ФИЛОЛОГИЧЕСКИЕ НАУКИ / PHILOLOGICAL SCIENCES

УДК 81-139: 808.5: 821.161.1: 159.9-612.821

РУБРИКАЦИЯ И СЕГМЕНТАЦИЯ ХУДОЖЕСТВЕННОГО ТЕКСТА (НА ПРИМЕРЕ РУССКОЙ ПРОЗЫ XIX–XX В. В.)

SUBJECT HEADINGS AND ARTISTIC TEXT SEGMENTATION (RUSSIAN PROSE OF XIX–XX CENTURIES)

©Кузина Н. В.
канд. филол. наук
Российский научно–исследовательский институт культурного и природного наследия им. Д. С. Лихачева
г. Москва, Россия
nvkuzina@mail.ru
©Kuzina N.
PhD
Likhachev Russian Research Institute of Cultural and Natural Heritage
Moscow, Russia
nvkuzina@mail.ru

Аннотация. Предлагается взгляд на конструкцию художественного повествовательного текста и методика, дающая возможность использовать статистический аппарат и аргументированно описывать (прежде всего при постановке диагноза или в случае экспертизы) тексты любого типа (как письменные, так и устные) с помощью составления модифицированного типа частотных словарей с адресами употребленных словоформ, с точки зрения деления на фрагменты на разных уровнях рубрикации, объединения частей текста с помощью доминантных и лейтмотивных вербальных средств, выделения универсальных по строению повествовательных блоков. Показана роль в организации итераций текста, определенной авторским замыслом, заданием говорения или сформированной иными причинами перво(прото)ситуации, ассоциативных рядов, предложен алгоритм выделения и описания элементарной единицы повествования (мотива), немотивированных заданием говорения элементов.

Abstract. It is proposed to look at the structure of artistic narrative text and methodology, making it possible to use the statistical apparatus and convincingly describe (especially in the diagnosis or in the case of examination) the texts of any type (both written and oral) with the preparation of a modified type of frequency dictionaries addresses the use of word forms, from the point of view of division into fragments at different levels of categorization, combining parts of the text with the dominant leitmotiv and verbal means, allocating universal in the structure of the narrative units. The role in the organization of the iterations of the text, a certain author's intention, the task of speaking, or formed by other causes of the first (proto) situation, associative series, an algorithm selection and description of the elementary narrative units (the motive), unmotivated job speaking elements.

Ключевые слова: модель текста, рубрикация, повествовательный блок, протоситуация, мотив, ассоциативно–вербальный ряд.

Keywords: model of text, headings, narrative flow, protosituation, motive, associative–verbal number.

Несистематизированная информация является во многом обесцененной до тех пор, пока исследователь не упорядочит данные по некоторым признакам, отражающим иерархические связи в пределах объекта исследования. Так, например, в лингвистике давно установлено, что «шапка» частотного словаря носителя языка характеризует его индивидуальную картину мира и ценностную модель (потребности), иными способами выявляемую трудоемко или не вполне достоверно.

Целью исследования было описать конструкцию, каркас текста на примере образцов кодифицированного русского литературного языка — художественных прозаических произведений. Был применен стандартный метод составления частотных словарей и введены некоторые его модификации [1–5]. Каждому графическому слову (СГ) присваивались данные: адрес (позиция) в целом тексте; адрес в выделенном автором фрагменте; указывался номер фрагмента. Приводился контекст (условная рабочая «строка» электронного варианта текста).

Дискретным был назван *текст* (ДТ), состоящий из фрагментов, выделенных и помеченных автором как самостоятельная повествовательная единица в составе целого (роман, цикл рассказов, рубрицированный рассказ). *Единицей рубрикации* (РЕ) — фрагмент текста, выделенный автором иначе, чем абзацным отступом («звездочек», озаглавливания или нумерации). РЕ может включать дальнейшее деление (*части* делятся на *главы*, *главы* — на *параграфы*). *Повествовательный блок* (ПБ) — единица уровня наррации в тексте, близкая *сверхфразовому единству, информационному блоку, анекдоту* в значении законченной истории; сегмент текста, выделяемый исследователем; является неэлементарной единицей повествования, более крупной, чем *мотив*. ПБ считается фрагмент текста, описывающий завершенное действие, движение и т. д. Граница ПБ — момент перехода действия в иной локус, смена времени действия, модальности, набора персонажей, эмоциональности лексики. Условие выделения блока — возможность свертывания его до словосочетания или назывного предложения (озаглавливание). *Ассоциативно–вербальной единицей* (АВЕ) был обозначен любой ряд повторяющихся варьируемых элементов, обладающих каждый семой (семами), оцениваемыми носителями языка как общие для ряда, вне зависимости от того, лексический ли это, корневой повтор, повтор слов одного семантического поля или ассоциативного ряда. При описании АВЕ обозначается ведущей лексемой ряда, объединяющей собой наибольшее число ассоциатов.

Такой тип обработки и представления текста позволяет решить следующий задачи:

1. Описать рубрикацию (деление на озаглавленные, пронумерованные и т. п. фрагменты самим автором), определить ее типы и параллели с содержанием и естественной логической структурой текста.
2. Описать зависимости между ролью главы / фрагмента в содержании и ее величиной (количеством входящих СГ).
3. Описать механизмы повторов в пределах фрагментов и целого текста.
4. Указать среднюю величину фрагмента (главы в повести и романе) как постоянную величину, характеризующую индивидуальный стиль.
5. Описать сегментацию текста (деление исследователем на фрагменты в соответствие с конструктивными особенностями фразы и семантики).

В рассматривавшемся материале (образцовая русская проза — выборочно романы, повести, циклы рассказов середины XIX–XX в. в. От И. Тургенева и Л. Толстого до А. Солженицына и Ю. Трифонова (более 40 текстов)) ДТ содержали от одного до четырех уровней рубрикации, что в случае художественного текста имеет жанровое, временнóе объяснение и может быть связано в том числе с особенностями мышления (усложнение–упрощение структуры) писателя. Особые функции «смыслового курсива», руководства восприятием текста имеет озаглавливание, обязательное для РЕ в циклах рассказов и факультативное для РЕ в ДТ других типов.

Количество РЕ одного уровня в исследуемых текстах колеблется от 239 (Л. Толстой, «Анна Каренина») до 3 (М. Пришвин, «Старые рассказы»). РЕ одного уровня могу иметь либо приблизительно равный вес (по количеству СГ) или существенно различаться по весу. Выделяются не менее двух разновидностей ДТ в зависимости от веса РЕ одного уровня: 1) для Тургенева, Толстого, Бабеля, Платонова, Набокова, например, характерно составление ДТ из РЕ (глав, рассказов), существенно различающихся (в десятки раз) по величине; 2) в некоторых ДТ М. Булгакова (цикл рассказов «Записки юного врача»), Б. Пастернака («Доктор Живаго»), Ю. Трифонова («Студенты») разница между весом РЕ одного уровня либо несущественна, либо они различаются по весу не более, чем в два раза. Средний вес РЕ (например, главы романа) является величиной, устойчивой на протяжении всего творчества и характеризующей поэтику автора. Наблюдается закономерность: если средний вес главы велик, то автор не использует в ДТ подчеркнуто контрастных по весу РЕ. Если средний вес мал, то автор предпочитает остранять текст посредством сочетания РЕ разной величины.

В первом случае средний вес РЕ составляет от трех до десяти тысяч СГ («Доктор Живаго» Б. Пастернака, «Записки юного врача» М. Булгакова, «Студенты» Ю. Трифонова и др.). Во втором случае средний вес РЕ в ДТ составляет от тысячи до двух тысяч СГ (1952 с амплитудой от 197 до 2073 в «Дворянском гнезде», 1939 с амплитудой от 449 до 4500 в «Отцах и детях» И. Тургенева; 1128 с амплитудой от 196 до 2324 в «Анне Карениной» Л. Толстого; 896 с амплитудой от 135 до 1919 в «Конармии» И. Бабеля и т. д.), Исключение составляет А. Платонов (средний вес — 4259 СГ с амплитудой от 438 до 11518), по многим показателям выпадающий из предложенных классификаций. Сюжетно маркированы (являются высшими точками «напряжения» в повествовании) в ДТ первой разновидности РЕ, резко выделенные в сторону уменьшения или увеличения веса (далее — *маргинальные*). По особенностям расположения РЕ разного веса выделяются ДТ, построенные по принципу контрастного сочетания вне динамики, и ДТ, выстроенные по принципу контрастного сочетания фрагментов с динамикой увеличения глав от начала к финалу («Чевенгур» А. Платонова, «Отцы и дети» И. Тургенева).

Основные типы АВЕ по распределению в РЕ, участвующие в организации ДТ

1. *АВЕ–лейтмотив* ДТ. Лексемы ряда (из числа тех, которые не являются высокочастотными в языке в целом и употребление которых не является приметой повествования в целом (т. е. исключая лексемы *быть, говорить, сказать* и др.) встречаются в большей части / половине РЕ ДТ. *АВЕ*–лейтмотив может состоять из любого количества элементов, если число употреблений лексем велико, он получает название *многокомпонентного*, если нет — *малокомпонентного.* Он может быть: а) *с доминантным текстом***,** если в одной из РЕ лексема ряда встречается значимо большее число раз, чем в остальных; б) *без доминантного текста***,** если употребления лексемы распределены по РЕ равномерно.

2. *АВЕ–доминанта* РЕ. Доминанты составляют лексемы, неоднократно встречающиеся, обладающие высокой частотой (3 и выше употребления; за исключением тех лексем, которые

обладают высокой частотой употребления в языке), фигурирующие в одной РЕ. В редких случаях доминанты могут поддерживаться однократными повторами лексемы в других РЕ. Выделяется, кроме основного, четыре вида доминант (виды, в разной степени близкие лейтмотивам):

а) Доминанта с обратной отсылкой. После РЕ, где лексема является доминантной, в цикле возможна РЕ, где лексема будет однократно употреблена и породит, таким образом, в сознании читателя эффект припоминания описанных ранее событий (РЕ, где лексема была доминантной);

б) Доминанта с приемом антиципации. РЕ с доминантной лексемой оказывается подготовлена ее однократным употреблением в одной из предшествующих РЕ;

в) Доминанта с приемом антиципации и обратной отсылкой. Такой вид доминанты граничит с лейтмотивом (лейтмотивным рядом с доминантным текстом) — РЕ с доминантной лексемой «подготовлена» ее опережающим однократным употреблением, а затем «припоминается» при обратной отсылке;

г) Парная доминанта. Лексема одинаково неоднократно встречается (и является маркирующей) в двух РЕ, граничит по функциям с двух–трехкратными употреблениями редких лексем. Это мощное средство создания в пределах цикла «текстов–двойников», задает параллелизм двух РЕ. Сложные типы доминант, как и лейтмотив, являются классическим примером руководства памятью читателя.

3. *АВЕ с минимальной частотностью* (состоят их редких, но выделенных в тексте, важных не только семантикой, но и звуковым обликом или экзотичностью обозначаемого понятия, лексем). Сходные для индивидуума явления описываются при помощи сходных лексических единиц. АВЕ с минимальной частотностью устанавливают параллели между неочевидно связанным ситуациями, персонажами, РЕ. Подобные повторы, при отдаленности друг от друга воспроизводимых слов и их уникальности, оказываются наиболее выделенными, бросаются в глаза при анализе словаря. На них обращает внимание читательская аудитория во время заданий на восприятие.

Универсальный механизм распределения информации в пределах выделенного автором текста фрагмента (главы / рассказа) — варьирование (многократное проговаривание) *протоситуации*. Каждый ПБ законченного повествовательного текста — один из вариантов описания протоситуации (в более или менее эксплицированной, полной / неполной форме). Иначе: при наличии общего развертывания действия от начала к финалу текста в каждом блоке наблюдается явление повтора универсального набора одних и тех же компонентов ситуации, ради которой начато говорение (движение как бы свернуто, отсутствует; каждый ПБ в свернутом виде содержит весь текст).

Последовательность ПБ, развертывание текста может быть условно названо «левополушарным», тогда как кодирование, инструментовка в АВР — «правополушарным».

Процесс написания текста может быть охарактеризован как ряд однородных циклических колебаний, каждое из которых соответствует одному ПБ, до наиболее удовлетворяющего автора проговаривания вербального набора, раскрывающего суть ситуации, до достижения катарсиса и завершения авторского задания. По специфике блоков (вес по числу СГ, синтаксис, отбор компонентов протоситуации) может быть описана динамика колебаний психофизиологического состояния автора.

Любая лексическая единица в тексте детерминирована, в нем действует механизм *сведéния*. Однако при первом чтении нередко читатель сталкивается с единицами, оцениваемыми как немотивированные или необязательные (избыточные).

Целесообразно определить типы мотивировки или критерии, согласно которым единица может восприниматься как обязательная / избыточная:

1) мотивировка с точки зрения принципа правдоподобия в описании объектов отражаемого материального мира (могло ли происходить называемое (вероятностный подход); является ли называемое необходимым и достаточным для описания данного локуса, персонажа и т. д.);

2) лингвистическая мотивировка (соблюдение–нарушение сочетаемости слов, условного стилистического единства; наличие и необходимость распространяющих и уточняющих конструкций);

3) фабульная мотивировка (насколько необходимо то, что названо, для последующего развития действия).

Операция *сведéния* мотивировки завершена, если в тексте нет «лишних» (избыточных сразу на всех уровнях) элементов. Можно также говорить по способу предъявления об *эксплицитной* и *подразумевающейся* мотивировке. Эксплицированная мотивировка возникает, когда в тексте открыто поясняется, что означает и какие функции имеет деталь (например, при описании экзотических топосов).

Уникальные составляющие ПБ — единицы с понижением внешней логической мотивации — маркеры индивидуальных травмирующих ситуаций, вызвавших наибольшие изменения состояния, о которых автор пытается рассказать, но не может в силу социальных ограничений (запрет на проявление чувств, эмоций) или ограничения в силу темы говорения. Те же единицы «проговаривания», но нарочито заданные, используются в суггестивных техниках.

Наиболее часто автор–писатель с разной степенью контроля сам косвенно благодаря ряду признаков указывает на то, что лексическая единица приобретает затекстовую (за пределами замысла) мотивировку. Лингвистическая и фабульная мотивированность единицы снижается следующими способами: нарушение сочетаемости (и возникновение у единицы переносного лексического значения); стилистический сбой (единица стилистически выделена по сравнению с окружающим ее текстом); уникальность синтаксической конструкции, куда входит единица; принадлежность единицы к редкой лексике (например, единица является именем собственным); немотивированность (избыточность) употребления единицы с точки зрения движения сюжета (например, излишнее распространение в описании детали, если деталь в дальнейшем оказывается не востребованной — не воспроизводящейся повторно). Во всех случаях маркированность единицы подтверждается экспериментом на восприятие с аудиторией.

Опираясь на традиции психоанализа (спонтанное ассоциирование) и закономерности, выявленные при разработке нейролингвистических методик, можно утверждать, что лексические средства с ослабленной мотивацией в тексте относятся к разряду синкретов, помечают ситуации жизненного пути автора, в которых был пережит существенный по значимости стресс, связанный с одно- или многократным критическим (или значительным, длительным, нестандартным) изменением физиологических показателей работы систем органов. Синкрет, по Л. С. Выготскому, — допонятийное образование, в котором элементы относятся к одной структуре субъективно, по принадлежности к эмоционально значимой ситуации; «в самом синкретизме нужно найти прообраз, прототип, зародыш будущих причинных связей». Синкреты отчетливо различимы как нарушения задания говорения при записи спонтанной речи, а при хранении в долговременной памяти программируют поведенческие реакции. Набор синкретов, являющихся стойкими образованиями, определяет уникальность ассоциативного ряда носителя сознания.

Частным случаем ПБ в полистилистичном тексте является *субтекст* — стилистически цельная, различающаяся стилистически с соседними, единица повествования, включающая от одного предложения. Наличие субтекстов может быть мотивировано формированием автора в разных стилистических и социальных средах, а в спонтанной речи — наличием разных

социальных ролей. Границей субтекстов являются меты в виде стилистически окрашенной лексики и изменений синтаксических, фонематических характеристик текста.

Рациональный тип чтения любого текста, содержащего повествовательные элементы — вслед за фабулой. Элементарной различимой строительной единицей повествования в данном случае является мотив. Выражением его в логике и в лингвистике является пропозиция; таким образом, при учете мотивов текста предполагается опора на субъектно–предикатную структуру предложения. Для описания мотивов текста и статистических операций над ними была разработана инфологическая ER–модель базы данных вида «сущность–связь». В качестве сущностных выбраны отношения «Контекст», «Предложение», «Пропозиция», «Модальность», «Мотив», «Семантика». Тип связей — «один–ко–многим» и «один–к–одному». Сущность «Пропозиция» представляет собой супертип модели, на основании которого строятся подтипы «Субъект», «Предикат», «Объект», «Время», «Место». Все подтипы наследуют атрибут «Текст» супертипа. Сущность «Модальность» имеет подтипы «Объективная», «Субъективная», «Аллюзии» и «Интертекст» без наследования. Даталогическое проектирование информационной модели выполнено в среде реляционной системы управления базами данных Microsoft Access. Номенклатура отношений базы данных: *Контекст* = (Имя произведения, *Номер абзаца*, Абзац); *Предложение* = (Номер абзаца, *Номер предложения*, Предложение); *Пропозиция* = (Номер предложения, *Номер пропозиции*, Пропозиция, Субъект, Субъект: выражение, Объект, Объект: выражение, Время, Время: выражение, Место, Место: выражение); *Модальность* = (Номер предложения, *Номер пропозиции*, Объективная (индикатив): время, Выражение в тексте, Объективная (ирр): наклонение, Выражение в тексте, Источник высказывания, Выражение в тексте, Прямое / косвенное выражение, Сигнал в тексте, Интертекст: источник, Интертекст: сигнал, Аллюзии: источник, Аллюзии: доп. значение, Аллюзии: сигнал); *Мотив* = (*Номер пропозиции*, Роль в фабуле, Тип мотива); *Семантика элементов* = (*Номер пропозиции*, Семантика субъекта, Семантика объекта, Семантика предиката, Семантика места, Семантика времени).

Программное сопровождение при работе над описанием мотива обеспечивалось канд. физ.–мат. наук, доц. Т. А. Самойловой, при формировании баз частотных словарей — инженером Б. С. Шаповаловым.

Среди частных задач, таким образом, на данный момент оказываются решенными следующие:

1) выявление соответствий между типом рубрикации текста и его жанром;

2) определение зависимости между ролью единицы рубрикации (главы, фрагмента — далее РЕ) в содержании ДТ и ее весом (количеством входящих СГ);

3) описание механизмов итерации в пределах РЕ и ДТ;

4) выявление закономерностей связи между повествовательными блоками (далее — ПБ) текста посредством ассоциативно–вербальных единиц (далее — АВЕ);

5) описание элементарной единицы повествования (мотива).

***

Предложенная выше модель (с указанием на те компоненты, которые могут быть принципиальными для практики) и методика анализа могут быть использованы (с некоторой модификацией, если речь идет о спонтанной устной или письменной речи) не только в лингвистике и литературоведении, но и в любых дисциплинах, где для постановки диагноза / проведения экспертизы необходим анализ когнитивной деятельности, познанной прежде всего через речевую продукцию респондента.

Для анализа образцов устной речи разработана модификация данной методики с учетом паузирования, акустических параметров речи и введения показателя долготы интервалов между

фрагментами текста, для анализа образцов письменного текста от руки — с учетом величины отступов, пробелов между словами и строками текста, величины текста и особенностей написания букв, наклона почерка, длины строки, наличия зачеркиваний (исправлений) и т. д.

Список литературы:

1. Кузина Н. В. Информационные модели в исследовании художественной прозы (спецкурс) // PRO=ЗА. Анализ текста: учебн. материалы. Смоленск, 2003. С. 164–199.
2. Кузина Н. В. Об одной особенности словаря М. Пришвина («Кладовая солнца») // Научная конференция, посвященная 129-летию со дня рождения писателя «М. Пришвин: актуальные вопросы изучения творческого наследия»: материалы. Елец, 2002. Вып. 1. С. 61–74.
3. Кузина Н. В. Предмет (деталь) в спонтанном и контролируемом (художественном) тексте: мотивированность, принципы отбора, слои семантики и проблема адекватности интерпретации исследователем // PRO=ЗА 3. Предмет. Материалы к обсуждению: сб. науч. тр. Смоленск. СГПУ, 2005. С. 11–14.
4. Кузина Н. В. Сигналы и стратегии вторичной семантизации в повествовательном тексте // Алфавит 2: строение повествовательного текста. Сб. науч. тр. Смоленск, 2004. С. 92–138.
5. Кузина Н. В., Сергеев А. А. От литературы к медицине: методика анализа текста на основе частотного словаря // М. А. Булгаков в потоке российской истории XX–XXI в. в. М.: ART–менеджер, 2011. Вып.1. Научная серия музея М. А. Булгакова. С. 28–64.

References:

1. Kuzina N. V. Informatsionnye modeli v issledovanii khudozhestvennoi prozy (spetskurs) [Information models in the study of fiction (special course)]. PRO=ZA. Analiz teksta [PRO=ZA. Text analysis]: educational materials. Smolensk, 2003, pp. 164–199.
2. Kuzina N. V. Ob odnoi osobennosti slovarya M. Prishvina (“Kladovaya solntsa”) [On one particular dictionary Prishvin (“Sun Pantry”)]. Nauchnaya konferetsiya, posvyashchennaya 129-letiyu so dnya rozhdeniya pisatelya “M. Prishvin: aktual'nye voprosy izucheniya tvorcheskogo naslediya” [Scientific conference dedicated to the writer, “Birth 129th Anniversary M. Prishvin: topical issues of the study of the creative heritage”]: materials. Elets, 2002, issue 1, pp. 61–74.
3. Kuzina N. V. Predmet (detal') v spontannom i kontroliruemom (khudozhestvennom) tekste: motivirovannost', printsipy otbora, sloi semantiki i problema adekvatnosti interpretatsii issledovatelem [Item (detail) in spontaneous and controlled (artistic) text: motivation, selection, and semantic layers of the issue of the adequacy of interpretation researcher]. PRO=ZA 3. Predmet. Materialy k obsuzhdeniyu [PRO=ZA 3. Subject. Materials for discussion]: collection of scientific works. Smolensk, SGPU, 2005, pp. 11–14.
4. Kuzina N. V. Signaly i strategii vtorichnoi semantizatsii v povestvovatel'nom tekste [Signals and Strategies secondary semantization in narrative text]. Alfavit 2: stroenie povestvovatel'nogo teksta [Alphabet 2: The structure of a narrative text]. Collection of scientific works, 2004, pp. 92–138.
5. Kuzina N. V., Sergeev A. A. Ot literatury k meditsine: metodika analiza teksta na osnove chastotnogo slovarya [From literature to medicine: method of text analysis based on frequency dictionary]. M. A. Bulgakov v potoke rossiiskoi istorii XX–XXI v. v. Moscow, ART–menedzher, 2011, issue 1. Nauchnaya seriya muzeya M. A. Bulgakova. pp. 28–64.

Работа поступила в редакцию
24.02.2016 г.

Принята к публикации
29.02.2016 г.