

КОРПУСНЫЙ ПОДХОД К ОПРЕДЕЛЕНИЮ СРЕДСТВ ВЫРАЖЕНИЯ ПРОСТРАНСТВЕННО-КАЧЕСТВЕННЫХ ХАРАКТЕРИСТИК ХУДОЖЕСТВЕННОГО ТЕКСТА

В настоящее время лингвистами при проведении корпусных исследований уделяется значительное внимание такой области знания как профессионально ориентированное программирование, которое позволяет оперативно решать широкий круг задач по автоматической и автоматизированной обработке текстов.

Целью проводимого исследования является определение метода моделирования пространственно-качественной характеристики художественного произведения, основанного на корпусе текстов. Основной теоретической базой исследования послужила монография Алексея Ивановича Горожанова «Прикладные аспекты анализа и интерпретации текстов».

На первом этапе исследования был определен тип исследуемого корпуса – письменный одноязычный литературный художественный исследовательский статический неразмеченный полнотекстовый синхронический корпус. В разрабатываемый корпус текстов были включены тексты рассказов из сборника Дж. Лондона «Smoke Bellew»(1912) на языке оригинала. На последующем этапе был проведен автоматизированный анализ частотности употреблений всех лексических единиц как во всем избранном художественном произведении, так и в его отдельных главах (рассказах). Для реализации был выбран язык программирования Python, так как он считается наиболее универсальным, популярным, гибким и доступным для понимания. Результа-

том работы программы является вывод 5358 уникальных словоупотреблений в порядке убывания частотности употребления в тексте. Следующим этапом проведения исследования выступает ручной отбор единиц, обозначающих и характеризующих сюжетные локации (существительные, прилагательные) и присвоение им категории «положительные», «отрицательные» и «нейтральные». С помощью программы конкорданс был определен контекст наиболее частотных локаций. Так, к примеру, наиболее частотным словоупотреблением, которое в большинстве случаев относится к категории пространства, является единица *boat*. Анализ ее основных характеристик позволил сделать вывод о преобладании положительных категорий (*dinky, good-lookin'*) над нейтральными (*afloat, home-made, lightly loaded*).

На этапе интерпретации полученных данных также были модифицированы и оптимизированы применяемые программные инструменты, минимизирована погрешность и проведен поиск альтернативных путей и программных способов интерпретации художественного текста. В перспективе планируется провести данный анализ с использованием библиотеки для расширенной обработки естественного языка spaCy.