Г. А. Дырхеева (Улан-Удэ)
an5dag1@mail.ru

О БОГАТСТВЕ ХУДОЖЕСТВЕННОГО ТЕКСТА: ЛИНГВОСТАТИСТИЧЕСКИЙ АНАЛИЗ НА ПРИМЕРЕ БУРЯТСКОГО ХУДОЖЕСТВЕННОГО ТЕКСТА

Язык писателя и авторский стиль всегда привлекали внимание лингвистов, поскольку изучение индивидуальных стилей писателей помогает глубже понять особенности художественной речи как функционального стиля. В подобных исследованиях одной из важных задач является определение множества лингвистических признаков, характерных и стабильных параметров текста, в которых отражаются стилевые особенности языка автора. Одним из путей решения этой задачи является статистический анализ лингвистических категорий в тексте.

В рамках указанной задачи нами проведен лингвостатистический анализ произведений классика бурятской литературы Х. Намсараева. Общий объем текста составил 271866 словоупотреблений. В результате получены частотные списки слов (11769 единиц), словоформ (36984 единицы), парных слов[1], фразеологизмов, русизмов.

Первоначально были определены квантитативно-типологические характеристики материала и проведено его сравнение с материалом других языков. Так, анализ выявил ряд статистических оценок в качестве типологических критериев: 1) статистическая покрываемость по зонам частотного словаря словоформ: в бурятском тексте 100 первых словоформ покрывают 23–28 % текста (во флективно-аналитических – 43–54 %); 2) средняя повторяемость словоформ по произведениям – 1,65–5,03, в общем корпусе – 7,35 (во флективно-аналитических – 9,8–18,04); 3) средняя длина словоформы в тексте 6,29 букв, в словаре – 8,59, что значительно больше средней длины словоформ в индоевропейских языках.

Типологической характеристикой языка могут быть слова высокочастотной зоны. Выявлено, что в бурятском языке в данную группу входят служебные и полнозначные стилистически нейтральные слова. При этом высока доля глагольных форм (из 200 самых частых слов – 23,5 %, которые покрывают 20 % текста), т. к. они могут использоваться в качестве служебных слов и вспомогательных глаголов в аналитических конструкциях. Существительные данной зоны образуют тематическую группу слов, в значениях которых проявляется доминирование компонентов человеческого микромира и его окружения. Слова высокочастотной зоны имеют удовлетворительную корреляцию в среднем до 25 ранга.

[1] Парные слова – бессоюзные сочетания слов, имеющие промежуточный статус между словосложением и сочинением, являются реальной особенностью, характеризующей монгольские языки: новое слово образуется склейкой двух слов с близким звучанием, одинаковой морфологией и соотнесенными значениями.

Статистическое моделирование помогает изучить факты, труднодоступные для прямого лингвистического наблюдения. Так, наиболее низкие показатели выявлены у повестей *«Алтан зэбэ»* и *«Эдиршууд»*. Процент покрываемого текста словами высокочастотной зоны у них больше, а значения коэффициента лексического разнообразия ниже. Подтверждением «бедности» этих произведений являются следующие факты: 1) значения К (концентрация словаря) в них выше, 2) значения коэффициента Сомерса, наоборот, ниже, 3) интенсивность использования знаменательных частей речи выше (в повести *«Эдиршууд»* в сравнении с повестью *«Цыремпил»* только числительные и частицы имеют более низкий индекс итерации), 4) параметр «однократные и низкочастотные слова» (или значения индекса исключительности) на уровне словаря также показывает преимущество рассказов *«Цыремпила»* и *«Нэгэтэ һуни»*, причем в последнем преобладают слова и словоформы с частотой менее 3, 5, использование парных слов, фразеологических выражений, а также пословиц и поговорок немногочисленно и однообразно. Все указанные индексы и коэффициенты взаимосвязаны, показывают зависимость от объема текста, поэтому результаты по текстам разного объема часто трудно сопоставлять.

При лингвистическом анализе важно рассмотреть распределение лексических элементов по частям речи. Выявлено, что слова, несущие основную смысловую нагрузку, составляют 88,1 % текста и 94,03 % словаря, распределение слов и словоформ по частям речи является стабильным для всей прозы Х. Намсараева, наиболее употребительными и стабильными частями речи, как и в других языках, являются существительные и глаголы, вычисление индекса итерации подтвердило мнение об описательности прозы.

Распределение частей речи по частотным зонам словаря показало, что знаменательные слова в основном распределены равномерно. Прирост словаря происходит за счет существительных, причастий, деепричастий и прилагательных. Служебные слова, а также местоимения и числительные показывают постепенное понижение их доли от высоких частот к низким. Подтвердилось мнение, что параметры «парные слова» и «фразеологические выражения» могут быть использованы в качестве характеристики лексического «богатства». Однако данные параметры и параметр «русские заимствования» не подтвердили гипотезу о временной градации произведений.

Лингвостатистический анализ является основой для дальнейшего вероятностно-статистического моделирования, которое предполагает: 1) определение основных информационных свойств естественного языка (энтропия, избыточность, структура слова, вес агглютинативной, флективной и аналитической морфологии), 2) построение базового языка, включающего наиболее важные единицы естественного языка. Подобный анализ позволяет теоретически сопоставить материал по различным языкам (флективно-аналитическим и агглютинативным). Результаты могут быть использованы для решения задачи о вероятностной природе «нормы» языка, для нормализации литературного языка, исследований по культуре речи, истории изменения и формирования бурятской лексики, в методике преподавания бурятского языка.