствами. В частности, Е. А. Гришина предлагает анализ иллокутивных значений с помощью жестов [Гришина 2017: 537—554]. Однако установление конкретных корреляций между просодическими конструкциями и движениями тела еще ждет своего часа. Цель нашей работы — продолжить исследования в направлении анализа жестов и рассмотреть устойчивые пары «просодия — жест». Ниже рассматриваются два типа иллокуций с богатой иллокутивной семантикой, которая выражается определенными типами просодий и регулярно сопровождается особой жестикуляцией. За анализируемыми иллокутивными типами стоят нетривиальные речевые действия, которые имеют в языке прежде всего не просодическое, а лексическое выражение. Это аргументация, совет, воспоминание, мечта. Поясним, что просодия, как правило, бывает связана с выражением более абстрактных языковых значений, таких, как тема и рема сообщения, указание на незавершенность повествования, выбор одного элемента из класса подобных. Абстрактным же понятиям, таким, как темы и ремы, в языке не соответствует никаких регулярных жестов. Тема и рема в высказывании есть, просодические средства выражения имеются, между тем характерные жесты отсутствуют. Речевые акты, о которых говорится ниже, без сопроводительных жестов практически не произносятся, то есть имеет место обязательное сопровождение просодической конструкции соответствующим жестом. Устоявшееся жестикуляционное сопровождение характерно только для нетривиальных иллокутивных смыслов, выражающихся просодически.

Материалом для анализа жестикуляции служат видеозаписи, размещенные в Мультимедийном подкорпусе Национального корпуса русского языка (НКРЯ). Аудиозаписи получены из того же корпуса и из корпуса «Рассказы о сновидениях» ([Spokencorpora 2017], см. детали в [Кибрик, Подлеская 2009]). Для анализа звучащих текстов используется компьютерная программа Praat [Boersma, Weenink 2018].

Аудио- и видеозаписи примеров, приводимых ниже, доступны по ссылке: [http://iling-ran.ru/yanko/prosodii\\_zhesty/](http://iling-ran.ru/yanko/prosodii_zhesty/).

В предложениях первого рассматриваемого типа просодия недифференцированно выражает целый ряд иллокутивных значений: ‘объясняю’, ‘советую’, ‘напоминаю’: *Вам **врача** надо вызвать* (совет)<sup>1</sup>; *Там **завтрак** на столе* (напоминание); *<Тише>. **Бабушка** спит* (объяснение того, почему не нужно шуметь); *<Оденься потеплее:> **ветер** пронзительный* (обоснование того, что требуется теплая одежда); *<В консерваторию он не попал:> **пел** плохо* (объяснение). В литературе такие предложения называются предложениями с неингерентной темой (по [Баранов, Кобозева 1983]), глобальными предложениями (по [Николаева 1982: 67]), *thetic* (по В. Матезиусу; см. [Sasse 1995]).

Обратимся к примеру (1) из Мультимедийного подкорпуса НКРЯ (фильм «Операция “С Новым годом”»):

<sup>1</sup> Здесь и ниже словоформа — носитель акцентного пика в предложении выделяется полужирным шрифтом.

(1) — *А чё ты в тулупе-то паришься?*

— **Наручники мешают.** [НКРЯ]

(См. аудиозаписи примера по ссылкам [http://iling-ran.ru/yanko/prosodii\\_zhesty/Naruchniki.wav](http://iling-ran.ru/yanko/prosodii_zhesty/Naruchniki.wav),

[http://iling-ran.ru/yanko/prosodii\\_zhesty/Naruchniki2.wav](http://iling-ran.ru/yanko/prosodii_zhesty/Naruchniki2.wav).)

Ответная реплика в диалоге (1) представляет собой предложение с неингерентной темой. Экстралингвистическая ситуация состоит в следующем. Патрульный полицейский пристегнул себя наручниками к задержанному, который при задержании проглотил ценный предмет — наградную медаль полицейского. Патрульный приводит задержанного в больницу в надежде, что врачи помогут ему вернуть медаль. Врач интересуется, почему патрульный сидит на приеме в зимней форме. Патрульный же в ответ удивляется, как доктор сам не догадался, что снять куртку ему мешают наручники. Перед нами речевой акт с иллюкутивной силой обоснования: ‘я не могу снять одежду, потому что я сцеплен наручниками с другим человеком’.

На ударном слоге акцентоносителя предложения — словоформе *наручники* — фиксируется падение частоты основного тона в существенном диапазоне частот, причем падению предшествует небольшой подъем частоты в начале ударного слога. Подъем начинается даже на исходе предударного слога. Этот подъем обеспечивает своего рода «набор высоты тона», необходимый для рельефного падения. Здесь предударный слог — это слог *на-*. Подъем продолжается и в начале ударного слога *-руч-*, а затем переходит в рельефное падение. Дальнейшее падение на заударных *-ники* завершает артикуляцию конструкции. Такая просодическая конфигурация соответствует интонационной конструкции ИК-2 по классификации Е. А. Брызгуновой [Брызгунова 1982: 107]. График изменения частоты основного тона предложения **Наручники мешают** см. по ссылке [http://iling-ran.ru/yanko/prosodii\\_zhesty/Naruchniki2.png](http://iling-ran.ru/yanko/prosodii_zhesty/Naruchniki2.png).

Возникает вопрос, как выбирается носитель акцента в предложениях с неингерентной темой. В предложении *Наручники мешают* примера (1) — это словоформа *наручники*, в предложении *Пел плохо* — это словоформа *пел*, а в предложении *Вам врача надо вызвать* — словоформа *врача*. При выборе акцентоносителя в предложениях с неингерентной темой действует несколько факторов. Основное влияние оказывают: 1) активация именных групп, которая исключает именные группы с активированным референтом из множества претендентов на роль акцентоносителя предложения; 2) синтаксическая иерархия, которая упорядочивает именные и наречные группы, служащие претендентами на роль акцентоносителя или содержащие акцентоносителя в своем составе; 3) иерархия зависимых членов в пределах синтаксической структуры более чем однословной группы. Иерархии устанавливаются приоритеты между группами и между словоформами внутри групп. Основная синтаксическая иерархия, ранжирующая группы, представляет собой следующее упорядочение:

*Данное (активированное) < Предикат (Р) < сирконстанты (С) < актанты* (в порядке, заданном актантной структурой предиката —  $A1 < A2 < A3 < A4 < A5 < A6$ ).

Иерархия говорит о том, что наименьшая вероятность стать в предложении с неингерентной темой акцентоносителем — у активированных компонентов предложения: активированные компоненты из множества возможных акцентоносителей исключаются. Далее по иерархии располагается предикат (Р), за которым следуют сирконстанты предложения (С). Завершают иерархию актанты (А) в соответствии с определенным упорядочением, присущим каждому конкретному предикату и зафиксированным в словаре. Упорядочение актантов говорит, например, о том, что дополнение имеет приоритет перед подлежащим, второе дополнение — перед первым, третий актант — перед вторым и т. д.

Примеры (2)—(4) ниже иллюстрируют некоторые приоритеты иерархии. Так, в примере (2) акцентоносителем ответной реплики служит подлежащее *отец*:

- (2) <... Что случилось?... Скажи толком!> — Твой **отец** (А1) приходил (Р) ночью (С). [НКРЯ]

В примере (3) акцентоноситель — это дополнение *сумку*:

- (3) <— Что случилось, граждане?> — **Сумку** (А2) порезали (Р) у иностранки (С). [НКРЯ]

В примере (4) в качестве акцентоносителя фигурирует не подлежащее, а глагол, потому что подлежащее *Боня* относится к известной информации: *Боня* — это имя собаки.

- (4) <— А что случилось с той... вашей собакой дальше?> — Ах, Боня **погиб**. [НКРЯ]

Кроме глобальной иерархии, внутри именных групп действуют внутренние (локальные) иерархии. Локальные иерархии устанавливают приоритет несогласованных определений над определяемым (*пачка сухариков*; *письмам отцу*; *отца дом*), определяемых — над согласованными определениями (*синяя птица*), отчества — над именем (*Марья Ивановна*), фамилии — над препозитивным именем (*Вася Иванов*). Локальные приоритеты говорят о том, что, если по глобальной иерархии приоритетной группой служит первое дополнение и оно содержит несогласованное определение<sup>2</sup>, то акцентоносителем предложения становится несогласованное определение<sup>2</sup>. Так, в примере (5) приоритетной группой служит группа дополнения *связь со временем*, а в качестве акцентоносителя в группе и в предложении в целом фигурирует зависимая словоформа *временем*:

- (5) <— Что с ним происходит?> — Утратил **связь со временем**. [НКРЯ]

Перейдем к жестикуляции, которая сопровождает реализацию иллюкуций, соответствующих просодии неингерентной темы. Произнесение предложений с неингерентной темой сопровождается выдвижением подбородка вперед по направлению к слушающему (жест «указание подбородком» по [Гришина 2017: 686]),

<sup>2</sup> Более подробно о выборе акцентоносителя см. [Янко 2001: 69—86].

широким открытием глаз (жест «открыть глаза»<sup>3</sup>) и быстрым подъемом бровей вверх (жест «поднять брови»). Таким образом, в соответствии с классификацией помет в НКРЯ совершаются жесты «выдвинуть подбородок», «поднять брови», «открыть глаза». Некоторые элементы артикуляции могут варьироваться: подбородок может выдвигаться вперед и вбок, в частности круговым движением; глаза необязательно раскрываются широко.

Обратимся к иллюстрации из фильма «Ночной продавец». Продавец магазина появляется в кадре с бейсбольной битой в руках. Бита предназначена для защиты от нападающего маньяка. Жена хозяина магазина, которая оставалась с маньяком, предлагает продавцу опустить биту: необходимость в защите отпала, потому что нападавший умер. Продавец не может в это поверить: он еще не видит лежащего нападавшего. Между продавцом и женщиной происходит такой диалог:

- (5) — *Опусти биту.*  
 — *Тихо. Тихо.*  
 — *Да опусти биту. Он умер!*  
 — *Как умер?*  
 — *Как все умирают.*

(См. видеозаписи примера по ссылкам: [http://iling-ran.ru/yanko/prosodii\\_zhesty/Nochnojprodavec.mp4](http://iling-ran.ru/yanko/prosodii_zhesty/Nochnojprodavec.mp4) и

[http://iling-ran.ru/yanko/prosodii\\_zhesty/Nochnojprodaveccenter.mp4](http://iling-ran.ru/yanko/prosodii_zhesty/Nochnojprodaveccenter.mp4).)

Предложение с неингерентной темой здесь — *Он умер*. Оно служит объяснением тому, что можно не бояться и опустить оружие. Перед нами иллюкутивный акт обоснования. Мы видим, что при произнесении предложения *Он умер* героиня — ее играет Ингеборга Дапкунайте — энергично выдвигает подбородок в сторону продавца (актер Павел Баршак), широко открывает глаза и поднимает брови. Просодия предложения сохраняет все параметры предложения с неингерентной темой.

Второй просодический тип отражает процесс погружения говорящего в его внутренний мир: говорящий мечтает о будущем, вспоминает о прошлом, предается размышлениям, сожалеет о несбыточном. По классификации интонационных конструкций Е. А. Брызгуновой перед нами конструкция типа ИК-6 с подъемом на ударном слоге словоформы-акцентоносителя и ровными или слабо нисходящими заударными [Брызгунова 1982: 107]. В анализируемых типах иллюкуций, отражающих погружение говорящего в мыслительный процесс, артикуляция конструкции сопровождается существенной растяжкой; время произнесения ударного слога увеличивается по сравнению со средним темпом в полтора-два раза.

Обратимся к примеру из корпуса «Рассказы о сновидениях»:

- (6) *Мы заехали на поля-янку, на березовую, погуля-яли там...*

(См. аудиозапись примера по ссылке [http://iling-ran.ru/yanko/prosodii\\_zhesty/Poljanka.wav](http://iling-ran.ru/yanko/prosodii_zhesty/Poljanka.wav).)

<sup>3</sup> Номенклатура жестов дается, где это возможно, по НКРЯ (<http://www.ruscorporu.ru/search-murco.html>). Номенклатура жестов разработана Е. А. Гришиной.



Предложение открывается темой *мы*, на которой наблюдается рельефный подъем. Вслед за подъемом на заударном слоге, который представлен первым слогом следующей словоформы *заехали*, частота резко падает. Это дефолтная просодия начальной темы ИК-3 по Е. А. Брызгуновой [Брызгунова 1982: 107]. ИК-3 здесь представляет собой контраст для другого подъема — подъема типа ИК-6. ИК-6 фиксируется на реме *заехали на полянку на березовую* с акцентоносителем *полянку* и второй реме *погуляли там* с акцентоносителем *погуляли*. На ударных слогах словоформ *полянку* и *погуляли* фиксируется более пологий подъем, чем на *мы*, и за ним частота не падает столь резко, как после словоформы *мы*, а снижается постепенно, держась некоторое время на относительно ровном уровне. Другой параметр — существенное замедление артикуляции. Речь моделирует мыслительный процесс, на который требуется известное время: говорящий «шевелит мозгами».

Выбор акцентоносителя в предложениях размышления и мечты подчинен тем же принципам, что и выбор акцентоносителя в предложениях с неингерентной темой.

Перейдем к анализу жестикуляции. При артикуляции предложений такого типа говорящий слегка откидывает голову назад и вбок (жест «откинуть голову» по классификации НКРЯ), прикрывает и закатывает глаза (жесты «прикрыть глаза» и «закатить глаза»), приподнимает брови (жест «поднять брови»). Движения относительно медленные.

Обратимся к сцене неудачного свидания героев Людмилы Гурченко и Никиты Михалкова из кинофильма «Вокзал для двоих»:

- (7) — Ну чего-то я не могу.  
 — Давай, давай, давай.  
 — Не знаю.  
 — Давай, давай, Верунчик.  
 — Ну как-то не по-человечески, наспех всё. Не могу.  
 — Наспех, наспех, наспех. Что делать, Верунчик? Жизнь такая, всё наспех.  
 Давай, давай, давай. Опа.  
 — Знаешь, я мечтаю, чтобы ты приехал на **неде-ельку**, мы б с тобой **погуля-яли**...  
 — Погуляем, погуляем, погуляем... ...по парку... Погуляем...  
 — ...сходили бы в **кино-о**...  
 (См. видеозапись примера по ссылке [http://iling-ran.ru/yanko/prosodii\\_zhesty/Mechtygurchenko.mp4](http://iling-ran.ru/yanko/prosodii_zhesty/Mechtygurchenko.mp4) и аудиозапись по ссылке [http://iling-ran.ru/yanko/prosodii\\_zhesty/Mechtygurchenko.wav](http://iling-ran.ru/yanko/prosodii_zhesty/Mechtygurchenko.wav).)

Мы видим, что слова о том, о чем мечтает героиня Гурченко (поиск в корпусе в данном случае проводился по ключевому слову *мечтать*), сопровождаются особой жестикуляцией. Вера в исполнении Гурченко оборачивает лицо к слушающему, приподнимает брови, морщит лоб, чуть прикрывает глаза, откидывает голову назад и вбок, говорит с существенной растяжкой: *я мечтаю, чтобы ты приехал на неде-ельку, мы б с тобой погуля-яли... сходили бы в кино-о....*

Итак, интонация погружения во внутренний мир, которая выражает иллюзии ‘мечтаю’, ‘вспоминаю’, ‘размышляю вслух’, сопровождается следующими значениями жестикуляционных параметров: откинуть голову, прикрыть глаза, закатить глаза, поднять брови, наморщить лоб, замедлить движения и речевую артикуляцию.

Возникает вопрос о методе анализа. Поиск в корпусе видеоматериалов проводился в тексте транскриптов по ключевым словам *что случилось?* и *что происходит?* для анализа жестикуляции, соответствующей неингерентной теме, и *помню, вспоминаю, мечтаю* — для анализа просодий, соответствующих погружению говорящего в размышления, воспоминания и мечты. Кроме того, при поиске использовались пометы, соответствующие параметру «значение жеста». При поиске в мультимедийном подкорпусе НКРЯ было найдено только одно значение, соответствующее нашему информационному запросу. Это одно из значений параметра «Значение жеста»: помета «аргумент»<sup>4</sup>. Теоретически этот термин соответствует прагматике неингерентной темы, которая действительно регулярно используется для аргументации. Однако ни один пример, который описывал бы жест аргументации (по классификации НКРЯ это жест «двинуть ладонью») не содержал одновременно жеста, который, по нашим данным, соответствует неингерентной теме. Просодия неингерентной темы, как мы показали, сопровождается совсем другим жестом. Это жест «выдвинуть подбородок». И жесты «выдвинуть подбородок» и «двинуть ладонью» в одном и том же дискурсе нам также не встретились. Стоит заметить, что жест «двинуть ладонью» вполне мог бы сопровождать артикуляцию предложений с неингерентной темой. Об этом говорят наблюдения и интроспекция. Мы предполагаем, что для анализируемой просодии релевантны оба жеста, но жест «выдвинуть подбородок» оказывается более устойчивым.

Подходящих значений жестов, которые сопровождали бы размышления, воспоминания и мечты, в номенклатуре жестов корпуса пока тоже не обнаруживается. Можно заключить, что разметка жестов, которая имеется в НКРЯ в настоящее время, в данной работе фактически не применялась.

Таким образом, следует признать, что имеющиеся здесь результаты получены, скорее, на основании интроспекции и наблюдения, чем на основании корпусного анализа. Рассмотренные здесь примеры служат не более чем подтверждением и иллюстрацией интроспекции и наблюдений. Естественно, в тексте этой статьи были рассмотрены далеко не все релевантные примеры. Между тем материала, полученного из корпуса, безусловно, недостаточно для проведения полноценного статистического анализа рассмотренных смыслов и соответствующих этим смыслам средств выражения: просодии во взаимодействии с жестами. Полученный корпусный материал подтверждает лишь исходную гипотезу о том, что корреляция

<sup>4</sup> Поясним, что для указания на значения жестов в НКРЯ используются такие пометы: аргумент, возражение, вопрос, беспокойство, благодарность и др.; классификацию значений и их полный список см. по ссылке: <http://www.ruscorpora.ru/attrs-murco-gestures-mixed.html>.

между просодическими и жестикуляционными средствами выражения нетривиальных иллокутивных значений может иметь место. Это основной результат проведенного исследования.

Будущее представленного здесь анализа видится в индексировании большого количества видеоматериала ключевыми словами и (или) пометами, которые описывают жесты и иллокутивные силы.

## Литература

- Брызгунова Е. А.* Интонация // Русская грамматика. Т. 1. М. : Наука, 1982.
- Баранов А. Н., Кобозева И. М.* Семантика общих вопросов в русском языке (категория установки) // Изв. АН СССР. Сер. лит. и яз. 1983. Т. 42. № 7. С. 263—275.
- Гришина Е. А.* Русская жестикуляция с лингвистической точки зрения (корпусные исследования). М. : ЯСК, 2017. 744 с.
- Кибрик А. А., Подлесская В. И.* Рассказы о сновидениях: Корпусное исследование устного русского дискурса / ред. А. А. Кибрик, В. И. Подлесская. М., 2009.
- Николаева Т. М.* Семантика акцентного выделения. М. : Наука, 1982.
- Янко Т. Е.* Коммуникативные стратегии русской речи. М. : ЯСК, 2001.
- Boersma P., Weenink D.* (2018). Praat: Doing phonetics by computer. Version 6.0.37. Online: <http://www.praat.org/>.
- Sasse H. J.* 'Theticity' and VS order: a case study. *Sprachtypologie und Universalienforschung* 48. 1/2. 1995.
- Spokencorpora 2017. Prosodically Annotated Corpus of Spoken Russian. Pilot version. Online: <http://spokencorpora.ru>.

**T. E. Yanko**

*Institute of Linguistics, Russian academy of sciences  
(Moscow, Russia)  
tanya\_yanko@list.ru*

## PROSODY AND GESTURES<sup>5</sup>

This paper is aimed at analyzing two types of prosodic constructions and the corresponding questures. The first prosodic type is characterized by a substantial fall of frequency. In the terminology of E. A. Bryzgunova, it is the Russian construction IK-2. In sentences with a certain accent placement, such prosody expresses a variety of illocutionary meanings: 'I explain', 'I recommend', 'I remind you': *Vam VRACHA nuzhno vyzvatj* 'I recommend that you should send for a doctor'; *Tishe, BABUSHKA spit* 'Keep quiet, grandmother is sleeping'; *Odenjsja poteplee: VETER pronzitel'nyj* 'Put something

<sup>5</sup> This study has been granted the financial support of the RFBR (project N 16-06-00226).

on: the wind is biting’. In linguistic literature, such sentences are called *thetic*. When articulated, they are accompanied by a rapid movement of a chin of a speaker towards the hearer, opening the eyes, and raising the eyebrows. The second prosodic type mirrors the process of thinking, dreaming, or recollecting the past events. In the terminology of E. A. Bryzgunova, it is the *IK-6* construction. Articulation of the construction is substantially lengthened: *Ja dazhe pomnju, tam DORO-OZHKA byla* ‘I am recollecting that there was a path’. When pronouncing, the speaker tosses his head back and aside. The prosodic parameters and gestures of the analyzed illocutionary types are exemplified by video fragments retrieved in the Russian National corpus. In the process of analyzing the spoken data, the computer system PRAAT is used. The audio and video recordings are available at: [http://iling-ran.ru/yanko/prosodii\\_zhesty/](http://iling-ran.ru/yanko/prosodii_zhesty/).

*Key words:* spoken speech, spoken corpora, prosody, gesture, illocution, PRAAT, speech acts

## References

- Baranov A.N., Kobozeva I.M. [Semantics of the *yes-no*-questions in Russian: (the category of an attitude)]. *Izv. AN SSSR. Ser. lit. i yaz.*, 1983, vol. 42, no. 7, pp. 263—275. (In Russ.)
- Boersma P., Weenink D. (2018). Praat: Doing phonetics by computer. Version 6.0.37. Available at: <http://www.praat.org/>.
- Bryzgunova E. A. [Intonation]. *Russkaya grammatika* [Russian Grammar], vol. 1. Moscow, Nauka Publ., 1982. (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 slavyanskoi kul'tury Publ., 2017. 744 p. (In Russ.)
- Kibrik A., Podlesskaya V. *Rasskazy o snovideniyah: Korpusnoe issledovanie ustnogo russkogo diskursa* [Nightdream stories: A corpus-based study of spoken Russian discourse]. Moscow, Yazyki slavyanskoi kul'tury Publ., 2009. (In Russ.)
- Nikolaeva T.M. *Semantika akcentnogo vydeleniya* [The semantics of accentual emphasis]. Moscow, Nauka Publ., 1982. (In Russ.)
- Sasse H.J. ‘Theticity’ and VS order: a case study. *Sprachtypologie und Universalienforschung* 48. 1/2. 1995.
- Spokencorpora 2017. Prosodically Annotated Corpus of Spoken Russian. Pilot version. Available at: <http://spokencorpora.ru>.
- Yanko T. E. *Kommunikativnye strategii russkoy rechi* [The communicative strategies of the Russian speech]. Moscow, Yazyki slavyanskoi kul'tury Publ., 2001. (In Russ.)

**О. В. Федорова**

*МГУ им. М. В. Ломоносова, Институт языкознания РАН*

*(Россия, Москва)*

*olga.fedorova@msu.ru*

## **О КОММУНИКАТИВНОЙ ФУНКЦИИ ВЗГЛЯДА\***

В статье с разных точек зрения рассматривается вопрос о коммуникативной функции взгляда. Сначала вводятся базовые сведения из области физиологии глазодвигательной активности, включая противопоставление фиксаций и саккад, основные используемые метрики, а также вопрос об уникальности человеческого глаза с точки зрения выполнения им коммуникативной функции. Затем описывается история айтрекинга — метода регистрации движений глаз, включающая четыре основных этапа. Отдельный раздел посвящен изучению движений глаз в повседневном общении, в том числе во время выполнения домашней работы, прогулки и общения в нелабораторных условиях. Далее описываются основные функции движений глаз в естественной коммуникации — функция мониторинга, регуляторная функция, экспрессивная функция и глазной контакт. Наконец, предлагается качественный и количественный анализ небольшого фрагмента длительностью около 8,5 мин естественной коммуникации трех собеседников, на двух из которых надеты очки-айтрекеры. Показано, что использование разных метрик — количества и продолжительности фиксаций или взглядов — дает принципиально разные результаты; в настоящем исследовании основной количественной характеристикой общения коммуникантов считается количество их взглядов. При сопоставлении взглядов трех коммуникантов оказывается, что они хорошо синхронизованы друг с другом и практически без исключений описываются несколькими правилами.

---

\* Исследование выполнено в Институте языкознания РАН при финансовой поддержке РФФИ, грант № 14-18-03819 «Язык как он есть: русский мультимодальный дискурс». Автор выражает благодарность всем участникам проекта, в первую очередь Н. А. Коротаеву, сделавшему ряд важных комментариев.

Отправной точкой для написания данной статьи послужил доклад автора на конференции «Слово и жест» (Гришинские чтения), которая состоялась 8 февраля 2018 г. В этом докладе, опираясь на главу «Грамматика взгляда» из недавно вышедшей монографии Е. А. Гришиной «Русская жестикуляция с лингвистической точки зрения» (2017), на материале небольшого отрывка

*Ключевые слова:* мультиканальная коммуникация, движения глаз, айтрекинг, фиксация, взгляд, зрительное внимание, коммуникативная функция.

Исключительная роль зрительного аппарата человека в коммуникации давно хорошо известна и не нуждается в дополнительной аргументации. Однако только в последние годы исследователи в полной мере осознали двойственность функционирования глаз в ходе естественного взаимодействия между людьми. Как образно отмечает Маттиас Гобель с соавторами, «уши не могут говорить, губы не могут слышать, но глаза могут как подавать сигнал, так и воспринимать его. Для человека эта двойственная функция делает глаза замечательным инструментом социального взаимодействия. Однако для психологов, пытающихся понять движения глаз, их двойственная функция вызывает фундаментальную неоднозначность» ([Gobel et al. 2015: 359], перевод мой. — О. Ф.; ср. аналогичную идею о “simultaneous input-output device” в работе [Jording et al. 2018]). В настоящей работе мы рассмотрим вопрос о функциях человеческого глаза в коммуникации, о методах их изучения, а также приведем пример одного конкретного исследования. В разделе 1 вводятся основные сведения из области физиологии зрения. Затем в разделе 2 кратко описана история айтрекинга — метода регистрации движений глаз. Небольшой раздел 3 посвящен изучению движений глаз в повседневном общении. В разделе 4 перечислены функции движений глаз в естественной коммуникации. Наконец, в разделе 5 мы предложим собственное исследование движений глаз в ходе естественного мультиканального общения.

## 1. Физиология зрения человека

Посмотрим, как человек рассматривает окружающие его объекты. Свет проходит через роговицу и хрусталик и фокусируется на сетчатке глаза, которая содержит светочувствительные рецепторы. Структура сетчатки неоднородна — в самом центре (fovea), где оптическая ось глаза пересекает поверхность сетчатки, рецепторы расположены наиболее плотно. Поэтому в процессе зрительного восприятия глаз ориентируется таким образом, чтобы изображение объекта локализовалось в фовеальной области, угловые размеры которой составляют около 2°. Для этой зоны характерен максимально высокий уровень обнаружения, опознания и идентификации объектов. Кроме центральной зоны, обычно выделяют ближнюю периферию ( $\pm 2^\circ$ – $15^\circ$ ), для которой характерно сравнительно высокое обнаружение, опознание и идентификация; среднюю периферию ( $\pm 15^\circ$ – $25^\circ$ ) — ограниченную способность опознания и идентификации кратковременных событий и трудности категоризации; дальнюю периферию ( $\pm 25^\circ$ – $35^\circ$ ) — хорошее обнаружение, но плохую идентификацию, опознание и классификацию объектов; наконец, экстремальную периферию (свыше  $\pm 35^\circ$ ) — только обнаружение (цит. по: [Барabanщиков,

---

общения трех коммуникантов мы провели пилотное исследование направления взгляда собеседников на границах реплик. Настоящая статья продолжает и углубляет эту работу.

Жегалло 2013]). Полное поле зрения здорового глаза человека — 150° по горизонтали и 130° по вертикали.

Глаз никогда долго не задерживается на одном месте благодаря чередованию быстрых перемещений (так называемых **саккад**, средняя длительность 30–60 мс) и относительно коротких остановок (**фиксаций** длительностью в среднем 200–500 мс, но бывает и в несколько раз длиннее). Даже при относительно неподвижном положении глаза (то есть когда не регистрируется саккад) происходят микродвижения, к числу которых относятся: тремор (мелкие частые колебания глаз, не поддающиеся произвольному контролю), дрейф (медленные микродвижения в ходе фиксации объекта, создающие наиболее благоприятные условия для обработки зрительной информации), микросаккады (быстрые движения глаз продолжительностью 10–20 мс, обеспечивающие устойчивые фиксации объекта), прослеживающие движения (плавные перемещения глаз, возникающие при движении объектов в поле зрения и обеспечивающие сохранение изображения объекта в зоне наиболее четкого видения), а также нистагм (чередование саккад и плавных прослеживающих движений).

В исследованиях глазодвигательной активности обычно используются следующие метрики: количество и продолжительность фиксаций, количество и продолжительность саккад, количество и продолжительность взглядов, а также кривые движения глаз (*scanpaths*). Рассмотрим две метрики немного подробнее. Одна из важных задач при изучении окуломоторной активности в процессе естественной коммуникации, еще не решенная на сегодняшний день, состоит в поиске оптимального минимально значимого уровня продолжительности фиксаций. Использование продолжительности фиксаций в качестве показателя, характеризующего когнитивную активность, базируется на гипотезе, что когнитивная обработка зрительной информации выполняется во время фиксаций, а во время саккад подавляется. В современных исследованиях минимальный значимый уровень продолжительности фиксаций колеблется от 30 мс в работах по изучению эмоциональных характеристик до 150 мс при изучении совместного решения когнитивных задач. Решая эту задачу, сначала мы вслед за авторами работ [Jacob, Karn 2003; Vertegaal et al. 2001; Brône, Oben 2015] выбрали оптимальным уровень в 120 мс. Однако, проанализировав аполученные таким образом данные, мы пришли к выводу, что при таком уровне теряются некоторые важные фиксации. Поэтому было принято решение считать оптимальным уровень в 100 мс, при котором подобные потери минимальны.

Второй важный вопрос касается взглядов<sup>2</sup> (*gaze*) — последовательности фиксаций и саккад, находящихся в одной области интереса. Исследования последних

<sup>2</sup> В работах В. А. Барабанщикова и его группы (например, см. [Барабанщиков, Жегалло 2013]) в качестве русскоязычного аналога английского термина *gaze* используется русская лексема *взор*. Однако, согласно данным из [Урысон 2003], лексема *взор* стилистически отмечена как устаревшее или поэтическое слово; кроме того, она невозможна в значении моментального действия, ср. *\*бросить взор*. По этой причине в настоящей работе мы будем использовать русский термин *взгляд*. Согласно [Урысон 2003], существительное *взгляд* может быть образовано как от глагола совершенного вида, так и от глагола несовершенного вида; в последнем случае лексема *взгляд*



двух десятилетий (в частности, см. работу [Jakob, Karn 2003]) показывают, что часто анализ количества и продолжительности взглядов оказывается более значимым, чем анализ количества и продолжительности отдельных фиксаций (подробнее об этом см. раздел 5). Считается, что информация о направлении взгляда собеседников дает нам информацию о распределении их **зрительного внимания** [Smith, Schenk 2012]. Кроме того, микросаккады являются мерой **скрытого внимания** [Engbert 2006] и обычно предшествуют саккадам [Deubel, Schneider 1996].

Наконец, рассмотрим вопрос об особенностях строения человеческого глаза по сравнению с глазами приматов. В 1997 г. Хироми Кобаяси и Сиро Кохсима опубликовали в журнале *Nature* исследование, посвященное одной интересной особенности человеческого глаза ([Kobayashi, Kohshima 1997], более подробная версия была представлена в их работе [Kobayashi, Kohshima 2001]). Они сравнили человеческие глаза с глазами почти половины известных науке приматов и пришли к выводу, что белая склера (более привычное название «белки глаз») — уникальная особенность человека. Авторы предположили, что эта физиологическая уникальность облегчает человеку определение направления взгляда другого человека на расстоянии и тем самым несет коммуникативную функцию ([Kobayashi, Kohshima 1997, 2001]. Это открытие было подтверждено в 2007 г. экспериментально в работе известного психолингвиста и антрополога Майкла Томаселло с коллегами, в которой авторы сравнили поведение человеческих младенцев 12 и 18 месяцев с поведением приматов [Tomasello et al. 2007]. Оказалось, что младенцы в первую очередь обращали внимание на направление взгляда экспериментатора, в то время как приматы следили за движениями его головы. По результатам этой работы Томаселло с коллегами сформулировали **гипотезу кооперативного глаза** [Tomasello et al. 2007]. По их мнению, тот факт, что человеческий глаз сообщает информацию о направлении взгляда, говорит об эволюционной адаптации с целью повышения согласованности совместных действий.

Однако в 2015 г. была опубликована новая статья на эту тему [Mayhew, Gómez 2015], в которой уникальность белой склеры человека была поставлена под сомнение. Хуан-Карлос Гомес заметил, что у гориллы Нади из Мадридского зоопарка вокруг зрачков есть белая оболочка. Совместно с Джессикой Мэйхью он рассмотрел большую выборку и пришел к выводу, что из 60 западных равнинных горилл лишь у 30 % были совершенно темные склеры; у остальных 70 % белый цвет в том или ином виде присутствовал, а у небольшой выборки в 7 % были совершенно белые, как у человека, склеры. Таким образом, утверждают авторы, белизна наших глаз не могла стать наиболее важным эволюционным фактором. По их мнению, основную роль сыграла удлинённая по горизонтали по сравнению с приматами форма глаза человека. Именно поэтому человеческий глаз лучше приспособлен к тому, чтобы определять по нему направление взгляда.

---

выражает один квант соответствующего действия. Как отмечает Е. В. Урысон, хотя в современном языке случаи такого употребления относительно редки (например, *любовь с первого взгляда*), в языке XVIII в. оно употреблялось в таком значении гораздо свободнее.

## **2. Методология исследования движений глаз**

История научного изучения движений глаз началась в XIX в. Согласно классификации известного специалиста в области изучения движений глаз при чтении Кита Райнера [Rayner 1998], первый период таких исследований берет начало с работ Луи Жавала, который в 1879 г. при помощи зеркала заметил, что движение глаз при чтении происходит не плавно, а наоборот, человек читает благодаря чередованию саккад и фиксаций (именно Жавал предложил такие названия). На протяжении первого периода было сделано много фундаментальных открытий относительно биомеханики движений глаз. В частности, стало известно, что во время саккад человеческий глаз практически не способен воспринимать окружающий мир, а длительность саккад в среднем равна 30–60 мс.

Второй период изучения движений глаз, согласно хронологии Райнера, начинается в 1920-х гг. В эти годы Гаем Бузвеллом был создан первый бесконтактный аппарат для регистрации движений глаз, при помощи которого он изучал движения глаз при чтении [Buswell 1935] и рассматривании картинок [Buswell 1937]. В 1960-х гг. широкую известность получили работы отечественного биофизика Альфреда Лукьяновича Ярбуса [Ярбус 1965, английский перевод Yarus 1967], который описал феномен избирательного рассматривания: при рассматривании картины большинство движений глаз направлены на наиболее информативные участки, то есть запись движений глаз можно назвать грубым зеркальным отражением этой картины. Более того, результаты рассматривания одной и той же картины меняются в зависимости от задания, которое получает испытуемый. Так, кривые движения глаз (scanpaths) при рассматривании картины Репина «Не ждали» менялись в зависимости от установок следующим образом: когда было необходимо оценить материальное положение семьи, особое внимание испытуемых привлекало убранство комнаты, которое, однако, практически не замечалось при определении возраста персонажей — в этом случае испытуемые направляли взгляд почти исключительно на лица людей [Ярбус 1965]. Аппаратура, которой пользовался Ярбус, также не знала аналогов — резиновая присоска с радиоантенной укреплялась непосредственно на склере глаза испытуемого, голова которого при этом фиксировалась в металлической рамке; сам испытуемый должен был во время эксперимента держать во рту специальную пластину, вылитую по форме его зубов. Несмотря на то что точность измерений, которой удалось достичь Ярбусу, была весьма хорошей, данная технология не получила в дальнейшем широкого распространения.

Третий период изучения движений глаз начался в середине 1970-х гг. и в большой степени связан с работами самого Райнера. Большая часть фактов, накопленных к сегодняшнему дню за всю историю изучения процессов чтения, объединена в популярной модели чтения E-Z Reader [Reichle et al. 2003].

Наконец, начало текущего и четвертого по счету периода в истории изучения движений глаз датируется серединой 1990-х гг., когда к ставшему уже традиционным изучению движений глаз при чтении добавилась возможность регистрации движений глаз испытуемых, движения головы которых практически

не ограничены. В настоящее время существует несколько разновидностей айтрекеров со свободным положением головы: полностью бесконтактная модель, когда камера монтируется в непосредственном окружении; модель в виде легкого шлема, который надевается на голову испытуемому; модель в виде очков, в которые вмонтированы миниатюрные видеокамеры — одна из них записывает то, на что смотрит испытуемый, а вторая при помощи отраженного света фиксирует изображение глаза.

### 3. Движения глаз в повседневном общении

За годы, прошедшие с момента появления первых айтрекеров, был накоплен большой багаж знаний об особенностях глазодвигательного поведения человека в различного рода экспериментальных заданиях — при чтении, рассматривании картин или сайтов в интернете и даже при игре в гольф. Несколько парадоксальным образом оказывается, что о распределении зрительного внимания в повседневном общении мы знаем гораздо меньше. Данный раздел основан на статье [Foulsham 2014], в которой приводится обзор подобных исследований.

В серии работ Майкла Ланда с коллегами [Land et al. 1999; Land, Hayhoe 2001; Land 2007] авторы, исследовав движения глаз людей во время приготовления чая и сэндвичей, сформулировали несколько общих принципов. Во-первых, люди обычно заново собирают необходимую информацию, а не достают ее из долговременной памяти ('just-in-time' selection). Во-вторых, человек направляет взгляд на целевой объект за 0,5–1 с до того, как к этому объекту начинает двигаться его рука. В-третьих, фиксации 'look ahead', характерные для лабораторных исследований, во время которых испытуемый собирает глазами информацию, чтобы использовать ее в будущем, для повседневного общения оказались не типичны.

В работах [Hollands, Marple-Horvat 2001; Patla, Vickers 2003; Foulsham et al. 2011] авторы проанализировали движения глаз людей во время прогулки. Оказалось, что человек по-разному смотрит на окружающие его объекты во время реальной прогулки и во время просмотра видеозаписи той же прогулки на экране компьютера; в частности, во время реальной прогулки люди заметно чаще смотрят на дорогу перед собой. Более того, человек смотрит на дорогу перед собой не случайным образом, а будто опережая свои движения примерно на два шага или 1 с. Кроме того, во время реальной прогулки человек заметно чаще поворачивает голову, координируя ее движения с движениями глаз, чем во время просмотра видеозаписи на экране монитора, когда он обычно использует более длинные глазные саккады, не поворачивая головы.

Наконец, в обзоре [Foulsham 2014] приводятся несколько примеров социального взаимодействия между людьми в нелабораторных условиях, которые оказываются отличными от общеизвестных данных, полученных в лабораторных условиях. Например, если в лабораторных условиях испытуемый автоматически следит за взглядом другого испытуемого, то в естественных условиях это оказывается зависимым от разных социокультурных обстоятельств. В частности, оказывается,

что человек чаще следит за направлением взгляда впереди идущего человека (ориентируясь по движениям его головы), чем того, кто идет рядом с ним или навстречу ему [Gallup et al. 2012].

#### 4. Функции направления взгляда в естественной коммуникации

В данном разделе мы рассмотрим три наиболее важные, на наш взгляд, публикации, посвященные исследованию функций направления взгляда в естественной коммуникации. Опишем эти работы в хронологическом порядке. Основы изучения распределения зрительного внимания собеседников в ходе естественного общения были заложены 50 лет назад классиком невербальной коммуникации Адамом Кендоном в работе под названием “Some functions of gaze direction in social interaction” [Kendon 1967]. Эти результаты были получены методом кинорегистрации с частотой 2 к/с. Проанализировав пятиминутное общение семи пар испытуемых, Кендон заключил, что:

- испытуемый чаще смотрит на собеседника, когда слушает его, чем когда сам говорит;
- фиксации на собеседнике длиннее, когда испытуемый молчит, чем когда говорит;
- когда испытуемый молчит, его фиксации на собеседнике длиннее, чем фиксации на окружении;
- когда испытуемый говорит, его фиксации на собеседнике короче, чем фиксации на окружении;
- наблюдаются сильные индивидуальные различия. Так, испытуемые-слушатели фиксировали взгляд на собеседнике от 32 до 81 % всего времени, а испытуемые-говорящие смотрели на собеседника от 20 до 68 % всего времени [Kendon 1967].

Кендон выделил следующие четыре функции направления взгляда:

- 1) функция мониторинга — глядя на собеседника, мы определяем, куда он смотрит, какова его мимика, в какой позе он находится;
- 2) регуляторная функция — в частности, говорящий регулирует смену реплик в диалоге, когда в конце своей реплики при помощи взгляда передает ход слушающему;
- 3) экспрессивная функция — например, взгляд в сторону во время речи собеседника может выражать неудовлетворенность его словами;
- 4) взаимный взгляд, глазной контакт — количество взаимных взглядов зависит от уровня эмоциональности.

Вторая классическая работа о функциях взгляда “The Different Functions of Gaze” была написана в 1973 г. В ней авторы описали основные функции движений взгляда таким образом [Argyle et al. 1973]:

- 1) функция поиска информации — говорящий смотрит на собеседника с целью немедленно получить реакцию на свои слова, слушающий подкрепляет слуховую информацию зрительной, следя за мимикой или направлением взгляда

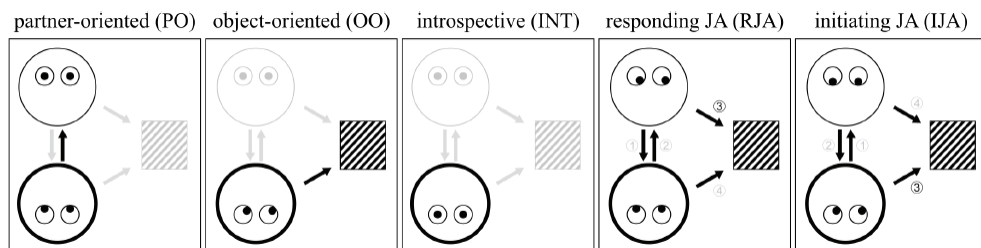


Рис. 1. Описание пространства социального взгляда, по [Jording et al. 2018]

собеседника; видимо, эта функция похожа на функцию мониторинга по Кендону [Kendon 1967];

2) сигнализирующая функция — говорящий или слушающий посылает различные межличностные сигналы, говорящий иллюстрирует свою речь; по-видимому, эта функция похожа на экспрессивную функцию по Кендону [Kendon 1967];

3) функция контроля за синхронизацией речи — аналогична регуляторной функции по Кендону [Kendon 1967];

4) взаимный взгляд, глазной контакт — аналогично Кендону [Kendon 1967].

Наконец, в только что вышедшей работе “A taxonomy for gaze-based communication in triadic interactions” авторы не просто описали основные функции направления взгляда коммуникантов, а предложили концептуальную схему описания пространства социального взгляда (‘Social Gaze Space’, [Jording et al. 2018]). Согласно этому подходу (рис. 1), выделяется пять возможных состояний взгляда одного коммуниканта (хотя в названии данной работы заявлено взаимодействие в триаде, реально в ней описываются два коммуниканта и один объект), схема затем расширяется до комбинации всех возможных состояний двух коммуникантов:

1) один коммуникант (выделен на схеме полужирным) смотрит на другого;

2) один коммуникант смотрит на объект и не смотрит на второго коммуниканта;

3) коммуникант погружен в себя и не смотрит ни на второго коммуниканта, ни на объект;

4) первый коммуникант смотрит на объект, отвечая на глазной призыв второго коммуниканта;

5) первый коммуникант инициирует взгляд второго коммуниканта на объект.

Ниже мы предложим свое представление о функциях направления взгляда коммуникантов в естественной коммуникации.

## 5. «Рассказы и разговоры о грушах»: грамматика взгляда

В настоящей работе мы, используя данные айтрекинга, представим качественный анализ фрагмента коммуникации трех собеседников. Насколько нам известно, существует только одно аналогичное исследование — недавняя работа Питера Ауэра “Gaze, addressee selection and turn-taking in three-party interaction” (2017), в которой автор показал, что при взаимодействии трех коммуникантов регуляторная

функция движения взгляда говорящего организована нетривиальным образом — одним и тем же взглядом на собеседника он может как выбрать его в качестве основного адресата, так и передать ему очередь хода [Auer 2017]. В нашем исследовании, однако, данная проблема не была столь актуальна, так как роли участников коммуникации были закреплены заранее (см. ниже).

Данное исследование было проведено на материале создаваемого нами ресурса «Рассказы и разговоры о грушах», проект по сбору которого осуществляется в Институте языкознания РАН (сайт [multidiscourse.ru](http://multidiscourse.ru)). Ресурс включает 40 записей суммарной длительностью около 15 ч. При сборе материала было использовано современное оборудование, в том числе промышленные видеокамеры с частотой 100 к/с и две пары очков-айтрекеров Tobii Glasses II с частотой 50 Гц.

Для проведения исследования была разработана новая оригинальная методика. В каждой записи принимали участие четыре человека с заранее распределенными ролями. Три участника — Рассказчик (Narrator, N), Комментатор (Commentator, C) и Пересказчик (Reteller, R) — участвовали в основной части записи, последний — Слушатель (Listener, L) — присоединялся в конце. Сначала N и C смотрели «Фильм о грушах» У. Чейфа [Chafe ed. 1980] и старались его запомнить. Затем к ним присоединялся R. Задача N состояла в том, чтобы рассказать сюжет фильма R; это был этап *рассказа* в режиме монолога. На следующем интерактивном этапе *разговора* C дополнял рассказ N, а R уточнял детали. Наконец, на этапе *пересказа* R в режиме монолога пересказывал сюжет фильма L. После этого L записывал услышанный пересказ. Таким образом, задача каждого участника состояла в том, чтобы максимально понятно донести до других полученную информацию, минимизировав эффект «испорченного телефона».

Аннотирование ресурса осуществляется по всем основным каналам, включая вербальный, просодический [Кибрик, Подлесская ред. 2009], кинетический и окулomotorный. Вокальная аннотация в программе PRAAT ([fon.hum.uva.nl/praat](http://fon.hum.uva.nl/praat)) состоит в членении речевого потока на значимые фрагменты (элементарные дискурсивные единицы — ЭДЕ, слова, (не)заполненные паузы), а также в приписывании свойств ЭДЕ и отдельным их частям. Для аннотации мануальных движений (движений рук) в программе ELAN ([tla.mpi.nl/tools/tla-tools/elan](http://tla.mpi.nl/tools/tla-tools/elan)) была разработана новая методика [Литвиненко и др. 2017]. В ходе окулomotorного аннотирования был произведен экспорт данных на видеосцену и с помощью программы Tobii Analyzer извлечены данные о временной развертке фиксаций длительностью выше 100 мс, на которые потом в ручном режиме была наложена аннотационная схема [Федорова 2017].

К настоящему моменту полностью аннотированы и выложены на сайт три записи (№04, 22 и 23) суммарной длительностью около 1 ч, которые составляют эталонный подкорпус. **Эталонный подкорпус** — это своего рода экспериментальная площадка, на которой мы тестируем различного рода гипотезы, чтобы потом верифицировать их на более обширном материале. Прежде чем переходить непосредственно к теме данной работы, надо убедиться, что три записи подкорпуса сопоставимы между собой. Как можно видеть из табл. 1, несмотря на то, что мы

не устанавливали для испытуемых временных ограничений, все три записи похожи друг на друга по длительности отдельных этапов — на *рассказ* приходится около 20 % от времени всей записи, на *разговор* — примерно 50 %, а на *пересказ* — 30 %.

Таблица 1

**Дескриптивная статистика эталонного подкорпуса по этапам**

№	Общая длительность записи	Рассказ	Разговор	Пересказ
04	24:36.240	05:22.640 (21,9 %)	12:37.920 (51,3 %)	06:35.680 (26,8 %)
22	18:04.960	03:37.960 (20,1 %)	08:48.280 (48,7 %)	05:38.700 (31,2 %)
23	16:26.520	03:52.400 (23,5 %)	07:41.240 (46,8 %)	04:52.880 (29,7 %)

Как было упомянуто в разделе 1, существуют разные метрики измерения окуломоторной активности. Ниже мы на конкретных примерах покажем, что использование разных метрик дает принципиально разные результаты. Поэтому особенно важно хорошо понимать, какую метрику в каких случаях нужно выбирать в качестве основной. В табл. 2 и 3 приведены суммарные длительности фиксаций для N и R по трем записям эталонного подкорпуса, с разбивкой на этапы. Можно видеть, что запись № 22 (выделена серой заливкой), которая прежде всего будет интересовать нас дальше, отличается от двух других несколькими особенностями. Во-первых, N22 мало смотрит на R22 на этапе *рассказа*, меньше других смотрит на окружение на этапе *разговора* и смотрит почти только на R22 на этапе *пересказа*. Во-вторых, R22 на этапе *рассказа* смотрит исключительно на N22, а на этапе *пересказа* много смотрит на L22.

В табл. 4 приведены частично те же цифры по длительности фиксаций для N и R на этапе *разговора* записи № 22, однако к ним добавлены данные по количеству фиксаций. Мы видим, например, что соотношение фиксаций N22 на R22 и окружении существенно меняется в зависимости от того, какой метрикой мы пользуемся (22 % на R22 vs. 24 % на окружении в случае подсчета количества фиксаций и 34 %

Таблица 2

**Распределение зрительного внимания N по этапам, данные анализа суммарной длительности фиксаций больше 100 мс в программе Tobii Analyzer, в %**

Фиксации на Этап		R	C	Окружение (включая L)
Рассказ	№ 04	61	0	39
	№ 22	21	0	79
	№ 23	72	0	28
Разговор	№ 04	68	5	27
	№ 22	34	56	10
	№ 23	21	57	22
Пересказ	№ 04	91	0	9
	№ 22	99	0	1
	№ 23	82	0	18

Таблица 3

**Распределение зрительного внимания R по этапам, данные анализа суммарной длительности фиксаций больше 100 мс в программе Tobii Analyzer, в %**

Фиксации на Этап		N	C	Окружение/L на этапе пересказа
Рассказ	№ 04	98	1	1
	№ 22	100	0	0
	№ 23	95	1	4
Разговор	№ 04	47	40	13
	№ 22	41	59	0
	№ 23	45	48	7
Пересказ	№ 04	0	0	47/53
	№ 22	2	0	38/60
	№ 23	4	1	61/32



на R22 vs. 10% на окружении в случае подсчета длительности фиксации). Данные различия вполне закономерны, так как фиксации на окружении обычно менее долгие по сравнению с фиксациями на собеседнике.

Таблица 4

**Распределение зрительного внимания N и R на этапе *разговора* записи № 22, данные анализа количества фиксаций и их длительности в программе Tobii Analyzer**

Участник		Фиксации на	R/N	С	Окружение
N	кол-во фиксаций/%		125/22	307/54	138/24
	длительность фиксаций, с/%		145 000/34	245 600/56	45 400/10
R	кол-во фиксаций/%		516/51	459/45	42/4
	длительность фиксаций, с/%		160 000/41	228 000/59	2000/0

Наконец, в табл. 5 мы сравнили то же количество фиксаций для N22 и R22 с количеством их взглядов, получив еще большие различия. Так, при переходе от количества фиксаций к количеству взглядов соотношение зрительного внимания к R22 и C22 меняется у N22 с 22 vs. 54 на 40 vs. 35, а у R22 — с 51 vs. 45 на 32 vs. 67.

Таблица 5

**Распределение зрительного внимания N и R на этапе *разговора* записи № 22, данные анализа количества фиксаций и количества взглядов**

Участник		Фиксации на	R/N	С	Окружение
N	кол-во фиксаций/%		125/22	307/54	138/24
	кол-во взглядов, с/%		35/40	31/35	22/25
R	кол-во фиксаций/%		516/51	459/45	42/4
	кол-во взглядов, с/%		64/32	132/67	2/1

Таким образом, тип выбранной метрики заметно влияет на получаемые результаты. В данном исследовании мы будем считать основной количественной характеристикой общения коммуникантов **количество их взглядов**. Ниже мы проанализируем небольшой фрагмент эталонного подкорпуса, включающий этап *разговора* между тремя собеседниками в записи № 22 длительностью около 8,5 мин. При аннотировании материала мы взяли за основу существующую полуавтоматическую аннотацию и объединили фиксации и саккады, направленные в одну область интереса, в один взгляд. В результате таких преобразований мы получили 88 взглядов N22, 198 взглядов R22 и 98 взглядов C22<sup>3</sup>. Если сопоставить взгляды трех коммуникантов, оказывается, что

<sup>3</sup> Взгляды участника C22, на которого не был надет айтрекер, были восстановлены приблизительно по видеозаписи. В целом подавляющее число взглядов на собеседника составляли взгляды на его лицо, однако точные цифры не релевантны для настоящей работы; тем не менее стоит отметить, что большое количество взглядов на лицо собеседника со стороны слушающего не является универсальным, ср. известную работу [Rossano 2009] о культуре народа чельталь, говорящего на одноименном языке в Мексике, в которой не принято смотреть на лицо говорящего собеседника. Сильные различия в количестве взглядов N22 и R22 обусловлены в первую очередь

они хорошо синхронизованы друг с другом и практически без исключений описываются несколькими интуитивно понятными правилами, приводимыми ниже<sup>4</sup>.

Напомним, что на этапе *разговора* в коммуникации активно участвуют все три участника. Задача коммуникантов состоит в том, чтобы подготовить R к последующему этапу *пересказа*, сообщив ему дополнительные подробности, забытые на предыдущем этапе, а также уточнив неясные места. Поэтому именно R является основным адресатом этого этапа. Основного адресанта этапа определить сложнее — несмотря на то, что N являлся единственным адресантом предыдущего этапа *рассказа*, он может остаться таковым и на этапе *разговора*.

**Правило 1<sup>5</sup>.** Когда говорящий (отличный от основного адресата, в данном примере C22) говорит, два других участника (N22 и R22) смотрят на него. Сам говорящий смотрит на основного адресата — на R22.

241.81 — 243.87	—одна /—почти полная, и в=    /третья —пустая.	
-----------------	---	--

**Правило 1а.** Когда говорящий (в данном примере C22) говорит, а два других участника (N22 и R22) смотрят на него, говорящий может некоторое время смотреть не на основного адресата R22, а на N22 — в случае, если он апеллирует к его словам.

288.24 — 290.80	Причём /мальчик /— сначала, видимо, хотел действительно взять несколько /груш,	
-----------------	---	--

**Правило 1б.** Когда говорящим является R22, а N22 и C22 смотрят на него, говорящий чередует взгляды на обоих коммуникантов — во время произнесения слов *Почему вы решили* R22 смотрит на C22, а во время произнесения слов *что он её \украл?* — уже на N22.

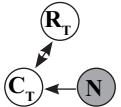
658.09 — 659.73	/Почему вы решили, что он её \украл?	
-----------------	---	--

различием в количестве взглядов на руки (75 для R22 и всего 4 для N22), которые разрывали взгляды на лицо.

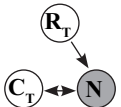
<sup>4</sup> Следует отметить, что проведенный нами ранее поиск закономерностей окуломоторного поведения коммуникантов при анализе количества и длительностей фиксаций не принес никаких интерпретируемых результатов.

<sup>5</sup> Ниже в иллюстративных примерах приводится время начала и конца данного фрагмента, вокальный транскрипт (подготовленный Н. А. Коротаевым), а также окуломоторная схема. На схеме три коммуниканта обозначены окружностями, говорящий коммуникант обозначен серой заливкой, стрелками обозначено направление взгляда. Идея подобного обозначения была взята нами из работ [Kendon 1967; Rossano 2012; Auer 2017].

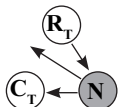
**Исключение из правила 1.** В анализируемом отрывке обнаружилось 5 случаев, в которых участники не переводили взгляд «на голос» говорящего коммуниканта. В данном примере перед анализируемой репликой один из коммуникантов (в данном примере С22) говорил и смотрел на основного адресата R22, два других (N22 и R22) смотрели на С22. Продолжая смотреть на С22, N22 произносит свое высказывание, однако оба коммуниканта никак на него не реагируют.

283.41 — 283.57	\Вообще не помню \↑этого!	
-----------------	---------------------------	---

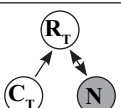
**Правило 2.** Пусть в некоторой коммуникативной ситуации говорит один из коммуникантов (в данном случае С22), а два других (N22 и R22) смотрят на него. Потом один из слушающих (N22), продолжая смотреть на говорящего, вмешивается в разговор, после чего два других участника переводят на него взгляд (в данном примере R22 переводит взгляд на N22 через 400 мс после начала звучания высказывания).

246.31 — 247.00	/Можно \вопрос?	
-----------------	-----------------	---

**Правило 3.** Во время своей реплики говорящий (в данном примере N22) иногда отводит взгляд от адресата сообщения (С22), некоторое время смотрит на окружение, а затем его взгляд возвращается обратно на С22.

250.76 — 254.08	если \один-н    (ц 0.26) (?0.11) \одну корзину — (полностью \полную,) — /увезли,	
-----------------	---	--

**Правило 4.** Пусть в некоторой коммуникативной ситуации говорящий (в данном примере С22) говорит (в данном примере описывая девочку на велосипеде и говоря: *и чёрные косы*), а два других коммуниканта (N22 и R22) смотрят на него. Потом другой коммуникант (N22), продолжая смотреть на говорящего (С22), подтверждает его слова, переводя (через 200 мс после начала звучания своего высказывания) взгляд на второго коммуниканта (R22), как бы передавая эту мысль дальше; в тот же момент говорящий (С22) также переводит взгляд на другого коммуниканта (R22).

696.66 — 697.83	\Да, вот /косы очень \длинные.	
-----------------	--------------------------------	---

## Заключение

В данной статье мы с разных точек зрения рассмотрели вопрос о коммуникативной функции взгляда. В разделе 1 были введены сведения из области физиологии зрения, включая ключевое противопоставление фиксаций и саккад, основные используемые метрики, а также вопрос об уникальности человеческого глаза с точки зрения выполнения им коммуникативной функции. В разделе 2 была описана история айтрекинга, включающая четыре основных этапа. Раздел 3 был посвящен изучению движений глаз в повседневном общении. В разделе 4, опираясь на три ключевых исследования, мы описали функции движений глаз в естественной коммуникации. Наконец, в разделе 5 мы предложили собственное исследование движений глаз в ходе естественного мультиканального общения.

Наконец, посмотрим на результаты нашей работы с точки зрения коммуникативных функций. В статье под названием “Functions of gaze in social interaction: Communication and monitoring” автор хотела ответить на вопрос, какая из двух функций важнее — функция мониторинга/поиска информации или экспрессивная функция, которую автор называет коммуникативной [Abele 1986]; по результатам проведенного в исследовании эксперимента коммуникативная функция оказалась менее важной. Однако, как представляется с современной точки зрения, во-первых, все выделяемые функции направления взгляда являются в широком смысле коммуникативными — они все используются для нужд коммуникации. Во-вторых, возвращаясь к идее о двойственности направления взгляда, мы почти никогда не можем точно понять, какую именно функцию выполняет тот или иной взгляд. В-третьих, на наш взгляд, сама по себе задача в каждой конкретной коммуникативной ситуации выделить какую-то одну, доминантную функцию является неправильной, так как в любой момент времени коммуникант является одновременно как адресатом, так и адресантом сообщения. Например, когда он смотрит на того или иного собеседника с целью мониторинга его реакции на свои слова, он своим взглядом именно на этого собеседника посылает сигнал о значимости для него именно этой информации.

Данная работа будет в дальнейшем продолжена на большем материале с применением статистических методов.

## Литература

Барабаничиков В. А., Жегалло А. В. Распознавание экспрессий лица в ближней периферии зрительного поля // Экспериментальная психология. 2013. Т. 6. № 2. С. 58–83.

Гришина Е. А. Русская жестикуляция с лингвистической точки зрения. Корпусные исследования. М. : Языки славянской культуры, 2017. 744 с.

Кибрик А. А. Русский мультиканальный дискурс. Часть I. Постановка проблемы // Психологический журнал. 2018а. Т. 39 (1). С. 70–80.

Кибрик А. А. Русский мультимедийный дискурс. Часть II. Разработка корпуса и направления исследований // Психологический журнал. 2018б. Т. 39 (2). С. 78–89.

Кибрик А. А., Подлеская В. И. (ред.) Рассказы о сновидениях: корпусное исследование устного русского дискурса. М. : ЯСК, 2009. 736 с.

Литвиненко А. О., Николаева Ю. В., Кибрик А. А. Аннотирование русских мануальных жестов: теоретические и практические вопросы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2017». М. : РГГУ, 2017. С. 271–286.

Урысон Е. В. Проблемы исследования языковой картины мира: Аналогия в семантике. М. : Языки славянской культуры, 2003. 224 с.

Федорова О. В. Распределение зрительного внимания собеседников в естественной коммуникации: 50 лет спустя // Когнитивная наука в Москве: новые исследования. Материалы конференции 15 июня 2017 г. / под ред. Е. В. Печенковой М. В. Фаликман. М. : БукиВеди : ИППиП, 2017. С. 370–375.

Ярбус А. Л. Роль движений глаз в процессе зрения. М. : Наука, 1965.

Abele A. Functions of gaze in social interaction: Communication and monitoring // Journal of Nonverbal Behavior. 1986. Vol. 10 (2). P. 83–101.

Argyle M., Ingham R., Alkema F. McCallin M. The Different Functions of Gaze // Semiotica. No. 7 (1). P. 19–32.

Brône G., Oben B. InSight Interaction. A multimodal and multifocal dialogue corpus // Language Resources and Evaluation. 2015. Vol. 49 (1). P. 195–214.

Buswell G. T. How people look at pictures. Chicago, 1935.

Buswell G. T. How adults read. Chicago, 1937.

Chafe W. (ed.). The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. Norwood, 1980. 327 p.

Darwin C. *The expression of the emotions in man and animals*. N. Y. : D. Appleton & Company, 1872. 366 p.

Deubel H., Schneider W. X. Saccade target selection and object recognition: Evidence for a common attentional mechanism // *Vision Research*. 1996. No. 36 (12). P. 1827–1837.

Engbert R. Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception // *Progress in Brain Research*. 2006. No. 154. P. 177–192.

Foulsham T. Eye movements and their functions in everyday tasks // *Eye*. 2015. No. 29 (2). P. 196–199.

Foulsham T., Walker E., Kingstone A. The where, what and when of gaze allocation in the lab and the natural environment // *Vision Research*. 2011. No. 51. P. 1920–1931.

Gallup A. C., Chong A., Couzin I. D. The directional flow of visual information transfer between pedestrians // *Biological Letters*. 2012. No. 8. P. 520–522.

Gobel M. S., Kim H. S., Richardson D. C. The dual function of social gaze // *Cognition*. 2015. Vol. 136. P. 359–364.

Jakob R. J. K., Karn K. S. Commentary on Section 4 — Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises // *The Mind's Eye*. North-Holland, Amsterdam, 2003. P. 573–605.

*Jording M., Hartz A., Bente G., Schulte-Rüther M., Vogeley K.* The “Social Gaze Space”: A Taxonomy for Gaze-Based Communication in Triadic Interactions // *Frontiers in Psychology* . No. 9. P. 226.

*Hollands M. A., Marple-Horvat D. E.* Coordination of eye and leg movements during visually guided stepping // *Journal of Motor Behavior*. 2001. No. 33. P. 205–216.

*Kendon A.* Some functions of gaze direction in social interaction // *Acta Psychologica*. 1967. Vol. 26. P. 22–63.

*Kobayashi H., Kohshima S.* Unique morphology of the human eye // *Nature*. 1997. No. 387. P. 767–768.

*Kobayashi H., Kohshima S.* Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye // *Journal of Human Evolution*. 2001. No. 40. P. 419–435.

*Land M., Mennie N., Rusted J.* The roles of vision and eye movements in the control of activities of daily living // *Perception*. 1999. No. 28. P. 1311–1328.

*Land M., Hayhoe M.* In what ways do eye movements contribute to everyday activities? // *Vision Research*. 2001. No. 41. P. 3559–3565.

*Land M.* Vision, eye movements, and natural behavior // *Visual Neuroscience*. 2009. No. 26. P. 51–62.

*Patla A. E., Vickers J. N.* How far ahead do we look when required to step on specific locations in the travel path during locomotion? // *Experimental Brain Research*. 2003. No. 148. P. 133–138.

*Rayner K.* Eye Movements in Reading and Information Processing: 20 Years of Research // *Psychological Bulletin*. 1998. No. 124 (3). P. 372–422.

*Reichle E. D., Rayner K., Pollatsek A.* The E-Z Reader model of eye-movement control in reading: Comparisons to other models // *Behavioral and Brain Sciences*. 2003. No. 26. P. 445–476.

*Rossano F., Brown P., Levinson S. C.* Gaze, questioning, and culture // *Comparative Studies in Conversation Analysis* / ed. by J. Sidnell. Cambridge University Press : Cambridge, 2009. P. 187–249.

*Rossano F.* Gaze behavior in face-to-face interaction. PhD MPI Psycholinguistics, Nijmegen, 2012.

*Smith D. T., Schenk T.* The premotor theory of attention: time to move on? // *Neuropsychologia*. 2012. Vol. 50. P. 1104–1114.

*Vertegaal R., Slagter R., Van der Veer G., Nijholt A.* Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes // *Proceedings of the Conference on Human Factors in Computing Systems*. 2001.

*Yarbus A. L.* Eye Movements and Vision. N. Y., 1967.

**O. V. Fedorova**

*Lomonosov Moscow State University, Institute of Linguistics RAS*

*(Russia, Moscow)*

*olga.fedorova@msu.ru*

## **ON THE COMMUNICATIVE FUNCTION OF THE GAZE**

The paper considers the communicative function of the gaze from different points of view. First, the author introduces basic information from the field of physiology of oculomotor activity, including the contraposition of fixations and saccades, the main metrics used, and the uniqueness of the human eye in terms of its communicative function. Then, the history of eyetracking — the method of the registration of eye movements, consisting of four main stages, — is described. A separate section is dedicated to the research of eye movements in everyday communication, including doing housework, walking and communication in non-laboratory settings. Then, the main functions of eye movements in natural communication are described: monitoring function, regulatory function, expressive function and eye contact. Finally, a qualitative and quantitative analyses of a small fragment with a duration of about 8.5 minutes of natural communication between the three interlocutors, two of which are equipped with eyetracking glasses, is proposed. It is shown that the use of different metrics — the number and duration of fixations or gazes — gives fundamentally different results. In this study, the main quantitative characteristic of interlocutors' communication is the number of their gazes. When comparing the gazes of three interlocutors, it turns out that they are well synchronized with each other and are described almost without exception by several rules.

*Keywords:* multichannel communication, eye movements, eyetracking, fixation, gaze, visual attention, communicative function.

## **References**

- Abele A. Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 1986, vol. 10 (2), pp. 83–101.
- Argyle M., Ingham R., Alkema F. McCallin M. The Different Functions of Gaze. *Semiotica*, 1973, vol. 7 (1), pp. 19–32.
- Barabanshchikov V. A., Zhegallo A. V. [Recognition of facial expressions in the near periphery of the visual span]. *Ekspperimental'naya psikhologiya*, 2013, vol. 6 (2), pp. 58–83. (In Russ.)
- Brône G., Oben B. InSight Interaction. A multimodal and multifocal dialogue corpus. *Language Resources and Evaluation*, 2015, vol. 49 (1), pp. 195–214.
- Buswell G. T. How adults read. Chicago, 1937.
- Buswell G. T. How people look at pictures. Chicago, 1935.
- Chafe W. (ed.). The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. Norwood, 1980. 327 p.



Darwin C. *The expression of the emotions in man and animals*. N. Y., D. Appleton & Company, 1872. 366 p.

Deubel H., Schneider W. X. Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 1996, vol. 36 (12), pp. 1827–1837.

Engbert R. Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception. *Progress in Brain Research*, 2006, vol. 154, pp. 177–192.

Fedorova O. V. [Distribution of the interlocutors' visual attention in natural communication: 50 years later]. *Kognitivnaya nauka v Moskve: novye issledovaniya. Materialy konferentsii 15 iyunya 2017 g.* [Cognitive science in Moscow: new research. The proceedings of the conference. 15 June 2017]. Eds. E. V. Pechenkova, M. V. Falikman. Moscow, Buki Vedi Publ., IPPiP, 2017, pp. 370–375. (In Russ.)

Foulsham T. Eye movements and their functions in everyday tasks. *Eye*, 2015, vol. 29 (2), pp. 196–199.

Foulsham T., Walker E., Kingstone A. The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, 2011, vol. 51, pp. 1920–1931.

Gallup A. C., Chong A., Couzin I. D. The directional flow of visual information transfer between pedestrians. *Biological Letters*, 2012, vol. 8, pp. 520–522.

Gobel M. S., Kim H. S., Richardson D. C. The dual function of social gaze. *Cognition*, 2015, vol. 136, pp. 359–364.

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

Hollands M. A., Marple-Horvat D. E. Coordination of eye and leg movements during visually guided stepping. *Journal of Motor Behavior*, 2001, vol. 33, pp. 205–216.

Jakob R. J. K., Karn K. S. Commentary on Section 4 — Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. *The Mind's Eye*. North-Holland, Amsterdam, 2003, pp. 573–605.

Jording M., Hartz A., Bente G., Schulte- R  ther M., Vogeley K. The “Social Gaze Space”: A Taxonomy for Gaze-Based Communication in Triadic Interactions. *Frontiers in Psychology*, 2018, vol. 9, pp. 226.

Kendon A. Some functions of gaze direction in social interaction. *Acta Psychologica*, 1967, vol. 26, pp. 22–63.

Kibrik A. A. [The Russian multichannel discourse. Part I. Problem statement]. *Psikhologicheskii zhurnal*, 2018a, vol. 39 (1), pp. 70–80. (In Russ.)

Kibrik A. A. [Russian multichannel discourse, Part II: A corpus and avenues of research]. *Psikhologicheskii zhurnal*, 2018b, vol. 39 (2), pp. 79–90. (In Russ.)

Kibrik A. A., Podlesskaya V. I. *Rasskazy o snovideniyakh: korpusnoe issledovanie ustnogo russkogo diskursa* [Night Dream Stories: A corpus study of spoken Russian discourse]. Moscow, Yazyki slavyanskoi kul'tury Publ., 2009. 736 p. (In Russ.)

Kobayashi H., Kohshima S. Unique morphology of the human eye. *Nature*, 1997, vol. 387, pp. 767–768.

Kobayashi H., Kohshima S. Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, 2001, no. 40, pp. 419–435.

Land M. Vision, eye movements, and natural behavior. *Visual Neuroscience*, 2009, vol. 26, pp. 51–62.

Land M., Hayhoe M. In what ways do eye movements contribute to everyday activities? *Vision Research*, 2001, vol. 41, pp. 3559–3565.

Land M., Mennie N., Rusted J. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 1999, vol. 28, pp. 1311–1328.

Litvinenko A. O., Nikolaeva Ju. V., Kibrik A. A. [Annotation of Russian manual gestures: theoretical and practical issues]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoi Mezhdunarodnoi konferentsii "Dialog 2017"* [Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue 2017"], iss. 16 (23), vol. 2, pp. 255–268. Moscow : RGGU Publ. (In Russ.)

Patla A. E., Vickers J. N. How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, 2003, vol. 148, pp. 133–138.

Rayner K. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 1998, no. 124 (3), pp. 372–422.

Reichle E. D., Rayner K., Pollatsek A. The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 2003, no. 26, pp. 445–476.

Rossano F. Gaze behavior in face-to-face interaction. PhD MPI Psycholinguistics, Nijmegen, 2012.

Rossano F., Brown P., Levinson S. C. Gaze, questioning, and culture. *Comparative Studies in Conversation Analysis*. Ed. by J. Sidnell. Cambridge University Press, Cambridge, 2009, pp. 187–249.

Smith D. T., Schenk T. The premotor theory of attention: time to move on? *Neuropsychologia*, 2012, vol. 50, pp. 1104–1114.

Uryson E. V. *Problemy issledovaniya yazykovoi kartiny mira: Analogiya v semantike* [Problems of the research of the linguistic picture of the world: Analogy in semantics]. Moscow, Yazyki slavyanskoi kul'tury Publ., 2003. 224 p. (In Russ.)

Vertegaal R., Slagter R., Van der Veer G., Nijholt A. Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. *Proceedings of the Conference on Human Factors in Computing Systems*, 2001.

Yarbus A. L. Eye Movements and Vision. New York, 1967.

Yarbus A. L. *Rol' dvizhenii glaz v protsesse zreniya* [The role of eye movements in the vision process]. Moscow, Nauka Publ., 1965. (In Russ.)

**Ю. В. Николаева**

*МГУ имени М. В. Ломоносова, Институт языкознания РАН*

*(Россия, Москва)*

*julianikk@gmail.com*

### **ЖЕСТОВЫЕ СБОИ В ДИАЛОГЕ\***

В статье на материале мультимодального корпуса «Рассказы и разговоры о грушах» ([www.multidiscourse.ru](http://www.multidiscourse.ru)) мы рассмотрели нарушения плавности жестикуляции. По аналогии с речевыми сбоями были выделены жестовые фальстарты (незавершенные жесты) и хезитации (замедленные фазы подготовки или ретракции жеста). В трех записях были изучены те фрагменты, в которых участники разговаривают между собой; общая длительность видеофрагментов составила примерно 30 минут. Мы разработали правила вычленения таких явлений и показали, что существует большая вариативность между участниками записи. Кроме того, не для всех жестов можно говорить о явных жестовых сбоях: жестовые ударения (beats) слишком короткие и простые по форме, и существуют вопросы относительно обоснованности выделения в них подготовительных фаз. Прагматические жесты (их также называют рекуррентными жестами или жестовыми семьями) демонстрируют некоторую устойчивость формы, в результате они часто встречаются в редуцированном виде, и различие фальстартов и полных реализаций в этом случае становится затруднительным. В изученном материале встретилось 70 жестовых сбоев, примерно 80 % из которых составили фальстарты, остальное — хезитации. Приблизительно две трети из них появились на границах реплик, что позволяет выдвинуть предположение об их важной роли в смене ролей в разговоре. Наибольшая доля таких жестовых сбоев (26 % от общего числа) — на последней ЭДЕ (элементарной дискурсивной единице) говорящего перед переходом слова к слушающему, однако они могут встречаться и в других контекстах.

---

\* Исследование выполнено в Институте языкознания РАН при финансовой поддержке РФФИ, грант № 14-18-03819 «Язык как он есть: русский мультимодальный дискурс».

Данная статья написана как продолжение работы, представленной в рамках доклада автора на конференции «Слово и жест» (Гришинские чтения), которая состоялась 8 февраля 2018 г. В докладе, развивая идеи, предложенные в главе «Грамматика взгляда» из недавно вышедшей монографии Е. А. Гришиной «Русская жестикуляция с лингвистической точки зрения» (2017), на материале одного разговора трех участников записи «Рассказы и разговоры о грушах» ([www](http://www)).

*Ключевые слова:* мультимодальные исследования, речевые сбои, жестикуляция, мультимодальный корпус, рассказы о грушах.

Один из ярких признаков разговорной речи — наличие речевых сбоев: пауз, фальстартов, хезитаций, самоисправлений. Довольно долго это явление оставалось вне сферы интересов лингвистики, однако со второй половины XX в. появились исследования самоисправлений говорящего ([Levelt, 1983, 1989; Schegloff et al. 1977]; см. также обзор в [Николаева 1970]). Изучение случаев, когда нарушается плавность порождения речи, помогает понять природу процессов, лежащих в основе порождения и понимания речи.

При этом известно, что коммуникация в ее самом распространенном и прототипическом варианте — мультимодальна: мы не только различаем слова, но еще слышим интонацию, с которой они были произнесены, учитываем особенности голоса и неречевые звуки, а если общение идет лицом к лицу — видим мимику, позы и движения собеседника [Кибрик 2008]. Исследование жестов, сопровождающих речь, как лингвистического феномена началось относительно недавно — несколько десятилетий назад, и для этой области еще нет вполне устоявшегося термина: говорят о кинесике [Birdwhistell 1952], невербальной семиотике [Крейдли-лин 2002], жестикуляции [Kendon 2004], мультимодальной лингвистике [Гришина 2017] и т. д. Очевидно, что разные термины предполагают и по-разному очерченный круг изучаемых явлений.

Известно, что речь и жестикуляция в коммуникации — два параллельных, но тесно переплетенных процесса [McNeill 1992; Kendon 2004]. В то время как про нарушения плавности речевого потока известно уже достаточно много, вопрос об особенностях и самом существовании подобного явления в жестикуляции еще предстоит изучить. Замечено, что при нарушении плавности речи наблюдается остановка жестикуляции [Mayberry, Jaques 2000; Seyfeddinipur, Kita 2001; Seyfeddinipur 2006; Esposito, Marinaro 2007; Graziano, Gullberg 2018], однако изучение явлений в жестикуляции, аналогичных речевым сбоям, взятых независимо от речи, нам не встретилось.

Мы исследовали случаи жестовых сбоев (замедления или незавершенные движения) в диалоге и сопоставили их с моментами смены ролей в разговоре. Определение жестовых сбоев и обоснование такого определения будут даны в разделе 1. В разделе 2 описываются мультимодальные подходы к описанию смены ролей в разговоре. В разделе 3 излагаются принципы отбора материала для анализа жестовых сбоев в рамках нашего исследования. Далее в разделе 4 описываются результаты анализа изучаемых явлений на корпусном материале. Наконец, в разделе 5 представлены выводы и дальнейшие направления исследований.

---

multidiscourse.ru) мы провели пилотное исследование движений рук собеседников на границах реплик [Николаева 2018]. В настоящей статье описывается продолжение этого исследования.

## 1. Жестовые и речевые сбои

Чаще всего работы в этой области сосредоточены на функциях жестов при речевых сбоях, показывая, что жесты так или иначе используются говорящим как средство компенсации или преодоления возникших трудностей либо регуляции взаимодействия между говорящим и слушающим [Yasinnik et al., 2005; Akhavan 2016; Stam, Tellier, 2017; Graziano, Gullberg 2018].

При этом существуют достаточно хорошо разработанные подходы к изучению речевых сбоев, включая корпусные методы (см., в частности, [Кибрик, Подлеская 2007]). Кроме того, существуют разные подходы к классификации речевых сбоев; так, в работе [Подлеская 2014] представлена подробная классификация с точки зрения нескольких независимых факторов: онлайн vs. офлайн и изоморфные vs. неизоморфные (то есть повторы vs. модификации, причем последние делятся, в свою очередь, на несколько типов). Большинство классификаций речевых сбоев, однако, так или иначе следуют подходу, предложенному в [MacLay, Osgood 1959], согласно которому речевые сбои делятся на четыре типа: филлеры, незаполненные паузы, повторы и, наконец, фальстарты или самоисправления. Мы изменили эту классификацию, чтобы применить ее к изучению аналогичных явлений в жестикуляции. При этом возникло несколько сложностей. Первая из них связана с тем, что при рассмотрении звучащей речи мы исходим из того, что паузы являются нарушением некоторых ожиданий, равно в диалоге и монологе. Однако этот подход неприменим к жестам, поскольку жестикуляция не обязательно будет непрерывной. В связи с этим незаполненные паузы, аналогом которых могло бы быть положение покоя для рук, не могут считаться сбоями.

Говоря о жестах, выделяют следующие фазы ([McNeill 1992], доработан подход [Kendon 1980]): подготовка (preparation — движение из положения покоя или положения, достигнутого после предыдущего движения), мах (stroke — основная и коммуникативно значимая фаза жеста), ретракция (recovery или return в [Kendon 1980], retraction в [McNeill 1992] — возвращение в положение покоя или некоторое промежуточное состояние). При этом обязательной является только маховая фаза, остальные могут быть пропущены — например, для цепочки жестов, следующих один за другим.

До и после маха возможны удержания (hold), которые служат для синхронизации речи и жестов, для передачи прагматических [Kendon 1995; Park-Doob 2010], семантических [Kita, Özyürek 2003], коммуникативных [Cibulka 2016] и других значений.

Аналогом заполненных пауз в жестикуляции можно считать удержания — остановки в реализации жеста перед или после основной (маховой) фазы. Однако известно, что маховая фаза жеста и соответствующие ей семантически слова обычно синхронизованы по времени (см., например, [Ferré 2010; Wagner, Malisz, Корр 2014]), а поскольку жест зачастую короче, чем вербальная единица, жестовые удержания могут служить для синхронизации жестов и речи. Следует отметить, что есть предположения о том, что паузы в жестах могут играть такую же

многоплановую роль, как паузы в речи [Esposito, Esposito 2010], однако эту гипотезу еще предстоит проверить. Кроме того, согласно приведенным выше исследованиям, остановки в жестикуляции объясняются в некоторых случаях нарушением плавности речи (опять же в связи с необходимостью синхронизации двух каналов), нашей же целью было изучение жестовых сбоев отдельно от речевых. Таким образом, жестовые удержания (полные остановки перед или после маховой фазы жеста) мы не включали в анализ жестовых сбоев, поскольку их когнитивную и коммуникативную роль как аналогов речевых пауз еще предстоит проверить.

В результате среди явлений в жестикуляции, сопоставимых с речевыми сбоями, мы можем выделить следующие: заполненные паузы и самоисправления. Аналогом заполненных пауз мы предлагаем считать замедление жестикуляции. При этом скорость маховой фазы может определяться очень многими факторами (например, семантическими, если в жесте важно показать замедленную скорость некоторого движения, или риторическими, аналогично замедленному произнесению какого-то речевого фрагмента), поэтому при рассмотрении замедлений в жестах мы ограничились вспомогательными фазами — подготовкой и ретракцией. В норме эти фазы выполняются по прямой траектории — кратчайшей от положения покоя до точки начала жеста для подготовки или в обратной последовательности для ретракции, и их скорость определяется общей скоростью речи и жестов на данном отрезке. Но иногда у говорящих встречаются очень медленные движения, нередко в сочетании с хезитациями в речи. В этих случаях, если амплитуда достаточно большая, можно отметить неравномерность, когда рука перемещается то быстрее, то медленнее, и это выглядит как ступенчатое движение. Можно предположить, что такое прерывистое движение во время вспомогательных фаз связано с нарушением удобного для говорящего темпа либо замедление и ускорение связаны с когнитивными процессами, обуславливающими возникновение и разрешение коммуникативных затруднений.

В [McNeill 1992] предлагается функциональная классификация жестов, наиболее распространенная на данный момент. Согласно ей, жесты делятся на жестовые ударения (beats) — простые по форме жесты, обычно выполняемые как короткие движения вверх-вниз; указательные жесты (указание на некоторого референта, присутствующего или воображаемого, или точку в пространстве); иконические (иллюстрации к описываемым референтам или событиям) и метафорические (соотнесенные с некоторым абстрактным понятием, не предполагающим физического воплощения). Мы предлагаем некоторые изменения этой классификации, опираясь на описания прагматических значений жестов, предложенные в [Kendon 1995, 2004]: последние две группы, иконические и метафорические жесты, выделять не на основании связи значений жестов и речи (буквальная и метафорическая иллюстрация, как это предлагает Д. Макнилл), а с точки зрения того, являются ли жесты иллюстрацией к содержанию слов (это будут изобразительные жесты) или же каким-то образом модифицируют значение вербального сообщения (например, уточняют достоверность/приблизительность или важность/второстепенность) либо отмечают, как слова соотносятся с более широким вербальным

и невербальным контекстом (например, центральность событий в рассказе или демонстрация того, что говорящий предполагает какую-то определенную реакцию от слушающего, эту группу мы предлагаем называть прагматическими [Nikolaeva 2017, 2018]). В любом случае, получаются четыре типа жестов, первый из которых (жестовые ударения) называют нереферентным, поскольку движение настолько простой формы крайне сложно связать с каким бы то ни было семантическим содержанием, а следующие три — референтными.

Фальстарты в жестикуляции выглядят как движение, прерванное до достижения семантически обусловленного или ожидаемого положения, и это явление в большинстве случаев достаточно хорошо заметно и не вызывает сомнений. Очевидно, что фальстарты можно заметить только для референтных жестов.

Кроме хезитаций и фальстартов в жестикуляции есть некоторые другие явления, которые иногда совпадают по времени с речевыми затруднениями и также выглядят как нарушение плавности. Во-первых, это многократные повторы сложного по форме и семантически нагруженного жеста, которые иногда можно интерпретировать как желание говорящего донести до адресата сложную или важную для него мысль, а иногда как повторы, аналогичные таковым при речевых сбоях. Во-вторых, среди кандидатов на жестовые сбои — наложение на референтные жесты простых по форме ритмических движений (жестовых ударений). Однако коммуникативные и когнитивные предпосылки для отнесения таких комплексов движений к интересующему нас классу явлений еще предстоит изучить.

## 2. Жесты и смена ролей

Развитие мультимодального направления в лингвистике меняет наш взгляд на то, какую роль жесты играют в коммуникации. При этом становится ясно, что все каналы общения важны для решения разных задач, в частности при смене ролей говорящего и слушающего в разговоре [Bohus, Horvitz 2010]. Известно, что в этом задействованы просодия и движения глаз (см., например, [Clemens, Diekhaus 2009]). Другие исследования, начиная с работы А. Кендона (1967), объясняют, как в таких ситуациях задействованы жесты. Согласно данным А. Кендона, в конце своей реплики говорящий останавливает жестикуляцию и (или) опускает руки в положение покоя. В работе [Duncan 1972] описываются сигналы, важные с точки зрения смены ролей в разговоре: это сигналы передачи слова (turn-yielding cues), сигналы обратной связи (back-channel cues) и сигналы сохранения роли говорящего (turn-maintaining cues). В дальнейшем исследовании Виманн и Кнапп [Wiemann, Knapp 1975] добавили к этому списку сигналы, запрашивающие право слова (turn-requesting cues). Сигналы передачи слова используются говорящим, чтобы сообщить адресату, что он или она сейчас закончат свою реплику и готовы передать слово другому. Эти сигналы могут быть просодическими, синтаксическими, лексическими или жестовыми. В последнем варианте это завершение жестикуляции. Сигналы обратной связи со стороны слушающего показывают, что он не собирается переходить к роли говорящего. В основном это просодические и лексические



знаки, но некоторые исследователи добавляют кивки, смены позы и жесты. Сигналы сохранения роли говорящего используются им, чтобы показать свое намерение продолжать говорить. Считается, что здесь жесты рук играют существенную роль, но часто встречаются и просодические средства (например, заполненные паузы). Запрос слова — это демонстрация намерения слушающего начать говорить, что часто выражается через наложение реплик или короткие слова или фразы, например «но эээ...» или «знаешь...» [Wei-dong 2007]. В работе [Duncan 1972] было показано, что удержания в жестах и расслабление напряженных рук заметно чаще появляются в конце реплики, таким образом сигнализируя слушающему о ее окончании.

Л. Мондада [Mondada 2007] изучала указательные жесты при смене ролей. Она показала, что это мультимодальный процесс, в котором участвуют языковой и вербальный каналы. Она также обнаружила, что смена ролей может быть инициирована с помощью жестов, иногда в опережение речи — до того, как текущая реплика говорящего будет завершена, или в начале реплики другого собеседника. В обоих случаях указание обозначает смену ролей участников, которую инициирует жестикулирующий, сообщая, что тот, на кого указывает, переходит из роли слушающего в роль возможного следующего говорящего. Помимо этого в работе [Cosnier 1996] говорится, что динамические жесты (противопоставленные жестовым удержаниям) в конце фразы могут показывать, что реплика не завершена, и говорящий продолжит после паузы.

Таким образом, разные авторы показали, что продолжающаяся жестикуляция служит слушающему сигналом, что реплика не закончена. Однако в реальном разговоре процесс передачи слова устроен гораздо сложнее. В книге Е. А. Гришиной (2017) описываются алгоритмы обмена взглядами, используемые говорящим и слушающим на потенциальных границах реплик. В нашем исследовании мы показываем, что применительно к жестовому каналу смена ролей также представляет собой сложно организованное и иногда многоэтапное взаимодействие участников разговора, а не просто однонаправленный сигнал от жестикулирующего к адресату, который слушающий может только принять, но не проигнорировать и не запросить.

### **3. Методика исследования и материал**

В рамках данного исследования мы использовали три записи «Рассказов и разговоров о грушах», которые были сделаны в рамках проекта «Русский мультимедийный дискурс» ([www.multidiscourse.ru](http://www.multidiscourse.ru)). Были выбраны те фрагменты записей, в которых участники разговаривают между собой, общей продолжительностью примерно 30 минут<sup>1</sup>.

В этих записях были отмечены все жесты, для которых можно было определить существенное замедление движения во время подготовки или ретракции (хезитации) или незавершенность жеста (фальстарты). Такие моменты жестовых сбоев были сопоставлены со сменой ролей в разговоре. Положение вербальной единицы

---

<sup>1</sup> Аннотация жестов, на основе которой размечали жестовые сбои, выполнена автором совместно с А. Л. Литвиненко. Вокальный транскрипт сделан В. И. Подлесской и Н. А. Коротаевым.

(ЭДЕ — элементарной дискурсивной единицы), сопровождавшейся отмеченными жестами, относительно границы реплик определялось по следующим правилам.

1. Переход от говорящего к слушающему (S-L) наблюдался тогда, когда ЭДЕ с жестовым сбоем была последней перед репликой другого собеседника или если после нее следовало только короткое слово или фраза, побуждающая адресата говорить (см. пример 1, строка R-vE032).

Пример 1.

R-vE029 И он весь такой /устал,

R-vE030 \Да?

R-vE031 /**потный**,<sup>2</sup>

R-vE032 /Нет?

2. Переход от слушающего к говорящему отмечался, если ЭДЕ с жестовым сбоем была первой или второй у данного говорящего в данной реплике (следует учитывать, что первые слова слушающего в разговоре нередко игнорируются говорящим либо они выступают скорее как прагматический маркер смены говорящего, чем как содержательная единица, см. пример 2, строка N-vE158).

Пример 2.

N-vE158 Нет,

N-vE159 коза **тоже была**,

N-vE160 **сначала** был пету= петух,

3. Сохранение роли говорящего (S-S) отмечалось, если одновременно с репликой говорящего (в которой встретился жестовый сбой) или сразу после нее адресат попытался включиться в разговор, но говорящий продолжил свою реплику.

4. Сохранение роли слушающего (L-L) отмечалось, если слушающий использовал жест, чтобы показать свое намерение участвовать в разговоре (вместе со словами или без), но говорящий, несмотря на это, продолжал свою реплику. Жестовые фальстарты без речи всегда считались как L-L, поскольку по умолчанию жестикулирует говорящий [McNeill 1992], а слушающему остаются адапторы и смены позы. То есть движение рук слушающего, которое интерпретируется как коммуникативно нагруженное (в отличие от адапторов и смен положения), — это попытка стать говорящим. Редкое исключение — такие жесты слушающего, которыми он дополняет слова говорящего (а не обращается к нему и не иллюстрирует собственный концепт). Иллюстраторы слушающего распознаются по согласованию, временно́му и семантическому, с речью.

Если жестовый сбой встретился в середине реплики говорящего и не было никаких попыток слушающего принять активное участие в разговоре, эти случаи классифицировались как «прочие» (табл. 1). Очевидно, что в эту категорию попали разнородные примеры. Такие жестовые сбои могут отмечать трудности подготовки вербального или жестового сообщения; сюда же относятся случаи, когда говорящий ожидает от слушающего ответной реплики, но слушающий не вступает в разговор.

<sup>2</sup> В транскриптах полужирным шрифтом выделен сегмент, на который наложен интересующий нас жест.

Вероятно, есть и другие причины, обуславливающие нарушения плавности жестикуляции, однако мы на данном этапе не располагаем инструментами для их различения.

Говоря об отборе корпусных примеров, следует отметить еще один тип жестов, которые заметно отличаются от прототипических, но не вошли в нашу выборку. В работе [Sekine 2017] предложен термин *disguised adaptors* для тех случаев, когда говорящий, начиная какой-то жест, меняет свое решение и вместо жеста исполняет адаптор [Ekman & Friesen, 1969] — движение, направленное на устранение дискомфорта или сдерживание эмоций, например поправить волосы или почесаться. Дословный перевод этого термина означает «скрытые адапторы», однако на самом деле в таких движениях оказываются скрыты жесты. Подробное изучение таких случаев еще предстоит осуществить.

Работа с корпусным материалом позволяет сделать некоторые выводы уже на этапе подготовки данных. В частности, можно говорить об индивидуальном стиле жестикуляции применительно к наличию, контекстам и форме жестовых сбоев. Так, среди девяти участников записи у одного (23R) не встретилось ни одного типичного жестового сбоя, однако часто в конце своих реплик он опускал руки очень быстро, таким образом жесты получались короткими и слабо артикулированными. По всей видимости, это не типичный сбой, когда говорящий или слушающий прерывает жест на середине или замедляет его, сообщая о своем намерении продолжить говорить или передать слово собеседнику, а знак, который можно было бы перевести как «Я закончил, теперь говорите вы». Эти сокращенные жесты встречались только в контекстах S-L, то есть в конце реплики говорящего, перед тем, как он стал слушающим.

Жесты, которые А. Кендон называет прагматическими [Kendon 1995], а К. Мюллер — рекуррентными [Bressem, Müller 2014], в речи используются немного иначе, чем иллюстраторы, поскольку форма последних зависит только от контекста и семантики сообщения. Типичный пример прагматического жеста, часто встречающегося в разговоре, — «метафора передачи» (“*conduit metaphor*” [McNeill 1992]): говорящий разворачивает ладонь вверх, пальцы направлены к собеседнику, и делает движение в сторону адресата, как бы предлагая ему на открытой ладони некоторое сообщение или объект. Поскольку такие жесты встречаются довольно часто, их форма становится узнаваемой, что делает допустимой их редуцированную реализацию, аналогично речевым клише: например, мы часто слышим «здрасьте» вместо произнесенного полностью «здравствуйте». В связи с этим такого типа жесты, даже выполненные в редуцированной форме, очень быстро и не доведенные до конца, не считались фальстартами.

#### 4. Результаты

В таблице 1 представлены сводные данные по всем трем разговорам с тремя участниками в каждом.

Как можно видеть из таблицы 1, в изученных нами разговорах жестовые сбои встречаются приблизительно 1 на 20 жестов, при этом фальстарты — в 4 раза чаще, чем жестовые гезитации. Самый частый контекст для жестовых сбоев — последняя

Жестовые сбои и смена ролей в разговоре

Участ- ник	Же- стов	Жесто- вых сбоев		Из жестовых сбоев				Из всех жестовых сбоев									
				Фальстар- тов		Хезита- ций		L-L		L-S		S-L		S-S		Прочие	
4N	170	4	2 %	4	100 %	0	0 %	0	0 %	0	0 %	3	75 %	1	25 %	0	0 %
4C	105	2	2 %	2	100 %	0	0 %	0	0 %	0	0 %	1	50 %	1	50 %	0	0 %
4R	227	29	13 %	23	79 %	6	26 %	8	28 %	4	14 %	8	28 %	0	0 %	9	31 %
22N	196	7	4 %	5	71 %	2	29 %	0	0 %	1	14 %	2	29 %	1	14 %	3	43 %
22C	175	6	3 %	6	100 %	0	0 %	0	0 %	0	0 %	0	0 %	1	17 %	5	83 %
22R	14	1	7 %	1	100 %	0	0 %	0	0 %	0	0 %	0	0 %	1	100 %	0	0 %
23N	406	14	3 %	10	71 %	4	29 %	0	0 %	4	29 %	4	29 %	3	21 %	3	21 %
23C <sup>3</sup>	167	7	4 %	4	57 %	3	43 %	0	0 %	2	29 %	0	0 %	1	14 %	4	57 %
Всего	1460	70	5 %	55	79 %	15	21 %	8	11 %	11	16 %	18	26 %	9	13 %	24	34 %

ЭДЕ в реплике, когда говорящий передает слово слушающему. Однако эти значения очень сильно зависят от говорящего.

В целом на границе реплик было две трети жестовых сбоев, что можно интерпретировать как подтверждение того, что они имеют большое значение при смене ролей в разговоре.

### Заключение

В данной статье мы рассмотрели жестовые сбои — явление, еще довольно мало изученное в мультимодальной лингвистике. Мы предложили критерии для выделения двух типов жестовых сбоев — фальстартов и хезитаций. Далее на корпусном материале были собраны данные о том, как жестовые сбои сочетаются с моментами смены ролей в разговоре. Наши данные показывают, что жестовые фальстарты и хезитации встречаются довольно часто (1 на 20 жестов), и две трети из них приходятся на границы реплик, особенно при передаче слова от говорящего к слушающему.

### Литература

Гришина Е. А. Русская жестикуляция с точки зрения. Корпусные исследования. М. : Языки славянской культуры, 2017. 744 с.

Кибрик А. А. Мультимодальная лингвистика: направления исследований // Обработка текста и когнитивные технологии (Когнитивное моделирование в лингвистике. Труды X международной конференции) / ред. В. Н. Поляков. Казань : Изд-во Казанского университета, 2008. С. 132–145.

<sup>3</sup> Участник 23R не включен в таблицу, поскольку у него не встретилось несомненных жестовых сбоев.

Кибрик А. А., Подлесская В. И. Самоисправления говорящего и другие типы речевых сбоев как объект аннотирования в корпусах устной речи // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2007. №2. С. 2–23.

Крейдлин Г. Е. Невербальная семиотика. М. : Новое литературное обозрение, 2002. 592 с.

Николаева Т. М. Новое направление в изучении спонтанной речи (о так называемых речевых колебаниях) [обзор] // Вопросы языкознания. 1970. №3. С. 117–123.

Николаева Ю. В. Жестовые сбои: контексты и функции // Слово и жест (Гришинские чтения). Материалы конференции. М., 2018.

Подлесская В. И. «То есть, не убили, а зарезали саблей»: самоисправления говорящего в устных рассказах // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». М. : РГГУ, 2014. Вып. 13 (20). С. 547–561.

Akhavan N., Goksun T., Nozari N. Disfluency production in speech and gesture // Proceedings of the 38th Annual Conference of the Cognitive Science Society / eds. A. Papafragou, D. Grodner, D. Mirman, J. C. Trueswell. Austin, TX : Cognitive Science Society, 2016. P. 716–721.

Birdwhistell R. L. Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture. Washington, DC : Department of State, Foreign Service Institute, 1952. 88 p.

Bohus D., Horvitz E. Facilitating Multiparty Dialog with Gaze, Gesture, and Speech // Proceedings of the ICMI'10, Beijing, China, 2010. P. 5.

Bressem J., Müller C. A repertoire of German recurrent gestures with pragmatic functions // Body — Language — Communication: An international handbook on multimodality in human interaction / eds. C. Müller et al. Berlin/Boston : De Gruyter Mouton, 2014. P. 1575–1592.

Cibulka P. How to do things with holds // Sign Language Studies. 2016. No. 16. P. 447–472.

Clemens C., Diekhaus C. Prosodic turn-yielding cues with and without optical feedback // Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL '09) / ed. Matthew Purver. Association for Computational Linguistics, Stroudsburg, PA, USA. 2009. P. 107–110.

Cosnier J. Les gestes du dialogue // Revue de Psychologie de la motivation. 1996. Nr. 21. P. 129–138.

Duncan S. Some Signals and Rules for Taking Speaking Turns in Conversations // Journal of Personality and Social Psychology. 1972. No. 23 (2). P. 283–292.

Ekman P., Friesen W. V. The repertoire of nonverbal behavior: Categories, origins, usage, and coding // Semiotica. 1969. No. 1. P. 49–98.

Esposito A., Esposito M. A. On Speech and Gestures Synchrony // Conference: Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues — COST 2102 International Conference, Budapest, Hungary, September 7–10 2010. P. 252–272.

*Esposito A., Marinaro M.* What pauses can tell us about speech and gesture partnership // *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue* / eds. A. Esposito, M. Bratanic, E. Keller, M. Marinaro. Amsterdam : IOS Press. 2007. P. 45–57.

*Ferré G.* Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French // *Language Resources and Evaluation, Workshop on Multimodal Corpora*, May 2010, Malta. W6, 2010. P. 86–91.

*Graziano M., Gullberg M.* When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons // *Frontiers in Psychology*. 2018. No. 9.

*Kendon A.* Gesticulation and speech: Two aspects of the process of utterance. The relationship of verbal and nonverbal communication. 1980. No. 25. P. 207–227.

*Kendon A.* Gestures as illocutionary and discourse structure markers in Southern Italian conversation // *Journal of Pragmatics*. 1995. No. 23 (3). P. 247–279.

*Kendon A.* *Gesture: Visible Action as Utterance*. UK : Cambridge University Press, 2004. 400 p.

*Kendon A.* Some Functions of Gaze-direction in Social Interaction // *Acta Psychologica*. 1967. No. 26. P. 22–63.

*Kita S., Özyürek A.* What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking // *Journal of Memory and Language*. 2003. No. 48 (1). P. 16–32.

*Levelt W.J.M.* Monitoring and self-repair in speech // *Cognition*. 1983. No. 14. P. 41–104.

*Levelt W.J.M.* *Speaking*. Cambridge, MA : MIT Press, 1989. 566 p.

*MacLay H., Osgood C.E.* Hesitation phenomena in spontaneous speech // *Word*. 1959. No. 14. P. 9–44.

*Mayberry R.I., Jaques J.* Gesture production during stuttered speech: insights into the nature of gesture-speech integration // *Language and Gesture* / ed. D. McNeill. Cambridge : Cambridge University Press. 2000. P. 199–214.

*McNeill D.* *Hand and mind: What gestures reveal about thought*. Chicago, IL, US : University of Chicago Press, 1992. 423 p.

*Mondada L.* Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers // *Discourse Studies*. 2007. No. 9 (2). P. 195–226.

*Nikolaeva Yu.V.* Pragmatic Gestures in Russian Retellings of “The Pear Stories” // *The Russian Journal of Cognitive Science*. 2017. Vol. 4 (2–3). P. 6–12.

*Nikolaeva Yu.V.* Pragmatic gestures reflect the speaker’s cognitive processes during story retelling // *Psychology. Journal of the Higher School of Economics*. In press. 2018.

*Park-Doob M.A.* *Gesturing Through Time: Holds and Intermodal Timing in the Stream of Speech*, 2010. 167 p.

*Schegloff E.A., Jefferson G., Sacks H.* The preference for self-correction in the organization of repair in conversation // *Language*. 1977. No. 53. P. 361–382.

*Sekine K.* Gestural hesitation reveals children’s competence on multimodal communication: Emergence of disguised adaptor // *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)* / eds. G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar. 2017. P. 3113–3118.

*Seyfeddinipur M.* Disfluency: Interrupting speech and gesture. Nijmegen : MPI-Series in Psycholinguistics, 2006. 196 p.

*Seyfeddinipur M., Kita S.* Gesture as an indicator of early error detection in self-monitoring of speech // Proceedings of the ISCA (International Speech Communication Association) Tutorial and Research Workshop. DiSS'01: Disfluency in spontaneous speech' University of Edinburgh, Scotland. 2001. P. 266–270.

*Stam G., Tellier M.* The sound of silence: the functions of gestures in pauses // Why Gesture? How the Hands Function in Speaking, Thinking and Communicating / eds. R. Breckinridge Church, M.W. Alibali, and S.D. Kelly. Amsterdam : Benjamins, 2017. P. 353–377.

*Wagner P., Malisz Z., Kopp S.* Gesture and speech in interaction: An overview // Speech Communication. 2014. No. 57. P. 209–232.

*Wei-dong Y.* Realizations of Turn-taking in Conversational interactions // US-China Foreign Language, USA (Serial No. 47). 2007. Vol. 5. No. 8. P. 19–30.

*Wiemann J., Knapp M.* Turn-taking in conversations // Journal of Communication. 1975. No. 25. P. 75–92.

*Yasinnik Y., Shattuck-Hufnagel S., Veilleux N.* Gesture Marking of Disfluencies in Spontaneous Speech, 2005. P. 173–178.

***Yu.V. Nikolaeva***

*Lomonosov Moscow State University, Institute of Linguistics of the Russian Academy of Sciences*

*(Russia, Moscow)*

*julianikk@gmail.com*

## **GESTURAL DISFLUENCIES IN DIALOGUE**

This paper presents a study concerning gesture disfluencies which is based on the multimodal corpus Russian Pears Chats and Stories ([www.multidiscourse.ru](http://www.multidiscourse.ru)). Similarly to speech disfluencies, we have described gesture falstarts (interrupted or aborted gestures) and hesitations (gestures with lengthened preparations or retractions) and suggested the procedure for marking these items in the corpus in relation to turn-taking. We have studied three interactions (total duration approximately 30 min). It has been shown that a huge variety of gesticulation among speakers is observed. Beat gestures have not been taken into account as they are too short and sharp. Pragmatic gestures (also called recurrent gestures or gesture families) display consistency of formal features in different contexts, so they often can be reduced or abbreviated. These two groups of gestures have not been included in the study.

We have found 70 cases of gesture disfluencies, about 80 % of which are falstarts, and others are hesitations. About two thirds of the disfluencies coincide with the end or the beginning of a turn. It allows us to speculate about an important role that gesture disfluencies play in turn-taking. Most of them (26 %) occur in the last phrase of the speaker in



the turn, marking passing turn to the listener, although gesture disfluencies can appear in other contexts too.

**Keywords:** multimodality, speech disfluencies, gesticulation, multimodal corpus, the pear stories

## References

Akhavan N., Goksun T., Nozari N. Disfluency production in speech and gesture. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Eds. A. Papafragou, D. Grodner, D. Mirman, J. C. Trueswell. Austin, TX, Cognitive Science Society, 2016, pp. 716–721.

Birdwhistell R. L. *Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture*. Washington, DC, Department of State, Foreign Service Institute, 1952. 88 p.

Bohus D., Horvitz E. Facilitating Multiparty Dialog with Gaze, Gesture, and Speech. *Proceedings of the ICMI'10, Beijing, China*, 2010, p. 5.

Bressem J., Müller C. A repertoire of German recurrent gestures with pragmatic functions. *Body — Language — Communication: An international handbook on multimodality in human interaction*. Eds. C. Müller et al. Berlin/Boston, De Gruyter Mouton, 2014, pp. 1575–1592.

Cibulka P. How to do things with holds. *Sign Language Studies*, 2016, no. 16, pp. 447–472.

Clemens C., Diekhaus C. Prosodic turn-yielding cues with and without optical feedback. *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL '09)*. Ed. Matthew Purver. Association for Computational Linguistics, Stroudsburg, PA, USA, 2009, pp. 107–110.

Cosnier J. Les gestes du dialogue. *Revue de Psychologie de la motivation*, 1996, no. 21, pp. 129–138.

Duncan S. Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology*, 1972, no. 23 (2), pp. 283–292.

Ekman P., Friesen W. V. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1969, no. 1, pp. 49–98.

Esposito A., Esposito M. A. On Speech and Gestures Synchrony. *Conference: Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues — COST 2102 International Conference*, Budapest, Hungary, September 7–10 2010, pp. 252–272.

Esposito A., Marinaro M. What pauses can tell us about speech and gesture partnership. *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*. Eds. A. Esposito, M. Bratanic, E. Keller, M. Marinaro. Amsterdam, IOS Press, 2007, pp. 45–57.

Ferré G. Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. *Language Resources and Evaluation, Workshop on Multimodal Corpora*, May 2010, Malta. W6, 2010, pp. 86–91.

Graziano M., Gullberg M. When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons. *Frontiers in Psychology*, 2018, no. 9.

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

Kendon A. Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, 1980, no. 25, pp. 207–227.

Kendon A. *Gesture: Visible Action as Utterance*. UK, Cambridge University Press, 2004. 400 p.

Kendon A. Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, 1995, no. 23 (3), pp. 247–279.

Kendon A. Some Functions of Gaze-direction in Social Interaction. *Acta Psychologica*, 1967, no. 26, pp. 22–63.

Kibrik A. A. [Multimodal linguistics: approaches of study]. V. N. Polyakov (ed.) *Obработка текста i kognitivnye tekhnologii (Kognitivnoe modelirovanie v lingvistike. Trudy X mezhdunarodnoi konferentsii)* [Text processing and cognitive technologies (Cognitive modeling in linguistics)]. Kazan', Kazan' university publishing company, 2008, pp. 132–145. (In Russ.)

Kibrik A. A., Podlesskaya V. I. [Speaker's self-repairs and other types of speech disfluencies as an object of annotation in spoken language corpora]. *Nauchno-tekhnicheskaya informatsiya. Seriya 2: Informatsionnye protsessy i sistemy* [Scientific-technical information. Series 2. Informational processes and systems]. 2007, no. 2, pp. 2–23. (In Russ.)

Kita S., Özyürek A. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 2003, no. 48 (1), pp. 16–32.

Kreidlin G. E. *Neverbal'naya semiotika'* [Nonverbal semiotics]. Moscow, Novoe literaturnoe obozrenie Publ., 2002. 592 p. (In Russ.)

Levelt W. J. M. Monitoring and self-repair in speech. *Cognition*, 1983, no. 14, pp. 41–104.

Levelt W. J. M. *Speaking*. Cambridge, MA, MIT Press, 1989. 566 p.

MacLay H., Osgood C. E. Hesitation phenomena in spontaneous speech. *Word*, 1959, no. 14, pp. 9–44.

Mayberry R. I., Jaques J. Gesture production during stuttered speech: insights into the nature of gesture-speech integration. *Language and Gesture*. Ed. D. McNeill. Cambridge, Cambridge University Press, 2000, pp. 199–214.

McNeill D. *Hand and mind: What gestures reveal about thought*. Chicago, IL, US, University of Chicago Press, 1992. 423 p.

Mondada L. Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 2007, no. 9 (2), pp. 195–226.

Nikolaeva T. M. [New approach to spontaneous speech studies (about so-called speech disfluencies)]. *Voprosy yazykoznaniya* [Topics in the study of language]. 1970, no. 3, pp. 117–123. (In Russ.)

Nikolaeva Yu.V. Pragmatic Gestures in Russian Retellings of “The Pear Stories”. *The Russian Journal of Cognitive Science*, 2017, vol. 4 (2–3), pp. 6–12.

Nikolaeva Yu.V. Pragmatic gestures reflect the speaker’s cognitive processes during story retelling. *Psychology. Journal of the Higher School of Economics*. In press. 2018.

Nikolaeva Yu.V. [Gestural disfluencies: contexts and functions]. *Slovo i zhest (Grishinskie chteniya). Materialy konferentsii* [Proc. of “Word and gesture (Grishina’s readings)”]. Moscow, 2018. (In Russ.)

Park-Doob M.A. Gesturing Through Time: Holds and Intermodal Timing in the Stream of Speech, 2010. 167 p.

Podlesskaya V. I. [“They Shot Him Dead, oh, no, They Knifed Him Dead with a Saber”: Self-Repairs in Oral Stories]. *Komp’yuternaya lingvistika i intellektual’nye tekhnologii: Po materialam ezhegodnoi Mezhdunarodnoi konferentsii «Dialog»* [Computational Linguistics and Intellectual Technologies. Proc. Of Annual conference “Dialogue”]. Moscow, Russian state university of humanities Publ., 2014, no. 13 (20), pp. 547–561. (In Russ.)

Schegloff E. A., Jefferson G., Sacks H. The preference for self-correction in the organization of repair in conversation. *Language*, 1977, no. 53, pp. 361–382.

Sekine K. Gestural hesitation reveals children’s competence on multimodal communication: Emergence of disguised adaptor. *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*. Eds. G. Gunzelmann, A. Howes, T. Tenbrink & E. Davelaar. 2017, pp. 3113–3118.

Seyfeddinipur M. Disfluency: Interrupting speech and gesture. Nijmegen, MPI-Series in Psycholinguistics, 2006. 196 p.

Seyfeddinipur M., Kita S. Gesture as an indicator of early error detection in self-monitoring of speech. *Proceedings of the ISCA (International Speech Communication Association) Tutorial and Research Workshop*. DiSS’01: Disfluency in spontaneous speech’ University of Edinburgh, Scotland, 2001, pp. 266–270.

Stam G., Tellier M. The sound of silence: the functions of gestures in pauses. *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*. Eds. R. Breckinridge Church, M. W. Alibali, and S. D. Kelly. Amsterdam, Benjamins, 2017, pp. 353–377.

Wagner P., Malisz Z., Kopp S. Gesture and speech in interaction: An overview. *Speech Communication*, 2014, no. 57, pp. 209–232.

Wei-dong Y. Realizations of Turn-taking in Conversational interactions. *US-China Foreign Language*, USA (Serial No. 47), 2007, vol. 5, no. 8, pp. 19–30.

Wiemann J., Knapp M. Turn-taking in conversations. *Journal of Communication*, 1975, no. 25, pp. 75–92.

Yasinnik Y., Shattuck-Hufnagel S., Veilleux N. Gesture Marking of Disfluencies in Spontaneous Speech, 2005, pp. 173–178.

<sup>1</sup>П.А. Бычкова, <sup>1,2</sup>Е.В. Рахилина, <sup>1</sup>Е.А. Слепак

<sup>1</sup>Школа лингвистики НИУ ВШЭ, <sup>2</sup>ИРЯ РАН

(Москва, Россия)

rakhilina@gmail.com

## ДИСКУРСИВНЫЕ ФОРМУЛЫ, ПОЛИСЕМИЯ И ЖЕСТОВОЕ МАРКИРОВАНИЕ\*

Лене Гришиной

*Да ну вас, да ну вас, да ну вас  
(Закат, предпоследний, погас).  
Житейская глупость и грубость  
Уже не касаются нас.*

*Касаются — влажные ветки  
Твоей побледневшей щеки.  
Заботы, как бурые белки,  
Умчались по веткам, легки.*

И. В. Чиннов. «Да ну вас, да ну вас, да ну вас...» (1978)

В статье представлено обсуждение маркирования семантики особых конструкций разговорной речи — дискурсивных формул. Вопрос о специальных средствах маркирования значения возникает в связи со свойством дискурсивных формул выступать в качестве законченных фиксированных реплик, без контекста внутри высказывания говорящего, и одновременно присущей им способности быть полисемичными. На примере русской многозначной дискурсивной формулы *Да ну!* показывается, как в случае полисемии таких единиц может происходить различение значений: помимо просодии смысловоразличительную функцию выполняет жестикация, сопровождающая высказывание. В тексте статьи приводится классификация контекстов употребления формулы *Да ну!* и выделяются пять основных ее значений: удивление, удивление-недоверие, девалоризация собеседника,

---

<sup>1</sup> Исследование выполнено при поддержке гранта РНФ № 16-18-02071

девальоризация речевого акта собеседника и «положительная» девальоризация опасения. Коротко освещаются предполагаемый путь конструкционализации формулы и развития каждого из значений. Далее производится количественный анализ жестикуляции, встретившейся при употреблении формулы в Мультимедийном корпусе русского языка, основанный на разметке всех движений говорящего в момент произнесения *Да ну!*. Выявляется корреляция между выделенными значениями и набором определенных жестов, что позволяет с одной стороны выстроить карту смежности существующих значений формулы, а с другой стороны получить представление о семантике свойственных разным значениям жестов.

*Ключевые слова:* грамматика конструкций, дискурсивные формулы, жестикуляция, полисемия, прагматика

## 1. Введение

Замысел этого исследования возник при работе над базой данных по русским конструкциям «Русский конструктикон», которая строится в рамках совместного проекта Школы лингвистики НИУ ВШЭ, Арктического университета Тромсе и университета Гётеборга. База «Русский конструктикон» аккумулирует неоднословные последовательности языковых единиц, которые в русском языке некомпозициональны, так что их семантика, так же как и у знаменитой английской конструкции *let alone* [Fillmore 1988], не складывается из семантики составляющих ее фрагментов. Важное свойство конструкций в том, что в их составе есть свободные слоты (переменные), которые позволяют конструкции встраиваться во внешний контекст, изменяться, изменять свойства заполняющих слоты единиц и т. д. [Рахилина, Кузнецова 2010]. Поэтому одна из задач базы «Русский конструктикон» — описать слоты каждой конструкции, разметить соответствующие им семантико-синтаксические роли и обозначить ограничения на их заполнение.

К настоящему времени база насчитывает уже около 700 конструкций с разметкой, толкованиями, примерами, а для части конструкций — переводами на английский и норвежский [описание базы см.: Janda et al. 2018]. Предполагается, что в результате в «Русском конструктиконе» будут собраны все русские конструкции — как классические, с фиксированными лексическими вставками, типа (1) или нетривиальным управлением, как (2), так и чисто грамматические, то есть состоящие из слотов с заданной грамматической информацией, ср. (3).

- (1) а/так что насчет XP? (а как насчет пятницы?)
- (2) NP-Асс звать NP-Ins /Nom (мальчика звать Юрой)
- (3) NP-Dat (не) VP-ся (мне не спится)

База «Русский конструктикон» решает важную практическую задачу, — ведь полного словаря конструкций пока нет ни для одного языка. Одновременно она выполняет и задачу теоретическую, связанную с уточнением границ понятия «конструкция»: если действительно удастся составить полный список конструкций

хотя бы для одного языка, среди них, помимо прототипических, наверняка обнаружатся и не вполне стандартные, периферийные и пограничные, имеющие нетривиальные свойства формы или семантики.

Один такой класс уже выделен, и он будет объектом описания в нашей статье. Это частный тип конструкций, который нуждается в особом формате представления в базе: изолированные составные реплики в диалоге, имеющие фиксированную форму и синтаксически равные предложению, типа: *Кто бы говорил! Что ты наделал! Еще бы!* и подобные. Вслед за Ч. Филлмором [Fillmore 1984] мы назвали этот класс конструкций *дискурсивными формулами* (*discourse formulas*). Дискурсивные формулы очень близки конструкциям: они неоднословны и всегда крайне идиоматичны. Особенность их в том, что они, в отличие от канонических конструкций, как правило, вовсе не имеют слотов. Это свойство вызвано их синтаксической изолированностью и неизменной, так сказать формульной, природой: они представляют собой диалоговые шаблоны, которые запоминаются и воспроизводятся говорящим как есть. Понятно, что такие формулы составляют важный пласт разговорного языка, а значит, должны быть собраны и полностью описаны. Эта проблема осознается лингвистическим сообществом, и похожие классы единиц разного объема и под разными названиями начинают постепенно вводиться в лингвистический оборот (ср. routines [Coulmas 1981; Aijmer 1996], коммуникативы [Шаронов 1996; Шаронов 2009] и др.).

Понятно, что для их описания требуются специальные исследования и специальные форматы представления данных, так что, например, в «Русском конструктиконе» этот тип конструкций выглядят как крайне периферийный, не похожий на прототип. Главное, что здесь упускает действующий формат «Русского конструктикона», — это роль жестикуляции и интонации. Именно на этой задаче мы остановимся подробнее. Чтобы понять ее значимость, мы специально выбрали здесь для анализа многозначную формулу *Да ну!*, имея в виду описать семантику ее основных употреблений и их корреляцию с сопутствующей жестикуляцией<sup>1</sup>.

Главным источником наших данных будет Мультимедийный подкорпус НКРЯ (МУРКО), собранный и сконструированный Леной (Еленой Александровной) Гришиной, и ее книга [Гришина 2017], в которой представлена разработанная ею методика анализа материала МУРКО<sup>2</sup>.

Лене Гришиной и ее беспрецедентному проекту мы и посвящаем эту статью.

<sup>1</sup> Как мы только что сказали, в разрешении многозначности дискурсивных формул может участвовать и интонация. Это обстоятельство нам важно, но для данной статьи — только как фон для описания жестикуляции, так что сведения об интонации будут даны лишь в самом общем виде [Дурягин, Рахилина 2019].

<sup>2</sup> Заметим, что и сама Е. А. Гришина когда-то решала похожую на нашу задачу описания многозначных дискурсивных маркеров *А!* и *О!* и их жестовых параметров [Гришина 2009; 2010].

## 2. Да ну: общая информация

Если судить по материалам НКРЯ, *да ну* является довольно частотной дискурсивной формулой: в основном корпусе нашлось 852 ее вхождения. Очевидно, что она является разговорной: в более формальном газетном корпусе, примерно равном по объему основному, имеется только 154 вхождения, в поэтическом корпусе, который, как правило, отражает более старшую норму и редко — современную

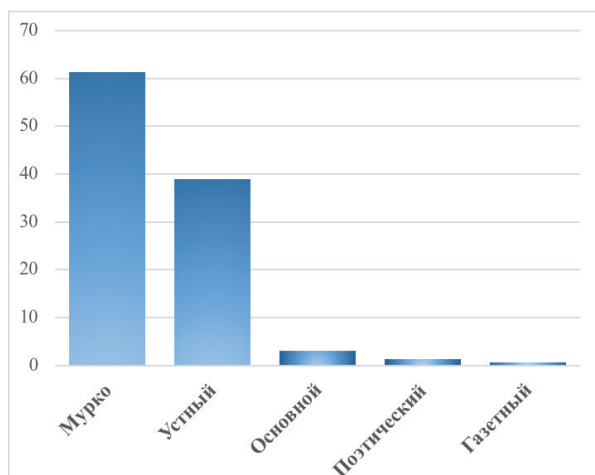


Рис. 1. Частотность *да ну* в подкорпусах НКРЯ, i.p.m.

разговорную [Рахилина, Плунгян 2012], встретилось всего 15 вхождений, а вот в устных корпусах таких вхождений много больше: в устном подкорпусе НКРЯ 471, а в МУРКО 291. Диаграмма (рис.1), которая отражает соотношение встречаемости формулы *Да ну!* в разных подкорпусах НКРЯ, подтверждает ее разговорный характер.

По данным НКРЯ, впервые *да ну* как стабильная конфигурация стало употребляться в начале XIX в., — но в уже устаревшем сейчас значении ускорения, ср. (4) и (5):

- (4) *Как ни лестно было для меня это приглашение, однако ж я долго отговаривался, извиняясь тем, что, не зная стихов, невозможно хорошо читать их, потому что легко дать им противосмысленную интонацию, но Гаврила Романович с нетерпением сказал: «Э, да ну, братец, читай!» (= ‘скорей читай’)* [С.П. Жихарев. Записки современника (1806–1809)].
- (5) *Да ну, брат, поскорее! Как ты копаешься!* [Н.В. Гоголь. Женитьба (1833–1842)].

Такие употребления почти исчезли — в современных текстах они заменились на *ну же*, *ну давай*, однако встречаются до сих пор, особенно в стилизациях, как речь эмигранта из романа Набокова (6) в переводе С. Ильина:

- (6) — *Чудно, Кармен, — ответила Долли и, обернувшись ко мне, продолжала по-русски. — По-моему, тебе не мешает выпить прямо сейчас. Да ну, пойдем же! И Бога ради, оставь ты здесь пиджак и жилет.* [Владимир Набоков. Смотри на арлекинов! (С. Ильин, 1999)].

Ср. здесь исходный англоязычный текст самого Набокова в (7), ясно указывающий на то, что выбранный эквивалент передает ровно этот уже устаревший смысл:



- (7) “*Righto, Carmen,*” replied Dolly, and turning to me continued in Russian: “*I think you need that drink right away. Oh, come along! And for God’s sake leave that jacket and waistcoat here.*” [Vladimir Nabokov. Look at the harlequins! (1974)].

Употребления (1)–(3) хорошо отражают исходную для *да ну* точку процесса, который по аналогии с *грамматикализацией*, то есть превращением лексемы в грамматический маркер, — называют *конструкционализацией* [Traugott, Trousdale 2013] — превращением семантически прозрачной последовательности лексических единиц в некомпозиционную конструкцию. Действительно, в исходном значении, которое иллюстрируется (1)–(3), *да ну* выступает как синоним *ну* (исходно — междометие, сопровождающее подстегивание лошади, ср. глагол *понукать*) с усиливающим его *да*: во всех этих контекстах *да* можно легко опустить. Таким образом, исходно, причем еще в начале XIX в., *да ну* представляло собой композиционную последовательность двух частиц *да* + *ну*. Теперь это некомпозиционная формула, развившая несколько значений — группы удивления и отрицания, к описанию которых мы переходим в следующих разделах.

#### Примечание

Косвенным свидетельством некомпозиционности *да ну* можно было бы считать дискурсивную формулу *ну да*, которая состоит из тех же, но собранных в другом порядке компонентов и обладает другим набором значений, — в основном согласия/подтверждения. Скорее всего, эти формулы просто совпали. Действительно, *ну да* восходит не к *ну*, а к *да*-согласия/подтверждения, интенсифицированному *ну*, которое добавляет ему идею ожидаемого, естественного следствия(8).

- (8) Ванька.<...>Ведь ты знаешь, барин мой звал ваших сюда на завтрак. Даша. **Ну, да!** мы затем и выехали. [И. А. Крылов. Пирог (1799–1801)]

Дополнительно о *ну да* см. [Кобозева и др. 2018].

### 3. *Да ну*: природа многозначности

Как мы уже говорили, основным материалом нашего исследования был Мультимедийный корпус русского языка. В нем было отобрано 190 видеофрагментов, содержащих формулу *да ну*, которая отделена восклицательным/вопросительным знаком, точкой, запятой или паузой. Все эти фрагменты были размечены: каждому клипу был приписан тег, соответствующий значению конструкции. Однако детальная семантическая разметка формул связана с трудностями: ведь прототипически дискурсивные формулы представляют собой отдельные предложения, и у них нет ближайшего контекста, который, как это обычно бывает, различался бы лексико-синтаксически и противопоставлял разные группы употреблений.

В поисках поверхностных маркеров семантических противопоставлений мы размечали жесты: движения рук, головы, корпуса, мимику и ориентацию взгляда — по тем фрагментам, на которых жестикуляция была хорошо видна, всего их

оказалось 78. Кроме того, клипы просматривались в программе PRAAT для определения интонационного контура. Косвенным способом проверки правильности деления на группы значений для нас были русские квазисинонимы формулы, а также ее переводные эквиваленты — в основном это были переводы на английский, подобранные по параллельному корпусу НКРЯ.

Таким образом, семантическая классификация принятых здесь значений *да ну* верифицировалась трояко: типологически, то есть переводом на другой или другие языки, а также типом жестикуляции и интонации<sup>3</sup>.

С большой долей очевидности употребление *да ну* делятся на две крупные семантические группы, которые мы условно назвали УДИВЛЕНИЕ (*Да ну? Неужели?*), как в (9), и ДЕВАЛОРИЗАЦИЯ — разновидность глубинного отрицания [Гришина 2017: 448], как в (10).

(9) А: *Эй, молодой человек, где ваш приятель?*

В: *А зачем?*

А: *Он увёл нашу Машу туда.*

В: *Да ну?! Вот молодец!*

[Иван Пырьев, Николай Погодин. Кубанские казаки, к/ф (1949)]

(10) А: *Надо в библиотеку сгонять.*

Б: *Да ну!!! Потом сгоняем.*

[Праздный разговор // Из материалов Ульяновского университета, 2007]

Эти две группы должны хорошо различаться интонационно: русскому удивлению/восхищению свойственна комбинация «восходящий + нисходящий акцент» (МН\* + НЛ\*) — [Князев, Пожарицкая 2011: 174; Касаткин 2007: 260], — тогда как контекстам девалоризации присущ нисходящий акцент НЛ\* [Дурягин, Рахилина 2019].

### Примечание

В НКРЯ нашлись переходные примеры, на основе которых можно восстановить семантический сдвиг, превративший маркер усиленного и желаемого ускорения в каждое из этих значений. Дело в том, что самым частотным и естественным контекстом для *да ну*-ускорения является (помимо исходного — ускорения движения) ускорение речи, когда говорящий торопит собеседника сообщить ему новую интересную информацию. В примере (11) середины XIX в. в этой функции сначала выступает просто *ну*, а потом, после того как уже получена некоторая новая информация, но все еще требуется новая, и *да ну*. В таком случае отделить семантику понукания (требования следующей порции информации) от семантики удивления (вызванного новыми сведениями) очень трудно, и с этой точки зрения (11) представляет промежуточный, переходный случай:

<sup>3</sup> Заметим, что разные диагностики не обязательно действуют синхронно — нас прежде всего будет интересовать роль жестикуляции.

- (11) — *Да пономарева-то могила... и того!*  
 — *Ну!*  
 — *И рассыпалась.*  
 — *Да ну!*  
 — *Обвалилась; дождем ее, знаешь, полило, она и провалилась.*  
 [Н. С. Лесков. Засуха (1862)]

Вторая траектория семантического развития *да ну*-ускорения связана с тем, что говорящий торопит собеседника перейти к разговору на интересующую его тему (или действию, как в примере (12)), а тот продолжает говорить о своем. Одновременно с желанием, чтобы собеседник скорее выполнил то, что от него требуется, говорящий испытывает раздражение по отношению к нему и к ситуации в целом. Это — потенциальный источник семантики девалоризации:

- (12) *Ну, что там? — заключает хозяйка, направляясь к двери и ведя за руку сына; но сделав два шага, она останавливается и продолжает, обернувшись к гостю боком: — Так я его тогда урезала — страсть! <...> «Это, говорю, подло, бесчестно, говорю, — четыре целковых». — Да ну, иди, что ль, к ней, — перебивает муж. — Пойду <...>* [Г. И. Успенский. Гость (1865)]

Дальнейшие наши исследования показывают, что противопоставления более дробны: в русском различается по крайней мере два типа удивления и несколько разных типов девалоризации. В свете данных МУРКО нам важно, что об этом свидетельствуют и жестовые диагностики.

## 4. Удивление

### 4.1. Собственно удивление

Удивление возникает как эмоция говорящего в ответ на новую информацию — именно ее маркирует конструкция *да ну*, которой в этом случае можно приписать следующее примерное толкование:

‘Говорящий не знал о том, что Р, собеседник ему сообщает, что Р. Неожиданность информации вызывает эмоцию удивления, говорящий принимает Р к сведению’.

Квазисинонимами к *да ну* в этих контекстах выступают такие реплики, как: *Ну и ну! Да ты что? Что вы говорите? Неужели? Серьезно?* Переводные эквиваленты, в сущности, тоже являются квазисинонимами, но в несколько более широком, типологическом смысле [Рахилина, Резникова 2013]. В английском это часто *Oh, my!* или просто *Oh!*, а также так называемые tag questions, ср. примеры (13) и (14) по параллельному корпусу: контекст показывает, что в обоих случаях говорящий сталкивается с новой для него информацией, удивлен ею и принимает ее как данное. Показателен последний пример, который демонстрирует нетождественность (по крайней мере для 60-х гг. XX в.) русского *О!* (в отличие от формально тождественного ему английского *Oh!*) группе маркеров удивления, которой принадлежит *да ну*:

- (13) *Conventions or not, but it was my birthday too, small difference of sixteen years, that's all. Oh my! Congratulations.* [Vladimir Nabokov. *Pale Fire* (1962)].  
Условность или нет, но это был также день моего рождения — каких-нибудь шестнадцать лет разницы, вот и все. **Да ну?** Поздравляю вас. [Владимир Набоков. *Бледный огонь* (Вера Набокова, 1983)].
- (14) *"No, Eeyore, I don't."* *"It means Learning, it means Education, it means all the things that you and Pooh haven't got. That's what A means."* *"Oh," said Piglet again. "I mean, does it?" he explained quickly.* [A. A. Milne. *The House at Pooh Corner* (1928)].  
— Нет, Иа, не знаю. — Оно означает Учение, оно означает Образование, Науки и тому подобные вещи, о которых ни Пух, ни ты не имеете понятия. Вот что означает А. — **О!** — снова сказал Пятачок. — Я хотел сказать «**Да ну?**» — поспешно пояснил он. [Алан Александр Милн. *Дом на Пуховой Опушке* (Б. Заходер, 1960)].

Как мы увидим дальше, хорошим поверхностным маркером этого класса употреблений *да ну* (как и следующих) оказывается жестикуляция: такая эмоция сопровождается характерными жестами и мимикой: корпус подается вперед, брови подняты, взгляд на собеседнике либо переходит (в терминологии Е. А. Гришиной) в зону коммуникации [Гришина 2017: 491]<sup>4</sup>.

#### 4.2. Удивление & отрицание (неверие)

Семантическая разметка формул часто дает промежуточные классы с комбинацией разных типов эмоций-реакций. В частности, одна и та же эмоция может фигурировать как в *да*-репликах, так и в *нет*-репликах, или положительных и отрицательных ответах. Таким образом, появление в разметке двух типов удивления вполне предсказуемо: если первое кодирует ситуацию, в которой говорящий принимает новую для него информацию, то должно быть второе, где говорящий по тем или иным причинам ставит ее под сомнение, не верит в сказанное:

- (15) *Костя почему-то поражен, и почему-то не верится ему, что у Майки была любовь с Женей: —Да ну, не может быть! — Было, — вздыхает мать.* [В. Ф. Панова. *Конспект романа* (1965)].

Ср. также аналогичный пример начала XX в.:

- (16) *Ты пойдешь, Данило, а я над вами воевуюю! —Да ну?— недоверчиво спросил Прилуков.* [А. Е. Зарин. *Кровавый пир* (1901)]

<sup>4</sup> Удивление может быть не только неподдельным, но и нарочитым, выражающим иронию говорящего по отношению к собеседнику и его новости. Для иронических контекстов жесты и мимика удивления не так релевантны, однако направление взгляда сохраняется и добавляется наклон головы, который в [Гришина 2017: 457, 458] служит одним из признаков отрицания.

Этот подтип удивления, который мы условно назвали НЕВЕРИЕ, можно истолковать примерно так:

‘Говорящий не знал о том, что Р, собеседник ему сообщает, что Р. Неожданность информации вызывает эмоцию удивления, говорящий не готов считать Р достоверным’.

Как видим, дополнительная семантика прямо противопоставляет этот тип удивления обычному. Квазисиноним *He может быть!* не передает этого противопоставления в полной мере, потому что в принципе он применим и к каким-то контекстам первого типа. Чтобы более точно отразить нужную семантику, здесь нужно значительно нарушить баланс удивления и отрицания в сторону отрицания, ср., например: *Никогда не поверю!*<sup>5</sup> Не случайно переводчики усматривают у этого типа *да ну* прямую связь с отрицанием, см. *No!* как его эквивалент в примере (17). В другом переводе с английского (18) *да ну* тоже возникает как маркер полного недоверия к сказанному собеседником, лишенного каких-либо положительных эмоций.

- (17) *He said he saw us with Eddie Harris at Martin's last night. "No!" They both giggled.* [Theodore Dreiser. *Sister Carrie* (1900)].  
— *Говорит, будто видел меня с Эдди Гаррисом у «Мартина».* — *Да ну? Девушки захихикали.* [Теодор Драйзер. *Сестра Керри* (М. Волосов, 1927)]
- (18) *No, lady, he said, I was at Buena Vista myself. Indeed, said Scarlett icily. Was I?* [Margaret Mitchell. *Gone with the Wind*, Part 1 (1936)]  
— *Ого! — воскликнул сержант.* — *Я сам дрался при Буэна-Виста, леди.* — *Да ну? — холодно произнесла Скарлетт.* [Маргарет Митчелл. *Унесенные ветром*, ч. 1 (Т. Озерская, 1982)]

Понятно, что никаких стандартных специальных контекстных маркеров для обозначения положительного/отрицательного отношения говорящего к новой информации нет, однако интонационный контур удивления-неверия (L\*, см. подробнее [Дурагин, Рахилина 2019]) отделяет его от обычного удивления и отождествляет со значениями девалоризации. Зато мимически и жестово эти типы контекстов

<sup>5</sup> Ср. также иронический вопрос *Да ладно?*, который имеет именно значение удивления-неверия (А) и невозможен в контексте простого удивления, как в (Б).

(А) — *Ну как пирожки?* — *Не хочу тебя расстраивать, но, по-моему, они испорченные.* — *Да ладно? <...> Да не, ну я бы сразу почувствовал.* [Дмитрий Дьяченко и др. *День радио*, к/ф (2008)]

(Б) *Вот что означает А.* — *О! — снова сказал Пятачок.* — *Я хотел сказать «Да ну/\*Да ладно?» — поспешно пояснил он.* [Алан Александр Милн. *Дом на Пуховой Опушке* (Б. Заходер, 1960)]

Исходное значение этой формулы совершенно иное, чем у *да ну*: его можно было бы описать как ‘безразличное согласие’, ср. (В). Однако в целом кластер значений *да ладно* очень близок к *да ну*.

(В) ...*Мстительные мероприятия отключаящих провайдеров вызвали скорее снисходительную к «Дождю» реакцию — «Да ладно, черт с ним».* [Максим Соколов. *Дубина общественного возмущения* // *Известия*. 2014.30 янв.]

и противопоставлены, и отделены от других. Так, в случае неверия брови могут быть нахмурены, а плечи подняты (пожимание плечами) — и эта мимика/жестикуляция для обычного удивления не характерна. А жест «махнуть рукой», свойственный в той или иной мере всем остальным классам девалоризации, отсутствует и в удивлении, и в удивлении-недоверии.

## 5. Девалоризация

Принципиально другое значение формулы *да ну* возникает при смене типа стимульной реплики. До сих пор мы обсуждали контексты, в которых речевым стимулом для говорящего было утверждение собеседника, содержащее новую информацию. Эту информацию он либо принимал, либо обесценивал. Между тем реакция девалоризации может возникать и в ответ на речевые стимулы других типов, относиться к разным параметрам ситуации и даже иметь разную степень негативности.

В отношении *да ну* мы предлагаем различать:

- девалоризацию предложения и отказ выполнить предлагаемое;
- девалоризацию решения, о котором сообщает собеседник (сопровождается отрицательной оценкой самого собеседника). Ср. *Да ну тебя!*;
- девалоризацию самого речевого акта высказывания собеседника;
- девалоризацию негативного суждения<sup>6</sup>.

Эти четыре типа употреблений *да ну!* мы рассмотрим в 5.1–5.4.

### 5.1. Отказы

Формула *да ну* как ответ на предложения, приглашения или советы не может выражать удивление — или только лишь удивление: естественным ответом на стимулы такого рода может быть либо положительная, либо отрицательная реакция: или согласие, или отказ. Для исходной семантики *да ну* (например, в свете семантики удивления-неверия, которая относится к группе отрицательных значений) более естественным выглядит развитие семантики отказа.

Предложение собеседника может быть прямое (10) или косвенное (19). Ответ *да ну!* здесь означает его девалоризацию и нежелание выполнить предлагаемое:

(19) [Сергей] *С днем рождения тебя!*

[Люда] *Спасибо, Сережка! А почему ты не захотел прийти вечером, когда соберутся все гости?*

[Сергей] *Да ну! Придут одни твой родственники. О чем мы с ними будем говорить?* [Карен Шахназаров и др. Исчезнувшая империя, к/ф (2008)].

<sup>6</sup> В просторечном и диалектном дискурсе можно найти и другие употребления *да ну* — например, в качестве отрицательного ответа на вопрос, как в (А).

(А) Интервьюер: *А за хмелем с мешками ходили?*

Респондент, 1940 г.р.: *Да ну, с мешками! Да сумочки возьмем да в сум... сколько там его надо. А он же легкий.* [Корпус говора села Роговатое].

Частая разновидность этого случая — когда в первой части обсуждается лицо или объект как некий, возможно, подходящий вариант для некоторой ситуации Р (эту часть можно рассматривать как вариант предложения разделить оценку обсуждаемого лица), а во второй он девалоризируется: оценивается говорящим как неприемлемый и неудачный (20):

- (20) [Даша, жен, 19] *Ой, все, разонравился мне Паша.*  
[Женя, жен, 19] *Как? Классный ведь парень!*  
[Даша, жен, 19] *Да ну! Зануда он!*  
[Разговор знакомых // Из материалов Ульяновского университета, 2007].

## 5.2. Девалоризация собеседника

Такого типа реплики — с отрицательной оценкой собеседника — засвидетельствованы еще в 60-е гг. XIX в. (см.(21))и до сих пор остаются достаточно частотными в разговорной речи.

- (21) *Пожалуйте, сударыня, Татьяна Марковна, ручку! Он схватил старушку за руку, из которой выскочил и покатился по полу серебряный рубль, пригтовленный бабушкой, чтоб послать к Ватрухину за мадерой. — Да ну, бог с тобой, какой ты беспокойный: сидел бы смирно! — с досадой сказала бабушка.* [И. А. Гончаров. Обрыв (1869)].

Наиболее характерным прагматическим контекстом для них, по нашим данным, является «трехчастный» (то есть состоящий из трех связанных между собой реплик) диалог, в котором первая реплика — это просьба или поручение говорящего, а вторая принадлежит его собеседнику и дает прямое или косвенное свидетельство его некооперативного поведения. Оно и служит наиболее частотной и естественной причиной раздражения или разочарования говорящего — в ситуации в целом и собеседнике в частности, ср:

[Говорящий предлагает/просит] — [отказ собеседника] — [говорящий недоволен, раздражен]

[Говорящий дал поручение] — [собеседник не выполнил] — [говорящий разочарован]

- (22) А: *Кисет-то мой принесла?*  
Б: *Ой, забыла, миленький Федот Евграфыч, забыла.*  
А: *Да ну! Да ладно уж! Махорка имеется. Спасибо сидр мой не забыли.*  
[Станислав Ростоцкий, Борис Васильев. ...А зори здесь тихие, к/ф (1972)].

С еще большей очевидностью оценка собеседника присутствует в варианте этой дискурсивной формулы, содержащей местоимение: *Да ну тебя/вас/его/ее/их!*, который семантически тождествен именно такому девалотивному употреблению *Да ну!* В принципе во всех случаях местоимение можно добавить или убрать без изменения смысла — это обстоятельство мы учитывали при поиске переводных



эквивалентов, — а их непросто найти, во всяком случае в английском. Заметим, что степень недовольства собеседником и раздражения против него недостаточна для того, чтобы заменить *да ну!* на прямую агрессию и ругань, адресованную непосредственно участнику диалога. Создается впечатление, что говорящий все-таки не перекладывает всю вину за провал своего плана на собеседника, а частично списывает ее за счет неудачно сложившихся обстоятельств. Поэтому даже в переводах на английский, в котором нет подходящей формулы, встречающиеся в качестве эквивалентов маркеры агрессии не имеют конкретного адресата:

- (23) *No, sah, I doan' want no sichdoin's. "Blame it, can't you TRY? I only WANT you to try — you needn't keep it up if it don't work."* [Mark Twain. The Adventures of Huckleberry Finn (1884)].

*Нет, сэр, ничего этого я не желаю. — Да ну тебя, неужели хоть попробовать не можешь? Ты только попробуй, а не выйдет, возьмешь и бросишь.* [Марк Твен. Приключения Гекльберри Финна (Н. Дарузес, 1950)].

- (24) *"You can't stay here. Why, with nobody to take care of you, you'd starve." Grampa cried, "Goddamn it, I'maol' man, but I can still take care a myself."* [John Steinbeck. The Grapes of Wrath (1939)].

*Как ты будешь жить? Кто о тебе позаботится? Ведь с голоду умрешь! Дед закричал: — Да ну вас всех! Я хоть и старик, а сумею сам о себе позаботиться.* [Джон Стейнбек. Гроздья гнева (Н. Волжина, 1940)]

Другой характерный вариант переводного эквивалента для *да ну*-девальоризации с акцентом на собеседнике (или *да ну PRON<sub>ACC</sub>*) — междометия, как если бы говорящему от возмущения и разочарования просто нечего было сказать:

- (25) *"I should have thought." Mr. Lorry began. "Pooh! You'd have thought!" said Miss Pross <...>* [Charles Dickens. The Tale of two Cities (1859)].

*Мне кажется... — начал было мистер Лорри. — Да ну вас, что это вам еще кажется! — перебила его мисс Просс<...>* [Чарлз Диккенс. Повесть о двух городах (С. Я. Бобров, М. П. Богословская, 1950–1960)].

Зато в самом русском есть квазисинонимы к этому типу девальоризации, которые заслуживают внимания, ср. *Иди к черту!* Механизм девальоризации, который они реализуют, состоит в исключении собеседника из так называемой «личной сферы говорящего» [Апресян 1986] — это именно тот тип манипуляции («наказания» собеседника), который в ответ на некооперативность маркируется *Да ну* (синонимичное *Да ну PRON<sub>ACC</sub>*).

Замечательно, что он иконично маркируется в русской жестикуляции и мимике: помимо оценочного «нахмурить брови» и движения взгляда от собеседника (по Гришиной, это маркеры отрицания и дистанцирования), ему соответствуют жесты «отвернуться от собеседника» и «махнуть рукой», который выражает досаду и одновременно служит эмблемой прощания и расставания (специалисты относят его к так называемой Away Gestures Family — см. [Bressem, Müller 2014]).

### 5.3. Девавторизация речевого акта

Этот класс употреблений *да ну!* реагирует не на суть сказанного собеседником — он претендует на то, чтобы дезавуировать сам речевой акт, указав на его неуместность. Понятно, что не любой речевой акт можно отменить, — самые естественные контексты для такого *да ну* — это извинения и комплименты, ср.:

- (26) А: *Извини, Алеш, он, видимо, очень устал.*  
В: *Да ну, Наташа, о чем ты говоришь.*  
[Юлий Райзман, Анатолий Гребнев. Частная жизнь, к/ф (1982)].
- (27) [Буфетчица] *А похорошела ты. Да-да, похорошела!*  
[Анна Веленстович] *Ой, да ну!*  
[Леонид Быков и др. Аты-баты, шли солдаты, к/ф (1976)].

В данном случае *Да ну!* говорит следующее: собеседник ошибся в выборе иллокуции — нет настоящей причины, чтобы извиняться, комплименты по данному поводу говорящему не нужны, он только испытывает неловкость.

В определенном смысле этот класс семантически противоположен предыдущему: поскольку в случае извинения и комплимента говорящий бенефициант, он не только не испытывает к собеседнику отрицательных эмоций, как в (26)–(28), но может испытывать и положительные, — например что-то вроде смущения. Поэтому если говорить о мимике и жестикуляции, то именно в этих контекстах, прежде всего, конечно, в ответ на комплимент, случается улыбка и смущение («глаза вниз») — помимо «отвернуться» и «нахмурить брови» или маркеров отрицания для извинения: «махнуть рукой», «склонить голову набок», «пожать плечами».

Обратим внимание на еще один более редкий, но тоже характерный случай девалторизации, близкий к только что рассмотренным. Это девалторизация излишней вежливости собеседника. Отмена формальностей в обращении (ср. разрешение обращаться к себе не по званию или должности, как *ваши благородие*, а по имени-отчеству: *Григорий Петрович*; или использовать краткую форму имени: не *Григорий*, а *Гриша*) может производиться говорящим с помощью *Да ну*:

- (28) [Рита] *Товарищ полковник.*  
[Курбатов] *Да ну... Тимур.*  
[Рита] *Тимур Курбатович.*  
[Курбатов] *Тимур.*  
[Валерий Тодоровский, Геннадий Островский. Мой сводный брат Франкенштейн, к/ф (2004)]

### 5.4. Девавторизация оценочного суждения

В этот класс употреблений попадают контексты, в которых собеседник дает негативную оценку ситуации Р. Эта оценка может быть связана с его страхом, предубеждениями, пессимистическими ожиданиями и подобным. Говорящий хочет

убедить собеседника ее снять, внушая ему оптимизм и говоря что-то вроде: ‘это не так страшно, не так опасно, не думай, не сосредотачивайся на этом неприятном Р’ и подобное. Ср. (29)–(30):

(29) А: *Организуем клумбу с самыми красивыми цветами!*

В: *Ассигнований нет!*

А: *Да ну, я попрошу шоферов, они мне принесут. Из Крыма роз, с Карпат эдельвейсов.*

[Алексей Мишурин и др. Королева бензоколонки, к/ф (1963)]

(30) — *Юшка, как твоя нога? — Да ну, фигня, — отмахнулась она. — Думали — растяжение, оказалось — ушиб. Заживет как на собаке.* [Дина Сабитова. Где нет зимы (2011)]

Круг русских квазисинонимов<sup>7</sup> (типа *Брось! Забудь!* и подобных) выделяет этот тип *да ну!* из прочих употреблений и связывает его с полем каузации прекращения ситуации (более точно — continuous prohibitive, см. [Рахилина 2013]). Такое *Да ну!* может маркироваться даже интонационно: наряду с обычным для других типов дева-лоризации падающим тоном оно может сопровождаться высоким, который как бы специально акцентирует незначимость предыдущего высказывания. Что касается жестикуляции, то этот тип *Да ну!* маркируется целым кластером жестов, отражающих отрицательную оценку («нахмурить брови»), собственно дева-лоризацию («махнуть рукой»), несогласие и дистанцирование («склонить голову», «поднять плечи»).

Теперь рассмотрим жестовые параметры *Да ну!* в разных его вариантах подробнее.

## 6. Жестикуляция и статистика<sup>8</sup>

### 4.1. Жестовые параметры и их соотношение

Как мы помним, из 190 видеозаписей с *Да ну!* в МУРКО только в 78 фрагментах жесты говорящего были достаточно хорошо видны во время произнесения формулы. Такие фрагменты размечались по семантическим классам *Да ну!* и по шести жестовым параметрам — ручные и головные жесты, движение корпуса, глазное поведение, улыбка или смех и движение бровей.

После этого к выборке были последовательно применены алгоритмы Random forest, CART (Classification and Regression Tree) и Multiple Correspondence Analysis.

Мы использовали алгоритм Random forest, чтобы ранжировать жестовые параметры по тому, насколько каждый из них влияет на выбор значения зависимой переменной, — в нашем случае — семантического класса формулы. Полученный результат представлен на рисунке 2. Наибольшая зависимость от семантического

<sup>7</sup> Хороших типологических коррелятов по параллельным корпусам для этого типа пока не найдено.

<sup>8</sup> Мы благодарим Г.А. Мороза за консультации по анализу данных.



Рис. 2. Ранжирование жестовых параметров

класса обнаружилась у направления взгляда говорящего. Следом за ним идет движение бровей, далее — уже значительно менее существенные головные и ручные жесты. Движение корпуса и улыбка, согласно этим данным, с семантикой формулы связаны минимально.

Применение к нашему материалу еще одного алгоритма — CART дает возможность выявить существенную корреляцию значений жестовых параметров и выделенных значений *Да ну!*, так что с его помощью мы сможем для каждого параметра (направление взгляда, движение корпуса, рук и проч.) проследить, какие группы употреблений *Да ну!* им маркируются.

### 6.1. Взгляд

При разметке движения глаз мы пользовались системой из [Гришина 2017: 491]. В этой системе противопоставлено направление взгляда внутри зоны коммуникации (ЗК) и вне ее, так что в рамках одного жеста описываются исходное положение взгляда и итоговое. В нашей разметке добавилась новая разновидность — направления взгляда вне ЗК: взгляд вниз.

На основании взаимодействия двух параметров — направления взгляда и семантики формулы — было построено дерево классификации данных (рис. 3).

Как видно из левой ветви графа (node 2), для реакций на извинение, отказов и девалоризации оценочного суждения нет разметки взгляда. Она отсутствует по чисто техническим причинам: в этих клипах взгляд не виден.

Семантический класс реакции на комплимент прекрасно выделяется жестами «Внутри ЗК → Опустить глаза» и «Взгляд вниз → Взгляд вниз» (node 4). Это довольно естественно, потому что для таких ситуаций характерен жест «потупить», который выражает смущение.

Оставшиеся данные делятся следующим интересным образом. *Да ну* в значении девалоризации провинившегося собеседника сопровождается либо выходом взгляда

a (apology) — реакция на извинение  
 c (compliment) — реакция на комплимент  
 da (disappointment) — досада на собеседника  
 db (disbelief) — неверие  
 e (encouragement) — девалоризация негативной оценки Р

ii..... Внутри ЗК → Внутри ЗК  
 io..... Внутри ЗК → Выход  
 oi..... Вне ЗК → Вход  
 oo..... Вне ЗК → Вне ЗК

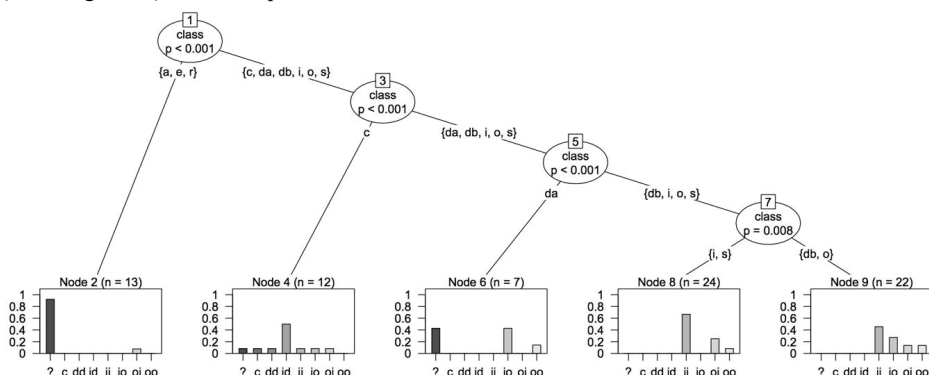


Рис. 3. Распределение значений параметра «взгляд» по группам употреблений *Да ну!*

из зоны коммуникации, либо пребыванием взгляда вне ее от начала и до конца реплики (node 6). Для удивления и производных от него иронических употреблений<sup>9</sup>, напротив, свойственно в первую очередь постоянное нахождение взгляда внутри ЗК либо также движение глаз из положения вне ЗК в сторону собеседника (node 8). В то же время объединенные в одну группу неверие и отказ разделить энтузиазм по поводу варианта X (см. 5.1) демонстрируют свойства как первого, так и второго: «Внутри ЗК → Внутри ЗК», «Внутри ЗК → Выход», «Вне ЗК → Вход». По крайней мере, для класса неверия это вполне объяснимый результат, поскольку семантически оно как раз и представляет собой сочетание удивления и девалоризации.

Заметим, что, по Е. А. Гришиной, стационарное положение взгляда внутри ЗК свойственно репликам подтверждения, а любой переход (в зону коммуникации или из нее наружу) — негации. Наша выборка дает несколько другую картину. Негативное высказывание — удивление — по-прежнему характеризуется комбинацией «Внутри ЗК → Внутри ЗК», но достаточно часто маркируется и переходом взгляда в зону коммуникации. Отрицательное — девалоризация — сочетается и с выходом взгляда из ЗК, и с фиксированным взглядом в сторону, но не с переводом глаз в ЗК.

Таким образом, более значимым фактором для отрицательности *Да ну!* является финальное направление взгляда, а не наличие или отсутствие движения.

Отрицательная полярность — не единственный признак, который маркируется взглядом. Жест «потупить глаза» характеризует группу употреблений с девалоризацией комплимента, отделяя ее среди других девалоризаций, — в том числе девалоризации извинения (табл. 1).

<sup>9</sup> Интересно, что по смыслу ирония совпадает с неверием, но маркируется так же, как удивление, которое эта ирония имитирует.

Направление взгляда при разных употреблениях *Да ну!*

Удивление (в том числе ироническое)	(1) Внутри ЗК → Внутри ЗК (2) Вне ЗК → Вход
Реакция на комплимент	Потупить глаза
Девалоризация (кроме комплимента), неверие	(1) Внутри ЗК → Выход (2) Вне ЗК → Вне ЗК

### 6.3. Движение бровей

Два лицевых жеста, встретившихся в выборке, — «нахмурить брови» и «поднять брови» — противопоставляют классы значений, содержащие компонент девалоризации, и удивление — единственный класс, который не содержит идеи отрицания. Как видно на рисунке 4, при удивлении брови часто подняты, но никогда не нахмурены; при девалоризации — бывают нахмурены, а подняты редко. Жест «поднять брови» в [Гришина 2017] отнесен к жестам дистанцирования, а «нахмуриться» — к эмфатическим [Гришина 2017: 457]. Именно эмфатические жесты коррелируют с контекстами, в которых присутствует отрицательная оценка (Гришина 2017: 458).

Мимика, таким образом, дублирует противопоставление, маркируемое интонационно, отделяя отрицательные употребления *Да ну!* от нейтральных.

a (apology) — реакция на извинение  
c (compliment) — реакция на комплимент  
da (disappointment) — досада на собеседника  
db (disbelief) — неверие  
e (encouragement) — девалоризация негативной оценки P  
i (irony) — ирония (из удивления)  
o (object) — девалоризация предложенного варианта  
r (refusal) — отказ на предложение  
s (surprise) — удивление

bigeyes..... жест «расширить глаза»  
browsdown... жест «нахмурить брови»  
browsup..... жест «поднять брови»

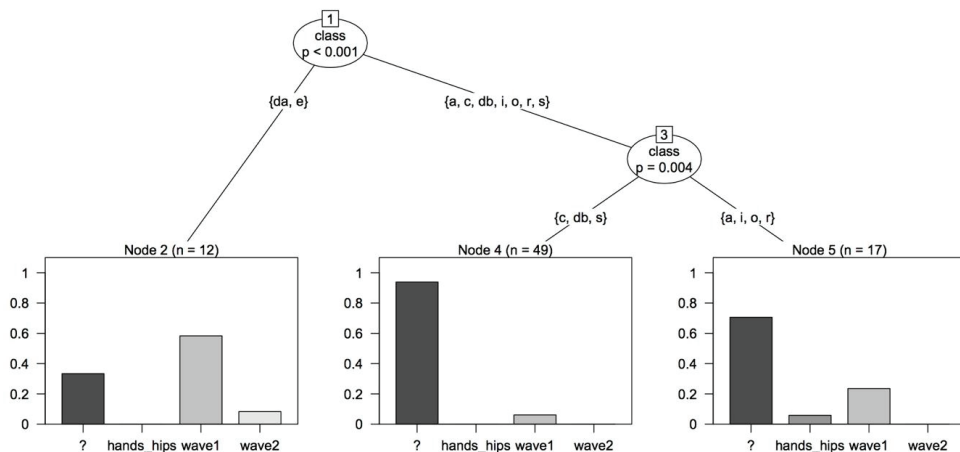


Рис. 4. Распределение параметра «движение бровей» по группам употреблений *Да ну!*

## 6.4. Голова

Классы реакции на комплимент, девалоризации собеседника и отказа на предложение маркируются специальным головным жестом (рис. 5, node 2), в котором говорящий отворачивает голову в сторону и задерживает там на некоторое время, прежде чем повернуться обратно к собеседнику<sup>10</sup>. Жест «отвернуться от собеседника» упоминается в [Гришина 2017] в одном ряду с жестом «закрыть глаза», который рассматривается там как способ блокировать поступающую информацию [Гришина 2017: 510]. Такое толкование жеста подходит и для *Да ну!*, особенно в ответе на комплимент, потому что девалоризация речевого акта подразумевает, что говорящий не собирается слушать то, что начал говорить собеседник.

Жест «склонить голову набок» характеризует другую группу значений — классы реакции на извинение, неверия, иронии и девалоризации предложенного варианта (node 5). Он классифицируется как жест дистанцирования [Гришина 2017: 457]; жесты этого типа часто сопровождают контексты несогласия [Гришина 2017: 458].

Как оказалось, для классов удивления и девалоризации негативного суждения характерно отсутствие головных жестов (node 4) — должно быть, потому, что это единственные два типа употребления *Да ну!*, обладающие положительной тональностью.

a (apology) — реакция на извинение  
c (compliment) — реакция на комплимент  
da (disappointment) — досада на собеседника  
db (disbelief) — неверие  
e (encouragement) — девалоризация негативной оценки P  
i (irony) — ирония (из удивления)  
o (object) — девалоризация предложенного варианта  
r (refusal) — отказ на предложение  
s (surprise) — удивление

KG..... жест «покачать головой»  
away..... жест «отвернуться»  
back..... жест «качнуть головой назад»  
down..... жест «опустить голову»  
nod..... жест «кивнуть»  
none..... отсутствие головного жеста  
tilt..... жест «склонить голову набок»  
to\_A..... жест «повернуть голову к собеседнику»  
turn..... жест «повести головой в сторону»

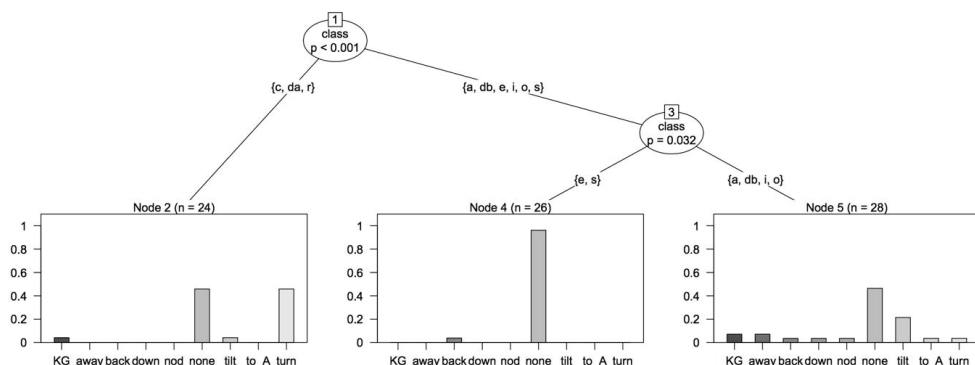


Рис. 5. Распределение параметра «движение головы» по группам употреблений *Да ну!*

<sup>10</sup> Иллюстрацию дает клип Мультимедийного корпуса, соответствующий следующему примеру: (А) [Пашка] *Попóзже-то выйдеишь?* [Катя] *А для чего?* [Пашка] *Да ну́, от́вернуться от собеседника Кáтя! А я́ откúда знáю? Тоскливо́ же одной-то!* [Василий Шукшин. Живет такой парень, к/ф (1964)].



В результате жесты головы разбивают кластер значений девалоризации на три части.

К первой относятся те, для которых более релевантно желание говорящего прекратить обсуждение вопроса, — а именно реакция на комплимент, девалоризация собеседника и отказ. Это проявляется в виде жеста «отвернуться от собеседника».

Ко второй относятся те, в которых более существен компонент несогласия. К ним относятся неверие, ирония, девалоризация предложенного варианта и извинение. Они маркируются жестом «склонить голову набок».

К третьей относится «мягкая» девалоризация — девалоризация негативного суждения, которая в данном случае примыкает к удивлению и не маркируется головными жестами.

## 6.5. Руки

Единственным ручным жестом, который присутствовал в выборке, был жест «махнуть рукой»<sup>11</sup>. Классы значений *Да ну!* разделились на три группы по тому, насколько этот жест им свойствен (рис. 6). Больше всего он встречался при

a (apology) — реакция на извинение  
 c (compliment) — реакция на комплимент  
 da (disappointment) — досада на собеседника  
 db (disbelief) — неверие  
 e (encouragement) — девалоризация негативной оценки Р  
 i (irony) — ирония (из удивления)  
 o (object) — девалоризация предложенного варианта  
 r (refusal) — отказ на предложение  
 s (surprise) — удивление

hands\_hips..... жест «руки в боки»  
 wave1..... жест «махнуть рукой»  
 wave2..... жест «взмахнуть двумя руками»

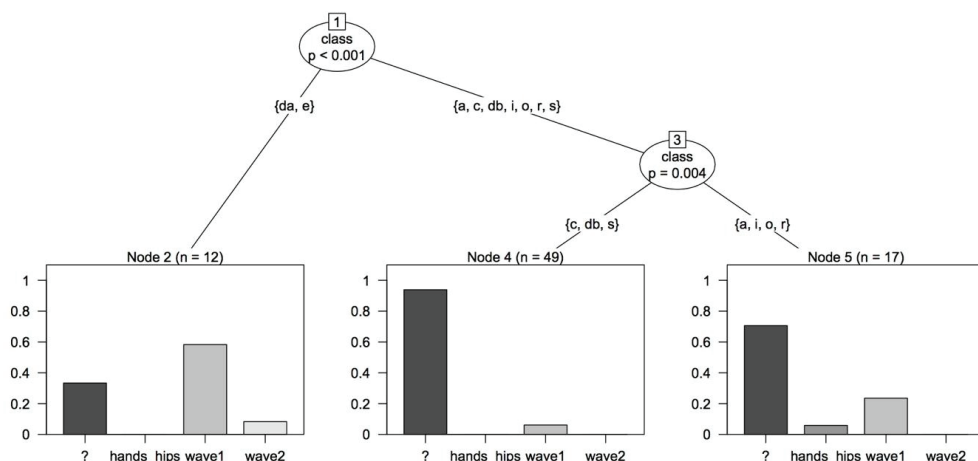


Рис. 6. Распределение параметра «ручные жесты» по группам употреблений *Да ну!*

<sup>11</sup> Один раз, в контексте, где формула была употреблена в значении девалоризации негативного суждения, встретила его разновидность, где то же движение было выполнено двумя руками.

девальоризации собеседника или негативного оценочного суждения (node 2). Менее характерен он оказался для ответа на извинение, девальоризации предложенного варианта или отказа (node 5) и, наконец, практически не появлялся при удивлении, неверии или реакции на комплимент (node 4) (табл. 2).

Таблица 2

**Движение руки при разных употреблениях *Да ну!***

Девальоризация собеседника; девальоризация негативного оценочного суждения	«Махнуть рукой»
Отказ, девальоризация варианта, ответ на извинение	«Махнуть рукой»
Удивление, неверие, реакция на комплимент	«Махнуть рукой»

Если считать, что жест «махнуть рукой» действительно родствен прощальному жесту и имеет семантику ‘неважно’, ‘бог с ним’, получившаяся градация приобретает смысл: с одной стороны оказываются классы «чистой» девальоризации, а с другой — удивление, которое исключает идею незначительности стимула. Именно поэтому классы удивления и недоверия в данном случае не противопоставлены, как в интонации, а демонстрируют сходное поведение.

## 6.6. Корпус

У движения корпуса также обнаружилась некоторая зависимость от семантики *Да ну!* (рис. 7). Для значения удивления характерны жесты «корпус вперед» и «корпус назад» (node 5), а классы реакции на извинение и неверия объединяет сочетаемость с жестом «пожать плечами» (node 4).

## 6.7. Multiple Correspondence Analysis

Применение алгоритма МСА позволило нам визуально оценить близость выделенных нами групп употреблений *Да ну!* на основании свойственных им жестовых параметров. На рисунке 8 видны три основные семантические зоны — удивление, неверие и класс девальоризаций, которые, если рассматривать соответствующую им жестикуляцию статистически, оказываются практически тождественны. Классу девальоризаций максимально противопоставлено удивление: оно отстоит от него пространственно, с пренебрежимо малой долей пересечения. Исключение составляет девальоризация негативного оценочного суждения (см. 5.4), которая выделяется из гомогенного класса других девальоризаций. Действительно, на представленном рисунке она существенно пересекается с удивлением, демонстрируя, что в таких употреблениях одновременно с отрицательным компонентом содержится некоторая позитивная задача. Тем временем значения жестовых переменных, которые коррелируют с классом неверия, не сосредоточены в одном кластере, а распространяются достаточно широко; так, он не только семантически, но и жестово объединяет удивление и девальоризацию.

a (apology) — реакция на извинение  
 c (compliment) — реакция на комплимент  
 da (disappointment) — досада на собеседника  
 db (disbelief) — неверие  
 e (encouragement) — девалоризация негативной оценки Р  
 i (irony) — ирония (из удивления)  
 o (object) — девалоризация предложенного варианта  
 r (refusal) — отказ на предложение  
 s (surprise) — удивление

back..... жест «отклонить корпус назад»  
 away..... жест «повести корпус вперед»  
 none.....отсутствие движения корпусом  
 shrug..... жест «пожать плечами»  
 slouch..... жест «сгорбиться»

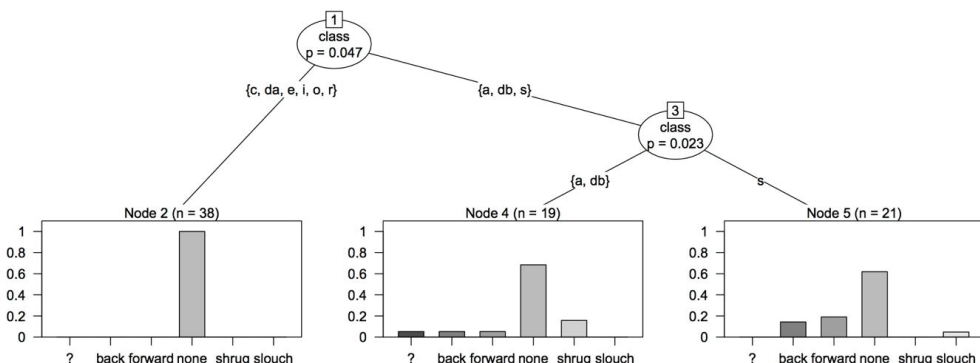


Рис. 7. Распределение параметра «движение корпуса» по группам употреблений *Да ну*

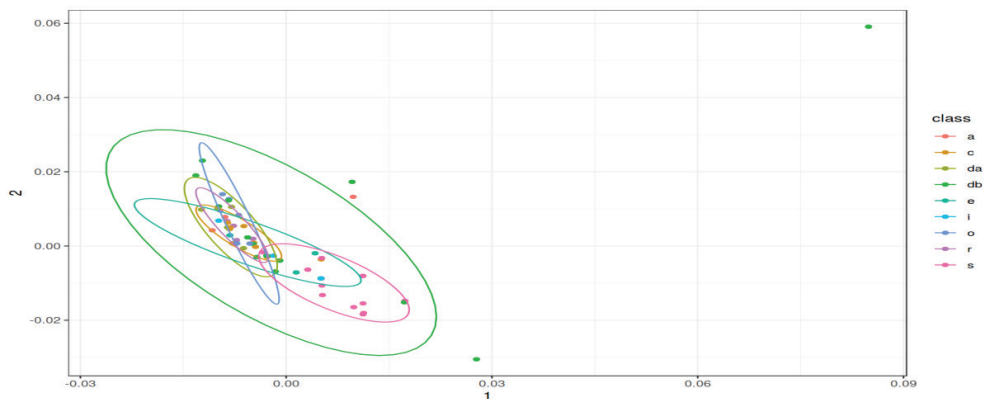


Рис. 8. Соотношение групп употреблений *Да ну!* с точки зрения жестового маркирования

## 7. Заключение

Настоящее исследование мы считаем пилотным для построения специализированной базы данных по дискурсивным формулам как особому типу конструкций в рамках проекта «Русский конструктик». Пример *Да ну!* демонстрирует специфику этих единиц и выявляет информацию, необходимую для их полного описания.

В нашей базе несколько полей. Проведенная работа позволяет оправдать их присутствие и заполнить их на материале *Да ну!*

1. Как языковая единица, или вход в базу, дискурсивная формула отличается от обычной конструкции и просто однословной предикатной единицы тем, что не содержит переменных — и в этом отношении кажется слишком простой.

На самом деле, как видно из только что представленного описания, важнейшей переменной для них является *тип предшествующего контекста*, который стимулирует формулу как ответную реплику говорящего: разные дискурсивные формулы реагируют на разные типы иллокуций, и они же структурируют значения многозначных формул. Действительно, если на вход *Да ну!* в значении удивления поступает утверждение, содержащее новую информацию, то на вход девалоризационного кластера, в зависимости от подтипа, поступают просьбы, приглашения, комплименты, извинения, опасения и проч. Мена стимула(который можно считать лингвистическим аналогом философского понятия речевого акта<sup>12</sup>) приводит к мене значения формулы.

2. Как мы видели на примере *Да ну!*, частотные формулы могут быть многозначны. У *Да ну!* мы выделили пять значений: удивление-доверие, удивление-неверие, девалоризация с оценкой собеседника, девалоризация речевого акта и девалоризация оценочного суждения. Можно ожидать, что полученные при описании *Да ну!* классы и подклассы употреблений (и соответствующие им *семантические пометы*) будут востребованы при дальнейшей классификации всего списка формул<sup>13</sup>.

3. Значения, представленные в кластере *Да ну!*, имеют *квазисинонимы*, — в том числе среди дискурсивных формул, ср., например, *Да ладно!* как маркер удивления-неверия.

4. *Типологические корреляты* дискурсивных формул расширяют наше представление и об источниках конструкционализации, и о дробности противопоставлений внутри реплик-реакций определенного типа. Материал английского показывает, что для некоторых значений прямых аналогов дискурсивных формул (то есть изолированных однословных реплик-реакций в ответ на те же стимулы) в каких-то языках может и не быть. В таком случае материал параллельных корпусов помогает найти их лексические или грамматические корреляты, расширяя наше представление о лексико-грамматической и семантико-синтаксической интеграции в языке.

5. Мы показали, что все употребления *Да ну!* восходят к композициональной структуре, «собранной» из двух частиц (*да* и *ну*) и имеющей семантику интенсификации ускорения. По нашему мнению, источники конструкционализации дискурсивных формул не бывают случайными, а соответствующие им процессы, приводящие к потере композициональности, должны быть регулярны, — поэтому они чрезвычайно интересны с теоретической и типологической точек зрения. Следовательно, *исходная для формулы структура* тоже должна отражаться в базе.

<sup>12</sup> Уделить достаточное внимание проблематике лингвистического и философского определения речевого акта, — и в том числе обзору работ на эту тему — в данной работе не представляется возможным. Этим вопросам будет посвящена отдельная готовящаяся нами публикация.

<sup>13</sup> Процедура составления списка описана в [Пужаева и др. 2018].

6. Важнейший для этой работы сюжет, выделяющий дискурсивные формулы среди прочих конструкций, — это особый способ маркирования семантических противопоставлений для формул. Поскольку формулы дискурсивны, то, как мы видели на примере *да ну*, они свойственны прежде всего разговорному языку и широко используют его преимущества: интонацию и жестикуляцию.

То, что в русском языке **интонация** служит для разрешения полисемии, хорошо известно — ср. пример с *вообще* в [Апресян 2004: 19]<sup>14</sup>, а также [Кодзасов 1993]. Наше исследование дает этому новые подтверждения: некоторые значения (группы значений) *Да ну!* отчетливо противопоставлены интонационно: ‘удивление’ (НЛ\*) — ‘девальоризация’ (L\*) — ‘девальоризация отрицательного суждения’ (Н\*/L\*).

7. Если интонационная тематика в целом достаточно традиционна, то обсуждение в этом контексте **жестикуляции**, в особенности в этом контексте, — во многом достижение Е. А. Гришиной. Она одной из первых, по крайней мере для русского языка, показала, что жестикуляция релевантна для этой задачи, предоставила значительный языковой материал, который нас в этом убеждает, и создала метаязык (язык разметки МУРКО), который позволяет этот материал структурировать, — в частности в нашей базе данных. Дробная разметка в базе фиксирует отдельно движение глаз, махи руки, мимику, движение корпуса и проч.

8. Все сказанное суммирует таблица 3, представляющая предварительные результаты нашего анализа *Да ну!*

Центральной в этой таблице является информация о жестикуляции. Видно, что жестикуляция не дублирует интонацию — у каждого из этих несегментных средств, как показывает таблица, своя функция. Мелодический контур набрасывает так сказать «крупную сетку» на значения *Да ну!* Он противопоставляет не-отрицательные (удивление) и отрицательные (девальоризация и неверие) реакции говорящего, особым образом маркируя своего рода промежуточный случай (девальоризацию отрицательного суждения). Жестикуляция эту классификацию подтверждает — и детализирует: таблица показывает, что значения разных параметров собираются в кластеры, которые маркируют несколько дополнительных значений: неверие, девальоризацию собеседника, ситуацию извинения и комплимента. Таким образом, жестовый материал горизонтальных строк независимо поддерживает семантические оппозиции, и это важный для нас вывод.

В свою очередь, столбцы таблицы 3 тоже подтверждают наши представления о структуре поля *Да ну!* Они свидетельствуют о непрерывности его семантического пространства. Действительно, каждый жестовый параметр как бы заполняет семантическую карту этого пространства, и в этом отношении они ведут себя так же, как ведет себя другой язык — английский, испанский, таджикский и проч., который по-своему структурирует определенную лексико-типологическую зону.

<sup>14</sup> Пример Ю. Д. Апресяна показывает, как фразовая интонация противопоставляет значения слова *вообще* в предложениях *Разжигать костры он ↓вообще запрещал* и *Разжигать костры он вообще ↓ запрещал*, [но <...>].

## Да ну! как многозначная дискурсивная формула

Входной речевой акт	Значение Да ну	Интонация	Жесты					Квазисинонимы	Переводы
			Взгляд	Мимика	Голова	Рука	Корпус		
новость	удивление							Неужели? Ну и ну! Надо же! Да ты что! Не может быть!	Oh, my! Oh, really? No way! Does it? You don't say? + tag questions
новость	ирония							Неужели? Надо же! Да ты что! Да ладно!	tag questions
извинение	девальоризация речевого акта							Да ладно! Что ты!	well, now!
новость	неверие							Да ладно! Не может быть! Никогда не поверю!	No! Really? Come on! Please, ...
комплимент	девальоризация речевого акта							Да ладно (тебе)!	—
предложение	отказ + девальоризация собеседника							—	Notreally?
невыполненная просьба	девальоризация собеседника							Да ну тебя! Иди к черту! Эх ты!	—
негативное оценочное суждение	девальоризация суждения							Чего ты! Да ладно (тебе)!	Come

## Литература

Апресян Ю. Д. Дейксис в лексике и грамматике и наивная модель мира // Семантика и информатика. 1986. Вып. 28. С. 5–33.

Апресян Ю. Д., Богуславская О. Ю., Левонтина И. Б., Урысон Е. В., Гловинская М. Я., Крылова Т. В. Новый объяснительный словарь синонимов русского языка. Первый выпуск. 2-е изд. М., 1999.

Гришина Е. А. 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., 2009. С. 80–90.

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

Гришина Е. А. Русская жестикуляция с лингвистической точки зрения (корпусные исследования). М.: Языки русской культуры, 2017. С. 217.

Дурягин П. В., Рахилина Е. В. Просодические средства маркирования дискурсивной формулы *да ну* // Тезисы VI Международной научной конференции «Культура русской речи» (21–23 февраля 2019 года) URL: <https://drive.google.com/file/d/1eXDZzrSpHqhYU3fPqVv0KMSPGnKkTXuM/view> (дата обращения: 06.07.2019).

Касаткин Л. Л. Русская интонация: тональные контуры // Проблемы фонетики. Вып. V. / Отв. ред. Р. Ф. Касаткина. М., 2007.

Князев С. В., Пожарицкая С. К. Современный русский литературный язык: Фонетика, орфоэпия, графика и орфография. М.: Академический Проект; Гаудеамус, 2012.

Кодзасов С. В. Интонация предложений с дискурсивными словами // Путеводитель по дискурсивным словам русского языка. М., 1993.

Национальный корпус русского языка. URL: <http://www.ruscorpora.ru/>.

Пуужаева С. Ю., Герасименко Е. А., Захарова Е. С., Рахилина Е. В. Автоматическое извлечение дискурсивных формул из текстов на русском языке // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2018. № 2.

Рахилина Е. В. Кондуктор, нажми на тормоза... // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая — 2 июня 2013 г.): в 2 т. Т. 1: Основная программа конференции. Вып. 12 (19). М., 2013. С. 665–673.

Рахилина Е. В., Кузнецова Ю. Л. Грамматика конструкций: теории, сторонники, близкие идеи // Лингвистика конструкций / ред. Е. В. Рахилина. М.: Азбуковник. 2010. С. 18–79.

Рахилина Е. В., Плунгян В. А. Из корпусных наблюдений над лексикой: о семантической эволюции и «лексических маркерах» // *La lettre et l'esprit — entrelangue et culture: Études à la mémoire de Jean Breuillard*. Vol. LXXXIII. Iss. 2–3. P.: Institut d'études slaves, 2012. P. 499–533.

Рахилина Е. В., Резникова Т. И. Фреймовый подход к лексической типологии // Вопросы языкознания. 2013. № 2. С. 3–31.

Шаронов И. А. Коммуникативы как функциональный класс и как объект лексикографического описания // Русистика сегодня. 1996. № 2. С. 89–112.

Шаронов И. А. Коммуникативы и методы их описания // Компьютерная лингвистика и интеллектуальные технологии (по материалам ежегодной Международной конференции «Диалог 2009»). Вып. 8 (15). М., 2009. С. 543–547.

Янко Т. Е. Интонационные стратегии русской речи в сопоставительном контексте. М., 2008.

Aijmer K. *Conversational Routines in English*. Harlow, 1996.

Bressem J. and Müller C. “The family of Away gestures: Negation, refusal, and negative assessment,” in *Body — Language — Communication. An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38.2)*. Berlin/Boston: De Gruyter Mouton, 2014. P. 1592–1604.

Coulmas F. (ed.). *Conversational routine*. The Hague: Mouton, 1981.

Fillmore C. J. Remarks on contrastive pragmatics // *Contrastive linguistics: Prospects and problems*. 1984. P. 101.



Fillmore C.J., Kay P., O'Connor M.C. Regularity and idiomaticity in grammatical constructions: The case of let alone // *Language*. 1988. Sep. 1. pp. 501-538.

Janda L.A., Ljashevskaya O., Neset T., Rakhilina E. & Tyers F.M. A Constructicon for Russian: Filling in the Gaps. B. Lyngfelt, T. T. Torrent, L. Borin & K. H. Ohara (eds.). *Constructicography: Constructicon development across languages*. (forthc.)

Kobozeva I., Ivanova O., Zakharov L. Towards multimodal modelling of verification discourse markers in russian dialog. // *Вестник Московского университета. Серия 9: Филология*, 2019, no. 1. pp. 36–49.

Traugott E.C., Trousdale G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.

**<sup>1</sup>P.A. Bychkova, <sup>1,2</sup>E.V. Rakhilina, <sup>1</sup>E.A. Slepak**

<sup>1</sup>*National Research University Higher School of Economics*

<sup>2</sup>*Vinogradov Institute of Russian Language, Russian Academy of Sciences  
Russia, Moscow*

*polyatomson@gmail.com, rakhilina@gmail.com, janenikel16@gmail.com*

## DISCOURSE FORMULAE, POLYSEMY AND GESTURE MARKING

The paper suggests a discussion of semantic marking for specific multiword colloquial constructions, discourse formulae. This type of linguistic unit is used in dialogical speech as isolated remark in response to the utterance of another member of conversation. The properties of discourse formulae include entire fixedness and non-transparent form on one hand and, on the other hand, the ability to develop various illocutionary meanings. In the paper it is argued that in case of polysemy not only the prosody allows to distinguish the intention of the speaker but also the gesture. The role of gesture for disambiguation of formulae meanings is illustrated by the case of Russian highly polysemantic discourse formula *Da nu!*. The paper covers the classification of contexts of its use which reveals five major meanings of *Da nu!*: expression of surprise, expression of disbelief, devaluation of the interlocutor, devaluation of the interlocutor's apology or compliment and "positive" devaluation of the interlocutor's concern. The process of historical development of the meanings is also briefly highlighted. A quantitative analysis of gesture most commonly accompanying *Da nu!* in each of the five meanings is further conducted, with the data retrieved from Multimodal Russian Corpus and then manually annotated. The analysis shows correlation between the identified meanings and particular gesture sets. The distribution of the gesture might allow to discover adjacency of the meanings of the formula and also provide a deeper insight into the semantics of the gesture involved.

**Keywords:** Construction grammar, discourse formulae, gesture, polysemy, pragmatics

## References

- Aijmer K. *Conversational Routines in English*. Harlow, 1996.
- Apresyan Yu.D. [Deixis in lexicon and grammar and naïve model of the world]. *Semiotika i informatika*. 1986. no. 28, pp. 5–33. (in Russ.)
- Apresyan Yu.D., Boguslavskaya O.Yu., Levontina I.B., Uryson E.V., Glovinskaya M.Ya., Krylova T.V. *Novyi ob'yasnitel'nyi slovar' sinonimov russkogo yazyka* [New Explanatory Dictionary of Russian Synonyms]. Issue 1, 2 ed. Moscow, 1999. (in Russ.)
- Bressem J. and Müller C. “The family of Away gestures: Negation, refusal, and negative assessment”. *Body — Language — Communication. An International Handbook on Multimodality in Human Interaction* (Handbooks of Linguistics and Communication Science 38.2). Berlin/Boston: De Gruyter Mouton, 2014. pp. 1592–1604.
- Coulmas F. (ed.) *Conversational routine*. The Hague: Mouton, 1981.
- Duryagin P.V., Rakhilina E.V. [Prosodic means for marking of discourse formula *da nu*]. *Tezisy VI Mezhdunarodnoi nauchnoi konferentsii “Kul'tura russkoi rechi”* [Abstracts of the 6<sup>th</sup> International scientific conference “Russian linguistic culture” (February 21–23, 2019)]. Available at: <https://drive.google.com/file/d/1eXDZzrSpHqhYU3fPqVv0KMSPGnKkTXuM/view> (access date: 06.07.2019).
- Fillmore C.J. Remarks on contrastive pragmatics // *Contrastive linguistics: Prospects and problems*. 1984. P. 101.
- Fillmore C.J., Kay P., O'Connor M.C. Regularity and idiomaticity in grammatical constructions: The case of let alone // *Language*. 1988. Sep 1. pp. 501–538.
- Grishina E.A. [On the question of word-gesture relationship (vocal gesture O in conversational speech)] // *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* (po materialam ezhegodnoi Mezhdunarodnoi konferentsii “Dialog 2009”) [Computational Linguistics and Intellectual Technologies (Annual International Conference “Dialogue 2009”), no. 8 (15). Moscow, RSGU Publ., 2009. pp. 80–90. (in Russ.)]
- Grishina, E. A. [Vocal gesture A in conversational speech] // *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* (po materialam ezhegodnoi Mezhdunarodnoi konferentsii “Dialog 2010”) [Computational Linguistics and Intellectual Technologies (Annual International Conference “Dialogue 2010”), no. 9 (16). Moscow, RSGU Publ., 2010, pp. 102–112 (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: YaSK Publishing House; Yazyki Slavyanskoi Kul'tury, 2017 (in Russ.)
- Janda L. A., Lyashevskaya O., Nessel T., Rakhilina E. & Tyers F. M. A Constructicon for Russian: Filling in the Gaps. B. Lyngfelt, T. T. Torrent, L. Borin & K. H. Ohara (eds.). *Constructicography: Constructicon development across languages*. (forthc.)
- Kasatkin L. L. [Russian prosody: pitch contours]. *Problemy fonetiki* [Issues in phonetics]. Iss. 5. R. F. Kasatkina (ed.). Moscow, 2007 (in Russ.)

Knyazev S. V., Pozharitskaya S. K. *Sovremennyi russkii literaturnyi yazyk: Fonetika, orfoepiya, grafika i orfografiya* [Modern Russian: Phonetics, orthoepy and orthography]. Moscow: Akademicheskii Proekt Publ.; Gaudeamus, 2012 (in Russ.)

Kobozeva I., Ivanova O., Zakharov L. Towards multimodal modelling of verification discourse markers in russian dialog. *Vestnik Moskovskogo universiteta*. Seriya 9: Filologiya, 2019. no. 1. pp. 36–49.

Kodzasov S. V. [The intonation of the sentences containing discourse words]. *Putevoditel' po diskursivnym slovam russkogo yazyka* [Guide to the discursive words of the Russian language]. Moscow, 1993. pp. 182–204 (in Russ.)

Russian National Corpus. URL: <http://www.ruscorpora.ru/>.

Puzhaeva Svetlana Yu., Gerasimenko Ekaterina A., Zakharova Elena S., Rakhilina Ekaterina V. [Automatic Extraction of Formulaic Expressions from Russian Texts]. *Vestnik NGU. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya*. [Vestnik NSU. Series: Linguistics and Intercultural Communication], 2018, vol. 16, no. 2, pp. 5–18. (in Russ.)

Rakhilina E. V. [Conductor, press the brake...] *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* (po materialam ezhegodnoi Mezhdunarodnoi konferentsii "Dialog 2013") [Computational Linguistics and Intellectual Technologies (Annual International Conference "Dialogue 2013")], no. 12 (19): In 2 vol. Vol. 1. Moscow, RSGU Publ., 2013, pp. 665–673 (in Russ.)

Rakhilina E. V., Kuznetsova Yu. L. [Construction Grammar: Theory, followers, similar ideas]. *Lingvistika konstruktivnaya* [Linguistics of constructions]. Rakhilina E. V. (ed.) Moscow: Azbukovnik Publ., 2010, pp. 18–79. (in Russ.)

Rakhilina E. V., Plungyan V. A. [Some corpus observations of lexicon: semantic evolution and "lexic markers"]. *La lettre et l'esprit — entrelangue et culture: Études à la mémoire de Jean Breuillard*. Vol. LXXXIII. Iss. 2–3. Paris, Institut d'études slaves, 2012, pp. 499–533. (in Russ.)

Rakhilina E. V., Reznikova T. I. [Frame approach to the lexical typology]. *Voprosy Jazykoznanija* [Topics in the study of language], 2013, no. 2, pp. 3–31. (in Russ.)

Sharonov I. A. [Communicatives as a functional class and a subject for lexicographic description]. *Rusistika segodnya*, 1996, no. 2, pp. 89–112. (in Russ.)

Sharonov I. A. [Communicatives and methods of their characterization]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* (po materialam ezhegodnoi Mezhdunarodnoi konferentsii "Dialog 2009") [Computational Linguistics and Intellectual Technologies (Annual International Conference "Dialogue 2009")], no. 8 (15). Moscow, 2009. pp. 543–547. (in Russ.)

Traugott E. C., Trousdale G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.

Yanko T. E. *Intonatsionnye strategii russkoi rechi v sopostavitel'nom kontekste* [International strategies of Russian speech in comparative context]. Moscow, 2008. (in Russ.)

**И. М. Кобозева, О. О. Иванова, Л. М. Захаров**

*МГУ им. Ломоносова*

*(Россия, Москва)*

*kobozeva@list.ru, wiola100@rambler.ru, leonid\_zakharov@mail.ru*

### **К МУЛЬТИМОДАЛЬНОМУ МОДЕЛИРОВАНИЮ ВЕРИФИКАТИВНЫХ ДИСКУРСИВНЫХ МАРКЕРОВ В РУССКОМ ДИАЛОГЕ**

В докладе представлены результаты мультимодального анализа ряда дискурсивных маркеров (ДМ) со значением позитивной верификации — *да, да-да, ну да, конечно, разумеется, естественно*, которые в русском диалоге маркируют истинность пропозиции Р, введенной в рассмотрение в предшествующей реплике собеседника. ДМ данного типа просодически вариативны и могут избыточно сопровождаться синонимичным жестом «кивок». Целью исследования было выявить факторы, которые влияют на паравербальное оформление таких маркеров. Анализ мини-диалогов, содержащих аудио- и видеоряд, из корпуса МУРКО позволил на уровне тенденций установить, какие синтаксические, прагматические и социолингвистические факторы влияют на жестовое сопровождение и просодическое оформление ДМ. Также в результате анализа собранных данных были выявлены маркеры, используемые для выражения обратной связи, имеющие одинаковое с верификативными ДМ вербальное оформление, но отличающиеся по своей функции. Полученные выводы вносят новые данные в систему существующих представлений о дискурсивных словах и сопровождающем их жестовом и просодическом оформлении, способствуют дальнейшему развитию Мультимедийного корпуса русского языка, а также могут найти применение при разработке анимированных компьютерных агентов и роботов.

*Ключевые слова:* мультимодальный анализ речи, диалог, дискурсивные маркеры, верификация, жест, кивок, просодия, семантика, прагматика.

#### **Теоретические предпосылки**

Устный дискурс как первичная форма существования языка в последние годы привлекает все большее количество исследователей. В устном дискурсе выделяется три канала передачи информации: вербальный (сегментные единицы устной речи), просодический (несегментные средства устной речи: интонация, громкость,

темп, тембр и проч.) и визуальный (жесты, мимика, язык тела и т. д.). Пропозициональная информация кодируется в основном вербальными средствами, просодические средства выражают информацию коммуникативно-прагматического характера (например, коммуникативную структуру высказывания, его цель, эмоциональное состояние говорящего), а жесты могут передавать как пропозициональную, так и коммуникативно-прагматическую информацию. Все три типа средств кодирования смысла тесно взаимосвязаны: они дополняют или дублируют смысл, передаваемый по другим каналам, а иногда и противоречат друг другу, как в случае иронии, когда интонация и мимика указывают на то, что говорящий имел в виду противоположное тому, что выразил вербально. При этом в передаче и понимании смысла высказывания важную роль играет контекст, вербальный и экстралингвистический. Все сказанное объясняет повышенное внимание лингвистов, изучающих устный дискурс, к паравербальным средствам коммуникации, в частности к просодическому оформлению речи и к сопровождающей ее жестикуляции.

Обычно исследователи сосредоточены на изучении взаимодействия вербальных средств (в том числе дискурсивных маркеров) либо с просодией (см., например, [Бонно, Кодзасов 1998]), либо с жестикуляцией (см., например, [Cienki, Müller 2008; Richter 2010]). Дискурсивные маркеры (ДМ) в устной речи, в отличие от полноточной лексики и грамматических показателей, нередко представляют собой мультимодальный кластер<sup>1</sup>. Программа описания ДМ с учетом всех трех типов средств, входящих в такие кластеры, была предложена уже в [Кобозева, Захаров 2004], однако опыты ее реализации практически единичны и касаются отдельных несвязанных между собой элементов системы ДМ (см. [Кобозева, Захаров 2007], а также [Гришина 2012]). Мы в данном случае предлагаем унифицированную мультимодальную репрезентацию для анализа семантической группы ДМ, которая позволяет отслеживать корреляции между вербальными, просодическими и жестовыми феноменами, фиксируя при этом факторы, которые предположительно могут влиять на их реализацию. В качестве объекта моделирования были выбраны верификативные дискурсивные маркеры, которые интересны тем, что их смысл в устном диалоге передается либо по слуховому каналу (при помощи произнесения слова с определенной интонацией), либо по зрительному каналу (при помощи кивка), либо по обоим каналам одновременно. В связи с этим возникает вопрос: от чего зависит способ реализации данного смысла? Кроме того, основные маркеры этого типа — *да* и *нет* — представлены рядом просодических вариантов, отражающих прагматические различия между их употреблениями (о семантико-просодической вариативности лексической единицы *да* см., например, [Гришина 2011; Кобозева, Захаров 2012]).

Верификативными дискурсивными маркерами (ВДМ) мы называем дискурсивные слова, которые могут самостоятельно употребляться в качестве верификативного высказывания. Так, *да* и *нет* принадлежат к ВДМ, поскольку они не только могут входить в состав верификативного высказывания, как в (1а) и (2а), но и сами по себе могут составлять такое высказывание, как показывают (1б) и (2б):

---

<sup>1</sup> Термин введен Е. А. Гришиной (см. [Гришина 2011а: 243]).

- (1) Это место свободно? — а. **Да**, свободно. / б. **Да**.  
(2) Петя уже выздоровел. — а. **Нет**, он еще не выздоровел. / б. **Нет**.

В синонимический ряд ВДМ *да* входят *конечно*, *разумеется*, *естественно* и целый ряд других ВДМ. Жестовыми синонимами для ВДМ *да* и *нет* являются соответственно жесты *кивнуть* и *покачать головой II* [Григорьева, Григорьев, Крейдлин 2001]<sup>2</sup>. Данные жесты, как и ВДМ, могут самостоятельно употребляться в функции коммуникативного акта верификации (жестового верификативного «высказывания»), но чаще производятся одновременно с произнесением ВДМ.

Термин «верификативное высказывание» был введен в [Добрушина 1993]. По сути, соглашаясь с данным Е. Р. Добрушиной определением, мы в целях моделирования диалога уточним его следующим образом.

Высказывание W является верификативным, если и только если:

1) высказывание W является реакцией говорящего на некоторое высказывание U собеседника;

2) в высказывании U содержится пропозиция ‘Р’ в одном из трех истинностных статусов<sup>3</sup>: «истина» (И), «ложь» (Л), «неопределенное истинностное значение» (Ø);

3) в высказывание W входит утверждение истинности или ложности пропозиции ‘Р’.

Из данного определения следует, что верификативные реплики могут быть либо позитивными, как в (1), либо негативными, как в (2). Далее рассматриваются только позитивные верификативные реплики. Обратим внимание на то, что верификативность не является иллокутивной характеристикой реплики, поскольку относится только к ее пропозициональному содержанию, тогда как цель говорящего и прочие условия успешности речевого акта она не затрагивает. Верификативные реплики могут иметь различные иллокутивные функции (ИФ). Так, реплики в (1) имеют ИФ ответа на вопрос, а в (2) — возражения. Возможен для данного семантического типа высказываний и ряд других ИФ, например подтверждение, согласие, отказ и их прагматические разновидности. Таким образом, ИФ высказывания — это один из возможных факторов варьирования паравербального оформления ВДМ.

### Материал и метод его аннотирования

Из корпуса МУРКО, входящего в НКРЯ, были извлечены видеозаписи с примерами диалогических единств, содержащих реактивную реплику с одним или более из шести ВДМ, прототипически выражающих позитивную верификацию, но имеющих тонкие прагмасемантические различия, как было показано в [Добрушина 1993; Киселева, Пайар 1998] (табл. 1).

<sup>2</sup> В [Крейдлин, 2004] данные жесты отнесены к классу жестов-регуляторов.

<sup>3</sup> Истинностный статус пропозиции — это истинностное значение, которое приписывает пропозиции говорящий, а не ее действительное истинностное значение в мире дискурса.

Таблица 1

## Количество реплик с различными ВДМ, вошедших в выборку

да	да-да <sup>4</sup>	ну да <sup>5</sup>	конечно	разумеется	естественно
46	46	42	45	46	41

Каждая реплика с таким ВДМ была проаннотирована по двум группам параметров. Первая группа — это формальные параметры всей реплики или только ВДМ: 1) речевой ряд (письменная форма реплики в нотации МУРКО); 2) просодия ВДМ (в нотации С. В. Кодзасова [Кодзасов 2009])<sup>6</sup>; 3) вербальное описание жестового ряда, сопровождающего речевой, с фиксацией расположения жестикуляции в декартовых координатах вслед за [Гришина 2013; Гришина 2014]; 4) прочие паравербальные характеристики реплики.

Вторая группа — это характеристики реплики или коммуникативной ситуации, которые предположительно могут оказывать влияние на просодию ВДМ или жестовый ряд. К ним относятся, во-первых, синтаксические характеристики: а) синтаксическая функция ВДМ: вводное слово или вводное предложение (например, *Конечно, приду*; *Да, он здесь*) / предложение, равное реплике (например, *Да; Естественно, малышка!*) / предложение в составе периода (например, *Конечно! Мы приглашаем только тех/ кто согласен пройти полное обследование*); б) наличие дискурсивных единиц (ДЕ) до/после ДЕ, содержащей верификацию; в) наличие в составе реплики других ДМ (например, *А! Конечно/ нас это очень интересует! Спасибо!*). Характеристики (а) и (б) важны прежде всего для просодии, но также и как фактор возможного сдвига жеста по отношению к ВДМ, а наличие в реплике других ДМ (признак (в)) позволяет уточнить прагматический аспект верификативной реплики, что важно, поскольку верификативные реплики, пропозициональная часть которых сводится к ВДМ, иллюкутивно неоднозначны.

Во-вторых, это прагматические характеристики. Они включают: а) ИФ верификативной реплики; б) ИФ реплики-стимула; в) прагматическую полярность высказывания: позитивную (+), если утверждение истинности со стороны верификатора (В) искренне и серьезно, или негативную (–), если В прибегает к иронии, то есть, утверждая истинность, имеет в виду ложность пропозиции, введенной собеседником (С). Эти параметры были включены в модель исходя из предположения, что кивок — это, в отличие от ВДМ, однозначное, но дополнительное при наличии ВДМ средство кодирования позитивной верификации и потому скорее будет присутствовать

<sup>4</sup> В [Добрушина 1993] было показано, что ВДМ с итерацией сегмента *да* (доходящей в нашем материале до 5) отличается по значению от ВДМ *да*.

<sup>5</sup> Хотя ВДМ *ну да* и формально и семантически представляет собой сочетание, в котором модальная частица *ну* в одном из ее просодико-семантических вариантов (см. [Баранов, Кобозева 1988]) композиционно соединяется с верификативным значением пропозициональной частицы *да*, для целей установления корреляций между вербальными и жестовыми единицами нам было удобнее рассматривать это сочетание как составной ВДМ.

<sup>6</sup> В отличие от работ Е. А. Гришиной, в которых просодия ДМ описывается на основе субъективной оценки, в нашем исследовании просодическая транскрипция опирается на данные инструментального анализа.



в таких репликах, ИФ которых предполагает позитивную верификацию со стороны В, и отсутствовать в случаях, когда ВДМ употреблен в несобственном значении, например в репликах с фатической ИФ (поддержания контакта) или репликах иронического согласия с мнением С. ИФ реплики-стимула введена в аннотацию с целью контроля правильности приписывания ИФ реплике, содержащей ВДМ.

В-третьих, это социолингвистические характеристики: а) соотношение социальных статусов С и В:  $V = C / V < C / V > C$ ; б) соотношение С и В по возрасту:  $V = C / V < C / V > C$ ; в) конфликтность ситуации (+ / -); г) пол В и С: м / ж. Введение в модель данных параметров связано с тем, что позитивная верификация в репликах с ИФ подтверждения и согласия соответствует принципу вежливости и подчиненному ему правилу согласия (с мнением С, с его желаниями и т. п.) [Leech 1983]. Отсюда можно предположить, что кивание, как знак, информативно избыточный при наличии ВДМ, будет функционировать как маркер вежливости, то есть присутствовать в ситуациях, предполагающих подчеркнутую степень вежливости В по отношению к С, и отсутствовать, когда в таком подчеркивании нет нужды. Теории вежливости сходятся на том, что на степень вежливости влияет социальная дистанция между коммуникантами, определяемая их социальным положением, возрастом и гендером. Конфликтность ситуации может снимать действие принципа вежливости, то есть можно ожидать, что в конфликтных ситуациях кивок будет отсутствовать. Кроме того, высказывались мнения, что у говорящих разного пола обнаруживаются разные предпочтения в выборе ДМ. Мы предположили, что наличие кивка при ВДМ может относиться к такого рода предпочтениям.

Приведем конкретный пример аннотированного фрагмента из полученной базы данных:

[Трентон, Владимир Басов, муж., 56, 1923]

[Леди Венделер, Елена Соловей, жен., 32, 1947]

Трентон: Минутку! У вас/ конечно/ есть список драгоценностей?

Леди Венделер: **Разумеется/** он у меня в кабинете.

[Евгений Татарский и др. Клуб самоубийц, или Приключения титулованной особы, к/ф (1979)]

<b>Речевой ряд</b>	Разумеется/	он у меня в кабинете.
<b>Просодия</b>	разуме́ется(ˈ – /ˈ)	
<b>Жестовый ряд</b>	–	Кивок
<b>Прочая паравербалика</b>	смотрит на собеседника	смотрит на собеседника
<b>Синтаксическая функция ВДМ</b>	вводное предложение	
<b>Наличие ДЕ до / после ВДМ</b>	/ +	
<b>Другие ДМ в реплике</b>	–	
<b>ИФ реплики с ВДМ</b>	Подтверждение	
<b>ИФ реплики-стимула</b>	Вопрос с предположением	
<b>Прагматическая полярность</b>	+	
<b>Соотношение социальных статусов В и С</b>	$V > C$	
<b>Соотношение В и С по возрасту</b>	$V < C$	
<b>Конфликтность</b>	–	
<b>Пол С и В</b>	С (м), В (ж)	

### Анализ данных

Проаннотированные данные были подвергнуты анализу. Было подсчитано распределение кивков по репликам с разными ВДМ (табл. 2).

Таблица 2

Процент реплик с конкретным ВДМ, сопровождаемых кивком

да	да-да	ну да	конечно	разумеется	естественно
69 %	60 %	55 %	75 %	66 %	9 %

Уже из этого простого подсчета очевидно, что ВДМ различаются по частоте сопровождения их кивком: если для большинства ВДМ кивок более вероятен, чем его отсутствие, то для *естественно* кивок не характерен. Это свидетельствует о том, что какие-то прагмасемантические компоненты, входящие в значение данного ВДМ, препятствуют употреблению кивка. Так, в данном случае речь идет о входящем в семантику ВДМ *конечно*, *разумеется* и *естественно* компоненте «гарант» [Киселева, Пайар 1998: 33]. В случае *естественно* в качестве гаранта истинности пропозиции Р, помимо говорящего, выступает природа вещей [Киселева 365]. Сам верификатор как бы отступает на второй план, и вследствие этого телесная вовлеченность в процесс верификации и дополнительное привлечение внимания к собственной персоне при помощи кивка становятся излишними.

Помимо кивков были выявлены следующие типы жестов, сопровождающие ВДМ и обогащающие их дополнительными значениями: «*щепоть*», «*указательный палец*», «*раскрытие пальцев вверх из положения неплотно сомкнутого кулака*», «*движение рукой от себя в сторону собеседника раскрытой ладонью вверх*», «*пожать плечами*», «*поднять брови*», «*наклонить голову на бок*», «*развести руками*» и др. Также в некоторых контекстах встретился жест «*закрыть глаза*», который ассоциируется со значением слова *да* [Крейдли 2004: 401] и может самостоятельно выступать в верификативном значении.

Таблица 3

Зависимость наличия кивка от выбора ВДМ

	есть кивок	нет кивка
да	32	14
да-да	22	20
ну да	28	18
конечно	33	12
разумеется	31	16
естественно	9	32
$\chi^2 = 30.673$ , $df = 5$ , $p\text{-value} = 1.087e-05$ Критическое значение $\chi^2$ при уровне значимости $p < 0.01$ составляет 15.086.		

Дополнительно фиксировалось установление зрительного контакта верификатора с собеседником при произнесении реплики с ВДМ.

Приведем выявленные корреляции.

Оказалось, что наличие/отсутствие сопровождающего кивка коррелирует с выбранным дискурсивным маркером (табл. 3).

При этом вне зависимости от используемого маркера кивок обычно отсутствует, если прагматическая полярность верификации отрицательная (ирония, скепсис), как в (3), и (или) ИФ реплики такова, что верификация

не является целью речевого акта, как в (4), где *Да-да* употреблено не для подтверждения истинности сказанного собеседником, а с целью сообщения ему, что сообщенная им информация принята к сведению:

- (3) Галя: Если честно/ я в Москву не для того приехала/ чтоб полы здесь мыть.  
Стасис: Ну **конечно**/ сразу звездой/ да? Только ведь/ как говорится/ все цели поражены. Надо стоять в очереди. А очередь длиннющая!  
[Андрей Михалков-Кончаловский, Авдотья Смирнова. Глянец, к/ф (2007)]
- (4) Боря: Эй / алё / алё / пацаны / его щас презентовать.  
Леша: **Да-да**. Мы щас решим всё.  
[Олег Фомин и др. День выборов, к/ф (2007)]

Обнаружилась также корреляция между направленностью взгляда верификатора и выбором ВДМ (табл. 4), а также наличием кивка (табл. 5).

Как показывают наши данные, ВДМ *разумеется* в подавляющем большинстве случаев сопровождается установлением зрительного контакта с С, что также можно объяснить особым типом «гаранта» в семантике этого ВДМ. У *разумеется* гарантом истинности высказывания, помимо самого верификатора, выступает любой разумный субъект [Киселева 1998: 355], и тем самым В, фиксируя взгляд на С, как бы передает ему функцию гаранта истинности пропозиции<sup>7</sup>.

Наличие зрительного контакта между собеседниками коррелирует с наличием или отсутствием кивка при произнесении ВДМ (см. табл. 5), что вполне естественно, учитывая, что кивок обращен к зрительному восприятию С.

В области социолингвистических переменных обнаружили следующие тенденции. Наличие или отсутствие кивка оказалось связанным с социальной дистанцией между собеседниками (табл. 6).

<sup>7</sup> Такая интерпретация зрительного контакта В с С была впервые предложена С. В. Татевосовым в устном общении.

Таблица 4

**Наличие зрительного контакта при разных ВДМ**

	есть зрительный контакт	нет зрительного контакта
да	24	22
да-да	29	12
ну да	31	11
конечно	25	20
разумеется	40	7
естественно	28	13
$\chi^2 = 12.413$ , $df = 5$ , $p\text{-value} = 0.1001$ Критическое значение $\chi^2$ при уровне значимости $p < 0.05$ составляет 11.07.		

Таблица 5

**Корреляция между зрительным контактом В с С и кивком**

	есть зрительный контакт	нет зрительного контакта
есть кивок	115	31
нет кивка	71	57
$\chi^2 = 15.929$ , $df = 1$ , $p\text{-value} = 6.575e-05$ Критическое значение $\chi^2$ при уровне значимости $p < 0.01$ составляет 6.635.		

Таблица 6

**Корреляция социальной дистанции между В и С с наличием кивка**

	есть кивок	нет кивка
$B < C$	33	16
$B > C$	39	20
$B = C$	76	74
$\chi^2 = 6.5893$ , $df = 2$ , $p\text{-value} = 0.03708$ Критическое значение $\chi^2$ при уровне значимости $p < 0.05$ составляет 5.991.		

Таблица 7

**Корреляция конфликтности ситуаций  
наличием кивка**

	есть кивок	нет кивка
конфликт	29	43
нет конфликта	121	74
$\chi^2 = 10.126$ , $df = 1$ , $p\text{-value} = 0.00234$ Критическое значение $\chi^2$ при $p < 0.01$ составляет 6.635.		

Данные говорят о том, что кивок существенно чаще сопровождает верификацию при наличии социальной дистанции между В и С. Это можно трактовать как свидетельство того, что данный жест является средством позитивной вежливости, направленной на сближение с собеседником, а не негативной вежливости дистанцирующего типа [Brown, Levinson 1987].

Наличие кивков при ВДМ зависит также от конфликтности коммуникативной ситуации (табл. 7).

Что касается просодического оформления ВДМ, то на исследованном материале наблюдается корреляция между ИФ реплики с ВДМ и ее интонационным контуром, а также интегральным просодическим параметром степени общей редукции ВДМ: в тех случаях, когда верификация Р не является целью речевого акта (например, в фатических употреблениях), ВДМ имеет меньшую длительность и громкость и в целом произносится менее отчетливо. В данном случае можно говорить о частном случае иконичности просодии: центральные, прототипические употребления ВДМ характеризуются нормальной артикуляцией, а периферийные — ослабленной. Другая корреляция — между конфликтностью ситуации и громкостью. Вопреки ожиданиям, конфликтная позитивная верификация характеризуется отклонением от нормы громкости как в сторону увеличения интенсивности звука, так и в сторону ее уменьшения.

### Употребление ВДМ в неверификативных целях

Рассматриваемые нами дискурсивные маркеры используются не только в верификативных высказываниях (см. выше), но и в речевых актах, выполняющих чисто фатическую функцию, то есть функцию поддержания контакта между собеседниками и регулирования дистанции между ними. Речевые и невербальные действия, совершаемые в качестве фатической реакции на слова собеседника, в англоязычной лингвистике обозначаются термином *backchanneling*, то есть осуществление обратной связи. В дальнейшем мы будем называть дискурсивные маркеры, используемые для осуществления обратной связи, дискурсивными маркерами обратной связи, сокращенно ДМОС (англ. backchannels [Knight 2009]). По [Coates 1986] ДМОС — это короткие вербальные или невербальные ответные реакции, которые используются Слушающим для выражения позитивного внимания к Говорящему без каких-либо попыток поменяться с Говорящим ролями. Используя ДМОС, Слушающий демонстрирует, что принцип коммуникативного сотрудничества (кооперативности) Грайса остается в силе [O’Keeffe and Adolphs, 2008: 74], и это помогает собеседникам выстраивать отношения друг с другом.

Отличить верификативную реплику от фатической реакции, осуществляемой при помощи того же самого маркера, например *да-да*, возможно далеко не всегда,

даже когда мы работаем с образцами мультиканальной коммуникации. Дело прежде всего в том, что верификативная функция реплики вполне совместима с фатической, например, когда говорящий А высказывает свое мнение по какому-то поводу, собеседник Б выражает согласие с этим мнением (и тем самым позитивное внимание к говорящему А), не пытаясь перехватить инициативу в разговоре, и говорящий А продолжает развивать свою мысль, ср.:

- (5) А: *Биштекс жесткий.*  
Б: *Да-да.*  
А: *Надо сказать об этом официанту.*

Нами был выделен ряд условий и признаков для отделения чисто верификативных ДМ от ДМОС, с которыми они совпадают по форме. Назовем последние квазиверификаторами. В первую очередь интерпретация ВДМ как квазиверификатора исключается в случае соответствия коммуникативной ситуации условиям успешности речевых актов, иллюкутивной целью которых является верификация суждения. Это, во-первых, те ситуации, когда Собеседник вводит в рассмотрение пропозицию, истинностное значение которой ему неизвестно или он в нем не уверен (задает вопрос, переспрашивает и т. п.), а Верификатор, более осведомленный или более компетентный в данной области, чем Собеседник, при помощи ВДМ сообщает ему, что данная пропозиция истинна (мы говорим здесь только о позитивных ВДМ, поскольку именно они могут совпадать по форме с ДМОС). Короче говоря, это иллюкутивно вынуждаемые ответы на информационный запрос. Во-вторых, это ситуации, в которых Собеседник высказывает суждение, истинность которого не очевидна третьему лицу или аудитории, а Верификатор, который в силу своей компетентности пользуется большим или, по крайней мере, не меньшим доверием, чем Собеседник, подтверждает истинность сказанного им, чтобы снять возможные сомнения. Таким образом, к ДМОС не имеют отношения ВДМ в репликах с ИФ ответа на вопрос (в широком смысле, включающем вопрос-предположение и переспрос) и с ИФ подтверждения.

Напротив, если реплика-стимул в принципе не подлежит верификации, как в случае перформативных РА, или не предполагает верификации со стороны адресата, как в случае прямых директивов, комиссивов, экспрессивов и деклараций [Серль 1986], то реакция на него при помощи ВДМ может интерпретироваться только как квазиверификация, то есть как ДМОС, или как псевдоверификация (см. ниже). Ср. квазиверификативные реакции в (6) и (7):

- (6) А, обращаясь к Б: Я вам так благодарен.  
Б: Да.
- (7) А, обращаясь к Б: Не волнуйтесь, я все устрою.  
Б. Конечно.

Целый ряд ситуаций допускает амбивалентную трактовку ВДМ. Так, в случае, когда участниками коммуникативной ситуации являются только собеседники А и Б и при этом оба равно компетентны в обсуждаемом вопросе, то произнесение

собеседником Б в качестве реакции на утверждение собеседника А в принципе может интерпретироваться тремя способами: 1) как намеренный РА согласия, цель которого — дать собеседнику знать, что говорящий придерживается того же мнения; 2) как чистый сигнал обратной связи (ДМОС), если на самом деле мнение говорящего не совпадает с мнением собеседника, но он по тем или иным причинам предпочитает его не озвучивать, например из вежливости (ср. максимум согласия как одну из максим принципа вежливости [Leech 1983]) или из-за нежелания делать свой речевой ход в тот момент, когда Собеседник, возможно, еще не успел привести аргументы в пользу высказанного мнения; 3) речевой акт согласия с мнением, совмещенный с сигналом обратной связи — выражением «позитивного внимания» к словам Собеседника без намерения перехватить инициативу в диалоге.

Признаки, характерные для чистых ДМОС, вытекают из приведенного выше определения:

1) ДМОС информационно избыточны, то есть они не добавляют в диалог новой информации (ни пропозициональной, ни модусной);

2) ДМОС могут употребляться тогда, когда предыдущее высказывание Собеседника еще не завершено в смысловом отношении, и будет продолжено им, то есть ДМОС может вставляться без получения от Собеседника сигнала о смене ролей;

3) соответственно ДМОС могут иметь особое таймирование: они либо отделены от реплики Собеседника сверхкороткой паузой, либо происходит наложение ДМОС на неоконченную реплику собеседника.

Однако даже на основании выделенных признаков не всегда удастся четко отделить ДМОС от ВДМ, к тому же признаки далеко не всегда присутствуют одновременно, поэтому мы принимаем решение не представлять отношение верификации и квазиверификации в виде бинарной оппозиции, а понимать его как шкалу, на одном полюсе которой находятся «чистые» квазиверификативные ДМОС, а на другом — «чистые» верификативные речевые акты. Между ними располагаются смешанные случаи, в большей или меньшей степени проявляющие свойства либо верификативных речевых актов того или иного типа, либо сигналов обратной связи.

Приведем пример квазиверификативной реплики из МУРКО:

- (8) [Мария (лингвист), жен.]  
[Евгений Головкин, муж.]  
Мария (лингвист): Ну вот как пример — допустим/ никарагуанский жестовый/ который возник на пустом месте буквально.  
Евгений Головкин: Да.  
Мария (лингвист): То есть субстрата нет...  
[Евгений Головкин. Как рождаются языки. Лекции Полит.ру (2011)]

Этот пример демонстрирует прототипический случай использования ВДМ *да* в функции ДМОС. В рассмотренном диалоге нет условий для верификативной интерпретации реплики *Да*, так как реплика-стимул содержит одно утверждение

(‘никарагуанский жестовый язык возник на пустом месте’), а Г (интервьюер) в силу своей меньшей компетентности по сравнению с интервьюируемым специалистом не соответствует условиям успешности РА подтверждения или согласия с этим утверждением. Также данное высказывание соответствует всем перечисленным выше признакам, характерным для сигналов обратной связи.

Помимо собственно верификативных и квазиверификативных ДМ (разновидности ДМОС), можно выделить также псевдоверификативные ДМ. Рассмотрим два примера из МУРКО:

- (9) [Марья Васильевна, Нонна Мордюкова, жен., 56, 1925]  
[Станислав, Юрий Богатырев, муж., 34, 1947]  
Станислав: А тепёрь серьёзно. Передáйте/ пожа́луйста/ Ни́не/ что отны́не но́ги моёй зде́сь не бу́дет. Нико́гда.  
Марья Васильевна: Да ниче́го/ переживём.  
Станислав: **Есте́ственно.**  
[Никита Михалков, Виктор Мережко. Родня, к/ф (1981)]

В (9) *Естественно*, несомненно, не ДМОС: в данном контексте этот ДМ не может служить фатическим сигналом позитивного внимания со стороны Станислава к словам собеседницы, поскольку ситуация явно конфликтная. Но нельзя здесь говорить и о подлинной верификации, хотя на первый взгляд Г как бы подтверждает ее слова. Дело здесь как в пропозициональном содержании реплики собеседницы — ‘мы переживем (то, чем ты нам угрожаешь)’, так и в ИФ ее высказывания — по сути, заявления о том, что косвенная угроза Станислава ее не испугала. Как было сказано выше, экспрессивы (выражение собственного психического состояния), по понятным причинам, не подлежат верификации со стороны адресата. В действительности ДМ *естественно* выступает здесь не как ВДМ, а как маркер того, что говорящий ожидал от собеседницы именно такой реакции. Его значение можно передать при помощи следующей перифразы: *Ничего другого я и не ожидал от вас услышать*.

- (10) [Перкинс, А. Абакаров, муж., 34, 1945]  
[Принц Флоризель, Олег Даль, муж., 38, 1941]  
Перкинс: Учтите/ на всё вопро́сы на́до отве́чать с по́льной откровенно́стью.  
Принц Флоризель: Ну разуме́ется.  
[Евгений Татарский и др. Клуб самоубийц, или Приключения титулованной особы, к/ф (1979)]

В (10) ДМ *Ну разуме́ется* используется не как ДМОС, поскольку данная реплика представляет собой иллокутивно вынуждаемую реакцию на директивный РА, а именно реакцию согласия (синонимичная эксплицитная перифраза: *Хорошо, я буду отвечать с полной откровенностью*). Но согласие с побуждением, хотя и может выражаться теми же средствами, что и согласие с мнением, не является верификативным РА, поскольку цель его состоит не в выражении мнения об истинности введенной ранее пропозиции, а в принятии на себя обязательства вести себя



в будущем в соответствии с высказанным собеседником желанием, предписанием и т. п.

Всего мы обнаружили 31 пример, соответствующий чистым или смешанным случаям сигналов обратной связи, а также псевдоверификативным репликам, среди которых 22 сопровождалась кивками (71 %). Однако на основании наших материалов сложно делать какие-то достоверные выводы о паравербальных характеристиках квази- или псевдоверификативных употреблений ВДМ, поскольку объем данных очень мал, так что этот вопрос остается открытым для дальнейших исследований.

### Выводы

Связь между жестами и тонкими различиями в семантике ВДМ очевидна, однако ее проявление — это скорее тенденция, чем жесткое правило. Зачастую жестикуляция представляет собой реализацию не одного, а нескольких жестов. Один и тот же жест может реализовываться в устном дискурсе вместе с разными дискурсивными маркерами. Причиной такой вариативности является многоплановая семантическая структура ВДМ, осложненная прагматическим компонентом каждого отдельного высказывания, которая и определяет пересечения в жестикуляции между различными маркерами.

Наличие или отсутствие кивка при употреблении ВДМ зависит от таких параметров, как социальный статус собеседников, их возраст, конфликтность ситуации, а также различия в сопровождении кивком прослеживаются между разными дискурсивными маркерами. Наличие кивка при произнесении высказывания не связано с полом собеседников и со статусом дискурсивного слова в высказывании. Наличие зрительного контакта зависит от используемого ВДМ и от сопровождающей этот маркер жестикуляции.

Положительная полярность верификации является необходимым, но не достаточным условием сопровождения верификативной реплики кивком.

Просодическая вариативность ВДМ коррелирует с ИФ реплики, со статусом верификативного компонента в смысловой структуре реплики (центральным или периферийным), с прагматической полярностью высказывания и конфликтностью ситуации.

### Литература

Баранов А. Н., Кобозева И. М. Модальные частицы в ответах на вопрос // Прагматика и проблемы интенциональности. М. : ИЯ АН СССР, 1988. С. 45–69.

Бонно К., Кодзасов С. В. Семантическое варьирование дискурсивных слов и его влияние на линеаризацию и интонирование (на примере частиц *же* и *ведь*) // Дискурсивные слова русского языка: опыт контекстно-семантического описания / под ред. К. Киселевой и Д. Пайара. М. : Метатекст, 1998. С. 382–443.

Григорьева С. А., Григорьев Н. В., Крейдлин Г. Е. Словарь языка русских жестов. М. : Языки русской культуры; Вена : Венский славистический альманах, 2001. 226 с.

Гришина Е. А. О мультимодальных кластерах в устной речи // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». 2011. Вып. 10 (17).

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

Гришина Е. А. Слово и жест: корпусные исследования устной речи. Saarbrücken : LAP, 2012.

Гришина Е. А. Жестикуляционные профили русских приставок // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». 2013. № 12. С. 19.

Гришина Е. А. Жесты и прагматические характеристики высказывания // Мультимодальные коммуникации: теоретические и эмпирические исследования. М., 2014. С. 25–47.

Добрушина Е. Р. Верификация в современной русской диалогической речи: дис. ... канд. филол. наук. М. : МГУ, 1993.

Киселева К. Л. РАЗУМЕЕТСЯ, или САМИ ПОНИМАЕТЕ // Дискурсивные слова русского языка: опыт контекстно-семантического описания / под ред. К. Киселевой и Д. Пайара. М. : Метатекст, 1998.

Киселева К. Л. ЕСТЕСТВЕННО, или DE RERUM NATURAE // Дискурсивные слова русского языка: опыт контекстно-семантического описания / под ред. К. Киселевой и Д. Пайара. М. : Метатекст, 1998.

Киселева К. Л., Пайар Д. Дискурсивные слова русского языка: Опыт контекстно-семантического описания. М. : Метатекст, 1998. 447 с.

Кобозева И. М., Захаров Л. М. Для чего нужен звучащий словарь дискурсивных слов русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международного семинара Диалог'2004. М. : Наука, 2004. С. 292–297.

Кобозева И. М., Захаров Л. М. «Как много в этом звуке!..» (просодико-семантические варианты русского междометия а) // Лингвистическая полифония: сборник статей в честь юбилея профессора Р. К. Потаповой. М. : Языки русских культур, 2007. С. 609–627.

Кобозева И. М., Захаров Л. М. Семантико-просодические вариации на тему «Да» (материалы к мультимедийному словарю русских дискурсивных слов) // Среди нехоженых путей: сборник научных статей к юбилею доктора филологических наук, профессора А. А. Кретьева. Воронеж : Изд-во ВГУ, 2012.

Кодзасов С. В. Комбинаторная модель фразовой просодии // Кодзасов С. В. Исследования в области русской просодии. М., 2009. С. 13–47.

Крейдлин Г. Е. Невербальная семиотика. М., 2004.

18. Серль Дж. Р. Классификация иллокутивных актов // Новое в зарубежной лингвистике. Теория речевых актов. М. : Прогресс, 1986. № 17. С. 170–194.

Brown P., Levinson S. Politeness: some universals in language usage. Cambridge : Cambridge University Press, 1987.

Cienki A., Müller C. (eds.) Metaphor and Gesture. Gesture Studies, 3. Amsterdam, Philadelphia : John Benjamins, 2009. P. 307.

Coates J. Women M. Language: A Sociolinguistic Account of Sex Differences in Language. L. ; N. Y. : Longman, 1986. P. 10–11.

Knight D. A multi-modal corpus approach to the analysis of backchanneling behavior. University of Nottingham, 2009.

Leech G. Principles of pragmatics. L. : Longman, 1983.

O'Keeffe A. and Adolphs S. Using a corpus to look at variational pragmatics: Response tokens in British and Irish discourse // Variational Pragmatics / eds. K. P. Schneider and A. Barron. Amsterdam, Netherlands : John Benjamins, 2008. P. 69–98.

Richter N. The gestural realization of some grammatical features in Russian // Gesture: evolution, brain and linguistic structures. Proceedings of the 4<sup>th</sup> conference of the international society for gesture studies. Frankfurt; Oder, 2003. P. 245.

**Irina Kobozeva, Olga Ivanova, Leonid Zakharov**

*Lomonosov Moscow State University*

*(Moscow, Russia)*

*kobozeva@list.ru, wiola100@rambler.ru, leonid\_zakharov@mail.ru*

## **TOWARDS MULTIMODAL MODELLING OF VERIFICATIONAL DISCOURSE MARKERS IN RUSSIAN DIALOG**

In the paper we present the results of a multimodal analysis of some Russian discourse markers (DMs) expressing positive verification — *da* ‘yes’, *da-da* ‘yes’, *nu da* ‘yes’, *конечно* ‘of course’, *разумеется* ‘of course’, *естественно* ‘naturally’. In a dialogical turn of a given speaker they mark the proposition P of the previous turn of the interlocutor as true. DMs of this type have variable prosody and may redundantly be accompanied with a synonymous gesture “nod”. The aim of the study was to identify the factors that influence the paraverbal aspects of such DMs. The analysis of mini-dialogs from the multimodal sub-corpus of RNC revealed the correlations between various syntactic, semantic, pragmatic and sociolinguistic factors on the one hand and gestural accompaniment and prosodical features of DMs on the other. Moreover, the survey showed some instances of backchanneling which is similar to verification under certain conditions. The results obtained may be useful for modelling the paraverbal behavior of embodied computer agents and robots.

*Key words:* multimodal speech analysis, dialog, discourse markers, verification, gesture, nod, prosody, semantics, pragmatics

## **References**

Baranov A. N., Kobozeva I. M. [Modal particles in answers to questions]. *Pragmatika i problemy intensional'nosti* [Pragmatics and Intensionality Problems]. Moscow, IY a AN SSSR, 1988, pp. 45–69. (In Russ.)

Bonnot C., Kodzasov S. V. [Semantic variation of discourse markers and its influence on linearization and intonation (the case of particles *zhe* and *ved'*)]. *Diskursivnye slova russkogo yazyka: opyt kontekstno-semanticheskogo opisaniya* [Discourse markers of Russian: a study in contextual-semantic description]. Moscow, Metatekst Publ., 1998, pp. 382–443. (In Russ.)

Brown P., Levinson S. Politeness: some universals in language usage. Cambridge, Cambridge University Press, 1987.

Cienki A., Müller C. (eds.) Metaphor and Gesture. *Gesture Studies*, 3. Amsterdam — Philadelphia, John. Benjamins, 2008. 307 p.

Coates J., Women M. *Language: A Sociolinguistic Account of Sex Differences in Language*. London and New York, Longman, 1986, pp. 10–11.

Dobrushina E. R. *Verifikatsiya v sovremennoi russkoi dialogicheskoi rechi*. Diss. kand. filol. nauk [Verification in modern Russian dialogical discourse. PhD phil. sci. diss]. Moscow, MSU, 1993. (In Russ.)

Grigor'eva S. A., Grigor'ev N. V., Kreidlina G. E. *Slovar' yazyka russkikh zhestov* [The dictionary of Russian gestures]. Moscow — Vienna, Yazyki russkoi kul'tury Publ.; Wiener slawistischer almanach. Sonderband 49, 2001. 226 p. (In Russ.)

Grishina E. A. [On multimodal clusters in oral speech]. *Komp'yuternaya Lingvistika i Intellektual'nye Tehnologii* [Computer linguistics and intellectual technologies]. Moscow, RSUH, 2011, no. 10 (17). (In Russ.)

Grishina E. A. [Da in Russian oral dialog]. *Russkaya lingvistika* [Russian Linguistics]. 2011, no. 35 (2), pp. 169–207. (In Russ.)

Grishina E. A. *Slovo i zhest: korpusnye issledovaniya ustnoi rechi* [Word and gesture: corpus investigation of oral speech]. Saarbrücken, LAP, 2012. (In Russ.)

Grishina E. A. [Gesture profiles of Russian prefixes]. *Komp'yuternaya Lingvistika i Intellektual'nye Tehnologii* [Computer Linguistics and Intellectual Technologies]. 2013, no. 12 (19), p. 19. (In Russ.)

Grishina E. A. [Gestures and pragmatic characteristics of an utterance]. *Mul'timodal'nye kommunikatsii: teoreticheskie i empiricheskie issledovaniya* [Multimodal communications: theoretical and empirical studies]. Moscow, 2014, pp. 25–47. (In Russ.)

Kiseleva K. L. [JESTESTVENNO 'naturally' or DE RERUM NATURAE]. *Diskursivnye slova russkogo yazyka: opyt kontekstno-semanticheskogo opisaniya* [Discourse markers of Russian: a study in contextual-semantic description]. Moscow, Metatekst Publ., 1998. (In Russ.)

Kiseleva K. L. [RAZUMEETSYA 'of course' or YOU UNDERSTAND]. *Diskursivnye slova russkogo yazyka: opyt kontekstno-semanticheskogo opisaniya* [Discourse markers of Russian: a study in contextual-semantic description]. Moscow, Metatekst Publ., 1998. (In Russ.)

Kiseleva K., Pajar D. *Diskursivnye slova russkogo yazyka: opyt kontekstno-semanticheskogo opisaniya* [Discourse markers of Russian: a study in contextual-semantic description]. Moscow, Metatekst Publ., 1998. 447 p. (In Russ.)

Knight D. A multi-modal corpus approach to the analysis of backchanneling behavior. University of Nottingham, 2009.

Kobozeva I. M., Zakharov L. M. [The sounding dictionary of discourse markers of Russian: what for?]. *Kompyuternaya Lingvistika i Intellektual'nye Tehnologii* [Computer linguistics and intellectual technologies]. Moscow, Nauka Publ., 2004, no. 3 (10), pp. 292–297. (In Russ.)

Kobozeva I. M., Zakharov L. M. [“How much is in this sound!” (prosodic-semantic variants of Russian interjection *a*)]. *Lingvisticheskaya polifoniya. Sbornik statei v chest' yubileya professora R. K. Potapovoi* [Linguistic polyphony. Collection of articles in honor of Professor R. K. Potapova's anniversary]. Moscow, Yazyki russkoi kul'tury Publ., 2007, pp. 609–627. (In Russ.)

Kobozeva I. M., Zakharov L. M. [Semantic-prosodic variations of the theme “Da” (materials for the multimedia dictionary of Russian discourse markers)]. *Sredi nekhozhennykh putei: Sbornik nauchnykh statei k yubileyu doktora filologicheskikh nauk, professora A. A. Kretova* [Among untrodden ways: Collection of scientific articles for the anniversary of Doctor of Philology, Professor AA. Kretov]. Voronezh, VSU Publ., 2012. (In Russ.)

Kodzasov S. V. *Kombinatornaya model' frazovoi prosodii* [Combinatorial model of phrasal prosody]. *Issledovaniya v oblasti russkoi prosodii* [Studies on Russian prosody]. Moscow, 2009, pp. 13–47. (In Russ.)

Kreidlin G. E. *Neverbal'naya semiotika* [Non-verbal semiotics]. Moscow, 2004. (In Russ.)

Leech G. Principles of pragmatics. L., Longman, 1983.

O'Keeffe A. and Adolphs S. Using a corpus to look at variational pragmatics: Response tokens in British and Irish discourse. In Schneider K. P. and Barron A. (Eds.) *Variational Pragmatics*. Amsterdam, Netherlands, John Benjamins, 2008, pp. 69–98.

Richter N. The gestural realization of some grammatical features in Russian, In Proceedings of the 4<sup>th</sup> conference of the international society for gesture studies. *Gesture: evolution, brain and linguistic structures*. Frankfurt/Oder, 2010, p. 245.

Searl' Dzh. R. [A classification of illocutionary acts]. *Novoe v zarubezhnoi lingvistike. Teoriya rechevykh aktov*. Moscow, Progress Publ., 1986, no. 17, pp. 170–194. (In Russ.)

**А. А. Зинина, Л. Я. Зайдельман, Н. А. Аринкин, А. А. Котов**  
НИЦ «Курчатовский институт»  
(Россия, Москва)  
zinina\_aa@nrcki.ru

## **ПОДХОДЫ К КЛАССИФИКАЦИИ КОММУНИКАТИВНЫХ ЖЕСТОВ: ТЕОРИЯ И ПРИЛОЖЕНИЕ К РОБОТАМ-КОМПАЬОНАМ\***

В работе представлен подход к классификации коммуникативных жестов — описаны жесты, у которых означаемое может быть выражено с помощью различных означающих. К классификации жестов мы подходим через выделение коммуникативных функций (означаемого) и инварианта выполнения жеста (означающего). Определенная коммуникативная функция может быть выражена жестами с различным набором исполнительных органов: с помощью мимики, движений головы, рук и тела. Путем выделения коммуникативных функций решается задача моделирования сложного коммуникативного поведения роботом-компаньоном. В работе обосновывается необходимость выделения коммуникативных функций при анализе реального поведения человека в эмоциональных ситуациях (на основе мультимодального корпуса REC), а также описываются контексты и инварианты коммуникативных функций. На примере коммуникативной функции *ожидание обратной связи* описывается инвариант «акцентирование внимания к ответной реакции собеседника» и различные варианты соответствующих жестов: (1) акцентирование взгляда на адресате, (2) демонстрация собственного бездействия при ожидании адресата, (3) демонстрация готовности перейти к последующим действиям, (4) кивки в подтверждение собственного тезиса или вопросительные движения с запросом реакции собеседника. Также описывается управляющая архитектура робота, с помощью которой моделируются основные особенности естественной коммуникации: коммуникативные функции и жесты. Описание инварианта и жестов для определенной коммуникативной функции позволяет более тонко настраивать коммуникативные реакции робота на входящее высказывание.

*Ключевые слова:* мультимодальный корпус, жестовая коммуникация, иконические жесты, робот-компаньон.

---

\* Разработка системы обработки речи и коммуникативных реакций для робота поддержана грантом РФФИ № 16-29-09601.

Современная корпусная лингвистика развивается сразу в двух самостоятельных сферах: с одной стороны, исследования в области жестикуляционной лингвистики обладают самостоятельной теоретической значимостью, с другой — широким практическим потенциалом — их результаты могут быть полезны при синтезе невербального коммуникативного поведения человекоподобным роботом. Более того, только в единстве «теории» и «практики» открывается возможность создания такого робота-компаньона, который будет способен формировать положительное впечатление у пользователя, поддерживать коммуникацию, длительное время оставаться интересным и привлекательным для собеседника.

В области социальной робототехники проводится множество исследований социального взаимодействия между роботом и человеком. Известно, что часть пользователей устанавливает эмоциональный контакт даже с бытовыми роботами-пылесосами [Sung et al. 2007]. Поэтому робот не обязательно должен выглядеть как человек, наоборот, чрезмерная схожесть робота с человеком вызывает у пользователей эффект «зловещей долины» [Mori 1970]. Наш подход заключается в том, что робот должен обладать минимальной конфигурацией (наиболее простой кинематической моделью), но при этом проявлять максимум эмоций, чтобы эффективно действовать в качестве робота-компаньона. Такие роботы обычно имеют голову, лицо, иногда — ручки или уши.

Результаты эксперимента [Leite, Pereira et al. 2008] показали, что именно эмоциональное поведение повышает эффективность взаимодействия между физическим роботом и пользователем. Например, робот Kismet обладает эмоциональной моделью и может выражать имитируемые эмоции с помощью ограниченного набора движений элементов лица (глаз, бровей, рта, ушей), а также с помощью речи, включая речевую просодию [Breazeal 2003]. Робот iCat также обладает эмоциональной моделью, позволяющей либо ориентироваться при ответе на эмоции адресата, либо предпочитать собственные реакции [Hindriks et al. 2012]. Этот робот имеет подвижную голову, глаза, брови и рот. Приведенные исследования показывают, что робот должен обладать не только развитой эмоциональной моделью, но и широким спектром эмоциональных выразительных средств — жестов и мимики. Исследователи отмечают существенную значимость гибкого комбинирования жестов и речи при «интуитивном» человеко-машинном взаимодействии [Beuter, Spexard et al. 2008].

Чтобы иметь возможность выполнять ориентированные жесты (например, поворачивать голову к собеседнику или указывать направление), а также общаться с несколькими собеседниками одновременно, роботу необходимо обладать компьютерным зрением. Кроме того, для лучшего установления контакта между роботом и собеседником робот должен имитировать поведение живого человека [Breazeal, Scassellati 2002]. В исследовании [Klamer 2011] отмечается значимость для робота речевых и невербальных средств поддержания коммуникации при установлении долгосрочного контакта. Согласно психологическим исследованиям, привлекательность робота обусловлена тем, что его коммуникативное поведение должно быть максимально приближено к коммуникативному поведению



человека. Очевидно, что жестовое поведение человекоподобного робота должно быть сложным и разнообразным, поэтому при разработке такого поведения необходимо ориентироваться на теоретические представления о природе жестового общения, в частности на существующие классификации жестов.

Движения человека доступны для наблюдения другими людьми и могут передавать информацию о состоянии человека, его целях и эмоциях. С этой точки зрения движения могут являться знаками, передающими информацию наблюдателю. Вместе с тем жесты не составляют такого же словаря знаков, как слова естественного языка. Означающие и означаемые жестов достаточно размыты, это затрудняет создание классификации. Жесты разнородны как по своей произвольности (намеренности выполнения), так и по типу отношений между означающим и означаемым. С одной стороны, произвольные жесты контролируются говорящим и направлены на передачу конкретного значения: *махать рукой* выражает 'приветствие', а ладонь с растопыренными пальцами может обозначать число 'пять' — говорящий сознательно выполняет эти жесты с целью транслировать сообщение адресату. С другой стороны, трясущиеся руки могут обозначать напряжение говорящего, прикрывание лица рукой может вызываться смущением — говорящий может не контролировать эти движения, что тем не менее не препятствует передаче адресату информации о состоянии говорящего.

Согласно Ч. Пирсу, отношение между означающим и означаемым можно характеризовать через базовые типы знаков: иконические, индексные и символные. Если символные знаки совсем не похожи на обозначаемый ими предмет — их действие основано на установленной по соглашению связи означающего и означаемого, у индексных знаков означающее и означаемое находятся в отношениях пространственной и временной смежности, то у иконических знаков означающее подобно означаемому. Жесты также занимают в этой классификации различные позиции. Жесты «фига» или «ОК» имеют символную природу — означаемое в них связано с передаваемым смыслом «по договору». Китайские жесты для обозначения числительных пальцами одной руки различаются: числа от 1 до 5 — иконические: человек просто демонстрирует нужное число пальцев, а обозначения от 6 до 10 — символные знаки, они отчасти похожи на соответствующие иероглифы, но догадаться об обозначаемом числе по самим жестам — невозможно. В качестве примеров индексных знаков приведем трясущиеся руки или прикрывание лица рукой: эмоциональное состояние адресанта вызывает соответствующее движение, в понимании Пирса можно говорить о том, что между означаемым и означающим имеется физическая связь. Это затрудняет классификацию означаемых жеста: означаемым может быть и заданный смысл, как у слов естественного языка, и эмоциональное состояние говорящего — в этом случае смысл вычленяется адресатом коммуникации, но не формируется говорящим.

Такое разнообразие жестов человека усложняет создание классификации: жест может являться как рефлекторной реакцией человека, так и намеренным знаковым действием с целью передать адресату заданный смысл. Сложность представляет и классификация означающих жеста. Слова естественного языка при написании

делятся на сегменты пробелами, знаками пунктуации, а при произнесении — фонетическими параметрами: тактами вдоха и выдоха, переходами голосового аппарата от гласных к согласным звукам. По контрасту с речевым сигналом жест является движением человека в пространстве и с трудом поддается сегментации. Более того, движение человека в пространстве является сложным комплексом, который может совмещать несколько движений: жест «пять», прикрывание лица рукой, трясущиеся руки и т. д. Отсутствие для жестов заданного набора сегментов, аналогичных алфавиту, словарю или набору фонем, — также затрудняет создание классификации. Данные трудности по-разному решаются в разных теоретических подходах.

Д. Эфрон [Efron 1941/1972] предлагает разделять следующие классы жестов. Во-первых, он выделяет жесты-эмблемы, кодирующие смысл независимо от лексического контекста, имеющие строго закрепленные за ними условия употребления, — например жест «ОК». Во-вторых, жесты-иллюстраторы, которые сопровождают определенный речевой фрагмент и связаны с его содержанием. В-третьих, жесты-регуляторы, управляющие ходом коммуникативного процесса, то есть устанавливающие, поддерживающие или завершающие коммуникацию [Григорьева и др. 2001].

Согласно классификации Д. МакНила [McNeill 2005], жест обладает некоторой размерностью: иконичностью, метафоричностью, деиктичностью, ритмичностью. Под иконическими понимаются такие жесты, которые дополняют речевое высказывание, — это изображения конкретных вещей или действий. Метафорические изображают абстрактные, а не конкретные предметы, деиктичные — указывают на определенный предмет или на местоположение какой-либо абстракции в пространстве, причем такие жесты могут выполняться не только рукой, но и головой, носом, локтем, ногой, глазами и др. Ритмические жесты подчеркивают значимые фрагменты речи.

А. Кендон различает жесты по трем параметрам: обязательность сопровождающей речи, наличие или отсутствие системных языковых черт в жестикуляции, а также степень регулярности жеста. Согласно его классификации, все жесты могут быть распределены на континууме по степени обязательности сопровождающей речи — системная организация в сочетании с произвольностью отношений между означаемым и означающим. Крайняя левая точка такого континуума будет описывать «речевые» жесты, которые не могут быть однозначно интерпретированы без сопровождающего речевого высказывания [Kendon 1988].

Е. А. Гришина обобщает, что на сегодняшний день нет единой и общепринятой классификации жестов [Гришина 2017: 12]. На практике исследователи сталкиваются с невозможностью отнести какой-либо жест к одной определенной категории, поэтому, руководствуясь специализированной практической задачей — задачей моделирования коммуникативного поведения человекоподобным роботом-компаньоном, — мы разрабатываем собственную классификацию жестов.

Чтобы объединить произвольные и спонтанные знаки в одну классификацию, мы используем следующий подход. Адресант может переживать различные состояния и проявлять их в своем поведении. Пусть состояние *C* (*condition*) адресанта

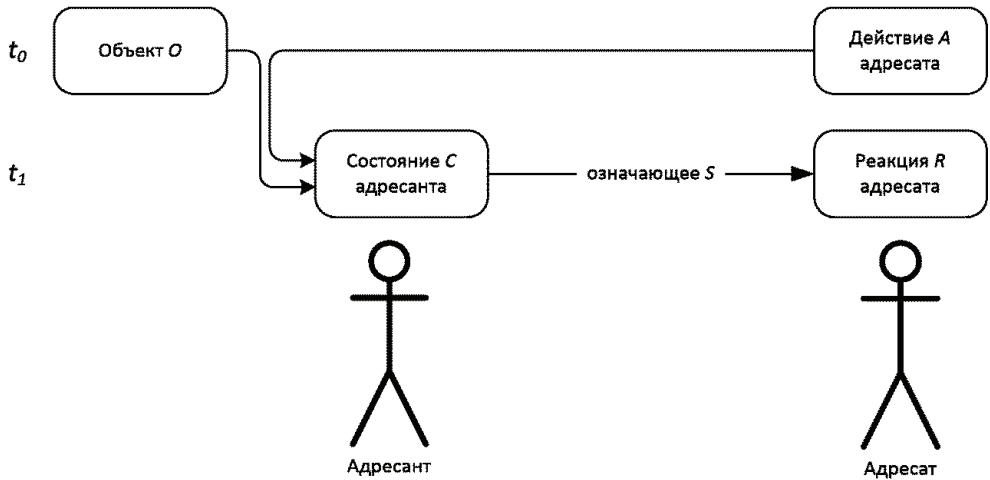


Рис. 1. Схема коммуникации для описания жестов

заставляет его выполнить действие  $S$  (*sign*) — (рис. 1). Например, адресант загрустил ( $C$ ) и глубоко вздохнул ( $S$ ). В этом случае  $C$  влечет  $S$ : имеет место  $C \Rightarrow S$  — грусть вызывает вздох. Адресат может наблюдать поведение адресанта и интерпретировать  $S$  как знак внутреннего состояния  $C$ , в этом случае  $S$  будет индексным знаком (означающим), указывающим на  $C$  (означаемое). Мы можем считать  $S$  сообщением в коммуникации, даже если  $S$  носит спонтанный характер и не контролируется адресантом. Интерпретируя знак  $S$ , адресат реконструирует состояние  $C$  адресанта, спросит того, как дела, и утешит его. В этом случае имеет место причинно-следственная связь  $S \Rightarrow R$ , где  $R$  — реакция адресата.

Состояние  $C$  могло быть вызвано внешними причинами. Некоторый объект  $O$  в окружении адресанта мог вызывать его состояние  $C$ , — например пробудить интерес или напугать. Демонстрируемые при этом адресантом действия  $S$  могут служить знаком, указывающим на присутствие в окружении привлекательного или опасного объекта. Такая простая схема поведения позволяет стайным живым существам реагировать на опасность или появление еды — при этом стимулом для действий ( $R$ ) живого существа становится поведение ( $S$ ) сородича. Поведение  $S$  можно рассматривать как индексный знак опасности  $O$ . Такую же реакцию живое существо может демонстрировать в ответ на действия  $A$  своего оппонента. Как показано в классических работах К. Лоренца [2001], агрессия является важным семиотическим механизмом, регулирующим поведение в популяции через ритуалы и знаковое взаимодействие. Животное может продемонстрировать форму агрессивного поведения  $A$ , на которое оппонент ответит встречной агрессией  $S$ . Однако тот из оппонентов, чья мотивация ниже, может в ходе конфликта продемонстрировать знак подчинения ( $S$ ), это останавливает ( $R$ ) агрессию оппонента и позволяет выбрать победителя конфликта без ущерба здоровью оппонентов.

У человека можно наблюдать набор поведенческих реакций, сходный с инстинктами животных из описаний Лоренца: объекты ( $O$ ) или действия ( $A$ ) других могут

вызывать у нас эмоции, проявляющиеся в поведении ( $S$ ). Однако человек как высшее разумное существо может контролировать собственные действия ( $S$ ) и моделировать реакции оппонента ( $R$ ). Пусть в простом случае запах тухлого яблока ( $O$ ) вызывал у адресанта отвращение ( $C$ ), что заставило адресанта закрыть нос и рот рукой ( $S$ ). Это простая реакция вида  $O \Rightarrow C \Rightarrow S$ . Сходную поведенческую реакцию могут вызвать неприятные действия  $A$  другого человека. В этом случае также имеет место причинно-следственная связь  $A \Rightarrow C \Rightarrow S$ . В произвольных реакциях объекты  $O$  и действия  $A$  будут оказывать непосредственное влияние на состояние  $C$  и вызывать действия  $S$ . Теперь представим, что входящий стимул вызывает у адресанта множество вариантов интерпретации, — то есть имеется множество состояний  $C$  и множество вариантов действий  $S$ , которые адресант может контролировать. Адресант имеет возможность прогнозировать реакции адресата на каждое такое  $S$  — будем считать, что реакции  $R$  подвержены большему контролю адресанта, и обозначать их  $R'$ . Прогнозируя различные реакции  $R'$ , адресант будет выбирать такие действия  $S$ , которые приведут к наиболее выгодным  $R'$ . Контролируемые действия будем обозначать  $S'$ . С одной стороны, это позволит адресанту выполнять референцию к объекту  $O$ : адресант может продемонстрировать интерес, внимание или эмоциональную оценку ( $S$ ) объекта  $O$ , чтобы нужным образом обратить на  $O$  внимание адресата (вызвать у адресата реакцию  $R'$  — внимание или эмоциональную оценку). С другой стороны, адресант сможет контролировать действия  $A$  адресата. Пусть адресат поступил некрасиво ( $A$ ), что вызвало осуждение ( $C$ ) адресанта. Он может проявить осуждение в форме, похожей на отвращение, — закрыть нос и рот рукой. Адресант при этом будет демонстрировать более контролируемое проявление ( $S'$ ) эмоций ( $C$ ), направленное на воздействие ( $R'$ ) на адресата. В диалоге люди, как правило, стараются снизить ущерб оппоненту [Brown, Levinson 1987]: если адресат своими действиями  $A$  доставил неудобство адресанту, то он будет стараться извиниться или избежать повторения неприятной ситуации, — то есть отвечать своими реакциями  $R$  на действия  $S$  собеседника. Такое изменение поведения адресата входит в задачу намеренного использования знаков  $S'$ . Это звено является ключевым для описания перехода от неконтролируемых действий  $S$  к произвольному поведению  $S'$ . Именно на этом шаге жесты из неконтролируемых индексных знаков становятся сообщениями, обладающими иллокутивной силой — направленными на воздействие на реакции  $R$  адресата.

Состояние  $C$  вызвано некоторым объектом  $O$ , поэтому, демонстрируя свое состояние ( $C$ ), адресант транслирует оценку объекта  $O$ . При этом адресант осуществляет референцию к объекту  $O$ , вкладывая в свое сообщение как оценку  $O$ , так и указание на  $O$ . Наблюдая  $S$ , адресат будет реконструировать внутреннее состояние  $C$  адресанта, а также объект  $O$ , вызвавший это состояние. Это ключевой переход от знаков-реакций (индексных знаков, обозначающих эмоции) к знакам, обладающим референцией и отсылающим к объекту  $O$ . Если же состояние  $C$  вызвано предшествующими действиями адресата  $A$ , то знак  $S$  получает референцию к предшествующим действиям адресата. К примеру,  $S$  может сообщить адресату, что его предшествующие действия  $A$  негативно повлияли на адресанта.

В предложенной схеме опора на те или иные компоненты коммуникации прямо соответствует функциям коммуникативного акта Р. Якобсона [Jakobson 1960]. Так, ориентация говорящего на реакцию адресата *R* соответствует *конативной* функции, ориентация на объект *O* — *референтивной* функции, а ориентация на собственное состояние *S* говорящего — *эмотивной* функции. При этом в одном жесте могут совмещаться несколько функций: иконический жест может иллюстрировать объект *O*, а особенности выполнения жеста могут добавлять оценку адресанта или воздействовать на реакции адресата. Важно отметить, что мимические движения имеют те же основные функции, что и жесты рук и ног. Как замечает Крейдлин, «лицо и многие связанные с ним мимические жесты и движения, такие как поднимать брови, закрывать глаза, надуть губы, поджать губы, закусить губу, наморщить лоб, нахмуриться, улыбка, поцелуи (в лоб, в щеку, в губы), и некоторые другие не только соотносятся с конкретными эмоциями, но и выполняют определенные коммуникативные и социальные функции» [Григорьева и др. 2001: 228]. Таким образом, чтобы указать место коммуникативных жестов в общей типологии, мы будем рассматривать более широкий спектр движений человека, относя мимику и коммуникативные движения к *жестам* в расширенном понимании этого термина.

### Объект исследования

В лаборатории нейрокогнитивных технологий Курчатовского комплекса НБИКС-технологий мы исследуем коммуникативное поведение людей в реальных эмоциональных ситуациях и стремимся предложить такое описание, которое может использоваться для управления поведением робота-компаньона. — При этом мы используем робота Ф-2<sup>1</sup> [Kotov et al. 2018; Зинина и др. 2018]. Классификацию коммуникативного поведения мы выполняем на основе мультимодального корпуса REC (Russian Emotional Corpus), содержащего размеченные в программе ELAN [Brugman, Russel 2004] видеозаписи различных эмоциональных диалогов на университетских экзаменах (295 фрагментов), в муниципальной службе одного окна (510 фрагментов), а также диалогов с информантами, которые занимаются каким-либо видом искусства, например хореографией или рисованием (10 фрагментов). Сбор и разметка корпуса осуществляются в нашей лаборатории на протяжении десяти лет [Kotov, Budyanskaya 2012]. В корпусе вручную размечаются речевые высказывания участников диалога. Для информанта (студента, клиента, респондента) размечаются движения его глаз, рук и губ. Мимикой глаз считаются коммуникативно значимые движения глаз и их окружения: век, бровей, мышц вокруг глаз, носа, например движения глазами вверх, вбок, расширение глаз, частые моргания и др. К движениям губ относятся улыбки, облизывания, прикусывания и др. Разметка движений рук выполняется на четырех слоях: это способ выполнения жеста, активный и пассивный органы, а также траектория движения. К жестам рук относятся всевозможные почесывания, поглаживания, манипуляции, иконические

<sup>1</sup> Сайт робота Ф-2: <http://f2robot.com/>.

знаки и др. На сегодняшний день в корпусе насчитывается более 304 000 аннотаций. При разметке видеозаписей мы специально отслеживаем согласованность экспертов — каждую видеозапись просматривают не менее двух человек, после чего обсуждаются спорные или неоднозначные моменты.

С точки зрения приведенного основания классификации для нас представляют интерес (а) означающие жестов — какие внутренние состояния  $S$  или цели  $R$  вызывают появление того или иного жеста, а также (б) означаемые жестов — какие формы выражения  $S$  позволяют выразить определенное означающее. Означающее жестов мы будем описывать через **коммуникативные функции**, а означаемые — через **инварианты** коммуникативных действий.

### Означаемое: коммуникативные функции

На основании результатов анализа базовых элементов разметки мы вводим дополнительный уровень разметки, который задает коммуникативную функцию мимического движения или жеста, если эта функция может быть определено установлена. Коммуникативные функции приписывают новые параметры аннотациям из базовой разметки корпуса, также классифицируют движения головы и тела человека, поскольку базовая разметка у этих элементов в корпусе отсутствует. Например, перпендикулярному движению руки или повороту головы может быть приписана функция *отрицание*. Это значит, что, двигая руку или поворачивая голову, адресант выражает смысл 'нет'. Мы выделяем 35 коммуникативных функций [Котов, Зинина 2015], которые частично уже перенесены на робота Ф-2.

Данная классификация является поисковой: мы продолжаем ее развитие на основе корпуса. Разработка классификации ведется по следующим принципам. Имеются коммуникативные функции, непосредственно связанные со структурой диалога (иллокутивными целями) и ориентированные на воздействие на говорящего, например *понимание-согласие-одобрение*, *апелляция*, *побуждение* и т. д. Для этих коммуникативных функций наиболее важен компонент коммуникации  $R$  — реакция адресата. Вторая группа коммуникативных функций связана с выражением внутреннего состояния адресата: *я-хезитация*, *я-смущение-фрустрация*, *я-радость* и т. д. Наиболее существенно для этой группы внутреннее состояние адресата  $S$ , это подчеркивается префиксом «я». Однако ключевой компонент коммуникации может меняться: часто говорящий демонстрирует фрустрацию, чтобы спровоцировать адресата на помощь, то есть в коммуникативной функции *я-смущение-фрустрация* ключевым компонентом становится реакция адресата  $R'$ , а внешнее выражение этой коммуникативной функции будет лучше контролироваться адресантом —  $S'$ . Третья группа коммуникативных функций связана с проявлением состояний ( $S$ ), преимущественно ориентированных на воздействие на адресата ( $R'$ ): *ты-пренебрежение*, *ты-позитивен-к-объекту*, *ты-попытка-успокоить*. Ключевым компонентом коммуникации здесь будет реакция адресата —  $R'$ , однако в схожих случаях адресант сам может «наплеватьски» относиться к некоторому объекту и выражать это в своих жестах (*ты-пренебрежение*). Четвертая

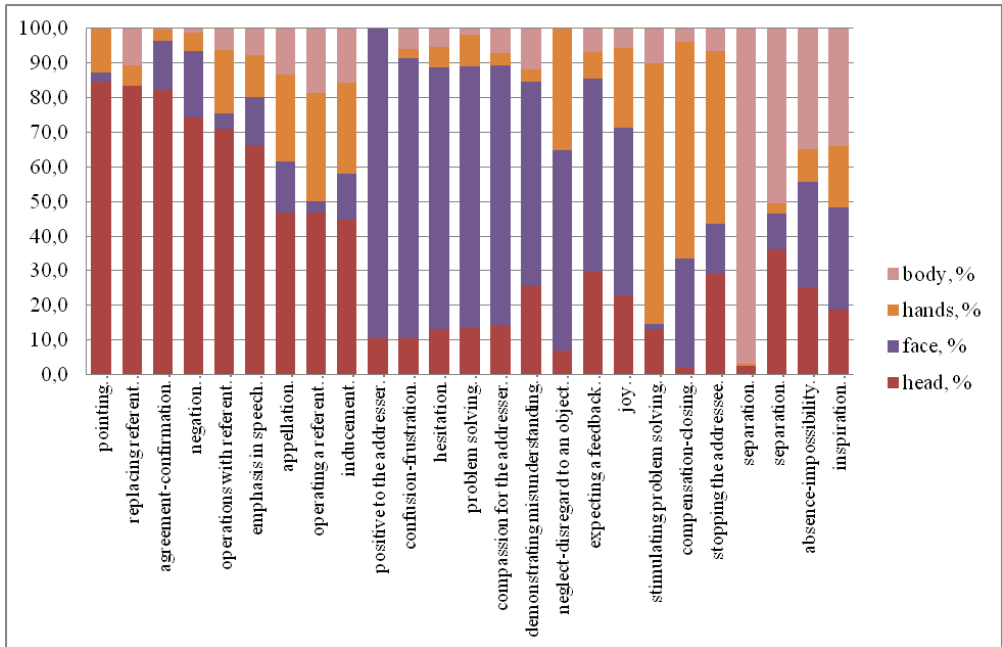


Рис. 2. Выражение коммуникативных функций

группа коммуникативных функций связана с обозначаемым объектом — реальным или воображаемым: *речь-референт*, *выбор-варианта* и т. д. Ключевым компонентом здесь будет объект *O*. Как видно из классификации, намерения выразить свое внутреннее состояние *C*, выполнить референцию к объекту *O* или воздействовать на адресата *R'* — могут в разном соотношении влиять на появление той или иной коммуникативной функции. Если же сумма этих составляющих превышает некоторый порог, коммуникативная функция считается активизировавшейся и далее может проявиться через те или иные жесты.

Каждая коммуникативная функция может выражаться с помощью различных движений — одно означаемое может иметь различные означающие. Более того, отдельное означающее может быть по-разному распределено между несколькими исполнительными органами — выражаться преимущественно с помощью головы, рук, мимики или тела (рис. 2). Например, такая функция, как *апелляция*, в 46,1% случаев выражается с помощью движений головы, в 24,5% — жестами рук, в 15,8% — с помощью мимики и в 13,6% — движениями тела. При этом *компенсация-закрытие* жестами рук выражается в 62,7% случаев, с помощью мимики — в 31,6% случаев, движениями тела — в 3,8%, а движениями головы — лишь в 1,9% случаев. *Я-hesитация*, наоборот, выражается преимущественно мимическими движениями — в 75,7% случаев, с помощью движений головы — в 13,1%, в 6% случаев — с помощью тела и лишь в 2,7% — движениями рук. Коммуникативная функция может выражаться комплексно — через мимику и движение нескольких элементов тела либо редуцированно — с помощью движения всего



Таблица 1

## Сочетаемость коммуникативных функций

	Функция	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	понимание-согласие-одобрение	15	0	2	7	2	2	1	27	1	5	0	10	7	0	10
2	отрицание-несогласие-возражение	0	12	1	4	3	0	4	9	1	19	1	4	10	0	7
3	апелляция	2	1	48	7	5	3	1	7	0	6	2	8	27	0	14
4	ожидание-обратной-связи	7	4	7	41	1	1	2	17	1	4	2	15	35	0	12
5	остановка-адресата	2	3	5	1	7	0	1	1	1	9	0	5	0	0	5
6	отсутствие-невозможность	2	0	3	1	0	10	0	12	0	15	1	4	3	1	6
7	демонстрация-непонимания	1	4	1	2	1	0	10	6	0	7	1	6	9	0	4
8	я-смущение-фрустрация	27	9	7	17	1	12	6	28	0	3	0	31	34	1	14
9	я-хезитация	1	1	0	1	1	0	0	0	3	1	1	8	18	5	4
10	я-размышление	5	19	6	4	9	15	7	3	1	58	3	20	47	13	7
11	ты-пренебрежение	0	1	2	2	0	1	1	0	1	3	5	9	12	0	1
12	дистанцирование	10	4	8	15	5	4	6	31	8	20	9	8	32	3	17
13	речь-эмфаза	7	10	27	35	0	3	9	34	18	47	12	32	82	2	34
14	стимулирование	0	0	0	0	0	1	0	1	5	13	0	3	2	7	3
15	компенсация-закрытие	10	7	14	12	5	6	4	14	4	7	1	17	34	3	111

одним органом [Зинина и др. 2018]. Из таблицы 1 видно, что коммуникативные функции, как правило, выражаются с помощью комплексного паттерна, например функция *компенсация-закрытие* в 111 случаях выполняется двумя или более исполнительными органами<sup>2</sup>. Это означает, что мимическое выражение *компенсации-закрытия* — сжатие губ, облизывание и др. — в реальном поведении сопровождается жестовым выражением — закрывающими жестами, манипулированием, почесываниями и др.

В реальном коммуникативном поведении наблюдаются случаи, когда информант одновременно выражает сразу несколько коммуникативных функций, в этом случае мы говорим об их наложении. Пересечения основных коммуникативных функций приведены в таблице 1. Всего таких пересечений в корпусе насчитывается 1810. При качественном анализе подобных пересечений можно выявить интересные тенденции, например, функция *я-размышление*, как правило, пересекается с такими функциями, как *речь-эмфаза*, *дистанцирование*, *отрицание-несогласие-возражение*, *отсутствие-невозможность*, *стимулирование*. Причем *я-размышление* в таких примерах выражается преимущественно с помощью мимических паттернов, а другие коммуникативные функции — с помощью движений рук, головы и тела. Говорящий при этом может смотреть в сторону, демонстрируя *я-размышление*, но с помощью рук продолжать взаимодействовать с собеседником: демонстрировать отрицание или эмфазу. Функция *компенсация* пересекается со следующими функциями: *отрицание-несогласие-возражение*, *я-размышление*, *понимание-согласие-одобрение*, *ожидание-обратной-связи*, *апелляция*, *я-смущение-фрустрация*,

<sup>2</sup> Руки рассматриваются здесь как один исполнительный орган.

*дистанцирование* и *речь-эмфаза*. Говорящий, к примеру, может демонстрировать отрицание головой и при этом закрываться рукой, как бы компенсируя ущерб социальному лицу адресата в терминах теории вежливости [Brown, Levinson 1987].

### Означающее: инварианты коммуникативных действий

Означающее лингвистических знаков допускает существенную вариативность: буква может выражаться через множество графических вариантов, а фонема — через множество аллофонов. Означающее жестов еще более разнообразно, однако на этом множестве можно выделить общую основу, которую мы будем называть **инвариантом** жеста. Инвариант — это общий принцип выполнения жеста, который необходим для передачи сообщения адресату и может проявляться в движениях разных исполнительных органов человека.

Рассмотрим понятие инварианта на примере *ожидания-обратной-связи*, где адресант выполнил некоторое действие и надеется получить обратную связь от адресата. Если обратная связь отсутствует, адресант старается подтолкнуть адресата на ответную реакцию. Инвариант действий адресанта можно описать так: адресант намерен получить от адресата некоторое сообщение для выполнения дальнейших действий и акцентирует это намерение. Этот инвариант может выражаться с помощью четырех групп жестов.

1. Адресант демонстрирует, что воспринимает реакцию адресата визуально — он хочет увидеть, что в ответ сделает адресат. В этом случае адресант может перевести взгляд на адресата в конце своего высказывания. В следующем примере адресант произносит ответ (в это время его голова и взгляд опущены), а затем поднимает голову и смотрит на адресата (рис. 3).<sup>3</sup>

Говорящий может смотреть в сторону и переводить взгляд на адресата из положения сбоку. Если к началу движения взгляд говорящего уже направлен на адресата, то говорящий вынужден акцентировать свое внимание по-другому. В следующем примере информант демонстративно поднимает брови (рис. 4).

Таким образом, если инвариант жеста не может быть выражен через направление взгляда (взгляд уже направлен на собеседника), адресант пытается выразить инвариант — акцентировать свое намерение — другими средствами, в данном случае — через движение бровей.

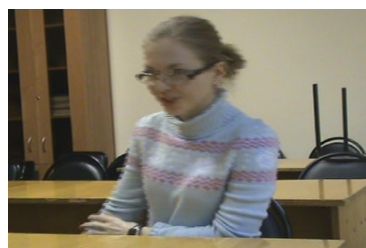
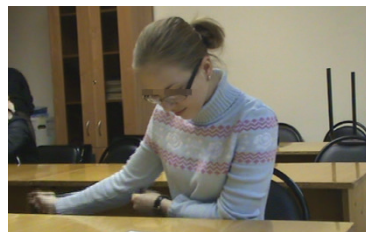


Рис. 3,  
а. 20081213-upr-a6 00:11.914<sup>3</sup>  
б. 20081213-upr-a6 00:13.134

<sup>3</sup> Корпус является закрытым, однако в научных целях можно ознакомиться с указанными примерами, если связаться с авторами работы.

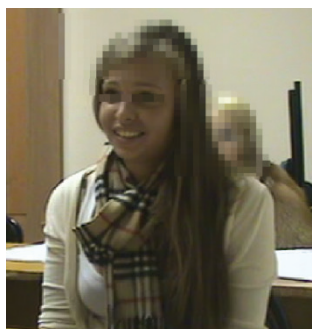
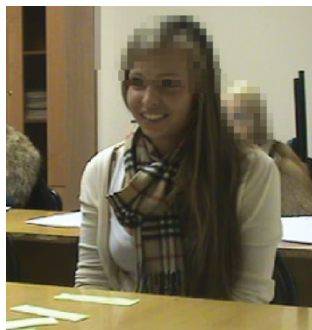


Рис. 4

а. 20081230-b4-furekl-m  
08:43.630, б. 20081230-b4-  
furekl-m 08:44.565

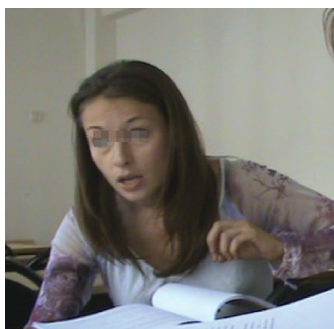
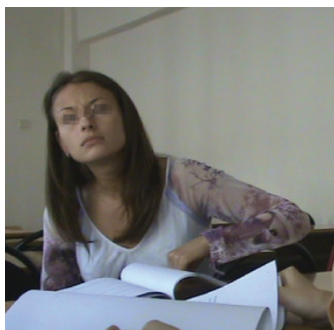


Рис. 5

а. 20080717-f-psy-m 05:15.570,  
б. 20080717-f-psy-m 05:01.890

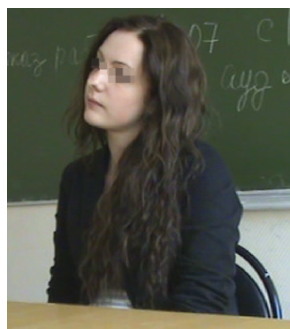
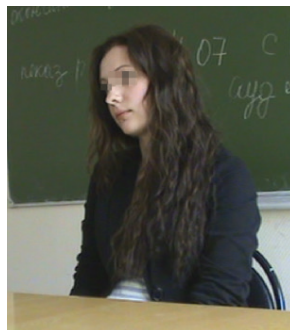


Рис. 6

а. 20080717-c16-psy-m  
04:12.920, б. 20080717-c16-  
psy-m 04:14.440

Акцентируя свое внимание на реакции собеседника, адресант может «присматриваться» и «прислушиваться»: прищуриваться, хмурить брови, сжимать губы и наклоняться немного вперед (рис. 5, а). Также говорящий может существенным образом сокращать дистанцию с собеседником (рис. 5, б). Мы предполагаем, что мимическое выражение ( $S$  — прищуривание глаз), характерное для рассматривания ( $C$ ) плохо различимого объекта, используется в данном случае как контролируемое действие  $S'$  для запроса обратной связи ( $R$ ) у адресата.

2. В ожидании реакции собеседника говорящий может демонстрировать свое бездействие: 'я сделал все, что нужно, и теперь жду тебя'. К примеру, продолжая рассматривать собеседника, говорящий может перекидывать голову с одного плеча на другое, как бы демонстративно меняя позу при бездействии (рис. 6).

Демонстрируя бездействие, говорящий также может дистанцироваться от адресата (рис. 7).

3. В инварианте указано, что обратная связь необходима адресанту, чтобы перейти к следующему действию. Стремясь продемонстрировать свою готовность, говорящий может останавливать жест в середине выполнения (рис. 8) или «группироваться», например складывать руки (рис. 9), показывая возможность перейти к действию.



Рис. 7

а. 20081220-zhurn-c2 00:51.357,  
б. 20081220-zhurn-c2 00:51.357

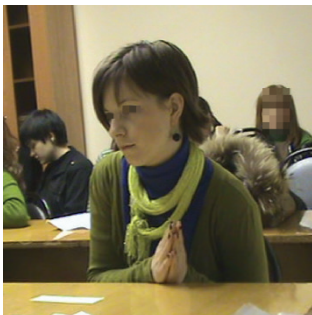


Рис. 8. 20081219-zhurn-a03  
01:45.363

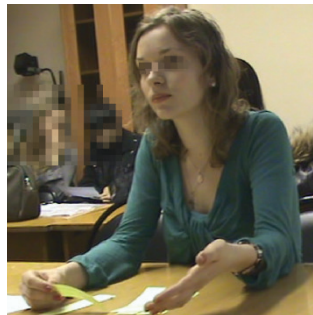


Рис. 9. 20081219-zhurn-a0.10  
00:41.065

4. Акцентируя свое намерение получить обратную связь, адресант может демонстрировать коммуникативные действия, близкие к подтверждению или вопросу. После своего высказывания говорящий может несколько раз кивать вперед или вбок<sup>4</sup>, как бы подтверждая адресату свое высказывание. Говорящий также может выполнять короткие движения головой вверх («кивки» вверх<sup>5</sup>), характерные для вопроса, но в данном случае — запрашивающие реакцию адресата.

Таким образом, анализ размеченных данных позволяет выделить спектр типичных представителей отдельной коммуникативной функции и их инвариант. Такое описание позволяет хранить для жеста несколько вариантов. При синтезе жестов несколько коммуникативных функций могут одновременно выполняться на разных исполнительных органах робота, что существенно разнообразит его коммуникативное поведение.

## Система управления роботом Ф-2

В рамках нашей работы мы стремимся сделать робота, коммуникативное поведение которого максимально приближено к коммуникативному поведению человека. Робот должен иметь возможность реализовывать основные особенности естественной коммуникации: выполнять комплексные и редуцированные жесты, гибко комбинировать мимические и жестовые движения, характерные для различных коммуникативных функций, а также фиксировать основные контексты для подбора релевантных ситуации мимических и жестовых реакций.

Проект робота Ф-2 разрабатывается в Курчатовском институте [Kotov et al. 2018; Зинина и др. 2018]. Робот обладает гибкой эмоциональной моделью, основанной

<sup>4</sup> Например, 20081219-zhurn-a08 01:34.

<sup>5</sup> Например, 20081219-zhurn-a09 2:03.

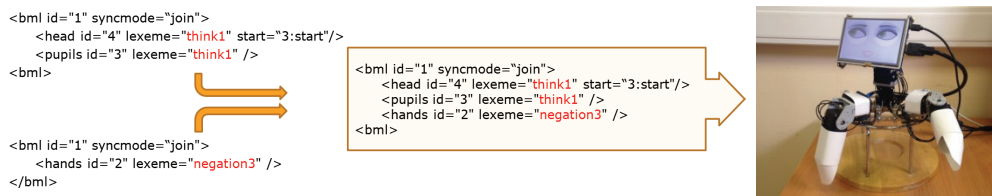


Рис. 10. Комбинирование тегов из различных пакетов BML в режиме join

на работе 13 негативных сценариев (Опасность, Обман, Неадекватность антагониста, Эмоциональность антагониста и т. д.) и 23 позитивных (Забота, Контроль над ситуацией, Служение и т. д.) [Kotov 2003, 2012]. В составе робота важную роль играет компонент понимания текста — семантический парсер [Kotov et al. 2018]. С помощью парсера восстанавливается семантическая структура (смысл) входящего высказывания, чтобы на ее основе робот смог выбрать ответную коммуникативную реакцию. Кроме того, проект робота Ф-2 включает разработку системы управления коммуникативным поведением для жестовых и мимических реакций робота на входящие тексты. На основании эмоциональной модели строится коммуникативное поведение робота-компаньона.

Для практической реализации этой модели мы разрабатываем такую архитектуру, которая в зависимости от активации определенного сценария (коммуникативной функции) продуцирует коммуникативные реакции, вызывающие BML-пакеты [Kopp, Krenn et al. 2006; Vilhjálmsón, Cantelmo et al. 2007]. В зависимости от активации определенного сценария робот может порождать различные ответы — использовать речь и комбинации движений разных исполнительных органов. Выражение коммуникативной функции снижает активизацию соответствующего сценария, однако, если эта активизация все еще находится выше порогового значения, робот может дополнительно выразить коммуникативную функцию с помощью других элементов поведения.

Для каждой коммуникативной функции в базе хранится несколько BML-пакетов, которые кодируют различные способы ее выражения. На сегодняшний день в базе сохранено 577 BML-пакетов. BML-пакет включает набор тегов, связанных с коммуникативной функцией (рис. 10). С помощью тега описываются исполнительные органы (глаза, голова, речь, левая и правая руки), задействованные в жесте, соответствующем определенной коммуникативной функции.

Робот в своем поведении гибко комбинирует определенные мимические и жестовые средства, что достигается благодаря особой архитектуре. Такая архитектура позволяет по-разному формировать очередь из BML-пакетов за счет того, что у каждого BML-пакета присутствует параметр сочетания с предшествующими пакетами — `syncmode`. Установка значений данного параметра позволяет по-разному группировать BML-пакеты. Например, режим `join` объединяет теги BML-пакета, и они вместе конкурируют в очереди за возможность занять необходимые исполнительные органы робота. Обработка тегов из пакета начинается, как только освободятся все из указанных в пакете исполнительных органов. При моделировании



коммуникативного поведения на роботе-компаньоне выбор определенного способа выражения коммуникативной функции зависит от свободного исполнительного органа, — например, если с помощью рук выражается остановка адресата, то апелляция, скорее всего, будет выражаться головой или мимикой — подобная комбинация коммуникативных функций характерна для естественного поведения.

### Заключение

Жесты могут быть описаны как коммуникативные знаки, распределенные от неконтролируемых рефлекторных или эмоциональных реакций ( $S$ ) до контролируемых действий ( $S'$ ), ориентированных на определенное воздействие ( $R'$ ). Жесты могут быть ориентированы на (а) выражение внутреннего состояния говорящего, (б) воздействие на адресата, в частности на регулирование ситуации коммуникации, (в) обозначение объекта реального мира или оценку этого объекта и (г) обозначение или оценку предшествующих действий адресата. По этому основанию в архитектуре компьютерного агента жесты можно представить как выражающие определенную коммуникативную функцию. Коммуникативная функция, таким образом, заменяет для жестов традиционное понятие означаемого.

Означающее жестов также имеет сложную структуру. Намеренный жест может использовать означающее, заимствованное у эмоционального бесконтрольного жеста. Жест может проявляться через разные исполнительные органы: брови, голову, руки, тело и т. д. В компьютерной модели целесообразно считать, что жест проявляется через те исполнительные органы, которые не заняты другими жестами и не блокированы другими ограничителями. Таким образом, жест может исполняться по-разному, однако в означающем жестов можно выделить общий инвариант — при различных типах исполнения жеста говорящий стремится выполнить некоторый общий компонент движения, который объединяет различные варианты исполнения жеста.

Предложенная модель выполнения коммуникативных жестов, основанная на активации коммуникативных функций и использовании инвариантов жестов, позволяет гибко сочетать в поведении различные жесты, использовать разные способы выражения коммуникативных функций в зависимости от свободных исполнительных органов. В целом это ведет к обогащению коммуникативного поведения робота Ф-2, реализующего данную модель. Подробнее познакомиться с архитектурой и функциями робота Ф-2 можно на сайте проекта: <http://f2robot.com/>.

### Литература

Григорьева С. А., Григорьев Н. В., Крейдлин Г. Е. Словарь языка русских жестов. М. : Языки русской культуры; Вена : Венский славистический альманах, 2001. 256 с.

Гришина Е. А. Русская жестикуляция с лингвистической точки зрения. М. : Издательский дом ЯСК, 2017. 744 с.

Зинина А. А., Аринкин Н. А., Зайдельман Л. Я., Котов А. А. Разработка модели коммуникативного поведения робота ф-2 на основе мультимодального корпуса «РЕС» // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 30 мая — 2 июня 2018 г.). Вып. 17 (24). 2018. С. 831–844.

Котов А. А., Зинина А. А. Функциональный анализ невербального коммуникативного поведения // Компьютерная лингвистика и интеллектуальные технологии. Вып. 14. Т. 1. М. : РГГУ, 2015. С. 299–310.

Лоренц К. Агрессия (так называемое зло). СПб. : Амфора, 2001.

Beuter N., Spexard T., Lutkebohle I., Peltason J., Kummert F. Where is this? — gesture based multimodal interaction with an anthropomorphic robot: 8th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2008). 2008 P. 585–591. DOI: 10.1109/ICHR.2008.4756009.

Breazeal C. Emotive qualities in lip-synchronized robot speech // Advanced Robotics / eds. R. D. Lane, L. Nadel. 2003. Iss. 2. Vol. 17. P. 97–113.

Breazeal C., Scassellati B. Robots that imitate humans // Trends in Cognitive Sciences. 2002. Vol. 6 (11). P. 481–487.

Brown P., Levinson S. C. Politeness: Some Universals in Language Usage (Studies in Interactional Sociolinguistics). Cambridge, 1987.

Brugman H., Russel A. Annotating Multimedia Multi-modal resources with ELAN: Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004). 2004. P. 2065–2068.

Efron D. Gesture and Environment. N. Y. : King's Crown Press, 1941 (2nd ed. 1972: Gesture, Race and Culture).

Hindriks K., Neerincx M. A., Vink M. The icat as a natural interaction partner: Playing go fish with a Robot // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012. Vol. 7068 LNAI. P. 212–231.

Jakobson R. Linguistics and poetics // Style in language. MA : MIT Press, 1960. P. 350–377.

Kendon A. How gestures can become like words // Cross-cultural perspectives in non-verbal communication / ed. F. Poyatos. Toronto, 1988. P. 131–141.

Klamer T., Allouch S. B., Heylen D. Adventures of Harvey: Use, acceptance of and relationship building with a social robot in a domestic environment // Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. 2011. Vol. 59 LNICST. P. 74–82.

Kopp S., Krenn B., Marsella S., Marshall A., Pelachaud C., Pirker H., Thórisson K., Vilhjálmsson H. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language // Intelligent Virtual Agents. 2006. P. 205–217.

Kotov A., Budyanskaya E. The Russian Emotional Corpus: Communication in Natural Emotional Situations // Computational Linguistics and Intellectual Technologies. 2012. Iss. 11 (18). Vol. 1. P. 296–306. RSUH, M.



Kotov A. A., Zaidelman L. Y., Arinkin N. A., Zinina A. A., Filatov A. A. Frames revisited: automatic extraction of semantic patterns from a natural text // *Computational Linguistics and Intellectual Technologies*. 2018. Iss. 17 (24). P. 357–367.

Leite I., Pereira A., Martinho C., Paiva A. Are Emotional Robots More Fun to Play With?: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1–3, 2008. P. 77–82.

McNeill D. *Gesture and Thought*. Chicago : University of Chicago Press, 2005. 328 p.

Mori M. The uncanny valley // *Energy*. 1970. Vol. 7 (4). P. 33–35.

Sung J.-Y., Guo L., Grinter R. E., Christensen H. I. “My Roomba Is Rambo”: Intimate Home Appliances // *International Conference on Ubiquitous Computing* / eds. J. Krumm et al. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007. P. 145–162.

Vilhjálmsón H., Cantelmo N., Cassell J. E., Chafai N., Kipp M., Kopp S., Mancini M., Marsella S., Marshall A., Pelachaud C., Ruttkay Z., Thórisson K., van Welbergen H., van der Werf R. The Behavior Markup Language // *Recent Developments and Challenges in Intelligent Virtual Agents*. 2007. P. 99–111.

**A. Zinina, L. Zaydelman, N. Arinkin, A. Kotov**

*Kurhcator Institute, Moscow, Russia*

*zinina\_aa@nrcki.ru*

## **APPROACHES TO THE CLASSIFICATION OF COMMUNICATIVE GESTURES: THEORY AND APPLICATION TO COMPANION ROBOTS**

The paper presents an approach to the classification of communicative gestures. We describe the gestures such that their signified can be expressed with the help of various signifiers. We classify gestures through the identification of communicative functions (the signified) and an invariant of the performance of the gesture (the signifier). A certain communicative function can be expressed by gestures with different sets of executive organs: with the help of facial expressions, movements of the head, hands or body. By distinguishing communicative functions, we solve the task of modeling complex communicative behavior of a companion robot. In the paper we explain the need of distinguishing communicative functions in the analysis of real human behavior in emotional situations (based on the multimodal corpus REC), and describe the contexts and invariants of communicative functions. Using the case of the *feedback expectation* communicative function we describe the “focus on the interlocutor’s response” invariant and different variations of the corresponding gestures: (1) eye focus on the addressee, (2) demonstration of one’s inaction while waiting for the addressee, (3) demonstration of the readiness to make a move, (4) nods in support of one’s own thesis or the interrogative movements which are requesting the interlocutor’s reaction. We also describe the robot’s control architecture which simulates the basic features of natural communication — communicative

functions and gestures. The description of the invariant and gestures for a certain communicative function allows us to adjust the robot's communicative responses to the incoming statement.

*Key words:* Multimodal corpus, gestural communication, iconic gestures, robot companions.

## References

Beuter N., Spexard T., Lutkebohle I., Peltason J., Kummert F. Where is this? — gesture based multimodal interaction with an anthropomorphic robot. *8th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2008)*, 2008, pp. 585–591. DOI: 10.1109/ICHR.2008.4756009.

Breazeal C. Emotive qualities in lip-synchronized robot speech. *Advanced Robotics*. Eds. R. D. Lane, L. Nadel, 2003, iss. 2, vol. 17, pp. 97–113.

Breazeal C., Scassellati B. Robots that imitate humans. *Trends in Cognitive Sciences*, 2002, vol. 6 (11), pp. 481–487.

Brugman H., Russel A. Annotating Multimedia Multi-modal resources with ELAN. *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*, 2004, pp. 2065–2068.

Efron D., van Veen S. Gesture, race and culture, 1972.

Grigor'eva S. A., Grigor'ev N. V., Krejdlin G. E. *Slovar' jazyka russkih zhestov* [A dictionary of the language of Russian gestures]. Moscow-Vienna, Languages of Russian culture Publ.; Viennese Slavic almanac Publ., 2001. 256 p. (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, LRC Publishing House, Languages of Slavic Culture Publ., 2017. 744 p. (In Russ.)

Hindriks K., Neerinx M. A., Vink M. The icat as a natural interaction partner: Playing go fish with a Robot. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7068 LNAI, pp. 212–231.

Jakobson R. Linguistics and poetics. *Style in language*. MA, MIT Press, 1960, pp. 350–377.

Kendon A. How gestures can become like words. *Cross-cultural perspectives in non-verbal communication*. Ed. F. Poyatos. Toronto, 1988, pp. 131–141.

Klamer T., Allouch S. B., Heylen D. Adventures of Harvey: Use, acceptance of and relationship building with a social robot in a domestic environment. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 2011, vol. 59 LNICST, pp. 74–82.

Kopp S., Krenn B., Marsella S., Marshall A., Pelachaud C., Pirker H., Thórisson K., Vilhjálmsson H. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. *Intelligent Virtual Agents*, 2006, pp. 205–217.

Kotov A. A., Zinina A. A. [Functional analysis of nonverbal communicative behavior]. *Computational Linguistics and Intellectual Technologies*. Moscow, RSUH Publ., 2015, iss. 14(1), pp. 299–310. (In Russ.)

Kotov A., Budyanskaya E. The Russian Emotional Corpus: Communication in Natural Emotional Situations. *Computational Linguistics and Intellectual Technologies*, 2012, iss. 11 (18), vol. 1, pp. 296–306. RSUH, M.

Kotov A. A., Zaidelman L. Y., Arinkin N. A., Zinina A. A., Filatov A. A. Frames revisited: automatic extraction of semantic patterns from a natural text. *Computational Linguistics and Intellectual Technologies*, 2018, iss. 17 (24), pp. 357–367.

Leite I., Pereira A., Martinho C., Paiva A. Are Emotional Robots More Fun to Play With? *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1–3, 2008*, pp. 77–82.

Lorenz K. *Agressiya (tak nazyvaemoe zlo)* [On aggression]. St. Petersburg, Amfora Publ., 2001. (In Russ.)

McNeill D. *Gesture and Thought*. Chicago, University of Chicago Press, 2005. 328 p.

Mori M. The uncanny valley. *Energy*, 1970, vol. 7 (4), pp. 33–35.

Sung J.-Y., Guo L., Grinter R. E., Christensen H. I. “My Roomba Is Rambo”: Intimate Home Appliances. *International Conference on Ubiquitous Computing*. Eds. J. Krumm et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 145–162.

Vilhjálmsón H., Cantelmo N., Cassell J. E., Chafai N., Kipp M., Kopp S., Mancini M., Marsella S., Marshall A., Pelachaud C., Ruttkay Z., Thórisson K., van Welbergen H., van der Werf R. The Behavior Markup Language, Recent Developments and Challenges in Intelligent Virtual Agents, 2007, pp. 99–111.

Zinina A., Arinkin N., Zaydelman L., Kotov A. [Development of communicative behavior model for f-2 robot basing on «REC» multimodal corpora]. *Computational Linguistics and Intellectual Technologies*, 2018, iss. 17 (24), pp. 831–844. (In Russ.)

**С. И. Переверзева, Н. А. Ермолаева, А. В. Зуева, Е. А. Слепак**

НИУ «Высшая школа экономики»

(Россия, Москва)

*P\_Sveta@hotmail.com, tasha.ermolaeva@gmail.com,*

*an5200@ya.ru, janenikel16@gmail.com*

## **DE PROFUNDIS: ПРОБЛЕМЫ ГЛУБОКОЙ РАЗМЕТКИ МУЛЬТИМЕДИЙНОГО РУССКОГО КОРПУСА И ПУТИ РЕШЕНИЯ\***

Статья написана по материалам доклада, представленного на конференции «Слово и жест», посвященной памяти Е. А. Гришиной (Гришинские чтения), 2018 г. Основным предметом внимания служит жестовая разметка Мультимедийного русского корпуса, над продолжением которой в настоящее время работают авторы статьи. Важным принципом разметки является единообразие решений, принимаемых для аннотирования сходных друг с другом явлений [Гришина 2008]; при этом единственным надежным источником для спорных случаев оказывается разметка, выполненная самой Е. А. Гришиной. Поскольку никакой источник не содержит прямых указаний, как следует поступать разметчику МУРКО в каждом спорном случае, авторы статьи пытаются, опираясь на имеющиеся данные, обнаружить те закономерности, которыми, возможно, руководствовалась Е. А. Гришина при глубоком аннотировании.

В статье описываются три такие закономерности: 1) правило выбора между жёстами *широко раскрыть глаза* и *поднять брови* (оба в значении ‘испуг’) — показано, что выбор здесь определяется главным образом относительной отчетливостью исполнения этих двух жестов; 2) правило выбора между значениями ‘подтверждение’ и ‘подчеркнуть эмфазу’ для жеста *закрыть глаза* и 3) правило выбора между теми же двумя значениями для жеста *кивнуть*. В обоих случаях ‘подтверждение’ выбирается, если реплика, которую сопровождает жест, является ответной; кроме того, для *кивнуть* также может быть важно наличие в сопровождаемой реплике смысловой отсылки к предыдущим словам жестикулирующего.

**Ключевые слова:** Мультимедийный русский корпус, жест, глубокая разметка, правила выбора, инструкция.

---

\* Работа выполнена в рамках Программы фундаментальных исследований НИУ ВШЭ.

## Введение

Мультимедийный русский корпус (МУРКО) был создан Е. А. Гришиной на базе Национального корпуса русского языка около 10 лет назад. Аннотирование входящих в него видеофрагментов (клипов) бывает трех видов — автоматическая разметка, полуавтоматическая и ручная (глубокая) [Гришина 2010]. Последний тип образует разметка разных параметров жестов, исполняемых в том или ином клипе, и разметка речевых актов.

С 2016 г. группа студентов Школы лингвистики НИУ ВШЭ под руководством к.ф.н., н.с. НИУ ВШЭ С. И. Переверзевой продолжает работу над глубокой жестовой разметкой, начатой Е. А. Гришиной.

Жестовая разметка включает в себя, наряду с прочим, параметры «название жеста» и «значение жеста». Задача разметчиков состоит в том, чтобы правильно установить соответствие между конкретным жестом и конкретным значением. Основным ориентиром и критерием правильности в настоящее время служит разметка клипов, сделанная Е. А. Гришиной, поскольку, как справедливо отмечала она сама, «большое значение для МУРКО будет иметь единообразие в работе будущих разметчиков. <...> Если при разметке какого-то одного явления принято неточное, если не сказать неправильное решение, то именно это решение, а никакое другое, должно быть принято и для всех остальных аналогичных явлений» [Гришина 2008: 208–209]. Главное затруднение связано с тем, что принципы, которыми следует руководствоваться разметчику в сложных случаях выбора (между синонимичными жестами или, наоборот, между разными значениями одного жеста), нигде не описаны в явном виде. На сегодняшний день сведения, необходимые для разрешения сомнений, можно получить из трех источников: (а) из работ Е. А. Гришиной и бесед с ее коллегами и учениками; (б) из так называемых «тропинок», по которым разметчика автоматически ведет специальная программа, обращающаяся к базе данных МУРКО («рабочее место разметчика» [Гришина 2008: 209]); (в) из собственных выводов, полученных при анализе материала уже имеющейся глубокой разметки МУРКО.

Важность источника (в) весьма велика, поскольку существуют такие вопросы, на которые ни работы Е. А. Гришиной, ни база данных не дают однозначного ответа. Например, «В чем разница между жестами *поднять брови* и *сделать большие глаза* (оба со значением ‘испуг’)?»; «Чем различаются жесты *ткнуть пальцем* и *показать пальцем*?»; «Когда жест *поднять брови* имеет значение ‘лукавство’, а когда ‘насмешка’?». Во всех подобных случаях разметчик оказывается примерно в той же ситуации, что и исследователь письменных памятников: опираясь на ограниченный материал, он пытается обнаружить определенные закономерности в изучаемой системе.

Ниже мы опишем три закономерности, которые нам удалось выявить при изучении глубокой жестовой разметки, выполненной Е. А. Гришиной. Первая связана с синонимией жестов *поднять брови* и *широко раскрыть глаза* (оба могут передавать значение ‘испуг’), вторая и третья — со значениями ‘подтверждение’ и ‘подчеркнуть эмфазу’, выделяемыми у жестов *закрыть глаза* и *кивнуть*.

## 1. 'Испуг': *поднять брови* vs. *широко раскрыть глаза*

Когда один из авторов, рассказывая коллегам о жестовом аннотировании МУР-КО, приводил в пример трудность выбора между *поднять брови* 'испуг' и *широко раскрыть глаза* 'испуг', те нередко удивлялись: «А разве можно широко раскрыть глаза, не поднимая брови?» Действительно, один мимический жест настолько тесно связан с другим — с точки зрения и физиологии, и семантики, — что естественно было бы ожидать, что в разметке клипов они всегда будут представлены вместе. Однако Е. А. Гришина приняла другое решение: из 17 клипов только в 2 указаны и жест *широко раскрыть глаза*, и жест *поднять брови*, в 1 клипе — только *поднять брови* (без *широко раскрыть глаза*), в 14 (sic!) — только *широко раскрыть глаза*.



Рис. 1



Рис. 2



Рис. 4



Рис. 3



Рис. 5



Рис. 6



Рис. 7

Наши совместные наблюдения показывают, что ключевым параметром здесь служит физическая реализация жестов, а именно степень поднятия бровей по сравнению со степенью раскрытия глаз. Среди 14 клипов, в разметке которых присутствует только *широко раскрыть глаза*, в отдельных случаях представлены такие разновидности мимики, когда глаза раскрываются, а брови (вопреки интуитивным ожиданиям!) остаются неподвижны (рис. 1–4).

Иногда брови могут немного подниматься,





Рис. 8



Рис. 9



Рис. 10

но главным компонентом реализации жеста остается ярко выраженное раскрытие глаз (рис. 5–7).

В тех же случаях, когда размечаются оба жеста, каждое движение — и раскрытие глаз, и поднятие бровей — видно отчетливо (рис. 8–10).

## 2. *Закрывать глаза*: ‘подтверждение’ vs. ‘подчеркнуть эмфазу’

Проблема разграничения этих, как кажется, не очень близких друг другу значений, возникла, когда нам понадобилось размечать диалоги типа (1) — *Но это еще не все*<sup>1</sup> (*закрывает глаза*). — *Не все?* — *Не все* (*снова закрывает глаза*) («Тот самый Мюнхгаузен», реж. М. Захаров). Здесь говорящий закрывает глаза, как бы давая понять слушающему, что положение дел именно таково, как он описывает («я действительно рассказал еще не все»). Неочевидно, как следует интерпретировать эти закрывания глаз — как подтверждение истинности собственных слов (‘подтверждение’) или как их усиление, выделение из контекста диалога (‘эмфаза’).

Чтобы понять, какое решение принимала в подобных случаях Е. А. Гришина, мы проанализировали все размеченные ею клипы, содержащие жест *закрывать глаза* в значениях ‘подтверждение’ (11 клипов) и ‘подчеркнуть эмфазу’ (7 клипов). Мы пришли к следующему выводу: выбор значения зависит от того, является реплика, сопровождаемая жестом *закрывать глаза*, ответной (или, в терминах статьи [Баранов, Крейдлин 1992: 87], иллокутивно зависимой) или не является. Если жестикулирующий закрывает глаза в ответ на какую-то реплику или какой-то жест (то есть если закрывание глаз выступает как аналог кивка<sup>2</sup>, своего рода невербальное «да»), то из двух вариантов следует выбрать ‘подтверждение’ (10 клипов из 11 в размеченном материале). *Закрывать глаза* ‘подчеркнуть эмфазу’ в таких ситуациях не используется (7 клипов из 7). Примечательно, что если жест *закрывать глаза*

<sup>1</sup> Здесь и далее подчеркиванием выделяются те фрагменты высказывания, которые сопровождаются соответствующей жестикуляцией.

<sup>2</sup> О том, что жест *закрывать глаза* может замещать кивок, см. в [Гришина 2017: 515–516].



‘подтверждение’ сопровождается репликой жестикулирующего, то он, как правило, приходится на ее начало, в то время как *закрыть глаза* ‘подчеркнуть эмфазу’ не употребляется в отрыве от речи и может встречаться в любой позиции реплики.

Исходя из этого наблюдения, мы приняли решение первому закрыванию глаз в диалоге:

- (1) (...*еще не все*) приписать значение ‘подчеркнуть эмфазу’, а второму (*Не все*) — значение ‘подтверждение’ (в ответ на переспрос).

### 3. *Кивнуть*: ‘подтверждение’ vs. ‘подчеркнуть эмфазу’

Поскольку, как было сказано выше, жест *закрыть глаза* в ряде ситуаций выступает как заместитель жеста *кивнуть*, следовало бы ожидать, что выбор между значениями ‘подтверждение’ (142 клипа) и ‘подчеркнуть эмфазу’ (197 клипов) здесь будет определяться той же закономерностью, что и в случае *закрыть глаза*: ответный, иллокутивно зависимый кивок означает ‘подтверждение’, а иллокутивно независимый эмфатически выделяет некий фрагмент реплики. Однако для кивков это правило выполняется лишь частично. Действительно, кивок как невербальное «да» в ответ на чью-то реплику никогда не встречается вместе со значением ‘подчеркнуть эмфазу’. Но если говорящий кивает на словах внутри своей собственной реплики, то, помимо 197 случаев подчеркивания эмфазы, есть еще 25 случаев, когда жесту приписывается значение ‘подтверждение’. Приведем несколько примеров:

- (2) *Да... не... не беспокойся, вон у нас сколько штофов. Во!*  
(3) — *Дядь Гриш! Ты бюрократ? — Я бюрократ? Вот те, кто вот эти бумаги выдумывают, вот это бюрократы. А я мученик!*  
(4) *Possible, possible, возможно! Вы для Мартынова все, абсолютно все. Ты понимаешь?*  
(5) *Пропала премьера! Дочери вашей нигде нет, выпускать некого, пороку у меня не хватает!*

Мы выдвинули гипотезу, что определяющим фактором здесь является наличие в словах, сопровождаемых кивком ‘подтверждение’, смыслового (а иногда и дословного) повтора, отсылающего к тому, что жестикулирующий уже сказал ранее. Так, в примере (2) *Во!* отсылает к *вон у нас сколько штофов*; ср. также (3) *бюрократы — вот те, кто эти бумаги выдумывает*, (4) *абсолютно все — вы для Мартынова все*, (5) *выпускать некого — дочери вашей нигде нет*.

Эта гипотеза помогает объяснить примерно половину случаев — 13 клипов из 25. Приведем примеры контекстов, для которых приведенное объяснение не подходит:

- (6) *А потом — хе! — как начали вешать за эту свободу, да как погнали в тюрьмы да в Сибирь. Перепугался... до смерти.*  
(7) *Родька! Я люблю железные дороги.*

- (8) — Ты не лег? — Нет. — А знаешь... Я тебя очень хорошо помню.
- (9) Мы будем только рады... что взялся за ум и занялся хорошим делом.

Для этих случаев мы пока не можем предложить никакой интерпретации; впрочем, вероятно, что 13 исключений из 142 примеров могут быть просто результатом ошибки, которой трудно избежать при ручной разметке такого большого количества клипов.

### Заключение

Лучшим подспорьем в работе разметчика, стремящегося к максимальному единообразию принимаемых решений, могла бы стать подробная инструкция, в которой сходные по названию (и, тем самым, по способу реализации) жесты (а также значения жестов) объединяются в группы и про каждый элемент внутри каждой группы дается пояснение, позволяющее разметчику понять, чем какой-то конкретный жест (или конкретное значение жеста) отличается от других, похожих на него. Мы надеемся, что наши исследования дадут возможность написать такую инструкцию для будущих разметчиков МУРКО.

### Литература

Баранов А. Н., Крейдлин Г. Е. Иллокутивное вынуждение в структуре диалога // Вопросы языкознания. 1992. № 2. С. 84–99.

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

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

Grishina E. A. Multimodal Russian Corpus (MURCO): First Steps // Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17–23 May 2010, Valletta, Malta. P. 2953–2960.

**S. I. Pereverzeva, N. A. Ermolaeva, A. V. Zueva, E. A. Slepak**

*National Research University Higher School of Economics*

*(Russia, Moscow)*

*P\_Sveta@hotmail.com, tasha.ermolaeva@gmail.com,  
an5200@ya.ru, janenikel16@gmail.com*

### DE PROFUNDIS: THE PROBLEMS OF THE DEEP ANNOTATION OF THE MULTIMODAL RUSSIAN CORPUS AND THE WAYS OF RESOLUTION

The paper focuses on the manual gesture annotation in the Multimodal Russian Corpus (MURCO), which was started by E. A. Grishina and is being developed by the authors of this paper. The important aspect of the annotation process is the attempt to provide

“the uniformity and commonality of the markup” [Grishina 2010] to the maximum degree possible. To do so, the annotator should carefully study the MURCO data which was marked by E. A. Grishina (since sadly it is no longer possible to ask the questions directly) and define the rules that govern the gesture annotation and that were probably implied by E. A. Grishina herself.

The paper describes three of such rules: 1) the choice between the gestures *to open one's eyes wide* and *to raise one's brows* (both meaning ‘fear’) — we state that the main factor here is the distinctness of gesture performance, 2) the choice between the meanings ‘confirmation’ and ‘emphasis’ of the gesture *to close one's eyes*, and 3) the choice between the same two meanings of the gesture *to nod*. In both cases the meaning ‘confirmation’ is preferable if the utterance accompanied by the gesture is the reply to somebody's remark. The other factor is the cohesion — if the utterance accompanied by the gesture conveys the same meaning as the previous utterance of the speaker, the meaning ‘confirmation’ should be preferred.

**Keywords:** Multimodal Russian Corpus, gesture, deep annotation, rules of choice, instruction.

### References

Baranov A. N., Kreydlin G. E. [Illocutive forcing in the dialogue structure]. *Voprosy yazykoznaniya*, 1992, no. 2, pp. 84–99. (In Russ.)

Grishina E. A. [Multimodal Russian Corpus (MURCO): problems of annotation]. *Natsional'nyi korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [Russian national corpus: 2006–2008. New results and perspectives]. St. Petersburg, 2009, pp. 175–214. (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, 2017. 744 p. (In Russ.)

Grishina E. A. Multimodal Russian Corpus (MURCO): First Steps. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17–23 May 2010, Valletta, Malta*. Valletta, 2010, pp. 2953–2960.



ТРУДЫ ИНСТИТУТА РУССКОГО ЯЗЫКА ИМ. В. В. ВИНОГРАДОВА

Выпуск 21

Национальный корпус русского языка: исследования и разработки

Оригинал-макет *Л.Е. Голод*  
Дизайн обложки *И.А. Тимофеев*

Подписано в печать 00.00.2019. Формат 70×100 <sup>1</sup>/<sub>16</sub>  
Бумага офсетная. Печать офсетная. Усл.-печ. л. 20,5. Заказ № 1961  
Тираж 300 экз.

Издательство «Нестор-История»  
197110 СПб., Петрозаводская ул., д. 7  
Тел. (812)235-15-86  
e-mail: [nestor\\_historia@list.ru](mailto:nestor_historia@list.ru)  
[www.nestorbook.ru](http://www.nestorbook.ru)

Отпечатано в типографии издательства «Нестор-История»  
Тел. (812)235-15-86