

Методологические проблемы создания Компьютерной антологии русского рассказа как языкового ресурса для исследования языка и стиля русской художественной прозы в эпоху революционных перемен (первой трети XX века)

Г.Я. Мартыненко, Т.Ю. Шерстинова, А.Г. Мельник, Т.И. Попова

Санкт-Петербургский государственный университет

g.martynenko@spbu.ru, t.sherstinova@spbu.ru,
st064458@student.spbu.ru, tipopova13@gmail.com

Аннотация

В контексте социальных потрясений первой трети XX в. рассматриваются революционные изменения в русском языке и художественной литературе на материале русского рассказа. Материалом для такого рода исследований является Компьютерная антология русского рассказа, разрабатываемая на кафедре математической лингвистики Санкт-Петербургского университета.

Рассмотрены теоретические и прикладные предпосылки данной масштабной задачи. В основе проекта лежат идеи русской формальной школы и, прежде всего, системные идеи Ю.Н. Тынянова. Привлечены критерии, обеспечивающие представительность авторов и их произведений на основе объективных формальных процедур, в частности, размера статьи в Краткой литературной энциклопедии, посвященной конкретному беллетристу. В дальнейшем предполагается расширение хронологических рамок антологии с охватом всего XX в., а также конца XIX и начала XXI столетия.

Ключевые слова: русский язык, русская литература, рассказ, революционная эпоха, компьютерная антология, первая треть XX века

«В русской литературе рассказ традиционно был сильным жанром. Пожалуй, лишь американская литература приближается в этом отношении к нашей»

Юрий Нагибин. О рассказе [22]

1. Введение

История России XX в. драматична. Драматичны и исторические судьбы русского языка и русской литературы этой эпохи. Особенно богата переломными событиями первая треть столетия, вместившая в себя три социальные революции: революцию 1905 года, Февральскую и Октябрьскую революции 1917 года и три войны (русско-японскую, первую мировую и гражданскую).

Упомянутые события привели к многократной перестройке русского языка и его стилевых парадигм. Эти изменения носили преимущественно стихийный характер, но был и опыт вмешательства государства и активной части общества в процесс формирования языковой политики и построения новой революционной эстетики, дабы противопоставить волю революционных масс достижениям прежних эпох. Имеется в виду реформа русской графики и орфографии, эстетические эксперименты революционных направлений в искусстве, например, Пролеткульта.

Часть языковых сдвигов заметна «невооруженным глазом», другие языковые изменения имеют латентный характер. Однако для осознания масштабности и тех и других трансформаций языка необходимо применение строгих количественных методов, анализ представительного объема языкового материала на разных лингвистических уровнях и сравнение разных хронологических срезов в динамическом аспекте. Только в результате такого анализа можно будет с уверенностью сказать, в какой степени одна из самых драматических эпох русской истории повлияла на трансформацию русского языка, какие языковые уровни были затронуты в первую очередь, в чем конкретно выразилось это изменение и какова реальная суть произошедших преобразований.

2. Теоретические предпосылки

Теоретическими предпосылками создания Компьютерной антологии являются три источника: теория совокупности выдающегося русского статистика начала XX века А.А. Чупрова, статистические представления русского писателя и ученого А. Белого о массовом создании словарей русских писателей и системные представления выдающегося представителя русской формальной школы Ю.Н. Тьянянов.

1) А.А. Чупров, основываясь на теории сообществ, выдвинул идею статистической совокупности, логика которой основана на собирательных понятиях [30]. Эта теория исходит из идеи сообщества Густава Рюмелина и идеи ценоза Карла Мёбиуса. Впоследствии логика Чупрова была применена к концепции корпуса [21].

2) Андрей Белый достиг важных качественных результатов, основываясь на массовом исследовании русской поэзии и открыв ряд принципиальных закономерностей в истории ритмики русского стиха. Впечатляющих результатов он достиг именно на акцентном уровне, так как только на нем можно было обеспечить значительную массовость наблюдения при состоянии скромных технологических средств начала прошлого века. На лексическом уровне А. Белому это не удалось, хотя он и поставил важную задачу «производственного» построения словарей русских писателей.

3) Ю.Н. Тьянянов выдвинул концепции синхронических и диахронических литературно-художественных систем. Под синхроническими системами он понимал совокупность произведений данной литературной эпохи, а под диахроническими — последовательность сменяющих друг друга синхронических систем. При этом автор сетует на то, что усилия большей части литературоведов устремлены на изучение произведений выдающихся писателей, тогда как периферия литературы и даже ее «центр» остаются за бортом исследовательского интереса. Это означает, что для Тьянянова крайне важным был вопрос максимальной представленности авторов в той или иной системе литературы и представительности корпуса текстов этих авторов. Любой текст Тьянянов рассматривал как литературный факт, который должен приниматься во внимание независимо от масштабов дарования автора и его роли в литературном процессе.

Проблема системного анализа литературы тесно переплетается с проблемой возрождения и сохранения литературного наследия, существенная часть которого до недавнего времени была вычеркнута из памяти народа, ведь профессиональная текстовая рефлексия предполагает обращение к информационным ресурсам, несоизмеримым с теми, с помощью которых удовлетворяются интересы массового читателя. Для сообщества исследователей современной формации необходим доступ ко всей литературной продукции, созданной в ту или иную историко-литературную эпоху. Наш подход позволяет заполнить лакуны и области представленности литературной периферии.

Следует обратить особое внимание на некоторые успешные практические шаги в изучении русского новеллистического наследия. Пальму первенства следует отдать замечательной антологии писателя Юрия Нагибина [22], который в теперь уже далеком 1987 г., то есть еще при советской власти, но уже в период назревавших перемен, на страницах «Книжного обозрения» опубликовал свою антологию, отражающую развитие

советской новеллистики за семьдесят лет, истекших со дня Великой Октябрьской революции. В предисловии к антологии он пишет: «Мы чужды каких-либо научных, литературоведческих претензий и потому не стремимся охватить как можно больше имен, но по возможности — как можно больше хороших рассказов разных писателей. Таким образом, у нас нет намерения создать всеобъемлющую антологию, да это и не по плечу одному составителю. Быть может, эта несовершенная попытка подвигнет наших специалистов на создание научной антологии советского рассказа в целом» [16]. Предлагаемая нами антология является откликом на призыв выдающегося писателя. Это первая попытка создания всеобъемлющей научной антологии русского рассказа на объективной основе, восходящей к идеям Тынянова.

Остановимся на еще одном важном пункте. Интерес к жанру рассказа в писательской среде, возможно, и среди читателей в течение XX столетия, был непостоянным. В значительной степени он был связан с социальной активностью различных слоев общества в России на различных этапах социально-политического развития — с его подъемами и спадами, взрывами и апатией. Ведь, как отмечает Юрий Нагибин, «в жизни страны бывают разные периоды, которые можно сравнить с боевыми действиями во время войны: затишье, отступление, бой местного значения, наступление... А какое же наступление без атаки? На острие атаки должен находиться рассказ» [22, с. 3]. Все эти размашистые колебания отражаются на мощности информационного потока жанра короткой строки. Эти волны позволяют выявить с помощью рассказа всплески и спады социальной динамики. То есть рассказ выступает в роли диагноста социальных процессов. Эти два вектора должны постоянно находиться в активной зоне исследовательского интереса.

3. Технологические предпосылки

К настоящему времени накоплен значительный опыт в осмыслении специфики работы с большими массивами данных в рамках компьютерной, количественной и корпусной лингвистики. Особенно значительные успехи характерны для последнего времени и связаны они с развитием усилий в области обработки очень больших массивов данных — так называемых *big data*. Эти усилия влились в традиционные статистические проблемы формирования и анализа малых и больших, в традиционном смысле, выборок.

Для традиционной статистики малая выборка — это несколько десятков единиц, а большая — несколько сотен. Однако в подходе, основанном на *big data*, технические средства и программное обеспечение позволяют работать с практически неограниченными объемами данных. Да, собственно, и сам человек в наши дни живет среди этих данных и сам стал частицей киберпространства. Любопытно, что такие данные нашли применение не только при решении актуальных современных задач гуманитаристики, но и стали внедряться в способы решения проблем традиционного языкознания и литературоведения при изучении стиля, сюжета, жанра, эмоциональной напряженности текста, траекторий нарратива и пр. [35, 37]. При этом возникают проблемы соотношения малых, больших и сверхбольших выборок.

Проблема эта не проста, так как согласно устоявшимся статистическим представлениям чрезмерно большие выборки, формируясь на основе практически бесконечной генеральной совокупности, весьма неоднородной, стирают различия между неоднородными фрагментами. То, что при этом выявляются какие-то сверхобщие закономерности, находится под большим вопросом: разве можно найти какие-то закономерности в конгломерате, напоминающем хаос? Такое исследовательское поведение с точки зрения классической статистики является абсурдным. Тем не менее, данный подход имеет некоторое оправдание с точки зрения лингвиста, поскольку при сверхбольших выборках вычерпываются ресурсы языка, а лингвиста волнует, прежде всего, это. Но если, лингвист — не только лингвист, но и словесник, то есть ориентирован на изучение жанрово-стилевой дифференциации литературного языка, то он не может

смириться с таким подходом. Большие числа ему тоже нужны, но только с учетом фактора неоднородности коммуникативного поведения человека. Литературоведа интересует, прежде всего, индивидуальное речевое поведение и поведение отдельных групп, разделяющих осознанно или инстинктивно свою принадлежность к группе.

При формировании Компьютерной антологии нужно иметь в виду, что такая работа предполагает крен в сторону литературоведения и той части лингвистики, которая использует лингвистические методы для решения проблем литературоведческих. Такое литературоведение ориентировано, прежде всего, в сторону той части поэтики, которая восходит к методам русской формальной школы XX в. [14, 28, 32]. Данный подход предполагает, прежде всего, ориентацию на исследование отдельных жанров и только затем — обращение ко всему жанровому разнообразию при учете самобытности каждого. В нашей концепции акцент сделан на жанре рассказа.

В пользу такого выбора говорит то, что рассказ выполняет функцию «разведчика» и жанра быстрого реагирования, отвечая на требования и вызовы эпохи и даже порой предвидя их. Следует также учитывать, что рассказ принадлежит к малым формам прозы, а это позволяет вовлечь в орбиту исследования большое количество прозаиков и их произведений. Важной особенностью рассказа является то, что он остро реагирует именно на современность, вбирая в себя характерные черты действительности и особенности языка современников. Причем этот жанр отражает все стороны реальной жизни, все ее богатство и огромное разнообразие индивидуальных стилевых систем.

При формировании антологии предполагается обеспечить достаточно полную представленность авторов с соблюдением принципа учета масштабов их роли в литературе на том или ином этапе эволюции жанра. В каждом периоде столетия должны себя обозначить и представители старшего поколения (архаисты), и молодые писатели. Часть из них следует в форватере прежних эпох, а другая проповедует новые тенденции (это новаторы). Необходимо также учитывать масштабы дарования каждого писателя и степень его влияния на коллег по писательскому цеху. Однако в любом случае нужно иметь в виду, что в пределах данного жанра художественная литература конкретного периода представляет собой монолитную целостность, спаянную «духом эпохи», единством общности социального, литературного и языкового существования.

Формирование списка авторов конкретной эпохи осуществляется на основе существующих библиографий, обзоров, энциклопедий, словарей писателей. Таких источников довольно много. Учитываются также интернет-коллекции и существующие корпуса. После составления списка авторов каталогизируются их произведения. При формировании выборки произведений будет использоваться критерий размера статьи, посвященной данному автору в Краткой литературной энциклопедии в 8 т. [18] и в Биографическом словаре русских писателей (1889–2007), а также объем списка литературы, относящейся к творчеству данного автора.

В работе используется широкий спектр филологической и исторической литературы, а также литературных обзоров и критических источников. Такие тексты можно разделить на две группы: лингвистические и литературоведческие. В пределах каждой группы рассматриваются труды, написанные в исследуемый период или «по свежим следам» [4-8, 11-13, 15, 17, 19, 23, 24, 27, 34, 36], а также публикации, подготовленные в более поздние годы вплоть до нашего времени [1-3, 10, 26]. При этом предпочтение отдается нами первой группе, так как она представляет собой непосредственную реакцию на процессы, протекающие на глазах у исследователя или критика, в том числе тех, которые принимают личное участие в этих процессах.

4. Заключение

В статье поставлена задача построения Компьютерной антологии русского рассказа первой трети XX века. Рассмотрены теоретические предпосылки и прикладные аспекты

этой масштабной задачи. В основе проекта лежат идеи русской формальной школы и прежде всего системные идеи Ю.Н. Тынянова, обеспечивающие представительность авторов и их произведений на основе объективных формальных критериев.

В дальнейшем предполагается расширение хронологических рамок антологии с охватом всего XX века а, возможно, и начала XXI столетия.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант № 17-29-09173 офи_м «Русский язык на рубеже радикальных исторических перемен: исследование языка и стиля предреволюционной, революционной и постреволюционной художественной прозы методами математической и компьютерной лингвистики (на материале русского рассказа)».

Литература

- [1] Аверин Б.В., Нитраур Э. (ред.) Русская литература XX века: Исследования американских ученых. СПб: Петрориф, 1993.
- [2] Аверин Б.В. От Толстого до Набокова. Из истории русской литературы. Издательство имени Н. И. Новикова, Gallina Scripsit, 2014.
- [3] Аверин Б.В., Козакевич А. Великая Война 1914–1918 гг. в поэзии и прозе. Антология / Сост., вступ. статья Б. В. Аверина. М.: Научно-образовательное культурологическое общество, 2014.
- [4] Белый Андрей. Мастерство Гоголя: Исследование. М.-Л.: ГИХЛ, 1934.
- [5] Верховской П.В. Письменная деловая речь. Словарь, синтаксис, стиль. Разбор бюрократических шаблонов и нарушений грамматики в языке документов. М.: Техника управления, 1930.
- [6] Винокур Г.О. Речевая практика футуристов // Винокур Г. Культура языка. М.: Изд-во «Федерация», 1929. С. 304 – 318.
- [7] Винокур Г.О. Язык НЭПа // Винокур Г. Культура языка. М.: Изд-во «Федерация», 1929. С. 115 – 139.
- [8] Винокур Г.О. О революционной фразеологии (Один из вопросов языковой политики). // ЛЕФ. 1923. № 2. С. 104 – 118.
- [9] Горнфельд А. Новые словечки и старые слова. М.: Колос, 1922.
- [10] Гречнев В.Я. Русский рассказ конца XIX–XX века. Л.: Наука, 1979.
- [11] Гринберг И. Широкое дыхание рассказа // Нева. 1968. № 8. С. 161 – 162.
- [12] Гринберг И. Энергия рассказа // Нева. 1978. № 1. С. 191 – 192.
- [13] Гусев Вл. «Левша» и стилевые тенденции в советской прозе // В мире Лескова: Сб. статей. М.: Советский писатель, 1983. С. 91 – 192.
- [14] Жирмунский В.М. Вопросы теории литературы. Статьи 1916–1926. Л.: Academia, 1928.
- [15] Карцевский С.И. Язык, война и революция. Берлин: Русское универсальное издательство, 1923.
- [16] Книжное обозрение. № 7, 13 февраля 1987 г. С. 3.
- [17] Крамов И. Достоинства рассказа // Дружба народов. 1977. № 8. С. 249 – 266.
- [18] Краткая литературная энциклопедия. М.: Сов. Энцикл., 1962—1978. Т. 1–9.
- [19] Мазон А. Лексика войны и революции в России (1914–1918). Введение. Аббревиация // Политическая лингвистика. 2013. № 1 (43). С. 203 – 210.
- [20] Мартыненко Г.Я. Введение в теорию числовой гармонии текста. СПб: Изд-во Санкт-Петербург. ун-та, 2009.
- [21] Мартыненко Г.Я. Основы стилеметрии. Л.: Изд-во ЛГУ, 1988.
- [22] Нагибин Ю.М. (сост.). Антология русского советского рассказа. Предисловие Ю. Нагибина. Библиотечка «Книжного обозрения». М.: Книжное обозрение, 1987.
- [23] Новиков Л.А. Стилистика орнаментальной прозы Андрея Белого. М.: Наука, 1990.

- [24] Поливанов Е.Д. Революция и литературные языки Союза ССР // За марксистское языкознание. М.: Федерация, 1931. С. 73 – 94.
- [25] Русские писатели 1800–1917. Биографический словарь. М.: Большая российская энциклопедия, 2007. Т. 1–5. 1989–2007.
- [26] Русский советский рассказ: очерки истории жанра / под ред. В.А. Ковалева, Академия наук СССР, Институт русской литературы, 1970.
- [27] Селищев А. Язык революционной эпохи. Из наблюдений над русским языком (1917–1926). М.: Работник просвещения, 1928.
- [28] Томашевский Б. Теория литературы: Поэтика. М.–Л.: ГИЗ, 1928.
- [29] Тынянов Ю.Н. Архаиста и новаторы. Л.: Прибой, 1929.
- [30] Чуковский К.И. От Чехова до наших дней. Литературные портреты. Характеристики. СПб-М.: Издание Т-ва Вольф, 1910.
- [31] Чупров А.А. Очерки по теории статистики. СПб: М. и С. Сабашниковы, 1910.
- [32] Шкловский Б.В. Теория прозы. М.: Федерация, 1929.
- [33] Шор Р. Язык и общество. М.: Работник просвещения, 1926.
- [34] Эйхенбаум Б. О прозе. О поэзии. Сб. статей. Л.: Художественная литература, 1986.
- [35] CLiC Dickens project. URL: <http://clic.bham.ac.uk> (дата обращения: 23.04.2018).
- [36] Jakobson R. Vliv revoluce na ruský jazyk. (Влияние революции на русский язык). Praha, 1921.
- [37] Reagan A.J., Mitchell L., Kiley D., Danforth C.M., Dodds P.S. The emotional arcs of stories are dominated by six basic shapes. arXiv: 1606.07772 [cs.CL] <https://arxiv.org/pdf/1606.07772.pdf> (дата обращения: 23.04.2018).

Methodological Issues Related with the Compilation of Digital Anthology of Russian Short Stories (the First Third of the 20th Century)

G.Y. Martynenko, T.Y. Sherstinova, A.G. Melnik, T.I. Popova

Saint-Petersburg State University

In the context of social upheavals of the first third of the 20th century, the revolutionary changes in the Russian language and fiction are to be studied on the basis of Russian short stories. The material for this kind of research is Digital Anthology of Russian short stories, which is currently being developed at the Department of Mathematical Linguistics of the St. Petersburg University. The theoretical basis and applications of this large-scale task are considered. The project is based on the ideas of the Russian formal school and, first of all, on the systemic ideas by Yury N. Tynyanov. Special criteria are developed to ensure the representativeness of authors and their works on the basis of objective formal procedures (e.g., the size of the article in the Concise Literary Encyclopedia devoted to a specific writer). In the future, an extension of the chronological framework of the Anthology with coverage of the entire 20th century, as well as the end of the nineteenth and the beginning of the 21st century, is proposed.

Keywords: Russian language, Russian literature, short story, revolutionary era, computer anthology, first third of the 20th century