

УДК 821.161.1

**А.С. ЯНОВСКАЯ**

(yanovska.nastya33@gmail.com)

Московский государственный лингвистический университет

## **КОРПУСНЫЙ ПОДХОД К АНАЛИЗУ ВЕРБАЛИЗАЦИИ КАТЕГОРИИ ПРОСТРАНСТВА В ХУДОЖЕСТВЕННОМ ПРОИЗВЕДЕНИИ\***

*Цель исследования заключается в том, чтобы проанализировать особенности выражения категории пространства в художественном произведении на материале цикла рассказов Дж. Лондона "Smoke Bellew", представленном на языке оригинала. Выявляются основные признаки лексических единиц с пространственным значением.*

*Используется программный инструмент, позволяющий получить статистические данные об избранных лексических единицах и на их основе интерпретировать художественное произведение с точки зрения вербализации категории пространства.*

*Ключевые слова: корпусный подход, сбалансированный корпус, художественное произведение, пространство, английский язык, Дж. Лондон.*

Необходимость оперативного анализа и обработки значительных объемов текстовой информации способствовала развитию такого направления языкознания, как компьютерная лингвистика.

С зарождения направления лингвистики текста – 60–70-х гг. прошлого века – возникла проблема определения текста. Анализ научной литературы по данному направлению позволяет сделать вывод о том, что текст по-разному понимается и по-разному определяется. С одной стороны, текст – это «любое высказывание, состоящее из одного или нескольких предложений, несущее в себе по замыслу говорящего законченный смысл», но в то же время текст понимается как «такое речевое произведение как повесть, роман, газетная или журнальная статья, научная монография, документы различного рода и т. п.» [8, с. 12]. В настоящем исследовании за основу принимается определение понятия текста И.Р. Гальперина, согласно которому, текст – «это произведение речетворческого процесса, обладающее завершенностью, объективированное в виде письменного документа, литературно обработанное в соответствии с типом этого документа, произведение, состоящее из названия (заголовка) и ряда особых единиц (сверхфразовых единств), объединенных разными типами лексической, грамматической, логической, стилистической связи, имеющее определенную целенаправленность и прагматическую установку» [2, с. 18].

Необходимо подчеркнуть различия, существующие между художественными и нехудожественными текстами (официальными документами, газетной информацией, научной литературой и т. д.). Цельность и разграничение художественного текста базируются не на строгой логике рассуждений, а на такой речемыслительной деятельности автора, которая, несмотря на наличие смыслового ядра, осуществляется в определенной структуре. Это структура определяется мотивами не только лингвистическими (языковыми), но экстралингвистического характера или индивидуально-психологического восприятия того или иного сюжета [7].

В нашем исследовании описание художественной реальности осуществляется посредством интерпретации категории пространства с использованием корпусного подхода.

Планируя проведение корпусного исследования для решения лингвистических задач необходимо, в первую очередь, определиться с характеристикой создаваемого корпуса, т. к. для решения какой-либо специфической лингвистической задачи требуется разработка специального лингвистического корпуса [10]. В научной литературе предлагаются различные классификации корпусов текстов. *Исследовательские корпуса* текстов создаются до проведения научных исследования и ориентирова-

\* Работа выполнена под руководством Горожанова А. И., доктора филологических наук, доцента, профессора кафедры грамматики и истории немецкого языка ФГБОУ ВО «МГЛУ».

ны на широкий круг задач, в том время как *иллюстративные корпуса* создаются после проведения исследования и используются для подтверждения выдвинутой теории. По типу данных традиционно формируются *письменные, речевые и смешанные* корпуса. Однако некоторые исследователи отмечают, что корпус – явление ни устное, ни письменное и ни печатное, он представляет собой четвертую форму – тексты на машинном носителе. Однако и эта точка зрения по праву оспаривается. Очевидно, что языковые данные могут быть представлены на нескольких языках. Отсюда корпуса текстов могут быть *одноязычные, двуязычные и многоязычные*. В некоторых случаях применение *параллельных* корпусов не исключается, например, с целью сопоставительного анализа оригинального текста и его перевода [5, с. 62]. По объему текстов корпуса разделяют на полнотекстовые и фрагментотекстовые. К полнотекстовым корпусам относится большинство современных корпусов, а также корпуса текстов определенного автора. По критерию динамичности выделяют *динамические* и *статические* корпуса текстов. *Динамические* корпуса называют также мониторинговыми или мониторинговыми [1, с. 18]. Они используются в исследовании функционирования языковых явлений на временной шкале – например, изменения значения слов, частоты использования тех или иных конструкций и т. д. *Статические* же корпуса охватывают какой-то определенной промежуток времени. К ним относятся и авторские корпуса – сборники текстов писателя.

В рамках нашего исследования целесообразно использовать сбалансированные корпуса, т. к. корпуса такого типа применяются для решения специализированных задач и не имеют универсальных жестких требований к объему [3, с. 204]. В рамках решения задачи по моделированию пространственной характеристики произведения художественной литературы представляется целесообразным создать *письменный одноязычный литературный художественный статический неразмеченный полнотекстовый синхронический сбалансированный (исследовательский)* корпус текстов, который включает тексты произведений одного писателя. При построении корпуса и разработки алгоритма автоматизированного анализа массива текстовых данных были отобраны англоязычные тексты рассказов американского писателя Дж. Лондона из цикла рассказов “Smoke Bellew” [11]. Данный сборник рассказов издан в 1912 г. и включает два сборника рассказов “Smoke Bellew” и “Smoke and Shorty”. На основании наибольшей целесообразности и релевантности нами были выбраны четыре рассказа, которые и были включены в создаваемый корпус: “The Stampede to Squaw Creek”, “The Man on the Other Bank”, “The Race for Number Three”, “The Townsite of Tra-Lee”.

Художественное пространство как объект данного исследования определяется как «модель мира данного автора, выраженная на языке его представления» [6, с. 252]. Пространство реализуется в тексте на различных языковых уровнях, а пространственные отношения формально характеризуются лексико-синтаксическими единицами. В нашей работе внимание акцентируется на лексическом уровне, который включает в себя пространственную лексику и специфические морфологические формы с локальным значением.

Восприятие категории пространства сводится к распределению единиц по следующим признакам:

- обобщенное/относительно ограниченное пространство (предусматривает наличие определенных границ, статус части от целого);
- открытое/замкнутое пространство;
- природное/искусственное пространство.

Анализ отобранного материала позволил распределить лексические единицы, относящиеся к категории пространства (см. табл. 1 на с. 110).

Многие лексические единицы представлены в нескольких оппозициях. Так, например, одной из наиболее частотных локаций является имя существительное *cabin*, которое характеризуется как относительно ограниченное открытое искусственное пространство, а единица *world* представлена только в подкатегории обобщенности.

Таблица 1

Распределение лексических единиц с признаками категории пространства

Обобщенное	Признак				
	Относительно ограниченное	Открытое	Замкнутое	Природное	Искусственное
world	cabin	hummock	saloon	hummock	saloon
everywhere	market	mountains	cabin	mountains	cabin
snow	mountains	canyon	hospital	canyon	hospitals
water	canyon	tributary	shelter	tributary	market
	tributary	river	room	river	canal
	river	sea	house	sea	room
	sea	trail		stream	house
	trail	country		trail	
	hummock	bedrock		bedrock	
	country	creek		creek	
	forest	claim		claim	
	valleys	forest		forest	
	shelter	water		water	
	woods	valley		valleys	
	bank	woods		woods	
	room	bank		bank	
	house				
	casino				

В процессе анализа были выделены 22 лексические единицы, представляющие собой уникальные локации – топонимы, которые, в первую очередь, выполняют назывную и выделительные функции, служат наименованиями географических пунктов, а также вызывают в сознании читателя исторические, социальные и культурные коннотации топонимов [9]. Можно заметить, что выбранные рассказы Дж. Лондона характеризуются значительным количеством реально существующих географических пунктов, расположенных в Северной Америке, которые включают в себя как территориальные, так и природные объекты (заливы, горные отвод, водные объекты). Автор детально описывает художественное пространство, используя следующие локации:

**Топонимы:** Dawson (небольшой город в Канаде), Yukon (название территории в Канаде), Elkhorn (город, США), Dyea (город, США, Аляска), Henderson (город, США), Tra-Lee, the camp of Two Cabins; Adam Creek, Squaw Creek, Swede Creek, Norway Creek, Mono Creek; the Discovery claim; the Indian River, Klondike streams, the Koyokuk, Surprise Lake, the Stewart River, the McQuestion River, the Yukon Basin, the Sea Lion, the Panama Canal.

Было принято решение также выделить шесть уникальных локаций, которые не связаны с непосредственными действиями, но встречаются в контексте воспоминаний: the Birch Creek, the Forty Mile, Tivoli, beach at Dyea, the Box Canyon and White Horse rapids, San Francisco.

В рамках проведенного исследования инструментом автоматического анализа частотности лексических единиц категории пространства является экспериментальная модель реляционной базы данных для оперирования сбалансированным лингвистическим корпусом художественного произведения, разработанная в лаборатории фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета [4].

В избранных рассказах Дж. Лондона программным способом выделены 3530 токенов, что включает в себя как лексические единицы, так и знаки препинания. Первые позиции списка занимают неко-

торые знаки препинания (точки, запяты, кавычки, из чего можно сделать вывод о широком использовании диалогической речи), артикли, предлоги и местоимения, что характерно для английского языка.

Полученные статистические данные позволяют проанализировать распределение частотности лексических единиц по элементам оппозиции (см. табл. 2, 3, 4 на с. 112):

Таблица 2

**Частотность лексических единиц оппозиции «обобщенное/относительно ограниченное»**

Обобщенное		Относительно ограниченное	
Лексическая единица	Частотность	Лексическая единица	Частотность
world	3	cabin	35
everywhere	1	market	1
snow	77	mountains	6
water	32	canyon	5
		tributary	1
		river	24
		sea	6
		trail	61
		hummock	2
		country	21
		forest	1
		valley	8
		shelter	2
		woods	4
		bank	55
		room	10
		house	1
Всего	113	Всего	243

Таблица 3

**Частотность лексических единиц оппозиции «открытое/замкнутое»**

Открытое		Замкнутое	
Лексическая единица	Частотность	Лексическая единица	Частотность
hummock	2	saloon	5
mountains	6	cabin	35
canyon	5	hospital	6
tributary	1	shelter	2
river	24	room	10
sea	6	house	1
trail	61		
country	21		
bedrock	3		
creek	63		
claim	46		
forest	1		
water	32		
valley	4		
woods	8		
bank	55		
Всего	338	Всего	59

Таблица 4

Частотность лексических единиц оппозиции «природное/искусственное»

Природное		Искусственное	
Лексическая единица	Частотность	Лексическая единица	Частотность
hummock	2	saloon	5
mountains	6	cabin	35
canyon	5	hospitals	6
tributary	1	market	1
river	24	canal	2
sea	6	room	10
trail	61	house	1
country	21		
bedrock	3		
creek	63		
claim	46		
forest	1		
water	32		
valley	4		
woods	8		
bank	55		
Всего	338	Всего	60

Таким образом, можно увидеть, что наиболее частотным словоупотреблением из уже отобранных ранее единиц является имя существительное *snow* (77 употреблений), которое может как относиться к обобщенной категории пространства, так и не иметь никакого отношения к локациям или являться глаголом. Преобладающей группой являются природные открытые пространства с общим количеством словоупотреблений 338, включающие в себя самые частотные единицы *trail* (61), *bank* (55), *claim* (46), *creek* (35), *river* (24).

Из числа топонимов выделяются такие единицы, как *Dawson* (60), *Yukon* (20) и *Klondike* (18) – все они относятся к одному географическому пространству.

Результаты проведенного исследования иллюстрируют факт активного поиска программных решений для автоматизации интерпретации художественных произведений и разработки инструментария, способствующего наиболее эффективному решению лингвистических задач. Статистические параметры позволяют сделать выводы о наиболее значимых локациях текста. В перспективе планируется составить семантические ряды, которые, при помощи модели базы данных, установят точное соотношение элементов указанных оппозиций между собой; предполагается провести анализ их качественных характеристик.

### Литература

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. 2-е изд., перераб. и дополн. СПб.: СПбГУ, 2013.
2. Гальперин И. Р. Текст как объект лингвистического исследования. 4-е изд. М.: КомКнига, 2006.
3. Горожанов А.И., Степанова Д.В. Интерпретация художественного произведения: корпусный подход // Филологические науки. Вопросы теории и практики. 2022. Т. 15. № 1. С. 203–208.
4. Горожанов А.И. Экспериментальное моделирование базы данных сбалансированного лингвистического корпуса // Филологические науки. Вопросы теории и практики. 2022. Т. 15. № 10. С. 3382–3386.
5. Крапивин Ю.Б., Степанова Д.В. Решение задачи автоматического перевода научно-технических текстов с английского языка на белорусский // Вестник Брестского гос. технич. ун-та. 2021. № 1(124). С. 61–65.

6. Лотман Ю. М. Сюжетное пространство русского романа XIX столетия // В школе поэтического слова. Пушкин. Лермонтов. Гоголь. М.: Просвещение, 1988. С. 235–249.
7. Махаева А. К. Текст как единица общения и обучения опосредованному межкультурному общению // Молодой ученый. 2010. Т. 2. № 1-2(13). С. 88–92. [Электронный ресурс]. URL: <https://moluch.ru/archive/13/1102/> (дата обращения: 15.02.2023).
8. Москальская О. И. Грамматика текста. М.: Высш. школа, 1981.
9. Пивоварова О. П. Вербализация категории пространства в семантическом и лингвокультурологическом аспектах: на материале лексем и фразеологизмов русского языка с пространственным значением: дисс. ... канд. филол. наук. Челябинск, 2005.
10. Gorozhanov A.I., Guseynova I.A. Corpus Analysis of the Grammatical Categories' Constituents in Fiction Texts Considering the Linguo-Regional Component // Journal of Siberian Federal University. Humanities and Social Sciences. 2020. Vol. 13. No. 12. P. 2035–2048.
11. London J. Smoke Bellew. NY: Merchant Books, 2009.

**ANASTASIYA YANOVSKAYA**  
*Moscow State Linguistic University*

**THE CORPUS APPROACH TO THE ANALYSIS OF THE VERBALIZATION  
OF THE SPACE CATEGORY IN THE FICTIONAL WORK**

*The aim of the study is to analyze the peculiarities of expressing the space category in the fictional work at the material of the circle of the stories "Smoke Bellew" by Jack London at the source language. There are revealed the basic traits of the lexical units with the space meaning. The author uses the programming tool that allows to get the statistics of the chosen lexical units and to interpret the fictional works on its basis from the perspective of the verbalization of the space category.*

*Key words: corpus approach, balanced corpus, fictional work, space, English language, Jack London.*