

# ПРИКЛАДНАЯ ЛИНГВИСТИКА

**Н. А. Богданова**

## ФОРМАЛИЗОВАННОЕ ПРЕДСТАВЛЕНИЕ ОБЪЕКТОВ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ: ЭВРИСТИЧЕСКИЙ ПОТЕНЦИАЛ

Для эффективного функционирования систем искусственного интеллекта требуется машинная модель реального мира, которая представляет собой упрощённое представление действительности с сохранением ключевых аспектов. Создание такой модели предполагает существенное сокращение количества входящих программ при сохранении необходимого объёма знаний о мире. Этот процесс известен как «сжатие знаний». Уменьшение сложности модели достигается за счёт фокусировки только на той информации, которая необходима для решения конкретной практической задачи. Такой подход называется «оптимизацией функций языка», который характерен для прикладной лингвистики.

Прикладная лингвистика использует «инженерный» подход к изучению лингвистических объектов, не стремясь к всестороннему и подробному описанию языка, а сосредотачиваясь на создании моделей, которые отвечают конкретным прикладным целям.

По отношению к моделируемому объекту, процессу или явлению модель должна удовлетворять целому ряду свойств. Важнейшими из них являются следующие [1, с. 13]:

1. Модель выступает в качестве упрощенного аналога изучаемого объекта (процесса, явления).
2. Модель не должна быть сложнее самого оригинала.
3. Метод изучения объекта (процесса, явления) путем его моделирования должен быть более экономным по сравнению с другими возможными методами изучения того же объекта.
4. Построенная модель должна быть предельно простой и логически корректной, не содержащей противоречий.
5. Модель должна по возможности иметь общий (универсальный) характер, позволяющий использовать ее для изучения других подобных объектов (процессов, явлений).
6. Модель должна отражать наиболее существенные черты реального объекта, процесса или явления, которые важны для проводимого в данный момент процесса моделирования.

Интерес к моделированию структуры художественных текстов начался с исследования В. Я. Проппа, который в работе «Морфология сказки» предпринял попытку выделить определённые инварианты в структуре волшебной сказки [2]. Для этого автором было проведено межсюжетное сравнение сказок: выделялись составные части волшебных сказок, а затем по этим частям сказки сравнивались, в результате чего получилась морфология, т.е. описание сказки по составным частям и отношению частей друг к другу.

Позднее теория В. Я. Проппа была дополнена и скорректирована в работах его последователей: были введены ввести с наименьшей (функцией) и наибольшей («ход» из 31 функции) повествовательными единицами промежуточные группировки, где функции объединялись бы в пары или в триады и были бы связаны парадигматическими и синтаксическими отношениями.

Другой возможностью построения модели художественного текста является лексико-семантическое моделирование, предполагающее создание на основе текстового авторского / жанрового корпуса алфавитно-частотного словаря, разбиение лексических единиц словаря по тематическим группам и последующий анализ наполняемости этих групп позволяет выявлять как авторские идеостилевые черты, так и формулировать некоторые выходящие за рамки авторского стиля жанровые особенности текстов.

Например, лингвистическая специфика текстов немецких коротких рассказов на лексико-семантическом уровне проявляется в сочетании трех признаков: денотативной антропоцентричности (внимание к человеку и его повседневной жизни), автологии (употребление лексических единиц преимущественно в их прямых значениях) и негативной эмоциональной окрашенности (преобладание пессимистического настроения произведения). Комбинаторика трех указанных признаков является достаточной для определения места жанра короткого рассказа среди других жанров малых форм на лексико-семантическом уровне.

Тематическая принадлежность текста на лексико-семантическом уровне может быть определена на основании таких трех факторов, как количественные характеристики отдельных тематических полей внутри класса лексических единиц, количественное соотношение отдельных тематических групп внутри одного и того же тематического поля, комбинации отдельных тематических групп различных тематических полей.

Тематическое ядро текстов на лексико-семантическом уровне может быть представлено как конъюнктивно-дизъюнктивная последовательность конкретных тематических групп в виде численно-буквенных кодов, что позволяет говорить о явлении художественной синонимии в жанре немецкой классической короткой прозы. Для минимизации информационного шума в процессе тематической классификации текстов на лексико-семантическом уровне необходимо учитывать не только наличие определенных тематических полей и тематических групп в семантической структуре текста, но их количественные характеристики и соотношения друг с другом.

Логико-семантическое моделирование художественного текста предполагает построение на основе частотного словаря таблицы основного статистического содержания текстов (выделение гласных и второстепенных опорных слов и определение их роли в сюжетодвижении: субъекты, объекты, места, времена, признаки, построение логико-семантических формул текста. Обобщенная логико-семантическая формула текста представляет собой конъюнкцию, имеющую иерархическую структуру и состоящую из регистро-

вых блоков и микротем – элементов логико-семантической структуры текста, которые выделяются с помощью таблиц основного статического содержания текстов, системы речевых регистров и рематических доминант.

Далее следует сравнительный анализ логико-семантических формул текстов на предмет доминирования отдельных микротем, наличия типичных последовательностей микротем, наличия точечной и /или линейной симметрии микротем. В логико-семантической структуре текстов коротких рассказов существует явление симметрии, которая имеет либо точечный, либо линейный характер и проявляется на двух уровнях – уровне регистровых блоков и уровне микротем. Типичной особенностью текстов коротких рассказов на логико-семантическом уровне является сочетание двух признаков: во-первых, двухуровневой симметрии (точечной симметрии на уровне регистровых блоков, точечной и линейной симметрии уровне микротем), во-вторых, событийной оси, лежащей вне текстовой ткани.

Тематическая принадлежность текста на логико-семантическом уровне может быть определена на основании пяти характеристик: количественные характеристики отдельных микротем различных речевых регистров, количественное соотношение отдельных микротем внутри одного и того же речевого регистра, комбинации отдельных микротем различных речевых регистров, количество речевых регистров и типов микротем, содержащих одну и ту же информацию, количественное соотношение повествовательных и описательных подрегистров изобразительного и информативного речевых регистров.

Логико-семантическое ядро текстов каждой тематики может быть представлено в формализованном виде как конъюнктивная последовательность конкретных микротем четырех речевых регистров в виде численно-буквенных кодов, что позволяет говорить о явлении художественной синонимии на логико-семантическом уровне.

Таким образом, эвристический потенциал формализованного представления содержания художественных текстов заключается в обнаружении инвариантных авторских, тематических и жанровых черт произведений, что дает дополнительные возможности интерпретации произведений, в том числе и через построение текстовых парадигм как нечетких множеств и определения коэффициентов принадлежности текстов к ядерной или периферийным зонам текстовых парадигм.

## ЛИТЕРАТУРА

1. *Зубов, А. В.* Информационные технологии в лингвистике : учебник для студ. Учреждений высш. проф. образования / А. В. Зубов, И. И. Зубова. – 2-е изд., испр. – М.Ж Издательский центр «Академия», 2012 – 208 с.
2. *Пропп, В. Я.* Морфология волшебной сказки / В. Я. Пропп. – Москва : Лабиринт, 2003 – 143 с.